Hypothesis Test for Upper Bound on the Size of Random Defective Set

Arkadii D'yachkov*, Ilya Vorobyev*, Nikita Polyanskii†, Vladislav Shchukin*
*Lomonosov Moscow State University, Moscow, Russia
†Huawei Technologies Co. Ltd., Moscow, Russia
agd-msu@yandex.ru, vorobyev.i.v@yandex.ru, polianskii.nikita@huawei.com, vpike@mail.ru

***Abstract*—Let $1 \le s < t$, $N \ge 1$ be fixed integers and a complex electronic circuit of size t is said to be an s-active, $s \ll t$, and can work as a system block if not more than s elements of the circuit are defective. Otherwise, the circuit is said to be an s-defective and should be replaced by a similar s-active circuit. Suppose that there exists a possibility to run N non-adaptive group tests to check the s-activity of the circuit. As usual, we say that a (disjunctive) group test yields the positive response if the group contains at least one defective element. In this paper, we will interpret the unknown set of defective elements as a random set and discuss upper bounds on the error probability of the hypothesis test for the null hypothesis $\{H_0 :$ the circuit is s-active$\}$ verse the alternative hypothesis $\{H_1 :$ the circuit is s-defective$\}$. Along with the conventional decoding algorithm based on the known random set of positive responses and disjunctive s-codes, we consider a T-weight decision rule which is based on the simple comparison of a fixed threshold T, $1 \le T < N$, with the known random number of positive responses p, $0 \le p \le N$.**

Keywords: Hypothesis test, group testing, disjunctive codes, maximal error probability, error exponent, random coding bounds.

I. Statement of Problem

Let $N \ge 2$, $t \ge 2$, s and T be integers, where $1 \le s < t$ and $1 \le T < N$. The symbol $\triangleq$ denotes the equality by definition, $|A|$ – the size of the set A and $[N] \triangleq \{1, 2, \ldots, N\}$ – the set of integers from 1 to N. A binary $(N \times t)$-matrix

$$X = \|x_i(j)\|, \quad x_i(j) = 0, 1, \quad i \in [N],\ j \in [t],$$
$$\boldsymbol{x}(j) \triangleq (x_1(j), \ldots, x_N(j)), \quad \boldsymbol{x}_i \triangleq (x_i(1), \ldots, x_i(t)),$$

with t columns (*codewords*) $\boldsymbol{x}(j)$, $j \in [t]$, and N rows $\boldsymbol{x}_i$, $i \in [N]$, is called a *binary code of length N and size $t = \lfloor 2^{RN} \rfloor$*, where a fixed parameter $R > 0$ is called a *rate* of the code X. The number of 1's in a binary column $\boldsymbol{x} = (x_1, \ldots, x_N) \in \{0,1\}^N$, i.e., $|\boldsymbol{x}| \triangleq \sum_{i=1}^N x_i$, is called a *weight* of $\boldsymbol{x}$. A code X is called a *constant weight code of weight* w, $1 \le w < N$, if for any $j \in [t]$, the weight $|\boldsymbol{x}(j)| = w$. The conventional symbol $\boldsymbol{u} \bigvee \boldsymbol{v}$ will be used to denote the disjunctive (Boolean) sum of binary columns $\boldsymbol{u}, \boldsymbol{v} \in \{0,1\}^N$. We say that a column $\boldsymbol{u}$ *covers* a column $\boldsymbol{v}$ if $\boldsymbol{u} \bigvee \boldsymbol{v} = \boldsymbol{u}$.

A. Disjunctive and Threshold Disjunctive Codes

Definition 1. [4]. A code X is called a *disjunctive s-code*, $s \in [t-1]$, if the disjunctive sum of any s-subset of codewords of X covers those and only those codewords of X which are the terms of the given disjunctive sum.

Let $\mathcal{S}$, $\mathcal{S} \subset [t]$, be an arbitrary fixed collection of defective elements of size $|\mathcal{S}|$. For a binary code X and collection $\mathcal{S}$, define the binary *response vector* of length N, namely:

$$\boldsymbol{x}(\mathcal{S}) \triangleq \begin{cases} \bigvee\limits_{j \in \mathcal{S}} \boldsymbol{x}(j) & \text{if} \quad \mathcal{S} \ne \emptyset, \\ (0, 0, \ldots, 0) & \text{if} \quad \mathcal{S} = \emptyset. \end{cases}$$

