New Results in the Theory of Superimposed Codes: Part II

A. D'yachkov[1], A. Macula[2], D. Torney[3], P. Vilenkin[1], S. Yekhanin[1]

Abstract — In the second part of the paper, we work out a constructive method for (s,ℓ)-codes [8] based on concatenated codes and MDS-codes [2, 3]. The method is a generalization of the constructive method for $(s,1)$-codes [1, 6]. In addition, we discuss the constructions of the list-decoding superimposed codes [4, 7], identified by a family of finite sets in which no union of L sets is covered by the union of s others.

1 Notations and Definitions

We use notations and definitions from "Part I" of the present paper [8]. Let t and N be positive integers, and $\mathcal{C}$ be a set of t binary codewords of length N:

$$\mathcal{C} \triangleq \{\mathbf{x}(1),\ldots,\mathbf{x}(t)\}, \quad \mathbf{x}(j) \triangleq \big(x_1(j),\ldots,x_N(j)\big) \in \{0,1\}^N. \qquad (1)$$

For any subset $\tau \subset [t]$ consider the disjunction and conjunction

$$\mathrm{V}(\tau) \triangleq \bigvee_{j\in\tau} \mathbf{x}(j), \quad \Lambda(\tau) \triangleq \bigwedge_{j\in\tau} \mathbf{x}(j). \qquad (2)$$

For positive integers s and ℓ, such that $t \geq s+\ell$, put

$$\pi(s,\ell,t) \triangleq \big\{(\mathcal{S},\mathcal{L}) \,:\, \mathcal{S},\mathcal{L}\subset[t],\, |\mathcal{S}|=s,\, |\mathcal{L}|=\ell,\, \mathcal{S}\cap\mathcal{L}=\varnothing\big\}. \qquad (3)$$

Definition 1 [8]. A *superimposed binary* (s,ℓ)*-code* of *length* N and *size* t is a set $\mathcal{C}$ (1), such that for any pair $(\mathcal{S},\mathcal{L}) \in \pi(s,\ell,t)$ the vector $\Lambda(\mathcal{L})$ is not covered by $\mathrm{V}(\mathcal{S})$.

[1] A. D'yachkov, P. Vilenkin and S. Yekhanin are with the Department of Probability Theory, Faculty of Mechanics & Mathematics, Moscow State University, Russia (dyachkov@mech.math.msu.su, paul@vilenkin.dnttm.ru, gamov@cityline.ru). Their work is supported by the Russian Foundation of Basic Research, grant 98–01–00241.

[2] A. Macula is with the Department of Mathematics, State University of New York, College at Geneseo, USA (macula@geneseo.edu). His work is supported by NSF Grant DMS–9973252.

[3] D. Torney is with the Theoretical Division T10, Los Alamos National Laboratories, USA (dct@lanl.gov). His work is supported by the US Department of Energy.

Definition 2. A *superimposed list-decoding code* of *strength* s and *list-size* L is a set $\mathcal{C}$ (1), such that for any pair $(\mathcal{S}, \mathcal{L}) \in \pi(s, L, t)$ the vector $\mathrm{V}(\mathcal{L})$ is not covered by $\mathrm{V}(\mathcal{S})$.

If $|\tau| = 1$, then $\mathrm{V}(\tau) = \Lambda(\tau)$. For this reason, for $\ell = L = 1$ definitions 1 and 2 are equivalent, and a set $\mathcal{C}$ in this case is called a *binary superimposed code* of *strength* s (or, briefly, s-*code*).

A codeword $\mathbf{x}(j)$ can be interpreted as a subset of the set $[N]$. Then $\mathrm{V}(\tau)$ is the union, and $\Lambda(\tau)$ is the intersection of corresponding sets. Taking this into account, a superimposed s-code can be identified by a family of sets in which no set is covered by a union of s others; a superimposed (s, ℓ)-code is identified by a family of sets in which no intersection of ℓ sets is covered by a union of s others; and a superimposed s-code with list-size L is identified by a family of sets in which no union of L sets is covered by a union of s others.

The applications of s-codes and (s, ℓ)-codes to the problem of identifying positive elements and positive sets in the group testing model are discussed in "Part I" of the present paper [8, Sec. 2]. Superimposed s-codes with list-size L can also be used in this model as follows: if $\mathsf{p} \subset [t]$ is a set of positive elements, $|\mathsf{p}| \le s$, and testing groups form an s-code with list-size L, then given the test results one can construct a set $\mathsf{p}' \subset [t]$, such that $\mathsf{p} \subseteq \mathsf{p}'$ and $|\mathsf{p}' \backslash \mathsf{p}| \le L - 1$. If $L = 1$, then one can decode an unknown set p exactly.

