RAPID COMMUNICATIONS

PHYSICAL REVIEW E VOLUME 48, NUMBER 6 DECEMBER 1993

# Bit-level correlations in some pseudorandom number generators

K. Kankaala,[1,2] T. Ala-Nissila,[1,3] and I. Vattulainen[1,3]

[1] *Department of Electrical Engineering, Tampere University of Technology, P.O. Box 692, FIN-33101 Tampere, Finland*

[2] *Centre for Scientific Computing, P.O. Box 405, FIN-02101 Espoo, Finland*

[3] *Research Institute for Theoretical Physics, P.O. Box 9 (Siltavuorenpenger 20 C), FIN-00014 University of Helsinki, Finland*

(Received 24 August 1993)

We present results of extensive bit-level tests on some pseudorandom number generators which are commonly used in physics applications. The generators have first been tested with an extended version of the $d$-tuple test. Second, we have developed a *cluster test* where a physical analogy of the binary numbers with the two-dimensional Ising model has been utilized. We demonstrate that this new test is rather powerful in finding periodic correlations on bit level. Results of both test methods are presented for each bit of the output of the generators. Some generators exhibit clear bit-level correlations but we find no evidence of discernible correlations for generators, which have recently produced systematic errors in Monte Carlo simulations.

PACS number(s): 02.70.Lq, 05.50.+q, 75.40.Mg

Vast amounts of random numbers are needed in several applications such as stochastic optimization [1] and Monte Carlo simulations [2]. Modern high speed computers have set rigorous demands for the quality of random numbers, which are usually produced by pseudorandom-number-generator algorithms. A prerequisite to the success of the methods is the quality of randomness of the output of the generators. It is usually determined by statistical tests [3]. Usually many such tests are needed, since there is no unique recipe for determining when a given sequence is "random enough."

Unfortunately, even comprehensive statistical testing cannot guarantee that a given random-number generator is reliable for all applications. In fact, tests are needed which would be more *physical*, based on the use of generators in solving actual physical problems. A few such application-specific tests have been performed [4–7]. In particular, intriguing results have been reported by Ferrenberg *et al.* [8], who employed some of the most commonly used random-number generators in simulations of the two-dimensional Ising model at criticality. When using the Wolff algorithm [9] they reported anomalously large errors with a particular generator, called R250. The same conclusion has been drawn from simulations of self-avoiding random walks [10], where also other similar generators failed.

Although there have been prior warnings against the use of shift register generators such as R250 [11,12], the results of Refs. [8,10] are surprising, since recent extensive statistical tests have found no discernible correlations in R250 [13]. In Ref. [8], the authors suggest that bit-level correlations in the most significant bits of R250 may be responsible for their results. If true, this casts serious doubt on the bit-level reliability of R250. More and also better tests are then needed to resolve the issue.

The purpose of the present work is to study bit-level correlations in some commonly used generators in more detail. To this end we have first extended the $d$-tuple test [12,14] to more efficiently find correlations. Second, we have developed a physical *cluster test* which is based on an analogy to the Ising model. The test is implemented on bit level and its effectiveness compared with the $d$-tuple test. We demonstrate that the cluster test is particularly powerful in finding periodic correlations. Both the $d$-tuple and the cluster tests are then applied to each bit of a number of generators, including some of the shift-register generators in Refs. [8,10]. Our results demonstrate that no discernible bit-level correlations can be found in the shift-register generators with the present test methods.

The pseudorandom-number generators used here include two linear congruential generators, LCG(16807,0,$2^{31}-1$) [15] known as GGL (CONG in Ref. [8]), and LCG(69069,1,$2^{32}$) [16] implemented as RAND [17]. Additionally, RAN3 [18] is a LF(55,24,$-$) based on a lagged Fibonacci algorithm, whereas RANMAR [19,20] is a combination generator. Finally, GFSR(250,103,$\oplus$) and GFSR(1279,216,$\oplus$) are generalized feedback shift-register generators known as R250 [4,21] and R1279, respectively. The details of the algorithms can be found, for example, in Ref. [13]. We note that the generators were implemented to produce integers except for RANMAR, whose 24-bit reals were multiplied by $2^{31}-1$. Initial seed values were chosen from the set $\{14159, 667790, 1415926535, 95141\}$, excluding R250 and R1279 which were initialized with GGL in double-precision accuracy.

