High Performance Parallel Pseudorandom Number Generator on Cellular Automata

Alla Levina[a], Daniyar Mukhamedjanov[a], Danil Bogaevskiy[a], Pavel Lyakhov[b], Maria Valueva[c] and Dmitry Kaplun[a,*]

[a] *Saint Petersburg Electrotechnical University "LETI", Professora Popova str., 5, St. Petersburg, 197022, Russia*
[b] *Department of Applied Mathematics and Mathematical Modeling, North-Caucasus Federal University, 1 Pushkin str., Stavropol, 355000, Russia*
[c] *Department of Number Theoretic Systems, North-Caucasus Federal University, 1 Pushkin str., Stavropol, 355000, Russia*

ARTICLE INFO

Keywords:
Cellular Automata, Pseudorandom Number Generation, Parallel Computing, Hardware-based Parallel Implementation

ABSTRACT

Nowadays the practice of developing algorithms to maintain the confidentiality of data shows that there is a lack of some features, such as velocity, predictability, etc. Generating pseudorandom numbers is one such problem that lies in the basement of many algorithms, even in hardware microprograms. An unreliable generator can cause cyberattacks on it, despite the security in the upper layers. At the same time, the algorithm should be fast enough to provide uninterrupted circuit work for the entire system. The paper presents a new algorithm generating pseudorandom numbers on cellular automata, which is not only fast and easy-repeating, but unpredictable enough and can be used in cryptographic systems. Using the NIST statistical test suite for random and pseudorandom number generators (PRNG), it is shown that the presented algorithm is more than three times superior to the state-of-the-art methods and algorithms in terms of $P - value$. A high level of the presented algorithm's parallelization allows for implementation it effectively on calculators with parallel structure. CPU-based architecture, FPGA-based architecture, CUDA-based architecture of PRNG and different PRNG implementations are presented to confirm high performance of the proposed solution.

1. Introduction

There is a large interest in the world research society in the security and ways of its realization in both hardware and software engineering. Nowadays, all areas of IT need to be protected from cyberattacks, some attacks presented in (1; 2; 3; 4; 5). Building secure software on many levels, such as network, application, etc., gave protection from variable attacks and increased level of confidence. To protect all layers of systems can be used ciphers (6; 7; 8) and/or error-detecting/error-correcting codes, some of them are based on random numbers. For an example random numbers used in PIN and password generation (PIN Protection Principles, ANSI X9.8:1, Password Generation, FIPS 181-1993), generation of primes (DSA, ANSI X9.30, RSA, ANSI X9.31, Prime Number Generation, ANSI X9.80) random challenges for authentication (Entity Authentication using PKC, FIPS 196) and key confirmation (NIST Key Schemes Recommendation). Many systems must have the ability to process the same numbers again under the same conditions and with data of the same origin. For this needs can be used pseudorandom generator that can repeat the same numbers as much as it will be needed. There are many examples of using various algorithms, such as linear shift or Yarrow (or Fortuna) (9), which are very good for their options, and some of them are even cryptographically secure, but there is no ideal algorithm; all of them have their own disadvantages compared to others.

In this paper, we present a new algorithm for generating pseudorandom numbers, based on cellular automata, and illustrate its implementation on FPGA and CUDA. Compared to the existing algorithm, the presented algorithm is different from the previous and similar ones in several respects:

1. better results in some NIST test;
2. work speed;
3. parallelization in hardware implementation;
4. flexibility;

*Dmitry Kaplun
alla_levina@mail.ru (A. Levina); mdonic92@gmail.com (D. Mukhamedjanov); dvbogaevskiy@etu.ru (D. Bogaevskiy); ljahov@mail.ru (P. Lyakhov); mriya.valueva@mail.ru (M. Valueva); dikaplun@etu.ru (D. Kaplun)
ORCID(s):

5. simplicity of architecture.

The paper is organized as follows: the first section presents an overview of ideas used in random number generators, the mathematical structure of homogeneous structures (HS) and cellular automata, the second section illustrates created algorithm, gives a mathematical explanation of the presented algorithm, the third section demonstrates test results, in section IV illustrates the implementation of the created algorithm on FPGA and section V gives CPU and CUDA implementation of the algorithm. In conclusion, a short overview of the presented work is given.

2. Random number generators overview

Random numbers are needed in a variety of applications, yet finding good random number generators (RNG) is a difficult task. In recent years, some new results of building pseudorandom number generators (PRNGs) based on chaos and chaotic map were presented (10), (11) (12) (13) and in work (14).

Actually, 'true' random generation is mostly based on some kind of natural physical processes, or in another way, the 'noise' source features of some machine hardware may be used. Simplifying, there are many ways to determine some noises as "1" or "0", giving a sequence of bits, and as a result a number. This number is exactly what we need.

But the difference between these and pseudorandom generators is in uniqueness, which means that there is no chance of getting the same number more than once. The great part of computations where a random number should be used is connected with cryptography. For simplicity, there is a need to control the calculation of the speed, effectiveness, and repetition generated numbers (PRNG). Some examples to understand the general idea of such generators will be presented in the following.

There exist many ways to generate pseudorandom numbers on a computer; the most popular one is the Linear Congruential Generator (LCG). In 1951 Lehmer generator (15) was published and LCG was published in 1958 (16), (17). Today, LCG can be named the method most commonly used for generating pseudorandom numbers. An advantage of LCGs is that the period is known and long with appropriate parameters. Although not the only criterion, a short period is a fatal flaw in a PRNG (18). LCGs based on the following recurrent formula:

$$X_{n+1} = (aX_n + c) \bmod m, n \geq 0, m > 0, 0 < a < m$$

The value m is the module, a is the multiplier, and c is an additive constant. The sequence has a maximum possible period m, after which it starts to repeat itself (18). LCGs are very popular among researchers, and most mathematical software packages include them. So-called lagged Fibonacci generators are also widely used. They are of the form:

$$X_n = (X_n r \circ X_n p) \bmod m$$

The numbers r and p are called "lags" and there are methods to choose them appropriately. The operator $\circ$ can be one of the following binary operators: addition, subtraction, multiplication, or exclusive or. However, it should be noted that from the point of view of hardware implementation, both congruential and lagged Fibonacci RNGs are not very suitable: they are inefficient in terms of silicon area and time when applied to fine-grained massively parallel machines, for built-in self-test or for other on-board applications (19).

There is a sufficiently large roadmap to achieve the PRNG objective, which has features such as simplicity, velocity, and unpredictability.

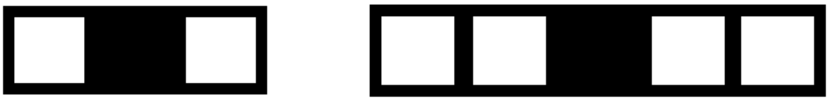
2.1. Homogeneous structures (HS) and cellular automata (CA)

Homogeneous structures (HS) (20) may be formally described as $\sigma = (Z^k, E_n, V, \phi)$, where Z^k - set of k-dimensional vectors, $E_n = 0, 1, \ldots, n-1$ - set of states of one cell in ϕ, $V = (\alpha_1, \ldots \alpha_{(h-1)})$ – neighborhood template (ordered set of distinct k-dimensional vectors from Z^k), $\phi = \phi(x_0, x_1, \ldots, x_{(h-1)})$, $\phi : (E_n)^h \rightarrow E_n$ – local transition function ($\phi(0, 0, ..., 0) = 0$).

