# Random Number Generation by Use of de Bruijn Sequence

*Hiroshi Harada*, Hiroshi Kashiwagi*, Kazuo Oguri***

* *Faculty of Engineering, Kumamoto University, Kumamoto, Japan*
** *Nippon Electric Company, 1-10, Nisshin-cho, Fuchu, Japan*

Abstract: This paper proposes a new method for generation of uniform random numbers using binary random sequences. These binary sequences are obtained from a de Bruijn sequence by random sampling method. Several statistical tests are carried out for the random numbers generated by the proposed method, and it is shown that the random numbers have good random properties.

## 1. Introduction

Random numbers, which have a uniform distribution function, are used in many kinds of application, such as computer simulations or Monte Carlo method, etc.. A subroutine for generating uniform random numbers (URN) is considered to be the most useful library program in a computer system.

The most popular method of URN generation was introduced by D.H.Lehmer in 1949. The method is called a linear congruential method, and a sequence of URN $\{X_i\}$ ( $i=0,1,2,\cdots$) is obatined by the next equation,

$$X_{i+1} = \lambda \cdot X_i + \mu \quad (\bmod p)$$

where $\lambda$, $\mu$ and $p$ are positive integers. However, the linear congruential method has the following defects.

(i) The generated URN fall mainly in the hyper-planes and do not distribute uniformly in n-dimensional space [1].

(ii) Let $L$ be the word length of a computer. The maximum length of the sequence of URN generated by this method is equal to $2^L$. If $L$ is small, the sequence length of URN becomes too short to be used in applications.

The defect (ii) can be solved by use of double precision integers, but the defect (i) cannot be solved by choosing $\lambda$, $\mu$ and $p$. Therefore, some other URN generator is required.

In this paper, a new method is proposed for genarating URN using binary random sequence, which is obtained from a de Bruijn sequence. The method for genarating the binary random sequence is described in section 2. In section 3, a new URN generator using the binary random sequences is proposed. The result of several statistical tests carried out for the URN generated by this method is shown in section 4.

## 2. Generation of binary random sequence

A new method has been proposed for generating binary random sequences based on an M-sequence [2],[3],[4]. The method is to sample an M-sequence randomly by use of a sequence of random numbers. However, the randomly sampled M-sequence is not suitable for genaration of URN. The reason is as follows. Let $\{a_i\}$ denote an $n$-th degree M-sequence and $\{q_i\}$ be the randomly sampled M-sequence. The period of $\{a_i\}$ is given by the next equation.

$$N_M = 2^n - 1$$

The number of 1's in a period of $\{a_i\}$ is equal to $2^{n-1}$ and that of 0's is equal to $2^{n-1} - 1$. Therefore, in a period of an $n$-th degree M-sequence, the probability that $a_i = 1$ is greater than that of $a_i = 0$.

$$\Pr\{a_i = 1\} = 2^{n-1}/N_M > (2^{n-1} - 1)/N_M = \Pr\{a_i = 0\}$$

Since $\{q_i\}$ is obtained from $\{a_i\}$ by random sampling, the sequence of URN generated from $\{q_i\}$ does not have good random properties.

In this paper, a de Bruijn sequence, which is derived from an M-sequence, is chosen as the original binary sequence. An $n$-th degree de Bruijn sequence can be generated by adding one 0 to the longest run of 0 in a period of an $n$-th degree M-sequence [5]. Then, the number of 1's in a period of a de Bruijn sequence is equal to that of 0's.

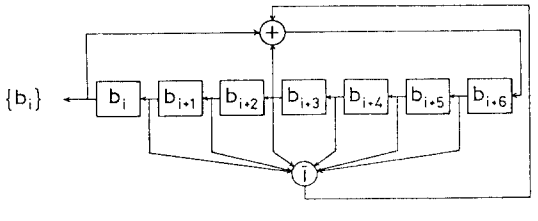



Fig.1 A circuit for generation of de Bruijn sequence $\{b_i\}$ when the characteristic polynomial is $f(x) = x^7 + x^3 + 1$

⊕ : mod 2 adder

ⓘ : NOR gate

□ : shift register

An example of the circuit for generating de Bruijn sequence is shown in Fig.1, where the characteristic polynomial of the sequence is written as eqn. (1).

$$f(x) = x^7 + x^3 + 1 \qquad (1)$$

Let $\{b_i\}$ denote the $n$-th degree de Bruijn sequence and $N$ be the period of $\{b_i\}$ .

