Antonín Culek; Jan Havel; Václav Přibyl

On a method of pseudo-random numbers generation

Kybernetika, Vol. 2 (1966), No. 3, (215)--225

Persistent URL: http://dml.cz/dmlcz/125728

KYBERNETIKA ČÍSLO 3, ROČNÍK 2/1966

On a Method of Pseudo-Random Numbers Generation

ANTONÍN CULEK, JAN HAVEL, VÁCLAV PŘIBYL

The present paper summarizes some knowledge, results and possibilities of use of the linear recurrence modulo 2 generation method of pseudo-random numbers. Characterization and properties of the method are briefly treated in the first part. Results of measurements and tests of the generator are presented in the second part.

It is shown, that although the method fulfills some basic properties, required for generators of random binomial sequences, it is not, as compared with the generator based on physical principles, too useful for further transformations (filtering including).

I. DESCRIPTION AND PROPERTIES OF THE METHOD

The above mentioned method is described in [1] and [2]. Nature of generating the sequence is represented on Fig. 1. Let us have an n-stage shift register and an arbitrary (with the only exception of the number 00 ... 000) n-place binary number

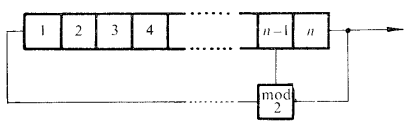

Fig. 1. Block schema of the pseudo-random numbers generation.

recorded in it. Let one or several adders modulo 2 $(0 + 0 = 1 + 1 = 0,\ 0 + 1 = 1 + 0 = 1)$ be connected to acceptable stages of the register and its or their output be connected to the input of the register. Then on the output of the register pseudo-random noncorrelated sequence of binary numbers (0,1) appears. Theory of generation of such sequences is treated in [3] and in the papers [4], [12] and it is a part of coding theory. Because of pseudo-random sequences, their period is to be determined. It may be shown that the length of the period depends on the degree of the polynomial, from which the sequence is generated and in the case of primitive polynomial we can, by means of acceptable selection of stages, from which we add, reach the maximal length of period $p = 2^n - 1$, where n is the degree of the primitive polynomial (n is the number of stages of the shift register). It is possible to write the i-th term a_i of the above mentioned sequence $\{a_i\}$

$$a_i = c_1 a_{i-1} + c_2 a_{i-2} + \ldots + c_n a_{i-n}$$

where c_j, $j = 1, 2, \ldots, n - 1$ is equal either 0 or 1 and $c_n = 1$. Then the sequence is of degree n. In accordance with [1], the necessary and sufficient condition for reaching the maximal period, i.e. $p = 2^n - 1$ is for polynomial

$$f(x) = 1 + c_1 x + c_2 x^2 + \ldots + x^n$$

to be primitive over GF(2) (Galois field modulo 2 cf. [3]). In this case, the output sequence has the following properties [5]:

1. For each period it holds, that the number of 0's and the number of 1's differ at most by one. From realization we exclude n-tuple containing all zeros. It is clear, that such an n-tuple generates 0's all the time.
2. In each period the number of groups of length k (where $k < n$) containing all zeros or all ones is twice as great as that of length $k + 1$.
3. Autocorrelation function of the output pseudo-random telegrafic signal (cf. Fig. 2) is:

$$R_{xx}(\tau) = \begin{cases} 1 - \dfrac{|\tau|\, 2^n}{T(2^n - 1)} & \text{for} \quad \tau \leqq T, \\ -\dfrac{1}{2^n - 1} & \text{for} \quad \tau \geqq T \end{cases}$$

   where $T = 1/f_c$ and f_c is the repetition frequency of the output pseudo-random sequence and n is the degree of the primitive polynomial.
4. Spectral power density is a discrete function:

$$S_{xx}(f) = \frac{\delta(f)}{(2^n - 1)^2} + \sum_{\substack{a=-\infty \\ a \neq 0}}^{+\infty} \frac{2^n}{(2^n - 1)^2} \left[\frac{\sin(\alpha\pi/2^n - 1)}{(\alpha\pi/2^n - 1)}\right]^2 \delta\left(f \frac{\alpha f_c}{2^n - 1}\right)$$

   and the distance between separate harmonic frequencies is $(f_c/2^n - 1)$ and the bandwidth of the function $S_{xx}(f)$ is approximately $0{,}32 f_c$.
5. This method of producing pseudo-random sequences may be used in digital computer techniques for gaining random numbers of more digits (maximal number of digits is n, where n is the length of the register) with uniform distribution in the interval $(0,1)$, what is proved in [4].