From the definition of HS, it can be seen that HS can be compared to the set of ordinary Moore automata (21) if their states depended on the states of neighboring countries. Actually, the neighborhood scheme can differ from each other (Neuman's scheme is more like a symbol plus with the changeable central cell, Moore's scheme is more like the square 3x3). There are a great number of combinations that can be used.

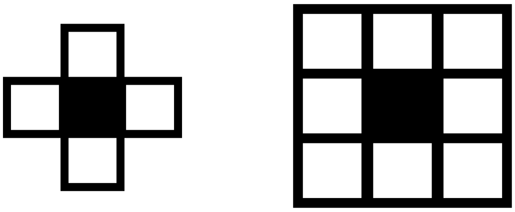
In this paper, only one class of HS will be used. It can be represented like $S = (k, n, m)$, where k and n are described above, and m is the set of $m_i > 0, i = 1, ..., k$. Another value needs to be determined — $P = V_m(\alpha)$. But with some restrictions for values $k, n, (k = n = 2)$, it's obvious, that $|S| = 2^{(2^{(h-1)})}$.

doi:10.20944/preprints202208.001



(a) r=1 (b) r=2

Figure 1: Neighborhood (1-dimensional)



(a) Neumann neighborhood (b) Moore neighborhood

Figure 2: Neighborhood (2-dimensional)

Actually, as a result of some HS, without loss of generality for this research, this object can be replaced by CA. Cellular automata (CA) (20) are dynamical systems in which space and time are discrete. A cellular automata consists of an array of cells, each of which can be in one of a finite number of possible states, updated synchronously in discrete time steps, according to a local, identical interaction rule. Here, we will only consider Boolean automata for which the cellular state $s \in 0, 1$. The state of a cell at the next time step is determined by the current states of the surrounding neighborhood of cells. Cellular array (grid) is d-dimensional, where $d = 1, 2, 3$ is used in practice; as an example for simplicity of understanding we shall concentrate on $d = 1$, i.e., one-dimensional grids.

Each rule in every single cell can be represented as a simple finite automata (or finite state machine) with its input states, transition function, and output state, which is the result of using the rule function with input state as an entry (22; 23). But in terms of cellular automata, there is an entry as a set of states, called neighborhood, when the next state is in formal dependence of states of the particular target cell and its surrounding cells' states (24; 25). For one-dimensional CA (Fig. 1), the target cell is surrounded by r neighbors (cells) at a discrete moment on the appropriate side (by template) where r cells are called radius (each cell has $2r + 1$ cells, which impact its next state).

Besides one-dimensional CAs, there are two main templates for two-dimensional CAs (Fig. 2): the first is so-called von Neumann neighborhood (26) (when the target cell is surrounded by four nondiagonal cells) and the second is called the Moore neighborhood (27) (when the target cell is surrounded by eight cells).

There are also two types of automata, which differ from each other by some kind of complexity in terms of rules. The fact is that one type, called uniform CA, determines one rule for all cells in the grid, but the second type, non-uniform (28) or inhomogenous may have different rules for cells. At the same time they both have the same features of simplicity, locality, and parallelism with the difference in realization (inhomogenous ones require more memory sources for describing rules). Additionally, when we think of finite grid of CAs, it should be highlighted that one-dimensional CAs grids are represented by the circular structure, when two-dimensional ones have a grid in toroidal form.

According to Wolfram's notation (29), firstly, it should be determined that the configuration is the set of ones and zeros (two-state CA, where the sequence is considered as a random number) at a particular discrete moment of time. Wolfram also suggested additional rule numbers. Let us describe the most popular Wolfram's *Rule 30* as an example:

In Boolean form rule 30 can be written as

$$f_0(t+1) = f_{(-1)}(t) \oplus (f_0(t) \vee f_1(t))$$

where the radius is $r = 1$ and $f(t)$ is the state of cell i at time t. As we can see from this formula, $f_0(t + 1)$ is the next state of the cell i at $t + 1$ step, which is a simple Boolean function of neighbor cells states and target cell state at the step t. That is how random sequence of bits is obtained, counting every cell as a target cell in the same moment for all the cells. There are several ways to make the sequence more complex and unpredictable, e.g. time spacing and site spacing.

Time spacing is a method of generating configurations, when not every configuration is a part of resulting sequence, it means that some configurations are generated to produce new ones only, but it will not be in the resulting sequence.

Site spacing is a method of generating sequences when only particular cells' states in the whole configuration (or a row) are bits of the resulting sequence. These methods are obviously empowered by unpredictability feature for whole sequences, but, of course, time or memory sources with these methods remain the same.

Another example in order to show how Wolfram's rules can be encoded. For an example: $f(111) = 1$, $f(110) = 0$, $f(101) = 1$, $f(100) = 1$, $f(011) = 1$, $f(010) = 0$, $f(001) = 0$, $f(000) = 0$, is denoted *Rule 184*.

As an example of non-uniform CA, it can be taken two rules: *Rule 90* and *Rule 150*, which are rules for particular cells in the grid. In Boolean form the *Rule 90* can be written as

$$f_0(t+1) = f_{(-1)}(t) \oplus f_1(t)$$

and *Rule 150*:

$$f_0(t+1) = f_{(-1)}(t) \oplus f_0(t) \oplus f_1(t)$$

These rules perform a sufficiently effective random number generator, in spite of the fact, that they can be simply described as linear Boolean functions.

3. Algorithm of generating pseudorandom numbers on cellular automata

The main idea of the presented algorithm is to combine the best practices of generating pseudorandom numbers by cellular automata and a number of improvements to enforce the general features of cellular automata, saving the best. Fig. 3 presents the flowchart of the proposed algorithm.

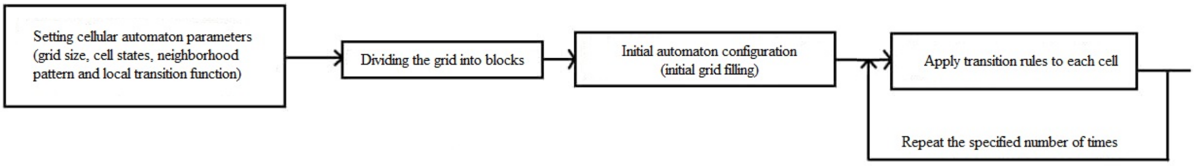

Setting cellular automaton parameters (grid size, cell states, neighborhood pattern and local transition function)
Dividing the grid into blocks
Initial automaton configuration (initial grid filling)
Apply transition rules to each cell
Repeat the specified number of times

Figure 3: The flowchart of the proposed algorithm

There are several subsections, which describe step-by-step all levels of algorithm evolution. In addition, there are some ideas for improving the presented algorithm with the help of genetic algorithms to expand the set of rules to be used to generate a pseudorandom number.

