Consecutive Positive Detectable Matrices and Group Testing for Consecutive Positives

Meinard Müller and Masakazu Jimbo

Keio University, Department of Mathematics, 3-14-1 Hiyoshi
Kohoku-ku, Yokohama 223-8522, Japan

Abstract

Colbourn (1999) developed some strategy for nonadaptive group testing when the items are linearly ordered and the positives items form a consecutive subset of all items. We improve his strategy by introducing the concept of 2-consecutive positive detectable matrices (2CPD-matrix) requiring that all columns and bitwise OR-sum of each two consecutive columns are pairwise distinct. Such a matrix is called maximal if it has a maximal possible number of columns with respect to some obvious constraints. Using a recursive construction we prove the existence of maximal 2CPD-matrices for any column size $m \in \mathbb{N}$ except for the case $m = 3$. Furthermore, we construct maximal 2CPD-matrices where each column is of some fixed constant weight. This leads to pooling designs, where each item appears in the same number of pools and all pools are of the same size.

1 Introduction

Let $C = \{c_1, \ldots, c_n\}$ a set of *items* and $\sigma : C \to \{0, 1\}$ a map indicating the *state* of each item. An item c_i is said to be *positive* if $\sigma(c_i) = 1$, otherwise *negative*. In applications such as DNA library screening (in this case, the items are *clones*) one has the goal to determine the set of all positive items in C, where a method is given to *test* the state of each item (e.g., by some chemical analysis). To reduce the number of tests, one chooses a subset $P \subset C$, also denoted as *group* or *pool*, and tests all items of P in one stroke. The state of a pool is *positive* if it contains at least one positive item, otherwise *negative*. This strategy is known as *group testing* which can be defined as the process of selecting pools and testing them to determine exactly which items are positive [1]. A group testing procedure is called *nonadaptive* if all pools are specified a priori without knowing the state of other pools. In this case, the *complexity* of the group testing algorithm is given by the number of its pools. Note that it must be ensured by the group testing procedure that every possible set of positive items is distinguished. Each nonadaptive group test with n items and m pools can be represented by some $m \times n$-matrix $H = (h_{ji})$ over GF(2), which we will refer to as *incidence matrix* of the group test. Here, the columns of H correspond to the items, the rows of H correspond to the pools, and $h_{ji} = 1$ means that the jth pool contains the ith item c_i, $1 \leq j \leq m$, $1 \leq i \leq n$.

For an overview of different group testing methods and some of their applications we refer to [2]. Colbourn [1] considers the setting were the set C is equipped with a linear order $c_i \prec c_{i+1}$, $1 \leq i < n$, and has the *d-consecutive positive property*, i.e., the set of positive items is a consecutive set with respect to the ordering $\prec$ and contains at most d items. His main result can be summarized as follows.

Theorem 1.1 *The complexity of nonadaptive group testing for a set C of n items having the d-consecutive positive property is $\Theta(d + \log_2 n)$.*

To prove the upper bound Colbourn designs a group testing algorithm which proceeds in two steps. In the first step, he considers the general case $d \geq 2$. The n items of C are partitioned into $\lceil n/(d-1) \rceil$ linearly ordered subpools of $(d-1)$ consecutive items respectively (except of the last subpool having possibly a smaller size). By assumption, at most two of these pools, which are then consecutive, are positive. The items of these positive pools can be tested individually in $O(d)$. Treating these subpools as items the general case can thus be reduced to the case $d = 2$ which is dealt with in the second step. To this means, Colbourn constructs an $m \times n$-matrix $H = (h_{ji})$ over GF(2) by adding three suitable rows to an incidence matrix of some Gray code and possibly deleting some columns. From this matrix H he gets a group test with $m = \lceil \log_2 n \rceil + 3$ pools which accomplishes the task for the case $d = 2$.

In this paper we improve the group testing method of Colbourn [1] described above. The main idea of our construction is that in the case $d = 2$ one can distinguish up to any two consecutive positive items if all columns of H as well as all vectors arising as bitwise OR-sum of two consecutive columns of H are pairwise distinct. We will denote such matrices as *2-consecutive positive detectable matrices* or, for short, as 2CPD-matrices. In Section 2 we prove by some recursive construction the existence of such matrices having a maximal number of columns for any column size $m \in \mathbb{N}$ except for the case $m = 3$ (Theorem 2.4). Based on these maximal 2CPD-matrices one gets a group testing procedure for the case $d = 2$ which needs $m = \lceil \log_2 n \rceil + 1$ pools to test n items. If the number m of pools is fixed, this allows a group test of up to $n = 2^{m-1}$ items. This improves

Colbourn's construction by a factor of four with respect to the number of items and is optimal under all possible group testing algorithms for a set C having the 2-consecutive positive property.

In view of the application it is desirable that each item has the same *replication number*, i.e., it appears the same number of times in the pools. In other words, all columns of the incidence matrix H should have some fixed constant weight. In Section 3, we investigate 2CPD-matrices of some constant column weight $r \in \mathbb{N}$. We give some recursive construction of such matrices having the maximal possible number of columns for any given column size $m \in \mathbb{N}$ and weight r with $1 \leq r \leq \lfloor \frac{m}{2} \rfloor$ (Theorem 3.6). We conclude with some final remarks and further references of related problems in Section 4.

