Des. Codes Cryptogr. (2017) 82:211–229
DOI 10.1007/s10623-016-0278-4

Symmetric disjunctive list-decoding codes

A. G. D'yachkov[1] · I. V. Vorobyev[1] · N. A. Polyanskii[1] · V. Yu. Shchukin[1]

Received: 20 September 2015 / Revised: 15 August 2016 / Accepted: 30 August 2016 /
Published online: 16 September 2016

Abstract In this work, 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 · Separating codes · Frameproof codes · Random coding bounds · Nonadaptive symmetric group testing · Underdetermined data

Mathematics Subject Classification 94B25 · 94B65

The material in this work was presented in part at the 2015 IEEE International Symposium on Information Theory [6]. This paper provides the proofs of all lemmas, which were formulated in [6], (refer to Sect. 2) and contains some improvements of the results, which were developed in [6] (refer to Sects. 1.3, 1.5 and 1.6).

This is one of several papers published in *Designs, Codes and Cryptography* comprising the "Special Issue on Coding and Cryptography".

✉ A. G. D'yachkov
agd-msu@yandex.ru

✉ V. Yu. Shchukin
vpike@mail.ru

[1] Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Moscow 119992, Russian Federation

1 Statement of problem and results

1.1 Notation and definitions

Let N, t, s, and L be integers, where $2 \le s < t$, $1 \le L \le t - s$. 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 $\lfloor a \rfloor$ ($\lceil a \rceil$) will be used to denote the largest (smallest) integer $\le a$ ($\ge a$).

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

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

$i \in [N]$, $j \in [t]$, 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 $\boldsymbol{x}(j)$, i.e., $|\boldsymbol{x}(j)| \triangleq \sum_{i=1}^{N} x_i(j)$, is called a *weight* of $\boldsymbol{x}(j)$, $j \in [t]$. A code X is called a *constant weight binary code of weight* w, $1 \le w < N$, 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* [21] (SDS) of $\boldsymbol{x}$ and $\boldsymbol{y}$. This operation will be denoted by ∇, that is $\boldsymbol{z} = \boldsymbol{x} \nabla \boldsymbol{y}$. 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 $\boldsymbol{u} \nabla \boldsymbol{v} = \boldsymbol{u}$. In other words, $\boldsymbol{u} = (u_1, u_2, \ldots, u_N)$ covers $\boldsymbol{v} = (v_1, v_2, \ldots, v_N)$ iff $u_i \ge v_i$ for each position $i \in [N]$, and $\boldsymbol{u}$ symmetrically covers $\boldsymbol{v}$ iff for each position $i \in [N]$ where u_i is binary we have that v_i is also binary and satisfies $u_i = v_i$.

1.2 Symmetric disjunctive list-decoding codes (SLD s_L-codes)

Definition 1 ([4,15]) A binary code X is called 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 $\mathscr{S}, \mathscr{L} \subset [t]$, $|\mathscr{S}| = s$, $|\mathscr{L}| = L$, $\mathscr{S} \cap \mathscr{L} = \varnothing$, there exists a row $\boldsymbol{x}_i$, $i \in [N]$, and a column $\boldsymbol{x}(j)$, $j \in \mathscr{L}$, such that

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

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:

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

Definition 2 ([5,8,11]) 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 SDS 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 $\mathscr{S}, \mathscr{L} \subset [t]$, $|\mathscr{S}| = s$, $|\mathscr{L}| = L$, $\mathscr{S} \cap \mathscr{L} = \varnothing$, there exists a row $\boldsymbol{x}_i$, $i \in [N]$, and a column $\boldsymbol{x}(j)$, $j \in \mathscr{L}$, such that

$$\begin{aligned} x_i(k) &= 0 \quad \forall k \in \mathscr{S} \qquad \text{and} \qquad x_i(j) = 1, \quad \text{or} \\ x_i(k) &= 1 \quad \forall k \in \mathscr{S} \qquad \text{and} \qquad x_i(j) = 0. \end{aligned}$$

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:

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

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

$$R_L^*(s+1) \le R_L^*(s) \le R_{L+1}^*(s). \tag{3}$$

Proof (Proposition 1) It immediately follows from Definition 2 that every SLD $(s+1)_L$-code is also an 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. □

1.3 Applications of symmetric disjunctive codes

Applications of SLD s_L-codes relate to the *non-adaptive symmetric group testing* which is based on the SDS of binary symbols.[1] 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 [21] 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 [22], SLD s_L-codes can be considered in connection with the problem of constructing of *two-stage non-adaptive symmetric group testing procedures*. In the first stage, one does N tests that can be depicted as a 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 SDS 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 $\le s + L - 1$ elements is selected. These units are separately tested in the second stage. Note that for $s \ge 2$ the rate $R_L^*(s)$ of SLD s_L-codes is a monotonically nondecreasing function of $L \ge 1$, 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 $\le s$ defects with the use of SLD s_L-codes.

In papers [8,9,11], we suggested another application of SLD codes called *reference communication system*, in which a set of stations transmits binary packets to the central station over a *multiple-access channel* (MAC). In general model of MAC the output sequence is an arbitrary function of the inputs. MAC can correspond to the *impulse modulation*, i.e., the

[1] The adaptive symmetric group testing for the search of binomial sample was considered in [21].

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 detail in [8,9]. If the model of MAC corresponds to the *frequency modulation*, i.e., the output ternary sequence is the SDS of the inputs, then coder and decoder can analogically use SLD s_L-codes.

Another application of SLD s_L-codes concerns with *undetermined data* [19,20], that arises in problems of pattern recognition, data compression and transmission, cryptography. Unlike the case of fully determined data, some symbols can take several values. Introduce formal definitions. Given an alphabet $A = \{a_1, a_2, \ldots, a_t\}$ 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 $a_{[t]}$, which can be specified by any basic symbol, is called *indefinite* and is denoted by $*$. Let $\mathscr{T}$ be a system of subsets $T \subseteq [t]$ and let $A^* \triangleq A^*_{\mathscr{T}} = \{a_T |\ T \in \mathscr{T}\}$ be an *undetermined alphabet* associated with the system.

Consider a problem of encoding of undetermined sequences such that the original undetermined sequence can be completely reconstructed from the encoded sequence. One encoding method refers to a *binary representation* [19,20] 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 $(N \times |\mathscr{T}|)$-matrix X^* with columns $\boldsymbol{x}(T) \in \{0,1,*\}^N$, $T \in \mathscr{T}$, where $\boldsymbol{x}(i)$ specifies $\boldsymbol{x}(T)$ if and only if $i \in T$. Let $\boldsymbol{x}(T)$ be a code of symbol a_T, then to decode the symbol it is sufficient to find the codewords of X, that specifies the given symbol code. 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 $X^* = (\nabla_{i \in T} \boldsymbol{x}(i), T \in \mathscr{T})$ give a fairly compact binary representation of undetermined alphabet associated with the system $\mathscr{T} = [t] \cup \{T \subset [t] |\ |T| \le s\}$ [20].