3.1. Grid

The main result of the presented algorithm is that it can generate pseudo-random sequences of numbers. Despite the fact that CA had already been used in such a way, we improved some characteristics by adding new conditions and ideas. There is a grid, where each cell has only two states: 0 and 1, and states are changed by some function. The result of steps in algorithm will be a binary performed number, which can be formed from the grid (or part of it). So, we have a grid with its sizes: p and q, which are both prime numbers (to improve periodic features). Changes start right from this stage: dividing all the grid into $m \geq 2$ blocks, which are formed as rectangles b_i with sizes l_{b_i} and w_{b_i} and each block consists of $l_{b_i} * w_{b_i}$ cells (Fig. 4). So, we get the equation:

$$p * q = \sum_{2 \leq i \leq m} (l_{b_i} * w_{b_i})$$

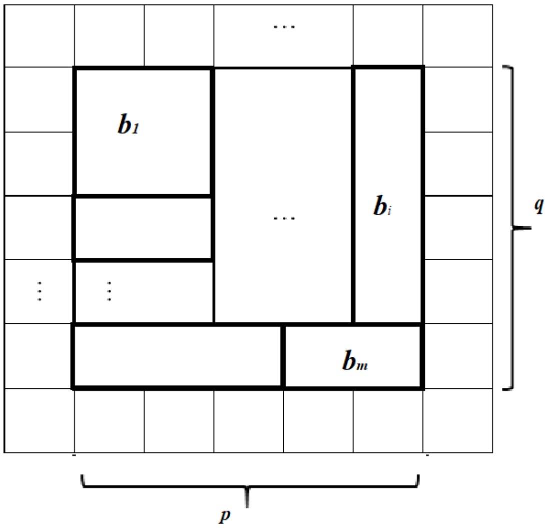


Figure 4: Whole grid view

This is quite a simple method, where each block needs to be placed in the grid, called like the compass (NESW-North, East, South, West) (Fig. 5). The matter is that there will be rule by which blocks will be placed from down left corner and move to the North (up) till the reaching of edge of the grid or to the next block, then to the East (right), South (down) and West (left), according to their size values. This rule may be various for more flexibility, i.e. "snake" method of filling the grid (filling rows from bottom to the top with switching to the next row from opposite sides).

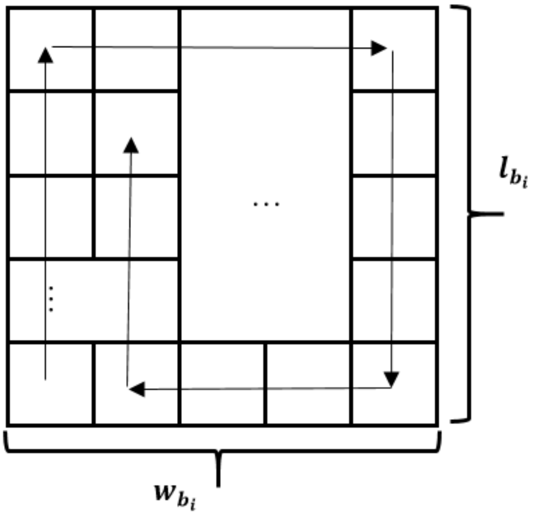
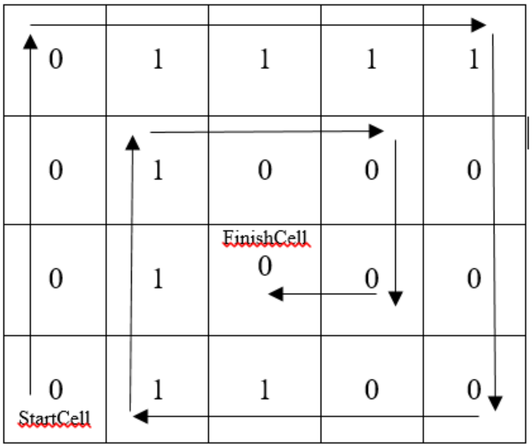


Figure 5: b_i block grid view with NESW scheme of filling

Here is an example of using the NESW method of rule. Let us take rule 15 (00001111 in binary representation) and fill the grid with sizes 4 ∗ 5 (Fig. 6).



0	1	1	1	1
0	1	0	0	0
0	1	FinishCell 0	0	0
0 StartCell	1	1	0	0

Figure 6: NESW method using example

3.2. Rules and neighborhood

We will work with only 2-dimensional space, it means that we take $d = 2, n = 2, (Z^2, E_2)$. Firstly, start configuration will be formed by the rule NESW, the same as blocks. It means, that number $l_{b_i} * w_{b_i}$ in binary form will fulfill the blockgrid along the narrowing edges and etc. What about the rules? The solution of this question may be very flexible because the algorithm is provided by two kinds of rules: numbered rules (like the rule 30 at the beginning of this article) and simple operations (their combination). Each block can have its own rule, which makes whole generating more non-uniform, which leads to unpredictability and better periodic features. Here comes an idea which involves the set of rules in the table, where each rule (both kinds) are encrypted in some combinations, depending on the number of using rules.

So, the rule for each block may be chosen by the simple *mod* operation from the $l_{b_i} * w_{b_i}$ in case of using every possible rule in the block. But for the best results, there should be chosen several rules with an equal number of ones and zeros (it may be used for the whole grid - if we have two blocks and 3 ones in first block rule, so the second block should have num-3 ones, where num is a length of rule-vector if both vectors are equal). Additionally, to avoid bad generations, we should determine ertain rules for the grid. Also, we could fill the "space" between the actual size of the grid and formal. The matter is the fact, that avoiding collisions of values we should make our grid less than its actual size for one cell (or two, depending on the neighborhood template) on each side.

The last question in generating our pseudorandom numbers is a neighborhood. The fact is that, as far as we have blocks, we can choose several samples for some blocks and its number and variety depending on the number of blocks. We can not use a number of rules more than the number of blocks, but we can use similar ones for several blocks.

3.3. "Repeat please"

Pseudorandom number generator must have the option of repeating the algorithm in order to get a similar number after the same operations with the same conditions and configurations. Actually, not to pass the result number, we can only send the key K – the sequence of our data in order to repeat operations on the other end, if needed. We use symbol | to mark concatenation.

$$K = p|q|l_{b_1}|w_{b_1}|...|l_{b_m}|w_{b_m}|V_1|...|V_m|T$$

Where V_i are sent like the sequence of 9 bits where each bit, if it is 1 than cell is in the sample, and when it is 0, it is out of sample.