In the classical problem of *non-adaptive group testing*, we describe N tests as a binary $(N \times t)$-matrix $X = \|x_i(j)\|$, where a column $\boldsymbol{x}(j)$ corresponds to the j-th element, a row $\boldsymbol{x}_i$ corresponds to the i-th test and $x_i(j) \triangleq 1$ if and only if the j-th element is included into the i-th testing group. The result of each test equals 1 if at least one defective element is included into the testing group and 0 otherwise, so the column of results is exactly equal to the response vector $\boldsymbol{x}(\mathcal{S})$. Definition 1 of disjunctive s-code X gives the important sufficient condition for the evident identification of any unknown collection of defective elements $\mathcal{S}$ if the number of defective elements $|\mathcal{S}| \le s$. In this case, the identification of the unknown $\mathcal{S}$ is equivalent to discovery of all codewords of code X covered by $\boldsymbol{x}(\mathcal{S})$, and its complexity is equal to the code size t. Note that this algorithm also allows us to check s-activity of the circuit defined in the abstract. Moreover, it is easy to prove by contradiction that every code X which allows to check s-activity of the circuit without error is disjunctive s-code. Indeed, if code X is not disjunctive s-code, then there exist a set $\mathcal{S} \subset [t]$, $|\mathcal{S}| = s$, and a number $j \in [t] \setminus \mathcal{S}$ such that $\boldsymbol{x}(\mathcal{S}) = \boldsymbol{x}(\mathcal{S} \cup \{j\})$.

Proposition 1. *The results of non-adaptive group tests specified by code X allow to check s-activity of the circuit if and only if X is disjunctive s-code.*

Definition 2. Let s, $s \in [t-1]$, and T, $T \in [N-1]$, be arbitrary fixed integers. A disjunctive s-code X of length N and size t is said to be a disjunctive s-code with *threshold* T (or, briefly, *s^T-code*) if the disjunctive sum of any $\le s$ codewords of X has weight $\le T$ and the disjunctive sum of any $\ge s+1$ codewords of X has weight $\ge T+1$.

Obviously, for any s and T, the definition of s^T-code gives a sufficient condition for code X applied to the group testing problem described in the abstract of our paper. In this case, only on the base of the known *number of positive responses* $|\boldsymbol{x}(\mathcal{S})|$, we decide that the controllable circuit identified by an unknown collection $\mathcal{S}$, $\mathcal{S} \subset [t]$, is s-active, i.e., the unknown size $|\mathcal{S}| \le s$ if $|\boldsymbol{x}(\mathcal{S})| \le T$ (s-defective, i.e., the unknown size $|\mathcal{S}| \ge s+1$, if $|\boldsymbol{x}(\mathcal{S})| \ge T+1$).

Remark 1. The concept of s^T-codes was motivated by troubleshooting in complex electronic circuits using a non-adaptive identification scheme which was considered in [8].

Remark 2. A similar model of special disjunctive s-codes was considered in [3], where the conventional disjunctive s-code is supplied with an additional condition: the weight $|\boldsymbol{x}(\mathcal{S})|$ of the response vector of any subset $\mathcal{S}$, $\mathcal{S} \subset [t]$, $|\mathcal{S}| \le s$, is at most T. Note that these codes have a weaker condition than our s^T-codes. In [3] authors motivate their group testing model with bounded weight of the response vector by a risk for the safety of the persons, who perform tests, in some contexts, when the number of positive test results is too large.

B. Hypothesis Test

Let a circuit of size t be identified by an unknown collection $\mathcal{S}_{un}$, $\mathcal{S}_{un} \subset [t]$, of defective elements of an unknown size $|\mathcal{S}_{un}|$. For a reasonable probabilistic interpretation, we assume that the different collections of defective elements of the same size are *equiprobable*. That is why, we set that the *probability distribution* of the random collection $\mathcal{S}_{un}$, $\mathcal{S}_{un} \subset [t]$, is identified by an unknown probability vector $\boldsymbol{p} \triangleq (p_0, p_1, \ldots, p_t)$, $p_k \ge 0$, $k = 0, 1, \ldots, t$, $\sum_{k=0}^{t} p_k = 1$, as follows:

$$\Pr\{\mathcal{S}_{un} = \mathcal{S}\} \triangleq \frac{p_{|\mathcal{S}|}}{\binom{t}{|\mathcal{S}|}} \quad \text{for any} \quad \mathcal{S} \subseteq [t]. \tag{1}$$

Let X be an arbitrary code of size t and length N. Given any fixed integer parameters s, $1 \le s < t$, and T, $1 \le T < N$, introduce the null hypothesis $\{H_0 : |\mathcal{S}_{un}| \le s\}$ (the circuit is s-active) verse the alternative $\{H_1 : |\mathcal{S}_{un}| \ge s+1\}$ (the circuit is s-defective), and consider the following *T-weight decision rule* motivated by Definition 2, namely:

$$\begin{cases} \text{accept } \{H_0 : |\mathcal{S}_{un}| \le s\} & \text{if } |\boldsymbol{x}(\mathcal{S}_{un})| \le T, \\ \text{accept } \{H_1 : |\mathcal{S}_{un}| > s\} & \text{if } |\boldsymbol{x}(\mathcal{S}_{un})| > T. \end{cases} \tag{2}$$