2 Construction of Maximal 2CPD-Matrices

We start with a formal definition of 2-consecutive positive detectable matrices mentioned in the introduction. In the following, $\vee$ will denote the OR operation of two bits in GF(2), i.e., $0 \vee 0 = 0$ and $0 \vee 1 = 1 \vee 0 = 1 \vee 1 = 1$. For vectors over GF(2) this operation is understood componentwise.

Definition 2.1 *Let $H = [x_1, x_2, \ldots, x_n]$ be an $m \times n$-matrix over* GF(2) *with column vectors x_i, $1 \leq i \leq n$. Define $y_i := x_i \vee x_{i+1}$, $1 \leq i \leq n-1$. Then H is called a* 2-consecutive positive detectable matrix *or, for short, a* 2CPD-matrix *iff the list*

$$x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_{n-1}$$

consists of pairwise distinct vectors. Define $y_n := x_n \vee x_1$. Then we say a 2CPD-*matrix H is* cyclic *iff*

$$x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$$

consists of pairwise distinct vectors.

Let H be a 2CPD-matrix as in Defintion 2.1. Then, we denote by $H^{\vee}$ the $m \times (2n-1)$-matrix $H^{\vee} := [x_1, y_1, x_2, y_2, x_3, \ldots, x_{n-1}, y_{n-1}, x_n]$. In the cyclic case we similarly define $H^{\circledvee} := [x_1, y_1, x_2, y_2, x_3, \ldots, x_{n-1}, y_{n-1}, x_n, y_n]$. Obviously, from the definition follows that all vectors x_i and y_i are nonzero. Furthermore, since there are 2^m vectors in $\mathrm{GF}(2)^m$ one gets $2n-1 \leq 2^m - 1$, i.e., $n \leq 2^{m-1}$. A 2CPD-matrix H is called *maximal*, or simply an M2CPD-matrix, iff $n = 2^{m-1}$. In this case any nonzero vector of $\mathrm{GF}(2)^m$ appears exactly once as a column of $H^{\vee}$. Therefore, any M2CPD-matrix cannot by cyclic at the same time. However, cyclic 2CPD-matrices will play a crucial role in Section 3. In the following, let M2CPDM(m) denote the class of M2CPD-matrices of column size m. We will give some examples in the next lemma.

Lemma 2.2 *For convenience, we write the OR-sums in $H^\vee$ in italics.*

(i) The following matrix is an M2CPD*-matrix of column size $m = 2$:*

$$H = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad H^\vee = \begin{bmatrix} 0 & \mathit{1} & 1 \\ 1 & \mathit{1} & 0 \end{bmatrix}.$$

(ii) There is no M2CPD*-matrix of column size $m = 3$.*

(iii) The following matrix is an M2CPD*-matrix of column size $m = 4$:*

$$\begin{aligned} H &= \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \\ H^\vee &= \begin{bmatrix} 1 & \mathit{1} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 0 \\ 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 \\ 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{1} & 1 \\ 0 & \mathit{1} & 1 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{0} & 0 \end{bmatrix} \end{aligned}$$

(iv) The following matrix is an M2CPD*-matrix of column size $m = 5$:*

$$H = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Proof: One directly checks that the matrices in (i), (iii) and (iv) meet the conditions of the definition. To show (ii), note that if there were an M2CPD-matrix of column size $m = 3$ the columns $x_1, \ldots, x_4$ must contain the three vectors of weight 1 and one vector of weight 2. Then, one easily checks that any such combination does not lead to an M2CPD-matrix. □

To find the M2CPD-matrices in the cases $m = 4$ and $m = 5$ we first reduced the number of possible candidates by utilizing necessary conditions on the weight distribution of the column vectors x_i, $1 \leq i \leq n$. (For example, all vectors of weight 1 must obviously be among the x_i's.) Then, the M2CPD-matrices were constructed by assembling "locally defined building blocks". For higher dimensions, the following theorem gives some recursive construction for M2CPD-matrices.

Proposition 2.3 *The existence of some $H \in$* M2CPDM*(m), $m > 2$, implies the existence of some $G \in$* M2CPDM*$(m+2)$.*

Proof: Let $H = [x_1, x_2, \ldots, x_n]$ be an M2CPD-matrix, $m > 2, n = 2^{m-1}$, and let $y_i = x_i \vee x_{i+1}$, $1 \le i \le n-1$. Then, we define the $(m+2) \times 4n$-matrix G as follows:

$$G = \left[\begin{array}{cccccccccc} \mathbf{0} & \mathbf{0} & x_n & x_{n-1} & \ldots & x_2 & x_1 & x_1 & x_2 & \ldots \\ 0 & 1 & 0 & 0 & \ldots & 0 & 0 & 1 & 1 & \ldots \\ 1 & 0 & 1 & 1 & \ldots & 1 & 1 & 0 & 0 & \ldots \end{array}\right.$$

$$\left.\begin{array}{cccccccccccc} \ldots & x_{n-1} & x_n & x_{n-1} & \ldots & x_2 & x_1 & x_2 & \ldots & x_{n-1} & x_n \\ \ldots & 1 & 1 & 1 & \ldots & 1 & 0 & 0 & \ldots & 0 & 0 \\ \ldots & 0 & 0 & 1 & \ldots & 1 & 0 & 0 & \ldots & 0 & 0 \end{array}\right]$$

