New Constructions of Non-Adaptive and Error-Tolerance Pooling Designs

Hung Q. Ngo* Ding-Zhu Du*

Abstract

We propose two new classes of non-adaptive pooling designs. The first one is guaranteed to be d-error-detecting and thus $\lfloor \frac{d}{2} \rfloor$-error-correcting, where d, a positive integer, is the maximum number of defectives (or positives). Hence, the number of errors which can be detected grows linearly with the number of positives. Also, this construction induces a construction of a binary code with minimum Hamming distance at least $2d + 2$. The second design is the q-analogue of a known construction on d-disjunct matrices.

1 Introduction

The basic problem of group testing is to identify the set of defectives in a large population of items. As it is becoming more standard to use the term *positive* instead of *defective*, we shall use the former throughout the paper. We assume some testing mechanism exists which if applied to an arbitrary subset of the population gives a *negative outcome* if the subset contains no positive and *positive outcome* otherwise. Objectives of group testing vary from minimizing the number of tests, limiting number of pools, limiting pool sizes to tolerating a few errors. It is conceivable that these objectives are often contradicting, thus testing strategies are application dependent.

Group testing algorithms can roughly be divided into two categories : *Combinatorial Group Testing* (CGT) and *Probabilistic Group Testing* (PGT). In CGT, it

*Department of Computer Science and Engineering, University of Minnesota, 200 Union street, EE/CS Building, room 4-192, Minneapolis, MN 55455, USA. e-mail: {hngo, dzd}@cs.umn.edu. Support in part by by the National Science Foundation under grant CCR-9530306.

is often assumed that the number of positives among n items is equal to or at most d for some given positive integer d. In PGT, we fix some probability p of having a positive. Group testing strategies can also be either *adaptive* or *non-adaptive*. A group testing algorithm is non-adaptive if all tests must be specified without knowing the outcomes of other tests. A group testing algorithm is *error tolerant* if it can detect or correct some e errors in test outcomes. Test errors could be either $0 \to 1$, i.e. a negative pool is identified as positive, or $0 \to 1$ in the contrast.

In this paper, we propose two new classes of non-adaptive and error-tolerance CGT algorithms. Non-adaptive algorithms found its applications in a wide range of practical areas such as DNA library screening [2, 5] and multi-access communications [16], etc. For a general reference on CGT, the reader is referred to a monograph by Du and Hwang [6]. Recently, Ngo and Du [14] gave a survey on non-adaptive pooling designs.

The rest of the paper is organized as follows. Section 2 presents basic definitions, notations and related works. Section 3 provides our results and section 4 concludes the paper.

2 Preliminaries

Throughout this paper, for any positive integer v we shall use $[v]$ to denote $\{1, 2, \dots v\}$. Also, given any set X and $k \in \mathbb{N}$, $\binom{X}{k}$ denotes the collection of all k-subsets of X. Naturally, $[0] = \emptyset$ and $\binom{X}{k} = \emptyset$ if $k > |X|$.

2.1 The Matrix Representation

Consider a $v \times n$ 01-matrix M. Let R_i and C_j denote row i and column j respectively. Abusing notation, we also let R_i (resp. C_j) denote the set of column (resp. row) indices corresponding to the 1-entries. The *weight* of a row or a column is the number of 1's it has.

Definition 2.1. M is said to be d-*disjunct* if the union of any d columns does not contain another column.

A d-disjunct $v \times n$ matrix M can be used to design a non-adaptive group testing algorithm on n items by associating the columns with the items and the rows with the pools to be tested. If $M_{ij} = 1$ then item j is contained in pool i (and thus test i). If there are no more than d positives and the test outcomes are error-free, then it is easy to see that the test outcomes uniquely identify the set of positives. We simply identify the items contained in negative pools as *negatives* (good items) and the rest as *positives* (defected items). Notice that d-disjunct property implies that each set of at most d positives corresponds uniquely to a test outcome vector, thus decoding test outcomes involves only a table lookup. The design of a d-disjunct matrix is thus naturally called a *non-adaptive pooling design*. We shall use this term interchangeably with the long "non-adaptive combinatorial group testing algorithm".

