Almost Cover-Free Codes and Designs

A.G. D'Yachkov, I.V. Vorobyev, N.A. Polyanskii, V.Yu. Shchukin

Almost Cover-Free Codes and Designs

A. G. D'yachkov, I.V. Vorobyev, N.A. Polyanskii and V.Yu. Shchukin

Lomonosov Moscow State University, Moscow, Russia
agd-msu@yandex.ru, vorobyev.i.v@yandex.ru, nikitapolyansky@gmail.com, vpike@mail.ru

Abstract. An s-subset of codewords of a binary code X is said to be (s,ℓ)-*bad* in X if the code X contains a subset of other ℓ codewords such that the conjunction of the ℓ codewords is covered by the disjunctive sum of the s codewords. Otherwise, the s-subset of codewords of X is called (s,ℓ)-*good* in X. A binary code X is said to be a cover-free (CF) (s,ℓ)-code if the code X does not contain (s,ℓ)-bad subsets. In this paper, we introduce a natural *probabilistic* generalization of CF (s,ℓ)-codes, namely: a binary code X is said to be an almost CF (s,ℓ)-code if the relative number of its (s,ℓ)-good s-subsets is close to 1. We develop a random coding method based on the ensemble of binary constant weight codes to obtain lower bounds on the capacity of such codes. Our main result shows that the capacity for almost CF (s,ℓ)-codes is essentially greater than the rate for ordinary CF (s,ℓ)-codes.

Keywords: Almost cover-free codes and designs, capacity, random coding bound

1 Statement of Problem and Results

1.1 Notations and Conventions

In what follows, the symbol $\triangleq$ denotes definitional equalities. For any positive integer n put $[n] \triangleq \{1,2,\ldots,n\}$. Let N and t be positive integers, $|A|$ – the size of set A. The standard symbol $\lfloor a \rfloor$ ($\lceil a \rceil$) will be used to denote the largest (least) integer $\le a$ ($\ge a$). Introduce a binary $N \times t$ matrix $X = \|x_i(j)\|$ having N rows $\boldsymbol{x}_i \triangleq \big(x_i(1), x_i(2), \ldots, x_i(t)\big)$, $i \in [N]$, and t columns $\boldsymbol{x}(j) \triangleq \big(x_1(j), x_2(j), \ldots, x_N(j)\big)$, $j \in [t]$. Any such matrix X is called a *binary code of length N and size $t = \lfloor 2^{RN} \rfloor$* (briefly, (N,R)-*code*), where a fixed parameter $R > 0$ is called the *rate* of code X. A column $\boldsymbol{x}(j) \in \{0,1\}^N$ is called a *j-th codeword*. 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 $x(j)$, $j \in [t]$.

For binary vectors $\mathbf{u} \triangleq (u_1,\ldots,u_N) \in \{0,1\}^N$ and $\mathbf{v} \triangleq (v_1,\ldots,v_N) \in \{0,1\}^N$, we will use the standard notations of component-wise *disjunction* $\mathbf{u} \bigvee \mathbf{v}$ and *conjunction* $\mathbf{u} \bigwedge \mathbf{v}$. We say that $\mathbf{u}$ is *covered* by $\mathbf{v}$ ($\mathbf{v} \succeq \mathbf{u}$) if $\mathbf{u} \bigvee \mathbf{v} = \mathbf{v}$.

1.2 Almost Cover-Free Codes

Let s and ℓ be positive integers such that $s+\ell \le t$ and $\mathcal{P}_s(t) \triangleq \{\mathcal{S} : \mathcal{S} \subset [t], |\mathcal{S}| = s\}$ is the collection of all s-subsets of the set $[t]$. Note that $|\mathcal{P}_s(t)| = \binom{t}{s}$.

Definition 1. Let $X = (\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(t))$ be an arbitrary binary code of length N and size t. A set $\mathcal{S} \in \mathcal{P}_s(t)$ is said to be (s,ℓ)-*bad* in X if there exists a set $\mathcal{L}$, $\mathcal{L} \subset [t] \setminus \mathcal{S}$ of size $|\mathcal{L}| = \ell$ such that

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

Otherwise, the set $\mathcal{S} \in \mathcal{P}_s(t)$ is called an (s,ℓ)-*good* set in X.

Let the symbol $\mathbf{B}(s,\ell,X)$ ($\mathbf{G}(s,\ell,X)$) denote the collection of all (s,ℓ)-bad ((s,ℓ)-good) sets $\mathcal{S} \in \mathcal{P}_s(t)$ in X and $|\mathbf{B}(s,\ell,X)|$ ($|\mathbf{G}(s,\ell,X)|$) is the size of the corresponding collection. Obviously, $|\mathbf{B}(s,\ell,X)| + |\mathbf{G}(s,\ell,X)| = \binom{t}{s}$.

Note an evident statement.