The $d$-tuple test is based on studying the properties of random numbers on bit level [12]. Our realization follows Ref. [14]. The main difference here is the improvement to calculate the $\chi^2$-distributed test statistics a total of $N$ times and submit their empirical distribution to a Kolmogorov-Smirnov (KS) test. The final test variables are therefore the values $K^+$ and $K^-$ of a KS-test statistic $K$ [3]. In each test the sequence of bits was considered to fail if the observed descriptive level $\delta = P(K \leq \{K^+, K^-\}|H_0)$ was less than 0.05 or larger than 0.95.

Based on previous work [13] the $d$-tuple test seems to find correlations more efficiently than the rank test [11,12], for example. In order to determine the quantitative effectiveness of the test we have first studied its

TABLE I. Results of the $d$-tuple test with inserted correlations in the bits, with a period of $\xi$. The probability for the test to observe correlations is denoted by $p$, which equals 1 up to $\xi_c \approx 43$.

| $\xi$ | 40 | 43 | 52 | 60 | 70 | 80 | 90 | 100 | 110 | 120 |
|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 1.000 | 0.889 | 0.778 | 0.333 | 0.667 | 0.222 | 0.333 | 0.111 | 0.222 | 0.000 |

ability to observe correlations inserted into the output of GGL, which passes the standard bit tests [13]. The correlations have been inserted periodically by setting the $i$th bit ($i = 1, 2, \ldots, 31$) of every $\xi$th number always equal to 1. By systematically varying $\xi$, we can then find the maximum approximate distance $\xi_c$ within which the $d$-tuple test can detect periodic correlations. The test was repeated three times with parameters $d = t = 3$, $n = 5000$, and $N = 1000$, where $d$ and $t$ are taken from Refs. [13,14] and $n$ is the number of samples in a single $\chi^2$ test. The results are shown in Table I, where the parameter $p$ gives the probability of observing correlations. Thus, the $d$-tuple test can always detect periodic correlations up to $\xi_c \approx 43$ bits apart. The same test was repeated with $d = 9$ and $t = 1$ to consider single bits only, which gave $\xi_c \approx 50$. We also note that we performed similar systematic tests for the rank test, which was found to be inferior to the $d$-tuple test.

To improve the detection range of the $d$-tuple test we have performed its *extended version.* This can be realized by testing bits from every $k$th number and then testing all $k$ such subsequences. This way all periodic correlations may be detected up to about $k\xi_c$ (assuming $k < \xi_c$). We have applied this extended test to GGL, R250, R1279, and RAN3, which were all tested twice. The results are summarized in Table II. The most remarkable result is that, up to $k = 20$, which corresponds to a distance of about 860 bits apart, no discernible correlations were observed for the 16 most significant bits of either R250 or R1279. In addition, we tested R250 with $k = 50$, $k = 100$, and $k = 1000$ where only one subsequence was studied in each case. No evidence of correlations was found. This result is in contrast to Ref. [10], where it was estimated that for R250 a typical range of correlations is about 400. However, when initialized with RAN3, which itself contains correlated bits, both R250 and R1279 display clear bit-level correlations, although the longer feedback of R1279 seems to be less sensitive to initial correlations.

There is a natural analogy between binary numbers and the Ising model, which can be made use of in constructing a physical *cluster test* in the following way. We take $i$th bits from every successive number and put them on a two-dimensional lattice of size $L^2$. By identifying zeros and ones with the "down" and "up" spins of the Ising model, the resulting configuration—if truly random—should be one of the $2^{L^2}$ equally weighted configurations corresponding to infinite temperature. The easiest quantity that one can then compute from this analogy is the magnetization. However, a better measure of *spatial* correlations can be obtained if we study the distribution of connected spins, or clusters of size $s$ on the lattice. The cluster size distribution $\langle n_s \rangle$ is given by [22]

$$\langle n_s \rangle = s p^s D_s(p), \tag{1}$$

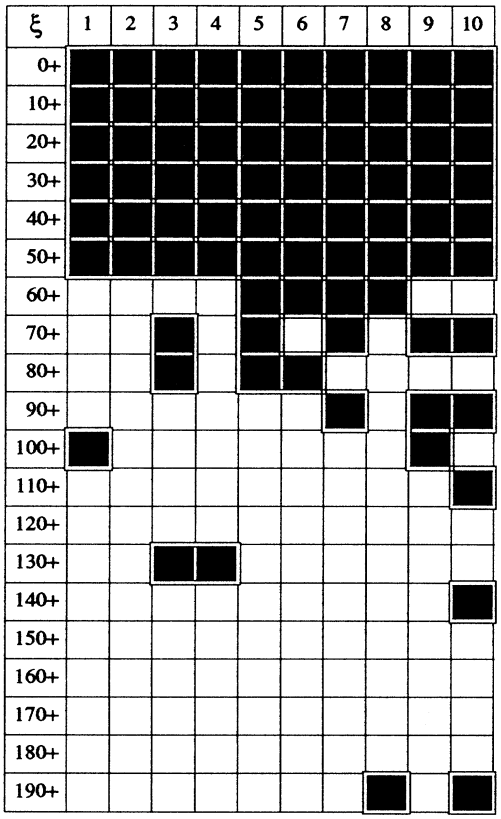
where $D_s(p)$'s are polynomials in $p = 1/2$. The normalization condition is $\sum_{s=1}^{\infty} \langle n_s \rangle = 1$. Enumeration of the polynomials $D_s(p)$ has been done up to $s = 17$ [22].