Let $S(\bar{d}, n)$ denotes the set of all subsets of n items (or columns) with size at most d, called the set of *samples*. For $s \in S(\bar{d}, n)$, let $P(s)$ denote the union of all columns corresponding to s, i.e. $P(s) = \bigcup_{i \in s} C_i$. A pooling design is e-error-detecting (correcting) if it can detect (correct) up to e errors in test outcomes. In other words, if a design is e-error-detecting then the test outcome vectors form a v-dimensional binary code with minimum Hamming distance at least $e+1$. Similarly, if a design is e-error-correcting then the test outcome vectors form a v-dimensional binary code with minimum Hamming distance at least $2e+1$. The following remarks are simple to see, however useful later on.

Remark 2.2. Suppose M has the property that for any $s, s' \in S(\bar{d}, n), s \neq s'$, $P(s)$ and $P(s')$ viewed as vectors have Hamming distance $\geq k$. In other words, $|P(s) \oplus P(s')| \geq k$ where $\oplus$ denotes the symmetric difference. Then, M is $(k-1)$-error-detecting and $\lfloor \frac{k-1}{2} \rfloor$-error-correcting.

Remark 2.3. M being d-disjunct is equivalent to the fact that for any set of $d+1$ distinct columns $C_{j_0}, \ldots C_{j_d}$ with one column (say C_{j_0}) designated, C_{j_0} has a 1 in some row where all C_{j_k}'s, $1 \leq k \leq d$ contain 0's.

2.2 Related Works

Previous works on error-tolerance designs are those of Dyachkov, Rykov and Rashad [8], Aigner [1], Muthukrishnan [13], Balding and Torney [3] and Macula [12, 11]. Dyachkov, Rykov and Rashad [8] derived upper and lower bounds for the test to item ratio given the number of tolerable errors, maximum number of positives, and the size of the population. Aigner [1] and Muthukrishnan [13], discussed optimal strategies when $d = 1$ and the number of errors is small, although in a slightly more general setting where each test outcome could be q-ary instead of binary. Balding and Torney [3] studied several instances of the problem when $d \leq 2$. In some specific case, they showed that an optimal strategy is possible if and only if certain Steiner system exists. In [12] Macula showed that his construction is error-tolerant with high probability, while in [11] he constructed e-error-tolerant d-disjunct matrices for certain values of e.

On construction of disjunct matrices, the most well-known method is to construct the matrix from *set packing designs*. This method was introduced by Kautz and Singleton [9] in the context of superimposed codes. A t-(v, k, λ) packing is a collection $\mathcal{F}$ of k-subsets of $[v]$ such that any t-subset of $[v]$ is contained in at most λ members of $\mathcal{F}$. When $\lambda = 1$ we can construct a $v \times |\mathcal{F}|$ d-disjunct matrix M from a t-$(v, k, 1)$ packing if $k > d(t-1)$. We simply index M's rows by members of $[v]$ and M's columns by members of $\mathcal{F}$, where there is a 1 in row $i \in [v]$ and column $F \in \mathcal{F}$ iff $i \in F$. Little is known about optimal set packing designs except for the case $t < 4$ (see, for example, [4, 14] for more details). Besides taking results directly from Design Theory, other works known on directly constructing d-disjunct matrices are those of Macula [10], Dýachkov, Macula, and Rykov [7].

3 Main Results

We first describe our d-disjunct matrices. Given integers $m \geq k > d \geq 1$. A matching of size l (i.e. it has l edges) is called an l-matching.

Definition 3.1. Let $M(m, k, d)$ be the 01-matrix whose rows are indexed by the set of all d-matchings on K_{2m}, and whose columns are indexed by the set of all k-matchings on K_{2m}. All matchings are to be ordered lexicographically. $M(m, k, d)$ has a 1 in row i and column j if and only if the i^{th} d-matching is contained in the j^{th} k-matching.

For q being a prime power, let $\mathbb{F}_q$ denote $GF(q)$. Let $\begin{bmatrix}\mathbb{F}_q^m\\ l\end{bmatrix}$ denote the set of all l-dimensional subspaces (l-subspaces for short) of the m-dimensional vector space on $\mathbb{F}_q$.

