

# Almost Optimal Construction of Functional Batch Codes Using Extended Simplex Codes

Lev Yohananov<sup>1</sup>, Student Member, IEEE, and Eitan Yaakobi<sup>2</sup>, Senior Member, IEEE

**Abstract**—A functional  $k$ -batch code of dimension  $s$  consists of  $n$  servers storing linear combinations of  $s$  linearly independent information bits. Any multiset request of size  $k$  of linear combinations (or requests) of the information bits can be recovered by  $k$  disjoint subsets of the servers. The goal under this paradigm is to find the minimum number of servers for given values of  $s$  and  $k$ . A recent conjecture states that for any  $k = 2^s - 1$  requests the optimal solution requires  $2^s - 1$  servers. This conjecture is verified for  $s \leq 5$  but previous work could only show that codes with  $n = 2^s - 1$  servers can support a solution for  $k = 2^{s-2} + 2^{s-4} + \lfloor \frac{2^{s/2}}{\sqrt{24}} \rfloor$  requests. This paper reduces this gap and shows the existence of codes for  $k = \lfloor \frac{5}{8} 2^{s-1} \rfloor - s$  requests with the same number of servers. Another construction in the paper provides a code with  $n = 2^{s+1} - 2$  servers and  $k = 2^s$  requests, which is an optimal result. These constructions are mainly based on extended Simplex codes and equivalently provide constructions for parallel Random I/O (RIO) codes.

**Index Terms**—Batch codes, Simplex codes, codes with availability, Private Information Retrieval (PIR).

## I. INTRODUCTION

MOTIVATED by several applications for load-balancing in storage and cryptographic protocols, *batch codes* were first proposed by Ishai *et al.* [7]. A batch code encodes a length- $s$  string  $x$  into  $n$  strings, where each string corresponds to a *server*, such that each batch request of  $k$  different bits (and more generally symbols) from  $x$  can be decoded by reading at most  $t$  bits from every server. This decoding process corresponds to the case of a single-user. There is an extended variant for batch codes [7] which is intended for a multi-user application instead of a single-user setting, known as the *multiset batch codes*. Such codes have  $k$  different users and each requests a single data item. Thus, the  $k$  requests can be represented as a *multiset* of the bits since the requests of different users may be the same, and each server can be accessed by at most one user.

Manuscript received 9 September 2021; revised 20 December 2021; accepted 1 March 2022. Date of publication 24 March 2022; date of current version 15 September 2022. This work was supported in part by Israel Science Foundation (ISF) under Grant 1817/18 and in part by the Technion Hiroshi Fujiwara Cyber Security Research Center and the Israel Cyber Directorate. An earlier version of this paper was presented in part at the 2021 International Symposium on Information Theory [DOI: 10.1109/ISIT45174.2021.9518215]. (Corresponding author: Lev Yohananov.)

The authors are with the Department of Computer Science, Technion—Israel Institute of Technology, Haifa 3200003, Israel (e-mail: levyohananov@cs.technion.ac.il; yaakobi@cs.technion.ac.il).

Communicated by G. Matthews, Associate Editor for Coding Theory.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TIT.2022.3162158>.

Digital Object Identifier 10.1109/TIT.2022.3162158

A special case of multiset batch codes, referred as *primitive batch codes*, is when each server contains only one bit. The goal of this model is to find, for given  $s$  and  $k$ , the smallest  $n$  such that a primitive batch code exists. This problem was considered in several papers; see e.g. [1], [2], [7], [8], [13]. By setting the requests to be a multiset of linear combinations of the  $s$  information bits, a batch code is generalized into a *functional batch code* [18]. Again, given  $s$  and  $k$ , the goal is to find the smallest  $n$  for which a functional  $k$ -batch code exists.

Mathematically speaking, an  $FB(n, s, k)$  functional  $k$ -batch code (and in short  $FB(n, s, k)$  code) of dimension  $s$  consists of  $n$  servers storing linear combinations of  $s$  linearly independent information bits. Any multiset of size  $k$  of linear combinations from the linearly independent information bits, can be recovered by  $k$  disjoint subsets of servers. If all the  $k$  linear combinations are the same, then the servers form an  $FP(n, s, k)$  functional  $k$ -Private Information Retrieval (PIR) code (and in short  $FP(n, s, k)$  code). Clearly, an  $FP(n, s, k)$  code is a special case of an  $FB(n, s, k)$  code. It was shown that functional  $k$ -batch codes are equivalent to the so-called linear *parallel random I/O (RIO) codes*, where RIO codes were introduced by Sharon and Alrod [10], and their parallel variation was studied in [11], [12]. Therefore, all the results for functional  $k$ -batch codes of this paper hold also for parallel RIO codes. If all the  $k$  linear combinations are of a single information bit (rather than linear combinations of information bits), then the servers form an  $B(n, s, k)$   $k$ -batch code (and in short  $B(n, s, k)$  code).

The value  $FP(s, k)$ ,  $B(s, k)$ ,  $FB(s, k)$  is defined to be the minimum number of servers required for the existence of an  $FP(n, s, k)$ ,  $B(n, s, k)$ ,  $FB(n, s, k)$  code, respectively. Several upper and lower bounds can be found in [18] on these values. Wang *et al.* [14] showed that for  $k = 2^{s-1}$ , the length of an optimal  $k$ -batch code is  $2^s - 1$ , that is,  $B(s, k = 2^{s-1}) = 2^s - 1$ . They also showed a recursive decoding algorithm. It was conjectured in [18] that for the same value of  $k$ , the length of an optimal functional batch code is  $2^s - 1$ , that is,  $FB(s, k = 2^{s-1}) = 2^s - 1$ . Indeed, in [16] this conjecture was proven for  $s = 3, 4$ , and in [18], by using a computer search, it was verified also for  $s = 5$ . However, the best-known result for  $s > 5$  only provides a construction of  $FB(2^s - 1, s, 2^{s-2} + 2^{s-4} + \lfloor \frac{2^{s/2}}{\sqrt{24}} \rfloor)$  codes [18]. This paper significantly improves this result and reduces the gap between the conjecture statement and the best-known construction. In particular, a construction of  $FB(2^s - 1, s, \lfloor \frac{5}{8} \cdot 2^{s-1} \rfloor - s)$  codes is given. To obtain this important result, we first show

an existence of  $FB-(2^s - 1, s, \lfloor \frac{3}{4} \cdot 2^{s-1} \rfloor)$  code. Moreover, we show how to construct  $FB-(2^s + \lceil (3\alpha - 2) \cdot 2^{s-2} \rceil - 1, s, \lfloor \alpha \cdot 2^{s-1} \rfloor)$  codes for all  $2/3 \leq \alpha \leq 1$ . Another result that can be found in [18] states that  $FP(s, 2^s) \leq 2^{s+1} - 2$ . In this case, the lower bound is the same, i.e., this result is optimal, see [5]. In this paper we will show that this optimality holds not only for functional PIR codes but also for the more challenging case of functional batch codes, that is,  $FB(s, 2^s) = 2^{s+1} - 2$ . Lastly, we show a non-recursive decoding algorithm for  $B-(2^s, s, k = 2^{s-1})$  codes. In fact, this construction holds not only for  $k$  single bit requests (with respect to  $k$ -batch codes) but also for  $k$  linear combinations of requests under some constraint that will be explained in the paper. All the results in the paper are achieved using a generator matrix  $G$  of an extended Simplex code [3] of length  $2^s$  and dimension  $s$ , where the matrix's columns correspond to the servers of the  $FB-(n, s, k)$  code.

There are three reasons we consider the conjectured case  $k = 2^s - 1$  for functional  $k$ -batch codes. First, this case has already been solved for PIR codes, functional PIR codes, and  $k$ -batch codes. Second, it provides a construction for parallel Random I/O (RIO) codes. Third, this is the case in which every linear combination of the  $s$  information bits is stored in one of the servers.

The rest of the paper is organized as follows. In Section II, we formally define functional  $k$ -batch codes and summarize the main results of the paper. In Section III, we show a construction of  $FB-(2^s + \lceil (3\alpha - 2) \cdot 2^{s-2} \rceil - 1, s, \lfloor \alpha \cdot 2^{s-1} \rfloor)$  for  $\alpha = 2/3$ . This result is extended for all  $2/3 \leq \alpha \leq 1$  in Section IV. In Section V, a construction of  $FB-(2^{s+1} - 2, s, 2^s)$  is presented. In Section VI, we present our main result, i.e., a construction of  $FB-(2^s - 1, s, \lfloor \frac{5}{6} \cdot 2^{s-1} \rfloor - s)$  codes. In Section VII a construction of  $B-(2^s - 1, s, 2^{s-1})$  is presented. Finally, Section VIII concludes the paper.

## II. DEFINITIONS

For a positive integer  $n$  define  $[n] = \{0, 1, \dots, n - 1\}$ . All vectors and matrices in the paper are over  $\mathbb{F}_2$ . We follow the definition of functional batch codes as it was first defined in [18].

**Definition 1:** A **functional  $k$ -batch code** of length  $n$  and dimension  $s$  consists of  $n$  servers and  $s$  information bits  $x_0, x_1, \dots, x_{s-1}$ . Each server stores a nontrivial linear combination of the information bits (which are the coded bits), i.e., for all  $j \in [n]$  exists  $\ell_j$  such that the  $j$ -th server stores a linear combination

$$y_j = x_{i_0} + x_{i_1} + \dots + x_{i_{\ell_j-1}},$$

such that  $i_0, i_1, \dots, i_{\ell_j-1} \in [s]$ . For any request of  $k$  linear bit combinations  $v_0, v_1, \dots, v_{k-1}$  (not necessarily distinct) of the information bits, there are  $k$  pairwise disjoint subsets  $R_0, R_1, \dots, R_{k-1}$  of  $[n]$  such that the sum of the linear combinations in the related servers of  $R_i, i \in [k]$ , is  $v_i$ , i.e.,

$$\sum_{j \in R_i} y_j = v_i.$$

Each such  $v_i$  will be called a **requested bit** and each such subset  $R_i$  will be called a **recovery set**.

Equivalently, a functional  $k$ -batch code is a **linear code** with an  $s \times n$  generator matrix

$$G = [g_0, g_1, \dots, g_{n-1}]$$

in  $\mathbb{F}_2^{s \times n}$  in which the vector  $g_j$  has ones in positions  $i_0, i_1, \dots, i_{\ell_j-1}$  if and only if the  $j$ -th server stores the linear combination  $x_{i_0} + x_{i_1} + \dots + x_{i_{\ell_j-1}}$ . Using this matrix representation, a functional  $k$ -batch code is an  $s \times n$  generator matrix  $G$ , such that for any  $k$  request vectors  $v_0, v_1, \dots, v_{k-1} \in \mathbb{F}_2^s$  (not necessarily distinct), there are  $k$  pairwise disjoint subsets of columns in  $G$ , denoted by  $R_0, R_1, \dots, R_{k-1}$ , such that the sum of the column vectors whose indices are in  $R_j$  is equal to the **request vector**  $v_j$ . The set of all recovery sets  $R_i, i \in [k]$ , is called a **solution** for the  $k$  request vectors. The sum of the column vectors whose indices are in  $R_j$  will be called the **recovery sum**.

A functional  $k$ -batch code of length  $n$  and dimension  $s$  over  $\mathbb{F}_2^s$  is denoted by  $FB-(n, s, k)$ . Every request of  $k$  vectors will be stored as columns in a matrix  $M$  which is called the **request matrix** or simply the **request**.

In Definition 1, we showed both scalar and vector definitions for  $k$ -batch codes, which are equivalent to each other. While the scalar definition in previous works is the more common one, the vector definition is used mainly in this work because of the unique properties of the generator matrices of these codes.

A  **$k$ -batch code** of length  $n$  and dimension  $s$  over  $\mathbb{F}_2^s$ , is denoted by  $B-(n, s, k)$  and is defined similarly to functional  $k$ -batch codes as in Definition 1 except of the fact that each request vector  $v_j \in \mathbb{F}_2^s$  is a unit vector. A **functional  $k$ -PIR code** [18] of length  $n$  and dimension  $s$ , denoted by  $FP-(n, s, k)$ , is a special case of  $FB-(n, s, k)$  in which all the request vectors are identical. We first show some preliminary results on the parameters of  $FB-(n, s, k)$  and  $FP-(n, s, k)$  codes which are relevant to our work. For that, another definition is presented.

**Definition 2:** Denote by  $FB(s, k), B(s, k), FP(s, k)$  the minimum length  $n$  of any  $FB-(n, s, k), B-(n, s, k), FP-(n, s, k)$  code, respectively.

Most of the following results on  $FB(s, k), B(s, k)$  and  $FP(s, k)$  can be found in [18], while the result in (c) was verified for  $s = 3, 4$  in [16].

**Theorem 3:** For positive integers  $s$  and  $t$ , the following properties hold: (a)

- 1)  $FP(s, 2^{s-1}) = 2^s - 1$ .
- 2)  $FP(st, 2^s) \leq 2t(2^s - 1)$ .
- 3) For  $s \leq 5$  it holds that  $FB(s, 2^{s-1}) = 2^s - 1$ .
- 4) An  $FB-(2^s - 1, s, 2^{s-2} + 2^{s-4} + \lfloor \frac{2^{s/2}}{\sqrt{24}} \rfloor)$  code exists.
- 5) For a fixed  $k$  it holds that

$$\lim_{s \rightarrow \infty} \frac{FB(s, k)}{s} \geq \frac{k}{\log(k+1)}.$$

- 6)  $B(s, 2^{s-1}) = 2^s - 1$  [14].
- 7)  $B(s, k) = s + \Theta(\sqrt{s})$  for  $k = 3, 4, 5$  [1], [13].
- 8)  $B(s, k) = s + \mathcal{O}(\sqrt{s} \log s)$  for  $k > 6$  [13].

Note that the result from Theorem 3(d) improves upon the result of  $FB-(2^s - 1, s, 2^{s-2} + 2^{s-4} + 1)$  functional batch

codes which was derived from a WOM codes construction by Godlewski [6]. This is the best-known result concerning the number of queries when the number of information bits is  $s$  and the number of encoded bits is  $2^s - 1$ .

The goal of this paper is to improve some of the results summarized in Theorem 3. The result in (c) holds for  $s \leq 5$ , and it was conjectured in [18] that it holds for all positive values of  $s$ .

*Conjecture 1:* [18] For all  $s > 5$ ,  $FB(s, 2^{s-1}) = 2^s - 1$ .

The reader can notice the gap between Conjecture 1 and the result in Theorem 3(d). More precisely, [18] assures that an  $FB-(2^s - 1, s, 2^{s-2} + 2^{s-4} + \lfloor \frac{2^{s/2}}{\sqrt{24}} \rfloor)$  code exists, and the goal is to determine whether an  $FB-(2^s - 1, s, 2^{s-1})$  code exists. This paper takes one more step in establishing this conjecture. Specifically, the best-known value of the number of requested bits  $k$  is improved for the case of  $s$  information bits and  $2^s - 1$  encoded bits. The next theorem summarizes the contributions of this paper.

*Theorem 4:* For a positive integer  $s$ , the following constructions exist: (a)

- 1) A construction of  $FB-(2^s - 1, s, \lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor)$  codes.
- 2) A construction of

$$FB-(2^s + \lceil (3\alpha - 2) \cdot 2^{s-2} \rceil - 1, s, \lfloor \alpha \cdot 2^{s-1} \rfloor)$$

codes where  $2/3 \leq \alpha \leq 1$ .

- 3) A construction of  $FB-(2^{s+1} - 2, s, 2^s)$  codes.
- 4) A construction of  $FB-(2^s - 1, s, \lfloor \frac{5}{6} \cdot 2^{s-1} \rfloor - s)$  codes.

We now explain the improvements of the results of Theorem 4. The construction in Theorem 4(a) improves upon the result from Theorem 3(d), where the supported number of requests increases from  $\frac{1}{2}2^{s-1} + 2^{s-4} + \lfloor \frac{2^{s/2}}{\sqrt{24}} \rfloor$  to  $\lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor$ . Note that by taking  $\alpha = 2/3$  in the result of Theorem 4(b), we immediately get the result of (a). However, for simplicity of the proof, we first show the construction for (a) separately, and afterwards, add its extension. The result of Theorem 4(d) is based on the result of Theorem 4(a) and improves it to  $\lfloor \frac{5}{6} \cdot 2^{s-1} \rfloor - s$  requests. Moreover, according to the second result of Theorem 3(b) if  $t = 1$  then  $FP(s, 2^s) \leq 2^{s+1} - 2$ . Based on the result in [5] it holds that  $FP(s, 2^s) \geq 2^{s+1} - 2$ . Therefore,  $FP(s, 2^s) = 2^{s+1} - 2$ . The construction in Theorem 4(c) extends this result to functional batch codes by showing that  $FB(s, 2^s) \leq 2^{s+1} - 2$ , and again, combining the result from [5], it is deduced that  $FB(s, 2^s) = 2^{s+1} - 2$ .

A special family of matrices that will be used extensively in the paper are the generator matrices of extended Simplex codes [3], as defined next.

*Definition 5:* A matrix  $G = [g_0, g_1, \dots, g_{2^s-1}]$  of order  $s \times 2^s$  over  $\mathbb{F}_2$  such that  $\{g_0, g_1, \dots, g_{2^s-1}\} = \mathbb{F}_2^s$  is called an *extended Simplex generator matrix* and in short *ES-matrix*.

We will use *ES*-matrices as the generator matrices of the linear codes that will provide the constructions used in establishing Theorem 4. More specifically, given a linear code defined by a generator *ES*-matrix  $G$  of order  $s \times n$  and a request  $M$  of order  $s \times k$ , we will show an algorithm that finds a solution for  $M$ . This solution will be obtained by rearranging the columns of  $G$  and thereby generating a new *ES*-matrix  $G'$ . This solution is obtained by showing all

TABLE I  
TABLE OF DEFINITIONS AND NOTATIONS

Notation	Meaning	Remarks
$FB-(n, s, k)$	A func. $k$ -batch code of length $n$ and dimension $s$	Sec. II
$B-(n, s, k)$	A $k$ -batch code of length $n$ and dimension $s$	Sec. II
$R_i$	The $i$ -th recovery set	Sec. II
$(\mathcal{G}, \mathcal{B}, \mathcal{R})$	A triple-set	Def. 7
$\mathcal{M}(\mathcal{G}, \mathcal{B}, \mathcal{R})$	A triple-matrix of $M$	Def. 7
$e$	A unit vector of length $s$ with 1 at its last index	Sec. III
$M$	A request matrix	Sec. III
$v_i, w_i$	The $i$ -th request/column vector in $M, \mathcal{M}$	Sec. III
$G$	An <i>ES</i> -matrix	Sec. III
$g_i$	A column vector in $G$ representing the $i$ -th server	Sec. III
$G_x(G)$	An $x$ -type graph of $G$	Def. 10
$C_x(G)$	The partition of simple cycles of $G_x(G)$	Def. 10
$P_x(g_i, g_j)$	A simple path between $g_i$ and $g_j$ in $G_x(G)$	Def. 12
$d_{P_x}(g_i, g_m)$	The sub-length from $g_i$ to $g_m$ in $P_x(g_i, g_j)$	Def. 12
$\mathcal{F}_x(g_i, g_j)$	A reordering function for a good-path $P_x(g_i, g_j)$	Def. 12

the disjoint recovery sets for the request  $M$ , with respect to indices of columns of  $G'$ . Although such a solution is obtained with respect to  $G'$  instead of  $G$ , it can be easily adjusted to  $G$  by relabeling the indices of the columns. Thus, any *ES*-matrix whose column indices are partitioned to recovery sets for  $M$  provides a solution. Note that *ES*-matrices store the all-zero column vector. Such a vector will help us to simplify the construction of the algorithm and will be removed at the end of the algorithm.

