Proceedings of the 1992 Winter Simulation Conference
ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson

FAST AND RELIABLE RANDOM-NUMBER GENERATION

Aaldert Compagner

Department of Physics, the Catholic University of America,
Washington D.C. 20064, U.S.A.

ABSTRACT

Based on an ensemble-theoretical definition of randomness and on the full hierarchy of correlation coefficients, a general strategy for the generation of random numbers is described. Use is made of well-tempered pseudorandom sequences, in which the remaining correlations are of very high order or cover extreme distances and thus are beyond the reach of usual testing procedures. In particular, the binary sequences produced by a cascade of a small number of shiftregisters characterized by primitive trinomials with Mersenne-prime degrees are suitable for large-scale Monte Carlo simulations.

1 INTRODUCTION

Due to the increasing availability of ever faster and more efficient machinery, only one obstacle remains in large-scale Monte Carlo simulation: the unreliable quality of the usual recipes for the generation of vast amounts of random numbers. Indeed, perfect recipes do not exist. For efficiency and general control all recipes have to be based on simple deterministic methods, in conflict with any notion of randomness.

Many different methods have been suggested; see the reviews given by Knuth (1981), Marsaglia (1985), Ripley (1990), James (1990), L'Ecuyer (1990), and Niederreiter (1991). The most common recipes are based on the linear-congruence, the lagged-Fibonacci or the binary-shiftregister method, or variants thereof. These have often been applied succesfully, depending on the special parameters of a particular recipe and on the simulation problem. Yet, also a number of accidents have occurred, in which hidden correlations between the random numbers interfered constructively with unknown correlations in the simulated system.

Testing a particular random-number generator is a rather time-consuming and unsatisfactory activity, especially because tests tend to be arbitrary and to fall short of what is required for large-scale Monte Carlo simulations. For instance, in the so-called spectral test the independence of only t consecutive numbers of, say, 32 bits is studied. Since t has to be restricted to values below 10 for practical reasons, at best only the correlations between 320 consecutive bits are proven to be small. This is much less than is needed to exclude the presence of systematic errors in the simulation results. Similar comments apply to other tests. In principle, a random-number generator should be tested in the context of the simulated system it is used for, but this advice is not very helpful. Improved methods for random-number generation will have to be based on theoretical arguments rather than tests.

In this paper, a new strategy for random-number generation by means of binary shiftregister sequences is reviewed (Compagner, 1991). In Section 2 general correlation coefficients are defined and in Section 3 it is shown how unavoidable correlations can be made harmless. Practical recipes based on results by Kurita and Matsumoto (1991), Heringa *et al.* (1992), and Wang and Compagner (1992) are given in Section 4. Concluding remarks are made in Section 5, where also a generalization to other methods is suggested.

2 ENSEMBLES AND CORRELATIONS

A convenient background to discuss random binary sequences is provided by ensemble theory. Consider a sequence $\{a_i(j)\}$ of N bits $a_1, ..., a_N$. It is identified by the number j of which it is the binary expansion ($j = 0, ..., 2^N-1$). An ensemble is defined by giving a weight p_j to each sequence j; the weights are zero or positive, and add up to 1. A usual measure for the amount of randomness is the ensemble entropy

$$S = - \sum_{\text{all } j} p_j \ln p_j . \qquad (1)$$

Three particular ensembles are of interest. The gambling ensemble is defined by $p_j = 2^{-N}$ for all j. It contains all sequences, with equal weights. The singular ensemble is given by $p_j = \delta_{jk}$ and consists of

the single sequence k. Finally, the scanning ensemble is defined by $p_{j'} = 1/N$ for all the N sequences j' that under cyclic translations are identical with a given sequence j. To avoid end effects the sequences are assumed to be periodic, $a_i(j') = a_{i+N}(j')$, with N as shortest period.

In analogy with the Ising model in statistical mechanics, a general correlation coefficient is defined in the following manner. First, the binary sequence j with elements $a_i(j) = 0, 1$ is replaced by the corresponding parity sequence with elements $b_i(j) = 1, -1$ respectively (the advantage is that the expected mean parity is 0 and absent from the formulae). Next, an ordered set $I(q, s) = \{i_1, ..., i_q\}$ of q different and fixed positions in the sequence is introduced, where q is the order and $s = i_q - i_1 + 1$ the size of the set. Finally, the correlation coefficient of order q and size s pertaining to set I is defined by

$$C_{I(q,s)} = \sum_{\text{all } j} p_j \prod_{i \in I} b_i(j) , \qquad (2)$$

the value of which lies on the segment $[1, -1]$.