$$\{b_i\} = b_0, b_1, \cdots, b_{N-1}$$

$$N = 2^n$$

The random sampling method is as follows. First, successive $k$-tuple of $\{b_i\}$ is generated,

$$\boldsymbol{b}_{ki} = ( b_{ki}, b_{ki+1}, \cdots, b_{ki+k-1} )^T$$

where $^T$ denotes transposition. Then, a random number $X_i$ ( $i$ =0,1,2,----), which is distributed uniformly between 0 and 1, is generated by the linear congruential method or some other methods. Using $X_i$ , the ( $[k \cdot X_i]$+1 )-th element of $\boldsymbol{b}_{ki}$ is chosen. Here, $[k \cdot X_i]$ is the maximum integer less than $k \cdot X_i$ . In this case, when the tuple length $k$ is longer, the random properties of the obtained binary sequence becomes good. In this paper, the tuple length $k$ is equal to the period of the de Bruijn sequence. Then, the $N$-tuple $\boldsymbol{b}$ is given as

$$\boldsymbol{b} = ( b_0, b_1, \cdots, b_{N-1} )^T \qquad (2)$$

Let $\{r_i\}$ be the randomly sampled sequence generated by this method, then $\{r_i\}$ is expressed by the original binary sequence $\{b_i\}$ as

$$\{r_i\} = b_{[N \cdot X_0]}, b_{[N \cdot X_1]}, \cdots, b_{[N \cdot X_i]}, \cdots$$

An example of the autocorrelation function (ACF) of the original de Bruijn sequense $\{b_i\}$ and the ensemble averaged autocorrelation function (EACF) of the randomly sampled sequence $\{r_i\}$ are shown in Fig.2 (a) and (b), respectively. In this paper, ACF of $\{b_i\}$ and $\{r_i\}$ are defined by the next equations.

$$\phi_{bb}(\tau) = \frac{1}{N} \sum_{i=0}^{N-1} (-1)^{b_i} \cdot (-1)^{b_{i+\tau}}$$

$$\phi_{rr}(\tau) = \frac{1}{N} \sum_{i=0}^{N-1} (-1)^{r_i} \cdot (-1)^{r_{i+\tau}}$$

From Fig.2(b), it is clear that EACF of $\{r_i\}$ has a sharp peak at delay 0, while at the other delays it is almost equal to 0. Therefore, the binary random sequence generated by the proposed method has good random properties.

## 3. URN generator based on the binary random sequence

Tausworthe proposed the method for generating URN based on M-sequence [6]. A sequence $\{x_i\}$ ( $i$ =0,1,2,--) of URN is constructed from M-sequence $\{a_i\}$ by the next equation.

$$x_i = 0.a_{iq+1}\ a_{iq+2} \cdots a_{iq+L} \quad \text{(base 2)}$$

where $L$ is the bit length of URN and $q$ is a positive integer satisfying the next conditions.

$$q \geq L$$

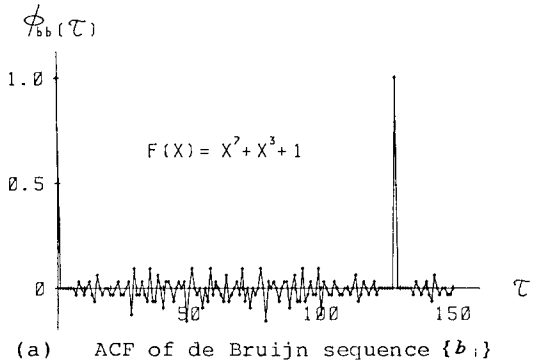
$$\text{gcm}\ (q, 2^n - 1) = 1$$

The period of the sequence $\{x_i\}$ called Tausworthe sequence is equal to that of the original M-sequence, and $x_i$ ( $i$ =0,1,2,--) distribute uniformly in $[n/L]$-dimensional space, where $n$ is the degree of M-sequence.

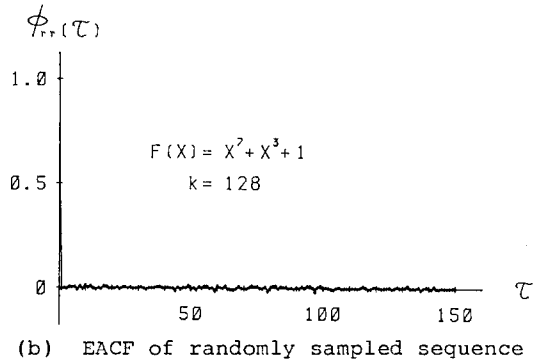
The method proposed in this paper uses the binary random sequence as is stated in section 2 in stead of M-sequence. Let $\{w_i\}$ denote a sequence of URN generated by the proposed method. Successive $L$ bits of the sequence $\{r_i\}$ are placed from the most significant bit of the element of $\{w_i\}$ to the least significant bit. Then, the $i$-th element of the URN sequence is represented as

$$w_i = 0.r_{Li}\ r_{Li+1} \cdots r_{Li+L-1} \quad \text{(base 2)}$$

Here, $L$ is the bit length of the random number and is equal to 16 in this paper.




