AD-A158 003 RANDOM SEQUENTIAL CODING BY HAMMING DISTANCE(U) STANFORD UNIV CA DEPT OF STATISTICS Y ITOH ET AL. 11 JUL 85 TR-360 N00014-76-C-0475

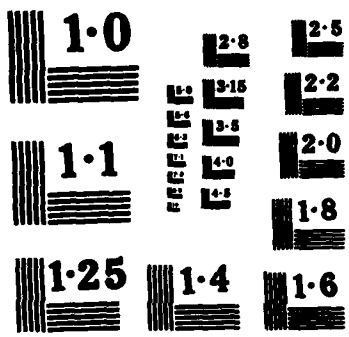
1/1

UNCLASSIFIED F/G 12/1 NL

END

FILMED

DTIC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

RANDOM SEQUENTIAL CODING BY HAMMING DISTANCE

BY

YOSHIAKI ITOH and HERBERT SOLOMON

TECHNICAL REPORT NO. 360

JULY 11, 1985

Prepared Under Contract

N00014-76-C-0475 (NR-042-267)

For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

§1. Introduction

The packing of spheres has been a subject of interest for many scientists (Sloane (1984)). Many investigators have sought the densest way to arrange identical spheres in space. An asymptotic formula that provides an upper bound for the packing density, σ_d, of d-dimensional spheres in a d-dimensional space as the space volume approaches infinity is due to H.E. Daniels and reported in Rogers (1958, 1964). It is

$$\sigma_d \sim \frac{(d+1)! e^{(d/2)-1}}{\sqrt{2}\Gamma(1+\frac{d}{2})(4d)^{d/2}}$$
$$\sim \frac{d}{e}\left(\frac{1}{\sqrt{2}}\right)^d$$

In addition several investigators have looked into the random packing of spheres. Random sequential packing of spheres has been applied by Bernal (1958) to study the structure of liquids and since then it has been discussed by other authors, for example, Higuti (1960), Solomon (1967), Tanemura (1979) and others. Because of analytical difficulties when $n \geq 2$, one dimensional random sequential packing has received attention. It is an interesting subject in probability and has been discussed by several authors, for example, Rényi (1958), Dvoretsky and Robbins (1964), Solomon (1967), and Itoh (1980). The one dimensional model can be extended to random sequential packing of cubes. Consider random sequential packing of hypercubes of sidelength 1 into a larger hypercube of sidelength x with rigid boundaries. It seems to be natural to expect that the limiting density β_d as x tends to infinity in a space of dimension d equals β^d where β is the limiting packing density for $d = 1$.

This conjecture was posed and discussed by Palasti (1960). as a follow up to Rényi's pioneering article and his results for one dimensional sequential random packing. Several simulations and analyses by Blaisdell and Solomon (1970, 1982) and Akeda and Hori (1975, 1976) refute her conjecture. Asymptotic behavior of the random packing density as the dimension tends to infinity is of interest, but there is no analytical result for dimensions higher than one. Computer simulatons are difficult and expensive as dimension increases. Here we introduce two simple models: simple cubic random packing and random packing by Hamming distance, to get some feel for random packing density in higher dimension.

We improve the previous simulation results for cubic random packing (Itoh and Ueda (1983)) by increasing the number of runs. Also, we give simulation results for random packing by Hamming distance and discuss the behavior of packing density when dimensionality is increased. For the case of Hamming distances of 2 or 3, $d^{-\alpha}$ fits the simulation results of packing density where d represents the dimension and $\alpha>0$ is an empirical constant estimated by the least squares method, see Table 3.
The variance of packing density is larger when k is even and smaller when k is odd, see Table 2, where k represents Hamming distance.

§2. Simple cubic random sequential packing

Blaisdell and Solomon (1982) gave an experimental formula

$$\beta_d^{1/d} - \beta_1 = (d-1)(\beta_2^{1/2} - \beta_1) \quad \text{for}$$

$d = 3, 4$, which may direct our attention to packing density in higher dimensions. But for the more general situation discussed by the above authors, it is difficult to get values for more than four dimensions. For a simple random packing model, Itoh and Ueda (1983) gave packing densities up to 11 dimensions. In this model consider a cube in d dimensions with sidelength 4 and a cubic lattice with unit sidelength. Cubes of sidelength 2 are put sequentially at random into the cube of sidelength 4 so that each vertex coincides with one of the lattice points. We continue until no place can be found in the large cube to place the smaller one. Consider the packing density γ_d of dimension d. The experimental formula by Blaisdell and Solomon (1982) fits the results for $d = 3, 4, 5$ as we can easily see from the results by Itoh and Ueda (1983). From computer simulations up to dimension 11, γ_{d+1}/γ_d seems to approach one (Itoh and Ueda (1983)), which may suggest β_{d+1}/β_d of the general model also approaches one as d tends to infinity, see Table 1.