*Definition 6:* Let  $M = [v_0, v_1, \dots, v_{n/2-1}]$  be a request of order  $s \times n/2$ , where  $n = 2^s$ . The matrix  $M$  has an **extended Simplex solution** if there exists an *ES*-matrix  $G = [g_0, g_1, \dots, g_{n-1}]$  of order  $s \times n$  such that for all  $i \in [n/2]$ ,

$$v_i = g_{2i} + g_{2i+1}.$$

In this case, we say that  $G$  is an extended Simplex solution for  $M$ .

Next, an example is shown.

*Example 1:* For  $s = 3$ , let

$$G = \begin{pmatrix} g_0 & g_1 & g_2 & g_3 & g_4 & g_5 & g_6 & g_7 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

be an *ES*-matrix. Given a request,

$$M = \begin{pmatrix} v_0 & v_1 & v_2 & v_3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

an extended Simplex solution for this request may be

$$G' = \begin{pmatrix} g'_0 & g'_1 & g'_2 & g'_3 & g'_4 & g'_5 & g'_6 & g'_7 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Lastly, for the convenience of the reader, the relevant notations and terminology that will be used throughout the paper is summarized in Table I.

### III. A CONSTRUCTION OF $FB-(2^s - 1, s, \lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor)$ CODES

In this section a construction of  $FB-(2^s - 1, s, \lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor)$  codes is presented. Let the request  $M$  be denoted by

$$M = [v_0, v_1, \dots, v_{\lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor - 1}].$$

Let  $e = (0, 0, \dots, 0, 1) \in \mathbb{F}_2^s$  be the unit vector with 1 at its last index. The solution for the request  $M$  will be derived by using two algorithms as will be presented in this section. We start with several definitions and tools that will be used in these algorithms.

*Definition 7:* Three sets  $\mathcal{G}, \mathcal{B}, \mathcal{R} \subseteq [2^{s-1}]$  are called a **triple-set** (the good, the bad, and the redundant), and are denoted by  $(\mathcal{G}, \mathcal{B}, \mathcal{R})$ , if the following properties hold,

$$\begin{aligned} \mathcal{G} &\subseteq \left[ \left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor \right], \\ \mathcal{B} &= \left[ \left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor \right] \setminus \mathcal{G}, \\ \mathcal{R} &= [2^{s-1}] \setminus (\mathcal{G} \cup \mathcal{B} \cup \{2^{s-1} - 1\}). \end{aligned}$$

Given a matrix  $M = [v_0, v_1, \dots, v_{\lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor - 1}]$  of order  $s \times \lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor$ , the matrix  $\mathcal{M}(\mathcal{G}, \mathcal{B}, \mathcal{R}) = [w_0, w_1, \dots, w_{2^s - 1}]$  of order  $s \times 2^{s-1}$  is referred as a **triple-matrix of  $M$**  if it holds that

$$w_t = \begin{cases} v_t & t \in \mathcal{G} \\ v_t + e & t \in \mathcal{B} \\ e & t \in \mathcal{R} \end{cases}.$$

Note that, we did not demand anything about the vector  $w_{2^s - 1}$ , i.e., it can be any binary vector of length  $s$ . Furthermore, by Definition 7, the set  $\mathcal{B}$  uniquely defines the triple-set  $(\mathcal{G}, \mathcal{B}, \mathcal{R})$ . We proceed with the following claim.

*Claim 1:* For any triple-set  $(\mathcal{G}, \mathcal{B}, \mathcal{R})$  if  $|\mathcal{B}| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$  then  $|\mathcal{B}| \leq |\mathcal{R}|$ .

The proof of this claim is shown in Appendix A. As mentioned above, our strategy is to construct two algorithms. We start by describing the first one which is the main algorithm. This algorithm receives as an input the request  $M$  and outputs a set  $\mathcal{B}$  and an extended Simplex-solution for some triple-matrix  $\mathcal{M}(\mathcal{G}, \mathcal{B}, \mathcal{R})$  of  $M$ . Using the matrix  $\mathcal{M}(\mathcal{G}, \mathcal{B}, \mathcal{R})$ , it will be shown how to derive the solution for  $M$ . This connection is established in the next lemma. For the rest of this section we denote  $n = 2^s$  and for our ease of notations both of them will be used.

*Lemma 8:* If there is an extended Simplex solution for  $\mathcal{M}(\mathcal{G}, \mathcal{B}, \mathcal{R})$  such that  $|\mathcal{B}| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$ , then there is a solution for  $M = [v_0, v_1, \dots, v_{\lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor - 1}]$ .

*Proof:* Let the  $ES$ -matrix  $G = [g_0, g_1, \dots, g_{n-1}]$  be an extended Simplex solution for  $\mathcal{M}(\mathcal{G}, \mathcal{B}, \mathcal{R})$ . Our goal is to form all disjoint recovery sets  $R_t$  for  $t \in \mathcal{G} \cup \mathcal{B} = \left[ \left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor \right]$  for  $M$ . Since  $G$  is an extended Simplex solution for  $\mathcal{M}(\mathcal{G}, \mathcal{B}, \mathcal{R})$ , for all  $t \in [2^{s-1}]$ , it holds that

$$w_t = g_{2t} + g_{2t+1}.$$

By definition of  $\mathcal{M}(\mathcal{G}, \mathcal{B}, \mathcal{R})$

$$w_t = \begin{cases} v_t & t \in \mathcal{G} \\ v_t + e & t \in \mathcal{B} \\ e & t \in \mathcal{R} \end{cases}.$$

Thus, if  $t \in \mathcal{G}$  then

$$v_t = w_t = g_{2t} + g_{2t+1},$$

and each recovery set for  $v_t$  is of the form  $R_t = \{2t, 2t+1\}$ . If  $t \in \mathcal{B}$  then

$$v_t + e = w_t = g_{2t} + g_{2t+1},$$

and if  $t' \in \mathcal{R}$  then

$$e = w_{t'} = g_{2t'} + g_{2t'+1}.$$

Therefore, for all  $t \in \mathcal{B}$  and  $t' \in \mathcal{R}$ ,

$$v_t = g_{2t} + g_{2t+1} + g_{2t'} + g_{2t'+1}.$$

By Claim 1, since  $|\mathcal{B}| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$ , it holds that  $|\mathcal{B}| \leq |\mathcal{R}|$ . Thus, for all  $t \in \mathcal{B}$ , each recovery set  $R_t$  for  $v_t$  will have a different  $t' \in \mathcal{R}$  such that

$$R_t = \{2t, 2t+1, 2t', 2t'+1\}.$$

■

In Lemma 8, it was shown that obtaining  $\mathcal{M}(\mathcal{G}, \mathcal{B}, \mathcal{R})$  which holds  $|\mathcal{B}| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$  provides a solution for  $M$ . Therefore, if the first algorithm outputs a set  $\mathcal{B}$  for which  $|\mathcal{B}| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$ , then the solution for  $M$  is easily derived. Otherwise, the first algorithm outputs a set  $\mathcal{B}$  such that  $|\mathcal{B}| > \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$ . In this case, the second algorithm will be used in order to reduce the size of the set  $\mathcal{B}$  to be at most  $\lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$ . For that, more definitions are required, and will be presented in the next section.

#### A. Graph Definitions

In the two algorithms of the construction, we will use undirected graphs, simple paths, and simple cycles that will be defined next. These graphs will be useful to represent the  $ES$ -matrix  $G$  in some graph representation and to make some swap operations on its columns.

*Definition 9:* An **undirected graph** or simply a **graph** will be denoted by  $\mathbf{G} = (V, E)$ , where  $V = \{u_0, u_1, \dots, u_{m-1}\}$  is its set of  $m$  nodes (vertices) and  $E \subseteq \{\{u_i, u_j\} \mid u_i, u_j \in V\}$  is its edge set. A finite **simple path** of length  $\ell$  is a sequence of distinct edges  $e_0, e_1, \dots, e_{\ell-1}$  for which there is a sequence of vertices  $u_{i_0}, u_{i_1}, \dots, u_{i_\ell}$  such that  $e_j = \{u_{i_j}, u_{i_{j+1}}\}, j \in [\ell]$ . A **simple cycle** is a simple path in which  $u_{i_0} = u_{i_\ell}$ . The **degree** of a node  $u_i$  is the number of edges that are incident to the node, and will be denoted by  $\deg(u_i)$ .

Note that in Definition 9 we did not allow parallel edges, i.e., different edges which connect between the same two nodes. By a slight abuse of notation, we will use graphs in which at most 2 parallel edges are allowed between any two nodes. That implies that cycles of length 2 may appear in the graph. In this case, we will use some notations for

distinguishing between two parallel edges as will be done in the following definition.

*Definition 10:* Given an  $ES$ -matrix  $G = [g_0, g_1, \dots, g_{n-1}]$ , and a non-zero vector  $x \in \mathbb{F}_2^s$ , denote the  $x$ -**type graph**  $\mathbf{G}_x(G) = (V, E_x(G))$  of  $G$  and  $x$  such that  $V = \mathbb{F}_2^s$  and a multi-set

$$E_x(G) = \left\{ \{g_i, g_i + x\} \mid i \in [n] \right\} \cup \left\{ \{g_{2t-1}, g_{2t}\} \mid t \in [n/2] \right\}.$$

For all  $t \in [n/2]$ , we say that  $g_{2t-1}$  and  $g_{2t}$  are a **pair**. An edge  $\{g_{2t-1}, g_{2t}\}$  will be called a **pair-type edge** and will be denoted by  $\{g_{2t-1}, g_{2t}\}_P$ . An edge  $\{g_i, g_i + x\}$  will be called an  **$x$ -type edge** and will be denoted by  $\{g_i, g_i + x\}_x$ . Note that for any  $g \in V$ , it holds that  $\deg(g) = 2$ . Thus, the graph  $\mathbf{G}_x(G)$  has a partition of  $\ell \geq 1$  disjoint simple cycles, that will be denoted by  $\mathbf{C}_x(G) = \{C_i\}_{i=0}^{\ell-1}$ , where every  $C_i$  is denoted by its set of edges.

For the rest of the paper, when the graph  $\mathbf{G}_x(G)$  is considered, it is assumed that  $x \neq 0$ . Note that  $E_x(G)$  is a multi-set since in case that  $g_{2t-1} = g_{2t} + x$ , we have two parallel edges  $\{g_{2t-1}, g_{2t}\}_P$  and  $\{g_{2t-1}, g_{2t}\}_x$  between  $g_{2t-1}$  and  $g_{2t}$ . For the following definitions assume that  $G = [g_0, g_1, \dots, g_{n-1}]$  is an  $ES$ -matrix of order  $s \times n$ .

*Definition 11:* Given an  $x$ -type graph  $\mathbf{G}_x(G)$  such that  $x \in \mathbb{F}_2^s$ , let  $g_i, g_j$  be two vertices connected by a simple path  $P_x(g_i, g_j, G)$  of length  $\ell - 1$  in  $\mathbf{G}_x(G)$  which is denoted by

$$g_i = g_{s_0} - g_{s_1} - \dots - g_{s_{\ell-1}} = g_j.$$

The path  $P_x(g_i, g_j, G)$  will be called a **good-path** if the edges  $\{g_{s_0}, g_{s_1}\}$  and  $\{g_{s_{\ell-2}}, g_{s_{\ell-1}}\}$  are both  $x$ -type edges. For all  $g_t$  and  $g_m$  on  $P_x(g_i, g_j, G)$ , denote by  $d_{P_x}(g_t, g_m, G)$  the length of the simple sub-path from  $g_t$  to  $g_m$  on  $P_x(g_i, g_j, G)$ . This length will be called the **sub-length from  $g_t$  to  $g_m$  in  $P_x(g_i, g_j, G)$** . When the graph  $G$  will be clear from the context we will use the notation  $P_x(g_i, g_j)$ ,  $d_{P_x}(g_t, g_m)$  instead of  $P_x(g_i, g_j, G)$ ,  $d_{P_x}(g_t, g_m, G)$ , respectively.

We next state the following claim.

*Claim 2:* Given a good-path  $P_x(g_i, g_j)$  of length  $\ell - 1$  in  $\mathbf{G}_x(G)$

$$g_i = g_{s_0} - g_{s_1} - \dots - g_{s_{\ell-1}} = g_j,$$

where  $x \in \mathbb{F}_2^s$ , the following properties hold.

- 1) The value of  $\ell$  is even.
- 2) For all  $m \in [\ell/2 - 1]$  the edge  $\{g_{s_{2m+1}}, g_{s_{2m+2}}\}_P$  is a pair-type edge.
- 3) For all  $t \in [\ell/2]$ ,  $g_{s_{2t}} = g_{s_{2t+1}} + x$ .
- 4) If  $g_i, g_j$  is not a pair, then the pair of  $g_i$  and the pair of  $g_j$  are not in  $P_x(g_i, g_j)$ .

*Proof:* We prove this claim as follows.

- 1) Since  $P_x(g_i, g_j)$  is a good-path, by definition the edge  $\{g_{s_0}, g_{s_1}\}_x$  is an  $x$ -type edge. We also know that for all  $t \in [\ell]$  it holds that  $\deg(g_{s_t}) = 2$ . Thus, the edge  $\{g_{s_1}, g_{s_2}\}_P$  is a pair-type edge, the edge  $\{g_{s_2}, g_{s_3}\}_x$  is an  $x$ -type edge, and so on. More formally, for all  $t \in [\ell/2]$  the edge  $\{g_{s_{2t}}, g_{s_{2t+1}}\}_x$  is an  $x$ -type edge and for all  $m \in [\ell/2 - 1]$  the edge  $\{g_{s_{2m+1}}, g_{s_{2m+2}}\}_P$  is a pair-type edge. Since the last edge  $\{g_{s_{\ell-2}}, g_{s_{\ell-1}}\}_x$

is also an  $x$ -type edge, we deduce that  $\ell - 1$  is odd or equivalently  $\ell$  is even.

- 2) The proof of this part holds due to a).
- 3) In a) we proved that for all  $t \in [\ell/2]$  the edge  $\{g_{s_{2t}}, g_{s_{2t+1}}\}_x$  is an  $x$ -type edge. Thus, by definition  $g_{s_{2t}} = g_{s_{2t+1}} + x$ .
- 4) Let  $g_m$  be a pair of  $g_i$  and we will prove that  $g_m \notin P_x(g_i, g_j)$ . Note that  $g_m \neq g_j$  and  $\deg(g_m) = 2$ . Therefore, if  $g_m \in P_x(g_i, g_j)$ , then  $g_i$  has to appear more than once in  $P_x(g_i, g_j)$ . This is in contradiction to the fact that  $P_x(g_i, g_j)$  is a simple path. ■

Another useful property on good-paths in  $x$ -type graphs is proved in the next claim.

*Claim 3:* If  $g_i, g_j$  is a pair, then there is a good-path  $P_x(g_i, g_j)$  in  $\mathbf{G}_x(G)$ .

*Proof:* We know that all nodes in  $\mathbf{G}_x(G)$  are of degree 2. Therefore, there is a simple cycle in  $C \in \mathbf{C}_x(G)$  including the edges  $\{g_i, g_m\}_x$  and  $\{g_j, g_p\}_x$  for some  $m, p \in [n]$ , and the edge  $\{g_i, g_j\}_P$ . By removing the edge  $\{g_i, g_j\}_P$  from  $C$  we get a simple path  $P$  starting with the edge  $\{g_i, g_m\}_x$  and ending with the edge  $\{g_j, g_p\}_x$ . Thus, by definition,  $P$  is a good-path  $P_x(g_i, g_j)$ . ■

The next definition will be used for changing the order of the columns in  $G$ .

*Definition 12:* Let  $\mathcal{H}_s$  be the set of all  $ES$ -matrices of order  $s \times n$ . Let  $\mathcal{P}_s \subseteq \mathbb{F}_2^s \times \mathbb{F}_2^s$  be the set of all couples of column vectors  $g_m, g_p$  of  $G$  such that there is a good-path  $P_x(g_m, g_p)$ . For every two column vectors  $g_i, g_j$  with a good-path  $P_x(g_i, g_j)$  of length  $\ell - 1$  in  $\mathbf{G}_x(G)$

$$g_i = g_{s_0} - g_{s_1} - \dots - g_{s_{\ell-1}} = g_j,$$

denote the **reordering function**  $\mathcal{F}_x : \mathcal{P}_s \times \mathcal{H}_s \rightarrow \mathcal{H}_s$  that generates an  $ES$ -matrix  $\mathcal{F}_x(g_i, g_j, G)$  from  $G$  by adding  $x$  to every column  $g_{s_m}, m \in [\ell]$ . We will use the notation  $\mathcal{F}_x(g_i, g_j)$  for shorthand.

The following claim proves that the function  $\mathcal{F}_x$  is well defined.

*Claim 4:* The matrix  $\mathcal{F}_x(g_i, g_j)$  is an  $ES$ -matrix of order  $s \times n$ .

*Proof:* Let  $P_x(g_i, g_j)$  be a good-path of length  $\ell - 1$  in  $\mathbf{G}_x(G)$  denoted by

$$g_i = g_{s_0} - g_{s_1} - \dots - g_{s_{\ell-1}} = g_j.$$

By using the function  $\mathcal{F}_x(g_i, g_j)$ , the vector  $x$  is added to every column  $g_{s_m}, m \in [\ell]$ . In Claim 2(3) it was shown that for all  $t \in [\ell/2]$ ,

$$g_{s_{2t}} = g_{s_{2t+1}} + x.$$

Therefore, adding  $x$  to all the columns  $g_{s_m}, m \in [\ell]$ , is equivalent to swapping the column vectors  $g_{s_{2t}}, g_{s_{2t+1}}$  for all  $t \in [\ell/2]$  in  $G$ . Since after rearranging the columns of  $G$ , it is still an  $ES$ -matrix, it is deduced that  $\mathcal{F}_x(g_i, g_j)$  is an  $ES$ -matrix. ■

To better explain these definitions and properties, the following example is presented.

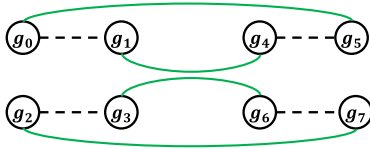
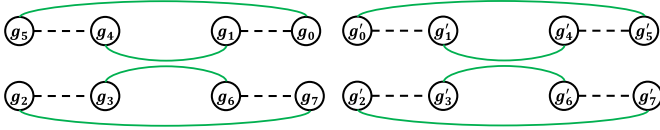


Fig. 1. The  $\mathbf{G}_x(G)$  graph. The green edges are the  $\mathbf{x}$ -type edges and the dashed edges are the pair-type edges.



(a) The graph  $\mathbf{G}_x(G')$  represented by the nodes  $\{g_0, g_1, \dots, g_7\}$ . (b) The graph  $\mathbf{G}_x(G')$  represented by the nodes  $\{g'_0, g'_1, \dots, g'_7\}$ .

Fig. 2. The graph  $\mathbf{G}_x(G')$ .

*Example 2:* For  $s = 3$ , let  $G$  be the following  $ES$ -matrix

$$G = \begin{pmatrix} g_0 & g_1 & g_2 & g_3 & g_4 & g_5 & g_6 & g_7 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Let  $\mathbf{x} = (1, 0, 1)$ . The graph  $\mathbf{G}_x(G)$  will be defined as in Figure 1. Note that in this case, the graph  $\mathbf{G}_x(G)$  is partitioned into two disjoint cycles. While the path  $g_0 - g_5$  is a good-path between  $g_0$  and  $g_5$ , the path

$$g_0 - g_1 - g_4 - g_5$$

is not a good-path between  $g_0$  and  $g_5$ . Note that there is no good-path between  $g_0$  and  $g_4$ . Let  $P_x(g_0, g_1)$  be the good-path between  $g_0$  and  $g_1$ ,

$$g_0 - g_5 - g_4 - g_1.$$

Thus,  $G' = \mathcal{F}_x(g_0, g_1)$  is the following  $ES$ -matrix

$$G' = \begin{pmatrix} g'_0 & g'_1 & g'_2 & g'_3 & g'_4 & g'_5 & g'_6 & g'_7 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ g_5 & g_4 & g_2 & g_3 & g_1 & g_0 & g_6 & g_7 \end{pmatrix},$$

with a new graph  $\mathbf{G}_x(G')$  as depicted in Figure 2.