Similarly to the construction of two-stage group testing procedures let us suggest another method of encoding of undetermined sequences with complete reconstruction, that is based on SLD s_L-codes with $L \ge 1$. In this case it is necessary to append the vector, which indicates the correct codewords among the decoding list, to the SDS. Suppose, $X = (\boldsymbol{x}(i), i \in [t])$ is an SLD s_L-code, and the size of the set $T \subset [t]$ does not exceed s. According to the definition of SLD s_L-code, $\nabla_{i \in T} \boldsymbol{x}(i)$ symmetrically covers not more than $s + L - 1$ codewords of X, those are $\boldsymbol{x}(d_1), \ldots, \boldsymbol{x}(d_k)$, $k \le s + L - 1$, where $d_1 < d_2 < \cdots < d_k$. Let the code of symbol a_T be the ternary sequence $\boldsymbol{z}(T) = \boldsymbol{x}(T)\boldsymbol{y}(T)$, where $\boldsymbol{x}(T)$ is the SDS of $\boldsymbol{x}(i)$, $i \in T$, and $\boldsymbol{y}(T) = (y_1, \ldots, y_{s+L-1})$ is a binary column of size $s + L - 1$ defined as follows

$$\forall j \in [s + L - 1] \quad y_j = \begin{cases} 1, & \text{if } j < k \text{ and } d_j \in T, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the symbol a_T can be completely reconstructed from $\boldsymbol{z}(T)$. Moreover, the memory cost and the time complexity of the decoding procedure are the same as in the previous method.

1.4 Relations between parameters of LD s_L-codes and SLD s_L-codes

The following obvious propositions from [5,8,11] relate the rate of LD s_L-codes (1) to the rate of SLD s_L-codes (2).

Proposition 2 ([5,8,11]) *Any LD s_L-code is also an SLD s_L-code.*

Proposition 3 ([5,8,11]) *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*

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

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 ([5,8,11]) *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 4 *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 (Corollary 2) Let X be an arbitrary SLD s_L-code of length N and size t. The code X_1 obtained in Proposition 3 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 4 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 [7,12]. The use of the inequalities (4) and (5), the lower bound $\underline{R}_L(s)$ [7] and the upper bound $\overline{R}_L(s)$ [7] on the rate of LD s_L-codes yields the results below.

Theorem 1 (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 $R_L(s)$ satisfy the inequalities*

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

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*

$$\underline{R}_L(s) \le R^*_L(s) \le \overline{R}^*_L(s) \triangleq \min\{\, 2\overline{R}_L(s),\ \overline{R}_L(s-1) \,\}. \tag{6}$$

1.5 q-ary frameproof list-decoding codes

Let $q \geq 2$ be integer. In this section we define a family of codes, which is a q-ary extension of the family of binary SLD s_L-codes, and give bounds on the rate of such codes, which generalize the bounds on the rate of SLD s_L-codes presented in Sect. 1.4.

Definition 3 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 $\mathscr{S} \subset [t]$, $|\mathscr{S}| = s$, the convex hull $\{\boldsymbol{u} \in \{0, 1, \ldots, q-1\}^N : \forall\, i \in [N]\ \exists\, j \in S,\ u_i = x_i(j)\} \subset \{0, 1, \ldots, q-1\}^N$ 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:

$$R_L^{(q)}(s) \triangleq \varlimsup_{N\to\infty} \frac{\log_q t_{fld}^{(q)}(N, s, L)}{N} = \varlimsup_{t\to\infty} \frac{\log_q t}{N_{fld}^{(q)}(t, s, L)}. \tag{7}$$

Remark 1 A q-ary FLD s_1-code is the special case of separating codes [13]. More specifically, for $L = 1$, Definition 3 is equivalent to the definition of $(s, 1)$-*separating* code. Some results and applications of $(s, 1)$-separating codes are presented in the survey [2].

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

Lemma 1 *For any integer $c \geq 1$, the existence of a q-ary FLD s_L-code of length N and size t yields the existence of a q^c-ary FLD s_L-code of length $\lceil N/c \rceil$ and size t.*

Proof (Lemma 1) Suppose there exists a q-ary FLD s_L-code X of length N and size t. For any integer $c \geq 1$, one can construct a one-to-one correspondence F between symbols $a^{(c)}$ of q^c-ary alphabet and c-length sequences $(a_1, \ldots, a_c)$ of q-ary symbols. Represent $N = \lambda c + r$, where $\lambda \geq 0$ and $1 \leq r \leq c$, and consider a q^c-ary code X' obtained from X by the replacement of every codeword $\boldsymbol{a} = (a_1, \ldots, a_N)$ in code X by the codeword

$$\boldsymbol{a}^{(c)} \triangleq \big(F^{-1}(a_1, \ldots, a_c), F^{-1}(a_{c+1}, \ldots, a_{2c}), \ldots, F^{-1}(a_{\lambda c+1}, \ldots, a_{\lambda c+r}, 0, \ldots, 0)\big)$$

of length $\lceil N/c \rceil$. One can easily prove by contradiction that the code X' is an FLD s_L-code. □

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

Corollary 3 *The rates of LD s_L-codes and q-ary FLD s_L-codes satisfy the following inequalities:*

$$\frac{\lfloor \log_2 q \rfloor}{\log_2 q} R_L(s) \leq R_L^{(q)}(s) \leq \frac{q}{\log_2 q} R_L(s). \tag{8}$$

Corollary 4 *The rate of q-ary FLD s_L-codes has the following upper bound:*

$$R_L^{(q)}(s) \leq \frac{q-1}{\log_2 q} R_L(s-1). \tag{9}$$

Proof (Corollary 3) By Proposition 2 and Lemma 1 the existence of LD s_L-code of length N and size t yields the existence of $2^{\lfloor \log_2 q \rfloor}$-ary FLD s_L-code of length $\lceil N/\lfloor \log_2 q \rfloor \rceil$ which can be interpreted as a q-ary code. Therefore, the left equality in (8) is true. To prove the right inequality in (8) consider an arbitrary q-ary FLD s_L-code X. Introduce the binary $(Nq \times t)$ matrix X_1 obtained by the standard replacement of each q-ary symbol $x, x \in \{0, 1, \ldots, q-1\}$, 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 (Corollary 4) Let X be an arbitrary q-ary FLD s_L-code and X_1 be the binary LD s_L-code of length qN, size t and weight N obtained from code X in the proof of Corollary 3. Then the code X_2 obtained in Proposition 4 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 (9) is correct. □

The next evident Corollary 5 is the consequence of the two previous corollaries, the lower bounds $\underline{R}_L(s)$ and the upper bounds $\overline{R}_L(s)$ on the rate of LD s_L-codes [7]. For a fixed $L \geq 1$ and $s \to \infty$ the mentioned bounds have the following asymptotic behaviours [7]

$$\underline{R}_L(s) = \frac{L}{s^2 \log_2 e}(1 + o(1)), \quad \overline{R}_L(s) = \frac{2L \log_2 s}{s^2}(1 + o(1)).$$

Corollary 5 *Let q, L be fixed and $s \to \infty$. The following bounds on the rate of q-ary FLD s_L-codes hold:*

$$\frac{L\lfloor \log_2 q \rfloor}{s^2 \log_2 q \ \log_2 e}(1 + o(1)) \leq R_L^{(q)}(s) \leq \frac{2L(q-1)\log_q s}{s^2}(1 + o(1)), \quad s \to \infty. \quad (10)$$

1.6 Random coding bounds on the rates of SLD s_L-codes and FLD 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 [7,12].

Theorem 2 (Lower random coding bound $\underline{R}_L^*(s)$).
The following three statements hold.