Introduce a *maximal error probability* of the decision rule (2):

$$\begin{aligned} \varepsilon_s(T, \boldsymbol{p}, X) \triangleq \max \big\{ & \Pr\{\text{accept } H_1 \big| H_0\}, \\ & \Pr\{\text{accept } H_0 \big| H_1\} \big\}, \end{aligned} \tag{3}$$

where the conditional probabilities in the right-hand side of (3) are defined by (1)-(2). Note that the number $\varepsilon_s(T, \boldsymbol{p}, X) = 0$ if and only if the code X is s^T-code. Denote by $t_s(N, T)$ the maximal size of s^T-codes of length N. For a parameter τ, $0 < \tau < 1$, introduce the rate of $s^{\lfloor \tau N \rfloor}$-codes:

$$R_s(\tau) \triangleq \overline{\lim_{N\to\infty}} \frac{\log_2 t_s(N, \lfloor \tau N \rfloor)}{N} \ge 0.$$

Definition 3. Let τ, $0 < \tau < 1$, and a parameter R, $R > R_s(\tau)$, be fixed. For the maximal error probability $\varepsilon_s(T, \boldsymbol{p}, X)$, defined by (1)-(3), consider the function

$$\varepsilon_s^N(\tau, R) \triangleq \max_{\boldsymbol{p}} \left\{ \min_{X} \varepsilon_s(\lfloor \tau N \rfloor, \boldsymbol{p}, X) \right\}, \tag{4}$$

where the minimum is taken over all codes X of length N and size $t = \lfloor 2^{RN} \rfloor$. The number $\varepsilon_s^N(\tau, R) > 0$ does not depend on the unknown probability vector $\boldsymbol{p}$ and can be called the *universal* error probability of the decision rule (2). The corresponding *error exponent*

$$E_s(\tau, R) \triangleq \overline{\lim_{N\to\infty}} \frac{-\log_2 \varepsilon_s^N(\tau, R)}{N}, \quad s \ge 1, \tag{5}$$

identifies the asymptotic behavior of the maximal error probability of the decision rule (2):

$$\exp_2\{-N\,[E_s(\tau, R) + o(1)]\}, \quad N \to \infty, \quad \text{if } E_s(\tau, R) > 0.$$

Along with (2) we introduce the *disjunctive decision rule* based on the conventional algorithm:

$$\begin{cases} \text{accept } H_0 & \text{if } \boldsymbol{x}(\mathcal{S}_{un}) \text{ covers} \le s \text{ columns of } X, \\ \text{accept } H_1 & \text{if } \boldsymbol{x}(\mathcal{S}_{un}) \text{ covers} > s \text{ columns of } X. \end{cases}$$

For a fixed code rate R, $R > 0$, the error exponent for disjunctive decision rule $E_s(R)$ is defined similarly to (3)-(5). The function $E_s(R)$ was firstly introduced in our paper [4], where we proved

Theorem 1. [4]. *If $R \ge 1/s$, then $E_s(R) = 0$.*

Remark 3. In our paper we will focus on the test of hypotheses H_0 and H_1, provided that the unknown defective set $\mathcal{S}_{un}$ is a random set with probability distribution (1). A similar statistical problem of constructing confidence interval for the size $|\mathcal{S}_{un}|$ of unknown (nonrandom) defective set $\mathcal{S}_{un}$ was considered in [2], [7]. The authors of [2] present a randomized algorithm that uses $G(\epsilon, c)\log_2 t$ non-adaptive tests and produces the statistic $\hat{s}$, that satisfies the following properties: probability $\Pr\{\hat{s} < |\mathcal{S}_{un}|\}$ is upper bounded by a small parameter $\epsilon \ll 1$ and the expected value of $\hat{s}/|\mathcal{S}_{un}|$ is upper bounded by a number $c > 1$. In [7] an adaptive randomized algorithm is proposed (algorithm is called adaptive if the next test is constructed based on the results of the previous tests). It uses at most $2\log_2\log_2|\mathcal{S}_{un}| + O(\frac{1}{\delta^2}\log_2\frac{1}{\epsilon})$ adaptive tests and estimates $|\mathcal{S}_{un}|$ up to a multiplicative factor of $1 \pm \delta$ with error probability $\le \epsilon$. Note that the estimating of $|\mathcal{S}_{un}|$ is a subtask of a non-standard group testing problem of identification $\mathcal{S}_{un}$ where is no restriction $|\mathcal{S}_{un}| \le s$. Another approach to solving this problem was considered in paper [1] where the authors propose to run a fixed non-adaptive tests on the first stage and to test individually each of the unresolved after stage 1 elements on the second stage. For probability distribution

$$\Pr\{j \in \mathcal{S}_{un}\} = p, \quad \forall i \in [t],$$

and some dependencies $p \triangleq p(t)$, the lower and upper bounds on asymptotics of the expected number of tests in described 2-stage procedure are obtained in [1].