It is easy to show that these coefficients obey a weak conservation law for the mean value (in which for most ensembles contributions of opposite sign largely cancel one another) and a strong one for the mean-square value (Compagner, 1991a):

$$\langle C_I \rangle = 2^{-N} \sum_{\text{all } I} C_I = p_0 , \qquad (3)$$

$$\langle C_I^2 \rangle = 2^{-N} \sum_{\text{all } I} C_I^2 = \sum_{\text{all } j} p_j^2 . \qquad (4)$$

The summations over all 2^N different sets I include the empty set $I(0, 0)$, the correlation coefficient of which is always 1.

In the gambling ensemble all true correlation coefficients are zero, as follows from (2) or (4), and the entropy (1) is maximal: $S = N \ln 2$. It is natural to identify this situation with complete randomness and total lack of information. The opposite is the case for the singular ensemble, in which all correlation coefficients are either 1 or -1; in a single sequence, everything is fixed and all sets are completely correlated or anticorrelated. The entropy is now equal to the minimum value $S = 0$. In this case, complete information about the sequence is available; in a given single sequence nothing is random.

A more subtle point of view is provided by the scanning ensemble, to which the attention is now restricted. In this case, a correlation coefficient measures the product of a particular set of parities at fixed relative positions, averaged over the full period of the sequence. For instance, for the sets $I = I(2, s)$ with $q = 2$ the quantity C_I as a function of s is just the usual paircorrelation function. The scanning-ensemble entropy is $S = \ln N$, which is far from maximal. Although the mean value (3) usually is exactly zero, the mean-square value (4) is $1/N$, which is small but not zero: not all correlation coefficients of a sequence can be arbitrarily small. In fact, a deterministic production rule for a sequence always corresponds to a completely correlated set, from which other such sets are generated. Eq. (4) shows that among the huge number of 2^N different sets that exist, these correlated sets are a small though crucial minority; finding them by means of testing is like looking at night for a needle in a haystack.

This is why tests of random-number generators are difficult and perhaps even misleading. The more one requires that a certain test is passed, the more one selects recipes and sequences such that the implicit correlations for which the test is sensitive are small, and the more one risks to make other correlations large for which the simulated system happens to be sensitive. To improve the quality control of random-number generators, the correlated sets that are tolerated need to be specified. Such a specification should in fact be part of any definition of randomness.

3 WELL-TEMPERED PSEUDORANDOMNESS

From these observations, an optimal strategy follows in just two steps. First, it is required that all sets below a certain size n are uncorrelated, where n should be as large as is possible. The maximum value is $n = \ln N / \ln 2$, which is the number of degrees of freedom (supposed to be integer) in a sequence of N bits; in a binary maximum-length sequence produced by a linear-feedback shiftregister, n is in fact equal to the length of the shiftregister. Consider a sequence of N bits in which all $2^n = N$ different strings of n bits occur just once, overlaps allowed; this is called here a pseudorandom sequence. For sets of size n or smaller, the scanning ensemble for a pseudorandom sequence of N bits is equivalent to the gambling ensemble for sequences of n bits. Therefore, the correlation coefficients for all these sets vanish. Pseudorandomness in this strict sense is almost identical with the usual condition of maximum length obeyed by many random-number generators. This condition is now seen to guarantee the vanishing of all correlation coefficients for sets of size n or smaller.

Many pseudorandom sequences produce random numbers of poor quality, and therefore a second step is necessary. Since n exhausts the available degrees of freedom, a pseudorandom sequence of N bits must

have non-zero correlation coefficients for sets of size $n+1$. A deterministic production rule is equivalent with a first completely correlated set of size $n+1$; iteration of the production rule to generate further elements of the sequence then leads in an increasingly stochastic fashion to all other correlated sets (all with correlation coefficient 1), which are of larger size. The best strategy is to require that the first correlated set, *i.e.* the production rule, is of high order. In addition, it should be sufficiently irregular to exclude that only after a few iterations other correlated sets arise that are of low order. Hence the definition:

> *A well-tempered pseudorandom sequence is a pseudorandom sequence with many degrees of freedom ($n>10^3$) in which all correlated sets are either of high order ($q>n^{1/2}$) or of very large size ($s>n^2$), or both.* (5)

The lower bounds given are rather arbitrary and depend on the circumstances. While $n>10^3$ may be too large for many purposes, it is too small for future large-scale simulations of great accuracy.