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

$$R_L^*(s) \geq \underline{R}_L^*(s) \triangleq \max_{0<Q\leq 1/2} \left(h(Q) + \frac{B_L(s, Q)}{s + L - 1} \right), \quad (11)$$

where

$$\begin{aligned} h(Q) &\triangleq -Q \log_2 Q - (1-Q)\log_2[1-Q], \\ B_L(s, Q) &\triangleq Q \log_2 \left[\frac{p(1-z)}{p(1-z) + q(1-z)} \right] + (1-Q)\log_2 \left[\frac{p(z)}{p(z) + q(z)} \right], \\ p(z) &\triangleq p_L(s, z) = z^s (z - z^s)^L, \\ q(z) &\triangleq q_L(s, z) = (z - z^s)(1 - z^s - (1-z)^s)^L, \end{aligned} \quad (12)$$

and $z \in (0, 1)$ is the unique root of the equation

$$Q(p(z) + q(z)) = (1-Q)(p(1-z) + q(1-z)). \quad (13)$$

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

$$\underline{R}^*_L(s) \geq \frac{L}{s^2 \log_2 e}(1 + o(1)). \tag{14}$$

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

$$\underline{R}^*_\infty(s) \triangleq \lim_{L\to\infty} \underline{R}^*_L(s) = \log_2 \left[\frac{(s-1)^{s-1}}{s^s} + 1 \right]. \tag{15}$$

If $s \to \infty$, *then*

$$\underline{R}^*_\infty(s) = \frac{\log_2 e}{es}(1 + o(1)) = \frac{0.5307\ldots}{s}(1 + o(1)). \tag{16}$$

The following theorem is given without a proof, the proof will be given in paper [18].

Theorem 3 ([18]) (Lower random coding bound $\underline{R}^{(q)}_L(s)$).
The following four statements hold.

1. *For any fixed* $q \geq 2$, $s \geq 2$ *and* $L \geq 1$ *the following lower bound holds:*

$$R^{(q)}_L(s) \geq \underline{R}^{(q)}_L(s) \triangleq \max_{q' \geq q} \frac{-\log_q P(q', s, L)}{(s + L - 1)k(q, q')}, \quad \textit{where} \tag{17}$$

$$P(q, s, L) \triangleq \sum_{m=1}^{\min\{q,s\}} \binom{q}{m} \left(\frac{m}{q}\right)^L \sum_{k=0}^{m} (-1)^k \binom{m}{k} \left(\frac{m-k}{q}\right)^s, \tag{18}$$

$$k(q, q') \triangleq \begin{cases} 1, & \textit{for} \quad q = q', \\ \lceil \frac{q'}{q-1} \rceil, & \textit{otherwise}. \end{cases} \tag{19}$$

2. *For any fixed* $q \geq 2$, $L \geq 1$ *and* $s \to \infty$

$$\underline{R}^{(q)}_L(s) = \frac{L(q-1)\log_q e}{s^2 (\log_2 e)^2}(1 + o(1)), \quad s \to \infty. \tag{20}$$

3. *For any fixed* $q \geq 2$ *and* $s \geq 2$ *there exists a limit* $\underline{R}^{(q)}_\infty(s) \triangleq \lim_{L\to\infty} \underline{R}^{(q)}_L(s)$, *which satisfies the following asymptotic equality:*

$$\underline{R}^{(q)}_\infty(s) = \frac{(q-1)\log_q e}{es}, \quad s \to \infty. \tag{21}$$

4. *For any fixed* $s \geq 2$ *and* $L \geq 1$ *there exists a limit*

$$\lim_{q\to\infty} \underline{R}^{(q)}_L(s) = \frac{L}{s + L - 1}.$$

The numerical values of the lower bound (11)–(13) are shown in Table 1, where the argument of maximum in (11) is denoted by $Q^*_L(s)$. Note that the lower bound (11)–(13) improves the random coding bound obtained in [16] using the ensemble with independent binary symbols of codewords. In addition one can see that for $q = 2$ and small values of $s \geq 2$ and $L \geq 1$, the lower bound (11)–(13) is greater than both the lower bound (17)–(19) and the lower bound (6) based the lower bound $\underline{R}_L(s)$ on the rate of LD s_L-codes from [7].

Note that, for $s \to \infty$, the asymptotic lower bound of $\underline{R}^*_L(s)$ (14) coincides both with the asymptotics (20) and with the asymptotic behavior of the random coding bound on the rate of

Table 1 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

LD s_L-codes [7]. In addition, for $L \to \infty$, the asymptotics of $\underline{R}_L^*(s)$ (15) coincides with the asymptotic behavior of the mentioned above bound from [7], and (16) coincides with (21).

The most recent bounds on the rate of q-ary FLD s_1-codes are obtained in paper [17]. As $s \to \infty$ the asymptotic behaviour of the lower bound derived by Shangguan C. et al. is equal to

$$\frac{(q-1)\log_q e}{s^2 e}(1+o(1)), \quad s \to \infty,$$

that is less than the asymptotics (20). The asymptotic behaviour of the upper bound obtained in [17] exceeds twice the asymptotics (10).

2 Proof of Theorem 2

The following Lemma will be used in the proof of Theorem 2.

Lemma 2 *The function*

$$\psi(z) \triangleq \frac{p(z)+q(z)}{p(1-z)+q(1-z)}, \quad 0 < z < 1, \tag{22}$$

where the functions $p(z)$ and $q(z)$ are defined in (12), continuously maps the interval $(0,1)$ into the interval $(0,+\infty)$ and strictly increases.

Proof (Lemma 2) Let us introduce the following function

$$g(z) \triangleq g(s,z) = \frac{z-z^s}{1-z-(1-z)^s}, \quad 0 < z < 1. \tag{23}$$

Rewrite the formula (22) using the monotonically increasing function $g(z)$ (23):

$$\psi(z) = \frac{z^s(g(z))^L + (z - z^s)(1 + g(z))^L}{(1-z)^s + (1 - z - (1-z)^s)(1 + g(z))^L}. \tag{24}$$

The division of the numerator and the denominator of (24) by $(z - z^s)(1 + g(z))^L$ leads to

$$\psi(z) = \frac{\left(\frac{g(z)}{1+g(z)}\right)^L \cdot \frac{z^s}{z-z^s} + 1}{\frac{(1-z)^s}{z-z^s} \cdot \frac{1}{(1+g(z))^L} + \frac{1}{g(z)}},$$

where the function $\frac{z^s}{z-z^s}$ is strictly increasing and the function $\frac{(1-z)^s}{z-z^s}$ is strictly decreasing. Thus, it is clear that $\psi(z)$ is strictly increasing.