II. Lower Bounds on Error Exponents

In this Section, we formulate and compare random coding lower bounds for the both of error exponents $E_s(R)$ and $E_s(\tau, R)$. These bounds were proved applying the random coding method based on the ensemble of constant-weight codes. A parameter Q in formulations of theorems 2-3 means the relative weight of codewords of constant-weight codes. Introduce the standard notations

$$h(Q) \triangleq -Q\log_2 Q - (1-Q)\log_2[1-Q],$$
$$[x]^+ \triangleq \max\{x, 0\}.$$

In [4], we established

Theorem 2. [4]. *The error exponent $E_s(R) \ge \underline{E}_s(R)$ where the random coding lower bound*

$$\underline{E}_s(R) \triangleq \max_{0<Q<1} \min_{Q\le q<\min\{1,sQ\}} \left\{\mathcal{A}(s,Q,q) + [h(Q) - qh(Q/q) - R]^+\right\},$$
$$\mathcal{A}(s,Q,q) \triangleq (1-q)\log_2(1-q) + q\log_2\left[\frac{Qy^s}{1-y}\right] + sQ\log_2\frac{1-y}{y} + sh(Q), \quad (6)$$

and y is the unique root of the equation

$$q = Q\frac{1-y^s}{1-y}, \quad 0<y<1. \quad (7)$$

In addition, as $s\to\infty$ and $R \le \frac{\ln 2}{s}(1+o(1))$, the lower bound $\underline{E}_s(R) > 0$.

In Section III we prove

Theorem 3. **1.** *The error exponent $E_s(\tau,R) \ge \underline{E}_s(\tau,R)$ where the random coding bound $\underline{E}_s(\tau,R)$ does not depend on $R>0$ and has the form:*

$$\underline{E}_s(\tau,R) \triangleq \max_{1-(1-\tau)^{1/(s+1)}<Q<1-(1-\tau)^{1/s}} \min\left\{\mathcal{A}'(s,Q,\tau),\, \mathcal{A}(s+1,Q,\tau)\right\}, \quad (8)$$

$$\mathcal{A}'(s,Q,\tau) \triangleq \begin{cases}\mathcal{A}(s,Q,\tau), & \textit{if } Q\le\tau\le sQ,\\ \infty, & \textit{otherwise},\end{cases} \quad (9)$$

where $\mathcal{A}(s,Q,\tau)$ is defined by (6)-(7).
2. *As $s\to\infty$ the optimal value of $\underline{E}_s(\tau,R)$*

$$\underline{E}_{Thr}(s) \triangleq \max_{0<\tau<1}\underline{E}_s(\tau,R) \ge \frac{\log_2 e}{4s^2}(1+o(1)),\ s\to\infty. \quad (10)$$

It is possible to use the decision rule (2) with any value of parameter T. The numerical values of the optimal error exponent $\underline{E}_{\text{Thr}}(s)$ along with the corresponding optimal threshold parameter $\tau=\tau(s)$ and the constant-weight code ensemble parameter $Q=Q(s)$ are presented in Table I. Table I contains the numbers $\underline{E}_s(0) \triangleq \lim\limits_{R\to0}\underline{E}_s(R)$ and $R_{\text{Thr}}(s) \triangleq \sup\{R : \underline{E}_s(R) > \underline{E}_{\text{Thr}}(s)\}$ as well. Theorems 1-3 show that, for large values of the rate parameter R, $R > R_{\text{Thr}}(s)$, the weight decision rule (2) has an advantage over the disjunctive decision rule as $N\to\infty$.

III. Proof of Theorem 3

Proof of Statement 1. For a fixed code X and parameters s and T, introduce the sets $B_k^i(T,X)$, $i=1,2$, $k=0,1,\dots,t$, of collections $\mathcal{S}$, $\mathcal{S}\subset[t]$, $|\mathcal{S}|=k$, as follows:

$$B_k^1(T,X) \triangleq \{\mathcal{S} : \mathcal{S}\subset[t],\ |\mathcal{S}|=k,\ |\boldsymbol{x}(\mathcal{S})|\ge T+1\},$$
$$B_k^2(T,X) \triangleq \{\mathcal{S} : \mathcal{S}\subset[t],\ |\mathcal{S}|=k,\ |\boldsymbol{x}(\mathcal{S})|\le T\}. \quad (11)$$

