Ann. Inst. Statist. Math.
38 (1986), Part B, 583-588

GOLAY CODE AND RANDOM PACKING

YOSHIAKI ITOH

(Received Aug. 9, 1985; revised May 21, 1986)

Summary

A random sequential packing by Hamming distance is applied to study Golay code. The probability of getting Golay code is estimated by computer simulation. A histogram of number of packed points is given to show the existence of several remarkable clusters.

1. Random sequential packing by Hamming distance

In this note we give a stochastic algorithm to generate Golay code. There is an effective and beautiful method by finite field theory to construct various codes. Hence our aim is mainly to show a method to study Golay code for the statistician who is not familiar with finite field theory.

Random sequential packing of spheres has been applied by Bernal [1] to study the structure of liquids and has been discussed by Higuti [4], Solomon [15], Tanemura [16] and others. Random sequential packing of one dimensional space is known as a car parking problem. Place unit intervals into the interval $[0, x]$ sequentially at random. Assume that the initial point ξ of the interval I is a random variable uniformly distributed in the interval $[0, x-1]$. If the intervals $I_1, I_2, \cdots, I_k$ have already been chosen, the next randomly chosen interval will be kept only if it does not intersect any of the intervals $I_1, I_2, \cdots, I_k$. In this case this interval will be denoted by I_{k+1}. If it does intersect any of the intervals $I_1, I_2, \cdots, I_k$ we neglect it and choose a new interval. The procedure is continued until none of the lengths of gaps generated by the intervals placed in $[0, x]$ is greater than 1. Mathematical studies are carried out by Rényi [12], Itoh [5], [6] and others.

Consider a set of 2^d points whose coordinates are 1 or 0 in a Euclidean space of dimension d. Euclidean distance is defined between two points of the 2^d points. The square of the Euclidean distance in this

Key words and phrases: Random sequential packing, Hamming distance, stochastic algorithm, Golay binary code, clusters of histogram.

case is called the Hamming distance in coding theory. At first we choose one point (d coordinates) at random and we record it. Choose another and record it if its Hamming distance is $\geqq k$, $(k<d)$, otherwise discard it. Now, choose the next point at random and record it if the Hamming distance from each of the 2 points is not less than k, otherwise discard it and choose another point at random. We continue this procedure until there is no possible point to record among the 2^d points and we now have the number of recorded points (Itoh and Solomon [8]). For the case of Hamming distances of 2 or 3, $d^{-\alpha}$ fits the simulation results of packing density where α is an empirical constant. The variance of packing density is larger when k is even and smaller when k is odd.