Proposition 1. *For $s \ge 2$ and $\ell \ge 1$, any $(s,\ell+1)$-good ((s,ℓ)-bad) set in X is (s,ℓ)-good ($(s,\ell+1)$-bad) set in X, i.e., the injections are true: $\mathbf{B}(s,\ell,X) \subset \mathbf{B}(s,\ell+1,X)$ and $\mathbf{G}(s,\ell+1,X) \subset \mathbf{G}(s,\ell,X)$.*

Definition 2. Let ϵ, $0 \le \epsilon \le 1$, be a fixed parameter. A code X is said to be an *almost cover-free* (s,ℓ)-*code* of *error probability* ϵ or, briefly, *CF* (s,ℓ,ϵ)-*code* if

$$\frac{|\mathbf{B}(s,\ell,X)|}{\binom{t}{s}} \le \epsilon \quad \Longleftrightarrow \quad |\mathbf{G}(s,\ell,X)| \ge (1-\epsilon)\binom{t}{s}. \tag{1}$$

Example 1. Consider 5×5 code X defined as: $\boldsymbol{x}(1) = (1,0,0,0,0)$, $\boldsymbol{x}(2) = (0,1,1,1,0)$, $\boldsymbol{x}(3) = (0,1,1,0,1)$, $\boldsymbol{x}(4) = (1,1,0,1,1)$, $\boldsymbol{x}(5) = (1,0,1,1,1)$. Then $\mathbf{G}(2,2,X) = \{\{1;2\}, \{1;3\}, \{1;4\}, \{1;5\}, \{2;3\}\}$ and X is a CF $(2,2,\frac{1}{2})$-code.

From Definition 2 and Proposition 1, it follows

Proposition 2. *Any CF $(s,\ell+1,\epsilon)$-code is a CF (s,ℓ,ϵ)-code.*

Monotonicity with respect to parameter s is provided by

Proposition 3. *If X is a CF (s,ℓ,ϵ)-code of size t and length N, then there exists a CF $(s-1,\ell,\epsilon)$-code X' of size $t-1$ and length N.*

Proof of Proposition 3. Let $\mathbf{B}(s,\ell,X,i) \triangleq \{\mathcal{S} : i \in S \in \mathbf{B}(s,\ell,X)\}$ denote the collection of all (s,ℓ)-bad sets $\mathcal{S}$ in X containing the element $i \in [t]$. Note that the cardinalities $|\mathbf{B}(s,\ell,X,i)|$, $0 \le |\mathbf{B}(s,\ell,X,i)| \le \binom{t-1}{s-1}$, $i \in [t]$, satisfy the equality:

$$\sum_{i=1}^{t} |\mathbf{B}(s,\ell,X,i)| = s \cdot |\mathbf{B}(s,\ell,X)| \le s\binom{t}{s}\epsilon,$$

where the last inequality follows from (1). This means that there exists $j \in [t]$, such that $|\mathbf{B}(s,\ell,X,j)| \le \binom{t-1}{s-1}\epsilon$. Then one can check that the code X' obtained from X by deleting the column $\boldsymbol{x}(j)$ is a CF $(s-1,\ell,\epsilon)$-code of size $t-1$ and length N. $\square$

For $\epsilon = 0$, the concept of CF (s,ℓ,ϵ)-code can be considered as a natural probabilistic generalization of the combinatorial concept of *CF* (s,ℓ)-*code* that

is defined in [1]-[2] as the *incidence matrix* of a *family of finite sets in which no intersection of ℓ sets is covered by the union of s others*. For the case $\ell = 1$, CF codes and their applications were introduced in [3]. For $\ell \geq 2$, CF (s, ℓ)-codes along with their applications to *key distribution patterns* were firstly suggested in [4].

Let $t(N, s, \ell)$ be the maximal size of CF (s, ℓ)-codes of length N and let $N(t, s, \ell)$ be the minimal length of CF (s, ℓ)-codes of size t. Then the number

$$R(s, \ell) \triangleq \varlimsup_{N \to \infty} \frac{\log_2 t(N, s, \ell)}{N} = \varlimsup_{t \to \infty} \frac{\log_2 t}{N(t, s, \ell)} \tag{2}$$

is called [2] the *rate* of CF (s, ℓ)-codes. In the recent papers [5, 6], one can find a detailed survey of the best known lower and upper bounds on the rate $R(s, \ell)$.

Using the conventional information-theoretic terminology accepted in the probabilistic coding theory [7], introduce

Definition 3. Let R, $R > 0$, be a fixed parameter. Taking into account inequality (1), define the *error* for almost CF (s, ℓ)-codes:

$$\epsilon(s, \ell, R, N) \triangleq \min_{X \,:\, t = \lfloor 2^{RN} \rfloor} \left\{ \frac{|\mathbf{B}(s, \ell, X)|}{\binom{t}{s}} \right\}, \tag{3}$$

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

$$\mathbf{E}(s, \ell, R) \triangleq \varlimsup_{N \to \infty} \frac{-\log_2 \epsilon(s, \ell, R, N)}{N}, \tag{4}$$

is said to be the error *exponent* for almost CF (s, ℓ)-codes, the number

$$C(s, \ell) \triangleq \sup\{R : \mathbf{E}(s, \ell, R) > 0\} \tag{5}$$

is said to be the *capacity* for almost CF (s, ℓ)-codes and rate $R(s, \ell)$ defined by (2) is called the *zero-error* capacity for almost CF (s, ℓ)-codes.

