Fourteenth International Workshop on Algebraic and Combinatorial Coding Theory
September 7–13, 2014, Svetlogorsk (Kaliningrad region), Russia *pp. 115–126*

Almost Disjunctive List-Decoding Codes (two talks)

A. G. D'yachkov agd-msu@yandex.ru
I.V. Vorobyev vorobyev.i.v@yandex.ru
N.A. Polyanskii nikitapolyansky@gmail.com
V.Yu. Shchukin vpike@mail.ru
Moscow State University, Faculty of Mechanics and Mathematics,
Department of Probability Theory, Moscow, 119992, Russia,

Abstract. A binary code is said to be a disjunctive list-decoding s_L-code, $s \geq 1$, $L \geq 1$, (briefly, LD s_L-code) if the code is identified by the incidence matrix of a family of finite sets in which the union of any s sets can cover not more than $L-1$ other sets of the family. In this paper, we introduce a natural *probabilistic* generalization of LD s_L-code when the code is said to be an almost disjunctive LD s_L-code if the unions of *almost all* s sets satisfy the given condition. We develop a random coding method based on the ensemble of binary constant-weight codes to obtain lower bounds on the capacity and error probability exponent of such codes. For the considered ensemble our lower bounds are asymptotically tight.

Index terms. Almost disjunctive codes, capacity, error probability exponent, random coding bounds, group testing, screening experiments, two-stage search designs.

1 Notations and Definitions

Let N, t, s, and L be integers, where $1 \leq s < t$, $1 \leq L \leq t-s$. Let $\triangleq$ denote the equality by definition, $|A|$ – the size of 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 (least) integer $\leq a$ ($\geq a$). A binary $(N \times t)$-matrix

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

$i \in [N]$, $j \in [t]$, with N rows $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ and t columns $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(t)$ (codewords) is said to be a *binary code of length N and size $t = \lceil 2^{RN} \rceil$* (briefly, *(N, R)-code*), where a fixed parameter $R > 0$ is called the *rate* of code X [1]-[2]. For any code X and any subset $\mathcal{S} \subset [t]$ of size $|\mathcal{S}| = s$, the symbol $\boldsymbol{x}(\mathcal{S}) \triangleq \{\, \boldsymbol{x}(j) \; : \; j \in \mathcal{S}\}$ will denote the corresponding s-subset of codewords (columns) of the code X. The number of 1's in column $x(j)$, i.e., $|\boldsymbol{x}(j)| \triangleq \sum\limits_{i=1}^{N} x_i(j)$, is called the *weight* of $x(j)$, $j \in [t]$. We say that X is a *constant-weight* binary code of weight w, $1 < w < N$, if for any $j \in [t]$, the weight $|\boldsymbol{x}(j)| = w$. The standard symbol $\bigvee$ denotes the *disjunctive* (Boolean) sum of two binary numbers:

$$0 \bigvee 0 = 0, \quad 0 \bigvee 1 = 1 \bigvee 0 = 1 \bigvee 1 = 1,$$

as well as the component-wise disjunctive sum of two binary columns. We say that a column $\mathbf{u}$ covers column $\mathbf{v}$ ($\mathbf{u} \succeq \mathbf{v}$) if $\mathbf{u} \bigvee \mathbf{v} = \mathbf{u}$.

Definition 1. An s-subset of columns $\boldsymbol{x}(\mathcal{S})$, $|\mathcal{S}| = s$, of a code X is said to be an s_L-*bad* subset of columns in the code X if there exists a subset $\mathcal{L} \subset [t]$ of size $|\mathcal{L}| = L$, such that $\mathcal{S} \cap \mathcal{L} = \varnothing$ and the disjunctive sum

$$\bigvee_{i \in \mathcal{S}} \boldsymbol{x}(i) \succeq \bigvee_{j \in \mathcal{L}} \boldsymbol{x}(j). \tag{2}$$

Otherwise, the s-subset $\boldsymbol{x}(\mathcal{S})$ is said to be an s_L-*good* subset of columns in the code X. In other words, for any s_L-good subset of columns in a code X, the disjunctive sum of its s columns can cover not more than $L-1$ columns of the code X that are not components of the given s-subset.

