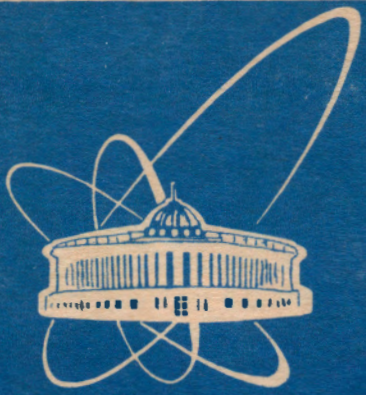

ОБЪЕДИНЕННЫЙ
ИНСТИТУТ
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ

Дубна

E11-95-198

G.A.Ososkov, E.A.Tikhonenko

NEW RANDOM NUMBER GENERATOR ON THE BASE OF 2D-CELLULAR AUTOMATON

Submitted to «Математическое моделирование»

1995

1 Introduction

For a software development in theoretical and experimental physics it is important to have rapid and reliable random numbers generators (RNG). The experience of previous efforts in this field [1] shows an obvious advantage of random number generators working in a group generating mode. The last assertion leads to an idea of the usage of cellular automata (CA).

A well-known paper of S.Wolfram [2] gives a detail example of RNG on the base of one-dimensional CA implementing a pseudorandom Fibonacci sequence. In this way the authors of given paper decided to use more extended possibilities of two-dimensional CA, desribed in section 2. The difficulties of a pure analytical way to study RNG properties were mentioned yet in [2]. It causes to apply statistical testing for determinating both a generated sequences period and its statistical properties. The latters usually includes investigations of uniformity of multi-dimensional distribution, correlation analysis etc. It demands to develop an adequate set of software tools for a confidential testing of obtained sequences. Besides of well-known statistical tests like monotony and gap tests [3] the authors propose in section 3 a novel easy-to-use and powerfull method for testing the uniformity of multidimentional random vectors distribution (with dimension up to 20 and more) called the imbedded histogram method.

Results of the study of proposed methods for generating of multidimensional random sequences by two-dimensional binary CA (TBCA) are given.

2 Cellular automata as random number generators

According to the classical T.Toffoli work [4]:"Cellular automata are discrete dynamic systems which behaviour is determined in terms of local dependencies".

The evolution of CA occurs in discrete spaces consisting of cells. Evolution laws are local i.e. system dynamics is given by an unchanged set of rules, by which a new state of cells is calculated in dependence of states of its neighbours. It is essential that this change of state occurs simultaneously and time is clocked.

Despite of the simplicity of their construction, CA can be capable of diverse and complex behaviour [5], which gives a possibility to use CA in simulating of nature systems and physical processes [6,7]. In particular CA can be used for generating of random numbers.

In this paper we propose one way of such TBCA applications. In the general case two-dimensional CA consists of $N * M$ cells a_{ij} filling $N*M$ matrix. The states of cells are integers between 0 and $k-1$. These values are updated in a parallel mode in discrete steps according to the following rule:

$$a'_{ij} = \phi \begin{pmatrix} a_{i-q,j-r}, & a_{i-q+1,j-r}, & \cdots & a_{i+q,j-r}, \\ a_{i-q,j-r+1}, & a_{i-q+1,j-r+1}, & \cdots & a_{i+q,j-r+1}, \\ \vdots & \vdots & \vdots & \vdots \\ a_{i-q,j+r}, & a_{i-q+1,j+r}, & \cdots & a_{i+q,j+r} \end{pmatrix} \quad (1)$$

We consider the case of $k = 2$, $r = q = 1$, that means that our cellular automaton is binary and next states of each cell is defined by states of both: itsefl and its eight nearest neighbours.

We investigate the properties of random vectors generators on the basis of two TBCAs, which are defined according to the rules:

$$\phi_1(i,j) = \left(\sum_{i=i-1}^{i+1} \sum_{j=j-1}^{j+1} a_{ij} \right) mod\, 2 \quad (2)$$

and

$$\phi_2(i,j) = \left(\sum_{i=i+1}^{i+1} \sum_{j=j-1}^{j+1} a_{ij} + a_{ij} * a_{i+1,j+1} \right) mod\, 2 \quad (3)$$

3 The method of imbedded histograms as a randomness test

The distribution unuformity of d-dimensional random vectors is a necessary and sufficient condition for a successful realization of Monte-Carlo algorithms with a dimension of d [8]. The obtained sequences of integer random numbers can be normalized by dividing each of these numbers by $\left(2^M - 1\right)$. The resulting numbers under some conditions supposed to be uniformly distributed in the range from 0 to 1 and can be considered

as d-dimensional vectors:

$$\begin{array}{lllll} \vec{X}_1 = & (x_1, & \ldots, & x_d) \\ \vec{X}_2 = & (x_{d+1}, & \ldots, & x_{2d}) \\ \ldots & \ldots & \ldots & \ldots \\ \vec{X}_n = & (x_{d(n-1)}, & \ldots, & x_{nd}) \\ \ldots & \ldots & \ldots & \ldots \end{array} \quad (4)$$

By splitting a single multidimensional cube to equal volumes (bins) we can find a frequency of random vectors hits in each of these bins. If binning to m is made for each dimension we obtain m^d equal volumes. Such the conventional approach to study statistical characteristics of multidimentional histograms is too time-consuming and requires computer memory growing as m^d .

