

An experimental exploration of Marsaglia's xorshift generators, scrambled

Sebastiano Vigna, Università degli Studi di Milano, Italy

Marsaglia proposed recently `xorshift` generators as a class of very fast, good-quality pseudorandom number generators. Subsequent analysis by Panneton and L'Ecuyer has lowered the expectations raised by Marsaglia's paper, showing several weaknesses of such generators, verified experimentally using the TestU01 suite. Nonetheless, many of the weaknesses of `xorshift` generators fade away if their result is scrambled by a non-linear operation (as originally suggested by Marsaglia). In this paper we explore the space of possible generators obtained by multiplying the result of a `xorshift` generator by a suitable constant. We sample generators at 100 equispaced points of their state space and obtain detailed statistics that lead us to choices of parameters that improve on the current ones. We then explore for the first time the space of high-dimensional `xorshift` generators, following another suggestion in Marsaglia's paper, finding choices of parameters providing periods of length $2^{1024} - 1$ and $2^{4096} - 1$. The resulting generators are of extremely high quality, faster than current similar alternatives, and generate long-period sequences passing strong statistical tests using only eight logical operations, one addition and one multiplication by a constant.

Categories and Subject Descriptors: G.3 [PROBABILITY AND STATISTICS]: Random number generation; G.3 [PROBABILITY AND STATISTICS]: Experimental design

General Terms: Algorithms, Experimentation, Measurement

Additional Key Words and Phrases: Pseudorandom number generators

1. INTRODUCTION

`xorshift` generators are a simple class of pseudorandom number generators introduced by Marsaglia [2003]. In Marsaglia's view, their main feature is speed: in particular, a `xorshift` generator with a 64-bit state space generates a new 64-bit value using just three 64-bit shifts and three 64-bit xors (i.e., exclusive ors), thus making it possible to generate hundreds of millions of values per second.

Subsequent analysis by Brent [2004] showed that the bits generated by `xorshift` generators are equivalent to certain *linear feedback shift registers*. Panneton and L'Ecuyer [2005] analyzed moreover in detail the generators using the TestU01 suite [L'Ecuyer and Simard 2007], finding weaknesses and proposing an increase in the number of shifts, or combination with another generator, to improve quality.

In the first part of this paper we explore experimentally the space of `xorshift` generators with 64-bit state space using statistical test suites. We sample generators at 100 equispaced points of their state space, to easily identify spurious failures. There are 2200 possible full-period `xorshift` generators, due to 275 possible values for its three shift parameters, and eight possible algorithms (see Figure 1); Marsaglia proposes some choice of parameters, that, as we will see, and as already reported by Panneton and L'Ecuyer [2005], are not particularly good. We report results that are actually worse than those of Panneton and L'Ecuyer as we use the entire 64-bit output of the generators. While we can suggest some good parameter choices, the result remains poor.

Thus, we turn to the idea of scrambling the result of a `xorshift` generator using a multiplication, as it is typical, for instance, in the construction of practical hash function due to the resulting *avalanching* behavior (bits of the result depend on several bits of the

This work is supported the EU-FET grant NADINE (GA 288956).

Author's addresses: Sebastiano Vigna, Dipartimento di Informatica, Università degli Studi di Milano, via Comelico 39, 20135 Milano MI, Italy.

	C code	
A_0	$x \hat{=} x \ll a; x \hat{=} x \gg b; x \hat{=} x \ll c;$	\mathbf{X}_1
A_1	$x \hat{=} x \gg a; x \hat{=} x \ll b; x \hat{=} x \gg c;$	\mathbf{X}_3
A_2	$x \hat{=} x \ll c; x \hat{=} x \gg b; x \hat{=} x \ll a;$	\mathbf{X}_2
A_3	$x \hat{=} x \gg c; x \hat{=} x \ll b; x \hat{=} x \gg a;$	\mathbf{X}_4
A_4	$x \hat{=} x \ll a; x \hat{=} x \ll c; x \hat{=} x \gg b;$	\mathbf{X}_5
A_5	$x \hat{=} x \gg a; x \hat{=} x \gg c; x \hat{=} x \ll b;$	\mathbf{X}_6
A_6	$x \hat{=} x \gg b; x \hat{=} x \ll a; x \hat{=} x \ll c;$	\mathbf{X}_7
A_7	$x \hat{=} x \ll b; x \hat{=} x \gg a; x \hat{=} x \gg c;$	\mathbf{X}_8

Fig. 1. The eight possible `xorshift64` algorithms. The list is actually derived from Panneton and L’Ecuyer [2005], as they correctly remarked that two of the eight algorithms proposed by Marsaglia were redundant, whereas two (A_6 and A_7) were missing. On the right side we report the name of the linear transformation associated to the algorithm as denoted by Panneton and L’Ecuyer [2005]. With our numbering, algorithms A_{2i} and A_{2i+1} are conjugate by reversal. Note that contiguous shifts in the same direction can be exchanged without affecting the resulting algorithm. We normalized such contiguous shifts so that their letters are lexicographically sorted.

input). This can be seen as the composition of a `xorshift` generator with a *multiplicative linear congruential generator*, and is actually suggested in passing in Marsaglia’s paper. The third edition of the classic “Numerical Recipes” [Press et al. 2007], indeed, proposes this construction for a basic, all-purpose generator. Since a lot of knowledge has been gathered in the last 50 years on multiplicative constants that give have *good spectral properties*, we use multipliers taken from [L’Ecuyer 1999].

From the wealth of data so obtained we derive generators with better statistical properties than those suggested in “Numerical Recipes”. We also investigate several interesting correlations, such as those between the weight of the characteristic polynomial and failures in statistical test suites.

In the last part of the paper, we follow the suggestion about high-dimensional generators contained in Marsaglia’s paper, and compute for the first time several choices of parameters that provide full-period `xorshift` generators with a state space of 1024 and 4096 bits. Once again, we propose generators that use a multiplication to scramble the result.

At the end of the paper, we apply our methodology to a number of popular non-cryptographic generators, and we discover that our high-dimensional generators are actually faster and of higher or equivalent statistical quality, as assessed by statistical test suites, than the alternatives.

The software used to perform the experiments described in this paper is distributed by the author under the GNU General Public License. Moreover, all files generated during the experiments are available from the author. They contain a large amount of data that could be further analyzed (e.g., by studying the distribution of p -values over the seeds). We leave this issue open for further work.

2. AN INTRODUCTION TO `xorshift` GENERATORS

The basic idea of `xorshift` generators is that the state space is modified by applying repeatedly a shift and an exclusive-or (xor) operation. In this paper we consider 64-bit shifts and state spaces of 2^n bits, with $n \geq 6$. We usually append n to the name of a family of generators when we need to restrict the discussion to a specific state-space size.

For `xorshift64` generators Marsaglia suggests a number of possible combination of shifts, shown in Figure 1. Not all choices of parameters give a full $(2^{64} - 1)$ period: there are 275 suitable choices of a , b and c and eight variants, totalling 2200 generators.

In linear-algebra terms, if L is the 64×64 matrix on $\mathbf{Z}/2\mathbf{Z}$ that effects a left shift of one position on a binary vector (i.e., L is all zeroes except for ones on the principal subdiagonal) and if R is the right-shift matrix (the transpose of L), each left/right shift/xor can be

described as a linear multiplication by $(1 + L^s)$ or $(1 + R^s)$, respectively, where s is the amount of shifting.¹ For instance, algorithm A_0 of Figure 1 is equivalent to the $\mathbf{Z}/2\mathbf{Z}$ -linear transformation

$$\mathbf{X}_1 = (1 + L^a)(1 + R^b)(1 + L^c).$$

It is useful to associate with a linear transformation M its *characteristic polynomial*

$$P(x) = \det(M - x).$$

The associated generator has maximum-length period if and only if $P(x)$ is primitive over $\mathbf{Z}/2\mathbf{Z}$. This happens if $P(x)$ is irreducible and if z has maximum period in the ring of polynomial over $\mathbf{Z}/2\mathbf{Z}$ modulo $P(x)$, that is, if the powers z, z^2, \dots, z^{2^n-1} are distinct modulo $P(x)$. Finally, to check this condition is sufficient to check that

$$x^{(2^n-1)/p} \neq 1 \pmod{P(x)}$$

for every prime p dividing 2^n-1 [Lidl and Niederreiter 1994].

The *weight* of $P(x)$ is the number of terms in $P(x)$, that is, the number of nonzero coefficients. It is considered a good property for generators of this kind that the weight is close to $n/2$, that is, that the polynomial is neither too sparse nor too dense [Compagner 1991]. For this reason, if a generator has characteristic polynomial $P(x)$ of degree d with weight W we define the *weight score* of the generator as

$$|W - d/2|,$$

so a low weight score is better.

Note that the family of algorithms of Figure 1 is intended to generate *64-bit values*. This means that the entire output of the algorithm should be used when performing tests. We will see that this has not always been the case in previous literature.

3. SETTING UP THE EXPERIMENTS

In this paper we want explore experimentally the space of a number of xorshift-based generators. Our purpose is to identify variants with full period which have particularly good statistical properties, and test whether claims about good parameters made in the previous literature are confirmed.

The basic idea is that of *sampling* the generators by executing a battery of tests starting with 100 different seeds that are equispaced in the state space. For instance, for a 64-bit state space we use the seeds $1 + i \lfloor 2^{64}/100 \rfloor$, $0 \leq i < 100$. The tests produce a number of statistics, and we decided to use the number of failed tests as a measure of low quality. Running multiple tests makes it easy to rule out spurious failures, as suggested also by Rukhin et al. [2001] in the context of cryptographic applications.

We use two tools to perform our tests. The first and most important is TestU01, a test suite developed by L'Ecuyer and Simard [2007] that contains several tests oriented towards the generation of uniform real numbers in $[0..1)$. We also perform tests using Dieharder, a suite of tests developed by Brown [2013], both as a sanity check and to compare the power of the two suites. Dieharder contains all original tests from Marsaglia's Diehard, plus many more other tests. The suite is more oriented towards the effective values assumed by each bit (e.g., it computes more statistics on subsequences). We refer frequently to the specific type of tests failed: the reader can refer to the TestU01 and Dieharder documentation for more information.