§3. Random sequential coding by Hamming distance

One of the possible applications of spherical random packing is in recognition theory, (Dolby and Solomon (1975)). To minimise the recognition error rate for a given rejection rate, the following decision rule is reasonable, if the objects are all equally likely to occur. If an arbitrary point in the observation space is within r distance units of the expected value point of one (and only one) of the given objects, identify the arbitrary point with that object. Otherwise reject the observation. Here distance is the usual Euclidean distance. Dolby and Solomon (1975) compared spherical random packing density with the ratio between existing monosyllabic words and possible monosyllabic forms in English. Here we simplify the situation and introduce a random sequential coding model.

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 case is called the Hamming distance in coding theory. Problems in random coding are discussed by several authors, (Shannon (1948), Abramson (1963)). Our model may be called random sequential coding. Consider a random sequential packing into the 2^d points. At first we choose one point (d coordinates) at random and we record it. Choose another and record it if its Hamming distance is $\geq k, (k < d)$, 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 $X(d, k)$.

Define a packing density by $X(d, k)/2^d$ and label $E(X(d, k)/2^d) = \delta_{d,k}$. Computer simulations suggest $\delta_{d+1,k}/\delta_{d,k}$ approaches one as d tends to infinity, as in the case of the simplest cubic random packing. Another interesting fact is the variation of the variance of $X(n, k)$. For k even the variance is large and for odd k the variance is small; for fixed d. see Table 2.

Our model will help to examine the coding system arising in nature (as well as linguistic problems). For example, in the case of the amino acid code, 64 words are theoretically possible in the triplet coding system by four species of nucleotides i.e. 4^3. The actual number of amino acids plus chain terminator in the code is only 21. The ratio of the number of elements actually used to the number possible is $21/64 = 0.328125$. It should be noted that this value is quite close to 0.3263, the packing value for $d = 6$ and $k = 2$.

Note that $2^6 = 64$. A random sequential packing model by Dolby and Solomon (1975) was applied to this problem by Itoh and Hasegawa (1980). But our present model may be more amenable for this situation.

§4. Packing density and experimental formula

From the inequality for Hamming bounds, we have

$$\frac{1}{\sum_{i=1}^{2e}\binom{d}{i}} < \delta_{d,k} < \frac{1}{\sum_{i=1}^{e}\binom{d}{i}}$$

for $k = 2e+1$, see Liu (1968). Hence, if the limit of $\delta_{d+1,k}/\delta_{d,k}$ exists, it should be one. The Hamming upper bound can be strengthened as follows. Let $M(d,k)$ be the maximum possible number of code words in a binary code with words of length d and minimum distance k, if $k/d < \frac{1}{2}$, then

$$M(d,k) \leq \frac{d \cdot 2^d K(p)}{\sum_{r=0}^{[tpd]}\binom{d}{r}\left(\frac{tk}{2} - r\right)}$$

where $p = k/2d$, $t = (1/2p(1-\sqrt{1-4p})$, $K(p) = p/\sqrt{1-4p}$ (Wyner (1964), Ash (1969)). Consider a regular packing for $k = 2$. Consider a set of points on the hyperplanes $x_1 + x_2 + \cdots x_d = m$, where m is odd. Then the mutual distance between each pair of points on the set is not less than 2. Take every point on the hyperplane for odd m and do not take any point of the hyperplanes for even m. This makes a regular packing of density 0.5, which does not depend on dimension. The form $d^{-\alpha}$ fits our random packing density reasonably well for $k = 2, 3$, as we shall see in Table 3, where α is estimated from the logarithmic plot of the simulation result for $10 \leq d \leq 17$ by the least squares method. For the case $k \geq 4$, $d^{-\alpha}$ does not fit our simulation results.