Since $H \in \mathrm{M2CPDM}(m)$, the $8n-1$ nonzero vectors of $\mathrm{GF}(2)^{m+2}$ can be represented by

$$\begin{bmatrix} \mathbf{0} \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} x_i \\ a \\ b \end{bmatrix}, \begin{bmatrix} y_j \\ a \\ b \end{bmatrix}, 1 \le i \le n, 1 \le j \le n-1, a, b \in \{0,1\}.$$

Since G has $4n$ columns, it suffices to show that all of these $8n-1$ vectors appear either as a column of G or as the OR-sum of two consecutive columns of G. This follows immediately for the vectors $[x_i, a, b]^T$, $2 \le i \le n-1, a, b \in \{0,1\}$, and $[y_j, a, b]^T$, $2 \le j \le n-2, a, b \in \{0,1\}$. For the remaining vectors this follows from the following representation of the matrix $G^\vee$:

$$G^\vee = \left[\begin{array}{cccccccccccccccc} \mathbf{0} & \mathbf{0} & \mathbf{0} & x_n & x_n & y_{n-1} & x_{n-1} & \ldots & x_2 & y_1 & x_1 & x_1 & x_1 & y_1 & x_2 & \ldots \\ 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \ldots & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 1 & \ldots \\ 1 & \mathit{1} & 0 & \mathit{1} & 1 & \mathit{1} & 1 & \ldots & 1 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \ldots \end{array}\right.$$

$$\left.\begin{array}{cccccccccccccccc} \ldots & x_{n-1} & y_{n-1} & x_n & y_{n-1} & x_{n-1} & \ldots & x_2 & y_1 & x_1 & y_1 & x_2 & \ldots & x_{n-1} & y_{n-1} & x_n \\ \ldots & 1 & \mathit{1} & 1 & \mathit{1} & 1 & \ldots & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \ldots & 0 & \mathit{0} & 0 \\ \ldots & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \ldots & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \ldots & 0 & \mathit{0} & 0 \end{array}\right]$$

□

Note that in the case $m = 2$, i.e., $n = 2$, the columns of the matrix $G^\vee$ are not any longer pairwise distinct. For example, the vector $[y_{n-1}, 1, 0]^T$ appears in this case more than once as OR-sum. Therefore, the condition $m > 2$ is needed in the construction of Proposition 2.3. From Lemma 2.2 and Proposition 2.3 we get the following result.

Theorem 2.4 *There exists a maximal 2-consecutive positive detectable matrix of any column size $m \in \mathbb{N}$ except for $m = 3$.*

3 2CPD-Matrices of Constant Column Weight

As mentioned in the introduction any M2CPD-matrix H of column size m defines an optimal nonadaptive group testing procedure with m pools and $n = 2^{m-1}$ items having the 2-consecutive positive property. In view of applications, however, M2CPD-matrices have the following two drawbacks. Firstly, the pool sizes (weight of the rows of H) are roughly between $\frac{n}{3}$ and $\frac{n}{2}$ which is too big for most applications. Secondly, the replication numbers of the items (weight of the corresponding columns of H) differ considerable among each other. For example, in the matrix H of Lemma 2.2, (iv), the first item appears in two pools, the second one in one pool, and the third one in three pools. This is not acceptable for many applications where one demands some constant replication number independent of the respective item. To this means, we investigate in this section 2CPD-matrices with some constant column weight.

Let $H = [x_1, x_2, \ldots, x_n]$ be a 2CPD-matrix of column size m where each column x_i is of weight r for some fixed $1 \leq r \leq m$, $1 \leq i \leq n$. In this case, the number n of columns is obviously bounded by $\binom{m}{r}$. Let $\mathrm{M2CPDM}(m, r)$ denote the class of 2-consecutive positive detectable matrices of column size m and of constant column weight r having the *maximal* possible number of columns $n = \binom{m}{r}$. The subclass of *cyclic* (see Definition 2.1) matrices in $\mathrm{M2CPDM}(m, r)$ will be denoted by $\mathrm{CM2CPDM}(m, r)$. Let $y_i := x_i \vee x_{i+1}$, $1 \leq i \leq n-1$, be defined as in Definition 2.1 and, in the cyclic case, $y_n := x_n \vee x_1$. Since the weight of each y_i is at least $r+1$, one gets the following necessary condition.

Lemma 3.1 *Let $H \in \mathrm{M2CPDM}(m, r)$ and $n := \binom{m}{r}$ be the number of columns of H. Then $n \leq \sum_{\ell=r+1}^{m} \binom{m}{\ell} + 1$. If, in addition, H is cyclic then $n \leq \sum_{\ell=r+1}^{m} \binom{m}{\ell}$.*

The following examples will illustrate the definitions and also constitute the starting matrices for the recursive constructions described below.