(-1,1)	(0,1)	(1,1)
(-1,0)	(0,0)	(1,0)
(-1,-1)	(-1,0)	(1,-1)

So this neighborhood mapping table becomes such a string

$$-1,-1;0,-1;1,-1;-1,0;0,0;1,0;-1,1;0,1;1,1$$

Here are the numbers of coordinates, instead of which it should be placed 1 or 0. T is a table, where each using rule may be encrypted, for example, a basic binary number. Thus, in spite of the fact that we must send such a long-sized key, we can put some data in the boundary cells, placed around our main grid in order to save correct transition of state. We can save $|V_1|\dots|V_m|T$ part of the key in these cells, depending on its complexity and size.

3.4. Addtitional function

To be sure, that our algorithm will generate high-quality random number sequences, we decided to use some nonlinear functions on this level of the development algorithm. The reason is in the CA rules. The fact is that only part of the rules has good random features (the ability to generate complex structures). But if we take only this part of the rules, we get too small a set of needed ones, there are two ways to improve it. The first one is to add a nonlinear function for XOR with the subset of the grid for making it more complex. The second is that we can choose "good" (having randomness features) packs of rules with the help of genetic algorithms or some kind of tutor algorithm. Therefore, the evolution of CAs will show us only those rules which can be used in practice.

4. Test results

Our algorithm was tested with the help of the NIST Test Suite, which was developed to test RNG and PRNG (31). The process of the test involved: generation sequences of bits by our algorithm (around 1000000 bits), testing.txt file with a sequence with NIST test suite. The 10000 variations of the rules pack and ways of filling the grid were tested, as sizes of grids. As a result, a number of the most "productive" packs of rules-grid size-neighborhood were highlighted.

So, we used the grid with sizes 307 $*$ 53, divided by 6 blocks of approximately equal lengths, while the width of each block was 53. Each block was filled with bits, generated by results of multiplication of blocks' sizes, with the NESW method. The neighborhood template was simple and even classical enough; it was a 3B rule (center cell change state in dependence of its previous state and left and right neighbors' ones). The package of rules was: {45, 75, 89, 101, 135, 86}.

Rule 45:

000	001	010	011	100	101	110	111
1	0	1	1	0	1	0	0

Rule 75:

000	001	010	011	100	101	110	111
1	1	0	1	0	0	1	0

Rule 89:

000	001	010	011	100	101	110	111
1	0	0	1	1	0	1	0

Rule 101:

000	001	010	011	100	101	110	111
1	0	1	0	0	1	1	0

Rule 135:

000	001	010	011	100	101	110	111
1	1	1	0	0	0	0	1

Rule 86:

000	001	010	011	100	101	110	111
0	1	1	0	1	0	1	0

4.1. Short Description of criteria with results

The test statistic is used to calculate a $P - value$ that summarizes the strength of the evidence against the null hypothesis. For these tests, each $P - value$ is the probability that a perfect random number generator would have produced a sequence less random than the sequence that was tested, given the kind of non-randomness assessed by the test. If a $P - value$ for a test is determined to be equal to 1, then the sequence appears to have perfect randomness. A $P - value$ of zero indicates that the sequence appears to be completely non-random. A significance level (α) can be chosen for the tests. If $P - value \geq \alpha$, then the null hypothesis is accepted; i.e., the sequence appears to be random. If $P - value < \alpha$, then the null hypothesis is rejected; i.e., the sequence appears to be non-random. The parameter α denotes the probability of the $TypeI$ error (if the data is, in truth, random, then a conclusion to reject the null hypothesis (i.e., conclude that the data is non-random) will occur a small percentage of the time) (31), α is equal to 0.01 (31).

4.1.1. *Frequency (monobit) test*

The focus of the test is the proportion of zeros and ones for the entire sequence. The purpose of this test is to determine whether the number of ones and zeros in a sequence is approximately the same as would be expected for a truly random sequence. The test assesses the closeness of the fraction of ones to 0.5, that is, the number of ones and zeros in a sequence should be about the same. All subsequent tests depend on the passing of this test (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.744146	99.6%	Frequency

4.1.2. *Frequency test within a block*

The focus of the test is the proportion of the ones within M-bit blocks. The purpose of this test is to determine whether the frequency of ones in an M-bit block is approximately M/2, as would be expected under an assumption of randomness. For block size $M = 1$, this test degenerates to the Frequency (monobit) test (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.380537	99.4%	BlockFrequency

4.1.3. *Runs test*

The focus of this test is the total number of runs in the sequence, where a run is an uninterrupted sequence of identical bits. A run of lengths k consists of exactly k identical bits and is bounded before and after with a bit of opposite value. The purpose of the run test is to determine whether the number of runs of ones and zeros of various lengths is as expected for a random sequence (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.428244	98.5%	Runs

4.1.4. *Longest run of ones in a block*

The focus of the test is the longest run of ones within M-bit blocks. The purpose of this test is to determine whether the length of the longest run of ones within the tested sequence is consistent with the length of the longest run of ones that would be expected in a random sequence. Note that an irregularity in the expected length of the longest run of ones implies that there is also an irregularity in the expected length of the longest run of zeroes. Therefore, only a test for ones is necessary (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.383827	99.1%	LongestRun

4.1.5. *Rank test*

The focus of the test is the rank of disjoint sub-matrices of the entire sequence. The purpose of the test is to check for linear dependence among fixed length substrings of the original sequence (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.702458	99.0%	Rank

4.1.6. *Discrete Fourier transform (spectral) test*

The focus of this test is the peak heights in the Discrete Fourier transform of the sequence. The purpose of this test is to detect periodic features (i.e. repetitive patterns that are close to each other) in the sequence tested that would indicate a deviation from the assumption of randomness. The intention is to detect whether the number of peaks exceeding the 95 % threshold is significantly different than 5 % (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.650637	99.5%	FFT

4.1.7. *Non-overlapping template matching test*

The focus of this test is the number of occurrences of prespecified target strings. The purpose of this test is to detect generators that produce too many occurrences of a given nonperiodic (aperiodic) pattern. For this test and for the Overlapping Template Matching test, an *m*-bit window is used to search for a specific *m*-bit pattern. If the pattern is not found, the window slides one bit position. If the pattern is found, the window is reset to the bit after the found pattern and the search is resumed (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.433358	98.7%	NonOverlappingTemplate

4.1.8. *Overlapping template matching test*

The focus of the Overlapping Template Matching test is the number of occurrences of prespecified target strings. Both this test and the Non-overlapping Template Matching test use an *m*-bit window to search for a specific *m*-bit pattern. As in the previous test, if the pattern is not found, the window slides one bit in the position. The difference between this test and the previous test is that, when the pattern is found, the window slides only one bit before resuming the search (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.632191	100%	OverlappingTemplate

4.1.9. *Maurer's "Universal statistical test"*

The focus of this test is the number of bits between the matching patterns (a measure that is related to the length of a compressed sequence). The purpose of the test is to detect whether the sequence can be significantly compressed without loss of information. A significantly compressible sequence is considered to be non-random (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.414862	98.3%	Universal