We consider a test failed if its p -value is outside of the interval $[0.001..0.999]$. This is the interval outside which TestU01 reports a failure by default. Sometimes a much stricter

¹A more detailed study of the linear algebra behind xorshift generators can be found in [Marsaglia 2003; Panneton and L'Ecuyer 2005].

threshold is used (For instance, L’Ecuyer and Simard [2007] use $[10^{-10} \dots 1 - 10^{-10}]$ when applying TestU01 to a variety of generators), but since we are going to repeat the test 100 times we can use relatively weak p -values: spurious failures will appear rarely, and we can catch borderline cases (e.g., tests failing on 50% of the seeds) that give us useful information. We call *systematic* a failure that happens for all seeds.

We remark that our choice (counting the number of failures) is somewhat rough; for example, we consider the same failure a p -value very close to 0 and a p -value just below 0.001. Indeed, other, more sophisticated methods might be used to aggregate the result of our samples: combining p -values, for instance, or computing a p -value of p -values [Rukhin et al. 2001]. However, our choice is very easy to interpret, and multiple samples partially compensate this problem (spurious failures will appear in few samples).

Of course, the number of experiments is very large—in fact, our experiments were carried using hundreds of cores in parallel and, overall, they add up to more than a century of computational time. Our strategy is to apply a very fast test to all generators and seeds, in the hope of isolating a small group of generators that behave significantly better. Stronger tests can then be applied to this subset. The same strategy has been followed by Panneton [2004] in the experimental study of `xorshift` generators contained in his Ph.D. thesis.

TestU01 offers three different predefined batteries of tests (SmallCrush, Crush and BigCrush) with increasing computational cost and increased difficulty. Unfortunately, Dieharder does not provide such a segmentation.

Note that Dieharder has a concept of “weak” success and a concept of “failure”, depending on the p -value of the test, and we used command-line options to align its behavior with that of TestU01: a p -value outside of the range $[0.001 \dots 0.999]$ is a failure. Moreover, we disabled the initial timing tests so that exactly the same stream of 64-bit numbers is fed to the two test suites.

In both cases we implemented our own `xorshift` generator. Some care is needed in this phase, as both TestU01 and Dieharder are inherently 32-bit test suites: since we want to test `xorshift` as a *64-bit* generator, it is important that all bits produced are actually fed into the test. For this reason, we implemented the generation of a uniform real value in $[0 \dots 1)$ by dividing the output of the generator by 2^{64} , but we implemented the generation of uniform 32-bit integer values by *returning first the lower and then the upper 32 bits of each 64-bit generated value*.²

An important consequence of this choice is that some of the bits are actually not used at all. When analyzing pseudorandom real numbers in the unit interval, there is an unavoidable bias towards high bits, as they are more significant. The very lowest bits have lesser importance and will be in any case perturbed by numerical errors. However, in our case the lowest eleven bits returned by the generator *are not used at all* due to the fact that the mantissa of a 64-bit floating-point number is formed by 53 bits only. For this reason, we will consistently run our tests both on a generator and on its reverse.³

We remark that in this paper we do not pursue the search for *equidistribution*—the property that all tuples of consecutive values, seen as vectors in the unit cube, are evenly distributed, as done, for instance, by Panneton and L’Ecuyer [2005]. Brent [2010] has already argued in detail that for long-period generators equidistribution is not particularly desirable, as it is a property of the whole sequence produced by the generator, and in the case of a long-period generator only a minuscule fraction of the sequence can be actually used. Moreover, equidistribution is currently impossible to evaluate exactly for long-period non-linear generators, and it is known to be biased towards the high bits [L’Ecuyer and Panneton 2005]: for instance, the WELL1024a generator has been designed to be *maximally equidistributed* [Panneton et al. 2006], and indeed it has measure of equidistribution $\Delta_1 = 0$,

²If a real value is generated when the upper 32 bits of the last value are available, they are simply discarded.

³That is, on the generator obtained by reversing the order of the 64 bits returned.

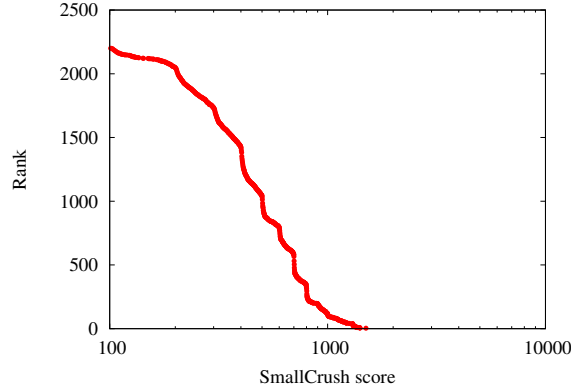


Fig. 2. Score-rank plot of the distribution of SmallCrush scores for the 2200 possible full-period xorshift64 generators.

but the generator obtained by reversing its bits has $\Delta_1 = 366$: a quite counterintuitive result, as in general we expect all bits to be equally important.

Another major problem with equidistribution is its “on/off” nature: it provides no measure of *how much* a generator is equidistributed. This leads to the following pathological behavior: if we take a maximally equidistributed sequence, no matter how long, and we flip the most significant bit of a single element of the sequence, the new sequence will have the *worst possible* Δ_1 . For instance, by flipping the most significant bit of a single chosen value out of the output of WELL1024a we can turn its equidistribution measure to $\Delta_1 = 4143$. But for any statistical or practical purpose the two sequences are indistinguishable—we are modifying one bit out of $2^5(2^{1024} - 1)$.

We note, however, that since multiplication by an invertible constant induces a permutation of the space of 64-bit values (and thus of t -tuples of such values), the choice of multiplication has the advantage, with respect to other scrambling techniques, of preserving some of the equidistribution properties of the underlying generator; more details will be given in the rest of the paper.

4. RESULTS FOR xorshift64 GENERATORS

In this section we report the results obtained for the 2200 possible variants of the xorshift64 generator with full period.

First of all, *all* generators fail at all seeds the MatrixRank test from TestU01’s SmallCrush suite. This is somewhat to be expected, as each new value is obtained by applying a linear transformation to the previous one. However, Panneton and L’Ecuyer [2005] report that *half* of the generators fail this test. Unfortunately, the authors do not detail the conditions (seed, implementation of the algorithm, etc.) of the experiments they performed, but TestU01 is available and contains an implementation of the xorshift64 family. A simple analysis of the code shows that the authors have chosen to use only 32 of the 64 generated bits as output bits, in practice applying a kind of *decimation* to the output of the generator. As we explained in Section 3, we actually feed *all* bits output by the generator to the test suite, which explains why we report a significantly worse performance.

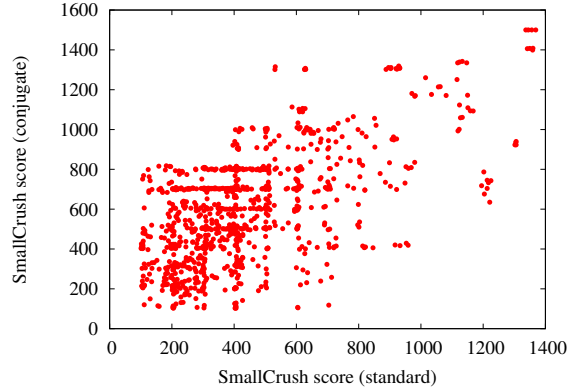


Fig. 3. Scatter plot of the SmallCrush score of conjugate generators.

A score-rank plot of the SmallCrush scores for all generators is shown in Figure 2. The plot associates with abscissa x the number of generators with x or more failures.⁴ We observe immediately that *there is a wide range of quality among the generators examined*. Indeed, inspecting more closely we would see that the best generator ($A_4(14, 13, 17)$, score 101) fails MatrixRank systematically and just Collision on one sample. The worst generators (e.g., $A_5(1, 1, 55)$, score 1500), instead, fails systematically *all* SmallCrush tests. The “bumps” in the plot corresponds to new tests failed systematically.

Figure 3 shows instead a scatter plot associating conjugate algorithms, which have just their bits reversed. While there is some weak correlation, we can see the bias towards high bits at work. Some generators fail systematically just a few tests while their reverse fail more than a dozen.

As we already remarked, to avoid high-bits bias we score each conjugate pair jointly, adding up the number of failures in SmallCrush: Table I reports the best four generators, which are the only ones failing systematically just the MatrixRank test. Any other choice fails more than half of the times the BirthdaySpacings test (data not shown here), and all generators with a rank lower than those shown in Table I fail systematically at least one test besides MatrixRank.

The table reports also results for the generator $A_0(13, 7, 17)$ suggested by Marsaglia in his original paper, claiming that it “will provide an excellent period $2^{64} - 1$ RNG, [...] but any of the above 2200 choices is likely to do as well”. Clearly, this is not the case: $A_0(13, 7, 17)/A_1(13, 7, 17)$ ranks 655 in the combined SmallCrush ranking and fails systematically several tests. Unfortunately, since this choice of parameters appeared in the original paper, other researchers have used it as well: this is what happens, for example, in the comparison table assembled by L’Ecuyer and Simard [2007] using TestU01 .

We remark that the triples suggested in “Numerical Recipes” are just of average quality when measured using TestU01. The authors suggest as best algorithm $A_3(4, 35, 21)$ (with its conjugate $A_2(4, 35, 21)$), as they notice that it is important to check conjugates), which however with score 389 ranks only 29 in our combined classification. The best result for the triples suggested at page 347 are those for $A_2(11, 29, 14)/A_3(11, 29, 14)$ (score 314, rank 9).

⁴Score-rank plots are the numerosity-based discrete analogous of the complementary cumulative distribution function of scores. They give a much clearer picture than frequency dot plots when the data points are scattered and highly variable.