Definition 2. Let ϵ, $0 \le \epsilon < 1$, be a fixed parameter. A code X is said to be a *disjunctive list-decoding* (s_L, ϵ)-*code* (or *almost disjunctive list-decoding* s_L-*code*) of *strength* s, *list size* L and *error probability* ϵ, $0 \le \epsilon < 1$, (briefly, *LD* (s_L, ϵ)-*code*), if the number $\mathbf{G}_L(s, X)$ of all s_L-good s-subsets of columns of the code X is at least $(1-\epsilon) \cdot \binom{t}{s}$. In other words, the number $\mathbf{B}_L(s, X)$ of all s_L-bad s-subsets of columns for LD (s_L, ϵ)-code X does not exceed $\epsilon \binom{t}{s}$, i.e.,

$$\mathbf{B}_L(s, X) \triangleq \binom{t}{s} - \mathbf{G}_L(s, X) \le \epsilon \cdot \binom{t}{s} \iff \frac{\mathbf{B}_L(s, X)}{\binom{t}{s}} \le \epsilon \tag{3}$$

The concept of LD (s_L, ϵ)-code can be considered as a natural "probabilistic" generalization of the classical superimposed s-code of Kautz-Singleton [3] corresponding to the case $L = 1$ and $\epsilon = 0$. For the case $L \ge 1$ and $\epsilon = 0$, disjunctive list-decoding codes (LD s_L-codes) were investigated in works [4]-[11] and the last detailed survey of the most important results obtained for LD s_L-codes is given in the recent paper [12] (see, also, preprint [13]).

Definition 3. Let $t_\epsilon(N, s, L)$ be the maximal size of LD (s_L, ϵ)-codes of length N and let $N_\epsilon(t, s, L)$ be the minimal length of LD (s_L, ϵ)-codes of size t. If $\epsilon = 0$, then the number

$$R_L(s) \triangleq \overline{\lim_{N\to\infty}} \frac{\log_2 t_0(N, s, L)}{N} = \overline{\lim_{t\to\infty}} \frac{\log_2 t}{N_0(t, s, L)} \tag{4}$$

is called [6] the rate of LD s_L-codes.

Observe [12] that at fixed $s \ge 2$, the number

$$R_\infty(s) \triangleq \lim_{L\to\infty} R_L(s), \qquad s = 2, 3, \ldots, \tag{5}$$

can be interpreted as the *maximal rate* for two-stage group testing in the disjunctive search model of any d, $d \le s$, defective elements based on LD s_L-codes. For the general two-stage group testing [9], the number $R_\infty(s)$ gives a lower bound on the corresponding rate.

Definition 4. Define the number

$$C_L(s) \triangleq \overline{\lim_{\epsilon\to 0}}\, \overline{\lim_{N\to\infty}} \frac{\log_2 t_\epsilon(N,s,L)}{N} = \overline{\lim_{\epsilon\to 0}}\, \overline{\lim_{t\to\infty}} \frac{\log_2 t}{N_\epsilon(t,s,L)} \ge R_L(s) \qquad (6)$$

called a *capacity* of almost disjunctive LD s_L-codes.

The definition (6) implies that if the parameter N is sufficiently large, then for any fixed ϵ, $\epsilon > 0$, and any fixed rate $R > 0$, there exists an LD (s_L, ϵ)-code X of length N and size $t = \lceil 2^{RN} \rceil$, i.e., (N, R)-code X, if and only if the rate $R < C_L(s)$. Obviously, $C_L(s) \le 1/s$ and the first open problem is: "how to improve this evident upper bound?"

Definition 5. Let R, $R_L(s) \le R < C_L(s)$, be a fixed parameter. Taking into account the inequality (3) from Definition 2, we introduce the concept of *error probability for almost disjunctive LD s_L-codes*:

$$\epsilon_L(s,R,N) \triangleq \min_{X\,:\,t=\lceil 2^{RN}\rceil} \left\{ \frac{\mathbf{B}_L(s,X)}{\binom{t}{s}} \right\}, \qquad (7)$$

where the minimum is taken over all (N, R)-codes X, and the function

$$\mathbf{E}_L(s,R) \triangleq \overline{\lim_{N\to\infty}} \frac{-\log_2 \epsilon_L(s,R,N)}{N}, \quad R_L(s) \le R < C_L(s), \qquad (8)$$

is said to be the *exponent* of error probability for almost disjunctive LD s_L-codes.

In Definitions 2-5 for the case $L = 1$, we use the terminology which is similar to a terminology for the concept of weakly separating designs introduced in [14]. Let X be a code of length N and size t and let $\Omega_\epsilon(X,s,t)$ be a collection of s-subsets of columns of the code X such that its size $|\Omega_\epsilon(X,s,t)| \ge (1-\epsilon)\cdot\binom{t}{s}$. The code X is said [14] to be a *disjunctive (s,ϵ)-design* (or *weakly separating s-design*), if there exists a collection $\Omega_\epsilon(X,s,t)$ such that the *disjunctive sums of any two s-subsets from the collection $\Omega_\epsilon(X,s,t)$ are different.* Weakly separating s-design can be considered [11] as an important example of information-theoretical model for the multiple-access channel [2]. It was proved [14] that the capacity of weakly separating s-designs is equal to $1/s$. For the case $L \ge 2$, the list-decoding weakly separating s-designs were suggested in the paper [15], where it was established that their capacity is equal to $1/s$ as well.