Let us now for illustration take an example of a 4-stage register and observe closely the technique of generating a pseudo-random sequence. Maximal length of the period is in this case $p = 2^4 - 1 = 15$. In Table 1 producing of this sequence is expressed in two cases. In the first one (column a) the modulo 2 adder is connected to the fourth (output) and the third stages of the register, in the second one (column b) the adder is connected to the fourth and the second stages of the register. In both cases the

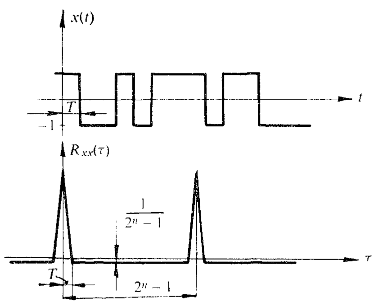

Fig. 2. Pseudo-random telegraph signal $x(t)$ (amplitude ± 1) and its autocorrelation function $R_{xx}(t)$.

Table 1.

Possibilities of producing the pseudo-random sequence with 4-stage register (a — complete period, *b* — incomplete period)

a			*b*		
step	register state	output	step	register state	output
1	1111	1	1	1111	1
2	0111	1	2	0111	1
3	0011	1	3	0011	1
4	0001	1	4	1001	1
5	1000	0	5	1100	0
6	0100	0	6	1110	0
7	0010	0			
8	1001	1	7	1111	1
9	1100	0			
10	0110	0			
11	1011	1	1	0110	0
12	0101	1	2	1011	1
13	1010	0	3	1101	1
14	1101	1	4	0110	0
15	1110	0			
16	1111	1			

group 1111 was selected for initial state of the register. In the first case we gain the maximal length of the period and it may be seen, that in the register all possible combinations except 0000 are recorded consecutively. We shall therefore call this period "complete". In the second case, the basic condition is not fulfilled and such a register does not produce the maximal period. The register takes only selected number of combinations and the output sequence has not formerly considered properties (cf. e.g. [1)]. It is also clear, that in this case the length of this "incomplete" period depends on the selection of the initial combination.

From the point of view of technical realization of the generator using this principle it is necessary to predetermine whether the way of connection (number of stages n, selection of outputs for adding) corresponds to the primitive polynomial. Polynomials up to the degree 34 are given in [3], several other works and measurements in this field are mentioned in the second part of this paper. However, it is advantageous for technical realization to use as simple polynomials as possible (trinomial), which implies the use of minimal number (one) of adders modulo 2.

If we compare the just described method of producing pseudo-random numbers with other methods used for digital computers, then its great advantage is its simplicity. It is in fact the only method of generation pseudo-random numbers usable for realization of special portable generators, working without the digital computer. At the same time no extreme length of the register is required, e.g. a 25-stage register can generate a sequence with the period equal to 33 554 431 bits, a 28-stage can produce already a sequence with the period equal to 268 435 455 bits. From the technical point of view, this method is capable of producing random numbers with very high repetition frequency because, except of adding modulo 2, it is necessary to ensure the reliable function of the shift register only. Thus in [6] a generator with output frequency $2 \cdot 10^8$ bit/sec is described and and in [7] a method of connection of pseudo-random numbers generator (operation frequency variable in the range $0 \div 4 \cdot 10^6$ bit/sec) with an analogue computer is introduced. The generator described in [8] may serve as an example of simple laboratory device. In the latter paper application of this principle for producing continuous realizations is presented and a low-frequency generator of pseudo-random noise is described. The author does not use the low-pass filter, usually used for producing continuous realizations, but generates the step function, the amplitude of which is directly proportional to the number of pulses of one kind (0) in the register. It is assumed that the distribution of amplitudes of this signal is binomial and in the case of register of sufficient length it approximates the Gaussian distribution.