Example 3.2 We use the notation H, $H^{\vee}$ and $H^{\circledvee}$ introduced in Section 2 and write, for convenience, the OR-sums in italics.

(i) The identity matrix Id_m of dimension $m \in \mathbb{N}$ is in $\mathrm{M2CPDM}(m, 1)$. Furthermore, one has $\mathrm{Id}_m \in \mathrm{CM2CPDM}(m, 1)$ for $m > 2$.

(ii) The following matrix H, given in the form $H^{\vee}$, is in $\mathrm{M2CPDM}(4, 2)$:

$$H^{\vee} = \begin{bmatrix} 1 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 \\ 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 1 \\ 1 & \mathit{1} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{1} & 1 \\ 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 1 & \mathit{1} & 1 & \mathit{1} & 0 \end{bmatrix}$$

Note that there is no cyclic matrix in M2CPDM$(4, 2)$. This follows directly from Lemma 3.1, since $n = \binom{4}{2} = 6 > 5 = \binom{4}{3} + \binom{4}{4}$.

(iii) The following matrix H, given in the form $H^{\circledvee}$, is in CM2CPDM$(5, 2)$:

$$H^{\circledvee} = \begin{bmatrix} 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{1} \\ 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 1 & \mathit{1} \\ 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} \\ 0 & \mathit{0} & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 1 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} \\ 1 & \mathit{1} & 1 & \mathit{1} & 1 & \mathit{1} & 1 & \mathit{1} & 0 & \mathit{0} & 0 & \mathit{0} & 0 & \mathit{0} & 0 & \mathit{0} & 0 & \mathit{0} & 0 & \mathit{1} \end{bmatrix}$$

(iv) The following matrix H is in CM2CPDM$(6, 3)$:

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Obviously, any permutation of the row vectors of a PD-matrix leads again to a PD-matrix. Furthermore, any cyclic shift of the column vectors of a cyclic PD-matrix will again define a cyclic PD-matrix. Since these observations will be useful in the later constructions, we note them down in the next lemma.

Lemma 3.3 *The class* CM2CPDM(m, r) *is invariant under row permutations and cyclic shifts of the column vectors. In other words, if $H = [x_1, x_2, \ldots, x_n]$ is in* CM2CPDM(m, r)*, then*

$$P \cdot [x_i, x_{i+1}, \ldots, x_n, x_1, \ldots, x_{i-1}]$$

is also in CM2CPDM(m, r) *for any $m \times m$-permutation matrix P and any $1 \leq i \leq n$.*

It is easy to check that the necessary condition of Lemma 3.1 is fulfilled for any r satisfying $1 \leq r \leq \lfloor \frac{m}{2} \rfloor$, $m \in \mathbb{N}$. In the following, our goal is to give some systematic construction of matrices in M2CPDM(m, r) for all $m \in \mathbb{N}$ and $1 \leq r \leq \lfloor \frac{m}{2} \rfloor$. We start with some simple recursive construction.

Lemma 3.4 *Let $A = [a_1, a_2, \ldots, a_k] \in$* M2CPDM$(m, r-1)$*, $k = \binom{m}{r-1}$, and $B = [b_1, b_2, \ldots, b_\ell] \in$* M2CPDM$(m, r)$*, $\ell = \binom{m}{r}$. If $a_k \vee b_1 \neq a_i \vee a_{i+1}$ for all $1 \leq i < k$ then*

$$C := \begin{bmatrix} a_1 & a_2 & \ldots & a_{k-1} & a_k & b_1 & b_2 & \ldots & b_{\ell-1} & b_\ell \\ 1 & 1 & \ldots & 1 & 1 & 0 & 0 & \ldots & 0 & 0 \end{bmatrix}$$

defines a matrix in M2CPDM$(m+1,r)$. *If, in addition,* $b_\ell \vee a_1 \neq a_i \vee a_{i+1}$ *for all* $1 \leq i < k$ *and* $b_\ell \vee a_1 \neq a_k \vee b_1$*, then* C *is cyclic, i.e.,* $C \in$ CM2CPDM$(m+1,r)$.

Proof: By the assumptions on A and B the columns of C are precisely the vectors of weight r of $GF(2)^{m+1}$. Observing the last component it follows immediately that all columns of C and all OR-sums of any two consecutive columns of C are pairwise different. In particular, for $[a_k \vee b_1, 1]^T$ this follows from the condition posed on $a_k \vee b_1$. The additional assumption on $b_\ell \vee a_1$ assures that $[b_\ell \vee a_1, 1]^T$ also differs from all other vectors in question. $\square$

For example, the matrix (iii) of Example 3.2 has been obtained by this construction using Id_4 as matrix A and the matrix (ii) of Example 3.2 as matrix B. Lemma 3.4 gives some recursive construction where the column size m increases. However, the weight r of the columns is kept fixed. The next proposition gives some recursive construction where r increases as well.

