Symmetric Disjunctive List-Decoding Codes

A.G. D'yachkov, I.V. Vorobyev, N.A. Polyanskii, V.Yu. Shchukin
Lomonosov Moscow State University, Moscow, Russia,
Email: agd-msu@yandex.ru, vorobyev.i.v@yandex.ru, nikitapolyansky@gmail.com, vpike@mail.ru

***Abstract*—In this paper, we consider *symmetric disjunctive list-decoding* (SLD) codes, which are a class of binary codes based on a *symmetric disjunctive sum* (SDS) of binary symbols. By definition, the SDS takes values from the ternary alphabet $\{0, 1, *\}$, where the symbol $*$ denotes "erasure". Namely: SDS is equal to 0 (1) if all its binary symbols are equal to 0 (1), otherwise SDS is equal to $*$. The main purpose of this work is to obtain bounds on the rate of these codes.**

Keywords:. Symmetric disjunctive codes, random coding bounds, nonadaptive symmetric group testing.

I. Statement of Problem and Results

A. Notations and Definitions

Let q, N, t, s, and L be integers, where $q \geq 2$, $2 \leq s < t$, $1 \leq L \leq t - s$. Let $\triangleq$ denote 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. The standard symbol $\lfloor a \rfloor$ will be used to denote the largest integer $\leq a$.

A binary $(N \times t)$-matrix

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

$i \in [N]$, $j \in [t]$, with N rows $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ and t columns $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(t)$ (codewords) 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 the codeword $x(j)$, i.e., $|\boldsymbol{x}(j)| \triangleq \sum\limits_{i=1}^{N} x_i(j)$, is called a *weight* of $x(j)$, $j \in [t]$. A code X is called a *constant weight binary code of weight w*, $1 \leq w < N$, if for any $j \in [t]$, the weight $|\boldsymbol{x}(j)| = w$.

Let $\boldsymbol{u} \bigvee \boldsymbol{v}$ denote the disjunctive sum of binary columns $\boldsymbol{u}, \boldsymbol{v} \in \{0, 1\}^N$. If $\boldsymbol{x}, \boldsymbol{y} \in \{0, 1, *\}^N$ are arbitrary *ternary* columns with components from the alphabet $\{0, 1, *\}$, then the ternary column $\boldsymbol{z} = (z_1, z_2, \ldots, z_N) \in \{0, 1, *\}^N$,

$$z_i \triangleq \begin{cases} 0, & \text{if} \quad x_i = y_i = 0, \\ 1, & \text{if} \quad x_i = y_i = 1, \\ *, & \text{otherwise}, \end{cases}$$

is called a *symmetric disjunctive sum* [1] of $\boldsymbol{x}$ and $\boldsymbol{y}$. This operation will be denoted by $\bigtriangledown$, that is $\boldsymbol{z} = \boldsymbol{x} \bigtriangledown \boldsymbol{y}$. We say that a binary column $\boldsymbol{u}$ *covers* a column $\boldsymbol{v}$ ($\boldsymbol{u} \succeq \boldsymbol{v}$) if $\boldsymbol{u} \bigvee \boldsymbol{v} = \boldsymbol{u}$, and a ternary column $\boldsymbol{u}$ *symmetrically covers* a column $\boldsymbol{v}$ ($\boldsymbol{u} \trianglerighteq \boldsymbol{v}$) if $\boldsymbol{u} \bigtriangledown \boldsymbol{v} = \boldsymbol{u}$.

B. Symmetric Disjunctive List-Decoding Codes (SLD s_L-codes)

Definition 1. [2], [3]. A binary code X is said to be a *disjunctive list-decoding code of strength s with list size L* (LD s_L-code) if the disjunctive sum of any s codewords of X covers not more than $L - 1$ other codewords of X that are not components of the given sum. In other words, for any two disjoint sets $\mathcal{S}, \mathcal{L} \subset [t], |\mathcal{S}| = s, |\mathcal{L}| = L, \mathcal{S} \cap \mathcal{L} = \varnothing$, there exist a row $\boldsymbol{x}_i, i \in [N]$, and a column $\boldsymbol{x}(j), j \in \mathcal{L}$, such that

$$x_i(k) = 0 \quad \forall k \in \mathcal{S} \qquad \text{and} \qquad x_i(j) = 1.$$