Other properties of the pseudo-random sequence generated by this method are equivalent to another methods of producing pseudo-random numbers, i.e. e.g.:

realization may be arbitrarily repeated,
it is possible to produce delayed realizations, etc.

For verification of some properties the prototype of the generator was constructed and the following measurements performed:

1. measurements of the maximal period with the minimal number of modulo 2 adders (further only one adder considered),
2. measurements of properties of generated sequence.

ad 1. Table 2 shows results of measurement of the length of period p of the output sequence as a function of both the length n of the register and the selection of places (n, k) for adding (only one adder modulo 2 is used). It means, that the adder is connected with the last (the n-th) stage and an arbitrary (say k-th) stage, where $1 \leqq k < n$, and the output of the adder is led to the input of the register. Measurements were carried out for $10 \leqq n \leqq 28$ and for the initial state of the register the combination 111 ... 1 was taken in all cases. As stated above, the necessary and

Table 2.

The length p of the period of the output sequence as a function of the number n of stages of the register and the selection of stages for adding (n, k), (initial combination 11 ... 1)

k \ n	10	11	12	13	14	15	16	17	18	19	20
1	889	1533	3255	7905	11811	32767	255	273	253921	413385	761763
2	42	2047	126	1785	254	4599	126	114661	146	129921	1778
3	1023	1953	45	8001	5115	63	57337	131071	189	491505	1048575
4	62	1533	28	7161	186	32767	60	1023	930	91749	84
5	15	595	819	6141	5461	35	16383	131071	32767	393213	75
6	62	595	18	7665	254	93	434	131071	42	520065	2046
7	1023	1533	819	7665	21	32767	63457	4599	262143	520065	779907
8	42	1953	28	6141	245	32767	24	35805	1022	47523	124
9	889	2047	45	7161	5461	93	63457	35805	27	174251	130305
10		1533	126	8001	186	35	434	4595	1022	174251	30
11			3255	1785	5115	32767	16383	131071	262143	47523	130305
12				7905	254	63	60	131071	42	520065	124
13					11811	4599	57337	1023	32767	520065	779907
14						32767	126	131071	930	393213	2046
15							255	114661	189	91749	75
16								273	146	491505	84
17									253921	129921	1048575
18										413385	1778
19											761763
$2^n - 1$	1023	2047	4095	8191	16383	32767	65535	131071	262143	524287	1048575

Table 2. (*Continuation*)

k \ n	21	22	23	24	25	26	27	28
1	5461	4194303	2088705	2097151	10961685	298936	125829105	17895697
2	2097151	3066	7864305	6510	25165821	15810	458745	23622
3	381	3670009	32767	189	33554431	2094081	219	268435455
4	406317	4094	2088705	2420	2158065	3570	5592405	508
5	5461	2752491	8388607	16766977	105	67074049	8877935	21082635
6	279	3906	458745	90	4185601	16002	1395	10230
7	49	4063201	2094081	1048575	33554431	13797	44564395	105
8	1966065	3066	2728341	56	8322945	14322	133693183	372
9	381	3899535	8388607	651	32247967	7469145	63	268435455
10	2088705	1190	87381	1638	155	12282	130023393	10922
11	2088705	33	126945	5586603	8257473	8371713	109226955	199753347
12	381	1190	126945	36	4161409	15330	1533	508
13	1966065	3899535	87381	5586603	4161409	39	130023393	268435455
14	49	3066	8388607	1638	8257473	15330	130023393	42
15	279	4063201	2728341	651	155	8371713	1533	268435455
16	5461	3906	2094081	56	32247967	12282	109226985	508
17	406317	2752491	458745	1048575	8322945	7449145	130023393	199753347
18	381	4094	8388607	90	33554431	14322	63	10922
19	2097151	3670009	2088705	16766977	4185601	13797	133693185	268435455
20	5461	3066	32767	2420	105	16002	44564395	372
21		4194303	7864305	189	2158065	67074049	1395	105
22			2088705	6510	33554431	3570	8877935	10230
23				2097151	25165821	2094081	5592405	21082635
24					10961685	15810	219	508
25						298936	458745	268435455
26							125829105	23622
27								17895697
$2^n - 1$	2097151	4194303	8388607	16777215	33554431	67108863	134217727	268435455