Definition 3.2. Let $M_q(m, k, d)$ be the 01-matrix whose rows (resp. columns) are indexed by elements of $\begin{bmatrix}\mathbb{F}_q^m\\ d\end{bmatrix}$ (resp. $\begin{bmatrix}\mathbb{F}_q^m\\ k\end{bmatrix}$). We also order elements of these set lexicographically. $M_q(m, k, d)$ has a 1 in row i and column j if and only if the i^{th} d-subspace is a subspace of the j^{th} k-subspace of $\mathbb{F}_q^m$.

We now show that $M(m, k, d)$ and $M_q(m, k, d)$ are d-disjunct.

Theorem 3.3. *Let* $g(m, l) = \binom{2m}{2l}\frac{(2l)!}{2^l l!}$, $v = g(m, d)$, *and* $n = g(m, k)$. *For* $m \geq k > d \geq 1$, $M(m, k, d)$ *is a* $v \times n$ d*-disjunct matrix with row weight* $g(m-d, k-d)$ *and column weight* $\binom{k}{d}$.

Proof. It is easy to see that $g(m,l)$ is the number of l-matchings of K_{2m}. Thus, $M(m,k,d)$ is a $v \times n$ matrix with row weight $g(m-d,k-d)$ and column weight $\binom{k}{d}$.

To show $M(m,k,d)$ is d-disjunct, we recall Remark 2.3. Consider $d+1$ distinct columns $C_{j_0}, C_{j_1}, \dots C_{j_d}$ of $M(m,k,d)$. Since all these columns are distinct k-matchings, for each $i \in [d]$ there exists an edge E_i of K_{2m} such that $E_i \in C_{j_0} \setminus C_{j_i}$. Hence, there exists a d-matching $R \subset C_{j_0}$ which contains all E_i's. To form R we simply add more edges in C_{j_0} to $\{e_i : 1 \in [d]\}$ if $|\{e_i : 1 \in [d]\}| < d$. Furthermore, since $R \not\subset C_{j_i}, \forall i \in [d]$, C_{j_0} has a 1 in row R where all other C_{j_i} contains 0.

□

Theorem 3.4. *Let* $\begin{bmatrix} m \\ l \end{bmatrix}_q := \frac{(q^m-1)(q^{m-1}-1)\dots(q^{m-l+1}-1)}{(q^l-1)(q^{l-1}-1)\dots(q-1)}$, $v = \begin{bmatrix} m \\ d \end{bmatrix}_q$, *and* $n = \begin{bmatrix} m \\ k \end{bmatrix}_q$. *For* $m \geq k > d \geq 1$, $M_q(m,k,d)$ *is a* $v \times n$ d*-disjunct matrix with row weight* $\begin{bmatrix} m-d \\ k-d \end{bmatrix}_q$ *and column weight* $\begin{bmatrix} k \\ d \end{bmatrix}_q$.

Proof. It is standard that the Gaussian coefficient $\begin{bmatrix} m \\ l \end{bmatrix}_q$ counts the number of l-subspaces of $\mathbb{F}_q^m$ (see, for example, Chapter 24 of [15]). The weight of any column C of $M_q(m,k,d)$ is the number of d-subspaces of C, hence it is $\begin{bmatrix} k \\ d \end{bmatrix}_q$. The weight $w(R)$ of any row R is the number of k-subspaces of $\mathbb{F}_q^m$ which contains the d-subspace R. To show $w(R) = \begin{bmatrix} m-d \\ k-d \end{bmatrix}_q$, we employ a standard trick, namely double counting. Let $I(m,k,d)$ be the number of ordered tuples $(v_1, \dots, v_{k-d})$ of $k-d$ vectors in $\mathbb{F}_q^m \setminus R$ such that each v_i is not in the span of R and other v_j's, $j \neq i$. Notice that $|\mathbb{F}_q^m| = q^m$ and $|R| = q^d$. Counting $I(m,k,d)$ directly, there are $q^m - q^d$ ways to choose v_1, then $q^m - q^{d+1}$ ways to choose v_2 and so on. Thus,