2 Lower Bounds on $R_L(s)$, $C_L(s)$ and $\mathbf{E}_L(s, R)$

The best known upper and lower bounds on the rate $R_L(s)$ of LD s_L-codes were presented in [12] (see, also, preprint [13]). For the classical case $L = 1$, these bounds have the form:

$$R_1(s) \le \overline{R}_1(s) = \frac{2\log_2 s}{s^2}(1 + o(1)), \quad s \to \infty, \tag{9}$$

$$R_1(s) \ge \underline{R}_1(s) = \frac{4e^{-2}\log_2 s}{s^2}(1 + o(1)) = \frac{0,542\,\log_2 s}{s^2}(1 + o(1)), \qquad s \to \infty. \tag{10}$$

If $s \ge 1$, $L \ge 2$, then our lower random coding bound on $R_L(s)$ was established [12] as

Theorem 1. [12] (Random coding bound $\underline{R}_L^{(1)}(s)$). **1.** *The rate*

$$R_L(s) \ge \underline{R}_L^{(1)}(s) \triangleq \frac{1}{s+L-1}\max_{0<Q<1} A_L(s,Q) = \frac{1}{s+L-1} A_L\left(s, Q_L^{(1)}(s)\right), \tag{11}$$

$$A_L(s,Q) \triangleq \log_2 \frac{Q}{1-y} - sK(Q, 1-y) - L\,K\left(Q, \frac{1-y}{1-y^s}\right), \tag{12}$$

$$K(a,b) \triangleq a \cdot \log_2 \frac{a}{b} + (1-a)\cdot \log_2 \frac{1-a}{1-b}, \quad 0 < a, b < 1, \tag{13}$$

where parameter y, $1 - Q \le y < 1$, is defined as the unique root of the equation

$$y = 1 - Q + Qy^s\left[1 - \left(\frac{y-y^s}{1-y^s}\right)^L\right], \qquad 1 - Q \le y < 1. \tag{14}$$

2. *For fixed $L = 2, 3, \ldots$ and $s \to \infty$, the asymptotic behavior of the random coding bound $\underline{R}_L^{(1)}(s)$ has the form*

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

3. *At fixed $s = 1, 2, 3, \ldots$ and $L \to \infty$, for the maximal rate $R_\infty(s)$ of two-stage group testing defined by* (5)*, the lower bound*

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

holds. If $s \to \infty$, then $\underline{R}_\infty^{(1)}(s) = \frac{\log_2 e}{e\cdot s}(1 + o(1)) = \frac{0,5307}{s}(1 + o(1))$.

In the given paper, we suggest a modification of the random coding method developed in [12] and obtain a lower bound on the capacity $C_L(s)$ along with a lower bound on the exponent of error probability $\mathbf{E}_L(s,R)$ for almost disjunctive s_L-codes. Let

$$[x]^+ \triangleq \begin{cases} x & \text{if } x \ge 0, \\ 0 & \text{if } x < 0, \end{cases} \quad \text{and} \quad h(a) \triangleq -a\log_2 a - (1-a)\log_2(1-a),\ 0 < a < 1,$$

be the standard notations for the positive part function and the binary entropy function.

Theorem 2. (Random coding lower bounds $\underline{C}(s)$ and $\underline{\mathbf{E}}_L(s,R)$). *The following three claims hold.* **Claim 1.** *The capacity $C_L(s)$ and the exponent of error probability $\mathbf{E}_L(s,R)$ for almost disjunctive LD s_L-codes satisfy inequalities*

$$C_L(s) \ \ge\ \underline{C}(s) \ \triangleq \max_{0<Q<1} C(s,Q) = C(s,Q(s)), \qquad s \ge 1, \quad L \ge 1, \tag{16}$$

$$C(s,Q) \triangleq h(Q) - [1-(1-Q)^s]\ h\left(\frac{Q}{1-(1-Q)^s}\right), \quad s \ge 1, \quad 0 < Q < 1, \tag{17}$$

and

$$\mathbf{E}_L(s,R) \ \ge\ \underline{\mathbf{E}}_L(s,R) \ \triangleq \max_{0<Q<1} E_L(s,R,Q), \qquad s \ge 1, \quad L \ge 1, \tag{18}$$

