数理解析研究所講究録
第607巻 1987年 76-82

A Stochastic Construction of Golay Code

Yoshiaki Itoh

The Institute of Statistical Mathematics
4-6-7 Minami-Azabu Minato-ku,
Tokyo 106, Japan

Masakazu Jimbo

Department of Information Sciences
Science University of Tokyo
Noda City, Chiba 278,
Japan

§1. Golay binary code and random packing

Leech (1964) gave the densest known packing of spheres in 24-dimensional space on the 24-digit binary sequences based on the Golay code, which is generated by 12 binary sequences. Hence the set is made up of 2^{12} or 4096 points of binary sequences called code words. From each point of the 4096 points, there are 759 points of Hamming distance 8, 2576 points of Hamming distance 12, 759 points of Hamming distance 16, and one point of Hamming distance 24. The proof is given for example in the book by MacWilliams and Sloane (1977). Here we will discuss a stochastic algorithm to generate Golay code introduced by Itoh (1986).

Random packing by Hamming distance is discussed by Itoh and Solomon (1986). Consider a random sequential packing into the 2^{24} points. At first we choose one point (24 coordinates) at random and

we record it. Choose another and record it if its Hamming distance is 8, 12, 16 or 24, otherwise reject it. Now, choose the next point at random and record it if the Hamming distance from each of the previously chosen 2 points is 8, 12, 16 or 24. When the points $I_1, I_2 \ldots, I_k$ are already chosen, the next point I_{k+1} will be chosen if each of the Hamming distance from each of the previously chosen $I_1, I_2, \ldots, I_k$ is 8, 12, 16 or 24. We continue this procedure until there is no possible point to record among the 2^{24} points and we now have a set S of recorded points. The histogram of the number of recorded points of S is given in Fig. 1. Out of 550 trials, 11 trials produce Golay code of 4096 points. Without loosing generality, we can choose the point $(0,0,\ldots, 0)$ as the first point. We get 12 rows from the second to the 13th by the random packing procedure of Hamming distance 8, 12, 16 or 24. Then we make the set of all possible sums of the binary sequences by the addition of modulo 2. If the set consists of 2^{12} points and the distribution of Hamming distance from each point has that of the Golay code, then the 12 rows produce a linear code of minimum distance 8, which is equivalent to the code used by Leech to construct the Leech lattice. Hence this will give a stochastic construction of Golay code.

§2. Random packing and finite field

Let V_n be an n-dimensional space over a finite field GF(2). The weight of a vector $\mathbf{a}$ is denoted by $\omega t(\mathbf{a})$. If the weight of every code word of a linear code C is a multiple of 4 then the code is said to be doubly even. The following lemmas is obvious.

Lemma 1. For two vectors $\mathbf{a}$ and $\mathbf{b}$ in V_n, we have $\omega t(\mathbf{a} + \mathbf{b}) = \omega t(\mathbf{a}) + \omega t(\mathbf{b}) - 2\mathbf{a}\cdot\mathbf{b}$, where $\mathbf{a}\cdot\mathbf{b}$ is the number of coordinates in which both elements of $\mathbf{a}$ and $\mathbf{b}$ are 1.

Lemma 2. Let $\mathbf{a}$ and $\mathbf{b}$ are vectors of V_n such that

$$\omega t(\mathbf{a}) \equiv \omega t(\mathbf{b}) \equiv d(\mathbf{a},\mathbf{b}) \equiv 0 (\mathrm{mod}\ 4).$$

Then we have $\mathbf{a}'\mathbf{b} \equiv \mathbf{a}\cdot\mathbf{b} \equiv 0 \pmod 2$.