$$I(m,k,d) = (q^m - q^d)(q^m - q^{d+1}) \dots (q^m - q^{k-1}) \tag{1}$$

On the other hand, $(v_1, \dots, v_{k-d})$ can be obtained by first picking a k-subspace C of $\mathbb{F}_q^m$ which contains R in $w(R)$ ways, then $(v_1, \dots, v_{k-d})$ is chosen from $C \setminus R$ in $I(k,k,d)$ ways. This yields

$$I(m,k,d) = w(R) I(k,k,d) \tag{2}$$

Combining (1) and (2) gives $w(R) = \begin{bmatrix} m-d \\ k-d \end{bmatrix}_q$ as desired. The fact that $M_q(m,k,d)$ is d-disjunct can be shown in a similar fashion to the previous theorem.

$\square$

The following lemma tells us how to choose k so that the test to item ratio ($\frac{v}{n}$) is minimized. The proof is easy to see and we omit it here.

Lemma 3.5. *For l goes from 1 to m, we have*

(i) The sequence $g(m, l)$ is unimodal and gets its peak at $l = \lfloor m - \sqrt{\frac{m+1}{2}} \rfloor$.

(ii) The sequence $\begin{bmatrix} m \\ l \end{bmatrix}_q$ is unimodal and gets its peak at $l = \lfloor \frac{m}{2} \rfloor$.

Before exploring further properties of $M(m, k, d)$, we need a definition and a lemma.

Definition 3.6. Let $C_{j_0}, C_{j_1}, \dots C_{j_d}$ be any $d + 1$ distinct columns of $M(m, k, d)$. A d-matching R is said to be *private for* C_{j_0} with respect to $C_{j_1}, \dots C_{j_d}$ if $R \in C_{j_0} \setminus \bigcup_{i \in [d]} C_{j_i}$. Let $p(C_{j_0}; C_{j_1}, \dots C_{j_d})$ denote the number of private d-matchings of C_{j_0} with respect to $C_{j_1}, \dots C_{j_d}$.

Lemma 3.7. *Given integers $m > d \geq 1$ and any labeled simple graph G with $|V(G)| = m$ and $|E(G)| = d$. Then, the number of vertex covers of size d (or d-covers for short) of G is at least $d + 1$.*

Proof. Decompose G into its connected components. Suppose $G_1, \dots, G_x$ are connected components which are not trees, and $G'_1, \dots, G'_y$ are the rest of the components. Isolated points are also considered to be trees, so that G'_i is a tree for all $i \in [y]$. For $i = 1, \dots x$, let $v_i = |V(G_i)|$ and $e_i = |E(G_i)|$. For $i = 1, \dots y$, let $v'_i = |V(G'_i)|$ and $e'_i = |E(G'_i)|$. The following equations are straight from the definitions :

$$\sum_{i \in [x]} v_i + \sum_{i \in [y]} v'_i = m \tag{3}$$

$$\sum_{i \in [x]} e_i + \sum_{i \in [y]} e'_i = d \tag{4}$$

hence,

$$0 \leq \sum_{i \in [x]} e_i - \sum_{i \in [x]} v_i = y - (m - d) \tag{5}$$

Observe that for any connected simple graph H, picking any $|V(H)|-1$ vertices out of $V(H)$ gives us a vertex cover. Hence, the number of $(|V(H)|-1)$-covers of H is at least $\binom{|V(H)|}{|V(H)|-1} = |V(H)|$. To this end, notice that a d-cover of G could be formed by two methods as follows.

(a) *Method 1*. For each $i \in [x]$, pick in v_i ways a $(v_i - 1)$-cover for G_i, then cover all other G_j, $j \neq i$, with all of their vertices. We have used up $(\sum_{i\in[x]} v_i) - 1$ vertices, and need $d - (\sum_{i\in[x]} v_i) + 1$ more to cover all edges of the G_i''s. Firstly, there should be enough number of vertices left. Indeed,

$$\sum_{i\in[y]} v_i' = m - \sum_{i\in[x]} v_i \geq d + 1 - \sum_{i\in[x]} v_i$$

Secondly, since each G_i' can be covered by $v_i' - 1$ vertices, to cover all G_i''s we need at most $\sum_{i\in[x]}(v_i' - 1)$ vertices. (3) and (5) assure that

$$\sum_{i\in[y]}(v_i' - 1) = m - \sum_{i\in[x]} v_i - y < d + 1 - \sum_{i\in[x]} v_i$$

In conclusion, this method gives us at least $(\sum_{i=1}^{x} v_i)$ d-covers for G.