Note that $g(z) \to \frac{1}{s-1}$ as $z \to 0$ and $g(z) \to s - 1$ as $z \to 1$. Therefore, by (24) the following limits are true:

$$\lim_{z \to 0+0} \psi(z) = 0,$$
$$\lim_{z \to 1-0} \psi(z) = +\infty.$$

Lemma 2 is proved. □

Proof (Statement 1) Fix $L \geq 1$, $s \geq 2$ and a parameter Q, $0 < Q \leq 1/2$. The bound (11)–(13) is obtained by the method of random coding over the ensemble of binary constant-weight codes [10] 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 $\lfloor QN \rfloor$. A pair of sets $(\mathscr{S}, \mathscr{L})$, $|\mathscr{S}| = s$, $|\mathscr{L}| = L$, $\mathscr{S} \cap \mathscr{L} = \varnothing$, is called an s_L^*-*bad pair* in code X if

$$\nabla_{i \in \mathscr{S}} \boldsymbol{x}(i) \trianglerighteq \nabla_{j \in \mathscr{L}} \boldsymbol{x}(j).$$

A codeword $\boldsymbol{x}(j)$ is called an s_L^*-*bad codeword* in code X if there exists s_L^*-bad pair of sets $(\mathscr{S}, \mathscr{L})$ such as $j \in \mathscr{L}$. For the ensemble $E(N, t, Q)$, denote by $P(N, Q, s, L)$ the probability of the event "the pair $(\mathscr{S}, \mathscr{L})$ is s_L^*-bad in code X", by $P_1(N, t, Q, s, L)$ the probability of the event "the codeword $\boldsymbol{x}(j)$ is s_L^*-bad in code X". Evidently,

$$P_1(N, t, Q, s, L) \leq \binom{t-1}{s+L-1}\binom{s+L-1}{s} P(N, Q, s, L) \leq \frac{t^{s+L-1}}{s!(L-1)!} P(N, Q, s, L).$$

Therefore, the expectation of the number of s_L^*-bad codewords in the code X is at most

$$t \cdot P_1(N, t, Q, s, L) \leq t \frac{t^{s+L-1}}{s!(L-1)!} P(N, Q, s, L),$$

that is why, for

$$t \leq \left[\frac{s!(L-1)!}{2P(N, Q, s, L)}\right]^{1/(s+L-1)},$$

the expectation of the number of s_L^*-bad codewords does not exceed half of the codewords, that is there exists an SLD s_L-code of length N and size $\lfloor t/2 \rfloor$. Thus, the maximal size of SLD s_L-codes satisfies the inequality

$$t_{sld}(N, s, L) \geq \left\lfloor \frac{1}{2} \left\lfloor \left[\frac{s!(L-1)!}{2P(N, Q, s, L}\right]^{1/(s+L-1)} \right\rfloor \right\rfloor.$$

Then, according to the definition (2) of the rate $R^*_L(s)$, the following inequality holds:

$$R^*_L(s) \geq \underline{R}^*_L(s) \triangleq \frac{1}{s+L-1} \max_{0<Q<1} A^*_L(s, Q),$$
$$A^*_L(s, Q) \triangleq \overline{\lim_{N\to\infty}} \frac{-\log_2 P(N, Q, s, L)}{N}. \tag{25}$$

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 $0 < Q \leq 1/2$.

To complete the proof of Statement 1 of Theorem 2, it is sufficient to compute the function $A^*_L(s, Q)$ (25).

Let us use the terminology of *types* [3]. Consider an arbitrary set of size s consisting of binary codewords of length N and weight $\lfloor QN \rfloor$: $(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(s))$, where $\boldsymbol{x}(i) \in \{0,1\}^N$, $\forall i \in [s]$. The set forms $(N \times s)$-matrix X_s. Let $\boldsymbol{a} \triangleq (a_1, \ldots, a_s) \in \{0,1\}^s$. Denote a *type* of the matrix X_s by $\{n(\boldsymbol{a})\}$, where $n(\boldsymbol{a})$, $0 \leq n(\boldsymbol{a}) \leq N$ is the number of $\boldsymbol{a}$-rows in the matrix X_s. Obviously, for any matrix X_s we have

$$\sum_{\boldsymbol{a}} n(\boldsymbol{a}) = N.$$

By $n(\boldsymbol{0})$ ($n(\boldsymbol{1})$) denote the number of the rows in X_s consisting of all zeros (ones). It allows to represent $P(N, Q, s, L)$ as

$$P(N, Q, s, L) = \sum_{\{n(\boldsymbol{a})\}\in\mathscr{N}} \frac{N!}{\prod_{\boldsymbol{a}} n(\boldsymbol{a})!} \binom{N - n(\boldsymbol{0}) - n(\boldsymbol{1})}{\lfloor QN \rfloor - n(\boldsymbol{1})}^L \binom{N}{\lfloor QN \rfloor}^{-s-L}, \tag{26}$$

where the set $\mathscr{N}$ consists of all possible types $n(\boldsymbol{a}), \boldsymbol{a} \in \{0,1\}^s$, such that:

$$0 \leq n(\boldsymbol{a}) \leq N \quad \forall \boldsymbol{a} \in \{0,1\}^s, \quad n(\boldsymbol{0}) \leq N - \lfloor QN \rfloor, \quad n(\boldsymbol{1}) \leq \lfloor QN \rfloor,$$
$$\sum_{\boldsymbol{a}} n(\boldsymbol{a}) = N, \quad \sum_{\boldsymbol{a}:\, a_i = 1} n(\boldsymbol{a}) = \lfloor QN \rfloor \quad \forall i \in [s]. \tag{27}$$

Let $N \to \infty$. For every type $n(\boldsymbol{a})$, $\boldsymbol{a} \in \{0,1\}^s$, let us consider the corresponding distribution $\tau \triangleq \{\tau(\boldsymbol{a})\} : \tau(\boldsymbol{a}) = \frac{n(\boldsymbol{a})}{N}$. Thus, for $N \to \infty$, the set $\mathscr{N}$ agrees with the set $\mathscr{T}$ consisting of the distributions with the following properties induced by (27):

$$\tau \in \mathscr{T} \Longleftrightarrow \left\{ \begin{array}{c} 0 \leq \tau(\boldsymbol{a}) \leq 1 \quad \forall \boldsymbol{a} \in \{0,1\}^s, \quad \tau(\boldsymbol{0}) \leq 1 - Q, \quad \tau(\boldsymbol{1}) \leq Q, \\ \sum\limits_{\boldsymbol{a}\in\{0,1\}^s} \tau(\boldsymbol{a}) = 1, \quad \sum\limits_{\boldsymbol{a}:\, a_i = 1} \tau(\boldsymbol{a}) = Q \quad \forall i \in [s]. \end{array} \right\} \tag{28}$$

Applying the Stirling approximation, we obtain the following logarithmic asymptotic behavior of the summand in the sum (26) for $\tau \in \mathscr{T}$:

$$-\log_2 \sum_{\{n(\boldsymbol{a})\}\in\mathscr{N}} \frac{N!}{\prod_{\boldsymbol{a}} n(\boldsymbol{a})!} \binom{N - n(\boldsymbol{0}) - n(\boldsymbol{1})}{\lfloor QN \rfloor - n(\boldsymbol{1})}^L \binom{N}{\lfloor QN \rfloor}^{-s-L}$$
$$= NF(\tau, Q)(1 + o(1)), \quad \text{where,}$$
$$F(\tau, Q) \triangleq \sum_{\boldsymbol{a}} \tau(\boldsymbol{a}) \log_2[\tau(\boldsymbol{a})] - (1 - \tau(\boldsymbol{0}) - \tau(\boldsymbol{1}))Lh\left(\frac{Q - \tau(\boldsymbol{1})}{1 - \tau(\boldsymbol{0}) - \tau(\boldsymbol{1})}\right)$$
$$+ (s + L)h(Q). \tag{29}$$

For the given Q, let the minimum of the function $F(\tau, Q)$ be attained at $\tau_Q = \{\tau_Q(\boldsymbol{a})\}$, then

$$A_L^*(s, Q) \triangleq \overline{\lim_{N\to\infty}} \frac{-\log_2 P(s, L, Q, N)}{N} = F(\tau_Q, Q) = \min_{\tau\in\mathscr{T}} F(\tau, Q). \tag{30}$$