We propose instead a substitution of this cumbersome way of multidimensional histogramming by a simple elegant procedure [9] which allows to split a cube of any dimension into m equal bins. A k-th bin $(k = 1, 2, \ldots, m)$ is located between two d-dimensional cubes with the lengths of sides equal correspondingly to $a_{k-1} = \sqrt[d]{\frac{k-1}{m}}$ and $a_k = \sqrt[d]{\frac{k}{m}}$. (In other words it is located inside a cube with a side a_k, but outside a cube with a side a_{k-1}.) For more detailed checking of uniformity of multi-dimensional distribution it ought to make both: binning from each corner of a d - dimensional single cube and from its center. In this way we obtain $(2^d + 1)$ histograms, each with an unchanged number of historam cells m.

Each vector belongs to a k-th volume:

$$k = integer\{m * [max\,(x_1, \cdots, x_d)]^d\} + 1 \quad (5)$$

In order to test the uniformity of the distribution of n d-dimensional random vectors we can use well-known χ^2-criterium.

For each of k volumes we obtain a frequency of hit ν_k and then calculate the value of χ^2:

$$\chi^2 = \frac{m}{n}\sum_{k=1}^{m} \nu_k^2 - n, \quad (6)$$

It is convenient to illustrate the method of imbedded histograms by the case $d = 2$, $m = 4$.

Let us split a square $1 * 1$ (Fig.1) to 4 blocks with equal areas $(1/m = 1/4)$; the first one is a square $a_1 * a_1$ (accordingly, $a_1 = \sqrt{1/m}$), and others

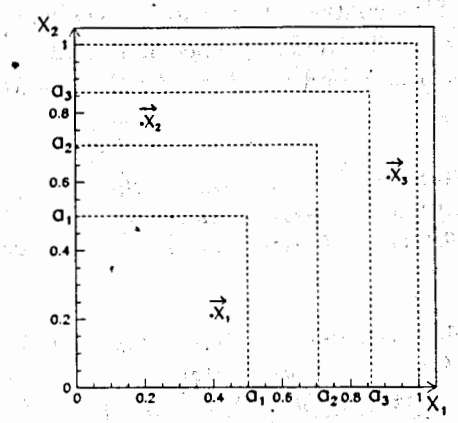

Figure 1: $2 - D$ example of the embedded histogramming.

are situated between 4 imbedded squares, in the general case:

$$a_k = \sqrt[d]{k/m},\ k = 1, \cdots, m \quad (7)$$

Accordingly, a hit into a k-th block is defined by the inequality:

$$a_k < max[x_1, x_2] < a_{k+1} \quad (8)$$

or, in the general case:

$$\sqrt[d]{k/m} < max[x_1, \cdots, x_d] < \sqrt[d]{(k+1)/m} \quad (9)$$

or

$$k < m * (max[x_1, \cdots, x_d])^d < k + 1, \quad (10)$$

from which the equation (5) follows.

4 Algorithms and program realization of the generator and statistical tests

Theoretically CAs are assumed as objects with infinite numbers of cells. However in any practical implementation it requires a finite number of

CA cells. This contradiction can be resolved on a way of creating of a cyclic structure by defining special boundary conditions. For the rule (1) for $r = q = 1$ and number of cells $N * M$ we can define boundary conditions as:

$$a'_{1j} = \phi(a_{N,j-1}, a_{1,j-1}, a_{2,j-1}, a_{N,j}, a_{1,j}, a_{2,j}, a_{N,j+1}, a_{1,j+1}, a_{2,j+1}) \qquad a'_{i1} = \phi(a_{i-1,M}, a_{i,M}, a_{i+1,M}, a_{i-1,1}, a_{i,1}, a_{i+1,1}, a_{i-1,2}, a_{i,2}, a_{i+1,2})$$

$$a'_{Nj} = \phi(a_{N-1,j-1}, a_{N,j-1}, a_{1,j-1}, a_{N_1,j}, a_{N,j}, a_{1,j}, a_{N-1,j+1}, a_{N,j+1}, a_{1,j+1}) \qquad a'_{iM} = \phi(a_{i-1,M-1}, a_{i,M-1}, a_{i+1,M-1}, a_{i-1,M}, a_{i,M}, a_{i+1,M}, a_{i-1,1}, a_{i,1}, a_{i+1,1}), \tag{11}$$

geometrically that means a simple convolution of a rectangle $N * M$ to a torus.