4.1.10. *Linear complexity test*

The focus of this test is the length of a linear feedback shift register (LFSR). The purpose of this test is to determine whether the sequence is complex enough to be considered random. Random sequences are characterized by longer LFSRs. An LFSR that is too short implies non-randomness (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.730485	100%	LinearComplexity

4.1.11. *Serial test*

The focus of this test is the frequency of all possible overlapping m-bit patterns throughout the sequence. The purpose of this test is to determine whether the number of occurrences of the 2*mm*-bit overlapping patterns is approximately the same as would be expected for a random sequence. Random sequences have uniformity; that is, every *m*-bit pattern has the same chance of appearing as every other *m*-bit pattern (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.651956	99.3%	Serial

4.1.12. Approximate entropy test

The focus of this test is the frequency of all possible overlapping m-bit patterns across the entire sequence. The purpose of the test is to compare the frequency of overlapping blocks of two consecutive/adjacent lengths (m and $m+1$) against the expected result for a random sequence (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.778903	99.1%	Approximate Entropy

4.1.13. Cumulative sums (Cusums) test

The focus of this test is the maximal excursion (from zero) of the random walk defined by the cumulative sum of adjusted (-1, +1) digits in the sequence. The purpose of the test is to determine whether the cumulative sum of the partial sequences occurring in the tested sequence is too large or too small relative to the expected behavior of that cumulative sum for random sequences. This cumulative sum may be considered as a random walk. For a random sequence, the excursions of the random walk should be near zero. For certain types of non-random sequences, the excursions of this random walk from zero will be large (31).

P-VALUE	PROPORTION	STATISTICAL TEST
0.734146	99.9%	CumulativeSums

4.1.14. Random excursions test

The focus of this test is the number of cycles that have exactly K visits in a cumulative sum random walk. The cumulative sum random walk is derived from partial sums after the (0,1) sequence is transferred to the appropriate (-1, +1) sequence. A cycle of a random walk consists of a sequence of steps of unit length taken at random that begin at and return to the origin. The purpose of this test is to determine if the number of visits to a particular state within a cycle deviates from what one would expect for a random sequence. This test is actually a series of eight tests (and conclusions), one test (conclusion) for each of the states -4, -3, -2, -1, +1, +2, +3, +4 (31).

P-VALUE	VALUE
0.673779	-4
0.640660	-3
0.636757	-2
0.906842	-1
0.614216	+1
0.042225	+2
0.570447	+3
0.802887	+4

4.1.15. Random Excursions variant test

The focus of this test is the total number of times a particular state is visited (that is, occurs) in a cumulative sum random walk. The purpose of this test is to detect deviations from the expected number of visits to various states in the random walk. This test is actually a series of 18 tests (and conclusions), one test (conclusion) for each of the states: -9…-1 and +1…+9 (31).

Table 1
RESULTS FOR THE UNIFORMITY OF P-VALUES AND THE PROPORTION OF PASSING SEQUENCES

P-VALUE	PROPORTION	STATISTICAL TEST
0.744146	99.6%	Frequency
0.380537	99.4%	BlockFrequency
0.734146	99.9%	CumulativeSums
0.428244	98.5%	Runs
0.383827	99.1%	LongestRun
0.702458	99.0%	Rank
0.650637	99.5%	FFT
0.433358	98.7%	NonOverlappingTemplate
0.632191	100%	OverlappingTemplate
0.414862	98.3%	Universal
0.778903	99.1%	ApproximateEntropy
0.651956	99.3%	Serial
0.730485	100%	LinearComplexity

P-VALUE	VALUE
0.148224	-9
0.293802	-8
0.799213	-7
0.810598	-6
0.589138	-5
0.525096	-4
0.584469	-3
0.659030	-2
0.878513	-1
0.878513	+1
0.469278	+2
0.502912	+3
0.711568	+4
0.910749	+5
0.860976	+6
0.702797	+7
0.532906	+8
0.543193	+9

The choice of rules is determined by good randomness and an equal number of ones and zeros in each of these particular rules. Table 1 presents the results of the NIST test suite package (31) with the input of the file, generated by the developed algorithm. All $P - values$ in this table are averaged from 10-100 rounds of tests.

Also, there is a comparison (Table 2) between the measurements of the NIST test suite of the presented method and the Linear Congruential method and it's seen that the presented method gets better results in some tests and $P - value$ in some tests is much higher.

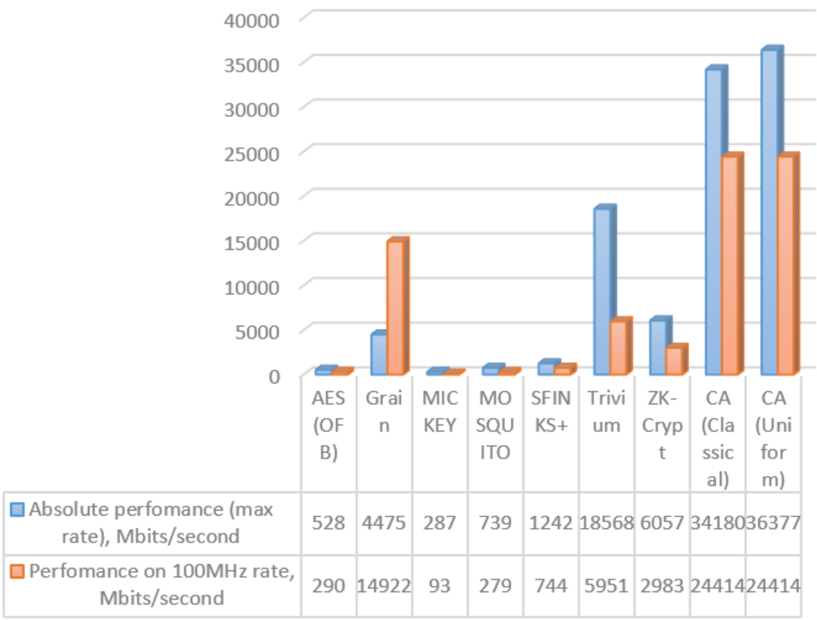
5. FPGA Implementation

The presented algorithm has two implementations, one on FPGA and one on CUDA. Here comes a chart (Fig. 7) from the paper (30), which shows the velocity of a pseudorandom numbers' generator, based on cellular automaton implemented on FPGA. There are several algorithms, which were measured on the same FPGA platform from eSTREAM (as pseudorandom generators, which were considered strictly and detailed enough for their process velocity and statistical features) (32; 33).