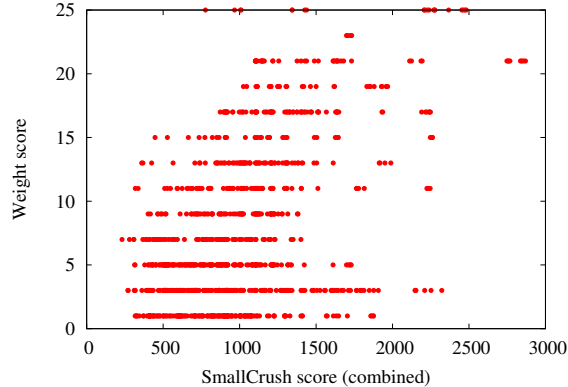


Fig. 4. Scatter plot of the combined SmallCrush score of conjugate `xorshift64` generators versus their weight score.

Finally, Figure 4 shows a scatter plot associating the combined (standard plus reverse) score of a generator with its weight score. While there is a very mild correlation (the very lower left and upper right corners are empty), it is definitely not the case that the SmallCrush score of a `xorshift64` generator is strictly dependent on its weight score.

SANITY CHECK 1. *Is the result of our experiments dependent on our seed choice? To answer this question, we repeated our experiments on `xorshift64` generators with SmallCrush on a different set of seeds, namely the integers in the interval $[1..100]$. The scatter plot in Figure 5 and Kendall’s⁵ $\tau \approx 0.98$ between the two rankings we obtain make rather clear that this is not the case. In particular, the four best conjugate pairs in Table I are the same with both seeds.*

To gather more information, we ran the full BigCrush suite and Dieharder on our four best generators, on Marsaglia’s choice and on the best choice from “Numerical Recipes”: the results are given in Table II and III. Even the four best generators fail now systematically the BirthdaySpacings, MatrixRank and LinearComp tests. The first two generators, however, turn out to perform better than other two. We also notice that BigCrush draws a much thicker line between our four best generators and the other ones, which now fail several more tests. Not surprisingly, Dieharder cannot really separate our four best generators from $A_2(4, 35, 21)/A_3(4, 35, 21)$.

4.1. Equidistribution

As we already discussed in the introduction, in this paper we do not pursue equidistribution of a generator. It is nonetheless interesting to compare the ranking provided by equidistribution properties and that provided by statistical tests. Note that a `xorshift64` generator is at least 1-dimensionally equidistributed, that is, every 64-bit value appears exactly once except for zero. We refer to the already quoted paper by Panneton and L’Ecuyer [2005] for a detailed description of the equidistribution statistics Δ_1 , the *sum of dimension gaps*: a lower value is better. A *maximally distributed* generator has $\Delta_1 = 0$, and we will refer to Δ_1

⁵We are using the generalization allowing ties and defined in [Kendall 1945], which is often called τ_b , reserving τ for the original coefficient [Kendall 1938]. But this distinction is pointless, as in [Kendall 1938] τ is defined only for rankings with no ties, and the definition given in [Kendall 1945] reduces exactly to the original definition if there are no ties.

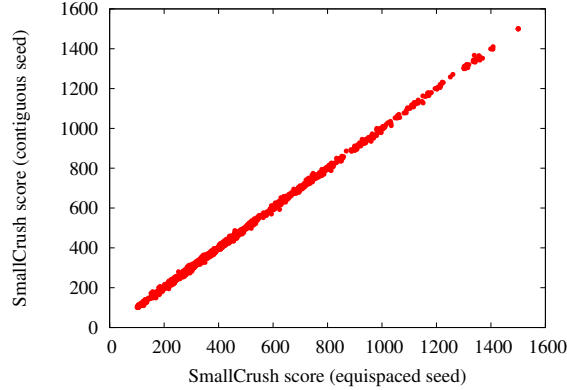


Fig. 5. Scatter plot of the scores obtained on each `xorshift64` generator using the seed $1 + i \lfloor 2^{64}/100 \rfloor$ or the seed $1 + i$, $0 \leq i < 100$.

Table I. Best four `xorshift64` generators following SmallCrush.

Algorithm	Failures	Conjugate	Failures	Overall	W
$A_2(11, 31, 18)$	111	$A_3(11, 31, 18)$	120	231	25
$A_2(8, 29, 19)$	155	$A_3(8, 29, 19)$	115	270	35
$A_0(8, 29, 19)$	159	$A_1(8, 29, 19)$	112	271	35
$A_0(11, 31, 18)$	130	$A_1(11, 31, 18)$	150	280	25
$A_2(4, 35, 21)$	209	$A_3(4, 35, 21)$	180	389	25
$A_0(13, 7, 17)$	276	$A_1(13, 7, 17)$	802	1078	25

Note: The only systematic failure is on the MatrixRank test. All other generators have an overall number of failures greater than 300, and fail systematically at least one test besides MatrixRank. $A_0(13, 7, 17)$ is the generator suggested in Marsaglia’s original paper, and ranks 655 on 1100 conjugate pairs. $A_3(4, 35, 21)$ is suggested as the best generator in this class in “Numerical Recipes” [Press et al. 2007], and ranks 29.

Table II. The generators of Table I tested with BigCrush.

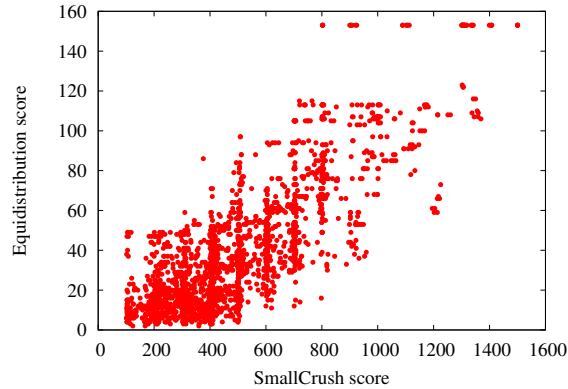
Algorithm	Failures	Conjugate	Failures	Overall	Systematic
$A_2(11, 31, 18)$	762	$A_3(11, 31, 18)$	750	1512	A
$A_2(8, 29, 19)$	747	$A_3(8, 29, 19)$	780	1527	
$A_0(8, 29, 19)$	749	$A_1(8, 29, 19)$	884	1633	
$A_0(11, 31, 18)$	748	$A_1(11, 31, 18)$	926	1674	
$A_2(4, 35, 21)$	961	$A_3(4, 35, 21)$	1444	2405	$A \cup B$
$A_0(13, 7, 17)$	1049	$A_1(13, 7, 17)$	5454	6503	$A \cup C$

Note: $A = \{\text{BirthdaySpacings, LinearComp, MatrixRank}\}$. $B = \{\text{RandomWalk1C, ClosePairsmNP, ClosePairsmNP1, ClosePairsmNP2S}\}$. $C = \{\text{RandomWalk1H, RandomWalk1J, RandomWalk1M, ClosePairsNJumps, ClosePairsNP, ClosePairsmNP2, ClosePairsmNP2S, CollisionOver, MaxOft, MaxOftAD, Permutation, Run, SampleMean, SampleProd, SerialOver, SumCollector}\}$. This table should be compared with Table I and with the experimental results by L’Ecuyer and Simard [2007].

Table III. The generators of Table I tested with Dieharder.

Algorithm	Failures	Conjugate	Failures	Overall	Systematic
$A_2(11, 31, 18)$	182	$A_3(11, 31, 18)$	162	344	dab_monobit2
$A_2(8, 29, 19)$	179	$A_3(8, 29, 19)$	181	360	
$A_0(8, 29, 19)$	176	$A_1(8, 29, 19)$	182	358	
$A_0(11, 31, 18)$	181	$A_1(11, 31, 18)$	186	367	
$A_2(4, 35, 21)$	189	$A_3(4, 35, 21)$	187	376	dab_monobit2
$A_0(13, 7, 17)$	183	$A_1(13, 7, 17)$	1352	1535	A

Note: $A = \{\text{dab_filltree, dab_monobit2, diehard_2dsphere, diehard_3dsphere, diehard_operm5, diehard_parking_lot, diehard_squeeze, rgb_minimum_distance, rgb_permutations}\}$. This table should be compared with Table II.

Fig. 6. Scatter plot of the SmallCrush score versus the equidistribution score of `xorshift64` generators.

as to the *equidistribution score*. We computed the equidistribution score for all generators using the implementation of Harase’s algorithm [Harase 2011] contained in the `MTToolBox` package from Saito [2013].

SANITY CHECK 2. *It is very easy to introduce hard-to-detect bugs in this kind of computations. However, the Ph.D. thesis of Panneton [2004] reports Δ_1 (therein called V) for all full-period `xorshift64` generators, and we checked that the results are the same. We will use `MTToolBox` for other computations, which explains the usefulness of recomputing and checking these values.*

Figure 6 shows that there is some correlation between the SmallCrush score and the equidistribution score of `xorshift64` generators. Nonetheless, in the “interesting” region (the lowest left corner) correlation is not very good—for instance, the generator $A_6(10, 7, 33)$, which has equidistribution score 2 (the best), fails systematically three types of tests, and ranks 757 in combination with its conjugate: it’s actually one of the *worst* generators. Its combined BigCrush score is 6691—even worse than the generator $A_0(13, 7, 17)$ suggested by Marsaglia.

The explanation is simple: similarly to SmallCrush score, equidistribution has high-bits bias, and a quite strong one [L’Ecuyer and Panneton 2005]. Indeed, Figure 7 reports a scatter plot of the equidistribution score of conjugate (i.e., reverse) generators, and the bias towards the high bits is very visible from the lack of correlation. There are apparently good generators whose reverse is actually worst: for instance, $A_2(12, 1, 31)$ has score 11 but its conjugate $A_3(12, 1, 31)$ has score 153—the highest value in the set, whereas $A_0(11, 5, 32)$

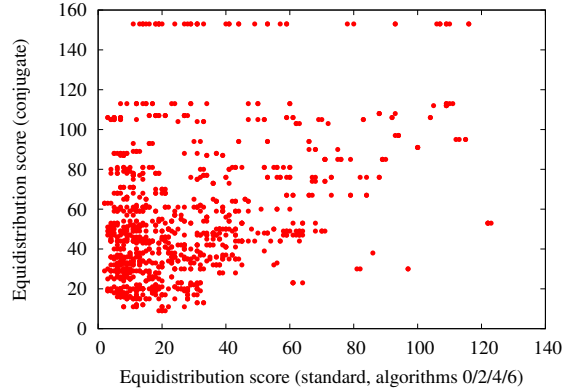


Fig. 7. Scatter plot of the equidistribution score of conjugate `xorshift64` generators.