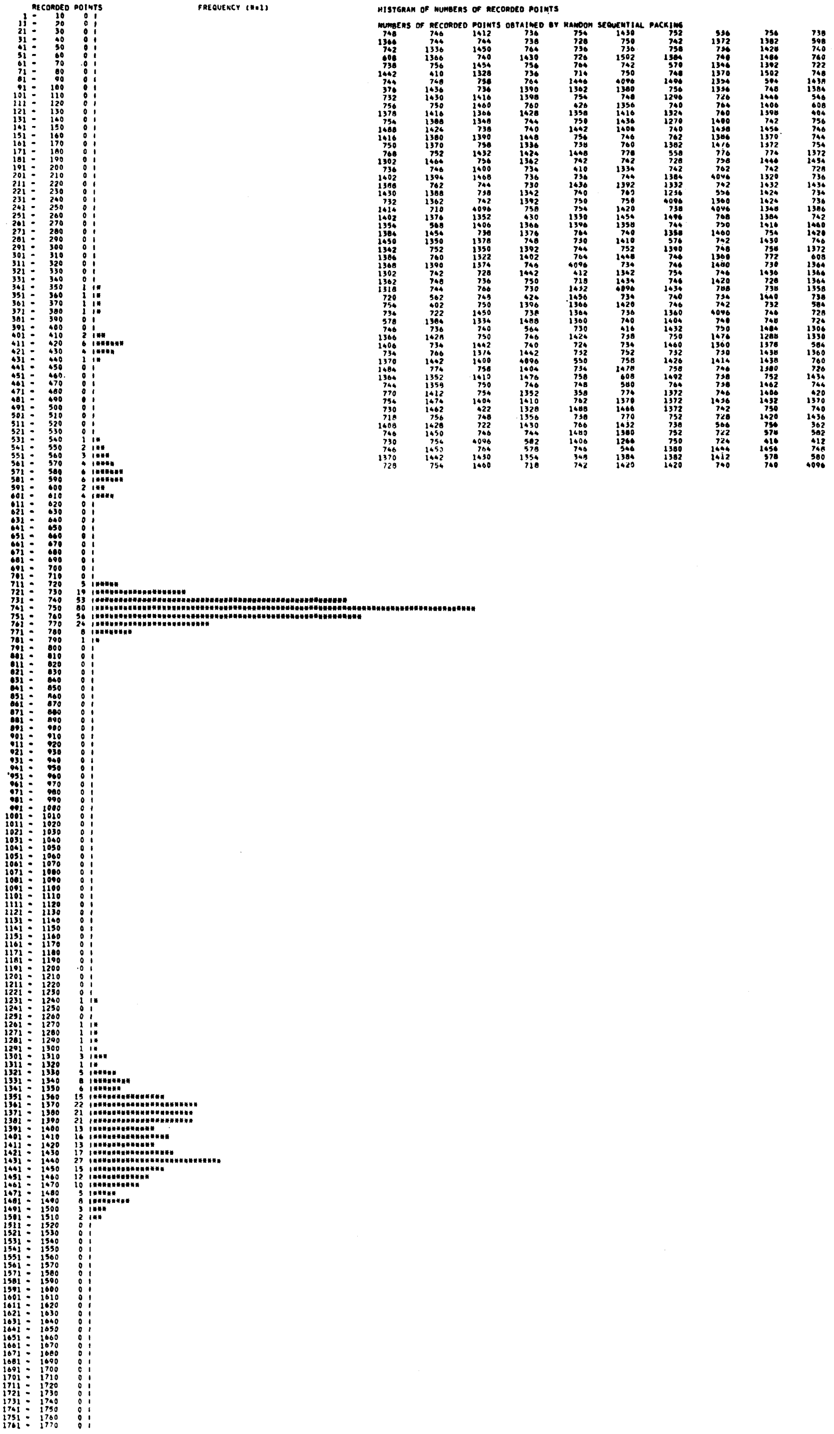
2. Frequency of getting Golay code by a random sequential packing

Leech [9] gave the densest known packing of spheres in 24-dimensional space based on the Golay code, as given by Sloane [14], which is known as one of the most important linear code. Leech used the set of all possible sums of the 12 binary sequences (Fig. 1) where the addition is carried out modulo 2. Hence the set is made up of 2^{12} or 4096 binary sequences called code words. From each point of the 4096 points, out of the remained 4095 points 759 points have Hamming distance 8, 2576 points have Hamming distance 12, 759 points have Hamming distance 16, and one point has Hamming distance 24. The distribution is the same for each of 4096 points. The proof is given in the book by Thompson [17].

Consider a restricted random sequential packing into the 2^{24} points.

$$\begin{matrix}
1&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1&1\\
0&1&0&0&0&0&0&0&0&0&0&0&1&1&0&1&0&0&0&1&1&1&0&1\\
0&0&1&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&1&1&1&0&1&1\\
0&0&0&1&0&0&0&0&0&0&0&0&1&1&0&0&0&1&1&1&0&1&1&0\\
0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&1&1&1&0&1&1&0&1\\
0&0&0&0&0&1&0&0&0&0&0&0&1&0&0&1&1&1&0&1&1&0&1&0\\
0&0&0&0&0&0&1&0&0&0&0&0&1&0&1&1&1&0&1&1&0&1&0&0\\
0&0&0&0&0&0&0&1&0&0&0&0&1&1&1&1&0&1&1&0&1&0&0&0\\
0&0&0&0&0&0&0&0&1&0&0&0&1&1&1&0&1&1&0&1&0&0&0&1\\
0&0&0&0&0&0&0&0&0&1&0&0&1&1&0&1&1&0&1&0&0&0&1&1\\
0&0&0&0&0&0&0&0&0&0&1&0&1&0&1&1&0&1&0&0&0&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&1&0&0&0&1&1&1&0
\end{matrix}$$

Fig. 1. Leech's 12 by 24 matrix.