Since F is continuous in the admissible compact space $\mathscr{T}$, finding the minimum of F under constraints (28) with excluded boundaries is sufficient to calculate (30). Let us write the minimization problem: $F \to \min$,

$$\text{Search domain}\mathbb{T}: \qquad 0 < \tau(\boldsymbol{a}) < 1 \quad \forall\, \boldsymbol{a} \in \{0,1\}^s, \quad \tau(\boldsymbol{1}) < Q, \quad \tau(\boldsymbol{0}) < 1 - Q, \tag{31}$$

$$\text{Restrictions:} \qquad \begin{cases} \sum\limits_{\boldsymbol{a}\in\{0,1\}^s} \tau(\boldsymbol{a}) = 1, \\ \sum\limits_{\boldsymbol{a}:\, a_i=1} \tau(\boldsymbol{a}) = Q \quad \forall\, i \in [s], \end{cases} \tag{32}$$

$$\text{Main Function:} \qquad F(\tau, Q) = (29): \ \mathbb{T} \to \mathbb{R}. \tag{33}$$

To find the extremal distribution τ_Q we apply the standard Lagrange multipliers method. Consider the Lagrangian:

$$\Lambda \triangleq F(\tau, Q) + \lambda_0 \left(\sum_{\boldsymbol{a}\in\{0,1\}^s} \tau(\boldsymbol{a}) - 1 \right) + \sum_{i=1}^{s} \lambda_i \left(\sum_{\boldsymbol{a}:a_i=1} \tau(\boldsymbol{a}) - Q \right). \tag{34}$$

The necessary conditions for the extremal distribution τ_Q are:

$$\begin{cases} \frac{\partial\Lambda}{\partial(\tau(\boldsymbol{a}))} = \log_2[\tau(a)] + \log_2 e + \lambda_0 + \sum_{i=1}^{s} \lambda_i a_i = 0, \quad \forall\, \boldsymbol{a} \in \{0,1\}^s \setminus \{\boldsymbol{0}, \boldsymbol{1}\}, \\ \frac{\partial\Lambda}{\partial(\tau(\boldsymbol{0}))} = \log_2[\tau(\boldsymbol{0})] + \log_2 e + \lambda_0 + L \log_2 \left[\frac{1-\tau(\boldsymbol{0})-\tau(\boldsymbol{1})}{1-Q-\tau(0)} \right] = 0, \\ \frac{\partial\Lambda}{\partial(\tau(\boldsymbol{1}))} = \log_2[\tau(\boldsymbol{1})] + \log_2 e + \lambda_0 + \sum_{i=1}^{s} \lambda_i + L \log_2 \left[\frac{1-\tau(\boldsymbol{0})-\tau(\boldsymbol{1})}{Q-\tau(\boldsymbol{1})} \right] = 0. \end{cases} \tag{35}$$

Let us show that the matrix of second derivatives of the Lagrangian is positive definite. Indeed, we have

$$\begin{aligned} \frac{\partial^2\Lambda}{\partial(\tau(\boldsymbol{a}))^2} &= \frac{\log_2 e}{\tau(\boldsymbol{a})} > 0, \quad \forall\, \boldsymbol{a} \in \{0,1\}^s \setminus \{\boldsymbol{0}, \boldsymbol{1}\}, \\ \frac{\partial^2\Lambda}{\partial(\tau(\boldsymbol{0}))^2} &= \frac{\log_2 e}{\tau(\boldsymbol{0})} + L \log_2 e \frac{Q - \tau(\boldsymbol{1})}{(1 - \tau(\boldsymbol{0}) - \tau(\boldsymbol{1}))(1 - Q - \tau(\boldsymbol{0}))} > 0, \\ \frac{\partial^2\Lambda}{\partial(\tau(\boldsymbol{1}))^2} &= \frac{\log_2 e}{\tau(\boldsymbol{1})} + L \log_2 e \frac{1 - Q - \tau(\boldsymbol{0})}{(1 - \tau(\boldsymbol{0}) - \tau(\boldsymbol{1}))(Q - \tau(\boldsymbol{1}))} > 0, \\ \frac{\partial^2\Lambda}{\partial(\tau(\boldsymbol{0}))\partial(\tau(\boldsymbol{1}))} &= -L \log_2 e \frac{1}{1 - \tau(\boldsymbol{0}) - \tau(\boldsymbol{1})} < 0, \end{aligned}$$

and the other elements of the matrix are zeros. That is why, this matrix is positive definite. Note that the matrix of second derivatives of the function $F(\tau, Q)$ coincides with the above matrix. Therefore [14], F is strictly $\cup$-convex in the domain $\mathbb{T}$. Hence a local minimum of F in $\mathbb{T}$ is global and unique. Due to the Karush-Kuhn-Tacker theorem [14], it is clear that if

there exists a solution satisfying the system (35) and the constraints (32), then it provides a local minimum of F, therefore, it gives the desired minimum distribution τ_Q.

Note that the symmetry of the problem yields equality: $\nu \triangleq \lambda_1 = \lambda_2 = \cdots = \lambda_s$. To prove this, we need to check that $\lambda_i = \lambda_j$ for $i \neq j$. Let $\bar{\boldsymbol{a}}_i \triangleq (0, \ldots, 1, \ldots, 0)$ be a row of length s, which has 1 at the i-th position and $0's$ at the other positions. A permutation of indices i and j leads to an equivalent problem. Hence, if τ_Q^1 is a solution, then τ_Q^2 is also a solution, where $\tau_Q^2(\boldsymbol{a}) \triangleq \tau_Q^1(\tilde{\boldsymbol{a}})$ and $\tilde{\boldsymbol{a}}$ is a row, obtained by permutation of indices i and j from the row $\boldsymbol{a}$. The uniqueness of the solution τ_Q implies that the distribution τ_Q^1 coincides with the distribution τ_Q^2. In particular, $\tau_Q^1(\bar{\boldsymbol{a}}_i) = \tau_Q^2(\bar{\boldsymbol{a}}_i) = \tau_Q^1(\bar{\boldsymbol{a}}_j)$. From the first equation of (35), it follows that $\lambda_i = \lambda_j$.

Introduce a parameter $\mu \triangleq e2^{\lambda_0}$. Then the Eq. (35) have the form:

$$\begin{cases} \log_2 \mu + \log_2[\tau(\boldsymbol{a})] + \nu \sum_{i=1}^s a_i = 0, \\ \log_2 \mu + \log_2[\tau(\boldsymbol{0})] + L \log_2 \left[\frac{1-\tau(\boldsymbol{0})-\tau(\boldsymbol{1})}{1-Q-\tau(\boldsymbol{0})}\right] = 0, \\ \log_2 \mu + \log_2[\tau(\boldsymbol{1})] + L \log_2 \left[\frac{1-\tau(\boldsymbol{0})-\tau(\boldsymbol{1})}{Q-\tau(\boldsymbol{1})}\right] + s\nu = 0. \end{cases} \tag{36}$$

After substitution $z \triangleq \frac{1}{1+2^{-\nu}}$, $0 < z < 1$, the first equation of (36) gives

$$\tau(\boldsymbol{a}) = \frac{2^{-\nu \sum a_i}}{\mu} = \frac{1}{\mu z^s}(1-z)^{\sum a_i} z^{s-\sum a_i} \quad \forall\, \boldsymbol{a} \in \{0,1\}^s \setminus \{\boldsymbol{0}, \boldsymbol{1}\}. \tag{37}$$