Denote by $t_{ld}(N, s, L)$ the maximal size of LD s_L-codes of length N and by $N_{ld}(t, s, L)$ the minimal length of LD s_L-codes of size t. Define the *rate* of LD s_L-codes:

$$R_L(s) \triangleq \varlimsup_{N \to \infty} \frac{\log_2 t_{ld}(N, s, L)}{N} = \varlimsup_{t \to \infty} \frac{\log_2 t}{N_{ld}(t, s, L)}. \quad (1)$$

Definition 2. [4], [5], [6]. A binary code X is said to be a *symmetric disjunctive list-decoding code of strength s with list size L* (SLD s_L-code) if the symmetric disjunctive sum of any s codewords of X symmetrically covers not more than $L - 1$ other codewords of X that are not components of the given sum. In other words, for any two disjoint sets $\mathcal{S}, \mathcal{L} \subset [t]$, $|\mathcal{S}| = s$, $|\mathcal{L}| = L$, $\mathcal{S} \cap \mathcal{L} = \varnothing$, there exist a row $\boldsymbol{x}_i, i \in [N]$, and a column $\boldsymbol{x}(j), j \in \mathcal{L}$, such that

$$x_i(k) = 0 \quad \forall k \in \mathcal{S} \qquad \text{and} \qquad x_i(j) = 1, \quad \text{or}$$
$$x_i(k) = 1 \quad \forall k \in \mathcal{S} \qquad \text{and} \qquad x_i(j) = 0.$$

Denote by $t_{sld}(N, s, L)$ the maximal size of SLD s_L-codes of length N and by $N_{sld}(t, s, L)$ the minimal length of SLD s_L-codes of size t. Define the *rate* of SLD s_L-codes:

$$R^*_L(s) \triangleq \varlimsup_{N \to \infty} \frac{\log_2 t_{sld}(N, s, L)}{N} = \varlimsup_{t \to \infty} \frac{\log_2 t}{N_{sld}(t, s, L)}. \quad (2)$$

Theorem 1. (Monotonicity properties). *The rate of SLD s_L-codes satisfies the following inequalities*

$$R^*_L(s + 1) \leq R^*_L(s) \leq R^*_{L+1}(s). \quad (3)$$

Proof of Theorem 1. It immediately follows from Definition 2 that every SLD $(s + 1)_L$-code is the corresponding SLD s_L-code, so the left inequality in (3) takes place. Simultaneously, every SLD s_L-code is SLD s_{L+1}-code, therefore the right inequality in (3) is true. □

C. Applications of Symmetric Disjunctive Codes

Applications of SLD s_L-codes relate to the *non-adaptive symmetric group testing* which is based on the symmetric disjunctive sum of binary symbols. Group testing deals with identification of defective units in a given pool. We use symmetric group tests, i.e., take a subset of the pool and check it. The outcome of a symmetric group test belongs to the ternary alphabet. It is equal to 0, 1 or $*$, if all tested units are not defective, all units are defective or at least one unit is defective and at least another one is not defective, respectively. The symmetric group testing was motivated by applications [1] in electrical devices testing and chemical analysis.

Suppose the size of the pool equals t and the number of defective units does not exceed s. As is the case with LD s_L-codes [7], SLD s_L-codes can be considered in connection with the problem of constructing *two-stage non-adaptive symmetric group testing procedures*. In the first stage, one does N tests that can be depicted as an binary $(N \times t)$-matrix $X = \|x_i(j)\|$, where a column $\boldsymbol{x}(j)$ corresponds to the j-th unit, a row $\boldsymbol{x}_i$ corresponds to the i-th test and $x_i(j) \triangleq 1$ if and only if the j-th unit is included into the i-th testing group. Then the ternary column y of the test results equals the symmetric disjunctive sum of the columns which correspond to the defective units. Let X be SLD s_L-code, after decoding of the result column y, i.e. search of codewords which are symmetrically covered by y, a set of $\leq s+L-1$ elements is selected. These units are separately tested in the second stage. Note that for $s \geq 2$ the rate $R_L^*(s)$ of SLD s_L-codes is a monotonically nondecreasing function of $L \geq 1$, and its limit

$$R_\infty^*(s) = \lim_{L \to \infty} R_L^*(s)$$

can be interpreted as the *maximum rate* of two-stage non-adaptive symmetric group testing procedures in a search for $\leq s$ defects with the use of SLD s_L-codes.