For the particular case $\ell = 1$, Definitions 1-3 were suggested in our paper [8], in which we introduce the concept of almost disjunctive list-decoding codes. The best presently known constructions of such codes were proposed in [9]. Bounds on the rate for these constructions were computed in the recent paper [10].

Definitions 1-3 and Proposition 1-3 lead to

Theorem 1. (Monotonicity properties.) *The following inequalities hold true*

$$R(s+1, \ell) \leq R(s, \ell) \leq R(s, \ell - 1), \quad C(s+1, \ell) \leq C(s, \ell) \leq C(s, \ell - 1),$$
$$\mathbf{E}(s+1, \ell, R) \leq \mathbf{E}(s, \ell, R) \leq \mathbf{E}(s, \ell - 1, R) \quad s \geq 1, \quad \ell \geq 2, \quad R > 0.$$

1.3 Almost Cover-Free Designs

By $\hat{\mathcal{P}}_s(\ell, t)$ denote the collection of *supersets* p, $\mathsf{p} \triangleq (P_1, P_2, \ldots, P_s)$, $P_i \subset \mathcal{P}_\ell(t)$, $i \in [s]$, where each p consists of s disjoint sets $P \subset [t]$ of size $|P| = \ell$, i.e.:

$$\hat{\mathcal{P}}_s(\ell, t) \triangleq \left\{ \mathsf{p} = (P_1, P_2, \ldots, P_s), \begin{array}{c} P_i \subset [t], \ |P_i| = \ell, \\ P_i \cap P_j = \varnothing \text{ for } i \neq j, \ i, j \in [s], \end{array} \right\}.$$

Obviously, the collection $\hat{\mathcal{P}}_s(\ell,t)$ has the cardinality

$$|\hat{\mathcal{P}}_s(\ell,t)| = \frac{1}{s!}\binom{t}{s\ell}\binom{s\ell}{(s-1)\ell}\cdots\binom{2\ell}{\ell}. \tag{6}$$

For a superset $\mathsf{p} \in \hat{\mathcal{P}}_s(\ell,t)$ and a code X, introduce the binary vector $\mathbf{r}(\mathsf{p},X)$:

$$\mathbf{r}(\mathsf{p},X) \triangleq \bigvee_{P\in\mathsf{p}}\bigwedge_{j\in P}\boldsymbol{x}(j), \quad \mathbf{r}(\mathsf{p},X) \triangleq \big(r_1,r_2,\ldots,r_N\big) \in \{0,1\}^N. \tag{7}$$

One can see that the i-th component of $\mathbf{r}(\mathsf{p},X)$ can be written in the form:

$$r_i = \begin{cases} 1, & \text{if there exists } P\in\mathsf{p} \text{ such that } x_i(j)=1 \text{ for all } j\in P,\\ 0, & \text{otherwise.}\end{cases} \tag{8}$$

Definition 4. Let $X=(\boldsymbol{x}(1),\boldsymbol{x}(2),\ldots,\boldsymbol{x}(t))$ be an arbitrary binary code of length N and size t. A superset p, $\mathsf{p}\in\hat{P}_s(\ell,t)$, is said to be an (s,ℓ)-*bad* superset in X, if there exists another superset $\mathsf{p}'\in\hat{P}_s(\ell,t)$, $\mathsf{p}\neq\mathsf{p}'$, such that $\mathbf{r}(\mathsf{p},X)=\mathbf{r}(\mathsf{p}',X)$. Otherwise, the superset p is said to be (s,ℓ)-*good* superset in X.

Let the symbol $\hat{\mathbf{B}}(s,\ell,X)$ ($\hat{\mathbf{G}}(s,\ell,X)$) denote the collection of all (s,ℓ)-bad ((s,ℓ)-good) supersets p, $\mathsf{p}\in\hat{P}_s(\ell,t)$, for the code X and $|\hat{\mathbf{B}}(s,\ell,X)|$ ($|\hat{\mathbf{G}}(s,\ell,X)|$) is the size of the corresponding collection. Obviously, $|\hat{\mathbf{B}}(s,\ell,X)|+|\hat{\mathbf{G}}(s,\ell,X)|=|\hat{\mathcal{P}}_s(\ell,t)|$.