2 Constructions of (s, ℓ)-codes

Trivial construction. Let $\mathcal{C}'$ be an $(s, 1)$-code of length N' and size t. Put $N \triangleq \binom{N'}{\ell}$ and let $\sigma_1, \ldots, \sigma_N$ be all ℓ-subsets of the set $[N']$. Construct a new code $\mathcal{C} \triangleq \{\mathbf{x}(1), \ldots, \mathbf{x}(t)\}$ of length N, for which

$$x_i(j) \triangleq \bigvee_{m \in \sigma_i} x'_m(j), \quad i \in [N], \quad j \in [t].$$

Then $\mathcal{C}$ is an (s, ℓ)-code. This yields the bound [8, (3)].

Concatenated construction. Consider an integer $q \ge 2$ and a set $\mathcal{C}$ (1), in which elements $x_i(j)$ are taken from the q-ary alphabet $[q] = \{1, \ldots, q\}$.

Definition 3. A q-ary set $\mathcal{C}$ defined above is called a *superimposed* q-*ary* (s, ℓ)-*code*, if for any pair $(\mathcal{S}, \mathcal{L}) \in \pi(s, \ell, t)$ there exists a coordinate

$i \in [n]$ for which the *coordinate sets*

$$\mathcal{L}_i \triangleq \{x_i(j) : j \in \mathcal{L}\} \subseteq [q] \quad \text{and} \quad \mathcal{S}_i \triangleq \{x_i(j') : j' \in \mathcal{S}\} \subseteq [q]$$

are disjoint, i.e., $\mathcal{S}_i \cap \mathcal{L}_i = \varnothing$. Integers t and n are called the *size* and *length* of code $\mathcal{C}$, respectively.

Proposition 1. (Concatenated construction) *Let $s \geq 1$, $\ell \geq 1$, $t \geq s + \ell$ and $q \geq s + \ell$ be integers. Assume that there exists a q-ary (s, ℓ)-code $\mathcal{C}^{(q)} = \|x_i^{(q)}(j)\|$ of size $t^{(q)}$ and length $n^{(q)}$ and an (s, ℓ)-code $\mathcal{C}' = \|x_i'(j)\|$ of size $t' \geq q$ and length n'. Then there exists a superimposed (s, ℓ)-code $\mathcal{C}$ of size $t = t^{(q)}$ and length $N = n^{(q)} n'$.*

Proof. The code $\mathcal{C}$ is constructed by the concatenation of codes $\mathcal{C}^{(q)}$ and $\mathcal{C}'$, i.e., each q-ary symbol $\theta \in [q]$ in the code $\mathcal{C}^{(q)}$ is replaced with the codeword $\mathbf{x}'(\theta)$ from the code $\mathcal{C}'$. The j-th codeword of the new code $\mathcal{C}$ has the form

$$\mathbf{x}(j) \triangleq \left(\mathbf{x}'\left(x_1^{(q)}(j)\right), \ldots, \mathbf{x}'\left(x_{n'}^{(q)}(j)\right)\right).$$

One can easily prove that this code $\mathcal{C}$ is really an (s, ℓ)-code.

Proposition 2. *Let $s = \ell = 2$. Then the minimum length $N(t, 2, 2)$ for $4 \leq t \leq 8$ has the form*

$$N(4, 2, 2) = \tbinom{4}{2} = 6, \quad N(5, 2, 2) = \tbinom{5}{2} = 10,$$
$$N(6, 2, 2) = N(7, 2, 2) = N(8, 2, 2) = 14.$$

Proof. For $t = s + \ell = 4$ the optimal (s, ℓ)-code is trivial [8, Prop. 2]. For $t = 5, 6$ we used a computer exhaustive search: for $t = 5$ the optimal $(2, 2)$-code is trivial, and for $t = 6$ the optimal code has length $N = 14$ (the trivial length for this case is $\binom{6}{2} = 15$).

Consider the following 3×8 quaternary matrix

$$C^{(4)} = \begin{pmatrix} 4 & 2 & 3 & 1 & 2 & 4 & 1 & 3 \\ 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \\ 2 & 4 & 1 & 3 & 2 & 4 & 1 & 3 \end{pmatrix}.$$

One can check that the columns of $C^{(4)}$ form a superimposed quaternary $(2, 2)$-code of size 8 and length 3. The concatenation of this code with the trivial $(2, 2)$-code of size 4 and length $\binom{4}{2} = 6$ leads to the binary $(2, 2)$-code of size $t = 8$ and length $N = 6 \cdot 3 = 18$. Examining this code, one can see that there is a pair of rows, which are repeated three times

in the given code. Obviously, we can remove two copies of each row, and get the binary $(2,2)$-code of size $t = 8$ and length $N = 14$. Since $N(6,2,2) = 14$, we have that $N(8,2,2) = N(7,2,2) = 14$.

Definition 4. The *Maximum Distance Separable code* (*MDS-code*) with parameters (q,k,n) is a q-ary code of size $t = q^k$, length n and the Hamming distance $d = n - k + 1$ [3].