The reasons behind (5) are the following. All translated sets are equivalent in the scanning ensemble, and only first sets $I(q, s) = \{i_1=1, ..., i_q=s\}$ are of interest. The total number of first sets $I(q, s)$ is $(s-2)!/(q-2)!(s-q)!$, which means that q is Gaussian-distributed, with a maximum at $q = s/2$ and a width $s^{-1/2}$. In the stochastic region it may be assumed, at least for maximum-length sequences (Compagner, 1991a), that $N/(N+1)$ of these sets are uncorrelated whereas the remaining fraction $1/(N+1)$ is correlated (with $C_I = 1$).

Therefore, the correlated sets of size s obey the same Gaussian as all sets of size s, only the amplitude is different. Of course, the deterministic production rule leads to deviations from this stochastic behavior, but for irregular high-order rules the stochastic region is entered close to the maximum of the Gaussian, which acts as a centre of attraction (the order of the production rule only needs to be much larger then 1, it doesnot need to have the optimal value $n/2$). Hence, deviations in the form of correlated sets with rather small q can be expected to become important only when the total number 2^s of sets of size s is so large that the lower-q tail of the Gaussian has an appreciable magnitude.

If recipes for well-tempered pseudorandom sequences can be found, the main possibilities to achieve optimal randomness are exhausted. Apart from size and order there are no other general parameters to characterize correlated sets. The main difficulty lies in the notion of irregularity.

4 PRODUCTION RULES

It may seem that little progress is made by reducing the randomness of a sequence to the irregularity of a high-order production rule, but the condition that is implied has an operational nature and is rather weak. It is much weaker than the requirement of maximum complexity for the whole sequence, of which it is reminiscent and which is part of the ingeneous definition of randomness in a sequence proposed by Kolmogorov and, independenty, by Chaitin (1987). However, that definition is based on the very absence of any production rule and hence is not constructive. In contrast, high-order production rules that are both simple and irregular are easily constructed.

A binary shiftregister with linear feedback generates a maximum-length sequence when its characteristic function is a primitive polynomial over GF(2). The degree of the polynomial is one less than the size of the first correlated set (which represents the production rule) and the number of terms is equal to the order of that set. Take for instance a two-bit feedback shiftregister, characterized by a trinomial $R(n; k) = 1+x^k+x^n$ with $0<k<n$, where k indicates the additional feedback position. Starting from a seed of n bits, the sequence is generated by iteration of the production rule $a_{i+n} = a_i \oplus a_{i+k}$. The order and size of the first correlated set $\{a_i, a_{i+k}, a_{i+n}\}$ are $q = 3$ and $s = n+1$, and its parity product is $b_i\, b_{i+k}\, b_{i+n} = (b_i\, b_{i+k})^2 = 1$. When the trinomial is primitive, the sequence may be called pseudorandom because it has the maximum length $L \equiv N-1 = 2^n-1$.

Many primitive trinomials are known. In particular for degrees that are Mersenne exponents they are easy to find. Zierler (1969) provided a list up to degree $n = 9689$ of all primitive Mersenne trinomials. That list was enlarged up to degree $n = 44497$ by Kurita and Matsumoto (1991), and then up to $n = 132049$ by Heringa *et al.* (1992). In all, 59 primitive Mersenne trinomials are known, of which there are 19 for the 8 Mersenne exponents between 10^3 and 10^4, and 15 for the first 9 Mersenne exponents above 10^4 (further Mersenne exponents are unknown). Thus, a great variety exists of two-bit feedback production rules that generate pseudorandom sequences with a first correlated set of large size.

However, two-bit feedback rules are not of high order. The ill-tempered sequences they generate contain third-order correlations of the smallest possible size $n+1$, which explains why the shiftregister method has often been mistrusted. Primitive polynomials of high degree with many terms would solve this problem, but already for moderate degrees

these are difficult to find; they would also be difficult to implement. An effective solution was found in terms of reducible polynomials (Compagner, 1991).