Definition 5. Let ϵ, $0\le\epsilon\le 1$, be a fixed parameter. A code X is said to be an *almost cover-free* (s,ℓ)-*design* of *error probability* ϵ or, briefly, CF (s,ℓ,ϵ)-design if

$$\frac{|\hat{\mathbf{B}}(s,\ell,X)|}{|\hat{\mathcal{P}}_s(\ell,t)|} \le \epsilon \quad\Longleftrightarrow\quad |\hat{\mathbf{G}}(s,\ell,X)| \ge (1-\epsilon)\,|\hat{\mathcal{P}}_s(\ell,t)|. \tag{9}$$

Example 2. For the code X described in Example 1, the collection of $(2,2)$-bad supersets

$$\hat{\mathbf{B}}(s,\ell,X) = \{(\{1;2\},\{4;5\}),\,(\{1;3\},\{4;5\}),\,(\{1;4\},\{2;3\}),\,(\{1;5\},\{2;3\})\}.$$

It follows that X is a CF $(2,2,\frac{4}{15})$-design.

Definition 6. Let R, $R>0$, be a fixed parameter. Taking into account inequality (9) define the *error* for almost CF (s,ℓ)-designs:

$$\hat{\epsilon}(s,\ell,R,N) \triangleq \min_{X\,:\,t=\lceil 2^{RN}\rceil}\left\{\frac{|\hat{\mathbf{B}}(s,\ell,X)|}{|\hat{\mathcal{P}}_s(\ell,t)|}\right\}, \tag{10}$$

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

$$\hat{\mathbf{E}}(s,\ell,R) \triangleq \varlimsup_{N\to\infty}\frac{-\log_2\hat{\epsilon}(s,\ell,R,N)}{N}, \tag{11}$$

is said to be the error *exponent* for almost CF (s,ℓ)-designs, the number

$$\hat{C}(s,\ell) \triangleq \sup\{R:\ \hat{\mathbf{E}}(s,\ell,R) > 0\}$$

is called the *capacity* for almost CF (s,ℓ)-designs.

For the particular case $\ell = 1$, Definitions 4-6 were already introduced in [11] to describe the model called *planning screening experiments*. In [11], it was proved that the capacity of almost CF $(s,1)$-designs $\hat{C}(s,1) = 1/s$. One can see that Definitions 4-6 represent a natural generalization of almost CF $(s,1)$-designs. We conjecture that the capacity $\hat{C}(s,\ell) = 1/(s\,\ell)$.

In Section 2, we establish

Theorem 2. *The capacities and the error exponents satisfy the inequality*

$$C(s,\ell) \le \hat{C}(s,\ell) \le 1/(s\,\ell), \quad \mathbf{E}(s,\ell,R) \le \hat{\mathbf{E}}(s,\ell,R).$$

However, in spite of the greater capacity, *using* of CF (s,ℓ,ϵ)-designs for the superset identification problem $\mathsf{p} \in \hat{P}_s(\ell,t)$ is *practically unacceptable*, since it requires much greater complexity, which is evidently equal to the complexity of exhaustive search $|\hat{\mathcal{P}}_s(\ell,t)| \sim t^{s\,\ell}$. It will be shown in Section 1.5 that CF (s,ℓ,ϵ)-codes are efficient CF (s,ℓ,ϵ)-designs and for such codes the algorithm of identification supersets $\mathsf{p} \in \hat{\mathcal{P}}_s(\ell,t)$, is essentially faster than the trivial one, and its complexity is proportional to t^{ℓ}.

1.4 Lower Bounds on $R(s,\ell)$, $C(s,\ell)$

The best presently known upper and lower bounds on the rate $R(s,\ell)$ of cover-free (s,ℓ)-codes were presented in [5, 6]. If $\ell \ge 1$ is fixed and $s \to \infty$, then these bounds have the following asymptotic form:

$$\frac{(\ell+1)^{\ell+1}}{e^{\ell+1}}\,\frac{\log_2 s}{s^{\ell+1}}\,(1+o(1)) \le R(s,\ell) \le \frac{(\ell+1)^{\ell+1}}{2e^{\ell-1}}\,\frac{\log_2 s}{s^{\ell+1}}\,(1+o(1)). \qquad (12)$$

In the present paper, we suggest a modification of the random coding method developed in [5] and [8], which permits us to obtain a lower bound on the capacity $C(s,\ell)$. Let $[x]^+$ and $h(x)$ denote the positive part function and the binary entropy function respectively. In Section 3, we prove

Theorem 3. (Random coding lower bound $\underline{C}(s,\ell)$).

Claim 1. *For $\ell \ge 2$, the capacity for almost CF codes satisfies inequality*

$$C(s,\ell) \ge \underline{C}(s,\ell) \triangleq \frac{1}{\ell}\max_{0\le Q\le 1}\mathcal{D}(\ell,Q,\hat{q}), \qquad (13)$$

where the function $\mathcal{D}(\ell,Q,\hat{q})$ is defined in the parametric form

$$\mathcal{D}(\ell,Q,\hat{q}) \triangleq (1-Q)\ell\,\log_2 z - (1-\hat{q})\log_2[1-(1-z)^{\ell}] + \qquad (14)$$

$$+\ell\left(\frac{(1-Q)}{z}(1-z) - \left(\frac{(1-Q)}{z} - \hat{q}\right)(1-z)^{\ell}\right)\log_2[1-z] + \ell h(Q),$$

and parameters z and $\hat{q}$ are uniquely determined by the following equations

$$Q = \frac{(1-z)(1-(1-z)^{\ell}) - (1-\hat{q})z(1-z)^{\ell}}{1-(1-z)^{\ell}}, \qquad \hat{q} = 1-(1-Q)^{s}. \tag{15}$$