The next lemma shows a very important property that will be used in the construction of the first algorithm. This algorithm will have a routine of  $\lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor$  iterations. In iteration  $t \leq \lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor$ , we will modify the order of the column vectors of  $G$  such that only the sums  $g_{2t} + g_{2t+1}$  and  $g_{n-2} + g_{n-1}$  will be changed by  $\mathbf{x} \in \mathbb{F}_2^s$ , and all other sums  $g_{2p} + g_{2p+1}$  where  $p \neq t, n/2 - 1$  will remain the same. The goal on the  $t$ -th iteration is to get that

$$g_{2t} + g_{2t+1} = v_t + \mathbf{1}_t e,$$

where  $\mathbf{1}_t \in \{0, 1\}$  and remember that  $e = (0, 0, \dots, 0, 1) \in \mathbb{F}_2^s$ .

*Lemma 13:* Let  $P_x(g_{r_1}, g_{r_2})$  be a good-path in  $\mathbf{G}_x(G)$  where  $\mathbf{x} \in \mathbb{F}_2^s$  and  $r_1, r_2 \in [n]$  such that  $g_{r_1}, g_{r_2}$  is not a

pair. If  $r_1 \in \{2i, 2i + 1\}$ ,  $r_2 \in \{2j, 2j + 1\}$  (and note that  $i \neq j$ ), then, the  $ES$ -matrix

$$G' = \mathcal{F}_x(g_{r_1}, g_{r_2}) = [g'_0, g'_1, \dots, g'_{n-1}]$$

satisfies the following equalities

$$\begin{aligned} g'_{2p} + g'_{2p+1} &= g_{2p} + g_{2p+1} + \mathbf{x} & p \in \{i, j\}, \\ g'_{2p} + g'_{2p+1} &= g_{2p} + g_{2p+1} & p \notin \{i, j\}, \end{aligned}$$

where  $p \in [n/2]$ .

*Proof:* We prove this lemma only for  $r_1 = 2i$  and  $r_2 = 2j$  where  $i < j$  while all other cases are proved similarly. Suppose that the good-path  $P_x(g_{2i}, g_{2j})$  is of length  $\ell - 1$  and denote it by

$$g_{2i} = g_{s_0} - g_{s_1} - \dots - g_{s_{\ell-1}} = g_{2j}.$$

Let  $S$  be the set  $S = \{s_0, s_1, \dots, s_{\ell-1}\}$ . Let

$$G' = [g'_0, g'_1, \dots, g'_{n-1}]$$

be an  $ES$ -matrix of order  $s \times 2^s$  generated by applying  $\mathcal{F}_x(g_{2i}, g_{2j}, G)$ . Thus, for all  $m \in [n]$

$$\begin{aligned} g'_m &= g_m & \text{if } m \notin S, \\ g'_m &= g_m + \mathbf{x} & \text{if } m \in S. \end{aligned}$$

Since  $P_x(g_{2i}, g_{2j})$  is as good-path and due to Claim 2(2), for all  $1 \leq t \leq \ell/2 - 1$ , it holds that  $\{g_{s_{2t-1}}, g_{s_{2t}}\}_{\mathbf{p}}$  is a pair-type edge. Thus, for all  $1 \leq t \leq \ell/2 - 1$

$$g'_{s_{2t-1}} + g'_{s_{2t}} = g_{s_{2t-1}} + \mathbf{x} + g_{s_{2t}} + \mathbf{x} = g_{s_{2t-1}} + g_{s_{2t}}.$$

Therefore, for all  $p \in [n/2] \setminus \{i, j\}$ , it holds that

$$g'_{2p} + g'_{2p+1} = g_{2p} + g_{2p+1}.$$

In case that  $p = i$  or  $p = j$ , by Claim 2(4) the columns  $g_{2i+1}$  and  $g_{2j+1}$  are not on the path  $P_x(g_{2i}, g_{2j})$ . Thus,  $g'_{2i+1} = g_{2i+1}$  and  $g'_{2j+1} = g_{2j+1}$ . Therefore,

$$g'_{2p} + g'_{2p+1} = g_{2p} + g_{2p+1} + \mathbf{x}.$$

Before proceeding to the next section, the following *FindShortPath*( $G, \mathbf{x}, t, m$ ) function is presented. Let  $G$  be an  $ES$ -matrix and  $\mathbf{G}_x(G)$  be its graph for some  $\mathbf{x} \in \mathbb{F}_2^s$ . Let  $\{g_{2t}, g_{2t+1}\}_{\mathbf{p}}$  be a pair-type edge in  $\mathbf{G}_x(G)$ . Assume that there is another pair-type edge  $\{g_{2m}, g_{2m+1}\}_{\mathbf{p}}$  in  $\mathbf{G}_x(G)$  such that  $m > t$ . The *FindShortPath*( $G, \mathbf{a}, t, m$ ) function will be used under the condition that there is a cycle  $C_i \in \mathbf{C}_x(G)$  such that both  $\{g_{2t}, g_{2t+1}\}_{\mathbf{p}}$  and  $\{g_{2m}, g_{2m+1}\}_{\mathbf{p}}$  are in  $C_i$ .

---

*FindShortPath*( $G, \mathbf{x}, t, m$ )

---

- 1:  $P_x \leftarrow$  the good-path  $P_x(g_{2t}, g_{2t+1}, G)$
  - 2:  $d_1 \leftarrow d_{P_x}(g_{2t+1}, g_{2m})$
  - 3:  $d_2 \leftarrow d_{P_x}(g_{2t+1}, g_{2m+1})$
  - 4: **if**  $d_1 < d_2$  **then**
  - 5:      $j \leftarrow 2m$
  - 6: **else**
  - 7:      $j \leftarrow 2m + 1$
  - 8: **Return**  $j$
- 

The *FindShortPath*( $G, \mathbf{x}, t, m$ ) function is presented since it will be used several times in this paper.

### B. The $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$ Function

Let  $G$  be an  $ES$ -matrix, let  $v \in \mathbb{F}_2^s$ , let  $t \in [n/2]$ , and let  $\mathcal{B}$  be a set. Denote  $\mathbf{y} = \mathbf{g}_{2t} + \mathbf{g}_{2t+1}$ . In this section we will show the function called  $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$ . This function will be used by the first algorithm which will be presented in the next section. The task of this function is to update the sum of the pair  $\mathbf{g}_{2t}, \mathbf{g}_{2t+1}$  to either  $v$  or  $v + \mathbf{y}$ . It also changes the sum of the last pair  $\mathbf{g}_{n-2}, \mathbf{g}_{n-1}$ , but, this pair is used as a “redundancy pair”, i.e., it is not important what the sum of this pair. Another important thing to mention, is that the algorithm  $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$  do not update the sum of the pairs on indices  $2p$  and  $2p + 1$  for all  $p \neq t$ , even though these columns could be reordered. The case  $\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = v$ ,  $\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = v + \mathbf{y}$  is called a good, bad case and  $t$  won't, will be inserted in  $\mathcal{B}$ , respectively. We now ready to present the function.

---

#### $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$

---

```

1:  $\mathbf{y} \leftarrow \mathbf{g}_{2t} + \mathbf{g}_{2t+1}$ 
2: if  $v = \mathbf{y}$  then
3:   Return  $G$  and  $\mathcal{B}$ 
4:  $\mathbf{u} \leftarrow \mathbf{g}_{2t+1} + \mathbf{g}_{n-2}$ 
5: for  $p = 1, 2, 3$  do
6:   if  $p = 1$  then
7:      $\mathbf{a} \leftarrow v + \mathbf{y}$ 
8:   if  $p = 2$  then
9:      $\mathbf{a} \leftarrow v + \mathbf{y} + \mathbf{u}$ 
10:    Swap the columns  $\mathbf{g}_{2t+1}$  and  $\mathbf{g}_{n-2}$  in  $G$ 
11:   if  $p = 3$  then
12:      $\mathbf{a} \leftarrow v + \mathbf{u}$ 
13:     Swap the columns  $\mathbf{g}_{2t}$  and  $\mathbf{g}_{n-2}$  in  $G$ 
14:    $P_a \leftarrow$  the good-path  $P_a(\mathbf{g}_{2t}, \mathbf{g}_{2t+1}, G)$ 
15:    $\mathbf{r} \leftarrow \{\mathbf{g}_{n-2}, \mathbf{g}_{n-1}\}_P$ 
16:   if  $\mathbf{r} \in P_a$  then
17:      $j \leftarrow \text{FindShortPath}(G, \mathbf{a}, t, n/2 - 1)$ 
18:      $G' \leftarrow \mathcal{F}_a(\mathbf{g}_{2t+1}, \mathbf{g}_j)$ 
19:     Return  $G'$  and  $\mathcal{B}'$ 
20:    $G' \leftarrow \mathcal{F}_a(\mathbf{g}_{2t}, \mathbf{g}_{2t+1})$ 
21:   Swap the columns  $\mathbf{g}'_{2t}$  and  $\mathbf{g}'_{n-2}$  of  $G'$ 
22:    $\mathcal{B}' \leftarrow \mathcal{B} \cup \{t\}$ 
23: Return  $G'$  and  $\mathcal{B}'$ 

```

---

An explanation of the  $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$  function is shown in the next example.

*Example 3:* In Fig 3 we illustrate three good situations in which Step 16 in the function  $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$  succeeds, and one bad case in which Step 16 in the function  $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$  fails. The solid green line in all figures is a sub-path of the good-path  $P_a$  (which is a path between the nodes  $\mathbf{g}_{2t}, \mathbf{g}_{2t+1}$  in  $\mathbf{G}_a(G)$ ). The dashed lines are the pair-type edges. The green dashed line is an edge on  $P_a$ . Without loss of generality, it is assumed that the closest node between  $\mathbf{g}_{n-2}$  and  $\mathbf{g}_{n-1}$  to  $\mathbf{g}_{2t+1}$  in  $P_a$  is  $\mathbf{g}_{n-2}$ . The labels of the edges represent the summation of the vectors of its incident nodes. Each of the three good cases illustrated in (a)-(c) lead to the fact that a pair

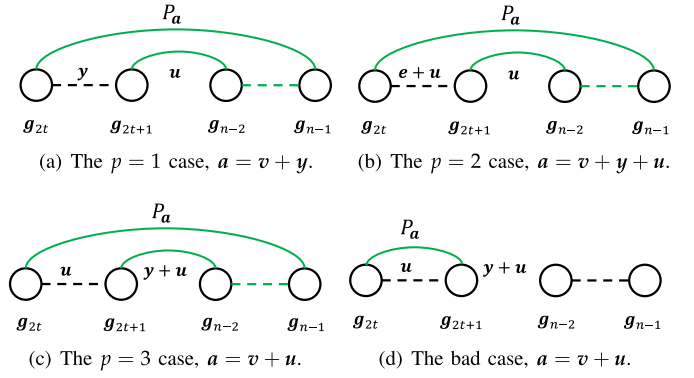


Fig. 3. Explanation of the function  $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$ . Each picture shows the graph after swapping the columns of  $G$ . In the  $p = 1$  and  $p = 2$  cases,  $\mathbf{u}$  is the sum of the vectors  $\mathbf{g}_{2t+1}$  and  $\mathbf{g}_{n-2}$ , and in the  $p = 3$  and  $p = 4$  cases,  $\mathbf{y} + \mathbf{u}$  is the sum of the vectors  $\mathbf{g}_{2t+1}$  and  $\mathbf{g}_{n-2}$ .

$\mathbf{g}'_{2t}, \mathbf{g}'_{2t+1}$  will be summed up to  $v$  (Step 18). In the bad case illustrated by (d), this pair will be summed up only to  $v + \mathbf{y}$  (Steps 20-21).

Denote by  $\mathbb{1} \in \{0, 1\}$  a binary indicator such that  $\mathbb{1} = 1$  if and only if the function  $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$  reaches Step 20. Our next goal is to prove the following important lemma.

*Lemma 14:* The function  $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$  will generate a matrix

$$G' = [\mathbf{g}'_0, \mathbf{g}'_1, \dots, \mathbf{g}'_{n-1}]$$

such that

$$\mathbf{g}'_{2p} + \mathbf{g}'_{2p+1} = \begin{cases} \mathbf{g}_{2p} + \mathbf{g}_{2p+1} & p \neq t, n/2 - 1 \\ \mathbf{v} + \mathbb{1}\mathbf{y} & p = t \end{cases}.$$

If  $\mathbb{1} = 1$  then the function  $\text{FindGoodOrBadRequest}(G, \mathcal{B}, t, v)$  will add  $t$  to the set  $\mathcal{B}$ .

*Proof:* First we show that if the function reaches Step 19, then

$$\mathbf{g}_{2t} + \mathbf{g}_{2t+1} + \mathbf{a} = v. \quad (1)$$

We separate the proof for the three cases of  $p \in \{1, 2, 3\}$ . To better understand these cases we refer the reader to Fig. 3(a)-(c). Remember that by Step 4,  $\mathbf{u} = \mathbf{g}_{2t+1} + \mathbf{g}_{n-2}$ .

- 1) If  $p = 1$ , then  $\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{y}$ . By Step 7,  $\mathbf{a} = v + \mathbf{y}$ , and therefore equality (1) holds.
- 2) If  $p = 2$ , then by Step 10, after swapping  $\mathbf{g}_{2t+1}$  and  $\mathbf{g}_{n-2}$ , it is deduced that

$$\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{y} + \mathbf{u}.$$

By Step 9,  $\mathbf{a} = v + \mathbf{y} + \mathbf{u}$ , which concludes the correctness of equality (1).

- 3) If  $p = 3$ , then by Step 13, after swapping  $\mathbf{g}_{2t}$  and  $\mathbf{g}_{n-2}$ , it is deduced that

$$\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{u}.$$

By Step 12,  $\mathbf{a} = v + \mathbf{u}$ , corresponding to (1).

Note that by Claim 3, there is always a good-path between  $\mathbf{g}_{2t}$  and  $\mathbf{g}_{2t+1}$ . Now, suppose that in one of these 3 cases, there

is a good-path  $P_a(\mathbf{g}_{2t}, \mathbf{g}_{2t+1})$  in  $\mathbf{G}_a(G)$  which includes the edge  $\{\mathbf{g}_{n-2}, \mathbf{g}_{n-1}\}_P$  (with respect to Step 16). By executing  $\text{FindShortPath}(G, \mathbf{a}, t, n/2)$  we find the closest node between  $\mathbf{g}_{n-2}$  and  $\mathbf{g}_{n-1}$  to  $\mathbf{g}_{2t+1}$  on the path  $P_a$ . This node is denoted by  $\mathbf{g}_j$ . Thus, the first and the last edges on the path  $P_a(\mathbf{g}_{2t+1}, \mathbf{g}_j)$  have to be  $x$ -type edges. By definition, it is deduced that  $P_a(\mathbf{g}_{2t+1}, \mathbf{g}_j)$  is a good-path. According to Step 18,  $G' = \mathcal{F}_a(\mathbf{g}_{2t+1}, \mathbf{g}_j)$ . By Lemma 13 this step changes the pair summations of only the pairs  $\mathbf{g}_{2t}, \mathbf{g}_{2t+1}$  and  $\mathbf{g}_{n-2}, \mathbf{g}_{n-1}$ . More precisely,

$$\mathbf{g}'_{2t} + \mathbf{g}'_{2t+1} = \mathbf{g}_{2t} + \mathbf{g}_{2t+1} + \mathbf{a} = \mathbf{v}$$

and the sum of the pair  $\mathbf{g}'_{n-2}, \mathbf{g}'_{n-1}$  does not matter.

Finally, if the function does not succeed to find any of these good-paths, we will show that it will create a matrix  $G'$  such that the pair  $\mathbf{g}'_{2t}, \mathbf{g}'_{2t+1}$  will be almost correct, that is,

$$\mathbf{g}'_{2t} + \mathbf{g}'_{2t+1} = \mathbf{v} + \mathbf{y}.$$

This will be done in Steps 20–21. First, the path  $P_a(\mathbf{g}_{2t}, \mathbf{g}_{2t+1})$  does not include the edge  $\{\mathbf{g}_{n-2}, \mathbf{g}_{n-1}\}_P$  since Step 16 has failed. Second, the function makes two swaps in Step 10 and Step 13 such that

$$\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{u},$$

and

$$\mathbf{g}_{2t+1} + \mathbf{g}_{n-2} = \mathbf{u} + \mathbf{y},$$

and by Step 12,  $\mathbf{a} = \mathbf{v} + \mathbf{u}$ . This is illustrated in Fig. 3(d). Thus, according to Step 20,  $G' = \mathcal{F}_a(\mathbf{g}_{2t}, \mathbf{g}_{2t+1})$  such that

$$\mathbf{g}'_{2t} + \mathbf{g}'_{2t+1} = \mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{u},$$

and the column vectors  $\mathbf{g}_{n-2}, \mathbf{g}_{n-1}$  are not changed. Therefore,

$$\mathbf{g}'_{2t+1} + \mathbf{g}'_{n-2} = \mathbf{g}_{2t+1} + \mathbf{g}_{n-2} + \mathbf{a} = \mathbf{v} + \mathbf{y}.$$

By Step 21 in which the columns  $\mathbf{g}'_{2t}, \mathbf{g}'_{n-2}$  are swapped we get

$$\mathbf{g}'_{2t} + \mathbf{g}'_{2t+1} = \mathbf{v} + \mathbf{y}.$$

We conclude that the function will generate an  $ES$ -matrix

$$G' = [\mathbf{g}'_0, \mathbf{g}'_1, \dots, \mathbf{g}'_{n-1}]$$

such that

$$\mathbf{g}'_{2p} + \mathbf{g}'_{2p+1} = \begin{cases} \mathbf{g}_{2p} + \mathbf{g}_{2p+1} & p \neq t, n/2 - 1 \\ \mathbf{v} + \mathbb{1}\mathbf{y} & p = t \end{cases}$$

where  $\mathbb{1} = 1$  if and only if the function reached Step 20. ■

### C. The First Algorithm

We start with the first algorithm which is referred by  $\text{FBSolution}(\tau, M)$ , where  $M$  is the request and  $\tau$  will be the number of iterations in the algorithm, which is the number of columns in  $M$ . We define more variables that will be used in the routine of  $\text{FBSolution}(\tau, M)$ , and some auxiliary results. The  $\tau$  iterations in the algorithm operate as follows. First, we demand that the initial state of the matrix

$$G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{n-1}]$$

will satisfy

$$\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{e}, \quad t \in [n/2]. \quad (2)$$

The matrix  $G$  exists due to the following claim.

*Claim 5:* There is an  $ES$ -matrix  $G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{n-1}]$  such that for all  $t \in [n/2]$

$$\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{e}.$$

*Proof:* Such an  $ES$ -matrix  $G$  is constructed by taking an order of its column vectors such that for all  $t \in [n/2]$ ,

$$\mathbf{g}_{2t} = (z_0, z_1, \dots, z_{s-2}, 0), \quad \mathbf{g}_{2t+1} = (z_0, z_1, \dots, z_{s-2}, 1).$$

The following corollary states that Claim 5 holds for all  $\mathbf{x} \in \mathbb{F}_2^s$  instead of  $\mathbf{e}$ . The proof of this corollary is similar to the one of Claim 5.

*Corollary 15:* For any  $M$  that has one kind of request  $\mathbf{v}_j$ , there is an extended Simplex solution for  $M$ .

Extending Corollary 15 to the cases where there are at most a fixed number of different requests  $d$  is an interesting problem by itself, which is out of the scope of this paper. In Section VI, we remark that  $d = 2$  can be solved by one of our constructions. For the case of  $d = 3$  we believe we have a proof, however it is omitted since we found it to be long and cumbersome. Finding a simple solution for this case and in general for arbitrary  $d$  is left for future research.