Acknowledgment

We would like to thank Joe Marhoul and Akira Yoshida for their help in programming the computer simulations.

References

Abramson, N. (1963) Information theory and coding, McGraw-Hill, New York, San Francisco, Toronto, London.

Akeda, Y. and Hori, M. (1975) Numerical test of Palasti's conjecture on two-dimensional random packing density. Nature, London 254, 318-319.

Akeda, Y. and Hori, M. (1976) On random sequential packing in two and three dimensions. Biometrika 63, 361-366.

Ash, R. (1965) Information Theory, Interscience Publishers, New York, London, Sydney.

Blaisdell, B.E. and Solomon, H. (1970) On random sequential packing in the plane and a conjecture of Palasti. J. Appl. Prob. 7, 667-698.

Blaisdell, B.E. and Solomon, H. (1982) Random sequential packing in Euclidean spaces of dimensions three and four and a conjecture of Palasti. J. Appl. Prob. 19, 382-390.

Dolby, J.L. and Solomon, H. (1975) Information density phenomena and random packing. J. Appl. Prob. 12, 364-370.

Dvoretzky, A. and Robbins, H. (1964) On the packing problem. Publ. Math. Inst. Hung. Acad. Sci. 9, 209-224.

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

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

Itoh, Y. and Hasegawa, M. (1981) Why twenty Amino Acids, Seibutsubutsuri. 21, 21-22 (in Japanese).

Itoh, Y. and Ueda, S. (1983) On packing density by a discrete random sequential packing of cubes in a space of n dimension. Proceedings of the Institute of Statistical Mathematics. 31, 65-69 (in Japanese with English summary).

Liu, C.L. (1968) Introduction to combinatorial mathematics, McGraw-Hill, New York.

Palasti, I. (1960) On some random space filling problems, Publ. Math. Inst. Hungar. Acad. Sci., Vol. 5, 353-360.

Rènyi, A. (1958) On a one dimensional problem concerning space filling. Publ. Math. Inst. Hung. Acad. Sci. 3, 109-127.

Rogers, C.A. (1958) The packing of equal spheres. Proc. London Math. Soc. 3, 609-620.

Rogers, C.A. (1964) Packing and covering. Cambridge Tracts in Mathematics and Mathematical Physics. No. 54, Cambridge University Press, Cambridge.

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

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

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

Tanemura, M. (1979) On random complete packing by discs. Ann. Inst. Statis. Math. 31, 351-365.

Wyner, A. (1964) Improved bounds for minimum distance and error probability in discrete channels. Bell Telephone Laboratories Internal Report, Murray Hill, N.J.

Simple cubic random packing

dimension	1	2	3	4	5	6	7	8	9	10	11
mean γ_d	0.8348	0.7112	0.6157	0.5481	0.4927	0.4508	0.4212	0.3958	0.3762	0.3631	0.3516
standard deviation	0.2352	0.2196	0.1790	0.1385	0.1044	0.0773	0.0548	0.0404	0.0277	0.0178	0.0131
number of trials	10000	10000	10000	10000	10000	10000	10000	3000	500	100	100
γ_{d+1}/γ_d	0.8519	0.8657	0.8902	0.8989	0.9150	0.9343	0.9397	0.9505	0.9652	0.9683	

Table 1

(The values for the dimension 11 are improvements over the results by Itoh and Ueda (1983) achieved by increasing the number of experiments.)