In papers [4], [5], we suggested another application of SLD codes called *reference communication system*. Let a system contain M *terminal stations* $\mathrm{S}_1, \mathrm{S}_2, ..., \mathrm{S}_M$ and let a *multiple-access channel* (MAC) connect these M stations to a *central station* (CS). Each terminal station has a *source*. In every time interval, the source can produce a binary *information packet* of length K. Introduce $t \triangleq 2^K$ and enumerate all 2^K possible information packets by integers from 1 to t. The packets are encoded into binary sequences of length N by a code $X = (\boldsymbol{x}(i), i \in [t])$, where the codeword $\boldsymbol{x}(i), i \in [t]$, is the encoded sequence corresponding to the information packet number i. Denote by $\mathcal{S}$ the set of numbers of generated packets and suppose $|\mathcal{S}| \leq s$.

The CS is interested only in the contents of the received packet and not in the senders. Using a *feedback broadcast channel* (FBC) the CS answers all M stations to all requests. The model of MAC corresponds to the *frequency modulation*, i.e., the output ternary sequence $\boldsymbol{y}$ is the symmetric disjunctive sum of the inputs. The scheme of reference communication system is represented on Figure 1.

Let the terminal stations use an SLD s_L-code X. Since the number of information packets produced by the terminal stations in the same time interval is not more than s, the CS is able to recover at most $s+L-1$ packets, which contain s transmitted packets.

Note that the model of MAC can also correspond to the *impulse modulation*, i.e., the output binary sequence is the disjunctive sum of the inputs. In this case, it is convenient to use LD s_L-codes for encoding and decoding information packets. The case of impulse modulation was considered in [5].

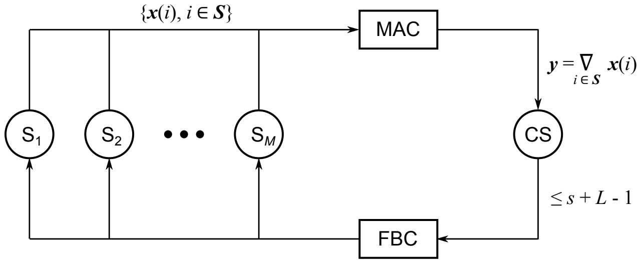

Fig. 1. Reference communication system

Another application of SLD s_1-codes concerns with *undetermined data* [8], [9]. Given an alphabet $A = \{a_1, a_2, ..., a_t\}$ of *basic symbols*, to every nonempty subset $T \subseteq [t]$, assign a symbol a_T, which is called *undetermined*. Its *specification* is any basic symbol $a_i, i \in T$. By a *specification* of a sequence of undetermined symbols we mean the result of replacing all its symbols by some of its specifications. The symbol $a_{[t]}$ that can be specified by any basic symbol is called *indefinite* and is denoted by $*$. Let $\mathcal{T}$ be a system of subsets $T \subseteq [t]$ and let $A^* = A^*_{\mathcal{T}} = \{a_T |\ T \in \mathcal{T}\}$ be an *undetermined alphabet* associated with the system.

Consider a problem of coding of undetermined sequences such that the original undetermined sequence can be completely reconstructed from the encoded sequence. One coding method refers to a *binary representation* [8], [9] of undetermined alphabet, which is defined as a pair (X, X^*) of $(N \times t)$-matrix X with columns $\boldsymbol{x}(i) \in \{0,1\}^N, i \in [t]$, and $(N \times |\mathcal{T}|)$-matrix X^* with columns $\boldsymbol{x}(T) \in \{0,1,*\}^N, T \in \mathcal{T}$, where $\boldsymbol{x}(i)$ specifies $\boldsymbol{x}(T)$ in undetermined alphabet $\{0,1,*\}$ if and only if $i \in T$. Advantages of such method are linear in t complexity of the symbol reconstruction and the fact that the mentioned condition allows to know only a small matrix X for reconstruction of the original undetermined sequence while the matrix X^* may contain up to 2^t columns. Obviously, an SLD s_1-code $X = (\boldsymbol{x}(i), i \in [t])$ and the matrix $X^* = (\bigtriangledown_{i \in T} \boldsymbol{x}(i), T \in \mathcal{T})$ give the fairly compact binary representation of undetermined alphabet associated with the system $\mathcal{T} = [t] \cup \{T \subset [t] |\ |T| \leq s\}$ [9].

D. Relations Between Parameters of LD s_L-codes and SLD s_L-Codes

The following evident propositions from [4], [5], [6] associate the rate of LD s_L-codes (1) with the rate of SLD s_L-codes (2).