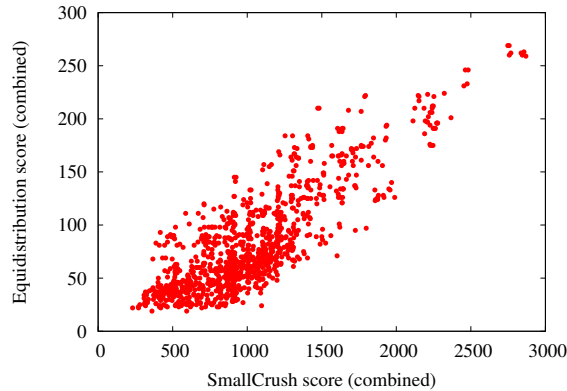


Fig. 8. Scatter plot of the combined SmallCrush score of conjugate `xorshift64` generators versus the combined equidistribution score. Notice the improvement in correlation with respect to Figure 6.

has score 3 (the best is 2) whereas its conjugate $A_1(11, 5, 32)$ has score 106. It is clearly necessary to combine the equidistribution score of a generator and of its reverse.

This approach improves somewhat the quality of the result: Figure 8 shows that there is a better correlation between combined SmallCrush scores and combined equidistribution scores. Nonetheless, even if equidistribution is able to detect bad generators, is not so good at detecting the *very best* generators. We already noticed that only four generators (Table I) fail systematically a single SmallCrush test. These generators, however, are not the best ones by equidistribution score: in display order, they rank 4, 11, 3 and 23 (the other two generators rank 37 and 570, respectively). The first two generators by combined equidistribution score, $A_4(8, 29, 19)$ and $A_6(8, 29, 19)$, rank 20 (combined score 361) and 170 (score 596) in the combined SmallCrush test. When analyzed with the more powerful lens of BigCrush, they have combined scores 3441 and 4082, respectively, and fail systematically almost *twenty* additional tests with respect to the set A of Table II. Definitely, choosing among `xorshift64` generators by equidistribution score alone is not a good idea.

As a final interesting observation, in Figure 9 we correlate the equidistribution score and the weight score of `xorshift64` generators, both in combined and non-combined form. It is somewhat fascinating that two mathematically defined features which are supposed to lead

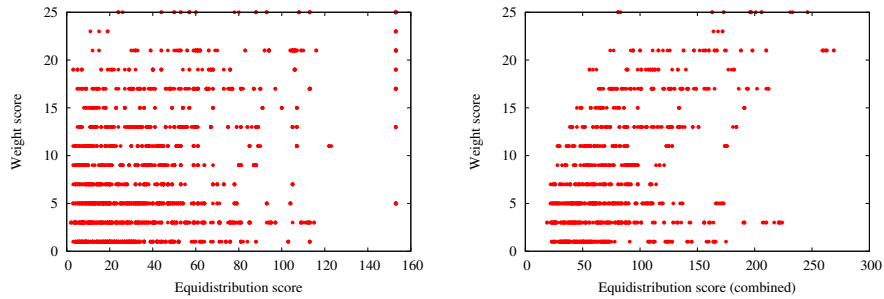


Fig. 9. Scatter plots of equidistribution scores versus weight scores for `xorshift64` generators.

Table IV. The three multipliers used in the rest of the paper. The subscripts recalls the t for which they have good figures of merit.

M_{32}	2685821657736338717
M_8	1181783497276652981
M_2	8372773778140471301

to good generators are entirely uncorrelated. Some mild correlation, as usual, appears if we use a combined equidistribution score.

5. RESULTS FOR `xorshift64*` GENERATORS

Since a `xorshift64` generator exhibits evident linearity artifacts, the next obvious step is to perturb its output using a nonlinear (in $\mathbf{Z}/2\mathbf{Z}$ sense) transformation. A natural candidate is multiplication by a constant, also because such operation is very fast in modern processors. Note that the current state of the generator is multiplied by a constant before returning it, but the state itself is not affected by the multiplication: thus, the period is the same.

We call such a generator `xorshift*`. By choosing a constant invertible modulo 2^{64} , we can guarantee that the generator will output a permutation of the sequence output by the underlying `xorshift` generator.

This approach was noted in passing in Marsaglia’s paper, and it is also proposed in a more systematic way in the third edition of “Numerical Recipes” [Press et al. 2007] to create a very fast, good-quality pseudorandom number generator. However, in the latter case the author *first* computes allegedly good triples for `xorshift` using DieHard (with results markedly different from ours, and in strident contrast with TestU01’s results, as discussed in Section 4) and *then* chooses a multiplier using a perfectly reasonable criterion (good spectral quality as a linear congruential generator). There is no reason why the best triples for a `xorshift64` generator (which are computed empirically) should continue to be such in a `xorshift64*` generator: and indeed, we will see that this is not the case.

We thus repeated the experiments of the previous section on `xorshift64*` generators. To further understand the dependency on the multiplier, we used three different multipliers, shown in Table IV. The first multiplier, M_{32} , is the one used in “Numerical Recipes”, which is suggested by L’Ecuyer [1999] as having good spectral properties. The goodness of the multiplier, however, is established by a *figure of merit* which is a normalized best distance between the hyperplanes of families covering tuples of length t given by successive outputs of the generators. The length t is an additional parameter, and M_{32} has the best figures of merit for $t = 32$. Clearly, if an alternative multiplier provides improvements on both t and the associated figure of merit, we have a hint that it could be chosen instead of M_{32} .

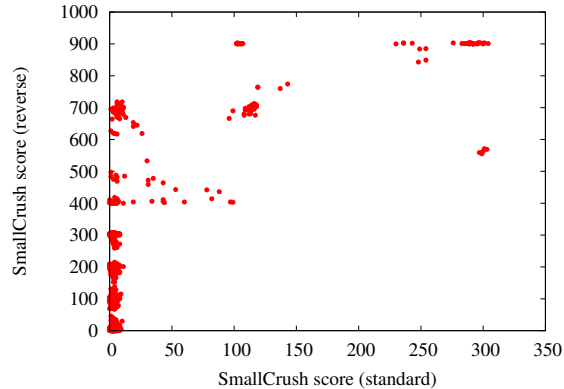


Fig. 10. Scatter plot of the SmallCrush score of `xorshift64*` generators and their reverse.

Lacking that possibility, what if we scramble `xorshift64`'s output with a multiplier that has *better* figures of merit for a *lower* t ? We thus also ran experiment with the multiplier $M_8 = 1181783497276652981$, which has a better figure of merit for $t = 8$ [L'Ecuyer 1999], and the multiplier $M_2 = 8372773778140471301$, which has a better figure of merit for $t = 2$ and was kindly provided by Richard Simard.

The landscape is now very different. Indeed, the scatter plot in Figure 10 shows that there is no correlation between the scores assigned by SmallCrush to a generator and its reverse.⁶

Another interesting observation on Figure 10 is that the lower right half is essentially empty. So bad generators have a bad reverse, but there are good generators with a very bad reverse. This suggests that the quality of a `xorshift64*` generator can vary wildly from the low to the high bits.

A score-rank plot of the SmallCrush scores for all generators shown in Figure 11 provides us with further interesting information: almost all generators have no systematic failure, but only about half of the reverse generators have no systematic failure. Moreover, the distribution of standard generators degrades smoothly, whereas the distribution of reverse generators sports again the “bump” phenomenon we observed in Figure 2.

Finally, Figure 12 is the analogous of Figure 4 for `xorshift64*` generators: the mild correlation between combined SmallCrush score and weight score of the underlying `xorshift64` is now completely absent.

Since we need to reduce the number of candidates to apply stronger tests, in the case of M_{32} we decided to restrict our choice to generators with 3 overall failed tests or less, which left us with 152 generators. Similar cutoff were chosen for M_8 and M_2 .

These generators were few enough so that we could apply both Crush and Dieharder. Once again, we examine the correlation between the score of a generator and its reverse by means of the scatter plots in Figure 13, which confirm the high-bits bias, albeit less so in the Dieharder case.

In Figure 14 we compare instead the two scores (Crush and Dieharder) available. The most remarkable feature is there are no points in the upper left corner: there is no generator that is considered good by Crush but not by Dieharder. On the contrary, Crush heavily penalizes (in particular because of the score on the reverse generator) a large number of generators.

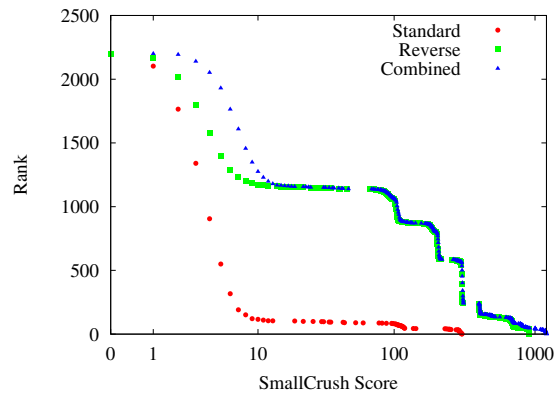


Fig. 11. Score-rank plot of the distribution of SmallCrush scores for the 2200 possible xorshift64* generators with multiplier M_{32} .

