

Minkowski Functionals Study of Random Number Sequences

Xinyu Zhang¹, Seth Watts², Yaohang Li³, and Daniel Tortorelli²

¹ College of Engineering, North Carolina A&T State University,
Greensboro, NC 27411
xzhang@ncat.edu

² Department of Mechanical Science and Engineering, University of Illinois at
Urbana-Champaign, Urbana, IL 61801
{watts2, dtortore}@uiuc.edu

³ Department of Computer Science, North Carolina A&T State University,
Greensboro, NC 27411
yaohang@ncat.edu

Abstract. Random number sequences are used in a wide range of applications such as simulation, sampling, numerical analysis, cryptography, and recreation. The quality of random number sequences is critical to the correctness of these applications. Many statistical tests have been developed to test various characteristics of random number generators such as randomness, independence, uniformity, etc. Most of them are based on testing on a single sequence. When multiple sequences are employed in an application, their potential correlations are also concerned. In this paper, we explore the techniques of using the Minkowski functionals and their extensions, the Minkowski valuations, to study the mathematical morphology of two dimensional binary image generated by pairwise random number sequences, and apply this method to describe and compare the properties of several well-known pseudo- and quasi-random number generators.

Keywords: Minkowski functionals, random number, random number test, point pattern.

1 Introduction

Random number sequences are desired to display no describable deterministic patterns, but follow a certain statistical distribution. The quality of a random number generator is usually measured by efficiency, uniformity, independence, randomness, reproducibility, and aperiodicity. To test the quality of random number sequences, many statistical tests suites are available [1, 2]. Most of these tests are designed for testing a single sequence. However, in many applications such as parallel Monte Carlo, multiple sequences are involved and the potential correlations among these sequences may also affect the correctness of these applications. To ensure random behavior across multiple random number sequences, studies of statistical testing on interleaving sequences [3] as well as averaging sequences have been used in the past [4]. In this paper, we explore the testing of sequence correlations in two dimensions (2D).

The study of the pair-wise random number sequence correlation is the foundation of the analysis of correlations among multiple sequences. To identify the correlations between sequences, a tool to describe point distribution in 2D quantitatively is required. Recently, the Minkowski functionals have been used to quantify patterns found in galaxies, neuronal cells, and metal foams as well as random point patterns [5, 6, 7]. In 2D, the Minkowski functionals correspond to area V_0 , perimeter V_1 , and Euler characteristic V_2 , i.e.:

$$V_0 = \int_P d^2 A, V_1 = \frac{1}{4} \int_{\partial P} d^1 S, V_2 = \frac{1}{2\pi} \int_{\partial P} \kappa d^1 S, \tag{1}$$

where κ denotes the curvature of along the boundary ∂P . The Minkowski functionals distill the complexity of an image into a small number of descriptors. However, they cannot fully describe morphological properties such as heterogeneity, symmetry, and anisotropy of images. To overcome the shortcoming, they are extended to the Minkowski valuations, which are the higher order moments of the Minkowski functionals [8]. The first- and second-order moments are of particular interest. Correspondingly in 2D, there are three first-order moments, \mathbf{V}_0 , \mathbf{V}_1 , and \mathbf{V}_2 , also referred to as the Minkowski vectors, given by:

$$\mathbf{V}_0 = \int_P \mathbf{x} d^2 A, \mathbf{V}_1 = \frac{1}{4} \int_{\partial P} \mathbf{x} d^1 S, \mathbf{V}_2 = \frac{1}{2\pi} \int_{\partial P} \kappa \mathbf{x} d^1 S, \tag{2}$$

where \mathbf{x} is the position vector. Moreover, the Minkowski vectors are usually normalized by their associated Minkowski functionals for a better geometric interpretation. These so-called centroids are defined as:

$$\mathbf{p}_i = \mathbf{V}_i / V_i \quad (i = 0,1,2 \text{ if } V_i \neq 0), \tag{3}$$

where \mathbf{p}_0 is the center of mass, \mathbf{p}_1 is the center of perimeter, and \mathbf{p}_2 is the center of curvature. The second-order moments define the second-order Minkowski tensors, $\mathbf{V}_0^{2,0}$, $\mathbf{V}_1^{r,s}$, and $\mathbf{V}_2^{r,s}$, viz:

$$\mathbf{V}_0^{2,0} = \int_P \mathbf{x} \otimes \mathbf{x} d^2 A, \mathbf{V}_1^{r,s} = \frac{1}{4} \int_{\partial P} \mathbf{x}^r \otimes \mathbf{n}^s d^1 S, \mathbf{V}_2^{r,s} = \frac{1}{2\pi} \int_{\partial P} \kappa \mathbf{x}^r \otimes \mathbf{n}^s d^1 S, \tag{4}$$

where $(r,s) = (2,0), (1,1), (0,2)$ indicates the degree of the tensor product of each vector with itself. However, only four of the above seven tensors carry independent information [9] and we will concentrate on the mass and perimeter tensors:

$$\text{mass tensor : } \mathbf{V}_0^{2,0} = \int_P \mathbf{x} \otimes \mathbf{x} d^2 A, \tag{5}$$

$$\text{perimeter tensor : } \mathbf{V}_1^{2,0} = \frac{1}{4} \int_{\partial P} \mathbf{x} \otimes \mathbf{x} d^1 S. \tag{6}$$

In this paper, we compare the Minkowski functionals of 2D binary images of the random number sequences generated by several well-known pseudo- and quasi-random number generators. We investigate the difference of the pseudo- and

quasi-random number sequences from the theoretical values, and their Minkowski valuations. We also study the Minkowski functionals in 2D random point distribution when correlation between random number sequences occurs. The remainder of the paper is organized as follows: Sect. 2 describes the general method of Minkowski functionals study of random number sequences, Sect. 3 introduces the measures of the Minkowski functionals and valuations, Sect. 4 analyzes the sequences generated by the random generators mentioned above, and Sect. 5 finalizes the conclusions.

2 Computation of the Minkowski Functionals and Valuations

The definitions of Minkowski functionals are given in equation (1). For binary images on a square lattice, the Minkowski functionals are linear combinations of elements including faces, edges, and vertices. The computation of Minkowski functionals is simply counting the total number of faces n_2 , edges n_1 , and vertices n_0 . The area, perimeter and Euler characteristic are computed by

$$V_0 = n_2, V_1 = -4n_2 + 2n_1, V_2 = n_2 - n_1 + n_0. \tag{7}$$

Michelsen *et al.* provided programming examples for counting n_2 , n_1 , and n_0 of 2D and 3D binary images [6]. Later Blasquez and Poraudeau gave a more efficient algorithm on 3D binary images by examining only half of a voxel’s neighbors and using binary decision diagrams [10]. Their method can also be applied to 2D images.

In 2D, two random number sequences are mapped to x and y coordinates of the points in a square lattice of $L \times L$. The double precision random numbers are multiplied by d and truncated to get an integer in $[0, L)$. Grains (discs or squares) are attached to the points and their sizes grow gradually. Figure 1 shows a square grain. As the lattices are square, it is reasonable to use square grains. A long sequence is divided into k subsequences and the Minkowski functionals are computed on each subsequence. As a result, this approach reveals both global and local properties.

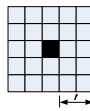


Fig. 1. Square grain with edge length of $2r+1$

We consider a collection of N points whose x and y coordinates are generated from uniform, uncorrelated random number sequences. In the bulk limit, when the volume approaches infinity and the density ρ is fixed, the averages $\langle V_i / N \rangle_N$ are given by [7]:

$$\langle V_0 / N \rangle_N = (1 - e^{-\rho m_0}) / \rho, \tag{8}$$

$$\langle V_1 / N \rangle_N = m_1 e^{-\rho m_0}, \tag{9}$$

$$\langle V_2 / N \rangle_N = (m_2 - m_1^2 \rho) e^{-\rho m_0}, \tag{10}$$

where $\langle V_i \rangle_N$ denotes the average of the Minkowski functionals of the point ensemble with density ρ and m_i denotes the mean values of the Minkowski functionals of a single grain.

By applying the normalized Minkowski functionals on a square lattice, and substituting the Minkowski functionals for a single square grain of edge length a , the theoretical values of point distribution of uniform, uncorrelated random number sequences can be derived as

$$\langle A \rangle = 1 - e^{-n}, \langle U \rangle = 4a\rho^{1/2} e^{-n}, \langle \chi \rangle_N = (1 - n) e^{-n}, \tag{11}$$

where $n = \rho a^2$ and A, U, χ , are the normalized Minkowski functionals ($A = V_0 / L^2, U = V_1 / LN^{1/2}, \chi = V_2 / N$) of a square [7]. Figure 2 shows the Minkowski functional curves with various densities. As the density decreases, the spans of the curves increase. While mapping a fixed length subsequence of random numbers to a lattice, a larger lattice size L yields a lower density and smoother curves, but demands more computation. In computation practice, one can control the density to be around 1% for a reasonable resolution.