$$E_L(s,R,Q) \ \triangleq \min_{Q \le q \le \min\{1,sQ\}} \left\{\mathcal{A}(s,Q,q) + L \cdot [h(Q) - q \cdot h(Q/q) - R]^+\right\}. \tag{19}$$

where the function $\mathcal{A}(s,Q,q)$, $Q < q < \min\{1,sQ\}$, is defined in the parametric form:

$$\mathcal{A}(s,Q,q) \triangleq (1-q)\log_2(1-q) + q\log_2\left[\frac{Qy^s}{1-y}\right] + sQ\log_2\frac{1-y}{y} + sh(Q), \tag{20}$$

$$q = Q\frac{1-y^s}{1-y}, \qquad 0 < y < 1. \tag{21}$$

Claim 2. *If $s \ge 1$ is fixed, then the random coding lower bound $\underline{C}(s) > \frac{\ln 2}{s}$ and at $s \to \infty$ the asymptotic behavior of $\underline{C}(s)$ and the asymptotic behavior of the optimal value $Q(s)$ in* (16) *are:*

$$\underline{C}(s) \ = \frac{\ln 2}{s}(1+o(1)), \quad Q(s) \ = \frac{\ln 2}{s}(1+o(1)). \tag{22}$$

Claim 3. *For any $s \ge 1$ and $L \ge 1$, the lower bound $\underline{E}_L(s,R)$ defined by* (18)-(21) *is a $\cup$-convex function of the rate parameter $R > 0$. If $0 < R < \underline{C}(s)$,* *then* $\underline{E}_L(s,R) > 0$. *If* $R \ge \underline{C}(s)$, *then* $\underline{E}_L(s,R) = 0$. *In addition, there exist a number* $R_L^{(cr)}(s)$, $0 \le R_L^{(cr)}(s) < \underline{C}(s)$, *such that*

$$\underline{E}_L(s,R) = (s+L-1)\underline{R}_L^{(1)}(s) - LR, \quad \text{if} \quad 0 \le R \le R_L^{(cr)}(s), \tag{23}$$

and

$$\underline{E}_L(s,R) > (s+L-1)\underline{R}_L^{(1)}(s) - LR, \quad \text{if} \quad R > R_L^{(cr)}(s), \tag{24}$$

where the random coding bound $\underline{R}_L^{(1)}(s)$ *is given by the formulas* (11)-(14).

In Sect. 4, we present a brief proof of Claim 1 only. We omit here proofs of Claims 2-3 which formulate the analytical properties of random coding bounds $\underline{C}(s)$ and $\underline{\mathbf{E}}_L(s,R)$). Table 1 gives some numerical values of the function

$$\underline{R}_L(s) \triangleq \max\left\{ \underline{R}_1(s)\,,\, \underline{R}_L^{(1)}(s) \right\}, \quad 2 \le s \le 10, \quad 2 \le L \le 10,$$

along with the corresponding values $Q_L(s)$ of the optimal relative weight $Q_L^{(1)}(s)$ in the right-hand side of (11) if $\underline{R}_L(s) = \underline{R}_L^{(1)}(s)$, or we put $Q_L(s) \triangleq *$ if $\underline{R}_L(s) = \underline{R}_1(s)$, where the values $\underline{R}_1(s)$ were calculated in [12], i.e,

$$Q_L(s) \triangleq \begin{cases} Q_L^{(1)}(s) & \text{if } \underline{R}_L(s) = \underline{R}_L^{(1)}(s) \text{ for } (2 \le s \le 6,\ L=2) \\ & \qquad\qquad\qquad\qquad \text{or } (2 \le s \le 10,\ 3 \le L \le 10)\ , \\ * & \text{if } \underline{R}_L(s) = \underline{R}_1(s) \ \text{ for } \ (7 \le s \le 10,\ L=2). \end{cases}$$

The function $\underline{R}_L(s)$, $L \ge 2$, $s \ge 2$, can be considered as the best presently known lower bound on the rate $R_L(s)$, $L \ge 2$, $s \ge 2$, of LD s_L-codes.

Figure 1 gives graphs of the exponent of error probability for some almost disjunctive LD s_L-codes.