Proposition 1. [4], [5], [6]. *Any LD s_L-code is the corresponding SLD s_L-code.*

Proposition 2. [4], [5], [6]. *Let $X = \|x_i(j)\|$ be an SLD s_L-code of length N and size t. Consider $(N \times t)$-matrix $X' = \|x'_i(j)\|$ with elements*

$$x'_i(j) \triangleq \begin{cases} 1, & \textit{if } \ x_i(j) = 0, \\ 0, & \textit{if } \ x_i(j) = 1. \end{cases}$$

Then the code of length $2N$ and size t composed of all rows of the codes X and X' is an LD s_L-code.

Corollary 1. [4], [5], [6]. *The rates of LD s_L-codes and SLD s_L-codes satisfy inequalities:*

$$R_L(s) \le R_L^*(s) \le 2R_L(s). \tag{4}$$

The next obvious proposition allows us to get another upper bound on the rate of SLD s_L-codes.

Proposition 3. *Let X be an LD s_L-code of length N and size t with a codeword $\boldsymbol{x}(j_0)$ of weight w. Then the code X'' of length $N-w$ and size $t-1$ constructed from the code X by removing the codeword $\boldsymbol{x}(j_0)$ and all rows x_i, for which $x_i(j_0)=1$, is an LD $(s-1)_L$-code.*

Corollary 2. *The rate of SLD s_L-codes has the following upper bound:*

$$R_L^*(s) \le R_L(s-1). \tag{5}$$

Proof of Corollary 2. Let X be an arbitrary SLD s_L-code of length N and size t. The code X_1 obtained in Proposition 2 from the code X is a constant weight LD s_L-code of length $2N$, size t and weight N. Then the code X_2 obtained in Proposition 3 from the code X_1 is an LD $(s-1)_L$-code of length N and size $t-1$. Hence as $N \to \infty$ the inequality

$$\frac{\log_2[t-1]}{N} \le R_L(s-1)(1+o(1))$$

holds. It means correctness of (5). □

The best presently known lower and upper bounds on the rate $R_L(s)$ were recently obtained in [10], [11]. The use of the inequalities (4) and (5), the lower bound $\underline{R}_L(s)$ [10] and the upper bound $\overline{R}_L(s)$ [10] on the rate of LD s_L-codes yields the results below.

Theorem 2. (Relationship between $R_L^*(s)$ and $R_L(s)$) *The following three statements hold.*

1. *For any fixed $s \ge 2$ and $L \ge 1$ the rates $R_L^*(s)$ and $R_L(s)$ have relationship*

$$R_L(s) \le R_L^*(s) \le \min\{\ 2R_L(s),\ R_L(s-1)\ \}.$$

2. *For any fixed $L \ge 1$ and $s \to \infty$*

$$R_L^*(s) = R_L(s)(1+o(1)).$$

3. *For any fixed $s \ge 2$ and $L \ge 1$ the rate of an SLD s_L-code satisfies the inequality*

$$\underline{R}_L(s) \le R_L^*(s) \le \overline{R}_L^*(s) \triangleq \min\{\ 2\overline{R}_L(s),\ \overline{R}_L(s-1)\ \}.$$

E. q-ary Frameproof List-Decoding Codes

Definition 2'. A q-ary code X is said to be a *q-ary frameproof list-decoding s_L-code* (q-ary FLD s_L-code) if, for any subset $\mathcal{S}$, $|\mathcal{S}|=s$, the q-ary coordinate set $\{x_i(j), j \in S\} \in [q]^N$ contains not more than $L-1$ other codewords of X that are not components of the subset $\{\boldsymbol{x}(j), j \in S\}$.

Denote by $t_{fld}^{(q)}(N,s,L)$ the maximal size of q-ary FLD s_L-codes of length N and by $N_{fld}^{(q)}(t,s,L)$ the minimal length of q-ary FLD s_L-codes of size t. Define the *rate* of q-ary FLD s_L-codes:

$$R_L^{(q)}(s) \triangleq \overline{\lim_{N\to\infty}} \frac{\log_q t_{fld}^{(q)}(N,s,L)}{N} = \overline{\lim_{t\to\infty}} \frac{\log_q t}{N_{fld}^{(q)}(t,s,L)}.$$

Remark 1. A q-ary FLD s_1-code is the special case of separating codes [12]. More specifically, for $L=1$, Definition 2' is equivalent to the definition of $(s,1)$-*separating* code. Some results and applications of $(s,1)$-separating codes are presented in the survey [13].