Proposition 3.5 *Let* $m \in \mathbb{N}$ *be even. If there is an* $H \in$ M2CPDM$(m, \frac{m}{2})$ *then there is also some* $G \in$ M2CPDM$(m+2, \frac{m}{2}+1)$. *Furthermore, if there is an* $H \in$ CM2CPDM$(m, \frac{m}{2})$ *then there is also some* $G \in$ CM2CPDM$(m+2, \frac{m}{2}+1)$.

Proof: We just consider the construction for the cyclic case which handles the noncyclic case as well. Let $H = [x_1, x_2, \ldots, x_n] \in$ CM2CPDM$(m, \frac{m}{2})$, $y_i := x_i \vee x_{i+1}$, $1 \leq i \leq n-1$, and $y_n := x_n \vee x_1$. Furthermore, let $a_1, \ldots, a_\ell$ and $b_1, \ldots, b_\ell$ the vectors in GF$(2)^m$ of weight $\frac{m}{2}-1$ and $\frac{m}{2}+1$ respectively, where $\ell := \binom{m}{m/2-1} = \binom{m}{m/2+1}$. By Corollary A.2 of the appendix and by relabelling the vectors b_j suitably, we may assume that there is some permutation $\pi \in \mathrm{Sym}(\{1,\ldots,n\})$ such that $a_j \subset x_{\pi(j)} \subset b_j$ for $1 \leq j \leq \ell$. Here, by $u \subset v$ for two vectors $u, v \in$ GF$(2)^m$ we mean that the support of u is contained in the one of v, i.e., $u \vee v = v$. Note that all vectors of GF$(2)^{m+2}$ of weight $\frac{m}{2}+1$ appear in the following list exactly once:

$$\begin{bmatrix} x_i \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} x_i \\ 0 \\ 1 \end{bmatrix}, 1 \leq i \leq n, \quad \begin{bmatrix} a_j \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} b_j \\ 0 \\ 0 \end{bmatrix}, 1 \leq j \leq \ell. \tag{1}$$

First, we define an $(m+2) \times 2n$-matrix F consisting of the first $2n$ vectors given in the list (1). Here, we use the fact that $n = \binom{m}{m/2}$ is even for even m. Let

$$F := \begin{bmatrix} x_1 & x_1 & x_2 & x_2 & x_3 & x_3 & x_4 & \ldots & x_{n-2} & x_{n-1} & x_{n-1} & x_n & x_n \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & \ldots & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & \ldots & 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Then F is obviously a cyclic 2CPD-matrix. Note that for even i the vectors $[y_i, 1, 0]^T$ and for odd i the vectors $[y_i, 0, 1]^T$ do not appear as OR-sum of two consecutive columns of F or the last and first column of F, $1 \leq i \leq n$. We now extend the matrix F by inserting the remaining column vectors of the list (1). For any j, $1 \leq j \leq \ell$, we insert the vectors $[a_j, 1, 1]^T$ and $[b_j, 0, 0]^T$ between $[x_{\pi(j)}, 0, 1]^T$ and $[x_{\pi(j)}, 1, 0]^T$. The vector b_j might appear among the vectors y_i, $1 \leq i \leq n$. However, from the note above it follows that either $[b_j, 1, 0]^T$ (Case 1) or $[b_j, 0, 1]^T$ (Case 2) does not appear as OR-sum of any two consecutive columns of F or the last and first column of F. Suppose $\pi(j)$ is even. Then we insert the columns as follows:

$$\text{Case 1:}\quad \begin{array}{cccc} x_{\pi(j)} & b_j & a_j & x_{\pi(j)} \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{array} \qquad \text{Case 2:}\quad \begin{array}{cccc} x_{\pi(j)} & a_j & b_j & x_{\pi(j)} \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{array}$$

If $\pi(j)$ is odd then we proceed in a similar way. (Just reverse the order of the columns in the above submatrices.) This defines an $(m+2) \times (2n+2\ell)$-matrix G consisting precisely of the vectors given in the list (1). Then the submatrix for Case 1 together with the OR-sums of two consecutive columns are given by

$$\text{Case 1:}\quad \begin{array}{ccccccc} x_{\pi(j)} & b_j & b_j & b_j & a_j & x_{\pi(j)} & x_{\pi(j)} \\ 1 & \mathit{1} & 0 & \mathit{1} & 1 & \mathit{1} & 0 \\ 0 & \mathit{0} & 0 & \mathit{1} & 1 & \mathit{1} & 1 \end{array}$$

and similarly for Case 2. From this it is not hard to show that G indeed defines a matrix in CM2CPDM$(m+2, \frac{m}{2}+1)$. □

From the last two recursive constructions we get the following main result of this section.