TABLE I
THE NUMERICAL VALUES OF $\underline{E}_{\text{THR}}(s)$ AND $R_{\text{THR}}(s)$

s	2	3	4	5	6
$\underline{E}_{\text{Thr}}(s)$	0.1380	0.0570	0.0311	0.0196	0.0135
$\tau(s)$	0.2065	0.1365	0.1021	0.0816	0.0679
$Q(s)$	0.1033	0.0455	0.0255	0.0163	0.0113
$\underline{E}_s(0)$	0.3651	0.2362	0.1754	0.1397	0.1161
$R_{\text{Thr}}(s)$	0.2271	0.1792	0.1443	0.1201	0.1027

Then the probability (3) is represented as

$$\varepsilon_s(T,\boldsymbol{p},X) \triangleq \max\left\{\sum_{k=0}^{s}\frac{p_k}{\sum\limits_{l=0}^{s}p_l}\frac{|B_k^1(T,X)|}{\binom{t}{k}},\ \sum_{k=s+1}^{t}\frac{p_k}{\sum\limits_{l=s+1}^{t}p_l}\frac{|B_k^2(T,X)|}{\binom{t}{k}}\right\}. \quad (12)$$

One can see that, for sets (11) and any k, $0\le k<t$, the inequalities

$$|B_{k+1}^1(T,X)| \ge \frac{t-k}{k+1}|B_k^1(T,X)| \quad \text{and}$$
$$|B_k^2(T,X)| \ge \frac{k+1}{t-k}|B_{k+1}^2(T,X)|$$

hold. Therefore, from (12) it follows that for any code X, the maximum $\max_{\boldsymbol{p}}\varepsilon_s(T,\boldsymbol{p},X)$ in the right-hand side of (4) is attained at the probability distribution $\boldsymbol{p}=(p_0,p_1,\dots,p_t)$ such that $p_s=p_{s+1}=1/2$ and $p_j=0$, $j\notin\{s,s+1\}$. Therefore, the definition (4) is equivalent to

$$\varepsilon_s^N(\tau,R) \triangleq \min_{X:\,t=\lfloor 2^{RN}\rfloor}\varepsilon_s(\lfloor\tau N\rfloor, X), \quad R>R_s(\tau), \quad (13)$$

where

$$\varepsilon_s(T,X) \triangleq \max\left\{\frac{|B_s^1(T,X)|}{\binom{t}{s}},\ \frac{|B_{s+1}^2(T,X)|}{\binom{t}{s+1}}\right\}. \quad (14)$$

Fix $s\ge2$, $0<\tau<1$, $R>R_s(\tau)$ and a parameter Q, $0<Q<1$. The bound (8) is obtained by the method of random coding over the ensemble of binary constant-weight codes [6] defined as the ensemble $E(N,t,Q)$ of binary codes X of length N and size $t=\lfloor 2^{RN}\rfloor$, where the codewords are chosen independently and equiprobably from the set consisting of all $\binom{N}{\lfloor QN\rfloor}$ codewords of a fixed weight $\lfloor QN\rfloor$.

For the ensemble $E(N,t,Q)$, denote the expectation of the error probability (14) by

$$\mathcal{E}_s^N(\tau,Q,R) \triangleq \mathbb{E}\Big[\varepsilon_s(\lfloor\tau N\rfloor, X)\Big]. \quad (15)$$

Note that there exists a code X of length N and rate R such that its maximal error probability (14) is upper bounded by $\mathcal{E}_s^N(\tau,Q,R)$, and due to (13) the following lower bound on the error exponent (5) of the decision rule (2) is given:

$$E_s(\tau,R) \ge \max_{0<Q<1}\lim_{N\to\infty}\frac{-\log_2\mathcal{E}_s^N(\tau,Q,R)}{N}. \quad (16)$$

Further we show that the limit in the right-hand side of (16) exists and its maximum by Q equals (8).

The cardinality of set $B_s^1(\lfloor \tau N\rfloor, X)$ can be presented through indicator functions:

$$\left|B_s^1(\lfloor \tau N\rfloor, X)\right| = \sum_{\mathcal{S}\in[t], |\mathcal{S}|=s} \mathbb{1}\{\mathcal{S}\in B_s^1(\lfloor \tau N\rfloor, X)\}.$$

Therefore, the expectation of the cardinality $\left|B_s^1(\lfloor \tau N\rfloor, X)\right|$ (and similarly, $\left|B_{s+1}^2(\lfloor \tau N\rfloor, X)\right|$) equals

$$\mathbb{E}\Big[\left|B_s^1\right|\Big] = \binom{t}{s} \Pr\left\{\mathcal{S}\in B_s^1 \,\middle|\, |\mathcal{S}|=s\right\}$$
$$\left(\mathbb{E}\Big[\left|B_{s+1}^2\right|\Big] = \binom{t}{s+1} \Pr\left\{\mathcal{S}\in B_{s+1}^2 \,\middle|\, |\mathcal{S}|=s+1\right\}\right). \tag{17}$$