When t sequences, each generated by a primitive Mersenne trinomial $R(n_i; k_i)$ of different degree n_i (increasing from n_1 to n_t), are added bitwise (mod 2), a sequence is obtained with a period Λ that is the product of the periods L_i of the original sequences, which are primes. The characteristic function of the resulting sequence is a reducible polynomial, with the primitive trinomials of the original sequences as factors. The degree n of this reducible polynomial is the sum of the degrees n_i of the trinomials. Its number of terms is $q = 3^t$, at least when this value is small compared with $n/2$, *i.e.* when the restriction to GF(2) causes only few chance cancellations; for larger values of 3^t, the order q of the reducible polynomial lags behind and usually levels off around $n/2$.

The length Λ of the resulting sequence is less than the maximum length 2^n-1 for a polynomial of degree n, but the relative difference is negligible when the smallest degree n_1 of the chosen trinomials is not too small. Then, the sequence effectively has at least the same randomness properties as a maximum-length sequence, and is at least almost pseudorandom (Wang and Compagner, 1992). Its production rule is of high order and very irregular: the exponents n_i (which are primes) and k_i that appear in the Mersenne trinomials are entirely accidental. Iteration of the production rule gives rise to further correlated sets of size $s > n$, but initially their order approaches $s/2$, where the Gaussian is maximal (see Section 3). Therefore, the resulting sequence may also be called well-tempered.

Choose $t = 4$ to 8 primitive trinomials $R(n_i; k_i)$ of different degrees from the table given by Heringa *et al.* (1992), say with $n_1 \approx 10^2$ and $n_t > 10^4$ while the other degrees are scattered in between. The reducible polynomial with $3^4 \approx 80$ to $3^8 \approx 6000$ terms that is formed by their product is the characteristic function of a well-tempered pseudorandom sequence with an extremely large period ($\Lambda > 10^{3000}$), in which bit-mixing takes place at many different length scales. An efficient implementation of the reducible polynomial is possible by means of a cascade (or, somewhat misleading in the present context, a binary tree) in which at each level pairs of sequences are 'exored'; 2 or 3 levels of the cascade suffice.

True, since only a small subsequence will ever be used in any simulation, the question of the uniform behavior of a well-tempered pseudorandom sequence arises. Uncorrelated sets, including low-order ones, may have a vanishing correlation coefficient when averaged over the whole period, but the same doesnot need to hold over a small subsequence. However, in principle also this problem is solved by irregular high-order production rules, which have a strong built-in tendency back to the normality of the Gaussian regime. Admittedly, this is not a proof. In fact, accidental and transient correlations of low order and rather small size can never be entirely excluded; their probability can be made small, but not smaller than befits the Gaussian regime.

In a few cases, the strategy outlined above has been compared with numerical data. For values of n below $\approx$ 100 the figures of merit introduced by Niederreiter and co-workers (Niederreiter, 1991) can be determined. Generators based on reducible polynomials of a certain degree with 3 primitive trinomials as factors, were found to have about the same figures of merit as optimal generators based on a single primitive polynomial of the same degree with relatively many terms (Wang and Compagner, 1992). For larger values of n, figures of merit are difficult to calculate and optimal polynomials are not available.

However, for n below $\approx 10^3$ many other numerical tests of generators based on reducible polynomials are possible. A suitable battery of tests for that purpose, consisting of more than 20 different tests (not counting variants) and including many well-known ones, was developed by Berdnikov and Turtia (1992). Preliminary results indicate that under these tests reducible polynomials consisting of a few Mersenne trinomials behave as expected.

5 CONCLUDING REMARKS

For large-scale Monte Carlo simulations values of n above at least 10^3 are necessary, but for larger n testing becomes increasingly difficult. A new type of test, based on a local entropy that is defined in terms of frequencies encountered in subsequences, will perhaps turn out to be useful. Also, a more detailed description, both theoretical and numerical, of the quasi-stochastic process in which correlated sets are generated by a production rule would be worthwile.

Random numbers are often generated by recipes in which sequences of integers modulo m are used. An analogy with a system of m-valued spins instead of the Ising model may now serve as a guideline. In a forthcoming paper (Compagner *et al.*, 1992) it will be shown that the definitions of Section 2 for the correlation coefficients can be generalized by using the m-th roots of the unity instead of the parities. It turns out that the general correlation coefficients, which are complex quantities lying on or in the unit circle, still obey conservation laws like (3) and (4), and that again order and size can be introduced as the

characteristic parameters of correlated sets.