Claim 2. *For $\ell \geq 2$ and $s \to \infty$, the lower asymptotic bound on $C(s,\ell)$ is*

$$C(s,\ell) \;\geq\; \frac{\log_2 e}{s^{\ell}} \cdot \frac{\ell^{\ell-1}}{e^{\ell}}(1+o(1)). \tag{16}$$

1.5 Boolean Model for Nonadaptive Search of Supersets

Definition 7. [2] A binary $N \times t$ matrix X is called a *cover-free (s,ℓ)-design* or, briefly, *CF (s,ℓ)-design* if for any $\mathsf{p}', \mathsf{p}'' \in \hat{\mathcal{P}}_s(\ell,t)$, $\mathsf{p}' \neq \mathsf{p}''$, the vectors $\mathbf{r}(\mathsf{p}', X) \neq \mathbf{r}(\mathsf{p}'', X)$.

Suppose a set of t samples is given. We identify it with the set $[t]$. In the present paper we consider a generalization of *the boolean search model for sets* [3] which is called *the boolean search model for supersets* [2]. Assume that a *positive superset* $\mathsf{p} \in \hat{\mathcal{P}}_s(\ell,t)$ is fixed. Our aim is to detect it using the minimal number of group tests, where *each test checks whether a testing group contains at least one set $P \in \mathsf{p}$ or not*. Now assume that we use N tests. They can be encoded by a code $X = \|x_i(j)\|$. A column $\boldsymbol{x}(j)$ corresponds to the j-th sample; a row $\boldsymbol{x}_i$ corresponds to the i-th test. We put $x_i(j) \triangleq 1$ iff the j-th sample is included into the i-th testing group; otherwise, $x_i(j) \triangleq 0$. The *outcomes* (8) of all N tests form the binary vector $\mathbf{r}(\mathsf{p}, X)$ (7), where $\mathsf{p} \in \hat{\mathcal{P}}_s(\ell,t)$ is the (*unknown*) positive superset. Thus, the code X should be designed in such a way that we should be able to detect a superset p given the vector $\mathbf{r}(\mathsf{p}, X)$. Obviously, it is possible if and only if X is a CF (s,ℓ)-design. Note that we deal with the *nonadaptive* search model arised from the needs of molecular biology and firstly suggested in [12].

Let X be a binary $N \times t$ matrix and $\mathsf{p}^{(\mathrm{un})} \in \hat{\mathcal{P}}_s(\ell,t)$ be an *unknown* superset. Any fixed set $P' \subset [t]$, $|P'| \leq \ell$, is called *acceptable* for the *known* vector $\mathbf{r}^{(\mathrm{kn})} \triangleq \mathbf{r}(\mathsf{p}^{(\mathrm{un})}, X)$ if the conjunction $\bigwedge\limits_{j \in P'} \boldsymbol{x}(j)$ is covered by $\mathbf{r}^{(\mathrm{kn})}$. In the given model, an effective decoding algorithm is based on the following

Proposition 4. [2] If X is an CF (s,ℓ)-code, then any superset $\mathsf{p} \in \hat{\mathcal{P}}_s(\ell,t)$ is composed of all acceptable sets for $\mathbf{r}^{(\mathrm{kn})}$. This means that the decoding complexity is proportional to $\binom{t}{\ell} \sim t^{\ell}$ and doesn't depend on s.

Note that in the general case of CF (s,ℓ)-design and the trivial decoding algorithm, we need to check all possible supersets $\mathsf{p} \in \hat{\mathcal{P}}_s(\ell,t)$. If s and ℓ are fixed and $t \to \infty$, then the number of such comparisons (decoding complexity) is proportional to $|\hat{\mathcal{P}}_s(\ell,t)| \sim t^{s\ell}$. Thus, CF (s,ℓ)-codes form a class of CF (s,ℓ)-designs for which the decoding algorithm based on Proposition 4 is strongly better than the trivial one.