Hardware modeling was performed on Artix 7 xc7a200tfbg676-3 in Vivado 2016.3. The synthesis parameters are shown in Table 3, the synthesis strategy is Flow$_{perfOptimized_high}.ThemodelingresultsarepresentedinTable4.Resourcesoft$

Table 2
COMPARISON OF ALGORITHM ON CA AND LINEAR CONGRUENTIAL METHOD

CA		Linear Congruential		STATISTICAL TEST
P-VALUE	PROPORTION	P-VALUE	PROPORTION	
0.744146	99.6%	0.739918	99.8%	Frequency
0.380537	99.4%	0.122325	99.0%	BlockFrequency
0.734146	99.9%	0.689918	100%	CumulativeSums
0.428244	98.5%	0.213309	98.7%	Runs
0.383827	99.1%	0.122325	98.1%	LongestRun
0.702458	99.0%	0.213309	99.0%	Rank
0.650637	99.5%	0.639918	100%	FFT
0.433358	98.7%	0.578346	98.4%	NonOverlappingTemplate
0.632191	100%	0.350485	100%	OverlappingTemplate
0.414862	98.3%	0.213309	98.3%	Universal
0.778903	99.1%	0.791468	100%	ApproximateEntropy
0.610977	98.8%	0.534146	99.2%	RandomExcursions
0.633388	99.2%	0.468312	100%	RandomExcursionsVariant
0.651956	99.3%	0.615983	100%	Serial
0.730485	100%	0.213309	100%	LinearComplexity



	AES (OFB)	Grain	MICKEY	MOSQUITO	SFINKS+	Trivium	ZK-Crypt	CA (Classical)	CA (Uniform)
🟦 Absolute perfomance (max rate), Mbits/second	528	4475	287	739	1242	18568	6057	34180	36377
🟧 Perfomance on 100MHz rate, Mbits/second	290	14922	93	279	744	5951	2983	24414	24414

Figure 7: Different generating pseudorandom numbers algorithms perfomance

We can compare the results of the FPGA-based implementation with the results presented in (34), (35). In (34) PRNG was implemented in...

6. CUDA implementation

Hardware used:

1. CPU: Intel Core i5-4670
2. GPU: GeForce GTX 750 Ti

Table 3
SYNTHESIS PARAMETER

PARAMETERS	RESULTS
Flatten hierarchy during LUT mapping	rebuilt
Convert clock gating logic to flop enable	off
Max number of global clock buffers used by synthesis	12
Fanout limit	400
Synthesis directive	Default
Retiming	Not checked
FSM Extraction Encoding	One_hot
Keep equivalent registers	Checked
Resource sharing	off
Control set optimization threshold	Auto
Disable LUT Combining	Checked
Min length for chain of registers to be mapped onto SRL	5
Max number of block RAM allowed in design	Max number allowed for the part in question
Max number of Ultra RAM blocks allowed in design	Max number allowed for the part in question
Max number of block DSP allowed in design	Max number allowed for the part in question
Max number of BRAM that can be cascaded by the tool	Max number allowed for the part in question
Max number of URAM that can be cascaded by the tool	Max number allowed for the part in question
Cascade DSP	Auto
Enable VHDL assert statements to be evaluated	Not checked

Table 4
HARDWARE MODELING RESULTS

SIZE	53*307	256*256	512*512
BLOCKS	4	16	16
DELAY, ns	4.275	4.681	4.681
LUTs	35334	142848	580608
FREQUENCY, MHz	234	214	214
ABSOLUTE PERFOMANCE, Mbit/s	3806081	14000427	56001709
PERFOMANCE ON 100MHz rate, Mbit/s	1627100	6553600	26214400

As noted above, one of the main advantages of the method is the possibility of parallelization of the calculations. Here will be described the implementation of the algorithm on CUDA (GPU) with the main differences from the CPU variant.

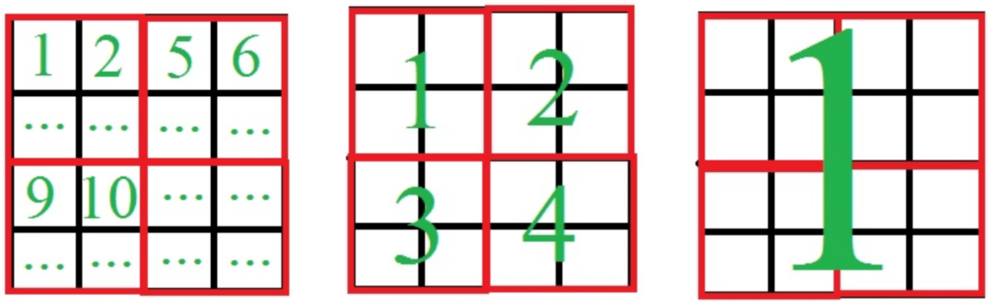
There are 2 approaches:

1. Sequential calculation of blocks, parallel calculation of cells in blocks.
   The cycle on the CPU is created in which each block will be calculated. To calculate a block on the GPU for each cell, a stream is created and the final state of the cell is calculated.
2. Parallel computation of blocks, parallel computation of cells in blocks.
   For each cell downstream will be run, in each stream the block to which the cell belongs is determined, and then the final state of the cell is calculated.

The most difficult part at this step is to calculate the index of the cell in the block when it is circled clockwise (NESW). Since the program is executed in parallel, there is no sequential round-trip in a clockwise direction, and the index is calculated using a formula that uses the perimeter of the rectangle.

It can be seen in Fig. 8 and Fig. 9 the difference between the three hardware-based methods of the described PRNG implementation. Table 5 presents the comparison between the results of the tests of these three methods. The difference chart is presented in Fig. 10).

After CUDA research, there were several tests with variable block size (Table 6).

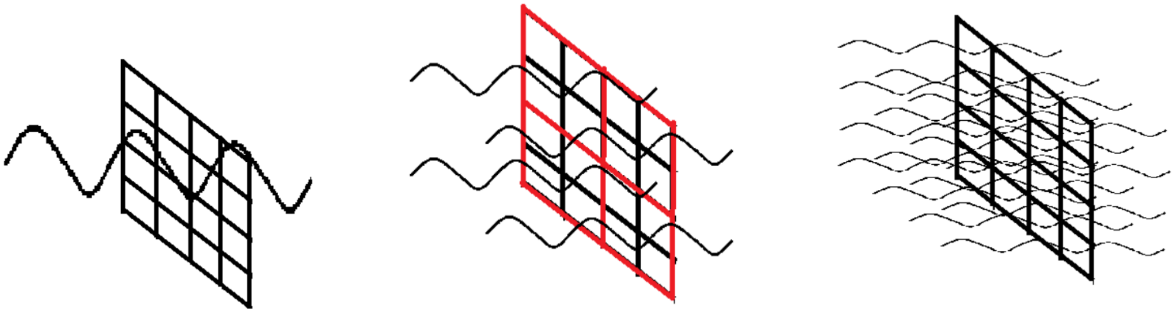


(a) Sequential calculations for both cells and blocks (CPU)

(b) Parallel calculations for cells in each block (GPU1)

(c) Parallel calculations for whole field (GPU2)

Figure 8: Calculations queue for CPU and GPU



(a) One stream for whole field (CPU)

(b) The stream for each block (GPU1)

(c) The stream for each cell (GPU2)

Figure 9: Using streams for calculations on CPU and GPU

Table 5
COMPARISON BETWEEN TESTS RESULTS OF THREE HARDWARE-BASED METHODS

SIZE	BLOCKS	CPU(μs)	GPU1(μs)	GPU2(μs)
53*307	4	1639	1948	4584
256*256	16	3816	5813	5786
512*512	16	17264	6792	7485
1024*1024	16	57986	8182	14583
2048*2048	16	257453	16887	43932
4096*4096	16	1128957	64226	171278

It is obvious that all the advantages and disadvantages of CPU and GPU-based implementations are concluded. But it is not in such a way for GPU1 and GPU2 implementations difference.