(a) ACF of de Bruijn sequence $\{b_i\}$




(b) EACF of randomly sampled sequence $\{r_i\}$

Fig.2 ACF of $\{b_i\}$ and EACF of $\{r_i\}$ when the characteristic polynomial is $f(x) = x^7 + x^3 + 1$

## 4. Result of statistical test

Several statistical tests are applied to the URN generated by the proposed method to show that the URN have good random properties. The method is as follows. The statistical tests [7] are shown in Table 1.

(i) A sequence $\{w_i\}$ of URN is generated by the proposed method and a chi-square value ($\chi^2$ value) for each test is calculated using the theoretical distribution and the observed distribution obtained from $\{w_i\}$. The number of samples is equal to 1000.

(ii) After calculating 100 $\chi^2$ values for each test, Kolmogorov-Smirnov test (K-S test) [7] is carried out. The values $K^{+}_{100}$ and $K^{-}_{100}$ are calculated using the next equations.

$$K^{+}_{100} = 10 \cdot \max(F_{100}(x) - F(x))$$

$$K^{-}_{100} = 10 \cdot \max(F(x) - F_{100}(x))$$

Here, $F(x)$ is the theoretical $\chi^2$ distribution function and $F_{100}(x)$ is the $\chi^2$ distribution function obtained from 100 $\chi^2$ values. If the values $K^{+}_{100}$ or $K^{-}_{100}$ are greater than 1.34, it is considered that these 100 $\chi^2$ values are not from $\chi^2$ distribution function and the hypothesis that $\{w_i\}$ is a random number sequence is rejected under the significance level of 5%.

(iii) K-S test is repeated 100 times using various $\{w_i\}$ ,and the rejected number of $K^{+}_{100}$ and $K^{-}_{100}$ is counted for each statistical test.

URN sequences $\{w_i\}$ are generated using various de Bruijn sequence $\{b_i\}$ and the statistical tests are carried out. The result of K-S test is shown in Table 2. Here, the degree $n$ of the de Bruijn sequence and the charactersitic polynomial $f(x)$ are

$$n = 5, \quad f(x) = x^5 + x^2 + 1 \qquad (3)$$

$$n = 6, \quad f(x) = x^6 + x + 1$$

$$n = 7, \quad f(x) = x^7 + x^3 + 1$$

The numbers in the table denote the rejected numbers of $K^{+}_{100}$ and $K^{-}_{100}$ values among 100 K-S tests. In this case, the significance level is 5%, so it is desirable that the numbers in the table are neary equal to 5. From this point of view, when the degree of the de Bruijn sequence is 5, the results of K-S test are not good concerning combination test. It is obvious that the URN sequences do not show good random properties even if the degree of the original de Bruijn sequence becomes higher. The reason is as follows. From eqn. (2), the $N$-tuple used for generating the randomly sampled sequence is constant. Therefore, if the uniformity of the distribution function of $\{X_i\}$ is slightly bad, the random properties of the generated URN becomes worse.

To solve this defect, three methods for changing $N$-tuple $b$ are tested.

(i) Method A

Method A is to change 1's in the $N$-tuple to 0's or 0's to 1's, every time one element of the randomly sampled sequence is generated. The $N$-tuples $b_{2i}$ and $b_{2i+1}$ are expressed as

$$b_{2i} = (b_0, b_1, \cdots, b_{N-1})^T$$

$$b_{2i+1} = (\bar{b}_0, \bar{b}_1, \cdots, \bar{b}_{N-1})^T$$

Table 1 Statistical tests

| tests | frequency | serial (2nd) | serial (3rd) | combination | run (above/below) | run (up/down) | poker |
|---|---|---|---|---|---|---|---|
| degrees of freedom | 15 | 15 | 63 | 10 | 7 | 3 | 3 |

Table 2 Results of K-S test when the significance level is 5%

| tests / degrees | frequency | serial (2nd) | serial (3rd) | combination | run (above/below) | run (up/down) | poker |
|---|---|---|---|---|---|---|---|
| 5 | 6 | 3 | 4 | 9 | 5 | 3 | 2 |
| 6 | 5 | 10 | 8 | 5 | 6 | 5 | 5 |
| 7 | 8 | 9 | 17 | 7 | 5 | 2 | 4 |