Remark 2. For $L=1$, the definition of q-ary FLD s_1-code is equivalent to the definition of *s-frameproof* code [14].

The q-ary extensions of Corollary 1 and Corollary 2 are given by

Corollary 1'. *The rates of LD s_L-codes and q-ary FLD s_L-codes satisfy inequalities:*

$$\frac{1}{\log_2 q} R_L(s) \le R_L^{(q)}(s) \le \frac{q}{\log_2 q} R_L(s). \tag{6}$$

Corollary 2'. *The rate of q-ary FLD s_L-codes has the following upper bound:*

$$R_L^{(q)}(s) \le \frac{q-1}{\log_2 q} R_L(s-1). \tag{7}$$

Proof of Corollary 1'. The left inequality in (6) is obvious, because every LD s_L-code is q-ary FLD s_L-code. To prove the right inequality in (6) consider an arbitrary q-ary FLD s_L-code. Introduce the binary $(Nq \times t)$ matrix X_1 obtained by the standard replacement of each q-ary symbol x, $x \in [q]$, in X by the $(0,1)$-binary column of length q and weight 1 containing the unique symbol 1 at the x-th position. One can easily check that the binary code X_1 is a binary LD s_L-code of length qN and size t. □

Proof of Corollary 2'. Let X be an arbitrary q-ary SLD s_L-code and X_1 be the binary LD s_L-code of length qN, size t and weight N that obtained from the code X in the proof of Corollary 1. Then the code X_2 obtained in Proposition 3 from the code X_1 is a binary LD $(s-1)_L$-code of length $(q-1)N$ and size $t-1$, so the upper bound (7) is correct. □

The next evident Corollary 3 is the consequence of the two previous corollaries and the bounds on the rate of LD s_L-codes [10].

Corollary 3. *Let q, L be fixed and $s \to \infty$. The following bounds on the rate of q-ary FLD s_L-codes hold:*

$$R_L^{(q)}(s) \ge \frac{4e^{-2}\log_q s}{s^2}(1+o(1)),$$
$$R_L^{(q)}(s) \le \frac{2L(q-1)\log_q s}{s^2}(1+o(1)), \quad s \to \infty.$$

F. Random Coding Bounds on the Rate of SLD s_L-codes

In the given paper, we develop a random coding method based on the ensemble of constant-weight codes and establish new lower random coding bounds on the rate of SLD s_L-codes. Some of the methods which are used in the proof of the next theorem are presented in [10], [11].

Theorem 3. (Lower random coding bound $\underline{R}_L^*(s)$). *The following three statements hold.*

1. *For any fixed $L \ge 1$ and $s \ge 2$ we have the inequality*

$$R_L^*(s) \ge \underline{R}_L^*(s) \triangleq \max_{0<Q\le 1/2} \left(h(Q) + \frac{B_L(s,Q)}{s+L-1} \right), \qquad (8)$$

where

$$\begin{aligned} h(Q) &\triangleq -Q\log_2 Q - (1-Q)\log_2[1-Q], \\ B_L(s,Q) &\triangleq Q\log_2\left[\frac{p(1-z)}{p(1-z)+q(1-z)}\right] + \\ &\quad + (1-Q)\log_2\left[\frac{p(z)}{p(z)+q(z)}\right], \qquad (9) \\ p(z) &\triangleq z^s(z-z^s)^L, \\ q(z) &\triangleq (z-z^s)(1-z^s-(1-z)^s)^L, \end{aligned}$$

and z is the unique root of the equation

$$Q(p(z)+q(z)) = (1-Q)(p(1-z)+q(1-z)). \qquad (10)$$

2. *For fixed $L = 1, 2, \dots$ and $s \to \infty$*

$$\underline{R}_L^*(s) \ge \frac{L}{s^2\log_2 e}(1+o(1)). \qquad (11)$$

3. *For fixed $s = 2, 3, \dots$ there exists a limit*

$$\underline{R}_\infty^*(s) \triangleq \lim_{L\to\infty} \underline{R}_L^*(s) = \log_2\left[\frac{(s-1)^{s-1}}{s^s} + 1\right]. \qquad (12)$$

If $s \to \infty$, then

$$\underline{R}_\infty^*(s) = \frac{\log_2 e}{es}(1+o(1)) = \frac{0.5307\dots}{s}(1+o(1)).$$

The numerical values of the lower bound (8)-(10) are shown in Table I, where the argument of maximum in (8) is denoted by $Q_L^*(s)$. Note that the lower bound (8)-(10) improves the random coding bound obtained in [15] using the ensemble with independent binary symbols of codewords. In addition one can see that for small values of $s \ge 2$ and $L \ge 1$, the lower bounds (8)-(10) are greater than the lower bounds $\underline{R}_L(s)$ on the rate of LD s_L-codes from [10].