Proof. By Lemma 1,

$$2\mathbf{a}'\mathbb{b} \equiv 2\mathbf{a}\cdot\mathbb{b} = \omega t(\mathbf{a}) + \omega t(\mathbb{b})$$
$$-\omega t(\mathbf{a} + \mathbb{b}) = \omega t(\mathbf{a}) + \omega t(\mathbb{b})$$
$$-d(\mathbf{a},\mathbb{b}) \equiv 0 \pmod 4.$$

Theorem 1. Let $\mathbb{I}_1, \ldots, \mathbb{I}_k$ be vectors of V_n. If the weight of every vector $\mathbb{I}_i$ is a multiple of 4 and the Hamming distance between any two vectors $\mathbb{I}_i$ and $\mathbb{I}_j$ $(i \neq j)$ is also a multiple of 4, then the code generated by $\mathbb{I}_1, \ldots, \mathbb{I}_k$ is doubly even.

Proof. Let S be a set of vectors including zero vector.

Assume the distance of any two vectors are multiples of 4 for any two vectors $\mathbf{a}$ and $\mathbb{b}$ of S, we have

$$\omega t(\mathbf{a} + \mathbb{b}) \equiv d(\mathbf{a},\mathbb{b}) \equiv 0 \pmod 4.$$

And

$$\mathbf{a}'\mathbb{b} \equiv \mathbf{a}\cdot\mathbb{b} \equiv 0 \pmod 2$$

is obtained by Lemma 2. Hence for any vector $\mathbb{C}$ of S,

$$\begin{aligned} d(\mathbf{a} + \mathbb{b}, \mathbf{c}) &= \omega t(\mathbf{a} + \mathbb{b} + \mathbf{c}) \\ &\equiv \omega t(\mathbf{a} + \mathbb{b}) + \omega t(\mathbf{c}) - 2(\mathbf{a} + \mathbb{b})'\mathbf{c} \\ &\equiv \omega t(\mathbf{a} + \mathbb{b}) + \omega t(\mathbf{c}) - 2(\mathbf{a}'\mathbf{c} + \mathbb{b}'\mathbf{c}) \\ &\equiv 0 \pmod 4 \end{aligned}$$

holds, which shows that the set $S' = S \cup \{\mathbf{a} + \mathbb{b}\}$ also has the property that the distance of any two vectors of S' are multiples of 4. Hence, by setting $S = \{I_1, \ldots, I_k, \mathbb{0}\}$ the linear code generated by S is shown to be doubly even.

Let S be a set of points (code words) obtained from the above random packing procedure and let $\mathbb{I}_1, \mathbb{I}_2, \ldots, \mathbb{I}_k$ be linearly independent code words which generate linear space C over a finite field GF(2).

Theorem 2.

$S \subset C^{\perp}$ where $C^{\perp}$ is the dual space of C

Proof. For any $\mathbb{J} \in S$

$\mathbb{J}\cdot\mathbb{I}_i \equiv \mathbb{J}\cdot\mathbb{I}_i \equiv 0 \pmod 2$, which means $\mathbb{J} \in C^{\perp}$.

We have $C \subset C^{\perp}$, as given in MacWilliams and Sloane (1977). Hence, $C = C^{\perp}$ if $\dim(C) = k = \frac{n}{2}$.

Hence when there are 12 linearly independent vectors $\mathbb{I}_1, \mathbb{I}_2, \ldots, \mathbb{I}_{12} \in S$, each of the remained vectors of S are represented by a linear combination of the 12 vectors. This gives a simple algorithm for our random packing procedure. We continue the random packing procedure until we get 12 linearly independent vectors $\mathbb{I}_1, \mathbb{I}_2, \ldots, \mathbb{I}_{12}$. Then we make the random packing into the 2^{12} linear combinations of $\mathbb{I}_1, \mathbb{I}_2, \ldots, \mathbb{I}_{12}$. This makes computational time of our simulation much shorter.

§3. Weight distributions and clusters

The enumeration of self-dual codes by Conway and Pless (1980) gives the following possible weight distribution for n = 0, 1, 2, 3, 4, 5, 7, 11,

0	4	8	12	16	20	24
1	6n	759-24n	2576-136n	759-24n	6n	1 .