The main differences between GPU1 and GPU2 implementations are that in the GPU2 implementation, the calculation of blocks and cells occurs in parallel, and in GPU1 - the cells in the block are calculated in parallel, and the blocks are sequential. In the second, there is a need for additional calculations, such as the calculation of the cell belonging to the unit. The second implementation will give an increase in performance with a large number of blocks and a large volume of cells.

It can be seen from the tables, on small data, that two implementations on the GPU lose their pros to the CPU. This is due to the fact that the amount of computation is so small that copying the device into memory and additional computations in each GPU stream take a lot of time for this amount of data.

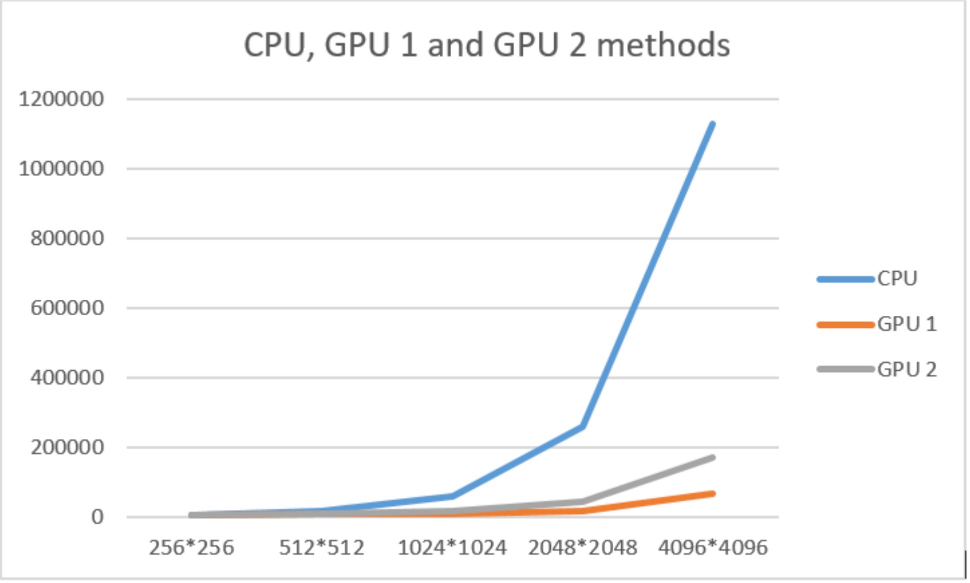


Figure 10: Number of cells (X) to execution time (Y) chart

Table 6
BLOCK SIZE VARIETY OUTPUT

FIELD	BLOCKSIZE	BLOCKS	CPU(μs)	GPU1(μs)	GPU2(μs)
256*256	16*16	256	43910	121671	19135
256*256	32*32	64	13683	30743	9567
256*256	64*64	16	6778	8157	7859
256*256	128*128	4	4873	2570	7001
512*512	32*32	256	53994	125554	52134
512*512	64*64	64	25109	36101	18919
512*512	128*128	16	17327	8826	10303
512*512	256*256	4	15189	2958	8167
1024*1024	64*64	256	95915	127036	175194
1024*1024	128*128	64	65996	38352	46781
1024*1024	256*256	16	58411	11874	20757
1024*1024	512*512	4	55890	5414	12556

However, on large amounts of data, implementation on GPU wins due to parallel computing. The use of additional calculations is due to the fact that they are performed in parallel and allow each cell and block to be calculated in parallel.

7. Discussion

It should be noted that from the point of view of hardware implementation, both linear congruent and retarded Fibonacci generators are not very suitable: they are not efficient in terms of microprocessors and time, when it is necessary to apply PRNG for distributed machines and parallel computing, for embedded testers, or for other applications at this level.

A third widely used type of generator is the so-called linear feedback generator (LFSR). Linear feedback shift registers are common among physicists and computer engineers. There are forms of LFSR that are suitable for hardware implementation.

However, it turns out that, compared to equivalent CA-based generators, they are less efficient; in addition, they are less applicable in terms of the possibility of implementation and debugging, although the area required for the CA cell is slightly larger than for the LFSR (24). Moreover, different sequences generated by the same CA are much less correlated than similar sequences generated by LFSR. This means that CA-generated bit sequences can be used in parallel, which implies clear advantages in using them in applications.

The proposed PRNG algorithm combines the advantages of the speed of cellular automata on a two-dimensional basis and ease of implementation, as a step-by-step series of Moore automata with initiated initial states. Compared to the classical algorithms generating pseudo-random numbers, cellular automaton presented algorithm has such changes:

- application of several neighborhood templates;
- the use of several independent cellular automata;
- implementation of transitions of various cellular automata according to various sets of rules;
- expansion of rule sets based on the complementarity of the number of ones and zeros for a statistically normal distribution.

8. Conclusion

This paper presents a pseudorandom generating algorithm on CA in a detailed form with new ideas to improve some features of an algorithm, which was before. Testing results show that the described generator has good perspectives because of its effectiveness, simplicity, velocity, and improved unpredictability.

Future research will be focusing on implementation presented generator in different cryptographic applications such as lightweight cryptography or coding theory.

9. Funding

This work was supported by the Ministry of Science and Higher Education of the Russian Federation ‘Goszadanie’ №075-01024-21-02 from 29.09.2021 (project FSEE-2021-0015).

10. Author Contributions

Conceptualization, A.L, D.M., P.L., D.B., D.K.; methodology, A.L., D.M., P.L. and D.K.; software, D.B., M.V.; validation, M.V., G.B. and A.L.; formal analysis, D.K., V.L.; investigation, D.M., D.B. and M.V.; resources, P.L. and D.K.; data curation, D.B., D. K. and P.L.; writing—original draft preparation, D.M., D.K. and P.L.; writing—review and editing, D.K., A.L and P.L.; visualization, M.V. and D.M.; supervision, A.L., D.M. and D.K.; project administration, D.K.; funding acquisition, A.L. and D.K. All authors have read and agreed to the published version of the manuscript.

11. Conflicts of Interest

The authors declare no conflict of interests.

References

[1] Daniel Genkin. Adi Shamir. Eran Tromer.: RSA Key Extraction via Low-Bandwidth Acoustic Cryptanalysis (extended version).IACR Cryptology ePrint Archive, 2013:857, 2013.

[2] Daniel Genkin. Itamar Pipman. Eran Tromer.: Get Your Hands Off My Laptop: Physical Side-Channel Key-Extraction Attacks on PCs. In CHES, pages 242–260, 2014.