For the ensemble $E(N,t,Q)$, denote the probabilities $\Pr\left\{\mathcal{S} \in B_s^1(\lfloor \tau N\rfloor, X)\,\middle|\,|\mathcal{S}| = s\right\}$ and $\Pr\left\{\mathcal{S} \in B_{s+1}^2(\lfloor \tau N\rfloor, X)\,\middle|\,|\mathcal{S}| = s\right\}$ by $P_s^1(\tau,Q,N)$ and $P_{s+1}^2(\tau,Q,N)$ correspondingly. It is obvious, that these probabilities depend only on s, τ, Q, N and do not depend on R. The formulas (17) yield that the expectation (15) satisfies the inequalities:

$$\begin{aligned}\max\left\{P_s^1(\tau,Q,N), P_{s+1}^2(\tau,Q,N)\right\} &\le \mathcal{E}_s^N(\tau,Q,R)\\ &\le P_s^1(\tau,Q,N) + P_{s+1}^2(\tau,Q,N).\end{aligned} \tag{18}$$

Given the code X, for a fixed subset $\mathcal{S}\subset[t]$, $|\mathcal{S}|=k$, of size k and a fixed integer w, consider a probability

$$P_k^N(Q,w) \triangleq \Pr\left\{\left|\bigvee_{j\in\mathcal{S}} \boldsymbol{x}(j)\right| = w\right\}.$$

Note that the probability $P_k^N(Q,w)$ does not depend on the choice of the set $\mathcal{S}$ and depends only on k, w, N and Q. To compute the logarithmic asymptotics of the probabilities in (18), we represent them in the following forms:

$$\begin{aligned}P_s^1(\tau,Q,N) &= \sum_{w=\lfloor \max\{\tau,Q\}N\rfloor+1}^{\min\{N, s\lfloor QN\rfloor\}} P_s^N(Q,w),\\ P_{s+1}^2(\tau,Q,N) &= \sum_{w=\lfloor QN\rfloor}^{\min\{\lfloor \tau N\rfloor, (s+1)\lfloor QN\rfloor\}} P_{s+1}^N(Q,w).\end{aligned} \tag{19}$$

The logarithmic asymptotics of the probability $P_k^N(Q,w)$ was calculated in [4], it equals

$$\lim_{N\to\infty} \frac{-\log_2 P_k^N(Q,\lfloor qN\rfloor)}{N} = \mathcal{A}(k,Q,q), \tag{20}$$

where the function $\mathcal{A}(k,Q,q)$ is defined by (6)-(7). Note that $P_s^1(\tau,Q,N)=0$ if $\tau > sQ$ and $P_{s+1}^2(\tau,Q,N)=0$ if $\tau<Q$. This remark, (19) and (20) yield

$$\begin{aligned}\lim_{N\to\infty} \frac{-\log_2 P_s^1(\tau,Q,N)}{N} &= \min_{q\in(\text{i1})} \mathcal{A}'(s,Q,q),\\ \lim_{N\to\infty} \frac{-\log_2 P_{s+1}^2(\tau,Q,N)}{N} &= \min_{q\in(\text{i2})} \mathcal{A}'(s+1,Q,q),\end{aligned} \tag{21}$$
$$(\text{i1}) \triangleq [\max\{\tau,Q\},1], \quad (\text{i2}) \triangleq [0,\min\{\tau,(s+1)Q\}],$$

where the function $\mathcal{A}'(k,Q,q)$ is defined by (9).

Therefore, (18) and (21) yield existence of limit

$$\begin{aligned}&\lim_{N\to\infty} \frac{-\log_2 \mathcal{E}_s^N(\tau,Q,R)}{N} =\\ &= \min\left\{\min_{q\in(\text{i1})} \mathcal{A}'(s,Q,q), \min_{q\in(\text{i2})} \mathcal{A}'(s+1,Q,q)\right\}.\end{aligned} \tag{22}$$

Let us recall some analytical properties of the function $\mathcal{A}(k,Q,q)$.

Lemma 1. [4]. *Function $\mathcal{A}(k,Q,q)$ as a function of the parameter q decreases in the interval $q\in[Q, 1-(1-Q)^k]$, increases in the interval $q\in[1-(1-Q)^k, \min\{1,kQ\}]$ and equals 0 at the point $q=1-(1-Q)^k$.*

Hence,

$$\begin{aligned}&\min_{q\in(\text{i1})} \mathcal{A}'(s,Q,q) = 0 && \text{if } \tau \le 1-(1-Q)^s,\\ &\min_{q\in(\text{i2})} \mathcal{A}'(s+1,Q,q) = 0 && \text{if } \tau \ge 1-(1-Q)^{s+1}.\end{aligned}$$

It establishes the equivalence of the bound (16)-(22) and the bound (8).

Proof of Statement 2. The full proof of Statement 2 is presented in [5]. Here we give only a sketch of the proof.