The code generated by the $\mathbb{I}_1, \mathbb{I}_2 \ldots, \mathbb{I}_{12}$ obtained in the previous section has one of the above 8 weight distributions. Hence the remained possible points after packing mutually independent 12 points should be not less than $4096-(12n+1)\times 12$, which will help to explain Fig. 2. The number of points obtained by our random packing procedure is not less than $4096/(n+1)$, if we can get the twelve independent vectors $\mathbb{I}_1, \mathbb{I}_2, \ldots, \mathbb{I}_{12}$. When n=0, the twelve vectors generate Golay code. Each of the other clusters may correspond to one of the other seven weight distributions. Smaller n may correspond to a cluster of larger recorded numbers.

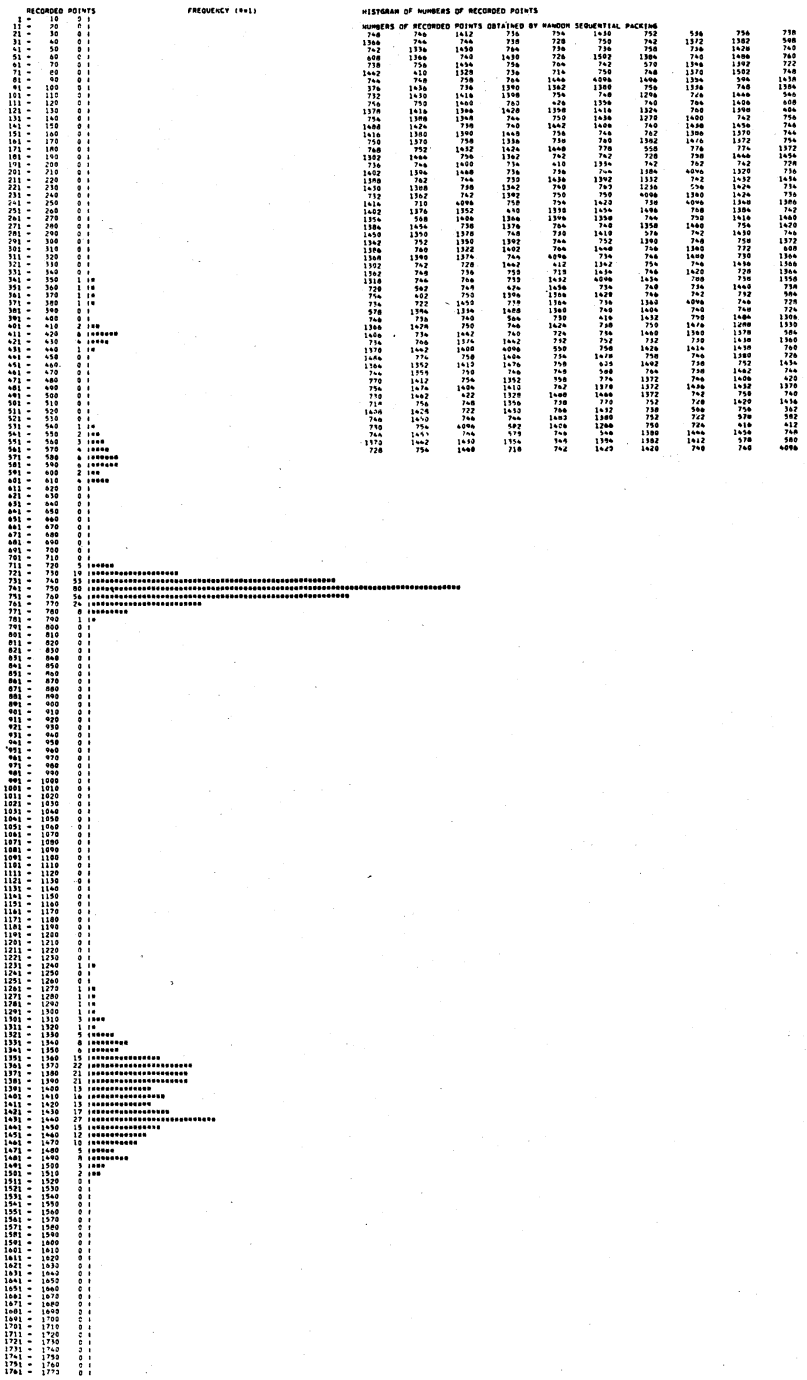

Fig. 1 Histogram

(The 11 trials of 4096 recorded points are not given here.)