Theorem 3.6 *For any $m \in \mathbb{N}$ and any r, $1 \leq r \leq \lfloor \frac{m}{2} \rfloor$, there exists a matrix in* M2CPDM(m, r). *There is also a matrix in* CM2CPDM(m, r) *except for the parameters $m = 2, r = 1$ and $m = 4, r = 2$.*

Proof: In the case $r = 1$ one has $\mathrm{Id}_m \in$ M2CPDM$(m, 1)$ for $m \in \mathbb{N}$ and $\mathrm{Id}_m \in$ CM2CPDM$(m, 1)$ for $m > 2$. The matrices (ii) and (iii) of Example 3.2 settle the cases M2CPDM$(4, 2)$ and CM2CPDM$(5, 2)$ respectively. Thus the claim of Theorem 3.6 is shown for $m \leq 5$. The matrix (iv) of Example 3.2 settles the case CM2CPDM$(6, 3)$. Now, using Proposition 3.5, we get matrices in CM2CPDM$(m, \frac{m}{2})$ for all even $m \in \mathbb{N}$.

It remains to show the claim for all $m > 5$ and $1 < r < \lfloor \frac{m}{2} \rfloor$ which will be done by induction on m using the above cases as starting matrices. Fix such an m and r and assume that the claim has already shown for $m-1$, i.e., we may assume the existence of some matrix $A = [a_1, a_2, \ldots, a_k] \in$ CM2CPDM$(m-1, r-1)$ and $B = [b_1, b_2, \ldots, b_\ell] \in$ CM2CPDM$(m-1, r)$. In case A and B satisfy the conditions

(i) $\{a_k \vee b_1, b_\ell \vee a_1\} \cap \{a_i \vee a_{i+1} \mid 1 \le i < k\} = \emptyset$ and

(ii) $b_\ell \vee a_1 \neq a_k \vee b_1$,

then Lemma 3.4 guarantees the existence of some matrix in $\mathrm{CM2CPDM}(m, r)$ which completes the proof of Theorem 3.6 in this case.

In general, however, conditions (i) and (ii) are not fulfilled. In the following we show that one can modify A and B so that these conditions are satisfied. This then finishes the proof. First, we assume that A contains two consecutive columns a_i and a_{i+1} such that $a_i \vee a_{i+1}$ has weight r, $1 \le i < k$, or $a_k \vee a_1$ has weight r. Similarly, we assume that B contains two consecutive columns b_j and b_{j+1} such that $b_j \vee b_{j+1}$ has weight $r+1$, $1 \le i < \ell$, or $b_\ell \vee b_1$ has weight $r+1$. By Lemma 3.3 we may permute the rows and cyclically shift the columns of A and B to get matrices of the same classes, respectively. Therefore, we may assume that A and B are of the following form where only the first and last columns are indicated:

$$A = \begin{array}{c} \begin{array}{ccc} a_1 & \qquad\qquad & a_k \end{array} \\ \begin{bmatrix} \begin{matrix} \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\}r-1 \\ 0\\ 0\\ 0\\ \vdots \end{matrix} & \cdots & \begin{matrix} 0 \\ r-1\left\{\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right. \\ 0\\ 0\\ \vdots \end{matrix} \end{bmatrix} \end{array} \qquad B = \begin{array}{c} \begin{array}{ccc} b_1 & \qquad\qquad & b_\ell \end{array} \\ \begin{bmatrix} \begin{matrix} \vdots\\ 0\\ 0\\ 0\\ \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\}r \end{matrix} & \cdots & \begin{matrix} \vdots\\ 0\\ 0\\ r\left\{\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right. \\ 0 \end{matrix} \end{bmatrix} \end{array}$$

From $r < \lfloor \frac{m}{2} \rfloor$ we get $2r \le m - 1$. Therefore, the vectors $a_k \vee b_1$ and $b_\ell \vee a_1$ have weight $2r - 1$. Since the weights of the vectors $a_i \vee a_{i+1}$, $1 \le i < k$, are at most $2r - 2$, condition (i) is fulfilled. Furthermore, $a_k \vee b_1 \neq b_\ell \vee a_1$ and condition (ii) holds as well.

Now, to finish the proof it suffices to show that any matrix $H = [x_1, \ldots, x_n] \in \mathrm{CM2CPDM}(m, r)$, $m \ge 5$, $1 \le r \le \lfloor \frac{m}{2} \rfloor$, constructed in the recursive process described above contains at least one pair of two consecutive columns x_i, x_{i+1}, $1 \le i < n$, or the pair x_n, x_1, such that the OR-sum $x_i \vee x_{i+1}$ or $x_n \vee x_1$ has weight $r + 1$. We denote the number of such pairs in H by $\mu(H)$. This number is clearly invariant under row permutations and cyclic shifts of columns of H. For $m = 5$ we gave an explicit matrix in Example 3.2 (iii), for which obviously $\mu(H) \ge 2$ holds. (The lower bound 2 suffices for our proof.) Furthermore, one clearly has $\mu(\mathrm{Id}_m) \ge 2$ for all $m > 2$ and also $\mu(H) \ge 2$ for all matrices H in $\mathrm{CM2CPDM}(m, \frac{m}{2})$, m even, $m \ge 6$, constructed as described in Proposition 3.5. Finally, it is easy to see that the matrix C constructed from A and B as described in Lemma 3.4 satisfies $\mu(C) \geq \mu(A) + \mu(B) - 2$. From this follows $\mu(H) \geq 2$ for all H constructed in the recursive process. $\square$

Finally, note that since any matrix $H \in \mathrm{M2CPDM}(m, r)$ contains each vector of weight r of $\mathrm{GF}(2)^m$ exactly once, it follows that each row of H has weight $\frac{r}{m} \cdot \binom{m}{r}$. In other words, the pool sizes of the corresponding group test all coincide. From Theorem 3.6 we get the following corollary.