The test procedure we have used is as follows. We first form a $L^2$ lattice as above and enumerate all the clusters in it. For such a configuration we calculate the (unnormalized) average size of clusters within $s = 1, 2, \ldots, 17$, denoted as $S_{17}^{(k)}$. This procedure is repeated $M$ times corresponding to configurational averaging, yielding $S_{17} = \sum_{k=1}^{M} S_{17}^{(k)}/M$. The theoretical counterpart of this quan-

TABLE II. Results of the extended $d$-tuple tests. $k$ denotes the extended range of the tests. See text for details.

| Random-number generator | $k$ | Failing bits in the $d$-tuple test | Comments |
|---|---|---|---|
| GGL | 1,5 | none | Double precision mode (return integers) |
| R250 | 1,5 | none | Integer mode, initialized with GGL in double precision |
| R250 | 20 | none | only 16 most significant bits were studied |
| R1279 | 1,5 | none | Integer mode, initialized with GGL in double precision |
| R1279 | 20 | none | only 16 most significant bits were studied |
| RAN3 | 1 | 1–5, 25–30 | Integer mode |
| R250 | 1 | 1–2, 27–31 | Integer mode, initialized with RAN3 producing integers |
| R1279 | 1 | 1 | Integer mode, initialized with RAN3 producing integers |

TABLE III. Results of the cluster test with correlations in the bits, with a period of $\xi$ from 1 to 200. Black squares denote corresponding distances at which correlations were found as explained in the text.




tity is given by $s_{17} = \sum_{s=1}^{17} s\langle n_s \rangle$. We also compute the empirical standard deviation $\sigma_{17}$ of the quantities $S_{17}^{(k)}$. For each $i$th bit the test statistic chosen in this work is

$$g_i = \frac{S_{17} - s_{17}}{\sigma_{17}}. \tag{2}$$

Using this statistic, tests were performed comparatively between several pseudorandom-number generators, with results from GGL assumed to be independent variables [23]. Therefore, the mean value of $g_i$ over all the 31 bits of GGL, denoted as $g_{\mathrm{GGL}}$ and the corresponding standard deviation $\sigma_{\mathrm{GGL}}$ were computed and the results for all other generators were compared with these values using

$$g_i' = \frac{|g_i - g_{\mathrm{GGL}}|}{\sigma_{\mathrm{GGL}}}. \tag{3}$$

The bit $i$ in question failed the test if $g_i'$ was greater than 3. We also considered other similar choices for the test parameters and criteria and obtained consistent results.

The effectiveness of the cluster test was first scrutinized by inserting periodic correlations as in the case of the $d$-tuple test. We chose $L = 200$, $M = 10\,000$ and the study was repeated for all values of $\xi = 1, 2, \ldots, L$. The results are shown in Table III, where filled squares denote distances where correlations were detected. With this choice of parameters the cluster test is able to find all periodic correlations up to $\xi_c \approx 60$. This shows that the cluster test performs somewhat better than either the $d$-tuple or rank tests.

Next, we have subjected each bit of the random-number generators to the cluster test. It was repeated twice with parameters $L = 200$ and $M = 10\,000$. Additional tests with $L = 500$ gave consistent results. Results are summarized in Table IV, where also results of the previous $d$-tuple and rank tests from Ref. [13] have been included. Although more powerful than the other methods, the cluster test still reveals no discernible correlations for either GGL, R250, or R1279. For RANMAR and RAN3, the cluster test gives results consistent with Ref. [13], but for RAND additional correlations are revealed in bits 8–12, which passed the $d$-tuple test.

For completeness, we also tested the distribution of bits. The bits failed the test if the deviation from the expected number of 1's (i.e., $L^2/2$) consecutively exceeded three times the standard deviation of the binomial distribution with $M$ samples. The test was repeated twice with $M = 4 \times 10^8$, and its results are also shown in Table IV. No correlations were found for GGL, R250, or R1279. Surprisingly, however, this rather simple test revealed that the first 11 bits of RAN3 fail (with standard deviations larger than 6.7) although only the first four or five bits fail in the other tests. This signals correlations in these additional bits. On the other hand, for RAND only bits 22–31 failed, which produced an exact 50-50 distribution of zeros and 1's.