Substitution (37) into the first and the second equations of the system (32) leads to

$$1 = \frac{1}{\mu z^s} \sum_{i=1}^{s-1} \binom{s}{i} z^i (1-z)^{s-i} + \tau(\boldsymbol{0}) + \tau(\boldsymbol{1}) = \frac{1 - z^s - (1-z)^s}{\mu z^s} + \tau(\boldsymbol{0}) + \tau(\boldsymbol{1}), \tag{38}$$

$$Q = \frac{1}{\mu z^s} \sum_{i=1}^{s-1} \binom{s-1}{i} z^i (1-z)^{s-i} + \tau(\boldsymbol{1}) = \frac{1 - z - (1-z)^s}{\mu z^s} + \tau(\boldsymbol{1}), \tag{39}$$

correspondingly. Subtraction (39) from (38) yields

$$1 - Q = \frac{z - z^s}{\mu z^s} + \tau(\boldsymbol{0}). \tag{40}$$

Due to (38)–(40) the second and third equations of the system (36) are equivalent to

$$\begin{aligned} \mu \left(1 - Q - \frac{z - z^s}{\mu z^s}\right) \left(\frac{1 - z^s - (1-z)^s}{z - z^s}\right)^L &= 1, \\ \mu \left(Q - \frac{1 - z - (1-z)^s}{\mu z^s}\right) \left(\frac{1 - z^s - (1-z)^s}{1 - z - (1-z)^s}\right)^L &= 1, \end{aligned} \tag{41}$$

respectively.

To shorten the formulas let us introduce the functions $p(z), q(z)$ (12) and the following $r(z)$:

$$r(z) \triangleq r_L(s, z) = z^s (1 - z^s - (1-z)^s)^L.$$

By using this notation, we obtain the following expressions of μ derived from the equations appearing in (41):

$$\mu = \frac{1}{1-Q}\frac{p(z)+q(z)}{r(z)}, \tag{42}$$

$$\mu = \frac{1}{Q}\frac{p(1-z)+q(1-z)}{r(z)}. \tag{43}$$

Equating of (42) and (43) leads to the equation on the parameter z:

$$Q(p(z)+q(z)) = (1-Q)(p(1-z)+q(1-z)),$$

which coincides with Eq. (13).

The substitutions (42) into (40) and (43) into (39) give:

$$\tau(\boldsymbol{0}) = (1-Q)\frac{p(z)}{p(z)+q(z)},$$

$$\tau(\boldsymbol{1}) = Q\frac{p(1-z)}{p(1-z)+q(1-z)}. \tag{44}$$

By Lemma 2, Eq. (13) has a unique solution. That is why there exists a solution of the system (35) and the minimum of $F(\tau, Q)$ (29) is attained at the distribution τ specified by (37) and (44). So, let us calculate the minimum of $F(\tau, Q)$ by substitution of (37) and (44). At the beginning, we compute the following sum:

$$\sum_{\boldsymbol{a}:\,\boldsymbol{a}\neq\boldsymbol{0},\boldsymbol{1}} \tau(\boldsymbol{a})\log_2[\tau(\boldsymbol{a})] = \{\text{by (37)}\}$$

$$= \sum_{i=1}^{s-1}\binom{s}{i}\frac{1}{\mu z^s}(1-z)^{s-i}z^i\left(\log_2\left[\frac{1}{\mu z^s}\right] + i\log_2 z + (s-i)\log_2[1-z]\right)$$

$$= \frac{1-z^s-(1-z)^s}{\mu z^s}\log_2\left[\frac{1}{\mu z^s}\right] + \frac{z-z^s}{\mu z^s}\log_2\left[z^s\right] + \frac{1-z-(1-z)^s}{\mu z^s}\log_2\left[(1-z)^s\right]$$

$$= \{\text{by (38), (40) and (39)}\}$$

$$= (1-\tau(\boldsymbol{0})-\tau(\boldsymbol{1}))\log_2\left[\frac{1}{\mu z^s}\right] + (1-Q-\tau(\boldsymbol{0}))\log_2\left[z^s\right] + (Q-\tau(\boldsymbol{1}))\log_2\left[(1-z)^s\right]$$

$$= (1-Q-\tau(\boldsymbol{0}))\log_2\left[\frac{1}{\mu}\right] + (Q-\tau(\boldsymbol{1}))\log_2\left[\frac{(1-z)^s}{\mu z^s}\right]. \tag{45}$$

Further, the use of (45) implies

$$\sum_{\boldsymbol{a}:\,\boldsymbol{a}\neq\boldsymbol{0},\boldsymbol{1}} \tau(\boldsymbol{a})\log_2[\tau(\boldsymbol{a})] - (1-\tau(\boldsymbol{0})-\tau(\boldsymbol{1}))Lh\left(\frac{Q-\tau(\boldsymbol{1})}{1-\tau(\boldsymbol{0})-\tau(\boldsymbol{1})}\right)$$

$$= (1-Q-\tau(\boldsymbol{0}))\left(-\log_2\mu - L\log_2\left[\frac{1-\tau(\boldsymbol{0})-\tau(\boldsymbol{1})}{1-Q-\tau(\boldsymbol{0})}\right]\right)$$

$$+ (Q-\tau(\boldsymbol{1}))\left(-\log_2\mu - \log_2\left[\frac{z^s}{(1-z)^s}\right] - L\log_2\left[\frac{1-\tau(\boldsymbol{0})-\tau(\boldsymbol{1})}{Q-\tau(\boldsymbol{1})}\right]\right)$$

$$= \{\text{by (36)}\}$$

$$= (1-Q-\tau(\boldsymbol{0}))\log_2[\tau(\boldsymbol{0})] + (Q-\tau(\boldsymbol{1}))\log_2\tau(\boldsymbol{1}). \tag{46}$$

Finally, the use of (46) and (44) leads to

$$\begin{aligned} F(\tau, Q) &= (s+L)h(Q) + (1-Q)\log_2[\tau(\boldsymbol{0})] + Q\log_2[\tau(\boldsymbol{1})] \\ &= (s+L-1)h(Q) + (1-Q)\log_2\left[\frac{p(z)}{p(z)+q(z)}\right] \\ &\quad + Q\log_2\left[\frac{p(1-z)}{p(1-z)+q(1-z)}\right]. \end{aligned}$$

The bound (11)–(13) is proved. □

Proof (Statement 2) For fixed $s \geq 2$ and $L \geq 1$, let us interpret Eq. (13) as a function $Q_L(s, z)$ of the argument z, $0 < z < 1$, i.e.,

$$Q_L(s, z) \triangleq \frac{p(1-z)+q(1-z)}{p(1-z)+q(1-z)+p(z)+q(z)}, \tag{47}$$

where the functions $p(z)$ and $q(z)$ are determined in (12).