Let $\ell \geq 1$ be fixed and $s \to \infty$. Taking into account (12), we conclude that for sufficiently large t the *use of CF (s,ℓ)-codes* gives the bounds:

$$\frac{\log_2 s}{s^{\ell+1}} \cdot \frac{(\ell+1)^{\ell+1}}{2e^{\ell-1}}\ (1+o(1)) \geq \log_2 t/N \ \geq\ \frac{\log_2 s}{s^{\ell+1}} \cdot \frac{(\ell+1)^{\ell+1}}{e^{\ell+1}}\,(1+o(1)).$$

The capacity $C(s,\ell)$ can be interpreted as the theoretical tightest upper bound on the information rate $\log_2 t/N$ of CF (s,ℓ,ϵ)-codes with error probability $\epsilon \to 0$. Therefore, the bound (16) means that for $\ell \geq 2$, $s \to \infty$ and sufficiently large t, *using of CF (s,ℓ,ϵ)-codes* guarantees the inequality:

$$\log_2 t/N \ \geq\ \frac{\log_2 e}{s^{\ell}} \cdot \frac{\ell^{\ell-1}}{e^{\ell}}(1+o(1)).$$

2 Proof of Theorem 2.

For any superset $\mathsf{p} \in \hat{P}_s(\ell,t)$, $\mathsf{p} = \{P_1, P_2, \ldots, P_s\}$, define the set $T(\mathsf{p})$:

$$T(\mathsf{p}) \triangleq \big\{\mathcal{S} \in \mathcal{P}_s(t):\quad S = \{a_1, a_2, \ldots, a_s\},\quad a_i \in P_i,\ P_i \in \mathsf{p},\ i \in [s]\big\}.$$

Observe that if all sets $\mathcal{S} \in T(\mathsf{p})$ are (s,ℓ)-good in X, then the superset p is also a (s,ℓ)-good superset in X.

Assume that a code X is a CF (s,ℓ,ϵ)-code. It means that the number (1) of bad (s,ℓ)-sets doesn't exceed $\epsilon \cdot \binom{t}{s}$. Given a bad (s,ℓ)-set $B \in \mathcal{P}_s(t)$ for the code X, one can check that the number of $\mathsf{p} \in \hat{P}_s(\ell,t)$ such that $B \in T(\mathsf{p})$ is at most $\binom{t-s}{s(\ell-1)}\binom{s(\ell-1)}{(s-1)(\ell-1)}\cdots\binom{2(\ell-1)}{\ell-1}$. This implies that the number of bad (s,ℓ)-supersets is at most $\epsilon \cdot \binom{t}{s}\binom{t-s}{s(\ell-1)}\binom{s(\ell-1)}{(s-1)(\ell-1)}\cdots\binom{2(\ell-1)}{\ell-1}$ or $\epsilon \cdot \ell^s \cdot |\hat{\mathcal{P}}_s(\ell,t)|$, where $|\hat{\mathcal{P}}_s(\ell,t)|$ is computed (6). Thus, X is also a CF $(s,\ell,\epsilon\cdot\ell^s)$-design. In other words, we proved the relations $C(s,\ell) \leq \hat{C}(s,\ell)$ and $\mathbf{E}(s,\ell,R) \leq \hat{\mathbf{E}}(s,\ell,R)$.

Now, fix $R > 0$ and $\epsilon > 0$ and suppose that the code X is a CF (s,ℓ,ϵ)-design of length N and size $t \triangleq \lfloor 2^{RN} \rfloor$. Observe that for any two different good (see Def. 4) supersets p, $\mathsf{p}' \in \hat{\mathbf{G}}(s,\ell,X)$, $\mathsf{p} \neq \mathsf{p}'$, two vectors $\mathbf{r}(\mathsf{p},X)$ and $\mathbf{r}(\mathsf{p}',X)$ defined by (7) are distinct, i.e., $\mathbf{r}(\mathsf{p},X) \neq \mathbf{r}(\mathsf{p},X)$. From the definition (9) of CF (s,ℓ,ϵ)-design, we get

$$(1-\epsilon)\cdot|\hat{\mathcal{P}}_s(\ell,t)| = (1-\epsilon)\cdot\frac{1}{s!}\binom{t}{s\ell}\binom{s\ell}{(s-1)\ell}\cdots\binom{2\ell}{\ell} \ \leq\ 2^N, \quad t = \lfloor 2^{RN} \rfloor. \quad (17)$$

The comparison of the left and right-hand sides of (17) leads to the bound

$$\hat{\epsilon}(s,\ell,R,N) \geq\ 1 - 2^N \cdot \Big(|\hat{\mathcal{P}}_s(\ell,t)|\Big)^{-1} \ =\ 1 - 2^{-N[(s\ell\cdot R-1)+o(1)]}, \quad N \to \infty.$$

This inequality means that the condition $R < 1/(s\ell)$ is necessary for $\hat{\mathbf{E}}(s,\ell,R) > 0$. It follows that $\hat{C}(s,\ell) \leq \frac{1}{s\ell}$. $\square$

3 Proof of Theorem 3

Here we present a sketch of the proof only. The preprint containing a full version of the given article is available at: arXiv: 1410.8566.