(from Itoh(1986))

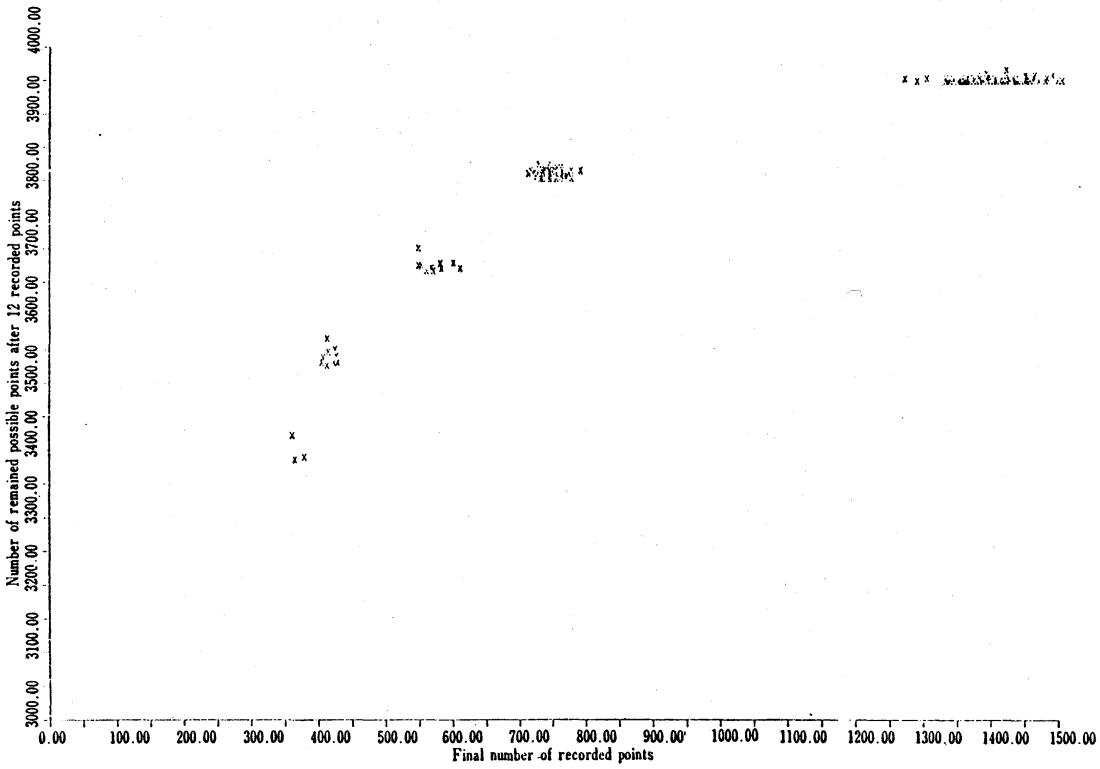

Fig. 2 The first 12 points and the clusters.

(from Itoh(1986))

References

Conway, J.H. and Pless, V (1980). On the enumeration of self-dual codes, Journal of Combinatorial Theory, Series A 28, 26-53.

Itoh, Y.(1986). Golay code and random packing, Ann. Inst. Statist. Math. 38, 583-588.

Itoh, Y. and Solomon, H.(1986). Random sequential coding by Hamming distance, J. Appl. Prob. 23, 688-695.

Leech, J.(1964). Some sphere packings in Higher space, Canad. J. Math., 16, 657-682.

MacWilliams, E.J. and Sloane, N.J.A.(1977). The theory of error-correcting codes, I, II, North-Holland.