Our aim is to offer the lower bound for the asymptotic behaviour of the expression

$$\begin{aligned}\underline{E}_{\text{Thr}}(s) \triangleq &\max_{0<\tau<1} \max_{1-(1-\tau)^{1/(s+1)}<Q<1-(1-\tau)^{1/s}}\\ &\min\left\{\mathcal{A}'(s,Q,\tau), \mathcal{A}(s+1,Q,\tau)\right\}, \quad s\to\infty.\end{aligned} \tag{23}$$

For any fixed τ, $0<\tau<1$, and any fixed Q, $1-(1-\tau)^{1/(s+1)} < Q < 1-(1-\tau)^{1/s}$, let us denote the solutions of the equation (7) for $\mathcal{A}(s,Q,\tau)$ and $\mathcal{A}(s+1,Q,\tau)$ by $y_1(Q,\tau)$ and $y_2(Q,\tau)$ correspondingly. Note that y_1 can be greater than 1. It follows from (7) that the parameter τ can be expressed in two forms:

$$\tau = Q\frac{1-y_1^s}{1-y_1} = Q\frac{1-y_2^{s+1}}{1-y_2}.$$

That is why the inequality $1-(1-\tau)^{1/(s+1)} < Q \Leftrightarrow \tau < 1-(1-Q)^{s+1}$ is equivalent to

$$\frac{1-y_2^{s+1}}{1-y_2} < \frac{1-(1-Q)^{s+1}}{1-(1-Q)}.$$

Note that, for any integer $n\ge 2$, the function $f(x)=\frac{1-x^n}{1-x}$ increases in the interval $x\in(0,+\infty)$. Hence, we have

$$1-(1-\tau)^{1/(s+1)} < Q \quad\Leftrightarrow\quad Q < 1-y_2,$$
and similarly,
$$Q < 1-(1-\tau)^{1/s} \quad\Leftrightarrow\quad Q > 1-y_1.$$

In conclusion, the pair of parameters (y_1,Q), $y_1>0$, $0<Q<1$, uniquely defines the parameters τ and y_2. Moreover, if the inequalities

$$0<\tau<1, \quad Q<1-y_2, \quad Q>1-y_1. \tag{24}$$

hold, then the parameters τ and Q are in the region, in which the maximum (23) is searched.

Let some constant $c>0$ be fixed, $s\to\infty$ and $y_1 \triangleq 1-c/s^2+o(1/s^3)$. Then, the asymptotic behavior of τ/Q equals

$$\frac{1-y_2^{s+1}}{1-y_2}=\frac{\tau}{Q}=\frac{1-y_1^s}{1-y_1}=s-\frac{c}{2}+o(1),$$

and, therefore,

$$y_2=1-\frac{c+2}{(s+1)^2}+o\left(\frac{1}{s^3}\right)=1-\frac{c+2}{s^2}+\frac{2}{s^3}+o\left(\frac{1}{s^3}\right).$$

To satisfy the inequalities (24) the parameter Q should be in the interval

$$\left(1-y_1=\frac{c}{s^2}+o\left(\frac{1}{s^3}\right), 1-y_2=\frac{c+2}{s^2}-\frac{2}{s^3}+o\left(\frac{1}{s^3}\right)\right).$$

Let us define the parameter Q as $Q\triangleq d/s^2$, where d, $c<d<c+2$, is some constant. Hence, Q is in the previous interval.

The full list of the asymptotic behaviors of the parameters is presented below:

$$\begin{aligned} y_1&=1-\frac{c}{s^2}+o\left(\frac{1}{s^2}\right), \qquad \tau=\frac{d}{s}-\frac{cd}{2s^2}+o\left(\frac{1}{s^2}\right),\\ y_2&=1-\frac{c+2}{s^2}+o\left(\frac{1}{s^2}\right), \quad Q=\frac{d}{s^2}, \quad s\to\infty, \end{aligned} \tag{25}$$

where c and d are arbitrary constants such that $c>0$, $c<d<c+2$. The parameters defined by (25) satisfy the inequalities (24), and, therefore, the substitution of asymptotic behaviors (25) into (23) leads to some lower bound on $\underline{E}_{\rm Thr}(s)$.