Proof of Claim 1. For a code X, the number $|\mathbf{B}(s,\ell,X)|$ of (s,ℓ)-bad sets in the code X can be represented in the form:

$$|\mathbf{B}(s,\ell,X)| \triangleq \sum_{\mathcal{S}\in\mathcal{P}_s(t)} \psi(X,\mathcal{S}), \qquad \psi(X,\mathcal{S}) \triangleq \begin{cases} 1 & \text{if the set } \mathcal{S}\in\mathbf{B}(s,\ell,X),\\ 0 & \text{otherwise.}\end{cases} \tag{18}$$

Let Q, $0<Q<1$, and R, $0<R<1$, be fixed parameters. Define the ensemble $\{N,t,Q\}$ of binary $(N\times t)$-matrices $X=(\boldsymbol{x}(1),\boldsymbol{x}(2),\dots\boldsymbol{x}(t))$, where columns $\boldsymbol{x}(i)$, $i\in[t]$, $t\triangleq\lfloor 2^{RN}\rfloor$, are chosen independently and equiprobably from the set consisting of $\binom{N}{\lfloor QN\rfloor}$ columns of the fixed weight $\lfloor QN\rfloor$. Fix two subsets $\mathcal{S},\mathcal{L}\subset[t]$ such that $|\mathcal{S}|=s$, $|\mathcal{L}|=\ell$ and $\mathcal{S}\cap\mathcal{L}=\varnothing$. From (18) it follows that for $\{N,t,Q\}$, the expectation $\overline{|\mathbf{B}(s,\ell,X)|}$ of the number $|\mathbf{B}(s,\ell,X)|$ is

$$\overline{|\mathbf{B}(s,\ell,X)|} \;=\; |\mathcal{P}_s(t)|\,\Pr\{\mathcal{S}\in\mathbf{B}(s,\ell,X)\}\,.$$

Thus, the expectation of the error probability for almost CF (s,ℓ)-codes is

$$\mathcal{E}^{(N)}(s,\ell,R,Q) \triangleq |\mathcal{P}_s(t)|^{-1}\,\overline{|\mathbf{B}(s,\ell,X)|} = \Pr\{\mathcal{S}\in\mathbf{B}(s,\ell,X)\}\,, \tag{19}$$

where $t=\lfloor 2^{RN}\rfloor$. The evident *random coding upper bound* on the error probability (3) for cover-free (s,ℓ)-codes is formulated as the following inequality:

$$\epsilon(s,\ell,R,N) \triangleq \min_{X\,:\,t=\lfloor 2^{RN}\rfloor}\left\{\frac{|\mathbf{B}(s,\ell,X)|}{|\mathcal{P}_s(t)|}\right\} \le \mathcal{E}^{(N)}(s,\ell,R,Q), \quad 0<Q<1. \tag{20}$$

The expectation $\mathcal{E}^{(N)}(s,\ell,R,Q)$ defined by (19) can be represented as follows

$$\mathcal{E}^{(N)}(s,\ell,R,Q) \;=\; \sum_{k=\lfloor QN\rfloor}^{\min\{N,\,s\lfloor QN\rfloor\}} \Pr\left\{\mathcal{S}\in\mathbf{B}(s,\ell,X)\,\Big/\,\left|\bigvee_{i\in\mathcal{S}}\boldsymbol{x}(i)\right|=k\right\}\times$$

$$\times\mathcal{P}_2^{(N)}(s,Q,k) \le \sum_{k=\lfloor QN\rfloor}^{\min\{N,\,s\lfloor QN\rfloor\}} \mathcal{P}_2^{(N)}(s,Q,k)\cdot\min\left\{1;\,\binom{t-s}{\ell}\mathcal{P}_1^{(N)}(\ell,Q,k)\right\}, \tag{21}$$

where we apply the total probability formula and the standard union bound for the conditional probability $\Pr\left\{\bigcup_i C_i\;/C\right\} \le \min\left\{1\,;\,\sum_i \Pr\{C_i/C\}\right\}$ and introduce the notations

$$\mathcal{P}_1^{(N)}(\ell,Q,k) \triangleq \Pr\left\{\bigvee_{i\in\mathcal{S}}\boldsymbol{x}(i)\succeq\bigwedge_{j\in\mathcal{L}}\boldsymbol{x}(j)\;\Big/\;\left|\bigvee_{i\in\mathcal{S}}\boldsymbol{x}(i)\right|=k\right\} \tag{22}$$

and

$$\mathcal{P}_2^{(N)}(s,Q,k) \triangleq \Pr\left\{\left|\bigvee_{i\in\mathcal{S}} \boldsymbol{x}(i)\right| = k\right\}, \quad \lfloor QN\rfloor \le k \le \min\{N, s\lfloor QN\rfloor\}. \quad (23)$$

Let $k \triangleq \lfloor qN\rfloor$ and the functions

$$\mathcal{D}(\ell,Q,q) \triangleq \lim_{N\to\infty} \frac{-\log_2\left[\mathcal{P}_1^{(N)}(\ell,Q,k)\right]}{N}, \quad (24)$$

$$\mathcal{A}(s,Q,q) \triangleq \lim_{N\to\infty} \frac{-\log_2\left[\mathcal{P}_2^{(N)}(s,Q,k)\right]}{N} \quad (25)$$

denote the exponents of the logarithmic asymptotic behavior for the probability of events (22) and (23) for $\{N,t,Q\}$ respectively. Define $\hat{q} \triangleq 1-(1-Q)^s$.