According to Corollary 15, for the rest of this section we assume that  $M$  has at least two kinds of requests  $\mathbf{v}_j$ . It is also assumed that  $\tau = \lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor$ . The  $ES$ -matrix at the end of the  $t$ -th iteration will be denoted by

$$G^{(t+1)} = [\mathbf{g}_0^{(t+1)}, \mathbf{g}_1^{(t+1)}, \dots, \mathbf{g}_{n-1}^{(t+1)}].$$

Now we are ready to present the  $\text{FBSolution}(\tau, M)$  algorithm.

---

#### Algorithm 1 $\text{FBSolution}(\tau, M)$

---

- 1:  $G^{(0)} \leftarrow G$
  - 2:  $\mathcal{B}^{(0)} = \emptyset$
  - 3: **for**  $t = 0, \dots, \tau - 1$  **do**
  - 4:  $G^{(t+1)}, \mathcal{B}^{(t+1)} \leftarrow \text{FindGoodOrBadRequest}(G^{(t)}, \mathcal{B}^{(t)}, t, \mathbf{v}_t)$
  - 5: **Return**  $G^{(\tau)}$  and  $\mathcal{B}^{(\tau)}$
- 

At the end of the  $\text{FBSolution}(\tau, M)$  algorithm we obtained the set  $\mathcal{B}_1 = \mathcal{B}^{(\tau)}$  and the matrix  $G^{(\tau)} = [\mathbf{g}_0^{(\tau)}, \mathbf{g}_1^{(\tau)}, \dots, \mathbf{g}_{n-1}^{(\tau)}]$ . By Definition 7, the set  $\mathcal{B}_1$  uniquely defines the triple-set  $(\mathcal{G}_1, \mathcal{B}_1, \mathcal{R}_1)$ . Since in our case  $\mathbf{y} = \mathbf{e}$ , by Lemma 14, for all  $t \in [n/2]$ , the matrix  $G^{(\tau)}$  satisfies that

$$\mathbf{g}_{2t}^{(\tau)} + \mathbf{g}_{2t+1}^{(\tau)} = \begin{cases} \mathbf{v}_t & t \in \mathcal{G}_1 \\ \mathbf{v}_t + \mathbf{e} & t \in \mathcal{B}_1 \\ \mathbf{e} & t \in \mathcal{R}_1 \end{cases}.$$

Let  $\mathcal{M}_1 = [\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{2^{s-1}-1}]$  be the matrix such that for all  $t \in [n/2]$

$$\mathbf{w}_t = \mathbf{g}_{2t}^{(\tau)} + \mathbf{g}_{2t+1}^{(\tau)}.$$

Therefore, it is deduced that  $\mathcal{M}_1 = \mathcal{M}_1(\mathcal{G}_1, \mathcal{B}_1, \mathcal{R}_1)$  is a triple-matrix of  $M$ . By definition of  $\mathcal{M}_1(\mathcal{G}_1, \mathcal{B}_1, \mathcal{R}_1)$ , the



matrix  $G^{(\tau)}$  is its extended Simplex solution. If the set  $\mathcal{B}_1$  satisfies  $|\mathcal{B}_1| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$  then by Lemma 8 there is a solution for  $M$ . Otherwise, we will make another reordering on the columns of  $G^{(\tau)}$  in order to obtain a new bad set  $\mathcal{B}_2$  for which  $|\mathcal{B}_2| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$ . This will be done in the next section by showing our second algorithm.

#### D. The Second Algorithm

From now on we assume that

$$G^{(\tau)} = G = [g_0, g_1, \dots, g_{n-1}]$$

and  $\mathcal{B} = \mathcal{B}_1$ . Before showing the second algorithm we start with the following definition.

*Definition 16:* Let  $\mathbf{C}_e(G)$  be a partition of simple cycles in  $\mathbf{G}_e(G)$ . A pair of distinct indices  $t_1, t_2 \in \mathcal{B}$  is called a **bad-indices pair** in  $\mathbf{C}_e(G)$  if both edges  $\{g_{2t_1}, g_{2t_1+1}\}_{\mathbf{P}}$  and  $\{g_{2t_2}, g_{2t_2+1}\}_{\mathbf{P}}$  are in the same simple cycle in  $\mathbf{G}_e(G)$ . Now we show the algorithm *ClearBadCycles*( $G, \mathcal{B}$ ) in which the columns of the *ES*-matrix  $G$  are reordered, and the set  $\mathcal{B}$  will be modified and its size will be decreased.

---

#### Algorithm 2 *ClearBadCycles*( $G, \mathcal{B}$ )

---

- 1:  $\mathbf{C}_e(G) \leftarrow$  The partition of simple cycles in  $\mathbf{G}_e(G)$
  - 2: **while**  $\exists t_1, t_2$  a bad-indices pair in  $\mathbf{C}_e(G)$  **do**
  - 3:    $j \leftarrow$  *FindShortPath*( $G, e, t_1, t_2$ )
  - 4:    $G \leftarrow \mathcal{F}_e(g_{2t_1+1}, g_j)$
  - 5:    $\mathbf{C}_e(G) \leftarrow$  The partition of simple cycles in  $\mathbf{G}_e(G)$
  - 6:   Remove  $t_1, t_2$  from  $\mathcal{B}$
  - 7: **Return**  $G$  and  $\mathcal{B}$
- 

Let  $G_2 = [g_0^*, g_1^*, \dots, g_{n-1}^*]$  be the *ES*-matrix and  $\mathcal{B}_2$  be the bad set output of the *ClearBadCycles*( $G, \mathcal{B}$ ) algorithm. We remind the reader that  $M = [v_0, v_1, \dots, v_{\lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor - 1}]$ . Let  $\mathcal{M}_2 = [w_0, w_1, \dots, w_{2^{s-1}-1}]$  be the matrix such that for all  $t \in [n/2]$

$$w_t = g_{2t}^* + g_{2t+1}^*.$$

Since  $\mathcal{B}_2$  uniquely defines the triple set  $(\mathcal{G}_2, \mathcal{B}_2, \mathcal{R}_2)$ , it is deduced that the matrix  $\mathcal{M}_2$  is a triple-matrix  $\mathcal{M}_2(\mathcal{G}_2, \mathcal{B}_2, \mathcal{R}_2)$  of  $M$ . Next, it will be shown that the algorithm *ClearBadCycles*( $G, \mathcal{B}$ ) will stop and  $|\mathcal{B}_2|$  will be bounded from above by  $\lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$  after the execution of the algorithm.

*Lemma 17:* The algorithm *ClearBadCycles*( $G, \mathcal{B}$ ) outputs a set  $\mathcal{B}_2$  such that  $|\mathcal{B}_2| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$ .

*Proof:* According to Step 2 if there is a simple cycle containing a bad-indices pair  $t_1, t_2$  in  $\mathbf{C}_e(G)$ , the algorithm will enter the routine. Thus, there is a good-path between  $g_{2t_1+1}$  and one of the nodes  $g_{2t_2}, g_{2t_2+1}$  (the closest one between them to  $g_{2t_1+1}$ ), and the index of this node is denoted by  $j$  (Step 3). Since  $t_1, t_2 \in \mathcal{B}$ , before the algorithm reaches Step 4, it holds

$$g_{2t_1} + g_{2t_1+1} = v_{t_1} + e, \quad g_{2t_2} + g_{2t_2+1} = v_{t_2} + e.$$

By executing  $\mathcal{F}_e(g_{2t_1+1}, g_j)$ , due to Lemma 13, the matrix  $G$  is updated to a matrix  $G'$  such that only the two following

pair summations are correctly changed to

$$g'_{2t_1} + g'_{2t_1+1} = v_{t_1}, \quad g'_{2t_2} + g'_{2t_2+1} = v_{t_2}.$$

Thus, the indices  $t_1$  and  $t_2$  are removed from  $\mathcal{B}$  (Step 6). Therefore, Step 2 will fail when each simple cycle will have at most one  $t \in \mathcal{B}$  such that

$$g_{2t} + g_{2t+1} = v_t + e,$$

and we will call it a “bad cycle”. Suppose that there are  $p$  bad cycles at the end of the algorithm. We are left with showing that  $p \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$ .

Observe that for all  $t \in \mathcal{R}$ , the nodes  $g_{2t}$  and  $g_{2t+1}$  are connected by two parallel edges, and therefore they create cycles of length 2. These cycles are not bad cycles by definition. Also note that if the pair  $g_{n-2}, g_{n-1}$  is in some bad cycle, then the bad pair  $g_{2t}, g_{2t+1}$  in this cycle can be corrected by  $\mathcal{F}_e(g_{2t+1}, g)$ , where  $g$  is the closest node between  $g_{n-2}$  and  $g_{n-1}$  to  $g_{2t+1}$  in this cycle. Therefore, we assume that  $g_{n-2}, g_{n-1}$  is not in any bad cycle. Since there are  $2|\mathcal{R}|$  such columns in  $G$  and together with the pair  $g_{n-2}, g_{n-1}$  and  $|\mathcal{R}| = 2^{s-1} - \lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor - 1$ , only the first

$$2 \cdot |\mathcal{B} \cup \mathcal{G}| = 2 \cdot \left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor$$

columns of  $G$  can be partitioned into bad cycles. Our next goal is to prove that the size of each bad cycle is at least 4. Assume to the contrary that there is a bad cycle of length 2. Since we are using the graph  $\mathbf{G}_e(G)$ , such a simple cycle of two nodes  $g_{2t}, g_{2t+1}$ ,  $t \in \mathcal{B}$ , satisfies that

$$g_{2t} + g_{2t+1} = e.$$

In that case  $g_{2t} + g_{2t+1} \neq v_t + e$  since  $v_t$  is non-zero vector, so  $g_{2t} + g_{2t+1} = v_t + e$ . According to Step 2 in the function *FindGoodOrBadRequest*( $G, \mathcal{B}, t, v$ ),  $t \notin \mathcal{B}$ , which results with a contradiction. Therefore, indeed all simple cycles are of size at least 4. Thus,

$$|\mathcal{B}| \leq p \leq \left\lfloor \frac{1}{4} \cdot \left( 2 \cdot \left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor \right) \right\rfloor = \left\lfloor \frac{1}{3} \cdot 2^{s-1} \right\rfloor,$$

where the last equality holds since by the nested division  $\left\lfloor \frac{\lfloor \frac{x}{y} \rfloor}{z} \right\rfloor = \left\lfloor \frac{x}{yz} \right\rfloor$  for real  $x, y$  and a positive integer  $z$ . ■

We are finally ready to prove the main result of this section.

*Theorem 18:* An  $FB$ -( $2^s - 1, s, \lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor$ ) code exists.

*Proof:* By Lemma 17 the algorithm *ClearBadCycles*( $G, \mathcal{B}$ ) outputs the set  $\mathcal{B}_2$  such that its size is at most  $\lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$ . The *ES*-matrix  $G_2$  is again an extended Simplex solution for a triple-matrix  $\mathcal{M}_2(\mathcal{G}_2, \mathcal{B}_2, \mathcal{R}_2)$  of  $M$ . Thus, by using Lemma 8, it is deduced that there is a solution for  $M$ . After removing the all-zero column vector from  $G$ , the proof of this theorem is completed. ■

#### IV. A CONSTRUCTION OF

$FB$ -( $2^s + \lceil (3\alpha - 2) \cdot 2^{s-2} \rceil - 1, s, \lfloor \alpha \cdot 2^{s-1} \rfloor$ ) CODES

In this section we show how to construct  $FB$ -( $2^s + \lceil (3\alpha - 2) \cdot 2^{s-2} \rceil - 1, s, \lfloor \alpha \cdot 2^{s-1} \rfloor$ ) codes where  $\frac{2}{3} \leq \alpha \leq 1$ . For convenience, throughout this section let  $n = 2^s$  and  $m =$

$2^s + \lceil (3\alpha - 2) \cdot 2^{s-2} \rceil$ . Note that since  $\alpha \geq 2/3$  it holds that  $m \geq n$ . Let  $\mathbf{e} = (0, 0, \dots, 0, 1) \in \mathbb{F}_2^s$ .

The following two definitions extend  $ES$ -matrices from Definition 5 and triple-matrices from Definition 7.

**Definition 19:** A matrix  $G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{m-1}]$  of order  $s \times m$  over  $\mathbb{F}_2$  is called an *extended-ES-matrix* if the matrix  $H_G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{n-1}]$  is an  $ES$ -matrix of order  $s \times n$  and for all  $n \leq i \leq m-1$  it holds  $\mathbf{g}_i = \mathbf{e}$ . The  $ES$ -matrix  $H_G$  will be called the  $H$ -part of  $G$ .

**Definition 20:** Three sets  $\mathcal{G}, \mathcal{B}, \mathcal{R} \subseteq [2^{s-1}]$  are called an  $\alpha$ -**triple-set**, and are denoted by  $\alpha$ - $(\mathcal{G}, \mathcal{B}, \mathcal{R})$ , if the following properties hold

$$\begin{aligned} \mathcal{G} &\subseteq \left[ \left[ \alpha \cdot 2^{s-1} \right] \right], \\ \mathcal{B} &= \left[ \left[ \alpha \cdot 2^{s-1} \right] \right] \setminus \mathcal{G}, \\ \mathcal{R} &= [2^{s-1}] \setminus (\mathcal{G} \cup \mathcal{B} \cup \{2^{s-1} - 1\}). \end{aligned}$$

Given a matrix  $M = [\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{\lfloor \alpha \cdot 2^{s-1} \rfloor - 1}]$  of order  $s \times \lfloor \alpha \cdot 2^{s-1} \rfloor$ , a matrix  $\mathcal{M}(\mathcal{G}, \mathcal{B}, \mathcal{R}) = [\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{2^{s-1}-1}]$  of order  $s \times 2^{s-1}$  is referred as an  $\alpha$ -**triple-matrix** of  $M$  if it holds that

$$\mathbf{w}_t = \begin{cases} \mathbf{v}_t & t \in \mathcal{G} \\ \mathbf{v}_t + \mathbf{e} & t \in \mathcal{B} \\ \mathbf{e} & t \in \mathcal{R} \end{cases}.$$

Note that by taking  $\alpha = 2/3$ , Definition 20 will be equivalent to Definition 7. Denote  $\mathcal{N} = [m] \setminus [n]$  and note that  $|\mathcal{N}| = \lceil (3\alpha - 2) \cdot 2^{s-2} \rceil$ . We seek to design an algorithm which is very similar to the  $FBSolution(\tau, M)$  algorithm in the following respect. This algorithm will output an  $ES$ -matrix  $H_G$  which will be the  $H$ -part of an extended- $ES$ -matrix  $G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{m-1}]$ , and a set  $\mathcal{B}$ . The set  $\mathcal{B}$  will define uniquely the  $\alpha$ -triple-set  $\alpha$ - $(\mathcal{G}, \mathcal{B}, \mathcal{R})$ . For all  $t \in \mathcal{G}$  we will get

$$\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{v}_t,$$

and for all  $t \in \mathcal{B}$  we will get an almost desirable solution, that is,

$$\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{v}_t + \mathbf{e}.$$

The summation of the last pair  $\mathbf{g}_{n-2}, \mathbf{g}_{n-1}$  will be arbitrary. Similarly to the technique that was shown in Section III, the set  $\mathcal{R}$  will be used to correct the summations  $\mathbf{v}_t + \mathbf{e}$  to  $\mathbf{v}_t$ , where  $t \in \mathcal{B}$ . For that, the  $ClearBadCycles(G, \mathcal{B})$  algorithm will be used as it was done in Section III in order to obtain an extended- $ES$ -matrix  $G$  and a set  $\mathcal{B}$  such that  $|\mathcal{B}| \leq \lfloor \frac{1}{2} \alpha \cdot 2^{s-1} \rfloor$ . Even though  $|\mathcal{B}| \leq \lfloor \frac{1}{2} \alpha \cdot 2^{s-1} \rfloor$ , the size of  $\mathcal{R}$  will not be bigger than the size of  $\mathcal{B}$  for  $\alpha > 2/3$ . Thus, in this case, not all bad summation can be corrected. For that, we define the set  $\mathcal{N}$  that is also used to correct the summations  $\mathbf{v}_t + \mathbf{e}$  to  $\mathbf{v}_t$ , where  $t \in \mathcal{B}$ . This will be done based on the property that for all  $t \in \mathcal{N}$  it holds that  $\mathbf{g}_t = \mathbf{e}$ . In case that  $\alpha < 1$ , together with  $\mathcal{R}$  and  $\mathcal{N}$ , the last pair  $\mathbf{g}_{n-2}, \mathbf{g}_{n-1}$  will be used for the correction of these summations. Thus, if the inequality  $|\mathcal{B}| \leq |\mathcal{R}| + |\mathcal{N}| + 1$  holds, then it is possible to construct a solution for  $M$ . In case that  $\alpha = 1$ , we obtain  $|\mathcal{R}| = 0$ . In this case, we will show how to get the inequality  $|\mathcal{B}| \leq |\mathcal{N}|$ ,

which will similarly lead to a solution for  $M$ . Even though the last pair  $\mathbf{g}_{n-2}, \mathbf{g}_{n-1}$  has an arbitrary summation, it will still be shown how to obtain the request  $\mathbf{v}_{n/2-1}$  from this pair. Therefore, our first goal is to show a condition which assures that either  $|\mathcal{B}| \leq |\mathcal{R}| + |\mathcal{N}| + 1$  or  $|\mathcal{B}| \leq |\mathcal{N}|$ . This is done in Claim 6.

**Claim 6:** Let  $(\mathcal{G}, \mathcal{B}, \mathcal{R})$  be an  $\alpha$ -triple-set where  $|\mathcal{B}| \leq \lfloor \frac{1}{2} \alpha \cdot 2^{s-1} \rfloor$ . If  $2/3 \leq \alpha < 1$ , then  $|\mathcal{B}| \leq |\mathcal{R}| + |\mathcal{N}| + 1$ , and if  $\alpha = 1$  then  $|\mathcal{B}| \leq |\mathcal{N}|$ .

**Proof:** Let  $2/3 \leq \alpha < 1$ . According to the definition of  $\alpha$ - $(\mathcal{G}, \mathcal{B}, \mathcal{R})$ , since  $\mathcal{G} \cup \mathcal{B} = [\alpha \cdot 2^{s-1}]$  it holds that

$$\begin{aligned} |\mathcal{R}| &= \left| [2^{s-1}] \setminus (\mathcal{G} \cup \mathcal{B} \cup \{2^{s-1} - 1\}) \right| \\ &= 2^{s-1} - \left\lfloor \alpha \cdot 2^{s-1} \right\rfloor - 1. \end{aligned}$$

We also know that  $|\mathcal{N}| = \lceil (3\alpha - 2) \cdot 2^{s-2} \rceil$ . Thus, in order to prove that  $|\mathcal{B}| \leq |\mathcal{R}| + |\mathcal{N}| + 1$ , since  $|\mathcal{B}| \leq \lfloor \frac{1}{2} \alpha \cdot 2^{s-1} \rfloor$ , it is enough to prove inequality (a) in

$$\begin{aligned} |\mathcal{R}| + |\mathcal{N}| + 1 &= 2^{s-1} - \left\lfloor \alpha \cdot 2^{s-1} \right\rfloor + \left\lceil (3\alpha - 2) \cdot 2^{s-2} \right\rceil \\ &\stackrel{(a)}{\geq} \left\lfloor \frac{1}{2} \alpha \cdot 2^{s-1} \right\rfloor \geq |\mathcal{B}|. \end{aligned}$$

Inequality (a) is equivalent to

$$2^{s-1} \geq \left\lfloor 2\alpha \cdot 2^{s-2} \right\rfloor + \left\lfloor \alpha \cdot 2^{s-2} \right\rfloor - \left\lceil (3\alpha - 2) \cdot 2^{s-2} \right\rceil,$$

which holds since

$$\begin{aligned} &\left\lfloor 2\alpha \cdot 2^{s-2} \right\rfloor + \left\lfloor \alpha \cdot 2^{s-2} \right\rfloor - \left\lceil (3\alpha - 2) \cdot 2^{s-2} \right\rceil \\ &\leq 2\alpha \cdot 2^{s-2} + \alpha \cdot 2^{s-2} - (3\alpha - 2) \cdot 2^{s-2} \\ &= 2^{s-2}(2\alpha + \alpha - (3\alpha - 2)) = 2 \cdot 2^{s-2} = 2^{s-1}. \end{aligned}$$

Now if  $\alpha = 1$ , then  $|\mathcal{B}| \leq \lfloor \frac{1}{2} \cdot 2^{s-1} \rfloor = 2^{s-2}$ . By the definition of  $\mathcal{R}$ , it holds that  $|\mathcal{R}| = 0$  and by the definition of  $\mathcal{N}$  it holds that  $|\mathcal{N}| = 2^{s-2}$ . Therefore  $|\mathcal{B}| \leq 2^{s-2} = |\mathcal{N}|$ . ■

Let  $M = [\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{\lfloor \alpha \cdot 2^{s-1} \rfloor - 1}]$  be a request of order  $s \times \lfloor \alpha \cdot 2^{s-1} \rfloor$ . Our goal is to construct an extended- $ES$ -matrix of order  $s \times m$  which will provide a solution for  $M$ . For that, the  $\alpha$ - $FBSolution(M)$  algorithm is presented. In this algorithm, the matrix  $H$  is represented by  $H = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{n-1}]$ .