[3] Daniel Genkin. Lev Pachmanov. Itamar Pipman.Eran Tromer.: Stealing Keys from PCs by Radio: Cheap Electromagnetic Attacks on Windowed Exponentiation. IACR Cryptology ePrint Archive, 2015:170, 2015.

[4] Levina A., Borisenko P., Mostovoy R., Orsino A.,Ometov A., Andreev S. Mobile Social Networking under Side-Channel Attacks: Practical Security Challenges. IEEE Access - 2017, Vol. 5, pp. 2591-2601

[5] Levina, A., Mostovoi, R., Sleptsova, D. et al. Physical model of sensitive data leakage from PC-based cryptographic systems. J Cryptogr Eng 9, 393–400, 2019.

[6] Toru Sasaki, Hiroyuki Togo, Jun Tanidaa and Yoshiki Ichiokab *Stream cipher based on pseudo-random number generation using optical affine transformation*, Applied Optics 39(14):2340-6 · June 2000

[7] Biryukov A., Shamir A *Cryptanalytic time/memory/data tradeoffs for stream ciphers. International Conference on the Theory and Application of Cryptology and Information Security.* Springer, Berlin, Heidelberg, 1-13, 2000.

[8] Cunsheng D. *The stability theory of stream ciphers*, 2011.

[9] Ferguson, Niels; Schneier, Bruce; Kohno, Tadayoshi, *Chapter 9: Generating Randomness*, Cryptography Engineering: Design Principles and Practical Applications. Wiley Publishing, Inc. ISBN 978-0-470-47424-2, 2010

[10] Moysis L, Rajagopal K, Tutueva AV, Volos C, Teka B, Butusov DN. *Chaotic Path Planning for 3D Area Coverage Using a Pseudo-Random Bit Generator from a 1D Chaotic Map*, Mathematics. 2021; 9(15):1821

[11] Tutueva, A.V., Karimov, T.I., Moysis, L. et al.*Improving chaos-based pseudo-random generators in finite-precision arithmetic*, Nonlinear Dyn 104, 727–737, 2021

[12] Palacios-Luengas, L., Pichardo-Méndez, J.L., Díaz-Méndez, J.A., Rodríguez-Santos, F., Vázquez-Medina, R.: *PRNG based on skew tent map*, Arab. J. Sci. Eng. 44(4), 3817–3830, 2019

[13] Datcu, O., Macovei, C., Hobincu, R.: *Chaos based cryptographic pseudo-random number generator template with dynamic state change*, Appl. Sci. 10(2), 451 (2020)

[14] P. L'Ecuyer, O. Nadeau-Chamard, Y.-F. Chen, and J. Lebar.: *Multiple Streams with Recurrence-Based, Counter-Based, and Splittable Random Number Generators*, Proceedings of the 2021 Winter Simulation Conference, invited paper, 2021.

[15] D. H. Lehmer, *Mathematical methods in large-scale computing units*, Proceedings of a Second Symposium on Large-Scale Digital Calculating Machinery, 1949, Harvard University Press, Cambridge, Mass., 1951, pp. 141—146.

[16] W. E. Thomson, A Modified Congruence Method of Generating Pseudo-random Numbers, The Computer Journal, Volume 1, Issue 2, 1958, Page 83, https://doi.org/10.1093/comjnl/1.2.83

[17] A. Rotenberg. 1960. A New Pseudo-Random Number Generator. J. ACM 7, 1 (Jan. 1960), 75–77. DOI:https://doi.org/10.1145/321008.321019

[18] Pierre l'Ecuyer. History of uniform random number generation. WSC 2017 - Winter Simulation Conference, Dec 2017, Las Vegas, United States.

[19] M. Tomassini, *Spatially Structured Evolutionary Algorithms: Artificial Evolution in Space and Time*, Springer, 2005

[20] V. B.Kudryavtsev, A.S.Podkolzin, "Cellular automata", *Intellectual systems* 1 - 4(10):657 - 692, 2006

[21] Moore, Edward F *Gedanken-experiments on Sequential Machines*, Automata Studies, Annals of Math. Studies. Princeton, N.J.: Princeton University Press (34): 129–153, 1956

[22] Stephen Wolfram *Random sequence generation by cellular automata. Advances in Applied Mathematics*, 1986

[23] Andrew Ilachinski *Cellular Automata: A Discrete Universe. World Scientific*, 2001.

[24] M. Tomassini, M. Sipper, and M. Perrenoud *On the generation of high-quality random numbers by two-dimensional cellular automata* IEEE Transactions on Computers, 49(10):1146–1151, Oct 2000

[25] M. Tomassini, M. Perrenoud *Cryptography with cellular automata*, Appl. Soft Comput., 1(2):151–160, 2001

[26] Weisstein, Eric W. *von Neumann Neighborhood* MathWorld–A Wolfram Web Resource. `http://mathworld.wolfram.com/vonNeumannNeighborhood.html`

[27] Weisstein, Eric W. *Moore Neighborhood* MathWorld–A Wolfram Web Resource. `http://mathworld.wolfram.com/MooreNeighborhood.html`

[28] Alberto Dennunzio, Enrico Formenti, Julien Provillard *Non-uniform cellular automata:Classes, dynamics, and decidability* Journal of Information and Computation, Elsevier, 2012

[29] Wolfram S. *Cellular Automat* Los Alamos Science 9: 2–21, 1983

[30] Zhukov A.E. Cellular Automata in Cryptography. Part 2. Voprosy kiberbezopasnosti [Cybersecurity issues], 2017, No 4 (22), pp. 47-66.

[31] Bassham, L. , Rukhin, A. , Soto, J. , Nechvatal, J. , Smid, M. , Leigh, S. , Levenson, M. , Vangel, M. , Heckert, N. and Banks, D., *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, 2010.

[32] Rogawski M. Hardware evaluation of eSTREAM candidates: Grain, Lex, Mickey128, Salsa20 and Trivium // The eSTREAM Project. – 2007. – 10 p

[33] Gurkaynak F. et al. Hardware evaluation of eSTREAM candidates: Achterbahn, Grain, MICKEY, MOSQUITO, SFINKS, Trivium, VEST, ZK-crypt // The eSTREAM Project. – 2006. – 12 p

[34] M. Bakiri, C. Guyeux, J. Couchot, L. Marangio and S. Galatolo, "A Hardware and Secure Pseudorandom Generator for Constrained Devices," in IEEE Transactions on Industrial Informatics, vol. 14, no. 8, pp. 3754-3765, Aug. 2018, doi: 10.1109/TII.2018.2815985.

[35] M. Bakiri, J. Couchot and C. Guyeux, "CIPRNG: A VLSI Family of Chaotic Iterations Post-Processings for $\mathbb{F}_2$ -Linear Pseudorandom Number Generation Based on Zynq MPSoC," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 65, no. 5, pp. 1628-1641, May 2018, doi: 10.1109/TCSI.2017.2754650.