NWC TP 4394

# RANDOM BINARY SEQUENCES

by

W. O. Alltop

Research Department

ABSTRACT. Two coding problems are considered. One is to obtain binary codes with desirable auto-correlation functions. The other is to obtain families of such codes with desirable cross-correlation functions. The results of a random computer search for the codes are described. An approximation to the expected number of trials in a search for families of codes is also presented.

NAVAL WEAPONS CENTER
CHINA LAKE, CALIFORNIA * AUGUST 1967

NAVAL WEAPONS CENTER
AN ACTIVITY OF THE NAVAL MATERIAL COMMAND

G. H. Lowe, Capt., USN ........................................ Commander
H. G. Wilson ........................................ Technical Director (Acting)

## FOREWORD

This report describes the results of a random computer search for binary codes. The search was motivated by consultation with T. A. Westaway of the Electronic Systems Branch, Systems Development Department, on the problem of obtaining binary codes of length 100 with desirable correlation properties. In anticipation of future problems in this area, data were obtained on codes of other lergths.

The work reported, as well as the preparation of the report, was supported by the Naval Air Systems Command Task Assignment A-04-535-207/216-1/F099-05-02.

This report is released at the working level. Because of the continuing nature of the research coding problem, refinements and modifications may later be made in this study.

Released by
D. E. ZILMER, Head
Mathematics Division
20 July 1967

Under authority of
HUGH W. HUNTER, Head
Research Department

NWC Technical Publication 4394

Published by............................Research Department
Collation......Cover, 3 leaves, DD Form 1473, abstract cards
First printing........................95 unnumbered copies
Security classification.........................UNCLASSIFIED

## INTRODUCTION

In electronic communications problems one is sometimes interested in obtaining finite length sequences of 1's and -1's with small auto- and cross-correlation functions. One way of acquiring such sequences, or families of such sequences, is to generate them randomly on a high-speed digital computer recording the desirable ones and discarding the rest. This report describes the results of such a random search with emphasis on sequences of length 100.

## THE SET OF SEQUENCES

Let $X_n$ be the set of all $2^n$ sequences of 1's and -1's of length $n$. Let $x$ denote the sequence $(x_1, x_2, \ldots, x_n)$. For $x, y \in X_n$ let

$$c_k(x,y) = \begin{cases} \sum\{x_i y_{i-k+n} : 1 \leq i \leq k\} & \text{if } 1 \leq k \leq n \\ \sum\{x_{i+k-n} y_i : 1 \leq i \leq 2n-k\} & \text{if } n+1 \leq k \leq 2n-1 \end{cases} \quad (1)$$

For $1 \leq k \leq n$, $c_k(x,y)$ is the value of the correlation of the left-most $k$ elements of $x$ with the right-most $k$ elements of $y$. For $n+1 \leq k \leq 2n-1$, $c_k(x,y)$ is the value of the correlation of the right-most $2n-k$ elements of $x$ with the left-most $2n-k$ elements of $y$.

The $(2n-1)$-tuple $c(x,y) = (c_1(x,y), c_2(x,y), \ldots, c_{2n-1}(x,y))$ of integers is called the cross-correlation function of $x$. The following integer-valued functions on $X_n \times X_n$ will be of interest.

$$M(x,y) = \begin{cases} \max\{|c_k(x,y)| : 1 \leq k \leq 2n-1\} & \text{if } x \neq y, \\ \max\{|c_k(x,y)| : 1 \leq k \leq 2n-1, k \neq n\} & \text{if } x = y \end{cases} \quad (2)$$

For simplification we let $m(x) = M(x,x)$. Estimates of $\mu_m$, the mean of the function $m$, and $\sigma_m$, the standard deviation of $m$ for values of $n$ ranging from 10 to 1,000 were obtained by random sampling in the spaces $X_n$. Table 1 lists for each $n$ the mean and standard deviation of the sample from $X_n$, as well as a measure of the difference of the mean from $n^{2/3}$. It is interesting to note how closely $n^{2/3}$ approximates the sample mean for $n$ between 40 and 300.