Figure 1:

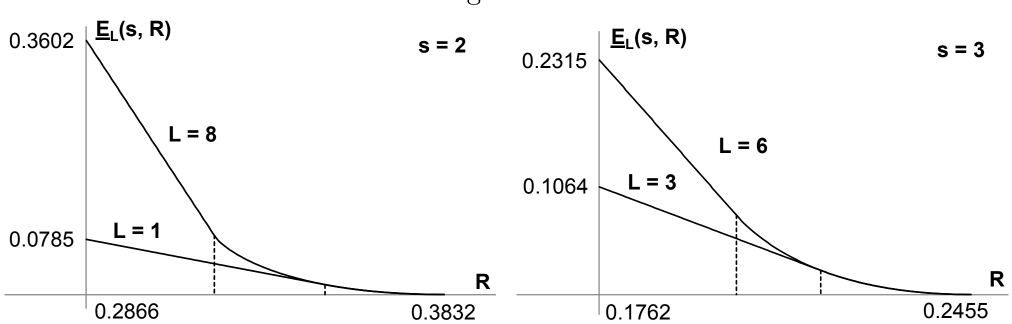

Table 1:

s_L	2_2	2_3	2_4	2_5	2_6	2_7	2_8	2_9
$Q_L(s)$	0.244	0.233	0.226	0.221	0.218	0.215	0.212	0.211
$\underline{R}_L(s)$	0.2358	0.2597	0.2729	0.2813	0.2871	0.2915	0.2948	0.2975
$R_L^{(cr)}(s)$	0.3355	0.3279	0.3242	0.3226	0.3218	0.3216	0.3215	0.3215
s_L	3_2	3_3	3_4	3_5	3_6	3_7	3_8	3_9
$Q_L(s)$	0.176	0.167	0.161	0.156	0.152	0.149	0.147	0.145
$\underline{R}_L(s)$	0.1147	0.1346	0.1469	0.1552	0.1611	0.1656	0.1690	0.1718
$R_L^{(cr)}(s)$	0.2177	0.2109	0.2065	0.2036	0.2017	0.2006	0.1998	0.1994
s_L	4_2	4_3	4_4	4_5	4_6	4_7	4_8	4_9
$Q_L(s)$	0.139	0.133	0.128	0.123	0.120	0.117	0.115	0.113
$\underline{R}_L(s)$	0.0684	0.0838	0.0941	0.1014	0.1068	0.1110	0.1143	0.1170
$R_L^{(cr)}(s)$	0.1632	0.1580	0.1542	0.1514	0.1494	0.1479	0.1468	0.1460
s_L	5_2	5_3	5_4	5_5	5_6	5_7	5_8	5_9
$Q_L(s)$	0.115	0.110	0.106	0.103	0.100	0.098	0.096	0.094
$\underline{R}_L(s)$	0.0456	0.0575	0.0660	0.0723	0.0771	0.0809	0.0840	0.0865
$R_L^{(cr)}(s)$	0.1311	0.1271	0.1240	0.1216	0.1197	0.1183	0.1171	0.1162
s_L	6_2	6_3	6_4	6_5	6_6	6_7	6_8	6_9
$Q_L(s)$	0.098	0.095	0.092	0.089	0.086	0.084	0.083	0.081
$\underline{R}_L(s)$	0.0325	0.0420	0.0490	0.0544	0.0587	0.0621	0.0649	0.0672
$R_L^{(cr)}(s)$	0.1098	0.1067	0.1041	0.1021	0.1004	0.0991	0.0980	0.0971
s_L	7_2	7_3	7_4	7_5	7_6	7_7	7_8	7_9
$Q_L(s)$	$*$	0.083	0.080	0.078	0.076	0.074	0.073	0.072
$\underline{R}_L(s)$	0.0260	0.0321	0.0380	0.0426	0.0463	0.0494	0.0519	0.0541
$R_L^{(cr)}(s)$	0.0945	0.0920	0.0899	0.0882	0.0868	0.0855	0.0845	0.0837
s_L	8_2	8_3	8_4	8_5	8_6	8_7	8_8	8_9
$Q_L(s)$	$*$	0.074	0.072	0.070	0.068	0.067	0.065	0.064
$\underline{R}_L(s)$	0.0213	0.0253	0.0303	0.0343	0.0376	0.0403	0.0426	0.0446
$R_L^{(cr)}(s)$	0.0830	0.0810	0.0793	0.0778	0.0765	0.0754	0.0745	0.0737
s_L	9_2	9_3	9_4	9_5	9_6	9_7	9_8	9_9
$Q_L(s)$	$*$	0.067	0.065	0.063	0.062	0.061	0.059	0.058
$\underline{R}_L(s)$	0.0178	0.0205	0.0248	0.0283	0.0312	0.0336	0.0357	0.0375
$R_L^{(cr)}(s)$	0.0741	0.0724	0.0709	0.0696	0.0685	0.0676	0.0667	0.0660
s	2	3	4	5	6	7	8	9
$\underline{C}(s)$	0.3832	0.2455	0.1810	0.1434	0.1188	0.1014	0.0884	0.0784
$Q(s)$	0.2864	0.2028	0.1569	0.1280	0.1080	0.0935	0.0824	0.0736
$R_1^{(cr)}(s)$	0.3510	0.2284	0.1705	0.1364	0.1137	0.0976	0.0855	0.0761