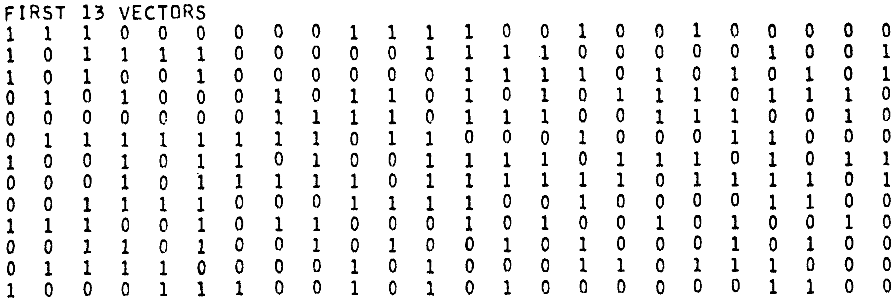
```
HISTGRAM OF NUMBERS OF RECORDED POINTS

NUMBERS OF RECORDED POINTS OBTAINED BY RANDOM SEQUENTIAL PACKING
    748     746    1412     736     754    1430     752     536     756     738
   1366     744     764     738     728     750     742    1372    1382     598
    742    1336    1450     764     736     736     758     736    1428     740
    608    1366     740    1430     726    1502    1384     740    1486     760
    738     756    1454     756     764     742     570    1346    1392     722
   1442     410    1328     736     714     750     748    1370    1502     748
    744     748     758     764    1446    4096    1496    1354     504    1438
    376    1436     736    1390    1362    1380     756    1356     748    1384
    732    1430    1416    1398     754     748    1296     726    1446     546
    756     750    1460     760     426    1356     740     764    1406     608
   1378    1416    1366    1428    1358    1416    1324     760    1398     404
    754    1388    1348     744     750    1436    1270    1400     742     756
   1488    1424     738     740    1442    1408     740    1458    1456     746
   1416    1380    1390    1448     756     746     762    1386    1370     744
    750    1370     758    1336     738     760    1382    1476    1372     754
    768     752    1432    1424    1448     778     558     776     774    1372
   1302    1464     756    1362     742     742     728     758    1446    1454
    736     746    1400     734     410    1334     742     762     742     728
   1402    1394    1468     736     736     744    1384    4096    1320     736
   1386     762     744     730    1436    1392    1332     742    1432    1434
   1430    1388     738    1342     740     760    1256     556    1424     734
    732    1362     742    1392     750     758    4096    1360    1424     736
   1414     710    4096     758     754    1420     738    4096    1348    1386
   1402    1376    1352     430    1330    1454    1496     768    1384     742
   1354     568    1406    1366    1396    1358     744     750    1416    1468
   1384    1454     738    1376     764     740    1358    1460     754    1428
   1450    1350    1378     748     730    1410     576     742    1430     746
   1342     752    1350    1392     744     752    1390     748     758    1372
   1386     760    1322    1402     764    1448     746    1368     772     608
   1368    1390    1374     746    4096     734     746    1480     738    1364
   1302     742     728    1442     412    1342     754     746    1436    1366
   1362     748     736     750     718    1434     746    1420     728    1364
   1318     744     766     730    1432    4096    1434     768     738    1358
    720     562     748     426    1456     734     740     734    1440     738
    754     402     750    1396    1366    1420     746     742     732     584
    734     722    1450     738    1364     736    1360    4096     746     728
    578    1384    1334    1488    1360     740    1404     748     748     724
    746     736     740     564     730     416    1432     750    1484    1306
   1366    1428     750     746    1424     738     750    1476    1288    1330
   1406     734    1442     740     724     734    1460    1360    1378     584
    734     766    1374    1442     732     752     732     730    1438    1360
   1370    1442    1408    4096     550     758    1426    1414    1438     760
   1484     774     758    1404     734    1478     758     746    1380     726
   1364    1352    1410    1476     758     608    1492     738     752    1434
    744    1359     750     746     748     580     764     738    1462     744
    770    1412     754    1352     358     774    1372     746    1486     420
    754    1474    1404    1410     762    1370    1372    1436    1432    1370
    730    1462     422    1328    1488    1466    1372     742     750     740
    718     756     748    1356     738     770     752     728    1428    1436
   1408    1428     722    1430     766    1432     738     566     756     362
    746    1450     746     744    1480    1380     752     722     578     582