---

### Algorithm 3 $\alpha$ - $FBSolution(M)$

---

- 1: **if**  $\alpha < 1$  **then**
  - 2:    $\tau \leftarrow \lfloor \alpha \cdot 2^{s-1} \rfloor$
  - 3: **else if**  $\alpha = 1$  **then**
  - 4:    $\tau \leftarrow 2^{s-1} - 1$
  - 5:  $H, \mathcal{B} \leftarrow FBSolution(\tau, M)$
  - 6:  $H, \mathcal{B} \leftarrow ClearBadCycles(H, \mathcal{B})$
  - 7: **if**  $\alpha < 1$  and  $\mathbf{g}_{n-2} \neq \mathbf{e}$  **then**
  - 8:    $H \leftarrow \mathcal{F}_{\mathbf{g}_{n-2} + \mathbf{e}}(\mathbf{g}_{n-2}, \mathbf{g}_{n-1})$
  - 9: **if**  $\alpha = 1$  and  $\mathbf{g}_{n-2} \neq \mathbf{v}_{n/2-1}$  **then**
  - 10:    $H \leftarrow \mathcal{F}_{\mathbf{g}_{n-2} + \mathbf{v}_{n/2-1}}(\mathbf{g}_{n-2}, \mathbf{g}_{n-1})$
  - 11: **Return**  $H$  and  $\mathcal{B}$
-

Denote by  $G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{m-1}]$  an extended- $ES$ -matrix of order  $s \times m$  such that the output matrix  $H$  from the  $\alpha$ - $FBSolution(M)$  algorithm is its  $H$ -part, i.e.,  $H_G = H$ . Note that Steps 5–6 define the set  $\mathcal{B}$ . This set is obtained using a similar technique to the one from Section III, except to the fact that here  $2/3 \leq \alpha \leq 1$ , while in Section III,  $\alpha = 2/3$ . It is important to note that the size of  $\mathcal{B}$  is bounded due to the execution of the  $ClearBadCycles(H, \mathcal{B})$  algorithm (Step 6). Therefore, we only state the following lemma since its proof is very similar to the one that was shown in Lemma 17.

**Lemma 21:** The  $\alpha$ - $FBSolution(M)$  algorithm outputs a set  $\mathcal{B}$  such that  $|\mathcal{B}| \leq \lfloor \frac{1}{2} \alpha \cdot 2^{s-1} \rfloor$ .

We will use Lemma 21 while proving the main theorem of this section.

**Theorem 22:** For any  $2/3 \leq \alpha \leq 1$ , a functional batch code

$$FB-(2^s + \lceil (3\alpha - 2) \cdot 2^{s-2} \rceil - 1, s, \lfloor \alpha \cdot 2^{s-1} \rfloor)$$

exists.

*Proof:* After finishing the  $\alpha$ - $FBSolution(M)$  algorithm, we obtain an  $ES$ -matrix  $H$  which is the  $H$ -part of the extended- $ES$ -matrix  $G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_m]$ . Remember that by the definition of  $G$  and  $\mathcal{N}$ , for all  $t \in \mathcal{N}$ , it holds that  $\mathbf{g}_t = \mathbf{e}$ . In Step 5 we invoke the algorithm  $FBSolution(\tau, M)$ . Therefore, for all  $t < \tau$ , there exists  $\mathbf{1}_t \in \{0, 1\}$  such that

$$\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{v}_t + \mathbf{1}_t \mathbf{e}.$$

Let  $\alpha$ - $(\mathcal{G}, \mathcal{B}, \mathcal{R})$  be an  $\alpha$ -triple-set that is uniquely defined by  $\mathcal{B}$  according to Definition 20. Clearly, for all  $t \in \mathcal{G}$ , the recovery set  $R_t$  is  $R_t = \{2t, 2t+1\}$ . By Lemma 21 it holds that  $|\mathcal{B}| \leq \lfloor \frac{1}{2} \alpha \cdot 2^{s-1} \rfloor$ . We separate this proof for two cases.

**Case 1:** Assume that  $\alpha < 1$ . Due to Lemma 21 and Claim 6, if  $\alpha < 1$ , it is deduced that  $|\mathcal{B}| \leq |\mathcal{R}| + |\mathcal{N}| + 1$ . Let  $t$  be the maximum number in  $\mathcal{B}$  and let  $\mathcal{B}' = \mathcal{B} \setminus \{t\}$ . Thus,  $|\mathcal{B}'| \leq |\mathcal{R}| + |\mathcal{N}|$ . Therefore, for all  $t \in \mathcal{B}'$ ,  $R_t$  will have a different  $t'$  such that  $R_t$  equals to either  $\{2t, 2t+1, 2t', 2t'+1\}$  where  $t' \in \mathcal{R}$ , or  $\{2t, 2t+1, t'\}$  where  $t' \in \mathcal{N}$ . Thus, we showed the recovery sets for all requests except of  $\mathbf{v}_t$ . Remember that  $\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{v}_t + \mathbf{e}$ , and note that if  $\mathbf{g}_{n-2} = \mathbf{e}$ , this case is finished. Otherwise,  $\mathbf{g}_{n-2} \neq \mathbf{e}$ . This is handled by Steps 7–8 as follows. By Claim 3, since  $\mathbf{g}_{n-2}$  and  $\mathbf{g}_{n-1}$  is a pair, we know that there is a good-path  $P_x(\mathbf{g}_{n-2}, \mathbf{g}_{n-1})$  in an  $x$ -type graph  $\mathbf{G}_x(G)$  for all  $x \in \mathbb{F}_2^s$ . Thus, if  $\mathbf{g}_{n-2} \neq \mathbf{e}$ , by taking  $x = \mathbf{g}_{n-2} + \mathbf{e}$ , the algorithm can use the reordering function  $\mathcal{F}_{\mathbf{g}_{n-2} + \mathbf{e}}(\mathbf{g}_{n-2}, \mathbf{g}_{n-1})$ , as it is done in Step 8. By Lemma 13, we obtain two new column vectors  $\mathbf{g}'_{n-2}$  and  $\mathbf{g}'_{n-1}$  such that

$$\begin{aligned} \mathbf{g}'_{n-2} &= \mathbf{g}_{n-2} + \mathbf{g}_{n-2} + \mathbf{e} = \mathbf{e}, \\ \mathbf{g}'_{n-1} &= \mathbf{g}_{n-1} + \mathbf{g}_{n-2} + \mathbf{e}, \end{aligned}$$

without changing the summations of all other pairs on this path. Therefore, the recovery set for  $\mathbf{v}_t$  will be  $R_t = \{2t, 2t+1, n-2\}$ , which concludes this case.

**Case 2:** Assume that  $\alpha = 1$ . Due to Lemma 21 and Claim 6 if  $\alpha = 1$  then  $|\mathcal{B}| \leq |\mathcal{N}|$ . Thus, similarly to Case 1, for all  $t \in \mathcal{B}$ , the recovery sets  $R_t$  can be obtained. However, we do not have a recovery set for  $\mathbf{v}_{n/2-1}$  since the sum

of the pair  $\mathbf{g}_{n-2}, \mathbf{g}_{n-1}$  is arbitrary. If  $\mathbf{g}_{n-2} = \mathbf{v}_{n/2-1}$ , then  $R_{n/2-1} = \{n-2\}$ . Otherwise, as in Case 1, by Step 10 it is deduced that  $\mathbf{g}'_{n-2} = \mathbf{v}_{n/2-1}$ . Again  $R_{n/2-1} = \{n-2\}$ , which concludes this case.

In both cases, after removing the all-zero column from  $G$ , we conclude the proof.  $\blacksquare$

## V. A CONSTRUCTION OF $FB-(2^{s+1} - 2, s, 2^s)$ CODES

In this section, a construction for  $FB-(2^{s+1} - 2, s, 2^s)$  codes will be shown by using the algorithm  $FBSolution(\tau, M)$ . Throughout this section let  $n = 2^{s+1}$  and let  $\mathbf{e} = (0, 0, \dots, 0, 1) \in \mathbb{F}_2^{s+1}$ . We start with the following definition.

**Definition 23:** A matrix  $G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{2^{s+1}-1}]$  of order  $s \times 2^{s+1}$  over  $\mathbb{F}_2$  such that each vector of  $\mathbb{F}_2^s$  appears as a column vector in  $G$  exactly twice, is called a *double-ES-matrix*.

Note that by removing the last row from any  $ES$ -matrix of order  $(s+1) \times n$ , we get a double- $ES$ -matrix of order  $s \times n$ . Also, note that each double- $ES$ -matrix has exactly two all-zero columns. These columns will be removed at the end of the procedure, obtaining only  $2^{s+1} - 2$  column vectors. Next, the definition of an extended Simplex solution is extended with respect to Definition 6.

**Definition 24:** Let  $M = [\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{2^s-1}]$  be a request of order  $s \times 2^s$ . The matrix  $M$  has an **extended Simplex solution** if there exists a double- $ES$ -matrix  $G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{2^{s+1}-1}]$  of order  $s \times n$  such that for all  $t \in [2^s - 1]$ ,

$$\mathbf{v}_t = \mathbf{g}_{2t} + \mathbf{g}_{2t+1},$$

and for  $t = n/2 - 1$  either  $\mathbf{v}_t = \mathbf{g}_{n-2} + \mathbf{g}_{n-1}$ , or  $\mathbf{v}_t = \mathbf{g}_{n-2}$ , or  $\mathbf{v}_t = \mathbf{g}_{n-1}$ .

Let  $M = [\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{2^s-1}]$  be a request of order  $s \times 2^s$ . Our goal is to construct a double- $ES$ -matrix of order  $s \times 2^{s+1}$  which will provide an extended Simplex solution for  $M$ . Let  $\mathcal{M} = [\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{2^s-1}]$  be a new matrix of order  $(s+1) \times 2^s$  generated by adding the all-zero row to  $M$ . Let  $\tau = n/2 - 1$  and  $\mathbf{0}_\ell$  be the all-zero vector of length  $\ell$ . We now show the algorithm  $OptFBSolution(\mathcal{M})$ , which receives as an input the matrix  $\mathcal{M}$  and outputs a double- $ES$ -matrix  $G$  that will be a solution for  $M$ . As mentioned in the Introduction the returned solution is optimal.

---

### Algorithm 4 $OptFBSolution(\mathcal{M})$

---

- 1:  $\tau \leftarrow n/2 - 1$
  - 2:  $G \leftarrow FBSolution(\tau, \mathcal{M})$
  - 3:  $\mathbf{y} \leftarrow \sum_{i=0}^{2^s-1} \mathbf{w}_i$
  - 4: **if**  $\mathbf{y} \neq \mathbf{0}_{s+1}$  and  $\mathbf{y} \neq \mathbf{g}_{n-2}$  **then**
  - 5:      $G \leftarrow \mathcal{F}_{\mathbf{g}_{n-2} + \mathbf{y}}(\mathbf{g}_{n-2}, \mathbf{g}_{n-1})$
  - 6: Remove the last row from  $G$
- 

The following lemma proves the correctness of Algorithm  $OptFBSolution(\mathcal{M})$ .

**Lemma 25:** The algorithm  $OptFBSolution(\mathcal{M})$  outputs a double- $ES$ -matrix  $G'$  which is an extended Simplex solution for  $M$ .

*Proof:* According to Step 2, the algorithm  $FBSolution(\tau, \mathcal{M})$  is used with  $\tau = n/2 - 1$ . Thus, by Lemma 14 we obtain an  $ES$ -matrix  $G = [g_0, g_1, \dots, g_{n-1}]$  such that for all  $t \in [n/2 - 1]$

$$g_{2t} + g_{2t+1} = w_t + \mathbb{1}_t e, \quad (3)$$

where  $\mathbb{1}_t \in \{0, 1\}$ . Let  $G' = [g'_0, g'_1, \dots, g'_{n-1}]$  be a double- $ES$ -matrix of order  $s \times n$  generated by removing the last row from  $G$  according to Step 6. Since  $\mathcal{M} = [w_0, w_1, \dots, w_{2^s-1}]$  is generated by adding the all-zero row to  $M$ , by removing the last row from  $G$  before Step 3, for all  $t \in [n/2 - 1]$  we could obtain  $G'$  such that

$$g'_{2t} + g'_{2t+1} = v_t. \quad (4)$$

However,  $G'$  would provide a solution for  $M$  except for the last request  $v_{n/2-1}$ . We handle the last request using Steps 3–5 that will be explained as follows.

Assume that  $\sum_{i=0}^{2^s-1} w_i = \mathbf{y}$  and note that

$$\sum_{i=0}^{2^{s+1}-1} g_i = \mathbf{0}_{s+1}. \quad (5)$$

Denote

$$\mathbb{1}_{n/2-1} = \sum_{t=0}^{2^s-2} \mathbb{1}_t(\text{mod } 2). \quad (6)$$

Thus, it is deduced that

$$\begin{aligned} g_{n-2} + g_{n-1} &\stackrel{(a)}{=} \sum_{i=0}^{2^{s+1}-3} g_i \stackrel{(b)}{=} \sum_{t=0}^{2^s-2} (w_t + \mathbb{1}_t e) \\ &\stackrel{(c)}{=} w_{n/2-1} + \mathbf{y} + \mathbb{1}_{n/2-1} e. \end{aligned}$$

Equality (a) holds due to (5), equality (b) holds according to (3), and equality (c) holds by the definition of  $\mathbf{y}$  and by (6). Now if  $\mathbf{y} = \mathbf{0}_{s+1}$ , according to Step 6, after removing the last row from  $G$ , we get  $G'$  such that equation (4) holds also for  $t = n/2 - 1$ , that is,

$$g'_{n-2} + g'_{n-1} = v_{n/2-1}.$$

Clearly, in this case  $G'$  is an extended Simplex solution for  $M$ . Otherwise, if  $\mathbf{y} \neq \mathbf{0}_{s+1}$  then the algorithm enters the if condition in Step 4. By Claim 3, since  $g_{n-2}$  and  $g_{n-1}$  is a pair, we know that there is a good-path  $P_x(g_{n-2}, g_{n-1})$  in an  $x$ -type graph  $\mathbf{G}_x(G)$  for all  $x \in \mathbb{F}_2^{s+1}$ . Thus, by taking  $x = g_{n-2} + \mathbf{y}$ , the algorithm will execute the reordering function  $\mathcal{F}_{g_{n-2}+\mathbf{y}}(g_{n-2}, g_{n-1})$  (Step 5). By Lemma 13, we obtain two new column vectors  $g'_{n-2}$  and  $g'_{n-1}$  such that

$$\begin{aligned} g'_{n-2} &= g_{n-2} + g_{n-2} + \mathbf{y} = \mathbf{y}, \\ g'_{n-1} &= g_{n-1} + g_{n-2} + \mathbf{y} = w_{n/2-1} + \mathbb{1}_{n/2-1} e, \end{aligned}$$

without changing the summation of all other pairs on this path. Again, by removing the last row from  $G$ , we obtain  $G'$  such that

$$g'_{2t} + g'_{2t+1} = v_t, \quad t \in [n/2 - 1]$$

and  $g'_{n-1} = v_{n/2-1}$ . Thus, all the recovery sets  $R_t, t \in [n/2 - 1]$  are of the form  $R_t = \{2t, 2t + 1\}$ , and the last recovery set

will be  $R_{n/2-1} = \{n-1\}$ , which concludes this case. In both cases,  $G'$  is a double- $ES$ -matrix with two all-zero columns that will be removed to provide an  $FB-(2^{s+1} - 2, s, 2^s)$  code.  $\blacksquare$

For the rest of the paper, we only state that it is possible to obtain the last recovery set from the redundancy columns  $g_{n-2}$  and  $g_{n-1}$ , as it was shown in the proof of Lemma 25. From the result of Lemma 25 we deduce the main theorem of this section.

*Theorem 26:* An  $FB-(2^{s+1} - 2, s, 2^s)$  code exists.

## VI. A CONSTRUCTION OF $FB-(2^s - 1, s, \lfloor \frac{5}{6} \cdot 2^{s-1} \rfloor - s)$ CODES

In this section we show how to improve our main result, i.e., we show a construction of  $FB-(2^s - 1, s, \lfloor \frac{5}{6} \cdot 2^{s-1} \rfloor - s)$  codes. Let  $M$  be a request denoted by

$$M = [v_0, v_1, \dots, v_{\lfloor \frac{5}{6} \cdot 2^{s-1} \rfloor - s}].$$

Remember that  $e = (0, 0, \dots, 0, 1) \in \mathbb{F}_2^s$ , and  $n = 2^s$ . The initial state of the matrix

$$G = [g_0, g_1, \dots, g_{n-1}]$$

will satisfy

$$g_{2t} + g_{2t+1} = e, \quad t \in [n/2].$$

Remember that for all  $x \in \mathbb{F}_2^s$ , the graph  $\mathbf{G}_x(G)$  has a partition of  $\ell \geq 1$  disjoint simple cycles, that will be denoted by  $\mathbf{C}_x(G) = \{C_i\}_{i=0}^{\ell-1}$  (Definition 10). Fix  $\tau \in [n/2]$ . The first ingredient in the solution of  $FB-(2^s - 1, s, \lfloor \frac{5}{6} \cdot 2^{s-1} \rfloor - s)$  codes will be presented in algorithm  $FBSolution2(\tau, M)$ , which is presented as Algorithm 5.