We omit the calculation of the asymptotic behavior of (23). The lower bound (10) is attained at $c\to\infty$ and $d=c+1$. If $s\to\infty$, then τ and Q are related by $\tau\sim s\cdot Q$. $\square$

IV. Simulation for Finite Code Parameters

For finite N and t, we carried out a simulation as follows. The probability distribution vector $\boldsymbol{p}$ (1) is defined by

$$p_s=p_{s+1}=1/2, \quad p_j=0, j\notin\{s,s+1\},$$

i.e. it is the distribution at which the maximum in the right-hand side of (4) is attained. Recall that it was proved in the beginning of the proof of Theorem 3. A code X is generated randomly from the ensemble of constant-weight codes, i.e. for some weight parameter w, every codeword of X is chosen independently and equiprobably from the set of all $\binom{t}{w}$ codewords. For every weight w and every decision rule, we repeat generation 1000 times and choose the code with minimal error probability. Note that for disjunctive decision rule $\Pr\{\text{accept } H_0|H_1\}=0$. The results of simulation are presented in Table II, where, for brevity, the probabilities $\Pr\{\text{accept } H_0|H_1\}$ and $\Pr\{\text{accept } H_1|H_0\}$ are denoted by $\Pr_{0|1}$ and $\Pr_{1|0}$ correspondingly. The best values of the maximal error probability (3) calculated using the formulas (11) and (14) for fixed parameters s, t and N are given in boldface.

If $s=2$, then for any code length N from Table II one can recommend to choose the corresponding code weight w, $1<w<N$, from Table II and generate an "optimal" random constant weight binary code of weight w, length N and arbitrary size t, $t>N$. In this case, for the corresponding threshold T, $w<T<N$, from Table II, an "optimal" error probability of the T-weight decision rule should be similar to the corresponding maximal error probability (3) indicated in Table II in boldface. As an example of such comparison, we put in Table II error probabilities (3) for $s=2$ and $t=100$ which were estimated by the Monte Carlo method, namely, subsets $\mathcal{S}$, $\mathcal{S}\subset[100]$, of size $|\mathcal{S}|=2$ and $|\mathcal{S}|=3$ were chosen randomly 1000 times.

TABLE II
Results of simulation

	T-weight decision rule				Disjunctive decision rule	
N	$\Pr_{1\|0}$	$\Pr_{0\|1}$	w	T	$\Pr_{1\|0}$	w
		$s=2$, $t=15$				
5	0.2571	**0.2571**	2	3	0.9333	2
8	**0.1619**	0.1604	3	5	0.7048	2
10	0	**0.1429**	1	2	0.4571	3
12	0	**0.0857**	1	2	0.1810	3
14	0	**0.0571**	1	2	0.0952	3
15	0	0.0462	2	4	**0.0286**	3
		$s=2$, $t=20$				
5	**0.2632**	0.2588	2	3	0.9579	2
8	0.1632	**0.1649**	3	5	0.8316	2
11	0.1053	**0.1509**	4	7	0.5158	3
12	**0.1158**	0.1123	4	7	0.4158	3
14	0	**0.0842**	2	4	0.2316	3
15	0	**0.0693**	2	4	0.1526	4
		$s=2$, $t=100$ (Estimated error probabilities)				
5	**0.2420**	0.2300	2	3	0.9980	2
8	0.1830	**0.1950**	3	5	0.9940	5
11	0.1570	**0.1630**	5	8	0.9830	4
12	0.1280	**0.1350**	4	7	0.9810	4
14	0	**0.1080**	2	4	0.9600	5
15	0	**0.0970**	2	4	0.9610	5

Acknowledgment

The research is supported in part by the Russian Foundation for Basic Research under Grant No. 16-01-00440 a.

References

[1] *Berger T.*, *Levenshtein V.I.*, Asymptotic Efficiency of Two-Stage Disjunctive Testing, *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1741-1749, 2002.

[2] *Damaschke P.*, *Muhammad A.S.*, Competitive group testing and learning hidden vertex covers with minimum adaptivity, *Discrete Math. Algorithm. Appl.*, vol. 2, no. 3, pp. 291-311, 2010.

[3] *De Bonis A.*, Constraining the number of positive responses in adaptive, non-adaptive, and two-stage group testing, *J. of Combinatorial Optimization*, vol. 32, no. 4, pp. 1254-1287, 2016.

[4] *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Almost Disjunctive List-Decoding Codes, *Problems of Information Transmission*, vol. 51, no. 2, pp. 110-131, 2015.

[5] *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Hypothesis Test for Upper Bound on the Size of Random Defective Set, Preprint, 2017. http://arxiv.org/pdf/1701.06201.

[6] *D'yachkov A.G.*, *Rykov V.V.*, *Rashad A.M.*, Superimposed Distance Codes, *Problems of Control and Inform. Theory*, vol. 18, no 4, pp. 237-250, 1989.

[7] *Falahatgar M.*, *Jafarpour A.*, *Orlitsky A.*, *Pichapati V.*, *Suresh A.T.*, Estimating the Number of Defectives with Group Testing, *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, pp. 1376-1380, 2016.

[8] *Zubashich V.F.*, *Lysyansky A.V.*, *Malyutov M.B.*, Block-randomized distributed trouble-shooting construction in large circuits with redundancy. *Izvestia of the USSR Acad. of Sci., Technical Cybernetics*, vol. 6, 1976.