Table 3 Results of K-S test when the significance level is 5%

| tests / methods | frequency | serial (2nd) | serial (3rd) | combination | run (above/below) | run (up/down) | poker |
|---|---|---|---|---|---|---|---|
| A | 6 | 3 | 4 | 8 | 5 | 10 | 2 |
| B | 8 | 5 | 3 | 9 | 3 | 5 | 1 |
| C | 8 | 4 | 4 | 4 | 2 | 7 | 1 |

Table 4 Results of K-S test when the significance level is 5%

| tests / methods | frequency | serial (2nd) | serial (3rd) | combination | run (above/below) | run (up/down) | poker |
|---|---|---|---|---|---|---|---|
| oroginal | 2 | 32 | 9 | 3 | 100 | 100 | 100 |
| proposed | 2 | 5 | 1 | 9 | 5 | 4 | 6 |

where $\bar{}$ denotes bit inverse operation.

(ii) Method B

In method B, the tuple length $k$ is equal to a half of the period of de Bruijn sequence. Then, the $k$-tuples $\boldsymbol{b}_{2i}$ and $\boldsymbol{b}_{2i+1}$ are given as

$$\boldsymbol{b}_{2i} = ( b_0, b_1, \cdots, b_{N/2-1} )^T$$
$$\boldsymbol{b}_{2i+1} = ( b_{N/2}, b_{N/2+1}, \cdots, b_{N-1} )^T$$

In this case, $\boldsymbol{b}_{2i}$ and $\boldsymbol{b}_{2i+1}$ are chosen on condition that the number of 1's in $\boldsymbol{b}_{2i}$ is equal to that in $\boldsymbol{b}_{2i+1}$.

(iii) Method C

Method C is to shift the de Bruijn sequence by one bit, every time the random sampling is carried out. Then the $N$-tuple $\boldsymbol{b}_i$ becomes

$$\boldsymbol{b}_i = ( b_i, b_{i+1}, \cdots, b_{i+N-1} )^T$$

Here, the subscripts are considered as mod $N$.

Using these methods, the sequences of URN are generated and the results of the statistical tests are shown in Table 3. In this case, the degree of de Bruijn sequence and the characteristic polynomial are the same as eqn. (3). Comparing these results, URN generated by method A and method B don't show good properties in run (up/down) test and in combination test, respectively, while URN generated by method C have good random properties. Therefore, it is shown that the method C is suitable for generating URN.

Another result of statistical test is shown in Table 4. In this case, the original URN $\{X_i\}$ is generated by the next equation [8].

$$X_i = \begin{cases} X_{i-1}+X_{i-2}+X_{i-3}+1357 & (X_{i-2} < 5\cdot 10^7) \\ X_{i-1}+X_{i-2}+X_{i-3} & (X_{i-2} \geq 5\cdot 10^7) \end{cases}$$

From Table 4, the original URN $\{X_i\}$ does not distribute uniformly in a high dimensional space. However, the URN generated by the proposed method shows good random properties.

## 5. Conclusion

A new method is proposed for generation of uniform random numbers using binary random sequence. The binary random sequence is generated from de Bruijn sequence by random sampling method and has good random properties. Several statistical tests are carried out for the random numbers generated by the proposed method. From the results of the tests, it is shown that these random numbers have good properties even when the sequence used for random sampling have poor random properties.

## Reference


[1] G.Marsaglia: Random Number Fall Mainly in the Planes, Proc. Nat. Acad. Sci., 61, 25/28 (1968)

[2] H.Harada, H.Kashiwagi, H.Honda and K.Oguri: Binary Random Sequence generation by Use of Random Sampling of M-sequence, Proc. '87 KACC., 832/835 (1987)

[3] H.Harada, H.Kashiwagi, H.Honda and K.Oguri: On Correlation Function of Randomly Sampled M-Sequence, Trans. SICE, 23-11, 1145/1150 (1987) (in Japanese)

[4] H.Harada, H.Kashiwagi, H.Honda and K.Oguri: On Some Properties of Randomly Sampled M-Sequence, Trans. SICE, 24-8, (1988) (in Japanese)

[5] G.Hoffmann de Visme: Binary Sequences, English Universities Press, (1971)

[6] R.C.Tausworthe: Random Numbers Generated by Linear Recurrence Modulo Two, Math. Comp., 19, 201/209 (1965)

[7] D.E.Knuth: The Art of Computer Programming Vol.2 Seminumerical Algorithms, Addison-Wesley (1981)

[8] C.G.Swain and M.S.Swain: A Uniform Random Number Generator that is Reproducible Hardware-independent and Fast, J. Chem. Inf. Sci., 20-1, 56/58 (1980)