---

### Algorithm 5 $FBSolution2(\tau, M)$

---

```

1:  $G^{(0)} \leftarrow G$ 
2: for  $t = 0, \dots, \tau - 1$  do
3:   for all  $2t \leq p, h \leq n - 1$  and  $t \leq m \leq \tau - 1$  do
4:     Swap  $g_p^{(t)}$  and  $g_{2t}^{(t)}$ 
5:     Swap  $g_h^{(t)}$  and  $g_{2t+1}^{(t)}$ 
6:     if  $g_{2t}^{(t)} + g_{2t+1}^{(t)} = v_m$  then
7:       Swap between  $v_t$  and  $v_m$  in  $M$ 
8:       Go to Step 2
9:      $\mathbf{a}_m \leftarrow v_m + g_{2t}^{(t)} + g_{2t+1}^{(t)}$ 
10:    Let  $C_i \in \mathbf{C}_{\mathbf{a}_m}(G)$  be a cycle s.t.  $\{g_{2t}^{(t)}, g_{2t+1}^{(t)}\}_{\mathbf{P}} \in C_i$ 
11:    if  $\{g_{2\ell}^{(t)}, g_{2\ell+1}^{(t)}\}_{\mathbf{P}} \in C_i$  s.t.  $\ell > t$  then
12:      Swap between  $v_t$  and  $v_m$  in  $M$ 
13:      Go to Step 16
14:     $G^{(t)}, M \leftarrow \text{BadCaseCorrection}(G^{(t)}, \mathbf{a}_m, M)$ 
15:    Go to Step 3
16:     $j \leftarrow \text{FindShortPath}(G^{(t)}, \mathbf{a}_m, t, \ell)$ 
17:     $G^{(t+1)} \leftarrow \mathcal{F}_{\mathbf{a}_t}(g_{2t+1}^{(t)}, g_j^{(t)})$ 
18: Return  $G^{(\tau)}$ 

```

---

In the internal routine starting on Step 3, on its  $t$ -th iteration, the algorithm will try to find two column vectors  $g_p^{(t)}$  and  $g_h^{(t)}$ , such that  $h, p \geq 2t$ , and a request  $v_m$ , where  $m \geq t$ ,

such that the sum of  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$  could be updated to  $\mathbf{v}_m$ , without corrupting the sums  $\mathbf{g}_{2t'}^{(t+1)} + \mathbf{g}_{2t'+1}^{(t+1)}$  such that  $t' < t$ . Our first task is to prove that if  $\tau \leq n/4$  then the algorithm will always find such  $\mathbf{g}_p^{(t)}$ ,  $\mathbf{g}_h^{(t)}$  and a request  $\mathbf{v}_m$ . In this case, the algorithm will provide  $2^{s-2}$  (when  $\tau = n/4$ ) requests and will never reach Step 14. Our second task is to prove that for the case  $n/4 < \tau \leq \lfloor n/3 \rfloor$ , the algorithm may reach Step 14, however by using the *BadCaseCorrection*( $G^{(t)}$ ,  $\mathbf{a}_m$ ,  $M$ ) function, which reorders the columns of  $G^{(t)}$ , the algorithm will succeed to construct  $\lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor$  recovery sets of size 2 (when  $\tau = \lfloor n/3 \rfloor$ ). Therefore, we are left to show how to construct additional  $\lfloor \frac{1}{6} \cdot 2^{s-1} \rfloor - s + 1$  recovery sets that will be of size 4 (remember that  $\lfloor x + y \rfloor \leq \lfloor x \rfloor + \lfloor y \rfloor + 1$ ). This part will be handled by the *FBSolution3*( $G, \tau, M$ ) algorithm. We notice that the *FBSolution3*( $G, \tau, M$ ) algorithm will be invoked only if

$$\left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor \leq \left\lfloor \frac{5}{6} \cdot 2^{s-1} \right\rfloor - s, \quad (7)$$

which holds for  $s \geq 7$ .

#### A. The $\tau \leq n/4$ Case

Before proving the correctness of the *FBSolution2*( $\tau, M$ ) algorithm, we start with an important definition.

*Definition 27:* On the  $t$ -th iteration, a path  $P_x$  between  $\mathbf{g}_p^{(t)}$ ,  $p \geq 2t$  and  $\mathbf{g}_h^{(t)}$ ,  $h \geq 2t$  will be called a **short-path** in  $\mathbf{G}_x(G^{(t)})$ , if it is a good-path, and all the pair-type edges  $\{\mathbf{g}_{2t'}^{(t)}, \mathbf{g}_{2t'+1}^{(t)}\}_P$  on  $P_x$  satisfy  $t' < t$ . The short-path  $P_{\mathbf{g}_p^{(t)} + \mathbf{g}_h^{(t)}}$  is called a trivial short-path.

Our first goal is to show that every  $\mathbf{g}_p^{(t)}$ ,  $p \geq 2t$  has  $n$  different short-paths ending on  $n$  columns  $\mathbf{g}_h^{(t)}$ ,  $h \geq 2t$ .

*Claim 7:* Fix some  $\mathbf{g}_p^{(t)}$  such that  $p \geq 2t$ . Then, for each  $\mathbf{G}_x(G^{(t)})$ , there exists  $\mathbf{g}_h^{(t)}$ ,  $h \geq 2t$ , such that there is a short-path between  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$ .

*Proof:* Given  $\mathbf{g}_p^{(t)}$  such that  $p \geq 2t$ , its pair  $\mathbf{g}_{p'}^{(t)}$  also satisfies  $p' \geq 2t$ . In Claim 3 we proved that for all  $\mathbf{x} \in \mathbb{F}_2^s$  every pair  $\{\mathbf{g}_{2m}, \mathbf{g}_{2m+1}\}_P$  has a good-path  $P_x(\mathbf{g}_{2m}, \mathbf{g}_{2m+1})$  in  $\mathbf{G}_x(G)$ . Therefore, by Claim 3, for all  $\mathbf{x} \in \mathbb{F}_2^s$ , there is a good-path  $P_x(\mathbf{g}_p^{(t)}, \mathbf{g}_{p'}^{(t)})$  in  $\mathbf{G}_x(G^{(t)})$ . If for all the edges  $\{\mathbf{g}_{2m}^{(t)}, \mathbf{g}_{2m+1}^{(t)}\}_P$  on  $P_x(\mathbf{g}_p^{(t)}, \mathbf{g}_{p'}^{(t)})$  it holds that  $m < t$ , then  $P_x(\mathbf{g}_p^{(t)}, \mathbf{g}_{p'}^{(t)})$  is a short-path. Otherwise, there exists an edge  $\{\mathbf{g}_{2m}^{(t)}, \mathbf{g}_{2m+1}^{(t)}\}_P$  on  $P_x(\mathbf{g}_p^{(t)}, \mathbf{g}_{p'}^{(t)})$ , such that  $m \geq t$ , and without loss of generality, we assume that this edge is the closest one to  $\mathbf{g}_p^{(t)}$  on  $P_x(\mathbf{g}_p^{(t)}, \mathbf{g}_{p'}^{(t)})$ . Let  $h \in \{2m, 2m+1\}$  such that the column  $\mathbf{g}_h^{(t)}$  is the closest node between  $\mathbf{g}_{2m}^{(t)}$  and  $\mathbf{g}_{2m+1}^{(t)}$  to  $\mathbf{g}_p^{(t)}$  on  $P_x(\mathbf{g}_p^{(t)}, \mathbf{g}_{p'}^{(t)})$ . Therefore  $h \geq 2t$  and this sub-path is a short-path by definition. ■

Next, we proceed to prove the correctness of the *FBSolution2*( $\tau, M$ ) algorithm. On Step 9, the algorithm will execute  $\mathbf{a}_m = \mathbf{v}_m + \mathbf{g}_p^{(t)} + \mathbf{g}_h^{(t)}$ . If  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$  have a non-trivial short-path between them in  $\mathbf{G}_{\mathbf{a}_m}(G^{(t)})$ , then our technique cannot update the sum of  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$  to be equal to  $\mathbf{v}_m$  without changing the sum of a pair  $\{\mathbf{g}_{2t'}^{(t)}, \mathbf{g}_{2t'+1}^{(t)}\}_P$  for some  $t' < t$ . So our goal is to find columns  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$  such that there are no (non-trivial) short-paths between them. We state

in the following claim that reordering the columns  $\mathbf{g}_p^{(t)}$  for  $p \geq 2t$  of  $G^{(t)}$  does not affect their short-paths.

*Claim 8:* The columns  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$  (before Steps 4–5) have no (non-trivial) short-paths between them, if and only if the columns  $\mathbf{g}_{2t}^{(t)}$  and  $\mathbf{g}_{2t+1}^{(t)}$  (after Steps 4–5) have no (non-trivial) short-paths between them.

*Proof:* We will prove only the first direction, while the second is proved similarly. In Claim 7 we proved that for all  $\mathbf{x} \in \mathbb{F}_2^s$ , there exists  $\mathbf{g}_{p'}^{(t)}$ ,  $p' \geq 2t$  such that there is a short-path between  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_{p'}^{(t)}$ . In this claim we assume that every such  $p'$  satisfies  $p' \neq h$ . By definition of the short-path, the edge  $\{\mathbf{g}_{2t}^{(t)}, \mathbf{g}_{2t+1}^{(t)}\}_P$  is not on any of these short-paths. Therefore, for all  $\mathbf{x} \in \mathbb{F}_2^s$ , the execution of Step 4 will not affect these short-paths. Similarly, for all  $\mathbf{x} \in \mathbb{F}_2^s$  the short-paths between  $\mathbf{g}_h^{(t)}$  and  $\mathbf{g}_{h'}^{(t)}$  will not be affected by the execution of Step 5 ( $h'$  is defined similarly to  $h$ ). Thus, the columns  $\mathbf{g}_{2t}^{(t)}$  and  $\mathbf{g}_{2t+1}^{(t)}$  (after Steps 4–5) will not have any (non-trivial) short-path between them. ■

Using Claim 8, we can make columns  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$  to be a pair. This is done by Steps 4 and 5, i.e., this pair is now  $\{\mathbf{g}_{2t}^{(t)}, \mathbf{g}_{2t+1}^{(t)}\}_P$ . In the next lemma we will use the properties of short-paths to prove the correctness of the algorithm.

*Lemma 28:* On the  $t$ -th iteration, if there are no (non-trivial) short-paths between  $\mathbf{g}_{2t}^{(t)}$  and  $\mathbf{g}_{2t+1}^{(t)}$ , then by the end of this iteration it holds that

$$\mathbf{g}_{2t}^{(t+1)} + \mathbf{g}_{2t+1}^{(t+1)} = \mathbf{v}_m,$$

and all the pair sums  $\mathbf{g}_{2t'}^{(t+1)} + \mathbf{g}_{2t'+1}^{(t+1)}$  such that  $t' < t$  will be unchanged.

*Proof:* If  $\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)} = \mathbf{v}_m$ , then due to Step 6 this lemma is correct. Otherwise, we know that there are no (non-trivial) short-paths between  $\mathbf{g}_{2t}^{(t)}$  and  $\mathbf{g}_{2t+1}^{(t)}$ . Therefore, Step 11 will succeed to find  $\{\mathbf{g}_{2\ell}^{(t)}, \mathbf{g}_{2\ell+1}^{(t)}\}_P \in C_i$  such that  $\ell > t$ . Thus, there is a good-path between  $\mathbf{g}_{2t+1}^{(t)}$  and one of the nodes  $\mathbf{g}_{2\ell}^{(t)}, \mathbf{g}_{2\ell+1}^{(t)}$  (the closest one between them to  $\mathbf{g}_{2t+1}^{(t)}$ ), and the index of this node is denoted by  $j$  (Step 16). By executing  $\mathcal{F}_{\mathbf{a}_m}(\mathbf{g}_{2t+1}, \mathbf{g}_j)$ , the matrix  $G^{(t)}$  is updated to a matrix  $G^{(t+1)}$  such that only the two following pair summations are correctly changed to

$$\begin{aligned} \mathbf{g}_{2t}^{(t+1)} + \mathbf{g}_{2t+1}^{(t+1)} &= \mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)} + \mathbf{a}_m = \mathbf{v}_m \\ \mathbf{g}_{2\ell}^{(t+1)} + \mathbf{g}_{2\ell+1}^{(t+1)} &= \mathbf{g}_{2\ell}^{(t)} + \mathbf{g}_{2\ell+1}^{(t)} + \mathbf{a}_m. \end{aligned}$$

Next, we will show that if  $t < n/4$  then the algorithm will find  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$  with no (non-trivial) short-paths between them.

*Lemma 29:* If  $t < n/4$ , then on the  $t$ -th iteration the algorithm will find  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$  with no (non-trivial) short-paths between them.

*Proof:* Fix some  $\mathbf{g}_p^{(t)}$ . By Claim 7, for each  $\mathbf{G}_x(G)$ , there exists  $\mathbf{g}_h^{(t)}$  such that there is a short-path between  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$ . Therefore  $\mathbf{g}_p^{(t)}$  has  $n$  different short-paths. Since  $t < n/4$  or  $2t < n/2$ , there are at least  $n/2 + 1$  options for choosing  $\mathbf{g}_h^{(t)}$ , and each of them has a trivial short-path with  $\mathbf{g}_p^{(t)}$ . Therefore, we are left with  $n/2 - 1$  short-paths, and at least

$n/2 + 1$  column vectors  $\mathbf{g}_h^{(t)}$ . Thus, there is at least one of them that has no (non-trivial) short-path with  $\mathbf{g}_p^{(t)}$ . ■

We are now ready to conclude with the following theorem.

*Theorem 30:* For  $\tau = n/4$  the algorithm  $\text{FBSolution2}(\tau, M)$  will construct recovery sums for the first  $2^{s-2}$  requests of  $M$ .

*Proof:* Since  $\tau = n/4$ , by Lemma 29, on the  $t$ -th iteration the algorithm  $\text{FBSolution2}(\tau, M)$  will find  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$  with no (non-trivial) short-paths between them. Therefore, by Claim 8, after executing Steps 4 and 5, the columns  $\mathbf{g}_{2t}^{(t)}$  and  $\mathbf{g}_{2t+1}^{(t)}$  have no (non-trivial) short-paths between them. By Lemma 28, since there are no (non-trivial) short-paths between  $\mathbf{g}_{2t}^{(t)}$  and  $\mathbf{g}_{2t+1}^{(t)}$ , by the end of this iteration it holds that

$$\mathbf{g}_{2t}^{(t+1)} + \mathbf{g}_{2t+1}^{(t+1)} = \mathbf{v}_m,$$

and all the pair sums  $\mathbf{g}_{2t'}^{(t+1)} + \mathbf{g}_{2t'+1}^{(t+1)}$  such that  $t' < t$  will be unchanged. ■

*Remark 1:* By using this technique we can obtain a solution for the case in which  $M$  has exactly two kinds of vectors ( $d = 2$ ). Suppose that  $\mathbf{v}_1$  appears  $t_1$  times and  $\mathbf{v}_2$  appears  $t_2$  times in  $M$ , and assume that  $t_1 \geq t_2$ . Then, initialize  $G$  such that each pair of columns will be summed up to  $\mathbf{v}_1$  and use this technique to update the first  $t_2 \leq n/4$  pairs to  $\mathbf{v}_2$ .

We proceed to the second case, i.e.,  $\tau \leq \lfloor n/3 \rfloor$ .

### B. The $\tau \leq \lfloor n/3 \rfloor$ Case

If  $n/4 < t \leq \lfloor n/3 \rfloor$ , then the algorithm may not be able to find  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$  with no (non-trivial) short-paths between them. However, at least one of these pairs will have at most one (non-trivial) short-path between them. Therefore, we may not be able to make some requests on the  $t$ -th iteration and reach Step 14. However, if we are left to handle more than one kind of request in  $M$ , we will succeed in this iteration. For that, we present the  $\text{BadCaseCorrection}(G^{(t)}, \mathbf{a}_t, M)$  function.

---

#### $\text{BadCaseCorrection}(G^{(t)}, \mathbf{a}_t, M)$

---

- 1: Let  $C_i \in \mathcal{C}_{\mathbf{a}_t}(G)$  be a cycle s.t.  $\{\mathbf{g}_{2t}^{(t)}, \mathbf{g}_{2t+1}^{(t)}\}_{\mathbf{p}} \in C_i$
  - 2: Find  $\{\mathbf{g}_{2t'}^{(t)}, \mathbf{g}_{2t'+1}^{(t)}\}_{\mathbf{p}} \in C_i$  s.t.  $\mathbf{g}_{2t'}^{(t)} + \mathbf{g}_{2t'+1}^{(t)} \neq \mathbf{v}_t$
  - 3:  $j \leftarrow \text{FindShortPath}(G^{(t)}, \mathbf{a}_t, t, t')$
  - 4:  $G^{(t)} \leftarrow \mathcal{F}_{\mathbf{a}_t}(\mathbf{g}_{2t+1}^{(t)}, \mathbf{g}_j^{(t)})$
  - 5: Swap  $\mathbf{g}_{2t'}^{(t)}$  and  $\mathbf{g}_{2t}^{(t)}$
  - 6: Swap  $\mathbf{g}_{2t'+1}^{(t)}$  and  $\mathbf{g}_{2t+1}^{(t)}$
  - 7: Swap  $\mathbf{v}_t$  and  $\mathbf{v}_{t'}$  in  $M$
  - 8: Return  $G^{(t)}, M$
- 

*Lemma 31:* On the  $t$ -th iteration such that  $t < \lfloor n/3 \rfloor$ , there exist  $p, h \geq 2t$  such that there is at most one (non-trivial) short-path between  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_h^{(t)}$ , and by the end of this iteration it holds that

$$\mathbf{g}_{2t}^{(t+1)} + \mathbf{g}_{2t+1}^{(t+1)} = \mathbf{v}_m,$$

and  $t$  recovery sums are satisfied.

*Proof:* Fix some  $\mathbf{g}_p^{(t)}$ . We already proved that  $\mathbf{g}_p^{(t)}$  has  $n$  different short-paths. Since  $t < \lfloor n/3 \rfloor$  or  $2t < \lfloor 2n/3 \rfloor$ ,

there are at least  $\lfloor n/3 \rfloor + 1$  options for choosing  $\mathbf{g}_h^{(t)}$ , and each one of them has a trivial short-path with  $\mathbf{g}_p^{(t)}$ . Therefore, we left with  $\lfloor 2n/3 \rfloor - 1$  short-paths and  $\lfloor n/3 \rfloor + 1$  column vectors  $\mathbf{g}_h^{(t)}$ . Therefore, in the worst case, there is a column vector  $\mathbf{g}_h^{(t)}$  such that there is exactly one (non-trivial) short-path  $P_x$  between  $\mathbf{g}_h^{(t)}$  and  $\mathbf{g}_p^{(t)}$ . If  $\mathbf{x} \neq \mathbf{g}_p^{(t)} + \mathbf{g}_h^{(t)} + \mathbf{v}_t$ , Step 11 will succeed, and we can construct a recovery set of size 2 for  $\mathbf{v}_t$  on the  $t$ -th iteration of the  $\text{FBSolution2}(\tau, M)$  algorithm, as proved in Lemma 28. If  $\mathbf{x} = \mathbf{g}_p^{(t)} + \mathbf{g}_h^{(t)} + \mathbf{v}_t$ , we cannot obtain  $\mathbf{v}_t$  on the  $t$ -th iteration. So from now we assume that  $\mathbf{x} = \mathbf{g}_p^{(t)} + \mathbf{g}_h^{(t)} + \mathbf{v}_t$ .

Assume that we have at least two different requests  $\mathbf{v}_j, j \geq t$  in  $M$ , and let  $\mathbf{v}_m \neq \mathbf{v}_t$ , for  $m \geq t$ . We prove that there does not exist a short-path  $P_y$  such that  $\mathbf{y} = \mathbf{g}_p^{(t)} + \mathbf{g}_h^{(t)} + \mathbf{v}_m$ . Therefore, Step 11 will succeed to find  $\{\mathbf{g}_{2\ell}^{(t)}, \mathbf{g}_{2\ell+1}^{(t)}\}_{\mathbf{p}} \in C_i$  such that  $\ell > t$ . As proved in Lemma 28 the algorithm  $\text{FBSolution2}(\tau, M)$  will construct a recovery set of size 2 for  $\mathbf{v}_m$  on the  $t$ -th iteration.