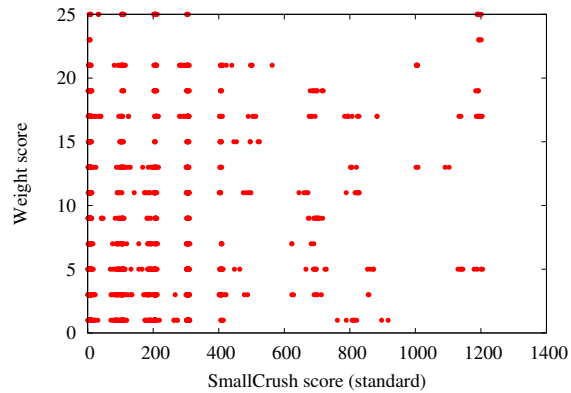


Fig. 12. Scatter plot of combined SmallCrush score or xorshift64* generators with multiplier M_{32} versus weights score of the underlying xorshift64 generators.

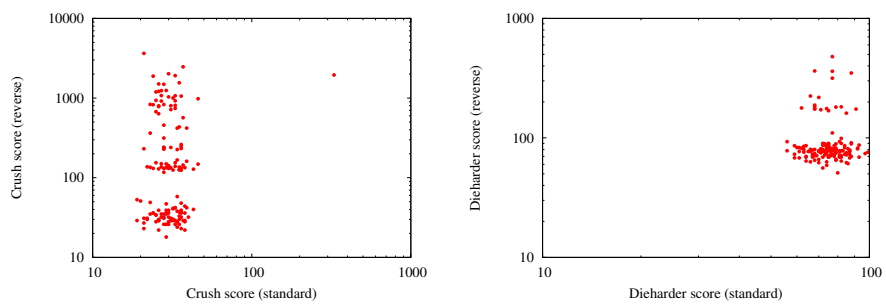


Fig. 13. Scatter plots for Crush (left) and Dieharder (right) scores on xorshift64* generators with multiplier M_{32} and their reverse, for the 152 best generators.

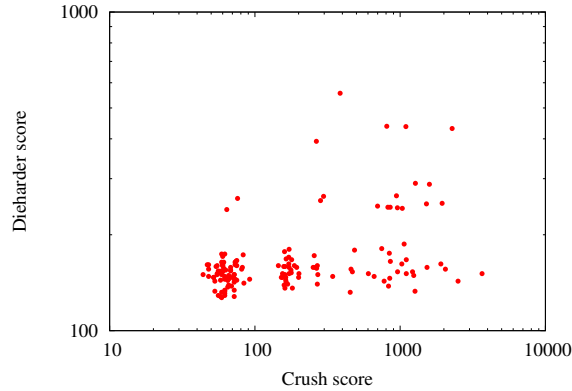


Fig. 14. A scatter plot of Crush and Diehard combined scores of the 152 SmallCrush-best `xorshift64*` generators. The plot is in log-log scale to accommodate some very high values returned by Crush on reverse generators. The “sweet spot” in the lower left corner contains generators that never fail systematically (not even reversed) in both test suites.

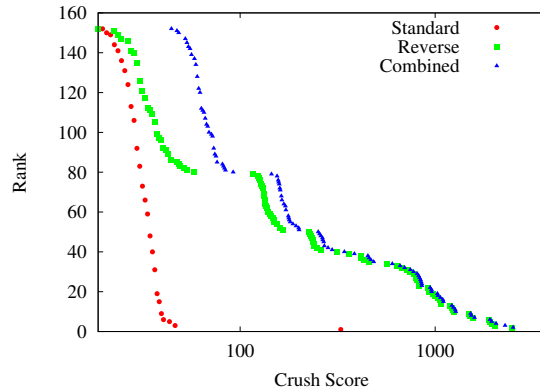


Fig. 15. Score-rank plot of the distribution of Crush scores for the 152 SmallCrush-best `xorshift64*` generators using multiplier M_{32} .

The generators we will select in the end all belong to the small cloud in the lower left corner, where the two test suite agree.

The score-rank plot in Figure 15 shows that our strategy pays off: we started with 152 generators with less than three failures, but analyzing them with the more powerful lens provided by Crush we get a much more fine-grained analysis: in particular, only 73 of them give no systematic failure, and they all belong to the “sweet spot” of Figure 14, that is, they do not give any systematic failure in Dieharder, too.

Finally, we selected for each multiplier the eight best generators with the best Crush score, and applied the BigCrush suite: we obtained several generators failing systematically the MatrixRank test only and shown in Table V (which should be compared with Table II). It is interesting to note that all generators have one additional systematic failure in the standard version with respect to the reversed version. This behavior is not surprising, as multiplication tends to make the lower bits more chaotic and of better quality than the

⁶In general, we report plots only for M_{32} , as the ones for the other multipliers are visually identical.

Table V. Results of BigCrush on the best eight xorshift64* generators found by SmallCrush and Crush in sequence. The generators fail systematically only MatrixRank.

Algorithm	Failures			W
	S	R	+	
<i>M</i> ₃₂				
<i>A</i> ₇ (11, 5, 45)	226	128	354	23
<i>A</i> ₇ (17, 23, 52)	232	130	362	25
<i>A</i> ₁ (12, 25, 27)	230	133	363	31
<i>A</i> ₁ (17, 23, 29)	229	137	366	21
<i>A</i> ₅ (14, 23, 33)	238	132	370	32
<i>A</i> ₅ (17, 47, 29)	231	141	372	24
<i>A</i> ₁ (16, 25, 43)	238	138	376	31
<i>A</i> ₇ (23, 9, 57)	242	134	376	19
<i>M</i> ₈				
<i>A</i> ₅ (11, 5, 32)	229	122	351	13
<i>A</i> ₂ (8, 31, 17)	229	126	355	21
<i>A</i> ₅ (3, 21, 31)	230	141	371	33
<i>A</i> ₃ (17, 45, 22)	241	133	374	27
<i>A</i> ₄ (8, 37, 21)	239	136	375	33
<i>A</i> ₃ (13, 47, 23)	232	144	376	27
<i>A</i> ₃ (13, 35, 30)	244	136	380	27
<i>A</i> ₄ (9, 37, 31)	243	141	384	27
<i>M</i> ₂				
<i>A</i> ₇ (13, 19, 28)	228	128	356	23
<i>A</i> ₃ (9, 21, 40)	228	132	360	35
<i>A</i> ₁ (14, 23, 33)	234	142	376	29
<i>A</i> ₇ (19, 43, 27)	239	137	376	23
<i>A</i> ₁ (17, 47, 28)	240	137	377	25
<i>A</i> ₅ (16, 11, 27)	234	144	378	25
<i>A</i> ₄ (4, 35, 15)	230	149	379	35
<i>A</i> ₇ (13, 21, 18)	238	144	382	31

upper bits. It also suggests that for these generators it is better to extract the lower bits, rather than the high bits, when just a subsequence is needed.

5.1. Equidistribution

Multiplication by an invertible element just permutes the elements of $\mathbf{Z}/2^{64}\mathbf{Z}$ leaving zero fixed, so a xorshift64* generator, like the underlying xorshift64 generator, is at least 1-dimensionally equidistributed. Since xorshift64* generators are not linear, analyzing more deeply equidistribution scores would require an enormous computing effort, which is not justified in consideration of our observations on xorshift64 generators. We tried nonetheless to correlate in Figure 16 the SmallCrush score of xorshift64* generators with the equidistribution score of the the underlying xorshift64 generators. No such correlation appears in the generators taken in isolation—there are even generators, such as $A_7(15, 1, 19) \cdot M_{32}$ with the best SmallCrush combined score (no failure) for which the equidistribution score of the underlying xorshift64 generator is the worst possible (153). However, once we combine both the SmallCrush scores and the equidistribution score some correlation appears. Once

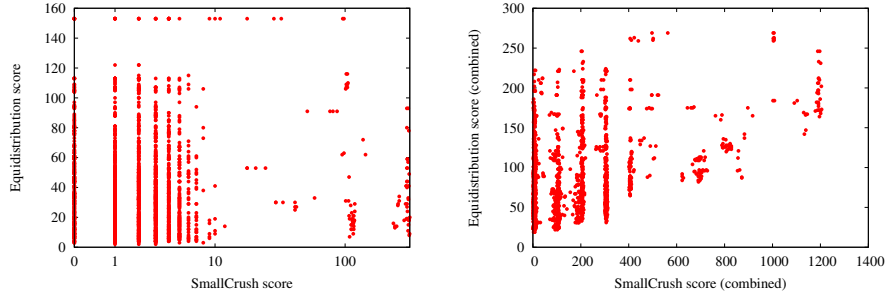


Fig. 16. Scatter plots of SmallCrush scores versus equidistribution scores of the underlying `xorshift64` generator for `xorshift64*` generators using multiplier M_{32} .

again, the combined equidistribution score is more useful to detect bad generators than to find the best ones, as the left part of the plot is quite chaotic.

6. HIGH DIMENSION

Marsaglia [2003] describes a strategy for `xorshift` generators in high dimension: the idea is to use always three low-dimensional shifts, but locating them in the context of a larger matrix of the form

$$M = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & (1 + L^a)(1 + R^b) \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & (1 + R^c) \end{pmatrix}$$

Marsaglia notes that even in this restricted forms there are matrices of full period (he provides examples for 32-bit shifts up to 160 bits). However, we could find no evidence in the literature that this route has been explored for high-dimensional (say, more than 1024 bits of state) generators. The only similar approach is that proposed by Brent [2007] with his `xorgens` generators, which however uses more shifts. The obvious question is thus: are these additional shifts really necessary? We are thus going to look for good, full-period generators with 1024 or 4096 bits of state using 64-bit basic shifts.⁷

The output of such generators will be given by the last 64 bits of the state space. It is well known [Niederreiter 1992] that every bit of the state space satisfies a linear recurrence (defined by the characteristic polynomial) with full period, so *a fortiori* the last 64 bits have full period, too.

Since we already know that some deficiencies of low-dimensional `xorshift` generators are well corrected by multiplication by a constant, we will follow the same approach, thus looking for good `xorshift*` generators of high dimension.⁸ Note that since multiplication by an integer invertible in $\mathbf{Z}/2^{64}\mathbf{Z}$ is a permutation of $\mathbf{Z}/2^{64}\mathbf{Z}$, a high-dimension `xorshift*` generator has the same period of the underlying `xorshift` generator.