(b) *Method 2*. This time, we are greedier by first taking all vertices in G_i's, $i \in [x]$ to cover them. $a = d - \sum_{i\in[x]} v_i$ vertices are needed to cover the rest. These a vertices can be chosen as follows. For each $(m-d)$-subset Y of $[y]$, cover each G_i', $i \in Y$ with $v_i' - 1$ vertices, then cover each G_i', $i \notin Y$ using all of its vertices. Indeed, the total number of vertices used is

$$\sum_{i\in Y}(v_i' - 1) + \sum_{i\notin Y} v_i' = \sum_{i\in[y]} v_i' - |Y| = (m - \sum_{i=1}^{x} v_i) - (m - d) = a$$

Moreover, obviously there are at least $\prod_{i\in Y} v_i'$ ways to pick d-covers for each particular Y. In total, the number of d-covers formed by Method 2 is at least $\displaystyle\sum_{Y\in\binom{[y]}{m-d}} \prod_{i\in Y} v_i'$. Noticing that $y \geq m - d \geq 1$, we have

$$\begin{aligned}\sum_{Y\in\binom{[y]}{m-d}}\prod_{i\in Y}v_i' &= \sum_{Y\in\binom{[y]}{m-d}}\prod_{i\in Y}(e_i'+1)\\ &\geq \sum_{i\in[y]}e_i' + \binom{y}{m-d}\\ &\geq (\sum_{i\in[y]}v_i' - y) + (y-m+d+1)\\ &= d+1-\sum_{i\in[x]}v_i\end{aligned}$$

Hence, methods 1 and 2 combined yields at least $(d+1)$ different d-covers for G.

□

Theorem 3.8. *Given $m > d \geq 1$, and any set of $d{+}1$ distinct columns $C_{j_0}, C_{j_1}, \ldots C_{j_d}$ of $M(m,m,d)$, then $p(C_{j_0}; C_{j_1}, \ldots C_{j_d}) \geq d+1$.*

Proof. Observe that for each $i \in [d]$, $C_{j_0} \cup C_{j_i}$ is a loopless multigraph which is 2-regular. $C_{j_0} \cup C_{j_i}$ consists of cycles with even lengths. Moreover, $C_{j_0} \neq C_{j_i}$ implies that $C_{j_0} \cup C_{j_i}$ must have a cycle of length at least 4; consequently, $|C_{j_0} \setminus C_{j_i}| \geq 2$, $\forall i \in [d]$.

For each $i \in [d]$, choose arbitrarily $E_i \subseteq C_{j_0} \setminus C_{j_i}$ so that $|E_i| = 2$. Let G be the graph with $V(G) = C_{j_0}$, $E(G) = \{E_1, \ldots E_d\}$. Then, G is a simple graph having m vertices and $\leq d$ edges. $|E(G)| \leq d$ because the E_i's are not necessarily distinct. Any d-subset R of C_{j_0} such that $R \cap E_i \neq \emptyset$, $\forall i$ is a private d-matching of C_{j_0} with respect to $C_{j_1}, \ldots C_{j_d}$. Note that R is nothing but a d-cover of G. To show $p(C_0; C_1, \ldots C_d) \geq d+1$, we shall show that the number of d-covers of G is at least $d+1$. Since adding more edges into G can only decrease the number of d-covers, we can safely assume that G has exactly d edges and apply Lemma 2.

□

Corollary 3.9. *Given integers $m > d \geq 1$, the following holds :*

(i) $M(m,m,d)$ is d-error-detecting and $\lfloor \frac{d}{2} \rfloor$-error-correcting.

(ii) Moreover, if the number of positives is known to be exactly d, then $M(m,m,d)$ is $(2d+1)$-error-detecting and d-error-correcting.

