LE MATEMATICHE
Vol. LIX (2004) – Fasc. I–II, pp. 215–224

COMBINATORIAL PROBLEMS ARISING FROM POOLING DESIGNS FOR DNA LIBRARY SCREENING

MASAKAZU JIMBO - MEINARD MÜLLER

Colbourn (1999) developed some strategy for nonadaptive group testing when the items are linearly ordered and the positive items form a consecutive subset of all items.

Müller and Jimbo (2004) improved 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 they proved the existence of maximal 2CPD-matrices for any column size $m \in \mathbb{N}$ except for the case $m = 3$. Moreover, maximal 2CPD-matrices such that each column is of some fixed constant weight are constructed. This leads to pooling designs, where each item appears in the same number of pools and all pools are of the same size.

Secondly, we investigate 2CPD-matrices of some constant column weight $\tau \in \mathbb{N}$. We give some recursive construction of such matrices having the maximal possible number of columns.

Thirdly, error correction capability of group testing procedures is essential in view of applications such as DNA library screening. We consider a error correcting 2CPD-matrices.

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 $d = 2$.

Mueller and Jimbo [3] improved the group testing method of Colbourn [1] described above. The main idea of their 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. Such matrices are denoted by 2-*consecutive positive detectable matrices* or, for short, as 2CPD-matrices. In Section 2, we describe 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 $\tau \in \mathbb{N}$. 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 τ with $1 \leq r \leq \lfloor \frac{m}{2} \rfloor$ (Theorem 3.6).

As is also pointed out in [1] or [5], error correction capability of group testing procedures is essential in view of applications such as DNA library screening. For a 2CPDM H, if the set of column vectors in $H^\vee$ together with the zero vector is a code with minimum distance d, then H is said to have minimum distance d. It is easy to see that if a 2CPDM H has minimum distance d, then it can correct $e = \lfloor \frac{d-1}{2} \rfloor$ errors of observations of pools. Therefore, extending the concept of 2CPD-matrices to error correcting codes is an interesting problem. A 2CPD-matrix with m pools, weight k and minimum distance d is denoted by 2CPDM(m, k, d).

In Section 4, we consider the existence of a maximal 2CPDM$(m, k, 2)$ in the case when $k = 2$ and 3.

We conclude with some open problems and final remarks in Section 5.

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 $H^\vee := [x_1, y_1, x_2, y_2, x_3, \ldots, x_{n-1}, y_{n-1}, x_n, y_n]$. Obviously, from the definition follows that all vectors x_i and y_i are nonzero. Furthermore, since there are 2^m vectors in 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 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 $m = 2$:

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

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

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

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

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

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

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

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 $H \in \mathrm{M2CPDM}(m)$, $m > 2$, implies the existence of some $G \in \mathrm{M2CPDM}(m + 2)$.*

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 $[y_{n-1}, 1, 0]^{\tau}$ 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 τ for some fixed $1 \leq \tau \leq m$, $1 \leq i \leq n$. In this case, the number n of columns is obviously bounded by $\binom{m}{\tau}$. Let M2CPDM(m, τ) denote the class of 2-consecutive positive detectable matrices of column size m and of constant column weight τ having the *maximal* possible number of columns $n = \binom{m}{\tau}$. The subclass of *cyclic* (see Definition 2.1) matrices in M2CPDM(m, τ) will be denoted by CM2CPDM(m, τ). 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 $H \in$ M2CPDM(m, τ) and $n := \binom{m}{\tau}$ be the number of columns of H. Then $n \leq \sum_{\ell=\tau+1}^{m} \binom{m}{\ell} + 1$. If, in addition, H is cyclic then $n \leq \sum_{\ell=\tau+1}^{m} \binom{m}{\ell}$.*

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 $H^{\vee}$ 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 M2CPDM$(m, 1)$. Furthermore, one has $\mathrm{Id}_m \in$ CM2CPDM$(m, 1)$ for $m > 2$.

(ii) The following matrix H, given in the form $H^{\vee}$, is in 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 $H^{\vee}$, 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, τ) *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, τ)*, then*

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

is also in CM2CPDM(m, τ) *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 τ 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, τ) 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$(m, \tau - 1)$, $k = \binom{m}{\tau - 1}$, *and* $B = [b_1, b_2, \ldots, b_\ell] \in$ M2CPDM(m, τ), $\ell = \binom{m}{\tau}$. *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$(m + 1, \tau)$. *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$(m + 1, \tau)$.