⁷The reason why the number 4096 is relevant here is that we know the factorization of Fermat's numbers $2^{2^k} + 1$ only up to $k = 11$. When more Fermat numbers will be factorized, it will be possible to design `xorshift` or `xorgens` generators with larger state space [Brent 2007]. Note that, however, in practice a period of $2^{1024} - 1$ is more than sufficient for any purpose. For example, even if 2^{100} computers were to generate sequences of 2^{100} numbers starting from random seeds using a generator with period 2^{1024} , the chances that two sequences overlap would be less than 2^{-724} .

⁸As in the `xorshift64` case, different choices for the shifts are possible. We will not pursue them here.

We cannot in principle claim full period if we look at a *single* bit of the output of a `xorshift*` generator; but this property can be easily proved by exquisitely combinatorial means:

PROPOSITION 6.1. *Let $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{2^n-2}$ be a list of t -bit values, $t \leq n$, such that every value appears 2^{n-t} times, except for 0, which appears $2^{n-t} - 1$ times. Then, for every fixed bit k the associated sequence has period $2^n - 1$.*

PROOF. Suppose that there is a k and a $p \mid 2^n - 1$ such that the k -th bit of $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{2^n-2}$ has period p (that is, the sequence of bits associated with the k -th bit is made by $(2^n - 1)/p$ repetitions of the same sequence of p bits). The k -th bit runs through $2^{n-1} - 1$ zeroes and 2^{n-1} ones (as there is a missing zero). This means that $(2^n - 1)/p \mid 2^{n-1}$, too, as the same number of ones must appear in every repeating subsequence, and since $(2^n - 1)/p$ is odd this implies $p = 2^n - 1$. \square

COROLLARY 6.2. *Every bit of the output of a full-period `xorshift*` generator has full period.*

6.1. Finding good shifts

The first step is identifying values of a , b and c for which the generator has maximum period using the primitivity check on the characteristic polynomial. We performed these computations using the algebra package Fermat [Lewis 2013], with the restriction that $a + b \leq 64$ and that a is coprime with b (see [Brent 2007] for the rationale behind this choices, which significantly reduce the search space). The resulting sets of values are those shown in Table VI and VIII.

For a state space of 1024 bits, we obtain 20 possible parameter choices, which we examined in combination with our three multipliers both through BigCrush and through Dieharder. The results, reported in Table VI and VII, are excellent: with the exception of two pathological choices, no test is failed systematically. For 4096-bit state space (Table VIII and IX) there are 10 possible parameter choices, and no generator fails a test systematically.

SANITY CHECK 3. *Very long and complex computations are prone to implementation, software and hardware errors. In particular, if no verification procedure exists, results of a search on a large state space like the one we just describe are very difficult to assess. We thus decided to compute again the same coefficient using an entirely different algorithm: instead of working on characteristic polynomials, we developed highly optimized Java software that exploits the particular structure taken by powers of the linear transformation M associated with a generator to compute such powers explicitly, and store them using space linear in the size of the state space: as it is known [Marsaglia and Tsay 1985], the generator has full period if and only if M has the same multiplicative period. It is thus sufficient to show that $M^{2^n-1} = 1$ and*

$$M^{(2^n-1)/p} \neq 1$$

for every prime p dividing 2^n-1 . The computation, in particular for the case of 4096 bits, turned out to be extremely intensive, requiring almost a month of computing time on a 40-core workstation and using more than half a terabyte of in-core memory, as each set of parameters can be checked in parallel, but for each such set we must compute and store the quadratures M^{2^k} , $0 < k < n$, so to be able to evaluate the condition above for all p 's. The results confirmed those obtained by using primitive polynomials.

6.2. Equidistribution

Looking at the shape of the matrix defining high-dimensional `xorshift` generators it is clear that if the state space is made of n bits the last $n/64$ output values, concatenated, are equal to the current state. This implies that such generators are at least $n/64$ -dimensionally

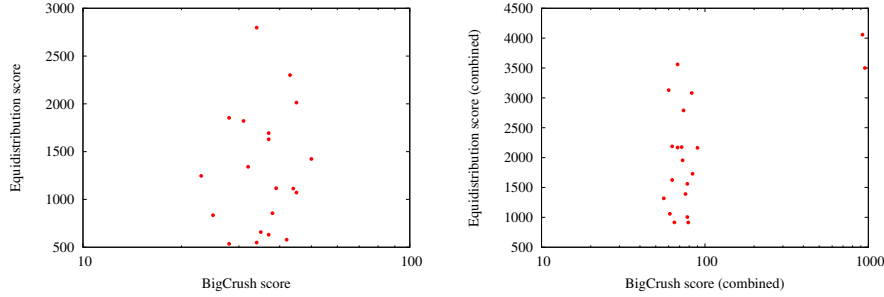


Fig. 17. Scatter plots of BigCrush scores versus equidistribution scores for `xorshift1024*` generators using multiplier M_{32} .

equidistributed (i.e., every $n/64$ -tuple of consecutive 64-bit values appears exactly once, except for a missing tuple of zeroes), so `xorshift1024` generators are at least 16-dimensionally equidistributed and `xorshift4096` generators are at least 64-dimensionally equidistributed. Since multiplication by a constant just permutes the space of tuples, the same is true of the associated `xorshift*` generators.

We now repeat the analysis of Section 5.1, always using the `MTTollBox` package, and show the results in Figure 17. No correlation between BigCrush scores of `xorshift1024*` generators and equidistribution scores of the underlying `xorshift1024` appears for the generators taken in isolation, but once we combine both the BigCrush scores and the equidistribution scores some correlation appears, as the two pathological generators have indeed a very high combined equidistribution score. The analogous graphs for `xorshift4096*` generators are omitted; they display no correlation at all.

7. COMPARISON

How do our best `xorshift*` generators score with respect to more complex generators in the literature? We decided to perform a comparison with the popular Mersenne Twister `MT19937` [Matsumoto and Nishimura 1998],⁹ with `WELL1024a`/`WELL19937a`, two generators introduced by Panneton et al. [2006] as an improvement over the Mersenne Twister, and with `xorgens4096`, a very recent 4096-bit generator introduced by Brent [2007] we mentioned in Section 6. All these generators are non-cryptographic and aim at fast, high-quality generation. As usual, 100 tests are performed at 100 equispaced points of the state space.

We choose generators from the `xorshift*` family that perform well on both BigCrush and Dieharder, have a good weight score and enough large parameters (which provide faster state change spreading): more precisely, the `xorshift64*` generator $A_1(12, 25, 27) \cdot M_{32}$ (Figure 19), `xorshift1024*` with parameters 31, 11, 30 and multiplier M_8 (Figure 20), and `xorshift4096*` with parameters 25, 3, 49 and multiplier M_2 (Figure 21).

7.1. Quality

Table X compares the BigCrush scores of the generators we discussed. We report also results on the Java standard random number generator, as a reality check with respect to stock generators currently found in computer languages.

The results are quite surprising. A simple 64-bit `xorshift*` generator has less linear artifacts than `MT19937`, `WELL1024a` or `WELL19937a` and, thus, a significantly better score.

⁹More precisely, with its 64-bit version.

Table VI. Results of BigCrush on the xorshift1024* generators. The last two generators fail systematically CouponCollector, Gap, HammingIndep, MatrixRank, SumCollector and WeightDistrib.

a, b, c	M_{32}				M_8				M_2					
	Failures			W	Failures			W	Failures			W		
	S	R	+		S	R	+		S	R	+			
27, 13, 46	25	31	56	275	1, 13, 7	28	19	47	113	3, 26, 35	29	24	53	89
31, 33, 37	28	32	60	79	3, 26, 35	29	22	51	89	27, 13, 46	41	20	61	275
22, 7, 48	37	24	61	223	40, 11, 31	24	33	57	77	25, 8, 15	38	24	62	281
7, 16, 55	37	26	63	65	15, 16, 19	30	32	62	255	31, 10, 27	36	31	67	233
9, 14, 41	23	40	63	167	22, 7, 48	29	33	62	223	9, 5, 60	24	43	67	227
41, 7, 29	28	37	65	265	9, 14, 41	32	30	62	167	1, 13, 7	28	42	70	113
1, 13, 7	34	34	68	113	41, 7, 29	25	38	63	265	15, 16, 19	36	34	70	255
10, 11, 61	32	36	68	155	31, 11, 30	33	32	65	363	2, 11, 61	40	30	70	81
9, 5, 60	44	28	72	227	2, 11, 61	25	41	66	81	41, 7, 29	36	34	70	265
16, 23, 30	37	36	73	59	10, 11, 61	42	25	67	155	9, 14, 41	33	37	70	167
3, 26, 35	45	29	74	89	7, 16, 55	32	35	67	65	22, 7, 48	37	35	72	223
25, 8, 15	42	34	76	281	16, 23, 30	35	34	69	59	31, 11, 30	45	27	72	363
31, 11, 30	35	43	78	363	25, 8, 15	25	45	70	281	7, 16, 55	36	39	75	65
40, 11, 31	38	40	78	77	27, 13, 46	39	32	71	275	31, 33, 37	37	39	76	79
31, 10, 27	34	45	79	233	31, 10, 27	40	32	72	233	10, 11, 61	41	37	78	155
2, 11, 61	43	40	83	81	9, 5, 60	40	36	76	227	16, 23, 30	44	37	81	59
15, 16, 19	45	39	84	255	31, 33, 37	39	39	78	79	40, 11, 31	38	48	86	77
10, 9, 63	39	51	90	69	10, 9, 63	31	49	80	69	10, 9, 63	48	48	96	69
51, 1, 46	31	890	921	111	51, 1, 46	60	896	956	111	51, 1, 46	31	799	830	111
47, 1, 41	50	902	952	99	47, 1, 41	67	907	974	99	47, 1, 41	47	799	846	99