Proof. For any $s, s' \in S(\bar{d}, n)$, $s \neq s'$, without loss of generality we can assume there exists $C_{j_0} \in s \setminus s'$. Theorem 3 implies $|P(s) \oplus P(s')| \geq d+1$, hence Remark 1 shows (i). If the number of positives is exactly d, we need to only consider $|s| = |s'| = d$; hence there exists $C_{j_0} \in s \setminus s'$ and $C'_{j_0} \in s' \setminus s$. This time, Theorem 3 implies $|P(s) \oplus P(s')| \geq 2d+2$. Again, Remark 1 yields (ii).

□

Corollary 3.10. *Given integers $m > d \geq 1$, then there exists a binary error-correcting code of dimension $g(m,d)$ and size $\binom{g(m,m)}{d}$ with minimum Hamming distance $2d+2$.*

Proof. The code can be constructed by taking all the unions of d columns in $M(m,m,d)$. Clearly, it is $(2d+1)$-error-detecting and d-error-correcting.

□

Borrowing an idea from Macula [12], we get the following algorithm which uses $M(m,k,2)$ for the at most d positive problem, and show that with very high probability, our algorithm gives the correct answer. Notice that each row of $M(m,k,2)$ is a 2-matching consisting of some two parallel edges $(e1, e2)$ of K_{2m}. We pay attention to $M(m,k,2)$ because it has good $\frac{v}{n}$ ratio.

Algorithm 3.11. Use $M(m,k,2)$ to design the pools as usual. For each edge $E \in E(K_{2m})$ such that the total number of positive outcomes involving E is $k-1$, i.e. $|\{(E,E') :$ the test (E,E') is positive $\}| = k-1$, identify the item $C = \{E\} \cup \{E' : (E,E')$ is positive $\}$ as a positive.

Theorem 3.12. *Algorithm 3.11 gives correct answer with probability $P(m,k,d)$ where*

$$P(m,k,d) \geq \left[\frac{\sum_{j=1}^{k} (-1)^{j+1} \binom{k}{j} \binom{\sum_{i=0}^{j} (-1)^i \binom{j}{i} g(m-i,k-i)}{d-1}}{\binom{g(m,k)-1}{d-1}} \right]^d$$

For example, $P(8,6,9) \geq 98.5\%$. This means that we could use $M(8,6,2)$, which has dimension 5460×18918900, to find at most 9 positives in a population of 18918900 items using only 5460 tests with 98.5% chance of success.

Proof. Given a set of d distinct columns $C_{j_1}, C_{j_2}, \ldots C_{j_d}$. $E \in E(K_{2m})$ is called a *mark* of C_{j_i} if E is a private 1-matching of C_{j_i} with respect to $\{C_{j_l}, l \in [d] \setminus \{i\}\}$, in which case C_{j_i} is said to be *marked*. If C_{j_i} is marked by E then exactly $k-1$ tests involving E and another edge in C_{j_i} is positive. Consequently, algorithm 3.11 gives correct answer if the set of d positives is a marked set, i.e. every element is marked.

The probability that algorithm 3.11 gives correct answer is thus the probability that a random d set of columns of $M(m,k,2)$ is marked. For a fixed C_{j_1}, there are $\binom{g(m,k)-1}{d-1}$ ways to pick the other $d-1$ columns. Let X_i be the event that C_i is marked relative to the other $d-1$ columns, then

$$P(m,k,d) = P(X_1)P(X_2|X_1)P(X_3|X_1,X_2)\ldots \geq (P(X_1))^d$$

To calculate $P(X_1)$, we count number of ways to pick $d-1$ columns other than C_{j_1} such that C_{j_1} is marked by some $E \in C_{j_1}$. Let A_i be the collection of all $(d-1)$-sets of columns other than C_{j_1} such that $E_i \in C_{j_1}$ marks C_{j_1} with respect to A_i. The answer is then $|\bigcup\{A_i, 1 \leq i \leq k\}|$. This number can be obtained by applying inclusion-exclusion principle twice. Dividing it by $\binom{g(m,k)-1}{d-1}$ gives us $P(X_1)$ and proves the theorem.