    730     754    4096     582    1406    1266     750     724     416     412
    746    1450     764     578     746     546    1380    1444    1456     748
   1370    1442    1430    1354     348    1384    1382    1412     578     580
    728     754    1460     718     742    1420    1420     740     740    4096
```

```
   RECORDED POINTS                   FREQUENCY (*=1)
   1 -   10    0 I
  11 -   20    0 I
  21 -   30    0 I
  31 -   40    0 I
  41 -   50    0 I
  51 -   60    0 I
  61 -   70    0 I
  71 -   80    0 I
  81 -   90    0 I
  91 -  100    0 I
 101 -  110    0 I
 111 -  120    0 I
 121 -  130    0 I
 131 -  140    0 I
 141 -  150    0 I
 151 -  160    0 I
 161 -  170    0 I
 171 -  180    0 I
 181 -  190    0 I
 191 -  200    0 I
 201 -  210    0 I
 211 -  220    0 I
 221 -  230    0 I
 231 -  240    0 I
 241 -  250    0 I
 251 -  260    0 I
 261 -  270    0 I
 271 -  280    0 I
 281 -  290    0 I
 291 -  300    0 I
 301 -  310    0 I
 311 -  320    0 I
 321 -  330    0 I
 331 -  340    0 I
 341 -  350    1 I*
 351 -  360    1 I*
 361 -  370    1 I*
 371 -  380    1 I*
 381 -  390    0 I
 391 -  400    0 I
 401 -  410    2 I**
 411 -  420    6 I******
 421 -  430    4 I****
 431 -  440    1 I*
 441 -  450    0 I
 451 -  460.   0 I
 461 -  470    0 I
 471 -  480    0 I
 481 -  490    0 I
 491 -  500    0 I
 501 -  510    0 I
 511 -  520    0 I
 521 -  530    0 I
 531 -  540    1 I*
 541 -  550    2 I**
 551 -  560    3 I***
 561 -  570    4 I****
 571 -  580    6 I******
 581 -  590    6 I******
 591 -  600    2 I**
 601 -  610    4 I****
 611 -  620    0 I
 621 -  630    0 I
 631 -  640    0 I
 641 -  650    0 I
 651 -  660    0 I
 661 -  670    0 I
 671 -  680    0 I
 681 -  690    0 I
 691 -  700    0 I
 701 -  710    0 I
 711 -  720    5 I*****
 721 -  730   19 I*******************
 731 -  740   53 I*****************************************************
 741 -  750   80 I********************************************************************************
 751 -  760   56 I********************************************************
 761 -  770   24 I************************
 771 -  780    8 I********
 781 -  790    1 I*
 791 -  800    0 I
 801 -  810    0 I
 811 -  820    0 I
 821 -  830    0 I
 831 -  840    0 I
 841 -  850    0 I
 851 -  860    0 I
 861 -  870    0 I
 871 -  880    0 I
 881 -  890    0 I
 891 -  900    0 I
 901 -  910    0 I
 911 -  920    0 I
 921 -  930    0 I
 931 -  940    0 I
 941 -  950    0 I
'951 -  960    0 I
 961 -  970    0 I
 971 -  980    0 I
 981 -  990    0 I
 991 - 1000    0 I
1001 - 1010    0 I
1011 - 1020    0 I
1021 - 1030    0 I
1031 - 1040    0 I
1041 - 1050    0 I
1051 - 1060    0 I
1061 - 1070    0 I
1071 - 1080    0 I
1081 - 1090    0 I
1091 - 1100    0 I
1101 - 1110    0 I
1111 - 1120    0 I
1121 - 1130    0 I
1131 - 1140    0 I
1141 - 1150    0 I
1151 - 1160    0 I
1161 - 1170    0 I
1171 - 1180    0 I
1181 - 1190    0 I
1191 - 1200    0 I
1201 - 1210    0 I
1211 - 1220    0 I
1221 - 1230    0 I
1231 - 1240    1 I*
1241 - 1250    0 I
1251 - 1260    0 I
1261 - 1270    1 I*
1271 - 1280    1 I*
1281 - 1290    1 I*
1291 - 1300    1 I*
1301 - 1310    3 I***
1311 - 1320    1 I*
1321 - 1330    5 I*****
1331 - 1340    8 I********
1341 - 1350    6 I******
1351 - 1360   15 I***************
1361 - 1370   22 I**********************
1371 - 1380   21 I*********************
1381 - 1390   21 I*********************
1391 - 1400   13 I*************
1401 - 1410   16 I****************
1411 - 1420   13 I*************
1421 - 1430   17 I*****************
1431 - 1440   27 I***************************
1441 - 1450   15 I***************
1451 - 1460   12 I************
1461 - 1470   10 I**********
1471 - 1480    5 I*****
1481 - 1490    8 I********
1491 - 1500    3 I***
1501 - 1510    2 I**
1511 - 1520    0 I
1521 - 1530    0 I
1531 - 1540    0 I
1541 - 1550    0 I
1551 - 1560    0 I
1561 - 1570    0 I
1571 - 1580    0 I
1581 - 1590    0 I
1591 - 1600    0 I
1601 - 1610    0 I
1611 - 1620    0 I
1621 - 1630    0 I
1631 - 1640    0 I
1641 - 1650    0 I
1651 - 1660    0 I
1661 - 1670    0 I
1671 - 1680    0 I
1681 - 1690    0 I
1691 - 1700    0 I
1701 - 1710    0 I
1711 - 1720    0 I
1721 - 1730    0 I
1731 - 1740    0 I
1741 - 1750    0 I
1751 - 1760    0 I
1761 - 1770    0 I
```