3 On Constructions of Almost Disjunctive Codes

For $L = 1$, constructions of LD s_1-codes (i.e classical disjunctive (superimposed) s-codes) based on the shortened Reed-Solomon codes were developed in [8]- [9]. The papers [8]- [9] significantly extend the optimal and suboptimal constructions of superimposed s-codes suggested in [3] and contain the detailed tables with parameters of the best known classical disjunctive (superimposed) s-codes. In addition, the table 3 from [9] along with the similar table presented in [10] gives a range of parameters (t, N, s, ϵ) corresponding to the best known LD (s_1, ϵ)-codes based on MDS codes. In the recent paper [16], it was proved that for the given parameters, the following parametric asymptotic equations

$$t = q^{\left\lfloor \frac{q}{\log_2 q} \right\rfloor},\ N = q(q+1),\ \epsilon = \epsilon(q) \to 0 \text{ if } s < q \cdot \ln 2,\ q \text{ -prime power},\ q \to \infty, \tag{25}$$

hold. Note that if $s \to \infty$ and $q \to \infty$, then the asymptotic behavior of the rate for LD (s_1, ϵ)-codes with parameters (25) is

$$\frac{\log_2 t}{N} = \frac{1}{q}(1 + o(1)) = \frac{\ln 2}{s}(1 + o(1))$$

and coincides with the asymptotic behavior of the random coding bound $\underline{C}(s)$ defined by (22).

4 Proof of Theorem 2

Proof of claim 1. For an arbitrary code X, the number $\mathbf{B}_L(s, X)$ of s_L-bad subsets of columns in the code X can be represented in the form:

$$\begin{aligned} \mathbf{B}_L(s, X) &\triangleq \sum_{\mathcal{S}\in[t],|\mathcal{S}|=s} \psi_L(X, \mathcal{S}), \\ \psi_L(X, \mathcal{S}) &\triangleq \begin{cases} 1, & \text{if the set } \boldsymbol{x}(\mathcal{S}) \text{ is } s_L\text{-bad in } X, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{26}$$

Let Q, $0 < Q < 1$, be a fixed parameter. Introduce the constant-weight ensemble $\{N, t, Q\}$ of binary $(N \times t)$-matrices X, where each column $\boldsymbol{x}(j)$, $j \in [t]$, of X is taken with replacement from the set containing $\binom{N}{w}$ binary columns of a given weight $w \triangleq \lfloor QN \rfloor$. From (26) it follows that for the ensemble $\{N, \lfloor 2^{RN} \rfloor, Q\}$, the expectation $\overline{\mathbf{B}_L(s, X)}$ of the number $\mathbf{B}_L(s, X)$ is

$$\overline{\mathbf{B}_L(s, X)} = \binom{t}{s} \Pr\left\{\boldsymbol{x}(S) \text{ is } s_L\text{-bad in } (N, R)\text{-code } X\right\}.$$

Therefore, the expectation of the error probability for almost disjunctive LD s_L-codes is

$$\mathcal{E}_L^{(N)}(s,R,Q) \triangleq \binom{t}{s}^{-1} \overline{\mathbf{B}_L(s,X)} = \Pr\{\boldsymbol{x}(S) \text{ is } s_L\text{-bad in } (N,R)\text{-code } X\}\,. \tag{27}$$

The evident *random coding upper bound* on the error probability (7) for almost disjunctive LD s_L-codes is formulated as the following inequality:

$$\epsilon_L(s,R,N) \triangleq \min_{X\,:\,t=\lfloor 2^{RN}\rfloor} \left\{ \frac{\mathbf{B}_L(s,X)}{\binom{t}{s}} \right\} \le \mathcal{E}_L^{(N)}(s,R,Q), \quad 0<Q<1. \tag{28}$$

The expectation $\mathcal{E}_L^{(N)}(s,R,Q)$ defined by (27) can be represented in the form

$$\mathcal{E}_L^{(N)}(s,R,Q) = \sum_{k=\lfloor QN\rfloor}^{\min\{N,s\lfloor QN\rfloor\}} \Pr\left\{ \begin{matrix} \boldsymbol{x}(S) \text{ is } s_L\text{-} \\ \text{-bad in } X \end{matrix} \Big/ \left|\bigvee_{i\in\mathcal{S}} \boldsymbol{x}(i)\right| = k \right\} \mathcal{P}^{(N)}(s,Q,k), \tag{29}$$

where we applied the total probability formula and introduced the notation

$$\mathcal{P}^{(N)}(s,Q,k) \triangleq \Pr\left\{ \left|\bigvee_{i\in\mathcal{S}} \boldsymbol{x}(i)\right| = k \right\}, \quad \lfloor QN\rfloor \le k \le \min\{N, s\lfloor QN\rfloor\}. \tag{30}$$