In fact, the general correlation coefficients are just a new interpretation of the characteristic function of a discrete multivariate distribution, where the elements of the sequence are the stochastic variables. A close relation exists also with the Fourier analysis of random sequencies carried out by Coveyou and MacPherson (1967) and with the spectral test that they introduced. It seems that the problem has turned full circle, to the next sheet of a Riemann surface.

The connection with the spectral test was noted in an interesting paper by O.E.Percus and J.K.Percus (1992), in which the correlation properties of sequences produced by two-bit feedback shiftregisters are compared with those of sequences generated by the linear-congruence method. Similar studies for sequences generated with the lagged-Fibonacci method would be very useful.

Most methods generate maximum-length sequences or good approximations thereof, which may be called pseudorandom. Whether the sequences are well-tempered remains to be seen. Probably, more terms have to be included in the linear-congruence rule, or multiple lags in the lagged-Fibonacci method.

A three-level cascade in which the bits generated by 8 widely different Mersenne trinomials are added (mod 2) is one of the best random-number generators available. A great variety of these trinomials exists, and the number-theoretical scatter in their exponents is excessive. The effective production rule is very irregular, of large order and of high degree. Such a cascade can be implemented in a single VLSI chip, the existence of which would allow Monte Carlo simulations that are at par with the machines used.

ACKNOWLEDGEMENTS

The author is indebted to A.S. Berdnikov, S.B. Turtia, O.E. Percus and J.K. Percus for helpful discussions, and to F. Berwald and Y.T.J.C. Fonk for technical assistance. This research was partially supported by NATO (grant CRG 900661) and by STW, the Dutch Technology Foundation (project DTN 22.2615/B).

REFERENCES

Berdnikov, A.S., and S.B. Turtia. 1992. *Vector random number generators: testing procedures and comparison of algorithms*. Report. St. Petersburg: Institute for Analytical Instrumentation.

Chaitin, G.J. 1987. *Information randomness and incompleteness*. Singapore: World Scientific.

Compagner, A. 1991a. Definitions of randomness. *Am. J. of Phys.* 59:700-705.

Compagner, A. 1991b. The hierarchy of correlations in random binary sequences. *J. of Stat. Phys.* 63:883-896.

Compagner, A., D. Wang and J.R. Heringa. 1992. The hierarchy of correlations in sequences of random integers. *J. of Comp. Phys.*, to be published.

Coveyou, R.R., and R.D. MacPherson. (1967). Fourier analysis of uniform random number generators. *J. Ass. Comp. Mach.* 14:100-119.

Heringa, J.R., H.W.J. Blöte and A. Compagner. 1992. New primitive trinomials of Mersenne-exponent degrees for random-number generation. *Int. J. of Modern Phys.* C3:561-564.

Knuth, D.E. 1981. *The art of computer programming.* Vol. 2, Ch. 3. Reading (Mass.): Addison-Wesley.

Kurita, Y., and M. Matsumoto. 1991. Primitive t-nomials (t = 3, 5) over GF(2) whose degree is a Mersenne exponent ≤ 44497. *Math. Comp.* 56:817-821.

L'Ecuyer, P. 1990. Random numbers for simulation. *Commun. ACM* 33(10):85-97.

Marsaglia, G. 1985. A current view of random number generators. In *Computer science and statistics*, ed. L.Billard. Amsterdam: Elsevier.

Niederreiter, H. 1991. Recent trends in random number and random vector generation. *Ann. Oper. Res.* 31:323-345.

Percus, O.E., and J.K. Percus. 1992. An expanded set of correlation tests for linear congruential random number generators. *Combinatorics, Probability and Computing* 1: to be published.

Ripley, B.D. 1990. Thoughts on pseudorandom number generators. *J. Comp. Appl. Math.* 31:153.

Wang, D., and A. Compagner. 1992. On the use of reducible polynomials as random number generators. *Math. of Comp.*, to be published.

Zierler, N. (1969). Primitive trinomials whose degree is a Mersenne exponent. *Inform. Control* 15:67-69.

AUTHOR BIOGRAPHY

AALDERT COMPAGNER is visiting professor at the Catholic University of America in Washington DC, on leave of absence from the Laboratory of Applied Physics of the Technological University in Delft, the Netherlands. He is one of the editors of the Dutch Journal of Physics, and his research includes problems in statistical mechanics and computational physics.