We are left with considering the case where all the requests  $\mathbf{v}_m$  in  $M$  for  $m \geq t$  are identical (and thus  $\mathbf{v}_m = \mathbf{v}_t$ ). Also assume that  $\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)} \neq \mathbf{v}_t$  which means that Step 6 has failed (otherwise this iteration will succeed). In this case the algorithm will use its  $\text{BadCaseCorrection}(G^{(t)}, \mathbf{a}_m, M)$  function. Let  $C_i \in \mathcal{C}_{\mathbf{a}_m}(G)$  be a cycle such that  $\{\mathbf{g}_{2t}^{(t)}, \mathbf{g}_{2t+1}^{(t)}\}_{\mathbf{p}} \in C_i$  (Step 1). First assume that for every pair-type edge  $\{\mathbf{g}_{2t'}^{(t)}, \mathbf{g}_{2t'+1}^{(t)}\}_{\mathbf{p}} \in C_i$  such that  $t' < t$  it holds that

$$\mathbf{g}_{2t'}^{(t)} + \mathbf{g}_{2t'+1}^{(t)} = \mathbf{v}_t.$$

Note that if a path

$$\mathbf{g}_{2t'}^{(t)} - \mathbf{g}_{2t'+1}^{(t)} - \mathbf{g}_{2t}^{(t)} - \mathbf{g}_{2t+1}^{(t)}$$

holds

$$\mathbf{g}_{2t'}^{(t)} + \mathbf{g}_{2t'+1}^{(t)} = \mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)} = \mathbf{v}_t,$$

then it has to be a cycle of length 4. Therefore, in our case,

$$\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)} = \mathbf{g}_{2t'}^{(t)} + \mathbf{g}_{2t'+1}^{(t)} = \mathbf{v}_t,$$

which is a contradiction to the assumption that  $\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)} \neq \mathbf{v}_t$ . Otherwise, we can assume that there is an edge  $\{\mathbf{g}_{2t'}^{(t)}, \mathbf{g}_{2t'+1}^{(t)}\}_{\mathbf{p}} \in C_i, t' < t$  such that

$$\mathbf{g}_{2t'}^{(t)} + \mathbf{g}_{2t'+1}^{(t)} = \mathbf{v}_{t'} \neq \mathbf{v}_t.$$

By executing Step 4,  $\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}$  will be updated to  $\mathbf{v}_t$ , which corrupts the sum of the pair  $\{\mathbf{g}_{2t'}^{(t)}, \mathbf{g}_{2t'+1}^{(t)}\}_{\mathbf{p}}$  such that

$$\mathbf{g}_{2t'}^{(t)} + \mathbf{g}_{2t'+1}^{(t)} = \mathbf{v}_{t'} + \mathbf{a}_t \neq \mathbf{v}_{t'}.$$

After executing Steps 5–6, we obtain two kinds of requests in  $M$  that are left to deal with, while still having  $t - 1$  valid recovery sets. In this case, the algorithm will return to Step 3. Since now we have two kinds of requests, as already proved, the algorithm will be able to construct a recovery set for either  $\mathbf{v}_t$  or  $\mathbf{v}_{t'}$  on the  $t$ -th iteration. ■

The next theorem follows directly from Lemma 31.

*Theorem 32:* If  $\tau \leq \lfloor n/3 \rfloor$ , then the algorithm  $\text{FBSolution2}(\tau, M)$  will construct recovery sums for the first  $\lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor$  requests of  $M$ .

Due to Theorem 32, we proved that the algorithm  $FBSolution2(\tau, M)$  provides an alternative construction for  $FB-(2^s - 1, s, \lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor)$  codes. However, the algorithm  $FBSolution2(\tau, M)$  is better than the algorithm  $FBSolution(\tau, M)$  since by using the algorithm  $FBSolution2(\tau, M)$ , all the recovery sets are of length 2 (and not 4). Therefore, we are left with  $\lceil n/3 \rceil$  unused column vectors. If (7) holds, we will put  $\lceil n/3 \rceil - 2$  of these columns, except the redundant columns  $\mathbf{g}_{n-2}, \mathbf{g}_{n-3}$ , as the first columns of the  $ES$ -matrix  $G$ , and similarly, the left requests of  $M$  that have no recovery sets yet are placed first. In the next section, we show how to obtain  $\lfloor \frac{1}{6} \cdot 2^{s-1} \rfloor - s + 1$  more recovery sets of size (at most) 4 from these  $\lceil n/3 \rceil$  unused columns of  $G$ .

### C. Constructing Recovery Sets of Size 4

According to the previous results as stated in Theorem 3(c), and due to (7), we can assume that  $s \geq 7$ . The  $FBSolution(\tau, M)$  algorithm uses the initialized  $ES$ -matrix  $G^{(0)}$  that satisfies (2). In fact, we can construct a similar algorithm that is initialized by any arbitrary  $ES$ -matrix  $G$ . Let  $\tau = \lfloor \frac{1}{6} \cdot 2^{s-1} \rfloor - s$ . The value of  $\tau$  represents the number of requests that will be handled. Note that

$$\begin{aligned} 4\tau &= 4 \left\lfloor \frac{1}{6} \cdot 2^{s-1} \right\rfloor - 4s \\ &= 4 \left\lfloor \frac{1}{12} \cdot 2^s \right\rfloor - 4s \leq \left\lfloor \frac{1}{3} \cdot 2^s \right\rfloor - 2, \end{aligned}$$

for  $s \geq 7$ , which is the number of unused columns in  $G$ . Our goal is to use either 2 or 4 columns of  $G$  for every recovery set. In other words, every  $\mathbf{v}_t$  will be equal to either  $\mathbf{g}_{4t}^{(t)} + \mathbf{g}_{4t+1}^{(t)}$  or  $\mathbf{g}_{4t}^{(t)} + \mathbf{g}_{4t+1}^{(t)} + \mathbf{g}_{4t+2}^{(t)} + \mathbf{g}_{4t+3}^{(t)}$ . To show this property we will prove that in every step of the algorithm, we have to have at least  $2(s+1)$  unused (or redundant) columns of  $G$ .

We start with the next definition, which is based on the fact that every  $s+1$  vectors in  $\mathbb{F}_2^s$  have a subset of  $h \leq s+1$  linearly dependent vectors.

**Definition 33:** Given an  $ES$ -matrix  $G^{(t)}$ , denote the set  $S_h^{(t)} \subseteq [s+1]$  of size  $h \leq s+1$ , such that

$$\sum_{i \in S_h^{(t)}} \left( \mathbf{g}_{4t+2i}^{(t)} + \mathbf{g}_{4t+2i+1}^{(t)} \right) = \mathbf{0}_s. \quad (8)$$

Denote the  $Reorder(G^{(t)})$  procedure that swaps arbitrarily between the columns of  $G^{(t)}$  presented in (8) and the columns indexed by  $\{4t \leq m \leq 4t+2h-1\}$  in  $G^{(t)}$  and returns the reordered matrix and  $h$  as an output.

By using the  $Reorder(G^{(t)})$  procedure which is defined in Definition 33, we can assume that

$$\sum_{i \in [h]} \left( \mathbf{g}_{4t+2i}^{(t)} + \mathbf{g}_{4t+2i+1}^{(t)} \right) = \mathbf{0}_s. \quad (9)$$

We are now ready to show the  $FBSolution3(G, \tau, M)$  algorithm, which is presented as Algorithm 6.

Note that since

$$\begin{aligned} \left\lfloor \frac{1}{3} \cdot 2^s \right\rfloor - 2 - 4\tau &= \left\lfloor \frac{1}{3} \cdot 2^s \right\rfloor - 4 \left\lfloor \frac{1}{12} \cdot 2^s \right\rfloor + 4s - 2 \\ &\geq 4s - 2 \geq 2(s+1), \end{aligned}$$

---

#### Algorithm 6 $FBSolution3(G, \tau, M)$

---

```

1:  $G^{(0)} \leftarrow G$ 
2:  $\mathcal{B}^{(0)} = \emptyset$ 
3: for  $t = 0, 1, \dots, \tau - 1$  do
4:    $G^{(t)}, h \leftarrow Reorder(G^{(t)})$ 
5:    $G^{(t)} \leftarrow FindEquivSums(G^{(t)}, t, h)$ 
6:    $G^{(t+1)}, \mathcal{B}^{(t+1)} \leftarrow FindGoodOrBadRequest(G^{(t)}, \mathcal{B}^{(t)}, 2t, \mathbf{v}_t)$ 
7: Return  $G^{(\tau)}$  and  $\mathcal{B}^{(\tau)}$ 

```

---

for  $s \geq 7$ , the  $2h \leq 2(s+1)$  column vectors of  $G^{(t)}$  presented in (9) are unused on the  $t$ -th iteration. By using these  $2h$  unused columns, the function  $FindEquivSums(G, t, h)$  will be able to reorder the columns of  $G^{(t)}$  such that

$$\mathbf{g}_{4t}^{(t)} + \mathbf{g}_{4t+1}^{(t)} = \mathbf{g}_{4t+2}^{(t)} + \mathbf{g}_{4t+3}^{(t)}, \quad (10)$$

without changing the previous valid recovery sums. Then, the function  $FindGoodOrBadRequest(G^{(t)}, \mathcal{B}^{(t)}, 2t, \mathbf{v}_t)$  will update the sum  $\mathbf{g}_{4t}^{(t)} + \mathbf{g}_{4t+1}^{(t)}$  to either  $\mathbf{v}_t$  or  $\mathbf{g}_{4t}^{(t)} + \mathbf{g}_{4t+1}^{(t)} + \mathbf{v}_t$ , again, without changing all the previous valid recovery sums. In the latter case, we are able to construct a recovery set  $R_t = \{4t, 4t+1, 4t+2, 4t+3\}$  of size 4 due to (10).

The  $FindEquivSums(G, t, h)$  algorithm is presented next.

---

#### $FindEquivSums(G, t, h)$

---

```

1: for  $i = 1, \dots, h-1$  do
2:    $\mathbf{x}_i \leftarrow \mathbf{g}_{4t+2i} + \mathbf{g}_{4t+2i+1}$ 
3:   if  $\mathbf{g}_{4t} + \mathbf{g}_{4t+1} \neq \mathbf{x}_i$  then
4:      $G \leftarrow FindGoodOrBadRequest(G, \mathcal{B}, 2t, \mathbf{x}_i)$ 
5:   if  $\mathbf{g}_{4t} + \mathbf{g}_{4t+1} = \mathbf{x}_i$  then
6:     Swap  $\mathbf{g}_{4t+2}$  and  $\mathbf{g}_{4t+2i}$ 
7:     Swap  $\mathbf{g}_{4t+3}$  and  $\mathbf{g}_{4t+2i+1}$ 
8:   Return  $G$ 

```

---

The proof of the correctness of the function  $FindEquivSums(G, t, h)$  is shown in the following theorem.

**Theorem 34:** If there are at least  $2(s+1)$  unused columns in  $G$ , then there is a function  $FindEquivSums(G, t, h)$  that can reorder the columns of  $G$  such that

$$\mathbf{g}_{4t} + \mathbf{g}_{4t+1} = \mathbf{g}_{4t+2} + \mathbf{g}_{4t+3},$$

without corrupting the previous valid recovery sums.

*Proof:* As explained before, the  $2h \leq 2(s+1)$  column vectors of  $G$  presented in (9) are unused. Therefore, the algorithm  $FindEquivSums(G, t, h)$  will not corrupt the previous valid recovery sums. Our goal is to prove that the algorithm will succeed on Step 5. On the  $i$ -th iteration, by executing Step 4, the sum of  $\mathbf{g}_{4t} + \mathbf{g}_{4t+1}$  will be either  $\mathbf{g}_{4t} + \mathbf{g}_{4t+1} + \mathbf{x}_i$  or  $\mathbf{x}_i$ , without changing other sums except of the redundancy sum  $\mathbf{g}_{n-2} + \mathbf{g}_{n-1}$ . If  $\mathbf{g}_{4t} + \mathbf{g}_{4t+1} = \mathbf{x}_i$ , Step 5 will succeed. Otherwise, we can assume that on the  $i$ -th iteration, the algorithm obtains  $\mathbf{g}_{4t} + \mathbf{g}_{4t+1} + \mathbf{x}_i$  on Step 4. Denote by  $\mathbf{x}_0$  the sum of  $\mathbf{g}_{4t} + \mathbf{g}_{4t+1}$  at the beginning of the algorithm.

Therefore, at the end of the  $i$ -th iteration, we obtain

$$\mathbf{g}_{4t} + \mathbf{g}_{4t+1} = \sum_{j=0}^i \mathbf{x}_j.$$

By (9),

$$\sum_{j=0}^{h-2} \mathbf{x}_j = \mathbf{x}_{h-1}.$$

Therefore, on the last iteration, i.e., when  $i = h - 1$ ,

$$\mathbf{g}_{4t} + \mathbf{g}_{4t+1} = \sum_{j=0}^{h-2} \mathbf{x}_j = \mathbf{x}_{h-1}.$$

Thus, Step 3 will fail and Step 5 will succeed, concluding the proof.  $\blacksquare$

We are ready to show the main result of this section.

*Lemma 35:* The  $\text{FBSolution3}(G, \tau, M)$  algorithm constructs  $\tau$  valid recovery sets  $R_t$  for  $\mathbf{v}_t$ , without corrupting previous recovery sums.

*Proof:* By Theorem 34 the  $\text{FindEquivSums}(G^{(t)}, t, M)$  algorithm will output  $G^{(t)}$  such that  $\mathbf{g}_{4t}^{(t)} + \mathbf{g}_{4t+1}^{(t)} = \mathbf{g}_{4t+2}^{(t)} + \mathbf{g}_{4t+3}^{(t)}$ , and all previous sums of requests are valid. Before the execution of the  $\text{FindGoodOrBadRequest}(G^{(t)}, \mathcal{B}^{(t)}, 2t, \mathbf{v}_t)$  function we denote the sums  $\mathbf{g}_{4t} + \mathbf{g}_{4t+1}$  and  $\mathbf{g}_{4t+2} + \mathbf{g}_{4t+3}$  by  $\mathbf{y}_t$ . The algorithm  $\text{FindGoodOrBadRequest}(G^{(t)}, \mathcal{B}^{(t)}, 2t, \mathbf{v}_t)$  will update only the sum  $\mathbf{g}_{4t}^{(t)} + \mathbf{g}_{4t+1}^{(t)}$  to either  $\mathbf{v}_t$  or  $\mathbf{v}_t + \mathbf{y}_t$  and the sum of the last pair  $\mathbf{g}_{n-2} + \mathbf{g}_{n-3}$  which is redundant, due to Lemma 14. The case  $\mathbf{g}_{4t}^{(t)} + \mathbf{g}_{4t+1}^{(t)} = \mathbf{v}_t$  is called a good case and we assume that all such  $t$ 's are inserted in a set  $\mathcal{G}$ . These pairs will be recovered by the recovery sets  $R_t = \{4t, 4t+1\}$ . The case  $\mathbf{g}_{4t}^{(t)} + \mathbf{g}_{4t+1}^{(t)} = \mathbf{v}_t + \mathbf{y}_t$  is called a bad case and all such  $t$ 's are assumed to be inserted in a set  $\mathcal{B}$ . By (10) for every  $t$  such that  $\mathbf{g}_{4t}^{(t)} + \mathbf{g}_{4t+1}^{(t)} = \mathbf{v}_t + \mathbf{y}_t$  we have that  $\mathbf{g}_{4t+2}^{(t)} + \mathbf{g}_{4t+3}^{(t)} = \mathbf{y}_t$ . Thus, in these cases, the requests will have the recovery sets  $R_t = \{4t, 4t+1, 4t+2, 4t+3\}$ .  $\blacksquare$

In Section V, we showed a technique to obtain another recovery set from the redundancy pair  $\mathbf{g}_{n-2}$  and  $\mathbf{g}_{n-1}$ . Using this technique, we are able to construct  $\lfloor \frac{1}{6} \cdot 2^{s-1} \rfloor - s + 1$  valid recovery sets. By combining the three cases above, the following theorem is deduced immediately.

*Theorem 36:* An  $FB-(2^s - 1, s, \lfloor \frac{5}{6} \cdot 2^{s-1} \rfloor - s)$  code exists.

## VII. A CONSTRUCTION OF $B-(2^s - 1, s, 2^{s-1})$ CODES

Wang *et al.* [14] showed a construction for  $B-(2^s - 1, s, 2^{s-1})$  codes, which is optimal, using a recursive decoding algorithm. In this section, we show how to achieve this result with the simpler, non-recursive decoding algorithm. Our solution solves even a more general case in which the requests  $\mathbf{v}_j$ 's satisfy some constraint that will be described later in this section. The idea of this algorithm is similar to the one of the  $\text{FBSolution}(\tau, M)$  algorithm. First, we slightly change the definition of an extended Simplex solution as presented in Definition 6 to be the following one.

*Definition 37:* Let  $M = [\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{n/2-1}]$  be a request of order  $s \times n/2$ , where  $n = 2^s$ . The matrix  $M$  has an

**extended Simplex solution** if there exists an  $ES$ -matrix  $G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{n-1}]$  of order  $s \times n$  such that for all  $i \in [n/2 - 1]$ ,

$$\mathbf{v}_i = \mathbf{g}_{2i} + \mathbf{g}_{2i+1},$$

and for  $i = n/2 - 1$  either  $\mathbf{v}_i = \mathbf{g}_{n-2} + \mathbf{g}_{n-1}$ , or  $\mathbf{v}_i = \mathbf{g}_{n-2}$ , or  $\mathbf{v}_i = \mathbf{g}_{n-1}$ . In this case, we say that  $G$  is an extended Simplex solution for  $M$ .

Let  $G$  be an  $ES$ -matrix. Let  $\mathbf{G}$  be the set of all matrices  $G'$  generated by elementary row operations on  $G$ . The following claim proves that elementary row operations on  $ES$ -matrices only reorder their column vectors.

*Claim 9:* Every  $G \in \mathbf{G}$  is an  $ES$ -matrix.

*Proof:* We will only prove that adding a row in  $G$  to any other row, generates an  $ES$ -matrix. By proving that, it can be inductively proved that doing several such operations will again yield an  $ES$ -matrix.

Without loss of generality, we assume that we add the  $i$ -th row, for some  $0 < i \leq s - 1$ , to the 0-th row of  $G$  and generate a new matrix  $G'$ . Assume to the contrary that  $G'$  is not an  $ES$ -matrix. Thus, there are two distinct indices  $\ell, m \in [n]$  such that  $\mathbf{g}'_\ell = \mathbf{g}'_m$ . Therefore, by definition of elementary row operations,  $G$  satisfies  $\mathbf{g}_\ell = \mathbf{g}_m$ , which is a contradiction.  $\blacksquare$

Let  $M$  be a request denoted by

$$M = [\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{2^{s-1}-1}].$$

Let  $\mathbf{M}$  be the set of all matrices  $M'$  generated by elementary row operations on request matrix  $M$ . We now present Lemma 38. Its proof follows directly from Claim 9.

*Lemma 38:* If there is an  $M \in \mathbf{M}$  such that there is an extended Simplex solution for  $M$ , then there is an extended Simplex solution for all  $M' \in \mathbf{M}$ .

*Proof:* Let  $M \in \mathbf{M}$  and let  $G$  be an extended Simplex solution for  $M$ . Let  $P$  be a set of elementary row operations, generating  $M'$  from  $M$ . By Claim 9, executing elementary row operations  $P$  on  $G$  generates an  $ES$ -matrix,  $G'$ . Since we applied the same elementary row operations  $P$  on both  $M$  and  $G$ , it is deduced that  $G'$  is an extended Simplex solution for  $M'$ .  $\blacksquare$

The constraint mentioned above is as follows. Given  $\mathbf{M}$ , we demand that there is a request  $M' \in \mathbf{M}$  having the 0-th row to be a vector of ones. Using Lemma 38, our algorithm will handle any request  $M'$  such that  $M' \in \mathbf{M}$  and  $\mathbf{M}$  holds this constraint. Note that if each request vector  $\mathbf{v}_j \in \mathbb{F}_2^s$  is a unit vector, then by summing up all of its rows to the 0-th one, it holds that there exists such a matrix in  $\mathbf{M}$  holds the constraint. Moreover, if every request vector is of odd Hamming weight, our algorithm will still find a solution. Therefore, from now on, we assume that the 0-th row of the request matrix  $M$  is a vector of ones.