Table VII. Results of Dieharder on xorshift1024* generators. No test is failed systematically.

a, b, c	M_{32}			W	M_8			W	M_2			W
	S	R	Failures +		S	R	Failures +		S	R	Failures +	
31, 33, 37	57	67	124	79	67	56	123	281	56	76	132	223
31, 11, 30	65	61	126	363	77	54	131	59	66	67	133	255
16, 23, 30	74	56	130	59	66	66	132	65	70	71	141	69
41, 7, 29	71	61	132	265	60	75	135	89	65	78	143	111
9, 14, 41	74	64	138	167	63	74	137	155	80	64	144	113
10, 9, 63	74	66	140	69	74	69	143	233	80	67	147	77
22, 7, 48	66	75	141	223	31, 33, 37	86	58	144	85	65	150	81
51, 1, 46	78	63	141	111	47, 1, 41	82	62	144	75	75	150	363
27, 13, 46	63	79	142	275	27, 13, 46	78	69	147	74	77	151	281
25, 8, 15	80	64	144	281	31, 11, 30	85	62	147	79	76	155	155
3, 26, 35	81	66	147	89	10, 9, 63	65	86	151	70	86	156	99
2, 11, 61	79	71	150	81	41, 7, 29	84	68	152	70	86	156	227
40, 11, 31	74	76	150	77	2, 11, 61	88	65	153	81	76	157	59
31, 10, 27	82	71	153	233	9, 14, 41	77	80	157	81	76	157	275
47, 1, 41	74	79	153	99	40, 11, 31	82	78	160	78	80	158	65
9, 5, 60	81	75	156	227	15, 16, 19	85	76	161	80	80	167	167
10, 11, 61	75	84	159	155	51, 1, 46	92	74	166	87	81	168	265
15, 16, 19	72	88	160	255	22, 7, 48	90	82	172	82	87	169	233
7, 16, 55	94	68	162	65	1, 13, 7	79	95	174	92	79	171	89
1, 13, 7	87	76	163	113	9, 5, 60	97	89	186	98	88	186	79

Table IX. Results of Dieharder on xorshift4096* generators.

M_{32}				M_8				M_2						
Algorithm	Failures		+	W	Algorithm	Failures		+	W	Algorithm	Failures		+	W
	S	R				S	R				S	R		
25, 3, 49	70	70	140	441	25, 3, 49	67	70	137	441	19, 34, 19	75	64	139	291
12, 11, 61	58	83	141	195	14, 41, 15	72	69	141	241	5, 22, 27	67	77	144	45
30, 29, 39	67	77	144	177	30, 29, 39	70	75	145	177	25, 3, 49	77	71	148	441
5, 22, 27	62	84	146	45	11, 9, 25	73	77	150	567	5, 27, 21	77	71	148	187
11, 9, 25	73	75	148	567	12, 11, 61	75	80	155	195	11, 9, 25	81	76	157	567
19, 34, 19	85	66	151	291	19, 34, 19	89	67	156	291	14, 41, 15	79	78	157	241
14, 41, 15	83	74	157	241	5, 22, 27	93	65	158	45	23, 26, 29	74	84	158	49
7, 12, 59	73	85	158	103	23, 26, 29	72	87	159	49	12, 11, 61	74	85	159	195
23, 26, 29	73	88	161	49	5, 27, 21	75	84	159	187	7, 12, 59	84	79	163	103
5, 27, 21	98	67	165	187	7, 12, 59	90	77	167	103	30, 29, 39	78	89	167	177

Table X. A comparison of generators using BigCrush.

Algorithm	Failures			W	Systematic
	S	R	+		
$A_1(12, 25, 27) \cdot M_{32}$	230	133	363	31	MatrixRank
$A_3(4, 35, 21) \cdot M_{32}$	240	223	463	25	MatrixRank, BirthdaySpacings
xorshift1024*	29	22	51	363	—
xorshift4096*	33	34	67	441	—
xorgens4096	42	40	82	961	—
MT19937	258	258	516	6750	LinearComp
WELL1024a	441	441	882	407	MatrixRank, LinearComp
WELL19937a	235	233	468	8585	LinearComp
java.util.Random	4078	9486	13564	—	Almost all

Table XI. A comparison of generators using Dieharder.

Algorithm	Failures		
	S	R	+
$A_1(12, 25, 27) \cdot M_{32}$	86	61	147
$A_3(4, 35, 21) \cdot M_{32}$	76	71	147
xorshift1024*	70	75	135
xorshift4096*	77	71	148
xorgens4096	80	78	158
MT19937	72	79	151
WELL1024a	81	61	142
WELL19937a	86	67	153
java.util.Random	4078	9486	13564

Note: No test is failed systematically, except for `java.util.Random`, which fails systematically `dab_bytedistrib`, `dab_dct`, `diehard_craps`, `diehard_dna`, `diehard_operm5`, `diehard_opso`, `diehard_oqso`, `diehard_squeeze`, `rgb_kstest_test`, `rgb_lagged_sum`, `rgb_minimum_distance` and `rgb_permutations`.

High-dimension `xorgens4096` and `xorshift*` generators perform significantly better, in spite of being extremely simpler, and have no systematic failure. The 64-bit `xorshift*` generator suggested by “Numerical Recipes” fails systematically the `BirthdaySpacings` test, contrarily to our selection. The Java standard generator is, in fact, unusable.¹⁰

Additionally, Table XI reports the results of Dieharder: the main observation is that at this level of quality Dieharder is unable to make any distinction between the generators, except for the case of the Java generator.

7.2. Escaping zeroland

We show in Figure 18 the speed at which a few of the generators of Table X “escape from zeroland” [Panneton et al. 2006]: purely linearly recurrent generators with a very large state space need a very long time to get from an initial state with a small number of ones to a state in which the ones are approximately half. The figure shows a measure of escape time given by the ratio of ones in a window of 4 consecutive 64-bit values sliding over the first 100 000

¹⁰Note that we report the number of *failed tests* on our 100 seeds. L’Ecuyer and Simard [L’Ecuyer and Simard 2007] report the number of *types of failed tests* (e.g., failing two distinct RandomWalk tests counts as one) on a single run, so some care must be taken when comparing the results we report and those reported by them.

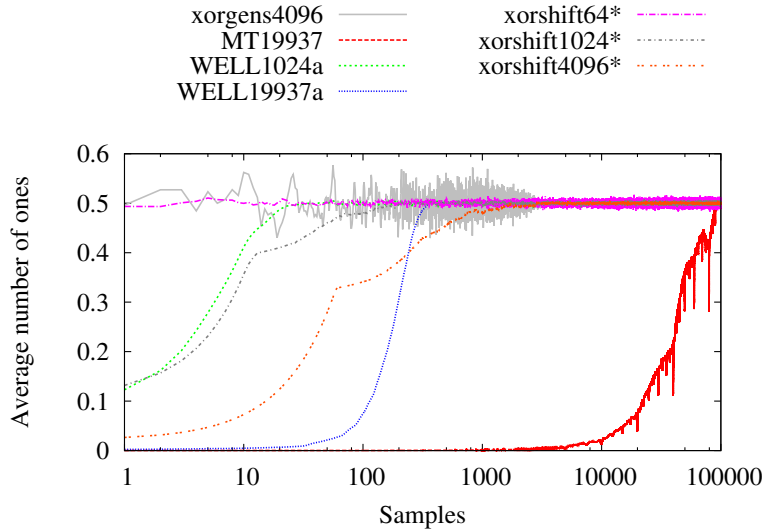


Fig. 18. Convergence to “half of the bits are ones in average” plot.

Table XII. Mean and variance for the data shown in Figure 18.

Algorithm	Mean	Variance
xorshift64*	0.5000	0.0039
xorgens4096	0.4999	0.0030
xorshift1024*	0.4999	0.0035
WELL1024a	0.4999	0.0036
xorshift4096*	0.4991	0.0110
WELL19937a	0.4991	0.0184
MT19937	0.3129	0.1689

generated values, averaged over all possible seeds with exactly one bit set (see [Panneton et al. 2006] for a detailed description).

As it is known, MT19937 needs hundreds of thousands of iterations to start behaving correctly. `xorshift4096*` and `xorgens4096` need a few thousand (but `xorgens4096` oscillates always around 1/2), `WELL19937a` and `xorshift1024*` a few hundreds, whereas `WELL1024a` just a few dozens, and `xorshift64*` is almost unaffected.

Table XII condenses Figure 18 into the mean and variance of the displayed values. Clearly, the multiplication step helps in reducing the correlation between the number of ones in the state space and the number of ones in the output values. Also, the slowness in recovering from states with too many zeroes is directly correlated to the size of the state space—a very good argument against linear generators with too large state spaces.

7.3. Speed

Finally, we benchmark the generators of Table X. Our tests were run on an Intel® Core™ i7-4770 CPU @3.40GHz (Haswell), and the results are shown in Table XIII (variance is undetectable, as we generate 10^{10} values in each test). We also report as a strong baseline results about SFMT19937, the *SIMD-Oriented Fast Mersenne Twister* [Saito and Matsumoto

Table XIII. Time to emit a 64-bit integer on an Intel® Core™ i7-4770 CPU @3.40GHz (Haswell).

Algorithm	Speed (ns/64 bits)
xorshift64*	1.60
xorshift1024*	1.36
xorshift4096*	1.36
xorgens4096	1.68
MT19937	4.09
SFMT19937	1.54
WELL1024a	5.18
WELL19937a	8.01

2008], a 128-bit version of the Mersenne Twister based on the SSE2 extended instruction set of Intel processors (and thus not usable, in principle, on other processors).

The highest speed is achieved by the high-dimensional `xorshift*` generators. Note that the timings in Table XIII include the looping logic, which we approximately benchmarked at 17ns/iteration. This means that the `xorgens4096` generator is actually 27% slower than a `xorshift1024*` or `xorshift4096*` generator. `SFMT19937` is a major improvement in speed over `MT19937`, albeit slightly slower than a high-dimensional `xorshift*` generator; it fails systematically, moreover, the same tests of `MT19937`.