In conclusion, we have performed extensive bit-level tests of several commonly used pseudorandom-number generators, including R250 which had been suggested to

TABLE IV. Results of the cluster test. $d$-tuple and rank test results are from Ref. [13]. The last column denotes bits which fail in testing the distribution of 1's.

| Random-number generator | Failing bits | | | |
|---|---|---|---|---|
| | Cluster test | $d$-tuple test | Rank test | Distribution of bits |
| GGL | none | none | none | none |
| R250 | none | none | none | none |
| R1279 | none | none | none | none |
| RANMAR | 25–31 | 25–31 | 25–31 | 25–31 |
| RAN3 | 1–4, 25–30 | 1–5, 25–30 | 1–5, 26–30 | 1–11, 24–30 |
| RAND | 8–31 | 13–31 | 18–31 | 22–31 |

contain bit-level correlations [8]. To this end, we have performed an extended version of the $d$-tuple test, and developed a physical *cluster test*, which is rather powerful in finding periodic correlations. Our results reveal significant bit-level correlations in some generators, such as RAN3 and RAND, but absolutely no discernible correlations in GGL, R250, or R1279. Thus, our results show that these generators should be good enough for many applications, where good bit-level properties *of their individual bits* are required. However, we note that it is still of crucial importance to further develop physical tests along the lines presented here to detect more subtle correlations, which may not be revealed by the present test methods.

[1] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines, A Stochastic Approach to Combinatorial Optimization and Neural Computing* (Wiley, Chichester, 1989).

[2] K. Binder, in *Monte Carlo Methods in Condensed Matter Physics*, edited by K. Binder (Springer-Verlag, Berlin, 1992).

[3] D. E. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, 2nd ed. (Addison-Wesley, Reading, MA, 1981).

[4] S. Kirkpatrick and E. P. Stoll, J. Comp. Phys. **40**, 517 (1981).

[5] C. Kalle and S. Wansleben, Comput. Phys. Commun. **33**, 343 (1984).

[6] J. Paulsen, J. Stat. Comput. Simul. **19**, 23 (1984).

[7] A. Milchev, K. Binder, and D. W. Heermann, Z. Phys. B **63**, 521 (1986).

[8] A. M. Ferrenberg, D. P. Landau, and Y. J. Wong, Phys. Rev. Lett. **69**, 3382 (1992).

[9] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).

[10] P. Grassberger, J. Phys. A **26**, 2769 (1993); Wuppertal University Report No. WU-B 93-03, 1993 (unpublished).

[11] G. Marsaglia and L.-H. Tsay, Linear Algebra Appl. **67**, 147 (1985).

[12] G. A. Marsaglia, in *Computer Science and Statistics: The Interface*, edited by L. Billard (Elsevier, Amsterdam, 1985), p. 3.

[13] I. Vattulainen, K. Kankaala, J. Saarinen, and T. Ala-Nissila, University of Helsinki Reports Nos. HU-TFT-93-1 (1993) and HU-TFT-93-22 (1993) (unpublished).

[14] N. S. Altman, SIAM J. Sci. Stat. Comput. **9**, 941 (1988).

[15] S. K. Park and K. W. Miller, Commun. ACM **31**, 1192 (1988); P. A. Lewis, A. S. Goodman, and J. M. Miller, IBM Syst. J. **8**, 136 (1969).

[16] G. Marsaglia, in *Applications of Number Theory to Numerical Analysis*, edited by S. K. Zaremba (Academic Press, New York, 1972), p. 249.

[17] *Convex Fortran Guide*, 1st ed. (Convex Computer Corp., Richardson, 1991), p. 553.

[18] W. H. Press, B. P. Flannery, S. A. Tenkolsky, and W. T. Vetterling, *Numerical Recipes, The Art of Scientific Computing*, Fortran version (Cambridge University Press, Cambridge, 1989), p. 198.

[19] F. James, Comput. Phys. Commun. **60**, 329 (1990).

[20] G. Marsaglia and A. Zaman, Stat. Prob. Lett. **8**, 329 (1990).

[21] T. G. Lewis and W. H. Payne, J. Assoc. Comput. Mach. **20**, 456 (1973).

[22] M. F. Sykes and M. Glen, J. Phys. A **9**, 87 (1976).

[23] We checked numerically that, to a good degree of approximation, the distribution for $S_{17}^{(k)}$ is given by a Gaussian for GGL.