TABLE 2

d	k	packing mean	variance	# of trials
3	2	3.493200	0.756829	10,000
4	2	6.203800	2.815747	10,000
5	2	11.080800	6.951967	10,000
5	3	3.879800	0.225975	10,000
6	2	20.171000	14.521211	10,000
6	3	6.213600	0.704846	10,000
6	4	3.363600	0.867882	10,000
7	2	37.082400	27.182128	10,000
7	3	9.938700	1.191661	10,000
7	4	5.139200	1.673591	10,000
8	2	68.899800	50.970457	10,000
8	3	16.461900	0.758824	10,000
8	4	8.237700	2.544253	10,000
8	5	3.810200	0.343610	10,000
9	2	129.102400	98.656380	10,000
9	3	28.378700	1.705657	10,000
9	4	13.091700	4.987190	10,000
9	5	4.861000	1.020781	10,000
9	6	3.288600	0.916802	10,000
10	2	242.240800	168.855301	10,000
10	3	49.495400	3.106489	10,000
10	4	21.141000	10.550774	10,000
10	5	7.588200	0.738095	10,000
10	6	4.434000	0.916536	10,000
11	2	457.204000	310.400785	1,000
11	3	87.313000	5.254285	1,000
11	4	34.859000	21.638758	1,000
11	5	11.676000	0.655680	1,000
11	6	6.417000	1.654766	1,000
11	7	3.782000	0.388865	1,000
12	2	867.509000	558.192111	1,000
12	3	155.635000	9.639414	1,000
12	4	58.842000	47.304340	1,000
12	5	18.122000	0.956072	1,000
12	6	9.853000	2.429821	1,000
12	7	3.986000	0.027832	1,000
12	8	3.310000	0.904805	1,000
13	2	1653.710000	959.137273	100
13	3	279.360000	22.394343	100
13	4	98.880000	116.328889	100
13	5	28.770000	1.451616	100
13	6	15.040000	4.705455	100
13	7	6.060000	0.279192	100
13	8	3.960000	0.079192	100

Table 2 (continued)

14	2	3155.140000	1373.091313	100
14	3	504.080000	27.872323	100
14	4	170.270000	243.128384	100
14	5	46.490000	2.131212	100
14	6	22.720000	10.203636	100
14	7	9.050000	0.654040	100
14	8	5.040000	1.048889	100
14	9	3.660000	0.570101	100
15	2	6029.100000	3039.525253	100
15	3	918.180000	45.280404	100
15	4	296.880000	471.056162	100
15	5	75.600000	3.474747	100
15	6	35.450000	16.654040	100
15	7	13.810000	0.620101	100
15	8	7.720000	1.880404	100
15	9	4.000000	0.000000	100
15	10	3.200000	0.969697	100
16	2	11585.050000	5824.371212	100
16	3	1675.980000	104.302626	100
16	4	513.730000	1140.300101	100
16	5	124.400000	4.606061	100
16	6	55.270000	31.855657	100
16	7	20.340000	1.297374	100
16	8	11.640000	3.424646	100
16	9	5.300000	1.080808	100
16	10	3.920000	0.155152	100
17	2	22306.100000	14989.211111	10
17	3	3074.400000	291.155556	10
17	4	895.500000	3013.166667	10
17	5	207.600000	5.155556	10
17	6	86.600000	65.600000	10
17	7	29.700000	1.344444	10
17	8	17.300000	4.011111	10
17	9	7.000000	1.111111	10
17	10	4.200000	0.400000	10
17	11	3.400000	0.933333	10

Table 3 Random packing density and $d^{-\alpha}$

d	k	packing density	$d^{-\alpha}$
3	2	0.43665	0.50333
4	2	0.38773	0.42052
5	2	0.34627	0.36578
6	2	0.31517	0.32540
7	2	0.28970	0.29642
8	2	0.26913	0.27269
9	2	0.25215	0.25334
10	2	0.23656	0.23720
11	2	0.22324	0.22349
12	2	0.21179	0.21166
13	2	0.20186	0.20133
14	2	0.19257	0.19222
15	2	0.18399	0.18411
16	2	0.17677	0.17683
17	2	0.17018	0.17026
4	3	0.12500	0.16059
5	3	0.12124	0.11964
6	3	0.09708	0.09406
7	3	0.07764	0.07675
8	3	0.06430	0.06435
9	3	0.05601	0.05509
10	3	0.04833	0.04794
11	3	0.04263	0.04228
12	3	0.03799	0.03769
13	3	0.03410	0.03392
14	3	0.03076	0.03076
15	3	0.02802	0.02808
16	3	0.02557	0.02579
17	3	0.02345	0.02381

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 360	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Random Sequential Coding By Hamming Distance		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Yoshiaki Itoh and Herbert Solomon		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 411SP		12. REPORT DATE July 11, 1985
		13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Sequential random packing, Cubic packing, Coding theory. gam sub d+1/gamma sub d

gamma sub d d

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Simple cubic random packing and random packing by Hamming distance are analyzed. Consider the packing density γ_d of dimension d by cubic random packing. From simulations up to dimension 11, γ_{d+1}/γ_d seems to approach one. Simulation results for random packing by Hamming distance are obtained. 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.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

END

FILMED

9-85

DTIC