sufficient condition for obtaining the maximal period $p = 2^n - 1$ is that the polynomial (in this special case the trinomial)

$$f(x) = x^n + x^k + 1$$

be primitive. Therefore all primitive trinomials (in corresponding range of degrees) can be determined from Table 2. For other degrees (up to 127) some of the primitive polynomials can be found in [9].

ad 2. To determine the distribution of 0's and 1's in the sequence during a period, the following measurements were carried out.

a) Output sequence was divided into consecutive couples (i.e. every particular number of the output sequence is contained just in one couple) and the distribution of relative frequencies of particular types of couples (00, 01, 10, 11) in the maximal period was tested. In this case period (20,17) was measured and results are shown in Table 3. In this measurement, the period was divided into 10 parts with 10^5 bits in each and Table 3 shows results from each part. For testing χ^2 — test was used.

Table 3.

Measurements of relative frequencies of couples in the period (20,17)

k	Δ 00	Δ 01	Δ 10	Δ 11	χ_3^2	P
1	+339	−254	−256	+171	21·3	10^{-4}
2	− 86	+128	−110	+ 68	3·3	0·35
3	+150	−168	+229	−211	11·8	0·007
4	− 70	− 65	+117	+ 18	1·8	0·6
5	+ 12	− 41	− 48	+ 77	0·8	0·85
6	−170	+100	+109	− 39	4·1	0·25
7	− 13	+ 14	+ 61	− 62	0·63	0·89
8	+ 1	+145	−124	− 22	2·9	0·39
9	−108	+ 48	− 10	+ 70	1·52	0·68
10	+122	− 32	−155	+ 65	3·5	0·33

$$\overline{\sum 00} = \overline{\sum 01} = \overline{\sum 10} = \overline{\sum 11} = \sum 12500$$

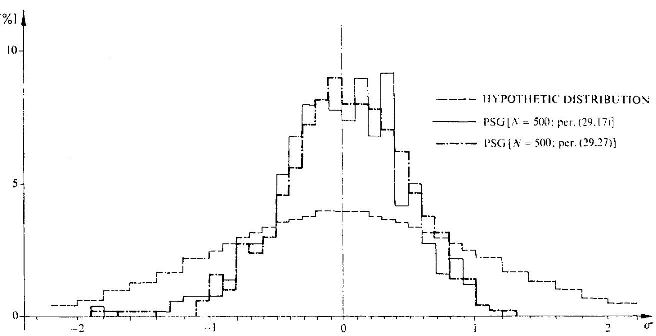

Fig. 3. Histogram of the distribution of relative frequencies of 1's in the complete period (29, 27) and in the incomplete period (29, 17).

b) The period was divided into intervals of length 10^4 bits each. In each interval, relative frequency of 1's was determined. Obtained values were normalised ($E = 0$, $D = 1$) and the hypothesis of normal distribution of these values was tested. Fig. 3 shows these results in histogram form compared with the theoretical normal distribution. In this figure results from periods (29, 27) and (29, 17) are stated. 500 trials were carried out in both cases. The period (29, 27) is a complete one. However, in

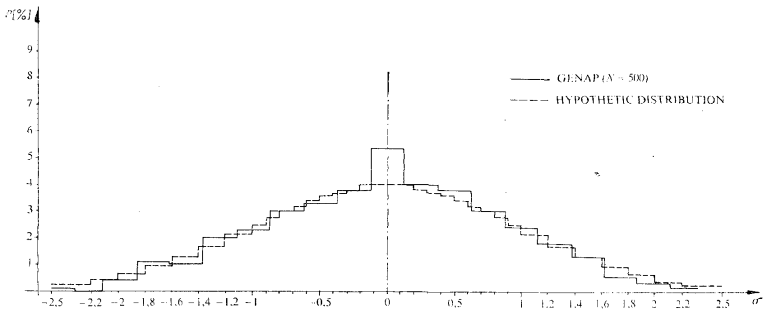

Fig. 4. Histogram of the distribution of relative frequencies of 1's in the sequence generated by GENAP (generator based on physical principles).