Fig. 2. Histogram

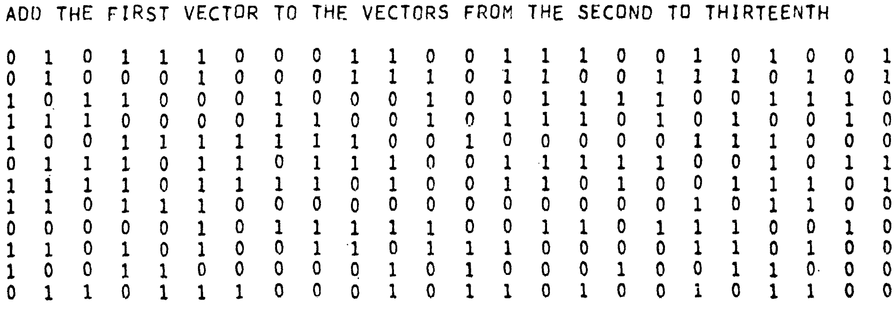
(The 11 trials of 4096 recorded points are not given here.)

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 discard 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. If the points $I_1, I_2, \cdots, I_k$ are already chosen, the next point I_{k+1} will be chosen if the Hamming distance from each of the previously chosen $I_1, I_2, \cdots, I_k$ 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 the number of recorded points N. This random sequential packing is introduced by Itoh [7]. Here we give histogram of the number of recorded points N in Fig. 2, where there are several clusters (see Itoh [7]). 11 trials out of 550 trials gave 4096 recorded points, that is to say, a code of 4096 code words of length 24 with minimum distance 8 was constructed by the above random sequential packing. Consider the first 13 vectors obtained by a random packing of 4096 recorded points (Fig. 3 (A)). Add the first vector to the other 12 vectors by modulo 2. Then we get the 12 vectors (Fig. 3 (B)). If



```
FIRST 13 VECTORS
1  1  1  0  0  0  0  0  0  1  1  1  1  0  0  1  0  0  1  0  0  0  0  0
1  0  1  1  1  1  0  0  0  0  0  1  1  1  1  0  0  0  0  0  1  0  0  1
1  0  1  0  0  1  0  0  0  0  0  0  1  1  1  1  0  1  0  1  0  1  0  1
0  1  0  1  0  0  0  1  0  1  1  0  1  0  1  0  1  1  1  0  1  1  1  0
0  0  0  0  0  0  0  1  1  1  1  0  1  1  1  0  0  1  1  1  0  0  1  0
0  1  1  1  1  1  1  1  1  0  1  1  0  0  0  1  0  0  0  1  1  0  0  0
1  0  0  1  0  1  1  0  1  0  0  1  1  1  1  0  1  1  1  0  1  0  1  1
0  0  0  1  0  1  1  1  1  1  0  1  1  1  1  1  1  0  1  1  1  1  0  1
0  0  1  1  1  1  0  0  0  1  1  1  1  0  0  1  0  0  0  0  1  1  0  0
1  1  1  0  0  1  0  1  1  0  0  0  1  0  1  0  0  1  0  1  0  0  1  0
0  0  1  1  0  1  0  0  1  0  1  0  0  1  0  1  0  0  0  1  0  1  0  0
0  1  1  1  1  0  0  0  0  1  0  1  0  0  0  1  1  0  1  1  1  0  0  0
1  0  0  0  1  1  1  0  0  1  0  1  0  1  0  0  0  0  0  0  1  1  0  0
```