TABLE 1.

| $n$ | Sample size | $\mu_m$ | $n^{2/3}$ | $\frac{\mu_m^{-2/3}}{n^{2/3}}$ | $\sigma_m$ |
|---|---|---|---|---|---|
| 10 | 500 | 4.24 | 4.64 | -0.086 | 1.31 |
| 20 | 500 | 6.97 | 7.37 | -0.054 | 1.79 |
| 30 | 500 | 9.28 | 9.66 | -0.039 | 2.12 |
| 40 | 500 | 11.49 | 11.70 | -0.018 | 2.61 |
| 50 | 500 | 13.54 | 13.57 | -0.002 | 2.86 |
| 60 | 500 | 15.49 | 15.33 | 0.010 | 3.25 |
| 70 | 500 | 17.05 | 16.99 | 0.004 | 3.42 |
| 80 | 500 | 18.54 | 18.57 | -0.002 | 3.67 |
| 90 | 500 | 20.17 | 20.08 | 0.004 | 4.00 |
| 100 | 500 | 21.53 | 21.54 | 0.000 | 4.05 |
| 110 | 500 | 22.91 | 22.96 | -0.002 | 4.32 |
| 120 | 500 | 24.49 | 24.33 | 0.007 | 4.30 |
| 130 | 500 | 25.69 | 25.66 | 0.001 | 4.53 |
| 140 | 500 | 27.22 | 26.96 | 0.010 | 4.58 |
| 150 | 500 | 28.32 | 28.23 | 0.003 | 4.74 |
| 200 | 500 | 33.74 | 34.20 | -0.013 | 5.07 |
| 300 | 500 | 44.19 | 44.81 | -0.014 | 6.28 |
| 500 | 300 | 60.02 | 63.00 | -0.047 | 7.68 |
| 1,000 | 200 | 91.69 | 100.00 | -0.083 | 11.35 |

## SELECTING SINGLE SEQUENCES

The first practical problem motivating this study was to find sequences $x \in X_{100}$ such that $m(x)$ is small. Generation of 500 random sequences indicated that the probability that $m(x) \leq 15$ is about 0.026. Among the sequences chosen none satisfied $m(x) \leq 10$. On the UNIVAC 1108 computer 500 members of $X_{100}$ can be chosen and the corresponding values of $m$ computed in a little less than one minute. If one is interested in obtaining only sequences $x$ such that $m(x)$

is less than or equal to some fixed bound $b$, then one discards the sequence $x$ immediately after computing the first $c_k(x,x)$ such that $|c_k(x,x)|$ exceeds $b$. This eliminates unnecessary calculation of the $c_k(x,x)$ for unacceptable sequences thus reducing the running time on the computer. Of course, if $b_1 \leq b_2$, then the number of sequences satisfying $m(x) \leq b_1$ will be less than the number satisfying $m(x) \leq b_2$. Table 2 shows the results of computer runs selecting sequences with upper bounds of 13, 15, and 20 for $m(x)$ in the set $X_{100}$. In the case of the 80 sequences found with an upper bound of 13, none had upper bound ≤ 10.

TABLE 2.

| Bound | No. of sequences tested | No. of good sequences | Running time |
|---|---|---|---|
| 13 | 30,000 | 80 | 10 min, 31 sec |
| 15 | 5,000 | 150 | 3 min, 1 sec |
| 20 | 900 | 374 | 2 min, 48 sec |

## SELECTING FAMILIES OF SEQUENCES

The second practical problem under consideration was to obtain a family $x^{(1)}, x^{(2)}, \ldots, x^{(f)}$ of sequences from $X_{100}$ such that $M(x^{(i)},x^{(j)})$ is less than some bound $b$ for $1 \leq i \leq j \leq f$. Here one is confronted with the problem of determining values of $f$ and $b$ for which the computer will generate a solution in a reasonable length of time.

Suppose the probability that $m(x) \leq b$ is $p$ for a random sequence $x$. The value $m(x)$ is the maximum of the first 99 values of $|c(x,x)|$ since

$$c_k(x,x) = c_{200-k}(x,x) \tag{3}$$

i.e., the function $c(x,x)$ is symmetric about $k = 100$. However, for $M(x,y)$ to be less than or equal to $b$, one must have $|c_k(x,y)| \leq b$ for $1 \leq k \leq 199$. If one assumes that the functions $c_1, c_2, \ldots, c_{199}$ are statistically independent, then the probability that $M(x,y) \leq b$ for two random sequences $x$, $y$ is approximately $p^2$.

Now suppose one wishes to select randomly a family $x^{(1)}, x^{(2)}, \ldots, x^{(f)}$ of sequences from $X_{100}$ as described above. First one chooses sequences until a sequence $x^{(1)}$ satisfies $m(x^{(1)}) \leq b$. Next one chooses sequences until a sequence $x^{(2)}$ satisfies $M(x^{(1)},x^{(2)}) \leq b$ and $m(x^{(2)}) \leq b$. Continuing recursively one chooses at stage $r$ a sequence $x^{(r)}$ satisfying

$$\begin{aligned} &M(x^{(1)}, x^{(r)}) \leq b \\ &M(x^{(2)}, x^{(r)}) \leq b \\ &\qquad\vdots \\ &M(x^{(r-1)}, x^{(r)}) \leq b \\ &m(x^{(r)}) \leq b. \end{aligned} \tag{4}$$

Having chosen sequences $x^{(1)}, x^{(2)}, \ldots, x^{(r-1)}$ the probability that a randomly chosen sequence $x$ satisfies the restrictions of Eq. 4 on $x^{(r)}$ is approximately

$$\underbrace{p^2p^2\ldots p^2}_{r-1}p = p^{2r-1}.$$

It should be noted here that $(x^{(1)}, \ldots, x^{(r-1)})$ is not actually the result of a random selection from the cross-product space

$$X_{100} \times X_{100} \times \ldots \times X_{100}$$

since the family $x^{(1)}, \ldots, x^{(r-1)}$ already satisfies inequalities similar to Eq. 4. Thus, $p^{2r-1}$ is possibly a poorer approximation than it might seem.