Due to existence and uniqueness of the root of the Eq. (13), continuity and monotonicity of the function (47) (by Lemma 2), one can rewrite the definition of the random coding bound (11)–(13) as

$$\underline{R}_L^*(s) \triangleq \max_{1/2 \leq z < 1} T_L(s, z), \tag{48}$$

where

$$T_L(s, z) \triangleq h(Q_L(s, z)) + B_L(s, Q_L(s, z)). \tag{49}$$

Let $L \geq 1$ be fixed and $s \to \infty$. If in definition (49) we put $z = 1 - \lambda/s$, where the parameter $\lambda = \lambda_L$ is independent of s, then (48) means that

$$\underline{R}_L^*(s) \geq T_L\left(s, 1 - \frac{\lambda}{s}\right). \tag{50}$$

Further, we will show that

$$T_L\left(s, 1 - \frac{\lambda}{s}\right) = \frac{L}{s^2}\left(-\lambda \log_2[1 - e^{-\lambda}]\right)(1 + o(1)). \tag{51}$$

Let us introduce the following notation:

$$\begin{aligned} U_L(s, z) &\triangleq \frac{p(1-z)}{p(1-z)+q(1-z)}, \\ V_L(s, z) &\triangleq \frac{p(z)}{p(z)+q(z)}. \end{aligned} \tag{52}$$

Then the function (49) can be represented as

$$T_L(s, z) = -Q\log_2 Q - (1-Q)\log_2[1-Q] + \frac{1}{s+L-1}\left(Q\log_2 U + (1-Q)\log_2 V\right), \tag{53}$$

where the shorthands $Q = Q_L(s, z)$, $U = U_L(s, z)$ and $V = V_L(s, z)$ are used.

Computation of two first terms of asymptotic expansions of $p(z), q(z), p(1-z), q(1-z)$ (12) for $z = 1 - \lambda/s$ and $s \to \infty$ leads to the equalities

$$p(1-z) = p\left(\frac{\lambda}{s}\right) = \left(\frac{\lambda}{s}\right)^{s+L} + o\left(\frac{1}{s^{s+L}}\right),$$
$$q(1-z) = q\left(\frac{\lambda}{s}\right) = \frac{\lambda(1-e^{-\lambda})^L}{s} + \frac{L\lambda^3 e^{-\lambda}(1-e^{-\lambda})^{L-1}}{2s^2} + o\left(\frac{1}{s^2}\right),$$
$$p(z) = p\left(1-\frac{\lambda}{s}\right) = e^{-\lambda}(1-e^{-\lambda})^L + \frac{\lambda e^{-\lambda}(1-e^{-\lambda})^L(\lambda + L\lambda - 2Le^{\lambda} - \lambda e^{\lambda})}{2(e^{\lambda}-1)s} + o\left(\frac{1}{s}\right),$$
$$q(z) = q\left(1-\frac{\lambda}{s}\right) = (1-e^{-\lambda})^{L+1} + \frac{\lambda e^{-\lambda}(1-e^{-\lambda})^L(\lambda + L\lambda - 2e^{\lambda})}{2s} + o\left(\frac{1}{s}\right). \quad (54)$$

Using (54), one can obtain the following asymptotics for the expressions (47), (52)

$$Q_L\left(s, 1-\frac{\lambda}{s}\right) = \frac{\lambda}{s} + \frac{L\lambda^2}{(e^{\lambda}-1)s^2} + o\left(\frac{1}{s^2}\right),$$
$$U_L\left(s, 1-\frac{\lambda}{s}\right) = \left(\frac{\lambda}{s}\right)^{s+L-1}(1-e^{-\lambda})^{-L}(1+o(1)),$$
$$V_L\left(s, 1-\frac{\lambda}{s}\right) = e^{-\lambda}\left(1 + \frac{\lambda - L\lambda - \lambda^2/2}{s} + o\left(\frac{1}{s}\right)\right). \quad (55)$$

Finally, equalities (55) yield the asymptotic behavior of (53) that coincides with (51). Taking derivative one can check that at $\lambda = \frac{1}{\log_2 e}$ the maximum

$$\max_{\lambda>0}\left\{-\lambda \log_2[1-e^{-\lambda}]\right\} = \frac{1}{\log_2 e} \quad (56)$$

is attained. Therefore, (50), (51) and (56) imply the asymptotic inequality (14) for the random coding bound (11)–(13). □

Proof (Statement 3) Let $s \geq 2$ and $L \geq 1$ be fixed. It is clear that $g(z)$ (23) 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

$$\left(\frac{g(z)}{1+g(z)}\right)^L = c(1-z), \quad \frac{1}{2} \leq z < 1, \quad (57)$$

exists and is unique, since the left-hand side of (57) monotonically increases and the right-hand side of (57) strictly decreases. Denote this root by $z_L(s, c)$.

Let $s \geq 2$ be fixed and $L \to \infty$. The definition (48) means that

$$\underline{R}^*_L(s) \geq T_L(s, z_L(s,c))(1+o(1)), \quad L \to \infty, \quad \forall\, c = c(s) > 0. \quad (58)$$

Further, we will show that

$$T_L(s, z_L(s,c)) \cdot (1+o(1)) = \log_2[s+c] - \frac{s+c-1}{s+c}\log_2[s+c-1] + \frac{1}{s+c}\log_2\left[\frac{(s-1)^{s-1}}{s^s}\right], \quad L \to \infty. \quad (59)$$

When $L \to \infty$, it is obvious that

$$z_L(s,c) = 1 + o(1), \quad \text{and hence,}$$
$$g(z_L(s,c)) = (s-1)(1+o(1)). \tag{60}$$

The use of definitions (12) and division of upper and lower parts of fractions (47), (52) by $(1-z-(1-z)^s)$ allow us to rewrite expressions Q, U and V (47), (52) in a more convenient form

$$\begin{aligned} Q_L(s,z) &= \frac{(1-z)^s + (1-z-(1-z)^s)(1+g(z))^L}{(1-z)^s + (1-z^s-(1-z)^s)(1+g(z))^L + z^s(g(z))^L}, \\ U_L(s,z) &= \frac{(1-z)^s}{(1-z)^s + (1-z-(1-z)^s)(1+g(z))^L}, \\ V_L(s,z) &= \frac{z^s(g(z))^L}{z^s(g(z))^L(z-z^s)(1+g(z))^L}. \end{aligned} \tag{61}$$

The equalities (60)–(61) imply the following asymptotics

$$\begin{aligned} Q_L(s, z_L(s,c)) &= \frac{1}{s+c}(1+o(1)), \\ U_L(s, z_L(s,c)) &= \left(\frac{(s-1)^{s-1}}{s^s}\right)^L (1+o(1)), \\ V_L(s, z_L(s,c)) &= \frac{1}{1+\frac{s}{c}}(1+o(1)). \end{aligned} \tag{62}$$

Next, the substitution (62) into the expression (53) involves (59).

Calculating the derivative in c, one can check that maximum of the right-hand side of (59) 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 (59), then the use of (58) establishes the following inequality for the random coding bound (11)–(13):

$$\underline{R}_L^*(s) \geq \log_2\left[\frac{(s-1)^{s-1}}{s^s} + 1\right](1+o(1)), \quad L \to \infty. \tag{63}$$

To prove the equality sign in (63), let us denote arbitrary sequence of argument of maximum (48) by $z = z_L(s)$, $1/2 \leq z_L(s) < 1$. By considering the different cases that can occur, we will find each time a contradiction with (63). First, assume that the sequence $z_L(s)$ is bounded by a constant $d < 1$, i.e., $1/2 \leq z_L(s) \leq d < 1$. Then due to (61) the asymptotic equalities

$$\begin{aligned} Q_L(s, z_L(s)) &= \frac{1}{1+g(z)}(1+o(1)), \\ U_L(s, z_L(s)) &= \frac{(1-z)^s}{1-z-(1-z)^s}\frac{1}{(1+g(z))^L}(1+o(1)), \\ V_L(s, z_L(s)) &= \frac{z^s}{z-z^s}\left(\frac{g(z)}{1+g(z)}\right)^L (1+o(1)), \quad L \to \infty, \end{aligned} \tag{64}$$

hold. However, the computation of asymptotic behavior of $T_L(s, z_L(s))$ (53), using (64), yields $\underline{R}_L^*(s) = T_L(s, z_L(s)) \to 0$ as $L \to \infty$. The current case involves the contradiction with (63). Hence, it is clear without less of generality that $z_L(s) \to 1$ ((60) holds).