For the ensemble $\{N,t,Q\}$ and any k, $\lfloor QN\rfloor \le k \le \min\{N,s\lfloor QN\rfloor\}$, the conditional probability of event (2) is

$$\Pr\left\{ \bigvee_{i\in\mathcal{S}} \boldsymbol{x}(i) \succeq \bigvee_{j\in\mathcal{L}} \boldsymbol{x}(j) \Big/ \left|\bigvee_{i\in\mathcal{S}} \boldsymbol{x}(i)\right| = k \right\} = \left[\frac{\binom{k}{\lfloor QN\rfloor}}{\binom{N}{\lfloor QN\rfloor}} \right]^L. \tag{31}$$

In addition, with the help of the *type* (or *composition*) terminology:

$$\{n(\mathbf{a})\}, \quad \mathbf{a} \triangleq (a_1,a_2,\ldots,a_s) \in \{0,1\}^s, \quad 0 \le n(\mathbf{a}) \le N, \quad \sum_{\mathbf{a}} n(\mathbf{a}) = N,$$

the probability of event (30) in the ensemble $\{N,t,Q\}$ can be written as follows:

$$\mathcal{P}^{(N)}(s,Q,k) = \binom{N}{\lfloor QN\rfloor}^{-s} \cdot \sum_{(33)} \frac{N!}{\prod_{\mathbf{a}} n(\mathbf{a})!}, \quad \lfloor QN\rfloor \le k \le \min\{N,s\lfloor QN\rfloor\}, \tag{32}$$

and in the right-hand side of (32), the sum is taken over all types $\{n(\mathbf{a})\}$ provided that

$$n(\boldsymbol{0}) = N - k, \qquad \sum_{\mathbf{a}:\, a_i = 1} n(\mathbf{a}) = \lfloor QN \rfloor \quad \text{for any } i \in [s]. \tag{33}$$

Let the function

$$\mathcal{A}(s, Q, q) \triangleq \lim_{N \to \infty} \frac{-\log_2 \mathcal{P}^{(N)}(s, Q, \lfloor qN \rfloor)}{N}, \quad Q \le q \le \min\{1, sQ\}, \tag{34}$$

denotes the exponent of the logarithmic asymptotic behavior for the probability of event (30) calculated by (32)-(33).

Further, the representation (29), the conditional probability (31) and the standard union bound

$$\Pr\left\{\bigcup_i C_i \,/C\right\} \le \min\left\{1\,;\, \sum_i \Pr\{C_i/C\}\right\}$$

lead to the upper bound

$$\mathcal{E}_L^{(N)}(s, R, Q) \le \sum_{k = \lfloor QN \rfloor}^{\min\{N, s\lfloor QN \rfloor\}} \mathcal{P}^{(N)}(s, Q, k) \min\left\{1\,;\, \binom{t-s}{L} \left[\frac{\binom{k}{\lfloor QN \rfloor}}{\binom{N}{\lfloor QN \rfloor}}\right]^L\right\}, \tag{35}$$

where the code size $t \triangleq \lfloor 2^{RN} \rfloor$. Inequality (35) and the random coding bound (28) imply that the error probability exponent (8) satisfies the inequality

$$\mathbf{E}_L(s, R) \ge \underline{\mathbf{E}}_L(s, R) \triangleq \max_{0 < Q < 1} E_L(s, R, Q), \tag{36}$$

$$E_L(s, R, Q) \triangleq \min_{Q \le q \le \min\{1, sQ\}} \left\{\mathcal{A}(s, Q, q) + L \cdot [h(Q) - q \cdot h(Q/q) - R]^+\right\}. \tag{37}$$

Lemma 1. *Let $\lfloor QN \rfloor \le k \le \min\{N, s\lfloor QN \rfloor\}$. For the conditional probability in the right-hand side of* (29)*, the lower bound*

$$\Pr\left\{\begin{matrix}\boldsymbol{x}(S) \text{ is } s_L\text{-} \\ \text{-bad in } X\end{matrix} \Bigg/ \left|\bigvee_{i \in \mathcal{S}} \boldsymbol{x}(i)\right| = k\right\} \ge D(s, L) \cdot \min\left\{1\,;\, \binom{t-s}{L} \left[\frac{\binom{k}{\lfloor QN \rfloor}}{\binom{N}{\lfloor QN \rfloor}}\right]^L\right\}, \tag{38}$$

holds, where $D(s, L)$ is some constant.