Lemma 1. *The function $\mathcal{A}(s,Q,q)$ of the parameter q, $Q<q<\min\{1,sQ\}$, defined by (25) can be represented in the parametric form*

$$\mathcal{A}(s,Q,q) \triangleq (1-q)\log_2(1-q) + q\log_2\left[\frac{Qy^s}{1-y}\right] + sQ\log_2\frac{1-y}{y} + sh(Q), \quad (26)$$

$$q = Q\frac{1-y^s}{1-y}, \qquad 0<y<1. \quad (27)$$

In addition, $\mathcal{A}(s,Q,q)$ as a function of q attains its unique minimal value which is equal to 0 at $q=\hat{q}\triangleq 1-(1-Q)^s$.

Lemma 2. *For $\ell \ge 2$, the value of the function $\mathcal{D}(\ell,Q,q)$ defined by (24) at point $q=\hat{q}$ is equal to*

$$\mathcal{D}(\ell,Q,\hat{q}) = (1-Q)\,\ell\,\log_2 z - (1-\hat{q})\log_2[1-(1-z)^\ell] +$$

$$+\ell\left(\frac{(1-Q)}{z}(1-z) - \left(\frac{(1-Q)}{z}-\hat{q}\right)(1-z)^\ell\right)\log_2[1-z] + \ell h(Q),$$

where z is uniquely determined by the following equation

$$Q = \frac{(1-z)(1-(1-z)^\ell) - (1-\hat{q})z(1-z)^\ell}{1-(1-z)^\ell}.$$

The inequality (21) and the random coding bound (20) imply that the error probability exponent (11) satisfies the inequality

$$\mathbf{E}(s,\ell,R) \ge \underline{\mathbf{E}}(s,\ell,R) \triangleq \max_{0\le Q\le 1} E(s,\ell,R,Q), \quad (28)$$

$$E(s,\ell,R,Q) \triangleq \min_{Q<q<\min\{1,sQ\}} \left\{\mathcal{A}(s,Q,q) + [\mathcal{D}(\ell,Q,q) - \ell R]^+\right\}. \quad (29)$$

Lemma 1 states that $\mathcal{A}(s,Q,q) > 0$ if $q \neq \hat{q}$. In particular, the condition $q \neq \hat{q}$ implies $E(s,\ell,R,Q) > 0$. Therefore, if $\ell R < \mathcal{D}(\ell,Q,\hat{q})$ then $E(s,\ell,R,Q) > 0$, what, in turn, means (see (5) and (28)) that

$$C(s,\ell) \geq \underline{C}(s,\ell) \triangleq \frac{1}{\ell} \max_{0 \leq Q \leq 1} \mathcal{D}(\ell,Q,\hat{q}), \quad \text{where } \hat{q} = 1-(1-Q)^s.$$

Thus, the lower bound (13) is established. □

Proof of Claim 2. Let $\ell \geq 2$ and $s \to \infty$. Substituting $z = s/(s+\ell)$ in (13)-(15) and computing the asymptotic behaviour complete the proof. □

References

1. Erdos P., Frankl P., Furedi Z., "Families of Finite Sets in Which No Set Is Covered by the Union of 2 Others". *Journal of Combinatorial Theory*, Series A, vol. 33, pp. 158-166, 1982.
2. D'yachkov A.G., Vilenkin P., Macula A., Torney D., "Families of Finite Sets in Which No Intersection of ℓ Sets Is Covered by the Union of s Others", *Journal of Combinatorial Theory*, Series A, vol. 99. pp. 195-218, 2002.
3. Kautz W.H., Singleton R.C., "Nonrandom Binary Superimposed Codes", *IEEE Trans. Inform. Theory*, vol. 10, no. 4, pp. 363-377, 1964.
4. Mitchell C.J, Piper F.C., "Key storage in Secure Networks", *Discrete Applied Mathematics*, vol. 21, pp. 215-228, 1988.
5. 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.
6. 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.
7. Csiszar I., Korner J. "Information Theory. Coding Theorems for Discrete Memoryless Systems", *Akademiai Kiado*, Budapest, 1981.
8. D'yachkov A.G., Vorobyev I.V., Polyanskii N.A., Shchukin V.Yu., "Almost Disjunctive List-Decoding Codes", *Proc. of International Conference on Algebraic and Combinatorial Coding Theory (ACCT)*, Svetlogorsk (Kaliningrad region), Russia, pp. 115-126, Sep. 7-13, 2014.
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, *Kluwer Academic Publishers*, pp. 265-282, 2000.
10. Bassalygo L.A., Rykov V.V., "Multiple-access hyperchannel", *Problems of Information Transmission*, vol. 49, no. 4, pp. 299-307, 2013.
11. Malyutov M.B., "The Separating Property of Random Matrices", *Mathematical Notes*, vol.23, no. 1, pp. 84-91, 1978.
12. Torney D.C., "Sets Pooling Designs", *Annals of Combinatorics*, vol. 3, pp. 95-101, 1999.