Computations were made for TBCA with various numbers of N of a matrix columns with a fixed number of rows $M = 31$. The content of each column is loading bit by bit into a 32-bit machine word, which is then normalized by dividing it by $\left(2^M - 1\right)$ after that. We choose $M = 31$ due to the following reasons: it is close to the word length of the majority of computers and it is a simple number that considerably increases a CA cycle length [3] . As the result of CA evolution we obtain N numbers in the range of (0,1).

While software implementing of formulae (2), (3) (11) it is important to use such properties of Fortran-90 as array assignment, masked array asignment and array sections. It gives an advantage for the program compilation with the vector optimization.

One of the main characteristics of RNG is its period, i.e. a coincidence of CA state 'B' with respect to any its previous state 'A'. It means an existence of a period because from state to state CA changes according a fixed rule. It ought to keep in a view that an initial conditions set-up can affect not only on a duration of a possible period but also on a moment of evolution of CA from which this periodicity begins. Due to that we develope a procedure of checking for a periodicity which takes in the account an influence of initial conditions and allows to set many control points (states of a matrix of CA which are cheked for a coincidence) with a varying step.

One of neccessary requirements to RNG is the absence of a correlation between obtained random numbers. In this way calculations of a coefficient of correlation were made according a formula:

$$\rho_k = \frac{\frac{1}{n-k}\sum_{i=1}^{n-k} x_i * x_{i+k} - \bar{x}^2}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}, \; where \; \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{12}$$

for $k = 1, 2, \ldots, 16$, where k is the distance between two numbers of a sequence which are tested.

5 Statistical Tests

Firstly we have investigated the properties of the generator built in accordance with the rule (2). As it was found this generator has comparatively short period. Only for lengths of a matrix column $N = 125, 150, 250, 500$ we determined that a period length is $> 10^8$. For other various column lengths we determined that a period length does not exceed 10^6. It ought to mention that a length of a period essentially depends on an initail state of a CA matrix.

Attempts to find a generator with a larger period led us to the generator based on the rule(3), which amplifies the rule (2) (called bellow as generator CARNG). The period of this generator is equal to $1,5 \cdot 10^7$ only for a CA matrix size $3 * 31$, starting from the special initial state of CA matrix with one non-zero row in a center of a matrix. For other cases being tested ($N = 2, 4, 5, \ldots, 10, 25, 50, 100$) a period length is sufficiently longer (at least $> 10^8$).

For both generators we have also satisfactory results on correlation checking.

Diagrams presented on Fig.2 demonstrate a good convergence of the coefficient correlation to zero with increasing of statistics.

The embedded histogram method shows the uniformity of distribution of obtained random vectors in spaces with dimension from 1 to 20.

Distributions of random vectors via splitting a single d- dimensional single cube into 50 equal bins are presented on Fig.3 (the quantity of obtained random numbers is equal $2 \cdot 10^6$).

It is known that if the degree of freedom is greater then 30 χ^2 should be normally distributed. Therefore 99% of χ^2's are located within 3σ limits, i.e., in our case, in the interval $(20, 80)$. The calculations carried out up to $2 \cdot 10^6$ random numbers show that the value of χ^2 keeps within this interval.

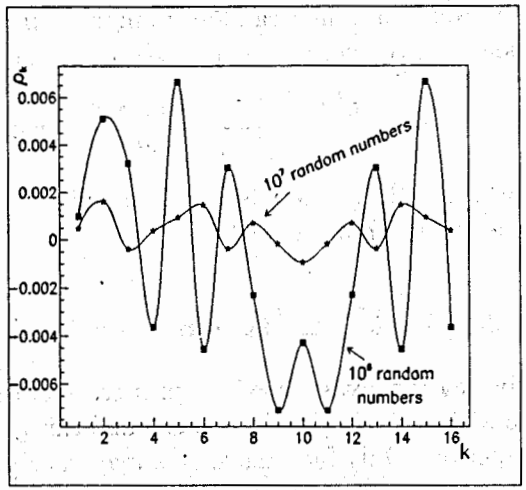

Figure 2: The values of correlation coefficient ρ_k for CARNG with a number of columns $N = 16$ of CA matrix for quantities of obtained numbers 10^6 and 10^7 accordingly.

As it was mentioned in [3], monotony test is one of the most powerfull RNG-properties tests. We have applied the modification of this test recommended in [3] when each element immediately following every serie of monotony is discarded to make successive runs statistically independent.

Classical gap test gives also quit satisfactory results (it is known that many generators fail on this test [10]).

From two described generators we prefer CARNG: it has quit satisfactory statistical properties, a long period and conforms to desirable properties of RNG given in [1].