this case it seems, that if the incomplete period does not differ too much in length from the complete one, then even its properties might be similar to those of the complete period.

c) Let us assume that the hypothesis of binomial distribution of output pseudo-random sequences holds and consider property 5 from part I., then it is possible to use so-called probability transformer (cf. [10]) by means of which the original random sequence of 0's and 1's (where we consider $P(0) = P(1) = 0{,}5$) may be transformed into the random sequence of the same type with probabilities $P(0^*) = p$, $P(1^*) = 1 - p$, where $p = m\,2^{-10}$, $m = 0, 1, \ldots, 2^{10}$. A sequence of 0^*'s and 1^*'s with probabilities $P(0^*) = p = 2^{-10}$, $P(1^*) = 1 - p$ was realized by means of this transformer*. The distribution of the lengths of 1^*-run was tested. Assuming the binomial distribution of the input sequence, the probability of 1^*-run of length T smaller than or equal to $k - 1$ is

* In this special case occurrence of 0^* in the new sequences represents in the original sequence the 10-tuple of 0's. The function of the probability transformer is namely based on the comparison of the 10-digit binary random number with the prescribed number, determining the value p (for details see [10]).

$$P(T \leqq k - 1) = \sum_{j=0}^{k-1} (1 - p)^j \, p = \frac{1 - (1 - p)^k}{1 - 1 - p} = 1 - (1 - p)^k ,$$

$$P(T \leqq k - 1) = 1 - \mathrm{e}^{-\lambda k} ,$$

where $\lambda = -\ln(1 - p)$; in this case $p = 2^{-10}$ yields $\lambda \doteq 0{,}001$.

Altogether 10^3 numbers were gained and these were distributed into the groups according to T as may be seen in Table 4, where results of the Kolmogorov-Smirnov test of a good fit of the empirical distribution function $F_n(x)$ and the theoretical distribution function $F(x)$ are stated.

Table 4.

Distribution function of 1*-runs

x	$F_n(x)$	$F(x)$	$F_n(x) - F(x)$
100	0·197	0·095	0·102
200	0·254	0·181	0·073
400	0·377	0·330	0·047
600	0·503	0·451	0·052
800	0·573	0·551	0·022
1000	0·645	0·632	0·013
1200	0·706	0·699	0·007
1400	0·753	0·753	0·000
1600	0·793	0·798	0·005
1800	0·820	0·835	0·015
2000	0·854	0·865	0·011

III. CONCLUSIONS AND COMPARISON WITH THE PHYSICAL SOURCE OF RANDOM SIGNAL

As may be seen form Table 4, the Kolmogorov-Smirnov function $1 - K(\lambda)$ yields values much smaller than the significance level 0,05, ($1 - K(\lambda)$being smaller than $5 \,.\, 10^{-7}$). In Fig. 3 bad fit of the approximation with the theoretical distribution can be seen too. In both cases it may be seen, that the form of the distribution of the maximal (complete) period is similar to that of incomplete period when the latter is suficiently long. All deviations are concentrated practically in the range of $\pm\sigma$. On the other hand, some other properties required for generators of random binomial sequences are fulfilled very well by this method (see 1 – 5, part I.) as shown by χ^2-values from Table 3, too.

Measurement data obtained using the generator of random binomial sequences based on physical principles (GENAP) (cf. [11]) are given in Fig. 4. Measurements were carried out under the same conditions as in the case of the pseudo-random sequences, distribution of 1*-runs′ lengths for the physical generator was tested. In this case the Kolmogorov-Smirnov function $1 - K(\lambda)$ equals 0,8643.