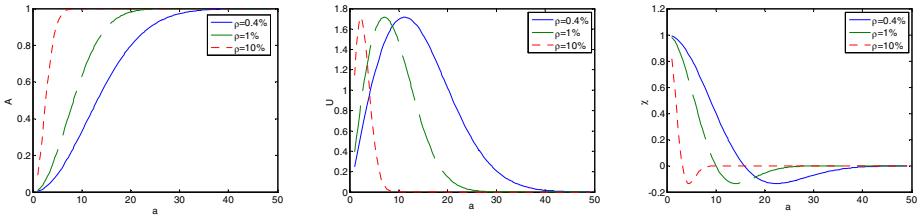


Fig. 2. Minkowski functionals as a function of square grain length

When computing the Minkowski valuations of 2D binary images, the equations (5) to (6) are used and the center of mass \mathbf{p}_0 and the mass tensor $\mathbf{V}_0^{2,0}$ are integrated on the pattern area, while the perimeter and curvature centroids \mathbf{p}_1 and \mathbf{p}_2 , and perimeter tensor $\mathbf{V}_1^{2,0}$ are integrated on the pattern boundary [9]. An efficient algorithm is given by Zhang *et al.* [11].

3 Measures of the Minkowski Functionals and Valuations

As the grain size grows, the image will eventually cover the whole lattice. The grain size of the full coverage is related to the largest gap. The perimeter curve first increases and then decreases to zero due to most grains at smaller size are isolated and their growth mainly contributes to increasing perimeter. When the growth of grains reaches a certain size and the overlapping dominate the process, as a result, the perimeter curve starts to decrease and eventually drops to zero if periodic boundary conditions are applied. The Euler characteristic curve can be explained in a similar

way. Initially, all of the grains are isolated and the Euler characteristic is 1. As the grains grow and overlapping occurs, the Euler characteristic decreases. When most of the grains overlap, the structure is dominated by holes and thus the Euler characteristic drops to negative. When the grains continue to grow, the holes start to be filled out and finally the Euler characteristic is 0 when the full coverage is reached.

Due to the discretization errors and the variations of random number generators, the experimental results may deviate from the theoretical values. To measure this difference, we compute the area between the two Minkowski functional curves using the trapezoid equation. Let $F(r)$ be the theoretical curve and $F_n(r)$ be the curve of a random number sequence, then the area is

$$D = \left(\sum_{i=1}^{r^*} |F(r_i) - F_n(r_i)| - \frac{(|F(r^*) - F_n(r^*)| + |F(r_1) - F_n(r_1)|)}{2} \right) \Delta r, \tag{12}$$

where r^* is the grain radius when the full coverage is reached for both curves.

In addition, we measure the distances dP of the centroids of the area, perimeter, or Euler characteristic \mathbf{P} from the image center \mathbf{C} ,

$$dP = |\mathbf{P} - \mathbf{C}| \tag{13}$$

For a normalized square lattice with length 1.0, \mathbf{C} is (0.5, 0.5).

Another important property of the random number sequences is the isotropy. The isotropy X is measured by the ratio of the two eigenvalues, λ_1 and λ_2 , of a Minkowski tensor,

$$X = \left| \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right|. \tag{14}$$

We measure the ratios of the eigenvalues of the area and perimeter tensors.

4 Random Number Generators

The middle-square method for random number generation was firstly suggested by John von Neumann. It has proved to be a comparatively poor source of pseudo-random numbers. The fundamental idea of the middle-square method is to take the square of the previous random integer and to extract the middle digits [2]. For example, one can take a four digit random integer 8653 and square it to get 74874409, and then take the middle 4 digits to create the next random integer of 8744. Most of the starting values will soon lead to a sequence with cycle of 6100, 2100, 4100, 8100, 6100 ... or degenerate to zero. With more digits, the period is larger and the quality is better.

Linear Congruential Generator (LCG), 64 bit Linear Congruential Generator (LCG64), Multiplicative Lagged Fibonacci Generator (MLFG), Lagged Fibonacci Generator (LFG), and Combined Multiple Recursive Generator (CMRG) are well-known “good” pseudo-random number generators provided by the SPRNG (Scalable Parallel Random Number Generators) library [12]. Parameterization is used in SPRNG library to generate parallel, independent random number sequences.