Lemma 1 (its proof is omitted) establishes the asymptotic accuracy of the upper bound in (35), i.e., there exists

$$\lim_{N \to \infty} \frac{-\log_2 \mathcal{E}_L^{(N)}(s, R, Q)}{N} = E_L(s, R, Q), \quad R > 0.$$

where the function $E_L(s, R, Q)$, $R > 0$, defined by (37) can be interpreted as the *exponent of random coding bound on error probability for almost disjunctive LD s_L-codes* in the ensemble $\{N, \lfloor 2^{RN} \rfloor, Q\}$ of constant-weight codes.

The analytical properties of the function (34) are formulated below (without proof) as

Lemma 2. *The function $\mathcal{A}(s, Q, q)$ of the parameter q, $Q < q < \min\{1, sQ\}$, defined by* (34) *can be represented in the parametric form* (20)-(21). *In addition, the function $\mathcal{A}(s, Q, q)$ is $\cup$-convex, monotonically decreases in the interval $(Q, 1-(1-Q)^s)$, monotonically increases in the interval $(1-(1-Q)^s, \min\{1, sQ\})$ and its unique minimal value which is equal to* 0 *is attained at $q = 1-(1-Q)^s$, i.e.,*

$$\min_{Q<q<\min\{1,sQ\}} \mathcal{A}(s, Q, q) \;=\; \mathcal{A}(s, Q,\, 1-(1-Q)^s) = 0, \quad 0 < Q < 1.$$

Claim 1 is an evident consequence of Lemma 2.

References

[1] *Gallager R.G.* Information Theory and Reliable Communication. J. Wiley, New York, 1968.

[2] *Csiszar I.*, *Korner J.* Information Theory. Coding Theorems for Discrete Memoryless Systems. Akademiai Kiado, Budapest, 1981.

[3] *Kautz W.H.*, *Singleton R.C.* Nonrandom Binary Superimposed Codes// IEEE Trans. Inform. Theory. 1964. V. 10. n 4. P. 363-377.

[4] *D'yachkov A.G.*, *Rykov V.V.* Bounds on the Length of Disjunctive Codes // Problems of Information Transmission. 1982. V. 18. n 3. P. 166-171.

[5] *Erdos P.*, *Frankl P.*, *Furedi Z.* Families of Finite Sets in Which No Set Is Covered by the Union of 2 Others // J. Combin. Theory. Ser. A. 1982. V. 33. P. 158-166.

[6] *D'yachkov A.G.*, *Rykov V.V.* A Survey of Superimposed Code Theory // Problems of Control and Inform. Theory. 1983. V. 12. n 4. P. 229-242.

[7] *D'yachkov A.G.*, *Rykov V.V.*, *Rashad A.M.* Superimposed Distance Codes // Problems of Control and Inform. Theory. 1989. V. 18. n 4. P. 237-250.

[8] *D'yachkov A.G.*, *Macula A.J.*, *Rykov V.V.* New Constructions of Superimposed Codes // IEEE Trans. Inform. Theory. 2000. V. 46. n 1. P. 284-290.

[9] *D'yachkov A.G.*, *Macula A.J.*, *Rykov V.V.* New Applications and Results of Superimposed Code Theory Arising from the Potentialities of Molecular Biology // In the book "Numbers, Information and Complexity". P. 265-282, Kluwer Academic Publishers, 2000.

[10] *D'yachkov A.G.*, *Vilenkin P.A.*, *Macula A.J.*, *Torney D.C.*, *Yekhanin S.M.* New Results in the Theory of Superimposed Codes // Proc. Seventh Int. Workshop on Algebraic and Combinatorial Coding Theory. Bansko, Bulgaria. 2000. pp. 126-136.

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

[12] *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.

[13] *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.* Bounds on the Rate of Superimposed Codes// arXiv: 1401.0050 [cs.IT].

[14] *Malyutov M.B.* The Separating Property of Random Matrices // Mathematical Notes. 1978. V.23. n 1. P. 84-91.

[15] *D'yachkov A.G.* Error probability bounds for the symmetrical model of the design of screening experiments// Problems of Information Transmission. 1981. V. 17 n. 4. pp. 245-263.

[16] *Bassalygo L.A.*, *Rykov V.V.* Multiple-access hyperchannel // Problems of Information Transmission, 2013. vol. 49. no. 4, pp. 299-307.