Fig. 3 (A)



```
ADD THE FIRST VECTOR TO THE VECTORS FROM THE SECOND TO THIRTEENTH

0  1  0  1  1  1  0  0  0  1  1  0  0  1  1  1  0  0  1  0  1  0  0  1
0  1  0  0  0  1  0  0  0  1  1  1  0  1  1  0  0  1  1  1  0  1  0  1
1  0  1  1  0  0  0  1  0  0  0  1  0  0  1  1  1  1  0  0  1  1  1  0
1  1  1  0  0  0  0  1  1  0  0  1  0  1  1  1  0  1  0  1  0  0  1  0
1  0  0  1  1  1  1  1  1  1  0  0  1  0  0  0  0  0  1  1  1  0  0  0
0  1  1  1  0  1  1  0  1  1  1  0  0  1  1  1  1  1  0  0  1  0  1  1
1  1  1  1  0  1  1  1  1  0  1  0  0  1  1  0  1  0  0  1  1  1  0  1
1  1  0  1  1  1  0  0  0  0  0  0  0  0  0  0  0  0  1  0  1  1  0  0
0  0  0  0  0  1  0  1  1  1  1  1  0  0  1  1  0  1  1  1  0  0  1  0
1  1  0  1  0  1  0  0  1  1  0  1  1  1  0  0  0  0  1  1  0  1  0  0
1  0  0  1  1  0  0  0  0  0  1  0  1  0  0  0  1  0  0  1  1  0  0  0
0  1  1  0  1  1  1  0  0  0  1  0  1  1  0  1  0  0  1  0  1  1  0  0
```

Fig. 3 (B)

the 12 vectors are linearly independent, the all possible sum of the 12 vectors make the same distribution of Hamming distance to the one generated by the 12 vectors given in Fig. 1. Hence the first 13 vectors are enough to construct a Golay code. It has been shown that any binary block code with 2^{12} code words of length 24 with minimum distance 8 is equivalent to the one generated by the vectors in Fig. 1 (see MacWilliams and Sloane [10]). Since the works by Golay [2], Hamming [3] and Shannon [13], many works are carried out on coding theory. Golay discovered the code by geometrical consideration. Golay's original idea introduced in the book by Thompson [17] is not based on finite field theory. His geometrical discussion is not easy to follow. The algebraic method to construct the Golay code is given in the book by Peterson [11]. It may be remarkable that Golay code is constructed at random without using deterministic construction.

3. Clusters in histogram

There are several evident clusters in the histogram of the recorded number N in Fig. 2. Here we give a remark on the clusters.

In our random packing the first point I_1 is recorded at random from the 2^{24} points. The second point I_2 is recorded at random from the points which have the Hamming distance 8, 12, 16 or 24 from the point I_1. The k-th point I_k is recorded at random from the points which have the Hamming distance 8, 12, 16 or 24 from each of the previously recorded $I_1, I_2, \cdots, I_{k-1}$. Consider the distribution of the number of points which have the Hamming distance 8, 12, 16 or 24 from each of the 12 points $I_1, I_2, \cdots, I_{12}$. Let X be the number. We consider the simulation of the 550 trials given in Fig. 2. Out of the 550 trials, 9 trials produced $X=4084$. The 9 trials have the recorded number of