Note that, for $s \to \infty$, the asymptotic lower bound of $\underline{R}_L^*(s)$ (11) coincides with the asymptotic behavior of the random coding bound on the rate of LD s_L-codes [10]. In addition, for $L \to \infty$, the asymptotics of $\underline{R}_L^*(s)$ (12) coincides with the asymptotic behavior of the mentioned above bound from [10].

II. Proof of Theorem 3

This Section contains five lemmas that are only stated. The proofs of Lemma 1-5 are presented in the preprint [17].

Proof of Statement 1. Fix $L \ge 1$, $s \ge 2$ and a parameter Q, $0 < Q \le 1/2$. The bound (8)-(10) is obtained by the method of random coding over the ensemble of binary constant-weight codes [16] defined as the ensemble $E(N,t,Q)$ of binary codes X of length N and size t, where the codewords are chosen independently and equiprobably from the set consisting of all $\binom{N}{\lfloor QN\rfloor}$ codewords of a fixed weight $\lfloor QN \rfloor$.

TABLE I
Numerical values of the lower bound $\underline{R}_L^*(s)$

s_L	2_1	2_2	2_3	2_4	2_5	2_6
$\underline{R}_L^*(s)$	0.2075	0.2457	0.2635	0.2744	0.2819	0.2874
$Q_L^*(s)$	0.5000	0.2764	0.2432	0.2297	0.2228	0.2180
s_L	3_1	3_2	3_3	3_4	3_5	3_6
$\underline{R}_L^*(s)$	0.0800	0.1153	0.1348	0.1470	0.1552	0.1611
$Q_L^*(s)$	0.2000	0.1794	0.1686	0.1613	0.1561	0.1524
s_L	4_1	4_2	4_3	4_4	4_5	4_6
$\underline{R}_L^*(s)$	0.0439	0.0684	0.0838	0.0941	0.1014	0.1068
$Q_L^*(s)$	0.1479	0.1391	0.1326	0.1275	0.1234	0.1201
s_L	5_1	5_2	5_3	5_4	5_5	5_6
$\underline{R}_L^*(s)$	0.0279	0.0456	0.0575	0.0660	0.0723	0.0771
$Q_L^*(s)$	0.1209	0.1150	0.1103	0.1064	0.1030	0.1003
s_L	6_1	6_2	6_3	6_4	6_5	6_6
$\underline{R}_L^*(s)$	0.0194	0.0325	0.0420	0.0490	0.0544	0.0587
$Q_L^*(s)$	0.1027	0.0983	0.0947	0.0915	0.0889	0.0865

A pair of sets $(\mathcal{S},\mathcal{L}), |\mathcal{S}| = s, |\mathcal{L}| = L, \mathcal{S}\cap\mathcal{L} = \varnothing$, we call an (s_L^*)-*bad* pair if

$$\bigvee_{i\in\mathcal{S}} \boldsymbol{x}(i) \trianglerighteq \bigvee_{j\in\mathcal{L}} \boldsymbol{x}(j).$$

For the ensemble $E(N,t,Q)$, denote by $P(N,Q,s,L)$ the probability of the event "the pair $(\mathcal{S},\mathcal{L})$ is (s_L^*)-bad". Note that the absence of (s_L^*)-bad pair of subsets in the code is the criterion of SLD s_L-code. Hence, similarly to the arguments in the proof of the lower random coding bound on the rate $R_L(s)$ (1) in [10], the rate $R_L^*(s)$ (2) satisfies the inequality

$$\begin{aligned} R_L^*(s) \ge \underline{R}_L^*(s) &\triangleq \frac{1}{s+L-1}\max_{0<Q<1} A_L^*(s,Q), \\ A_L^*(s,Q) &\triangleq \varliminf_{N\to\infty} \frac{-\log_2 P(N,Q,s,L)}{N}. \end{aligned} \qquad (13)$$

Note that the set of all s_L^*-bad pairs of any codeword weight is invariant under the binary negation operation, it implies the equality $P(N,Q,s,L) = P(N,1-Q,s,L)$. Therefore, it is enough to consider only $0 < Q \le 1/2$.