Further, let us assume that

$$\left(\frac{g(z)}{1+g(z)}\right)^L \frac{1}{1-z} \to 0, \quad L \to \infty. \tag{65}$$

Then using (60) and (65) one can achieve the following asymptotic behaviors of (61)

$$\begin{aligned} Q_L(s, z_L(s)) &= \frac{1}{s}(1+o(1)), \\ U_L(s, z_L(s)) &= \frac{(1-z)^{s-1}}{(1+g(z))^L}(1+o(1)), \\ V_L(s, z_L(s)) &= \frac{1}{s}\left(\frac{g(z)}{1+g(z)}\right)^L \frac{1}{1-z}(1+o(1)), \quad L \to \infty. \end{aligned} \tag{66}$$

However, the equalities (60) and (66) lead to $\underline{R}^*_L(s) = T_L(s, z_L(s)) \to 0$ as $L \to \infty$. So, the current case is in contradiction with (63).

Next, let us assume that

$$\left(\frac{g(z)}{1+g(z)}\right)^L \frac{1}{1-z} \to \infty, \quad L \to \infty. \tag{67}$$

The use of (60) and (67) leads to the following asymptotic behavior of (61)

$$\begin{aligned} Q_L(s, z_L(s)) &= \left(\frac{1+g(z)}{g(z)}\right)^L (1-z)(1+o(1)), \\ U_L(s, z_L(s)) &= \frac{(1-z)^{s-1}}{(1+(z))^L}(1+o(1)), \\ V_L(s, z_L(s)) &= 1+o(1), \quad L \to \infty. \end{aligned} \tag{68}$$

It is obvious that the equalities (60) and (68) yield

$$T_L(s, z_L(s)) = \frac{Q(s-1)}{s+L-1}\log_2[1-z] + o(1). \tag{69}$$

One can see that from the first equality in (68) it follows that

$$Q = O(1-z).$$

Therefore, the asymptotic equality (69) implies $\underline{R}^*_L(s) = T_L(s, z_L(s)) \to 0$ as $L \to \infty$ and the current case is in contradiction with (63).

Without loss of generality we can conclude that

$$\left(\frac{g(z)}{1+g(z)}\right)^L = c(1-z)(1+o(1)). \tag{70}$$

Note that (70) is similar to (57). Finally, using (60) and (70) one can obtain the equalities appearing in (62), and we obtain in this way Formula (59).

Statement 3 of Theorem 2 is proved. □

Acknowledgements A. G. Dyachkov, I. V. Vorobyev, N. A. Polyanskii and V. Yu. Shchukin have been supported in part by the Russian Foundation for Basic Research under Grant No. 16-01-00440a.

References

1. Boneh D., Shaw J.: Collusion-secure fingerprinting for digital data. IEEE Trans. Inf. Theory **44**(5), 1897–1905 (1998).
2. Cohen G.D., Schaathun H.G.: Asymptotic overview on separating codes. Technical Report 248, Department of Informatics, University of Bergen, Norway (2003).
3. Csiszar I., Korner J.: Information Theory: Coding Theorems for Discrete Memoryless Systems. Cambridge University Press, Cambridge (2011).
4. 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 (2003).
5. D'yachkov A.G., Rykov V.V.: An application of codes for the multiple access channel in the ALOHA communication system. In: Proccedings of the 6-th All-Union Seminar in Computing Networks, vol. 4, pp. 18–24. Moscow-Vinnitsa (1981) (in Russian).
6. D'yachkov A.G., Rykov V.V.: A survey of superimposed code theory. Probl. Inf. Transm. **12**(4), 229–242 (1983).
7. D'yachkov A.G., Rykov V.V.: Superimposed codes for multiple accessing of the OR-channel. In: Proceedings of the IEEE International Symposium on Information Theory (ISIT). IEEE, Boston (1998).
8. D'yachkov A.G., Rykov V.V., Rashad A.M.: Superimposed distance codes. Probl. Inf. Transm. **18**(4), 237–250 (1989).
9. D'yachkov A.G., Vorobyev I.V., Polyanskii N.A., Shchukin V.Yu.: Bounds on the rate of superimposed codes. In: Proceedings of the IEEE International Symposium on Information Theory (ISIT), pp. 2341–2345. IEEE, Honolulu (2014).
10. D'yachkov A.G., Vorobyev I.V., Polyanskii N.A., Shchukin V.Yu.: Bounds on the rate of disjunctive codes. Probl. Inf. Transm. **50**(1), 27–56 (2014).
11. D'yachkov A.G., Vorobyev I.V., Polyanskii N.A., Shchukin V.Yu.: Symmetric disjunctive list-decoding codes. In: Proceedings of the IEEE International Symposium on Information Theory (ISIT), pp. 2236–2240. IEEE, Hong Kong (2015).
12. D'yachkov A.G., Vorobyev I.V., Polyanskii N.A., Shchukin V.Yu.: Almost disjunctive list-decoding codes. Probl. Inf. Transm. **51**(2), 110–131 (2015).
13. Friedman A.D., Graham R.L., Ullman J.D.: Universal single transition time asynchronous state assignments. IEEE Trans. Comput. **18**(6), 541–547 (1969).
14. Galeev E.M., Tikhomirov V.M.: Optimization: Theory, Examples, Problems. Editorial URSS, Moscow (2000) (in Russian).
15. Kautz W.H., Singleton R.C.: Nonrandom binary superimposed codes. IEEE Trans. Inf. Theory. **10**(4), 363–377 (1964).
16. Rashad A.M.: On symmetrical superimposed codes. J. Inf. Process. Cybern. **25**(7), 337–341 (1989).
17. Shangguan C., Wang X., Ge G., Miao Y.: New Bounds For Frameproof Codes. Preprint, 2014. http://arxiv.org/pdf/1411.5782v1.
18. Shchukin V.Yu.: List Decoding for a multiple-access hyperchannel. Probl. Inf. Transm. (will be published).
19. Sholomov L.A.: Binary representation of underdetermined data. Dokl. Math. **87**(1), 116–119 (2013).
20. Sholomov L.A.: Binary representations of underdetermined data and superimposed codes. Appl. Discret. Math. **1**, 17–33 (2013) (in Russian).
21. Sobel M., Kumar S., Blumenthal S.: Symmetric binomial group-testing with three outcomes. In: Proceedings the Symposium on Statistical Decision Theory and Related Topics, pp. 119–160. Purdue University, Lafayette (1971).
22. Vilenkin P.A.: On constructions of list-decoding superimposed codes. In: Proceedings of the 6th International Workshop on Algebraic and Combinatorial Coding Theory (ACCT-6), pp. 228–231. Pskov, Russia (1998).