□

4 Discussions

We have given the constructions of two different classes of pooling designs. $M(m,k,d)$ has good performance when the number of positives is small comparing to the number of items. Deterministically, a larger ratio of positives to items is sometime preferred. Probabilistically, however, $M(m,k,2)$ could be used to solve the $S(\bar{d},n)$ problem with very high probability of success. The main strength of this construction is that $M(m,m,d)$ is d-error-detecting. It also yields the construction of a d-error-correcting code. $M_q(m,k,d)$ is the q-analogue of the construction given by Macula [10]. An interesting question is: "what is the q-analogue of a matching?"

One could think of several different variations of the matching idea. For example, a possible generalization is to index the rows (resp. columns) of a matrix $M(m,k,d,l)$ with all graphs having d (resp. k) edges whose vertex degrees are $\leq l$. $M(m,k,d)$ is nothing but $M(m,k,d,1)$. Further investigations in this direction might lead to better designs.

Lastly, in reality given a specific problem with certain parameters, m and k have to be chosen appropriately to suit one's need. More careful analysis need to be done to help pick the *best* m and k given n, d and/or any other constraints from practice. We need some reasonably good asymptotic formulas to estimate them.

References

[1] M. AIGNER, *Searching with lies*, J. Combin. Theory Ser. A, 74 (1996), pp. 43–56.

[2] D. J. BALDING, W. J. BRUNO, E. KNILL, AND D. C. TORNEY, *A comparative survey of non-adaptive pooling designs*, in Genetic mapping and DNA sequencing (Minneapolis, MN, 1994), Springer, New York, 1996, pp. 133–154.

[3] D. J. BALDING AND D. C. TORNEY, *Optimal pooling designs with error detection*, J. Combin. Theory Ser. A, 74 (1996), pp. 131–140.

[4] T. BETH, D. JUNGNICKEL, AND H. LENZ, *Design theory*, Cambridge University Press, Cambridge, 1986.

[5] W. J. BRUNO, E. KNILL, D. J. BALDING, D. C. BRUCE, N. A. DOGGETT, W. W. SAWHILL, R. L. STALLINGS, C. C. WHITTAKER, AND D. C. TORNEY, *Efficient pooling designs for library screening*, Genomics, (1995), pp. 21–30.

[6] D. Z. DU AND F. K. HWANG, *Combinatorial group testing and its applications*, World Scientific Publishing Co. Inc., River Edge, NJ, 1993.

[7] A. G. DYACHKOV, A. J. MACULA, JR., AND V. V. RYKOV, *New constructions of superimposed codes*, IEEE Trans. Inform. Theory, 46 (2000), pp. 284–290.

[8] A. G. DYACHKOV, V. V. RYKOV, AND A. M. RASHAD, *Superimposed distance codes*, Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform., 18 (1989), pp. 237–250.

[9] W. H. KAUTZ AND R. C. SINGLETON, *Nonrandom binary superimposed codes*, IEEE Trans. Inf. Theory, 10 (1964), pp. 363–377.

[10] A. J. MACULA, *A simple construction of d-disjunct matrices with certain constant weights*, Discrete Math., 162 (1996), pp. 311–312.

[11] ——, *Error-correcting nonadaptive group testing with d^e-disjunct matrices*, Discrete Appl. Math., 80 (1997), pp. 217–222.

[12] ——, *Probabilistic nonadaptive group testing in the presence of errors and dna library screening*, Annals of Combinatorics, (1999), pp. 61–69.

[13] S. MUTHUKRISHNAN, *On optimal strategies for searching in presence of errors*, in Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (Arlington, VA, 1994), New York, 1994, ACM, pp. 680–689.

[14] H. Q. NGO AND D. Z. DU, *A survey on combinatorial group testing algorithms with applications to dna library screening*, in DIMACS: Series in Discrete Mathematics and Theoretical Computer Science/DU2, Providence, RI, 2000, Amer. Math. Soc.

[15] J. H. VAN LINT AND R. M. WILSON, *A course in combinatorics*, Cambridge University Press, Cambridge, 1992.

[16] J. K. WOLF, *Born again group testing: multiaccess communications*, IEEE Transaction on Information Theory, IT-31 (1985), pp. 185–191.