To complete the proof of the theorem, it is sufficient to compute the function $A_L^*(s,Q)$ (13).

Lemma 1. *If there exists a solution $z, 0 < z < 1$, of the equation (10), then the function $A_L^*(s,Q)$ (13) equals*

$$\begin{aligned} (s+L-1)h(Q) + (1-Q)\log_2\left[\frac{p(z)}{p(z)+q(z)}\right] + \\ + Q\log_2\left[\frac{p(1-z)}{p(1-z)+q(1-z)}\right], \end{aligned}$$

where the functions $h(\cdot), p(\cdot)$ and $q(\cdot)$ are determined by (9).

Lemma 2. *The function*

$$\rho(z) \triangleq \frac{p(z)+q(z)}{p(1-z)+q(1-z)}, \quad 0 < z < 1, \qquad (14)$$

continuously maps the interval $(0,1)$ into the interval $(0,+\infty)$ and strictly increases.

By Lemma 2 the equation (10) has the unique solution. Thus, the condition of Lemma 1 is clear, it means that the bound (8)-(10) is proved. □

Proof of Statement 2. For fixed $s \geq 2$ and $L \geq 1$, let us interpret equation (10) as a function $Q_L(s,z)$ of the argument $z, 0 < z < 1$, i.e.,

$$Q_L(s,z) \triangleq \frac{p(1-z)+q(1-z)}{p(1-z)+q(1-z)+p(z)+q(z)}, \tag{15}$$

where the functions $p(\cdot)$ and $q(\cdot)$ are determined in (9).

Due to existence and uniqueness of the root of the equation (10), continuity and monotonicity of the function (15) (by Lemma 2), one can rewrite the definition of the random coding bound (8)-(10) as

$$\underline{R}^*_L(s) \triangleq \max_{1/2 \leq z < 1} T_L(s,z), \tag{16}$$

where

$$T_L(s,z) \triangleq h(Q_L(s,z)) + B_L(s, Q_L(s,z)). \tag{17}$$

Let $L \geq 1$ be fixed and $s \to \infty$. If in definition (17) we put $z = 1 - \lambda/s$, where the parameter $\lambda = \lambda_L$ is independent of s, then (16) means that

$$\underline{R}^*_L(s) \geq T_L\left(s, 1 - \frac{\lambda}{s}\right). \tag{18}$$

Lemma 3. *For a fixed $L \geq 1$ and $s \to \infty$, the next asymptotic equality holds:*

$$T_L\left(s, 1-\frac{\lambda}{s}\right) = \frac{L}{s^2}\left(-\lambda \log_2[1-e^{-\lambda}]\right)(1+o(1)). \tag{19}$$

Taking derivative one can check that at $\lambda = \frac{1}{\log_2 e}$ the maximum

$$\max_{\lambda > 0}\left\{-\lambda \log_2[1-e^{-\lambda}]\right\} = \frac{1}{\log_2 e} \tag{20}$$

is attained. Therefore, (18) and (20) imply for the random coding bound (8)-(10) the asymptotic inequality (11). □

Proof of Statement 3. For fixed $s \geq 2$ and $L \geq 1$, let us introduce the following function

$$g(z) \triangleq g_L(s,z) = \frac{z - z^s}{1 - z - (1-z)^s}, \quad \frac{1}{2} \leq z < 1. \tag{21}$$

It is clear that $g(z)$ (21) monotonically increases in the interval $[1/2, 1)$, attains 1 at the point $z = \frac{1}{2}$ and has the left limit $s-1$ as $z \to 1$.

For large enough parameter L and a fixed parameter $c > 0$ independent of L, one can see that the root of equation

$$\left(\frac{g(z)}{1+g(z)}\right)^L = c(1-z), \quad \frac{1}{2} \leq z < 1, \tag{22}$$

exists and is unique, since the left-hand side of (22) monotonically increases and the right-hand side of (22) strictly decreases. Denote this root by $z_L(s,c)$.

Let $s \geq 2$ be fixed and $L \to \infty$.