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 τ of the columns is kept fixed. The next proposition gives some recursive construction where τ increases as well.

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

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 $\mathrm{M2CPDM}(m, \tau)$. There is also a matrix in $\mathrm{CM2CPDM}(m, \tau)$ except for the parameters $m = 2, r = 1$ and $m = 4, r = 2$.*

Finally, note that since any matrix $H \in \mathrm{M2CPDM}(m, \tau)$ 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. Error correcting CPDMs.

As is also pointed out in [1] or [5], 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 [4] for $m \geq 10$. It would be interesting to know whether in this case there even exists a maximal 2CPD-matrix or not.

For a 2CPDM H, if the set of column vectors in $H^\vee$ together with the zero vector is a code with minimum distance d, then H is said to have minimum distance d. It is easy to see that if a 2CPDM H has minimum distance d, then it can correct $e = \lfloor \frac{d-1}{2} \rfloor$ errors of observations of pools. Therefore, extending the concept of 2CPD-matrices to error correcting codes is an interesting problem.

A 2CPD-matrix with m pools, weight k and minimum distance d is denoted by 2CPDM(m, k, d).

Thirdly, we consider the existence of a maximal 2CPDM(m, k, 2) in the case when $k = 2$ and 3. Let H be a 2CPDM(m, k, 2) constructed by arranging all vectors of length m and weight k as column vectors of the matrix so that the column vectors in $H^\vee$ are all distinct. In this case H is is obviously maximal. In case of $k = 2, 3$, we have the following theorem.

Theorem 4.1. *A maximal cyclic 2CPDM*(m, 2, 2) *exists for any* $m \geq 6$.

Theorem 4.2. *A maximal cyclic 2CPDM*(m, 3, 2) *exists for any* $m \geq 8$.

5. Open prpblems and final remark.

Finally, some open problems are listed below.

Problem 1. Theorem 2.4 holds for any m and k satisfying $1 \leq k \leq \lfloor \frac{m}{2} \rfloor$. But there are some other parameters satisfying the necessary condition in Lemma 3.1. For example, $m = 9$ and $k = 5$ is the smallest such example. Does there exist a maximal (cyclic) 2CPDM(m, k) for m and k which satisfy the condition of Lemma 3.1 but not satisfying $1 \leq k \leq \lfloor \frac{m}{2} \rfloor$?

Problem 2. Does there exists a (cyclic) 2CPDM(m, k, 2) for $k \geq 4$?

Let X and Y be subsets of F^m. If the columun vectors of a 2CPDM H with minimum distance d consists of all vectors in X and column vectors in $H^\vee$ are in $X \cup Y$, then H is denoted by 2CPDM[X, Y, d]. If there is no restriction to Y, then a 2CPDM[X, Y, d] is simply written by 2CPDM[X, d].

Problem 3. Let X be a constant weight code with length m, weight k, minimum distance $d = 3$ having the maximum number of codewords. Does there exists a (cyclic) 2CPDM[X, 3]? For example, a cyclic 2CPDM[X, 3] does not exist for X being the set of codewords of weight 3 in a Hamming code of length 3, 7, or 15. In general, when X is the set of codewords of weight 3 in a Hamming code of length m, does there exist a cyclic 2CPDM[X, 3]?

Sagols et. al. [4] solved the *dominance code problem*, i.e., ordering codewords so that every two consecutive codewords have one dominating the other, for the code of the even weight vectors when $m \geq 10$. Related to this problem, the following problem may be settled.

Problem 4. In the case when $X \cup Y$ is the set of all even weight non-zero vectors, does there exist a maximal 2CPDM[X, 2] with $n = 2^{m-2}$?

Problem 5. In the case when $X \cup Y$ is the set of non-zero codewords of the Hamming code of length $m = 2^k - 1$, does there exist a maximal 2CPDM$[X, 3]$ with $n = 2^{m-k-1}$?

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], [5].

REFERENCES

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

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

[3] M. Mueller - M. Jimbo, *Consecutive positive detactable matricies and group testing for consecutive positives,* Discrete Mathematics, 279 (2004), pp. 369–381.

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

[5] H.Q. Ngo - D.-Z. Du, *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..

M. Jimbo
Graduate School of Information Science,
Nagoya University (JAPAN)
e-mail: jimbo@is.nagoya-u.ac.jp

M. Müller
Department of Computer Science,
University of Bonn,
Römerstr. 164, 53117 Bonn (GERMANY)
e-mail: meinard@cs.uni-bonn.de