6 Conclusions

The given results of testing of the proposed CARNG on the basis of 2D-cellular automaton show its quite suitable properties: multi-dimensional uniformity of random sequences, the absence of correlative links and quite satisfactory results on various statistical tests. It was also determined that its period is sufficiently long (at least $> 10^8$). In our opinion, as though the best features of this generator (due to its nature) can

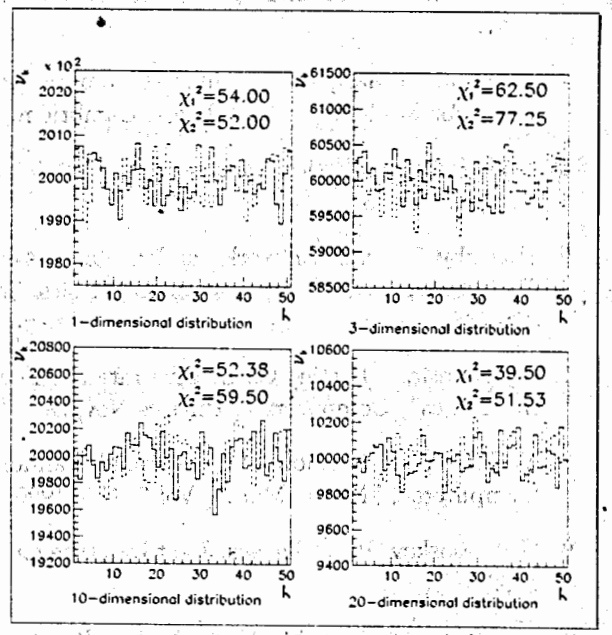

Figure 3: Frequency of hits of random vectors into $m = 50$ bins with dimensions $d = 1, 3, 10, 20$ for CARNG with a number of columns $N = 25$ of CA matrix (the solid line is for binning from a left low corner, the dashed line is for binning from a right top corner of a d-dimensional single cube).

be effectively achieved in its parallel realization as a highly integrated electronic chip, the CARNG implementation (vector or parallel) should be solved dependently of specifics of every concrete application.

This work was completed due to the support of the Russian Foundation for Fundamental Research.

REFERENCES

1. F.James "A Review of Pseudorandom Number Generators", Computer Physics Communications 60(1990), 329-344.

2. S.Wolfram "Ramdom Sequences Generation by Cellular Automata", Advances in Applied Mathematics 7, 123-169(1986).

3. D.E.Knuth "Seminumerical Algorithms", Vol.2, Addison-Wesley, Mass., 1981.

4. T.Toffoli, N.Margolus "Cellular Automata Machines: A New Environment For Modelling", MIT Press, Cambridge, Mass., 1987.

5. S.Wolfram "Statistical Mechanics of Cellular Automata", Rev.Modern Phys. 55 (1983).

6. B.Denby "Neural Networks and Cellular Automata in Experimential High Energy Physics", Computer Physics Communications 49(1988), 429-448.

7. B.Boghosian "Lattice Gases Illustrate the Power of Cellular Automata in Physics", Computers in Physics Nov/Dec, 1991, 585-590.

8. Yu.L.Levitan, I.M.Sobol "On a Pseudo-random Generator for Personal Computers", "Math. Mod.", Vol 2, N.8/1990.

9. G.A.Ososkov, Ph.D Thesys, Joint Institute for Nuclear Research, Dubna, 1986.

10. M.Lücher "A Portable High-Quality Random Number Generator for Lattice Field Theory Simulations", Computer Physics Communications 79(1994), 100-110.

Ососков Г.А., Тихоненко Е.А. E11-95-198

Новый генератор случайных чисел
на базе двумерного клеточного автомата

Опыт использования генераторов случайных чисел при моделировании (особенно в физике) показывает очевидные преимущества использования генераторов, вырабатывающих случайные векторы (группы независимых случайных чисел). Предложен новый эффективный метод генерации векторов случайных чисел на основе двумерного бинарного клеточного автомата. Проведены соответствующие проверки случайных последовательностей на периодичность, корреляционные свойства и равномерность многомерного распределения.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Препринт Объединенного института ядерных исследований. Дубна, 1995

Ososkov G.A., Tikhonenko E.A. E11-95-198

New Random Number Generator
on the Base of 2D-Cellular Automaton

The experience of generating of random numbers for various simulatings especially in physics shows the evident advantages of programmed random number generators emanating random vectors (groups of random numbers). In this way we proposed a new efficient method for generating of random vectors on the basis of two-dimensional binary cellular automaton. Due to its nature our method is suitable for the parallel implementation. The corresponding set of statistical criteria for testing the random vector sequence period, correlating properties and multi-dimensional distribution is developed.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Preprint of the Joint Institute for Nuclear Research. Dubna, 1995