Lemma 4. *The substitution of $z = z_L(s,c)$ into the function (17) yields*

$$T_L(s, z_L(s,c)) \cdot (1+o(1)) = \frac{1}{s+c}\log_2\left[\frac{(s-1)^{s-1}}{s^s}\right] - \\ - \frac{s+c-1}{s+c}\log_2[s+c-1] + \log_2[s+c], \quad L \to \infty. \tag{23}$$

The definition (16) means that

$$\underline{R}^*_L(s) \geq T_L(s, z_L(s,c))(1+o(1)), \quad L \to \infty, \\ \forall\, c = c(s) > 0. \tag{24}$$

Calculating the derivative in c, one can check that maximum of the right-hand side of (23) is attained at the point $c = c(s) = \frac{s^s-(s-1)^s}{(s-1)^{s-1}}$. If we substitute this value $c = c(s)$ into (23), then the use of (24) establishes for the random coding bound (8)-(10) the inequality

$$\underline{R}^*_L(s) \geq \log_2\left[\frac{(s-1)^{s-1}}{s^s} + 1\right](1+o(1)), \quad L \to \infty. \tag{25}$$

Lemma 5. *The asymptotic inequality (25) is an equality.*
Statement 3 of Theorem 3 is proved. □

References

[1] *Sobel M.*, *Kumar S.*, *Blumenthal S.*, Symmetric Binomial Group-Testing with Three Outcomes, *Purdue Symposium on Statistical Decision Theory and Related Topics*, 1971.

[2] *D'yachkov A.G.*, *Rykov V.V.*, A Survey of Superimposed Code Theory, *Problems of Control and Inform. Theory*, vol. 12, no. 4, pp. 229-242, 1983.

[3] *Kautz W.H.*, *Singleton R.C.*, Nonrandom Binary Superimposed Codes, *IEEE Trans. Inform. Theory*, vol. 10, no. 4, pp. 363-377, 1964.

[4] *D'yachkov A.G.*, *Rykov V.V.*, An Application of Codes for the Multiple Access Channel in the ALOHA Communication System, *Proccedings of the 6-th All-Union Seminar in Computing Networks, Moscow-Vinnitsa*, vol. 4, pp. 18-24, 1981 (in Russian).

[5] *D'yachkov A.G.*, *Rykov V.V.*, Superimposed Codes for Multiple Accessing of the OR-channel, *1998 IEEE International Symposium on Information Theory, Boston, USA*, Aug. 1998.

[6] *D'yachkov A.G.* Lectures on Designing Screening Experiments, *Lecture Note Series 10*, Combinatorial and Computational Mathematics Center, Pohang University of Science and Technology (POSTECH), Korea Republic, Feb. 2003 (survey, 112 pages).

[7] *Vilenkin P.A.*, On Constructions of List-Decoding Superimposed Codes, *Proc. 6th Int. Workshop on Algebraic and Combinatorial Coding Theory (ACCT-6), Pskov, Russia*, pp. 228-231, 1998.

[8] *Sholomov L.A.*, Binary Representation of Underdetermined Data, *Doklady Akademii Nauk*, vol. 448, no. 3, pp. 275-278, 2013.

[9] *Sholomov L.A.*, Binary Representations of Underdetermined Data and Superimposed Codes, *Prikl. Diskr. Mat.*, no. 1, pp. 17-33, 2013 (in Russian).

[10] *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Bounds on the Rate of Disjunctive Codes, *Problems of Information Transmission*, vol. 50, no. 1, pp. 27-56, 2014.

[11] *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Bounds on the Rate of Superimposed Codes, *2014 IEEE International Symposium on Information Theory*, pp. 2341-2345, Honolulu, HI USA, Jun.29-Jul.4, 2014.

[12] *Friedman A.D.*, *Graham R.L.*, *Ullman J.D.*, Universal single transition time asynchronous state assignments, *IEEE Trans. Comput.* , vol. 18, no. 6, pp. 541-547, 1969.

[13] *Cohen G.D.*, *Schaathun H.G.*, Asymptotic overview on separating codes, *Tech. Report 248*, Department of Informatics, University of Bergen, Bergen, Norway, 2003.

[14] *Boneh D.*, *Shaw J.*, Collusion-secure fingerprinting for digital data, *IEEE Trans. Inform. Theory*, vol. 44, no. 5, pp. 1897-1905, 1998.

[15] *Ahmed M. Rashad*, On Symmetrical Superimposed Codes, *J. Inf. Process. Cybern EIK 29*, vol. 7, pp. 337-341, 1989.

[16] *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.

[17] *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Symmetric Disjunctive List-Decoding Codes, *arXiv*:1410.8385 [cs.IT], 2014.