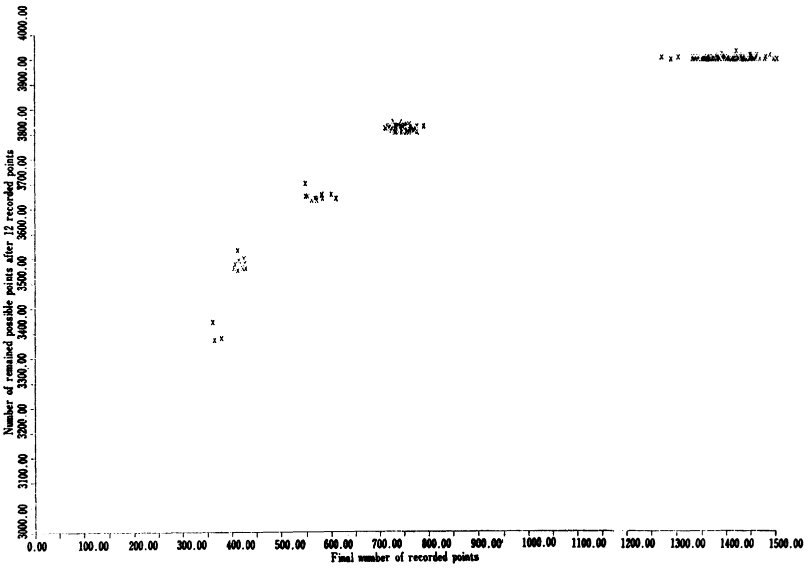


Fig. 4. The first 12 points and the clusters.

points, $N=4084+12=4096$. There were 306 trials for which $X<4084$, and 235 trials for which $X>4084$. When $X>4084$, it is supposed that the 12 vectors, which represent the first 12 points, are not linearly independent. Hence we limit our analysis to the trials for which $X<4084$. We plot (N, X) for each of the 305 trials in Fig. 4, which shows that the first 12 points roughly determine the recorded number N and the cluster in the histogram in Fig. 2.

Acknowledgement

The author is grateful to Dr. N. Niki who provided him with a computer program in machine language to calculate Hamming distance. He is also grateful to Miss Suzu Ishihara for programming the simulation.

This study was carried out under the ISM Cooperative Research Programs 85-ISMCRP-2, and 85-ISMCRP-13,

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Bernal, J. D. (1959). A geometrical approach to the structure of liquids, *Nature*, **17**, 141-147.

[2] Golay, M. J. E. (1949). Notes on digital coding, *Proc. I.R.E.* (*I.E.E.E.*), **37**, 657.

[3] Hamming, R. W. (1947). *Self-Correcting Codes-Case 20878*, Memorandum 1130-RWH-MFW, Bell Telephone Laboratories, July 27, 1947.

[4] Higuti, I. (1960). A statistical study of random packing of unequal spheres, *Ann. Inst. Statist. Math.*, **12**, 257-271.

[5] Itoh, Y. (1980). On the minimum of gaps generated by one-dimensional random packing, *J. Appl. Prob.*, **17**, 134-144.

[6] Itoh, Y. (1985). Note on a restricted random cutting of a stick, *Proc. Inst. Statist. Math.*, **33**, 97-99. (In Japanese with English summary).

[7] Itoh, Y. (1985). Abstract of research works in 1984 "random packing by Hamming distance", *Proc. Inst. Statist. Math.*, **33**, 156-157. (In Japanese).

[8] Itoh, Y. and Solomon, H. (1986). Random sequential coding by Hamming distance, *J. Appl. Prob.* (to appear).

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

[10] MacWilliams, E. J. and Sloane, N. J. A. (1977). *The Theory of Error-Correcting Codes*, I, II, North-Holland.

[11] Peterson, W. W. (1961). *Error-Correcting Codes*, M.I.T. Press, Cambridge, Massachusetts.

[12] Rényi, A. (1958). On a one-dimensional problem concerning random space filling, *Publ. Math. Inst. Hung. Acad. Sci.*, **3**, 109-127.

[13] Shannon, C. E. (1948). A mathematical theory of communication, *Bell System Tech. J.*, **27**, 379-423, 623-656.

[14] Sloane, N. J. A. (1984). The packing of spheres, *Scientific American*, January, 116-125.

[15] Solomon, H. (1967). Random packing density, *Proc. Fifth Berkeley Symp. Math. Stat. Prob.*, **3**, 119-134, Univ. of California Press.

[16] Tanemura, M. (1979). On random complete packing by discs, *Ann. Inst. Statist. Math.*, **31**, 351-365.

[17] Thompson, T. M. (1983). From error-correcting codes through sphere packing to simple groups, *The Mathematical Association of America*.