A `xorshift64*` generator is actually *slower* than its high-dimensional counterparts. This is not surprising, as the three shift/xors in a `xorshift64*` generator form a dependency chain and must be executed in sequence, whereas two of the shifts of a higher-dimension generator are independent and can be internally parallelized by the CPU. `WELL1024a` and `WELL19937a` are heavily penalized by their 32-bit structure.

```
#include <stdint.h>

uint64_t x;

uint64_t next() {
    x ^= x >> 12; // a
    x ^= x << 25; // b
    x ^= x >> 27; // c
    return x * 2685821657736338717LL;
}
```

Fig. 19. The suggested `xorshift64*` generator in C99 code. The variable `x` should be initialized to a nonzero seed before calling `next()`.

8. CONCLUSIONS

After our careful experimental analysis, we reach the following conclusions:

A `xorshift1024*` generator is an excellent choice for a general-purpose, high-speed generator. The statistical quality of the generator is very high (it has, actually, the best results in BigCrush), and its period is so large that the probability of overlapping sequences is practically zero, even in the largest parallel simulation. Nonetheless, the state space is reasonably small, so that seeding it with high-quality bits is not too expensive,

```

#include <stdint.h>

uint64_t s[ 16 ];
int p;

uint64_t next(void) {
    uint64_t s0 = s[ p ];
    uint64_t s1 = s[ p = ( p + 1 ) & 15 ];
    s1 ^= s1 << 31; // a
    s1 ^= s1 >> 11; // b
    s0 ^= s0 >> 30; // c
    return ( s[ p ] = s0 ^ s1 ) * 1181783497276652981LL;
}

```

Fig. 20. The suggested `xorshift1024*` generator in C99 code. The array `s` should be initialized to a nonzero seed before calling `next()`.

```

#include <stdint.h>

uint64_t s[ 64 ];
int p;

uint64_t next(void) {
    uint64_t s0 = s[ p ];
    uint64_t s1 = s[ p = ( p + 1 ) & 63 ];
    s1 ^= s1 << 25; // a
    s1 ^= s1 >> 3; // b
    s0 ^= s0 >> 49; // c
    return ( s[ p ] = s0 ^ s1 ) * 8372773778140471301LL;
}

```

Fig. 21. The suggested `xorshift4096*` generator in C99 code. The array `s` should be initialized to a nonzero seed before calling `next()`.

and recovery from states with a large number of zeroes happens quickly. The generator is also blazingly fast (it is actually the fastest generator we tested), providing a 64-bit value in slightly more than a nanosecond. The reasonable state space makes also more likely, in case a large number of generators is used at the same time, that their state can fit the cache. In any case, with respect to other generators, the state space is accessed in a more localized way, as read and write operations happen *at two consecutive locations*, and thus will generate at most one cache miss.

In case memory is an issue, or array access is expensive, a very good general-purpose generator is a `xorshift64*` generator. While the generator $A_1(12, 25, 27) \cdot M_{32}$ fails systematically the MatrixRank test, it has less linear artifacts than MT19937, WELL1024a or WELL19937a, which fail systematically even more tests. It is a very good choice if memory footprint is an issue and a very large number of generators is necessary. A `xorshift64*` generator can also actually be *faster* than a `xorshift1024*` generator if the underlying language incurs in significant costs when accessing an array: for instance, in Java a `xorshift64*` generator emits a value in 1.62 ns, whereas a `xorshift1024*` generator needs 2.08 ns.

Linear generators with an excessively long period have a number of problems that are not compensated by higher statistical quality. Generating 64 bits with

WELL19937a requires almost ten times the time required by a xorshift1024* generator, with no detectable improvement in the statistical quality of the output by means of test suites; moreover, recovery from state spaces with many zeroes, albeit enormously improved with respect to MT19937, is still very slow, and seeding properly the generator requires almost twenty thousands random bits. In the end, it is in general difficult to motivate state spaces larger than 2^{1024} . Similar considerations are made, for example, by L'Ecuyer and Panneton [2005].

Surprisingly simple and fast generators can produce sequences that pass strong statistical tests. The code in Figure 20 is extremely shorter and simpler than that of MT19937, WELL1024a or WELL19937a. Yet, it performs significantly better on BigCrush. It is a tribute to Marsaglia's cleverness that just eight logical operations, one addition and one multiplication by a constant can produce sequences of such high quality. xorgens generators are similar with this respect, but use several more operations due to the additional shift and to the usage of a *Weyl generator* to hide linear artifacts [Brent 2007].

The t for which the multiplier has a good figure of merit has no detectable effect on the quality of the generator. If our tests, we could not find any significant difference between the behavior of generators based on M_{32} , M_8 or M_2 . It could be interesting to experiment with multipliers having very *bad* figures of merit.

Equidistribution is more useful as a design feature than as an evaluation feature. While *designing* generators around equidistribution might be a good idea, as it leads in general to good generators, *evaluation* by equidistribution is a more delicate matter because of high-bits bias and of the failure to detect the generators having the best scores in statistical suites (actually, as we have seen, some of the *worst* generators could be chosen instead).

TestU01 has significantly more resolution than Dieharder as a test suite. In particular in the high-dimension case, TestU01 is able to provide useful information, whereas Dieharder scores flatten down. However, TestU01 should be applied always to the reverse generator, too, to account for its high-bits bias.

REFERENCES

- Richard P. Brent. 2004. Note on Marsaglia's Xorshift Random Number Generators. *Journal of Statistical Software* 11, 5 (2004), 1–5.
- Richard P. Brent. 2007. Some long-period random number generators using shifts and xors. In *Proceedings of the 13th Biennial Computational Techniques and Applications Conference, CTAC-2006 (ANZIAM J.)*, Wayne Read and A. J. Roberts (Eds.), Vol. 48. C188–C202.
- Richard P. Brent. 2010. The myth of equidistribution for high-dimensional simulation. *CoRR* abs/1005.1320 (2010).
- Robert G. Brown. 2013. Dieharder: A Random Number Test Suite (Version 3.31). (2013). Retrieved January 8, 2014 from <http://www.phy.duke.edu/~rgb/General/dieharder.php>
- Aaldert Compagner. 1991. The hierarchy of correlations in random binary sequences. *Journal of Statistical Physics* 63, 5-6 (1991), 883–896.
- Shin Harase. 2011. An efficient lattice reduction method for \mathbf{F}_2 -linear pseudorandom number generators using Mulders and Storjohann algorithm. *J. Comput. Appl. Math.* 236, 2 (2011), 141–149.
- Maurice G. Kendall. 1938. A New Measure of Rank Correlation. *Biometrika* 30, 1/2 (1938), 81–93.
- Maurice G. Kendall. 1945. The treatment of ties in ranking problems. *Biometrika* 33, 3 (1945), 239–251.
- Pierre L'Ecuyer. 1999. Tables of linear congruential generators of different sizes and good lattice structure. *Math. Comput* 68, 225 (1999), 249–260.
- Pierre L'Ecuyer and François Panneton. 2005. Fast random number generators based on linear recurrences modulo 2: overview and comparison. In *Proceedings of the 37th Winter Simulation Conference*. Winter Simulation Conference, 110–119.
- Pierre L'Ecuyer and Richard Simard. 2007. TestU01: A C library for empirical testing of random number generators. *ACM Trans. Math. Softw.* 33, Article 22 (August 2007). Issue 4.

- Robert H. Lewis. 2013. Fermat: A Computer Algebra System for Polynomial and Matrix Computation (Version 5.1). (2013). Retrieved January 8, 2014 from <http://home.bway.net/lewis/>
- Rudolf Lidl and Harald Niederreiter. 1994. *Introduction to finite fields and their applications*. Cambridge University Press, Cambridge.
- George Marsaglia. 2003. Xorshift RNGs. *Journal of Statistical Software* 8, 14 (2003), 1–6.
- George Marsaglia and Liang-Huei Tsay. 1985. Matrices and the structure of random number sequences. *Linear Algebra Appl.* 67 (1985), 147–156.
- Makoto Matsumoto and Takuji Nishimura. 1998. Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator. *ACM Trans. Model. Comput. Simul.* 8, 1 (1998), 3–30.
- Harald Niederreiter. 1992. *Random number generation and quasi-Monte Carlo methods*. CBMS-NSF regional conference series in Appl. Math., Vol. 63. SIAM.
- François Panneton. 2004. *Construction d'ensembles de points basé sur une récurrence linéaire dans un corps fini de caractéristique 2 pour la simulation Monte Carlo et l'intégration quasi-Monte Carlo*. Ph.D. Dissertation. Université de Montréal.
- François Panneton and Pierre L'Ecuyer. 2005. On the xorshift random number generators. *ACM Trans. Model. Comput. Simul.* 15, 4 (2005), 346–361.
- François Panneton, Pierre L'Ecuyer, and Makoto Matsumoto. 2006. Improved long-period generators based on linear recurrences modulo 2. *ACM Trans. Math. Softw.* 32, 1 (2006), 1–16.
- William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. 2007. *Numerical recipes: the art of scientific computing*. Cambridge University Press.
- Andrew Rukhin, Juan Soto, James Nechvatal, Miles Smid, Elaine Barker, Stefan Leigh, Mark Levenson, Mark Vangel, David Banks, Alan Heckert, James Dray, and San Vo. 2001. *A Statistical Test Suite For Random and Pseudorandom Number Generators for Cryptographic Applications*. National Institute for Standards and Technology, pub-NIST:adr. NIST Special Publication 800-22, with revisions dated May 15, 2001.
- Mutsuo Saito. 2013. MTToolBox (Version 0.2). (2013). Retrieved January 8, 2014 from <http://msaito.github.io/MTToolBox/en/>
- Mutsuo Saito and Makoto Matsumoto. 2008. SIMD-Oriented Fast Mersenne Twister: a 128-bit Pseudo-random Number Generator. In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, Alexander Keller, Stefan Heinrich, and Harald Niederreiter (Eds.). Springer, 607–622.