Remember that  $\mathbf{e} = (0, 0, \dots, 0, 1) \in \mathbb{F}_2^s$ . The initial state of the matrix

$$G = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{n-1}]$$

will satisfy

$$\mathbf{g}_{2t} + \mathbf{g}_{2t+1} = \mathbf{e}, \quad t \in [n/2].$$



Given  $\mathbf{g} \in \mathbb{F}_2^s$ , let  $f(\mathbf{g})$  be the bit value in the 0-th position in  $\mathbf{g}$ . Remember that for all  $\mathbf{x} \in \mathbb{F}_2^s$ , the graph  $\mathbf{G}_{\mathbf{x}}(G)$  has a partition to  $\ell \geq 1$  disjoint simple cycles, that will be denoted by  $\mathbf{C}_{\mathbf{x}}(G) = \{C_i\}_{i=0}^{\ell-1}$  (Definition 10). Let  $\tau = 2^{s-1}$ . We are now ready to show the following algorithm.

---

**Algorithm 7** *BSolution*( $\tau, M$ )

---

```

1:  $G^{(0)} \leftarrow G$ 
2: for  $t = 0, \dots, \tau - 2$  do
3:   if  $f(\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}) = 1$  then
4:     Find  $\mathbf{g}_p^{(t)}, \mathbf{g}_h^{(t)}$  s.t.  $p, h \geq 2t$  and  $f(\mathbf{g}_p^{(t)} + \mathbf{g}_h^{(t)}) = 0$ 
5:     Swap  $\mathbf{g}_p^{(t)}$  and  $\mathbf{g}_{2t}^{(t)}$ 
6:     Swap  $\mathbf{g}_h^{(t)}$  and  $\mathbf{g}_{2t+1}^{(t)}$ 
7:      $\mathbf{a}_t \leftarrow \mathbf{v}_t + \mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}$ 
8:     Let  $C_i \in \mathbf{C}_{\mathbf{a}_t}(G^{(t)})$  be a cycle s.t.  $\{\mathbf{g}_{2t}^{(t)}, \mathbf{g}_{2t+1}^{(t)}\}_{\mathbf{P}} \in C_i$ 
9:     Find  $\{\mathbf{g}_{2m}^{(t)}, \mathbf{g}_{2m+1}^{(t)}\}_{\mathbf{P}} \in C_i$  s.t.  $m > t$ 
10:     $j \leftarrow \text{FindShortPath}(G^{(t)}, \mathbf{a}_t, t, m)$ 
11:     $G^{(t+1)} \leftarrow \mathcal{F}_{\mathbf{a}_t}(\mathbf{g}_{2t}^{(t)}, \mathbf{g}_j^{(t)})$ 
12: if  $\mathbf{g}_{n-2}^{(\tau-1)} + \mathbf{g}_{n-1}^{(\tau-1)} \neq \mathbf{v}_{n/2-1}$  and  $\mathbf{g}_{n-2}^{(\tau-1)} \neq \mathbf{v}_{n/2-1}$  then
13:    $G^{(\tau)} \leftarrow \mathcal{F}_{\mathbf{g}_{n-2}^{(\tau-1)} + \mathbf{v}_{n/2-1}}(\mathbf{g}_{n-2}^{(\tau-1)}, \mathbf{g}_{n-1}^{(\tau-1)})$ 
14: Return  $G^{(\tau)}$ 

```

---

Our first goal is to prove that on the  $t$ -th iteration when the *BSolution*( $\tau, M$ ) algorithm reaches Step 7, it holds that  $f(\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}) = 0$ .

*Lemma 39:* On the  $t$ -th iteration when the *BSolution*( $\tau, M$ ) algorithm reaches Step 7, it holds that  $f(\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}) = 0$ .

*Proof:* If on the  $t$ -th iteration in Step 3,  $f(\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}) = 1$  then the *BSolution*( $\tau, M$ ) algorithm will try to find  $p, h \geq 2t$  such that  $f(\mathbf{g}_p^{(t)} + \mathbf{g}_h^{(t)}) = 0$ . Note that since  $t \leq \tau - 2$ , we have  $\mathbf{g}_{2t}^{(t)}, \mathbf{g}_{2t+1}^{(t)}$  and at least two more column vectors  $\mathbf{g}_{2m}^{(t)}, \mathbf{g}_{2m+1}^{(t)}$  such that  $m > t$ . By the pigeonhole principle, there exist two indices  $p, h \in \{2t, 2t+1, 2m, 2m+1\}$  such that  $f(\mathbf{g}_p^{(t)}) = f(\mathbf{g}_h^{(t)})$ . After executing Step 5 and Step 6 we obtain  $f(\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}) = 0$ . ■

Our next goal is to show that in Step 9 on the  $t$ -th iteration, the *BSolution*( $\tau, M$ ) algorithm will find  $\{\mathbf{g}_{2m}^{(t)}, \mathbf{g}_{2m+1}^{(t)}\}_{\mathbf{P}} \in C_i$  such that  $m > t$ . We start with the following claim.

*Claim 10:* Given a graph  $\mathbf{G}_{\mathbf{x}}(G)$  and its partition to cycles  $\mathbf{C}_{\mathbf{x}}(G) = \{C_i\}_{i=0}^{\ell-1}$ , for all  $C_i \in \mathbf{C}_{\mathbf{x}}(G)$  it holds that

$$\sum_{\{\mathbf{g}_{2m}, \mathbf{g}_{2m+1}\}_{\mathbf{P}} \in C_i} (\mathbf{g}_{2m} + \mathbf{g}_{2m+1}) \equiv \frac{1}{2} |C_i| \mathbf{x} \pmod{2}.$$

*Proof:* Assume that  $C_i$  is of length  $2\ell$ , and its cycle representation is given as follows

$$C_i = \mathbf{g}_{s_0} - \mathbf{g}_{s_1} - \dots - \mathbf{g}_{s_{2\ell-1}} - \mathbf{g}_{s_{2\ell}} - \mathbf{g}_{s_0},$$

where both of the edges  $\{\mathbf{g}_{s_0}, \mathbf{g}_{s_1}\}_{\mathbf{P}}, \{\mathbf{g}_{s_{2\ell-1}}, \mathbf{g}_{s_{2\ell}}\}_{\mathbf{P}}$  are pair-type edges. By Claim 2(3), for all odd  $t \in [2\ell]$  it holds that  $\mathbf{g}_{s_t} = \mathbf{g}_{s_{t+1}} + \mathbf{x}$ . Thus, by summing only the sums of the nodes of the pair-type edges in  $C_i$  we obtain

$$\sum_{\{\mathbf{g}_{2m}, \mathbf{g}_{2m+1}\}_{\mathbf{P}} \in C_i} (\mathbf{g}_{2m} + \mathbf{g}_{2m+1})$$

$$= \sum_{t \in [2\ell]} \mathbf{g}_{s_t} = \sum_{t \in [2\ell], t \text{ is odd}} (\mathbf{g}_{s_t} + \mathbf{g}_{s_t} + \mathbf{x}) \equiv \ell \mathbf{x} \pmod{2}.$$

Now we are ready to prove the following lemma. ■

*Lemma 40:* On the  $t$ -th iteration when the *BSolution*( $\tau, M$ ) algorithm reaches Step 9, it will find  $\{\mathbf{g}_{2m}^{(t)}, \mathbf{g}_{2m+1}^{(t)}\}_{\mathbf{P}} \in C_i$  such that  $m > t$ .

*Proof:* Remember that we assumed that the bit value in the 0-th position for all the requests  $\mathbf{v}_j$  is 1. Therefore, on the  $t$ -th iteration, for all  $m < t$  it holds that  $f(\mathbf{g}_{2m}^{(t)} + \mathbf{g}_{2m+1}^{(t)}) = 1$ , and by Lemma 39,  $f(\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}) = 0$ . Now, assume to the contrary that there are no  $m > t$  such that  $\{\mathbf{g}_{2m}^{(t)}, \mathbf{g}_{2m+1}^{(t)}\}_{\mathbf{P}} \in C_i$ . Note that

$$\begin{aligned} f(\mathbf{a}_t) &= f(\mathbf{v}_t + \mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}) \\ &= f(\mathbf{v}_t) + f(\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}) = 1 + 0 \equiv 1 \pmod{2}. \end{aligned}$$

By Claim 10, if  $|C_i| = 2\ell$  then,

$$\sum_{\{\mathbf{g}_{2m}, \mathbf{g}_{2m+1}\}_{\mathbf{P}} \in C_i} (\mathbf{g}_{2m}^{(t)} + \mathbf{g}_{2m+1}^{(t)}) \equiv \ell \mathbf{a}_t \pmod{2},$$

and therefore,

$$\sum_{\{\mathbf{g}_{2m}, \mathbf{g}_{2m+1}\}_{\mathbf{P}} \in C_i} f(\mathbf{g}_{2m}^{(t)} + \mathbf{g}_{2m+1}^{(t)}) \equiv \ell f(\mathbf{a}_t) = \ell \pmod{2}.$$

However, since only the edge  $\{\mathbf{g}_{2t}^{(t)}, \mathbf{g}_{2t+1}^{(t)}\}_{\mathbf{P}} \in C_i$  satisfies that  $f(\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}) = 0$ , it is deduced that

$$\sum_{\{\mathbf{g}_{2m}, \mathbf{g}_{2m+1}\}_{\mathbf{P}} \in C_i} f(\mathbf{g}_{2m}^{(t)} + \mathbf{g}_{2m+1}^{(t)}) = \ell - 1 \not\equiv \ell \pmod{2},$$

which violates Claim 10. ■

We are ready to show the main theorem of this section.

*Theorem 41:* Given a request matrix  $M$  having the 0-th row to be a vector of ones, the *BSolution*( $\tau, M$ ) algorithm finds an extended Simplex solution for  $M$ .

*Proof:* First, we will prove that the *BSolution*( $\tau, M$ ) algorithm generates  $2^{s-1} - 1$  recovery sets for the first  $2^{s-1} - 1$  requests  $\mathbf{v}_t$ . This is done by Steps 1–11. Note that the sums  $\mathbf{g}_{2m}^{(t)} + \mathbf{g}_{2m+1}^{(t)}$  for all  $m < t$ , might be changed only after Step 9. We will show that these sums will not be changed and the sum  $\mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)}$  will be equal to  $\mathbf{v}_t$  at the end of the  $t$ -th iteration. By Lemma 40, when the *BSolution*( $\tau, M$ ) algorithm reaches Step 9, it will find  $\{\mathbf{g}_{2m}^{(t)}, \mathbf{g}_{2m+1}^{(t)}\}_{\mathbf{P}} \in C_i$  such that  $m > t$ . Thus, there is a good-path between  $\mathbf{g}_{2t+1}$  and one of the nodes  $\mathbf{g}_{2m}, \mathbf{g}_{2m+1}$  (the closest one between them to  $\mathbf{g}_{2t+1}$ ), and the index of this node is denoted by  $j$  (Step 10). Due to Lemma 13, by executing  $\mathcal{F}_{\mathbf{a}_t}(\mathbf{g}_{2t+1}, \mathbf{g}_j)$ , the matrix  $G^{(t)}$  is updated to a matrix  $G^{(t+1)}$  such that only the two following pair summations are correctly changed to

$$\begin{aligned} \mathbf{g}_{2t}^{(t+1)} + \mathbf{g}_{2t+1}^{(t+1)} &= \mathbf{g}_{2t}^{(t)} + \mathbf{g}_{2t+1}^{(t)} + \mathbf{a}_t = \mathbf{v}_t \\ \mathbf{g}_{2m}^{(t+1)} + \mathbf{g}_{2m+1}^{(t+1)} &= \mathbf{g}_{2m}^{(t)} + \mathbf{g}_{2m+1}^{(t)} + \mathbf{a}_t. \end{aligned}$$

Lastly, Steps 12–13 handle the last recovery set in a similar way as was done in the proof of Theorem 22. ■

## VIII. CONCLUSION

In this paper, functional  $k$ -batch codes and the value  $FB(s, k)$  were studied. It was shown that for all  $s \geq 6$ ,  $FB(s, \lfloor \frac{5}{6}2^{s-1} \rfloor - s) \leq 2^s - 1$ . In fact, we believe that by using a similar technique, this result can be improved to  $\lfloor \frac{7}{8}2^{s-1} \rfloor - s$  requests, but this proof has many cases and thus it is left for future work. We also showed a family of  $FB-(2^s + \lceil (3\alpha - 2) \cdot 2^{s-2} \rceil - 1, s, \lfloor \alpha \cdot 2^{s-1} \rfloor)$  codes for all  $2/3 \leq \alpha \leq 1$ . Yet another result in the paper provides an optimal solution for  $k = 2^s$  which is  $FB(s, 2^s) = 2^{s+1} - 2$ . While the first and main result of the paper significantly improves upon the best-known construction in the literature, there is still a gap to the conjecture which claims that  $FB(s, 2^{s-1}) = 2^s - 1$ . We believe that the conjecture indeed holds true and it can be achieved using extended Simplex codes.

## APPENDIX A

*Theorem 1:* For any triple-set  $(\mathcal{G}, \mathcal{B}, \mathcal{R})$  if  $|\mathcal{B}| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$  then  $|\mathcal{B}| \leq |\mathcal{R}|$ .

*Proof:* According to the definition of  $(\mathcal{G}, \mathcal{B}, \mathcal{R})$  and since  $\mathcal{G} \cup \mathcal{B} = \lfloor \frac{2}{3} \cdot 2^{s-1} \rfloor$  it holds that

$$\begin{aligned} |\mathcal{R}| &= \left| [2^{s-1}] \setminus \left( \mathcal{G} \cup \mathcal{B} \cup \{2^{s-1} - 1\} \right) \right| \\ &= 2^{s-1} - \left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor - 1. \end{aligned}$$

Thus, in order to prove that  $|\mathcal{B}| \leq |\mathcal{R}|$ , since  $|\mathcal{B}| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$ , we will prove inequality (a) in

$$|\mathcal{R}| = 2^{s-1} - \left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor - 1 \stackrel{(a)}{\geq} \left\lfloor \frac{1}{3} \cdot 2^{s-1} \right\rfloor \geq |\mathcal{B}|.$$

This inequality is equivalent to

$$2^{s-1} - 1 \geq \left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor + \left\lfloor \frac{1}{3} \cdot 2^{s-1} \right\rfloor.$$

We separate the proof for the following two cases.

**Case 1:** If  $s$  is even, then

$$2^s \equiv 1 \pmod{3}, \quad 2^{s-1} \equiv 2 \pmod{3}.$$

Thus,

$$\begin{aligned} \left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor + \left\lfloor \frac{1}{3} \cdot 2^{s-1} \right\rfloor &= \left\lfloor \frac{2^s}{3} \right\rfloor + \left\lfloor \frac{2^{s-1}}{3} \right\rfloor \\ &= \frac{2^s - 1}{3} + \frac{2^{s-1} - 2}{3} = \frac{3 \cdot 2^{s-1} - 3}{3} = 2^{s-1} - 1. \end{aligned}$$

**Case 2:** If  $s$  is odd, then

$$2^s \equiv 2 \pmod{3}, \quad 2^{s-1} \equiv 1 \pmod{3}.$$

Thus,

$$\begin{aligned} \left\lfloor \frac{2}{3} \cdot 2^{s-1} \right\rfloor + \left\lfloor \frac{1}{3} \cdot 2^{s-1} \right\rfloor &= \left\lfloor \frac{2^s}{3} \right\rfloor + \left\lfloor \frac{2^{s-1}}{3} \right\rfloor \\ &= \frac{2^s - 2}{3} + \frac{2^{s-1} - 1}{3} = \frac{3 \cdot 2^{s-1} - 3}{3} = 2^{s-1} - 1. \end{aligned}$$

Therefore, it is deduced that in both cases if  $|\mathcal{B}| \leq \lfloor \frac{1}{3} \cdot 2^{s-1} \rfloor$  then  $|\mathcal{B}| \leq |\mathcal{R}|$ . ■

## ACKNOWLEDGMENT

The authors would like to thank Prof. Tuvi Etzion and Prof. Moshe Schwartz for their help in choosing a suitable name for the paper. They also thank the Associate Editor Dr. Gretchen Matthews and two anonymous reviewers whose comments improved the paper and its presentation.

## REFERENCES

- [1] H. Asi and E. Yaakobi, "Nearly optimal constructions of PIR and batch codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 151–155.
- [2] S. Buzaglo, Y. Cassuto, P. H. Siegel, and E. Yaakobi, "Consecutive switch codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2485–2498, Apr. 2018.
- [3] S. Arora and B. Barak, *Computational Complexity—A Modern Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [4] G. Cohen, P. Godlewski, and F. Merx, "Linear binary code for write-once memories," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 5, pp. 697–700, Sep. 1986.
- [5] A. Fazeli, A. Vardy, and E. Yaakobi, "PIR with low storage overhead: Coding instead of replication," 2015, *arXiv:1505.06241*.
- [6] P. Godlewski, "WOM-codes construits à partir des codes de Hamming," *Discrete Math.*, vol. 65, no. 3, pp. 237–243, Jul. 1987.
- [7] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Batch codes and their applications," in *Proc. 23rd Annu. ACM Symp. Theory Comput.*, 2004, pp. 262–271.
- [8] A. S. Rawat, Z. Song, A. G. Dimakis, and A. Gál, "Batch codes through dense graphs without short cycles," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1592–1604, Apr. 2016.
- [9] R. L. Rivest and A. Shamir, "How to reuse a write-once memory," *Inf. Control*, vol. 55, nos. 1–3, pp. 1–19, Dec. 1982.
- [10] E. Sharon and I. Alrod, "Coding scheme for optimizing random I/O performance," in *Proc. Non-Volatile Memories Workshop*, San Diego, CA, USA, Mar. 2013, pp. 1–5.
- [11] H. Sun and S. A. Jafar, "The capacity of private computation," 2017, *arXiv:1710.11098*.
- [12] M. Vajha, V. Ramkumar, and P. V. Kumar, "Binary, shortened projective Reed Müller codes for coded private information retrieval," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2648–2652.
- [13] A. Vardy and E. Yaakobi, "Constructions of batch codes with near-optimal redundancy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1197–1201.
- [14] Z. Wang, H. M. Kiah, Y. Cassuto, and J. Bruck, "Switch codes: Codes for fully parallel reconstruction," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2061–2075, Apr. 2017.
- [15] E. Yaakobi, S. Kayser, P. H. Siegel, A. Vardy, and J. K. Wolf, "Codes for write-once memories," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5985–5999, Sep. 2012.
- [16] A. Yamawaki, H. Kamabe, and S. Lu, "Construction of parallel RIO codes using coset coding with Hamming codes," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Kaohsiung, Taiwan, Nov. 2017, pp. 239–243.
- [17] L. Yohananov and E. Yaakobi, "Almost optimal construction of functional batch codes using Hadamard codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Melbourne, VIC, Australia, Jul. 2021, pp. 3139–3144.
- [18] Y. Zhang, T. Etzion, and E. Yaakobi, "Bounds on the length of functional PIR and batch codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4917–4934, Aug. 2020.

**Lev Yohananov** (Student Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in computer science from the Technion—Israel Institute of Technology, Haifa, Israel, in 2016, 2018, and 2022, respectively. He is currently a Post-Doctoral Student at the Computer Engineering Department, Ben Gurion University of the Negev. His research interests include algebraic error-correction codes, codes over graphs and trees, and combinatorics.

**Eitan Yaakobi** (Senior Member, IEEE) received the B.A. degree in computer science and mathematics and the M.Sc. degree in computer science from the Technion—Israel Institute of Technology, Haifa, Israel, in 2005 and 2007, respectively, and the Ph.D. degree in electrical engineering from the University of California, San Diego, in 2011. From 2011 to 2013, he was a Post-Doctoral Researcher with the Department of Electrical Engineering, California Institute of Technology, and the Center for Memory and Recording Research, University of California, San Diego. He is currently an Associate Professor at the Computer Science Department, Technion—Israel Institute of Technology. His research interests include information and coding theory with applications to non-volatile memories, associative memories, DNA storage, data storage and retrieval, and private information retrieval. He received the Marconi Society Young Scholar (2009) and the Intel Ph.D. Fellowship (2010–2011).