The presented results can be used for comparison of both types of generators. The advantages of pseudo-random numbers (their fast generation, reproducibility, etc.) are countervailed by the restrictions imposed on further transformations of the pseudorandom sequence. The pseudo-random generator will in general give worse results than the generator based on physical principles, provided the connection with a filter is considered. The transformation of the original pseudorandom sequence into the $0^* - 1^*$ sequence with probabilities p, $1 - p$ is not too good for small p.

Finally it is necessary to point out, that all the measurements were carried out on sequences generated on the basis of trinomials. Properties of sequences derived from polynomials with more terms were not investigated.

(Received October 16th, 1965.)

REFERENCES

[1] S. W. Golomb: Sequences with Randomness Properties. Glenn L. Martin Co., Baltimore, Md, June 14, 1955.

[2] D. C. J. Poortvliet: The Measurement of System Impulse Response by Cross-correlation with Binary Signals. Technical University, Delft 1962.

[3] W. W. Peterson: Error-Correcting Codes. M.I.T. Press, 1961 – Russian translation, Moskva 1964.

[4] R. C. Tausworthe: Random Numbers Generated by Linear Recurrence Modulo Two. Math. of Computation 19 (Apr. 1965), 90, 201–209.

[5] R. L. T. Hampton: A Hybrid Analog-Digital Pseudo-Random Noise Generator. Analog Hybrid Computer Laboratory, The University of Arizona, Tuscon, Arizona.

[6] Marolf R. A.: 200 Mbit/s Pseudo-Random Sequence Generators for very Wide Band Secure Communication Systems. Proc. Nat. Electron, Conf. Chicago III, (1963), 183–187.

[7] Hampton L., Korn G. A., Mittchell B.: Hybrid Analog-Digital Random Noise Generation. IEEE Trans. on Electronic Comp. (1963), 412–413.

[8] C. Kramer: A Low Frequency Pseudo-Random Noise Generator. Electr. Eng. (1965), 465–467.

[9] Watson E. J.: Primitive Polynomials (mod 2). Mathematics of Computation XV (1962), 368–369.

[10] J. Havel: Měnič pravděpodobnosti (Probability Transformer). Slaboproudý obzor 24 (1963), 2, 83–88.

[11] J. Havel: Elektronický generátor náhodných posloupností (An Electronic Generator of Random Sequences). Slaboproudý obzor 20 (1959), 12, 735–740.

[12] Huffman D. A.: The Synthesis of Linear Sequential Coding Networks. Proc. 3rd London Symp. on Inf. Theory.

O jedné metodě vytváření pseudonáhodných čísel

ANTONÍN CULEK, JAN HAVEL, VÁCLAV PŘIBYL

Článek shrnuje některé poznatky, výsledky a možnosti využití metody generování pseudonáhodných čísel lineárním rekurentním způsobem při použití sečítání modulo 2. Pseudonáhodná čísla jsou vytvářena pomocí posuvného registru, ve kterém je zapsána určitá kombinace nul a jedniček. Z předem zvolených míst registru jsou pak dvojková čísla sečítána modulo 2, obsah registru je posunut o jedno místo a výsledek sečítání zaveden na první místo registru.

V prvé části článku je ve stručnosti uvedena charakteristika a vlastnosti tohoto způsobu generování. Druhá část obsahuje výsledky měření a testů, které byly získány na funkčním vzorku. Ukazuje se, že ačkoliv tento způsob generování splňuje velmi dobře některé základní vlastnosti, které jsou požadovány od generátorů náhodných binomických posloupností, neřídí se relativní četnosti nul a jedniček v dostatečně dlouhých realizacích Gaussovým zákonem rozložení s rozptylem, který by odpovídal korelační funkci pseudonáhodné posloupnosti. V závěru je provedeno srovnání výsledků této metody s výsledky dosahovanými generátorem náhodné binomické posloupnosti, který pracuje na fyzikálním principu.

Antonín Culek, Ing. Jan Havel CSc., Ing. Václav Přibyl, Ústav teorie informace a automatizace ČSAV, Praha 2, Vyšehradská 49.