Proposition 3. *If $q^k \geq s + \ell$ and $n \geq s\ell(k-1) + 1$, then any MDS-code with parameters (q,k,n) is a superimposed q-ary (s,ℓ)-code.*

For any integer $\lambda \geq 1$ and a prime power $q \geq \lambda$ there exists an MDS-code with parameters $(q, \lambda+1, q+1)$ (*Reed–Solomon code*). The concatenation of this code with the optimal binary superimposed code of size q leads to the following

Proposition 4. *Let $s \geq 1$, $\ell \geq 1$ and $\lambda \geq 1$ be integers and $q \geq s\ell\lambda$ be a prime power. Then*

$$N(q^{\lambda+1}, s, \ell) \leq N(q, s, \ell) \left[s\ell\lambda + 1\right].$$

Table 1 gives several numerical values of upper bounds on $N(t,2,2)$ calculated with the help of propositions 2 and 4. For instance,

1. $N(16,2,2) = N(4^2,2,2) \leq N(4,2,2) \cdot [4 \cdot 1 + 1] \leq 6 \cdot 5 = 30$;
2. $N(512,2,2) = N(8^3,2,2) \leq N(8,2,2) \cdot [4 \cdot 2 + 1] \leq 14 \cdot 9 = 126$;

t	4	8	16	25	64	512	625	2^{12}	2^{16}	2^{20}
N	6	14	30	50	70	126	250	270	390	510

Table 1. Parameters of superimposed $(2,2)$-codes

3 On Constructions of List-Decoding Codes

For a set of codewords $\mathcal{C}$ (1) and a subset $\tau \subset [t]$ denote by $L(\tau, \mathcal{C}) \geq 0$ the number of indices $j \in [t] \backslash \tau$, such that the vector $\mathbf{x}(j)$ is not covered by $\mathrm{V}(\tau)$. Let $L_s(\mathcal{C})$ denote the maximum value of $L(\tau, \mathcal{C})$ over all $\tau \subset [t]$, $|\tau| = s$. The number $L_s(\mathcal{C})$ is the maximum list-size of the list-decoding superimposed code of strength s (see definition 2).

Along with $L_s(\mathcal{C})$ we study the number $L_s^*(\mathcal{C})$, which is the *average number* of codewords covered by a random s-subset $\tau \subset [t]$:

$$L_s^*(\mathcal{C}) \triangleq \sum_{\substack{\tau \subset [t] \\ |\tau\| = s}} L(\tau, \mathcal{C}) \Bigg/ \binom{t}{s}. \tag{4}$$

Further we claculate value of L_s^* and give the upper bound on L_s, for binary superimposed codes, that are obtained from q-nary MDS codes by trival concatenation. Those codes were studied in [1, 6, 7]

We say that the concatenation is trival if q-nary symbols are replaced with the columns of the $(q \times q)$ identity matrix.

Theorem 1: For a binary superimposed code, obtained from (q, k, n) MDS code by trival concatenation,

$$L_p^* = q^k \frac{\left(\binom{q^k-1}{p} - C(p)\right)}{\binom{q^k}{p}} \tag{5}$$

$$C(p) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} D(p, i)$$

$$D(p, v) = \begin{cases} \binom{q^{k-v}(q-1)^v}{p}, & \text{if } v \leq k; \\ \binom{A_v(v)}{p}, & \text{if } v > k. \end{cases}$$

$$A_v(v) = (q-1) \sum_{j=0}^{k-1} (-1)^j \binom{v-1}{j} q^{k-j-1}$$

Theorem 2: For a binary superimposed code, obtained from (q, k, n) MDS code by trival concatenation,

$$L(s) \leq min\{s^k - s, q^k - \frac{n * (q-s) * q^{k-1}}{w} - s\}, \tag{6}$$

where w is the greatest solution of the equatation

$$\prod_{i=1}^{k-1} (w - i) = (n-1)(n-k+1) \left(\frac{q-s}{q}\right)^{k-1}$$

Another construction of list-decoding superimposed codes based on the incidence of the finite sets was studied in [5].

References

[1] W.H. Kautz, R.C. Singleton, "Nonrandom Binary Superimposed Codes," *IEEE Trans. Inform. Theory*, **4** (1964), 363-377.

[2] R.S. Singleton, "Maximum Distance Q-Nary Codes," *IEEE Trans. Inform. Theory*, **2** (1964), 116-118.

[3] F.J. MacWilliams, N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North Holland, 1983.

[4] A. D'yachkov, V. Rykov, "A Survey of Superimposed Code Theory," *Problems of Control and Inform. Theory*, **4** (1983), 229–242.

[5] P. Vilenkin, "On Constructions of List–Decoding Superimposed Codes," *Proc. of ACCT-6*, Pskov, Russia, 1998, 228–231.

[6] A. D'yachkov, A. Macula, V. Rykov, "New Constructions of Superimposed Codes," *IEEE Trans. Inform. Theory*, **1** (2000), 284–290.

[7] A. D'yachkov, A. Macula, V. Rykov, "New Applications and Results of Superimposed Code Theory Arising from the Potentialities of Molecular Biology," *Numbers, Information and Complexity*, pp. 265–282, Kluwer Academic Publishers, 2000.

[8] A. D'yachkov, A. Macula, D. Torney, P. Vilenkin, S. Yekhanin, "New Results in the Theory of Superimposed Codes: Part I," *present volume*.