Suppose $x^{(r)}$ is the $N_r$th sequence tested after $x^{(r-1)}$. Then $N_r$ is a random variable, and $p^{1-2r}$ approximates the mean of $N_r$. Let $N$ denote the number of sequences tested in order to select the entire family $x^{(1)}, x^{(2)}, \ldots, x^{(f)}$. $N$ is also a random variable, and

$$N = N_1 + N_2 + \ldots + N_f. \tag{5}$$

Since the mean of $N$ is the sum of the means of the $N_i$, we have the following approximation to $\mu_N$, where $\mu_i$ is the mean of $N_i$.

$$\begin{aligned}\mu_1 + \ldots + \mu_f &= p^{-1} + p^{-3} + \ldots + p^{1-2f} \\ &= p^{-1}(1+p^{-2}+\ldots+p^{-2(f-1)}) \\ &= p^{1-2f}\left(\frac{1-p^{2f}}{1-p^2}\right). \qquad (6)\end{aligned}$$

Since $p^{2f}$ will in general be very small, we have

$$\mu_N \sim p^{1-2f}(1-p^2)^{-1}. \qquad (7)$$

Table 3 shows estimates of $\mu_N$ based on approximation (Eq. 7) for three values of b. In the table p is the probability that $m(x) \leq b$ as estimated by a sample of 500 members of $X_{100}$.

TABLE 3.

| b | $p_b$ | N |
|---|---|---|
| 15 | 0.026 | $9.25 \times 10^{45}$ |
| 20 | 0.464 | $5.98 \times 10^{8}$ |
| 25 | 0.840 | $5.34 \times 10^{2}$ |

On the basis of Table 3 it was decided to place the bound b for families of sequences from $X_{100}$ at 25. Table 4 shows the results of ten computer runs each selecting a family of 15 sequences with auto- and cross-correlation bounded by 25. At the bottom of Table 4 the means of the sample N's and running times are given. In view of the various assumptions made in approximating N in Table 3 the difference between the approximation 534 for $b = 25$ in Table 3, and the sample $\mu_N$ of 1,020 in Table 4 cannot be considered too surprising.

TABLE 4.

| N | Running time |
|---|---|
| 1,582 | 9 min 55 sec |
| 971 | 6 min 34 sec |
| 563 | 3 min 55 sec |
| 934 | 6 min 6 sec |
| 1,678 | 10 min 22 sec |
| 824 | 5 min 18 sec |
| 1,561 | 9 min 26 sec |
| 798 | 5 min 7 sec |
| 919 | 5 min 58 sec |
| 368 | 2 min 41 sec |
| 1,020 | 6 min 32 sec |

## DOCUMENT CONTROL DATA - R&D

*(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)*

1. ORIGINATING ACTIVITY *(Corporate author)*

Naval Weapons Center
China Lake, California 93555

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

None

3. REPORT TITLE

RANDOM BINARY SEQUENCES

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Research Report

5. AUTHOR(S) *(Last name, first name, initial)*

Alltop, W. O.

6. REPORT DATE

August 1967

7a. TOTAL NO. OF PAGES

6

7b. NO. OF REFS

0

8a. CONTRACT OR GRANT NO.

b. PROJECT NO.

c. AIRTASK A-05-535-207/216-1/
F099-05-02

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

NWC TP 4394

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. AVAILABILITY/LIMITATION NOTICES

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Naval Air Systems Command
Naval Material Command
Washington, D. C. 20360

13. ABSTRACT

Two coding problems are considered. One is to obtain binary codes with desirable auto-correlation functions. The other is to obtain families of such codes with desirable cross-correlation functions. The results of a random computer search for the codes are described. An approximation to the expected number of trials in a search for families of codes is also presented.

UNCLASSIFIED

Security Classification

| 14. KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|---|---|---|---|---|---|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| Binary coding | | | | | | |
| Random coding | | | | | | |

## INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or *in* the report. Enter *last* name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

__________________________________ ."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

__________________________________ ."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

__________________________________ ."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

UNCLASSIFIED

Security Classification