Corollary 3.7 *For any $m \in \mathbb{N}$ and any r, $1 \leq r \leq \lfloor \frac{m}{2} \rfloor$, there is an optimal group testing procedure for items having the 2-consecutive positive property with m pools of size $\frac{r}{m} \cdot \binom{m}{r}$ and $n = \binom{m}{r}$ items, where each item appears in exactly r pools.*

4 Final Remarks

Nonadaptive group testing has motivated many problems in combinatorial design theory. In this paper we have introduced and constructed certain classes of 2CPD-matrices which can be used in group testing procedures for items having the d-consecutive positive property (which can be reduced, as mentioned in the introduction, to the case $d = 2$). We want to emphasize that the problem, where one does not require the positives to be consecutive, is essentially different to the one discussed in this paper. The case, where one just assumes that the positive items are bounded by some number d, requires that the OR-sums of *any* d (not necessarily distinct) columns of the group testing incidence matrix are pairwise distinct. This problem has lead to the concept of *d-disjunctive* matrices. For an overview and further references concerning these matrices we refer the reader to [2, 4].

As is also pointed out in [1] or [4], error correction capability of group testing procedures is essential in view of applications such as DNA library screening. Therefore, extending the concept of 2CPD-matrices to error correcting codes is an ongoing research project of the authors. In general, it seems to be difficult to find maximal 2CPD-matrices, where the columns x_i and the OR-sums y_i cover all vectors of some error correcting code. For example, if one considers the code consisting of all even weighted vectors (which is a one-error decting code) non-existence of maximal 2CPD-matrices can be shown for all columns sizes $m \leq 8$. We note that any maximal 2CPD-matrix over such a code would also give a solution to the *dominance code problem* (i.e., ordering codewords so that every two consecutive codewords have one dominating the other) which was solved by Sagols et. al. in [3] for $m \geq 10$. It would be interesting to know whether in this case there even exists a maximal 2CPD-matrix or not.

Appendix

In this appendix we give some recursive construction to partition the vectors of $\mathrm{GF}(2)^m$ into dominating sequences of increasing weight for any $m \in \mathbb{N}$. This result is needed in form of Corollary A.2 in the proof of Proposition 3.5.

Let $u, v \in \mathrm{GF}(2)^m$, then we say that the vector v *dominates* the vector u if the support of u is contained in the one of v, i.e., $u \vee v = v$. In this case we also write $u \subset v$. A sequence of pairwise different vectors $S = (s_1, s_2, \ldots, s_k)$, $s_i \in \mathrm{GF}(2)^m$, $1 \leq i \leq k$, is called an *increasing dominating sequences* if $s_1 \subset s_2 \subset \ldots \subset s_k$. Obviously, the maximal length of such a sequence is $k = m + 1$, which contains the vector of weight 0 as first and the vector of weight m as last element. In the following, we identify a sequence with its underlying set of vectors. If we remove a maximal increasing dominating sequence S from $\mathrm{GF}(2)^m$, any increasing dominating sequences consisting of vectors in $\mathrm{GF}(2)^m \setminus S$ has length at most $m-1$ and there are at most $\binom{m}{1} - \binom{m}{0} = m - 1$ such pairwise disjoint sequences. Continuing the same way, one arrives at the problem if there is a partition of $\mathrm{GF}(2)^m$ into increasing dominating sequences such that among these sequences there are

- one sequence of length $m + 1$,
- $\binom{m}{1} - \binom{m}{0}$ sequences of length $m - 1$,
- $\binom{m}{2} - \binom{m}{1}$ sequences of length $m - 3$, and, in general,
- $\binom{m}{i} - \binom{m}{i-1}$ sequences of length $m + 1 - 2i$ for $0 \leq i \leq \lfloor \frac{m}{2} \rfloor$.

In the following we show that there is such a partition of $\mathrm{GF}(2)^m$ for any $m \in \mathbb{N}$, which will be refered to as *optimal partition* and denoted by $\mathcal{S}^m$. Furthermore, the subset of sequences of $\mathcal{S}^m$ having length $m + 1 - 2i$ will be denoted by $\mathcal{S}_i^m$, $0 \leq i \leq \lfloor \frac{m}{2} \rfloor$. For convenience, we also set $\mathcal{S}_i^m := \emptyset$ for all $i < 0$ or $i > \lfloor \frac{m}{2} \rfloor$, $i \in \mathbb{Z}$. Let $S = (s_1, s_2, \ldots, s_k)$ be an increasing dominating sequences in $\mathrm{GF}(2)^m$. For notational reasons we identifiy S with an $m \times k$-matrix having s_i as ith column, $1 \leq i \leq k$. Then, we define the sequences

$$S^- := \begin{bmatrix} s_1 & s_2 & \ldots & s_{k-1} \\ 1 & 1 & \ldots & 1 \end{bmatrix} \quad \text{and} \quad S^+ := \begin{bmatrix} s_1 & s_2 & \ldots & s_{k-1} & s_k & s_k \\ 0 & 0 & \ldots & 0 & 0 & 1 \end{bmatrix}.$$

Clearly, S^- and S^+ define increasing dominating sequences in $\mathrm{GF}(2)^{m+1}$. Now, we construct an optimal partition $\mathcal{S}^m$ by induction on m, $m \in \mathbb{N}$. For $m = 1$ an optimal partition is defined by $\mathcal{S}^1 := \{[01]\}$. Suppose we have constructed an optimal partition $\mathcal{S}^m$, then define

$$\mathcal{S}_i^{m+1} := \{S^- \mid S \in \mathcal{S}_{i-1}^m\} \cup \{S^+ \mid S \in \mathcal{S}_i^m\} \quad \text{for} \quad i \in \mathbb{Z}.$$

Note that one has $\mathcal{S}_i^{m+1} = \emptyset$ for $i < 0$ and $i > \lfloor \frac{m+1}{2} \rfloor$. Now, we set $\mathcal{S}^{m+1} := \bigcup_{i \in \mathbb{Z}} \mathcal{S}_i^{m+1}$. It is straightforward to show that $\mathcal{S}^{m+1}$ indeed defines an optimal partition of $\mathrm{GF}(2)^{m+1}$ into increasing dominating sequences. One just has to observe that in the construction of $\mathcal{S}^{m+1}$ each vector of $\mathcal{S}^m$ is extended exactly two times – one time by 0 and one time by 1. We leave the details to the reader and give some examples instead.

$$m = 1: \quad \mathcal{S}_0^1 = \left\{ \begin{bmatrix} 01 \end{bmatrix} \right\}$$

$$m = 2: \quad \mathcal{S}_0^2 = \left\{ \begin{bmatrix} 011 \\ 001 \end{bmatrix} \right\} \qquad \mathcal{S}_1^2 = \left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

$$m = 3: \quad \mathcal{S}_0^3 = \left\{ \begin{bmatrix} 0111 \\ 0011 \\ 0001 \end{bmatrix} \right\} \quad \mathcal{S}_1^3 = \left\{ \begin{bmatrix} 01 \\ 00 \\ 11 \end{bmatrix}, \begin{bmatrix} 00 \\ 11 \\ 01 \end{bmatrix} \right\}$$

$$m = 4: \quad \mathcal{S}_0^4 = \left\{ \begin{bmatrix} 01111 \\ 00111 \\ 00011 \\ 00001 \end{bmatrix} \right\} \quad \mathcal{S}_1^4 = \left\{ \begin{bmatrix} 011 \\ 001 \\ 000 \\ 111 \end{bmatrix}, \begin{bmatrix} 011 \\ 000 \\ 111 \\ 001 \end{bmatrix}, \begin{bmatrix} 000 \\ 111 \\ 011 \\ 001 \end{bmatrix} \right\} \quad \mathcal{S}_2^4 = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

We summarize the result in the following Theorem.

Theorem A.1 *There is an optimal partition of* $\mathrm{GF}(2)^m$ *into increasing dominating sequences for any* $m \in \mathbb{N}$.

From Theorem A.1 we immediately obtain the following corollary, which is used in the proof of Proposition 3.5.

Corollary A.2 *Let* $m \in \mathbb{N}$ *be an even number. Let* $x_1, \ldots, x_n$ *denote the vectors of weight* $\frac{m}{2}$ *in* $\mathrm{GF}(2)^m$, $n := \binom{m}{m/2}$. *Similarly, let* $a_1, \ldots, a_\ell$ *and* $b_1, \ldots, b_\ell$ *denote the vectors in* $\mathrm{GF}(2)^m$ *having weight* $\frac{m}{2} - 1$ *and* $\frac{m}{2} + 1$ *respectively, where* $\ell := \binom{m}{m/2-1} = \binom{m}{m/2+1}$. *Then,* $\ell < n$ *and there is some permutation* $\pi \in \mathrm{Sym}(\{1, \ldots, n\})$ *and some permutation* $\tau \in \mathrm{Sym}(\{1, \ldots, \ell\})$ *such that* $a_j \subset x_{\pi(j)} \subset b_{\tau(j)}$ *for* $1 \leq j \leq \ell$.

References

[1] Colbourn, C. J.: Group testing for consecutive positives. Annals of Combinatorics **3** (1999), 37–41.

[2] Du, D.-Z., Hwang, F. K.: Combinatorial group testing and its applications. World Scientific, Singapore, 1993.

[3] Sagols, F., Riccio, L. P., Colbourn, C. J.: Dominating error correcting codes with distance two. Journal of Combinatorial Designs **10** (2002), 294–302.

[4] Ngo, H. Q., Du, D.-Z.: A survey on combinatorial group testing algorithms with applications to DNA library screening. DIMACS Series in Discrete Mathematics and Theoretical Computer Science/DU2, Providence, RI, 2000, Amer. Math. Soc.