Quasi-random number sequences, also called low-discrepancy sequences, are designed to improve convergence rate of Monte Carlo integration. The sequences intend to provide high uniformity instead of randomness to achieve low discrepancy. The Halton, Faure', and Sobol are popular quasi-random number generators [13].

In this article, we study the Minkowski functionals of the pseudo-random number sequences generated by LCG, LCG64, LFG, MLFG, and CMRG provided by the SPRNG library and the middle-square generator as well as the Halton, Faure', and Sobol quasi-random number sequences. For each of the pseudo- or quasi-random number generator, we generate 100 pairs of sequences with length 1,024,000. Each pair of these sequences is divided to 100 pairs of subsequences and mapped to the two axes of a square lattice of size 1024×1024 to form a 2D binary image. Square grains are attached to the points. The growth is cut off at r equals 100. Periodic boundary conditions are used for all the cases. The Minkowski functional measures given in the last section are computed with the results given in Tables 1, 2, and 3.

Table 1. Area bounded by the Minkowski functional curve from random number sequences and the theoretical uniform uncorrelated point distribution

	D_A		D_U		D_χ	
	Mean	Std	Mean	Std	Mean	Std
Pseudo-random number sequences						
Middle-square 6 digits	60.0637	29.1301	57.8705	35.7068	26.0772	29.9295
Middle-square 8 digits	80.3811	20.2021	45.9626	26.5424	41.1672	40.6423
LCG	0.0122	0.0071	0.4052	0.0062	0.2946	0.0088
LCG64	0.0122	0.0072	0.4055	0.0063	0.2947	0.0087
LFG	0.0123	0.0072	0.4054	0.0063	0.2947	0.0088
MLFG	0.0123	0.0072	0.4053	0.0063	0.2949	0.0088
CMRG	0.0124	0.0072	0.4053	0.0063	0.2946	0.0088
Quasi-random number sequences						
Sobol	0.4451	0.2704	1.7844	0.7816	0.8696	0.3007
Halton	0.5408	0.2506	2.2789	0.8433	1.1447	0.3475
Faure	1.1830	0.9434	3.6864	2.1378	1.5648	0.3856

The middle-square generators have extreme large D values, since degeneration occurs in many cases. The generators in SPRNG show very close means and standard deviations in Table 1 and demonstrate good match with theoretical values of the Minkowski functionals. Figure 3 shows the distribution of D_A , D_U , and D_χ of the LCG samples.

The values of quasi-random numbers vary dramatically, since their high uniformity and poor randomness by design. As a result, some obvious patterns can be observed in Fig. 4 for Sobol and Faure' sequences. The averaged Minkowski functionals as a function of radius are plotted in Fig. 5 for the LCG, Sobol, Halton, and Faure' sequences. When degeneration happens in the middle-square generator, the density

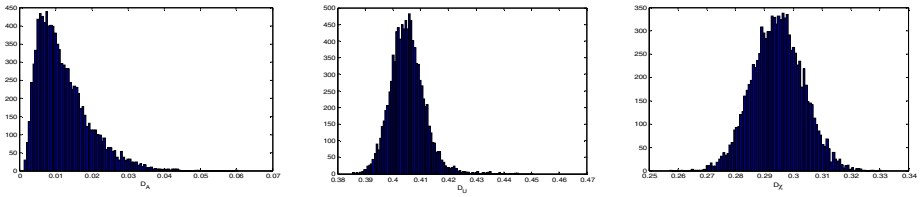


Fig. 3. Distributions of curve areas of the LCG sequences

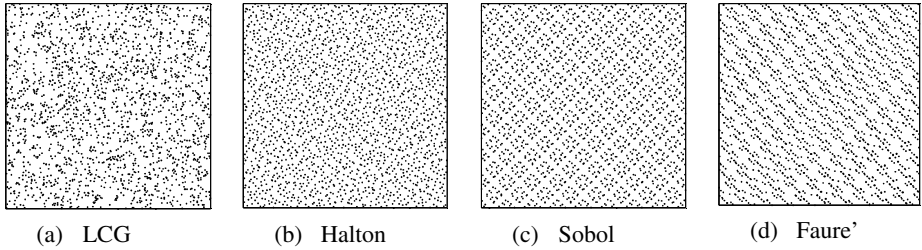


Fig. 4. 2D images of paired sequences generated from pseudo- and quasi-random number generators

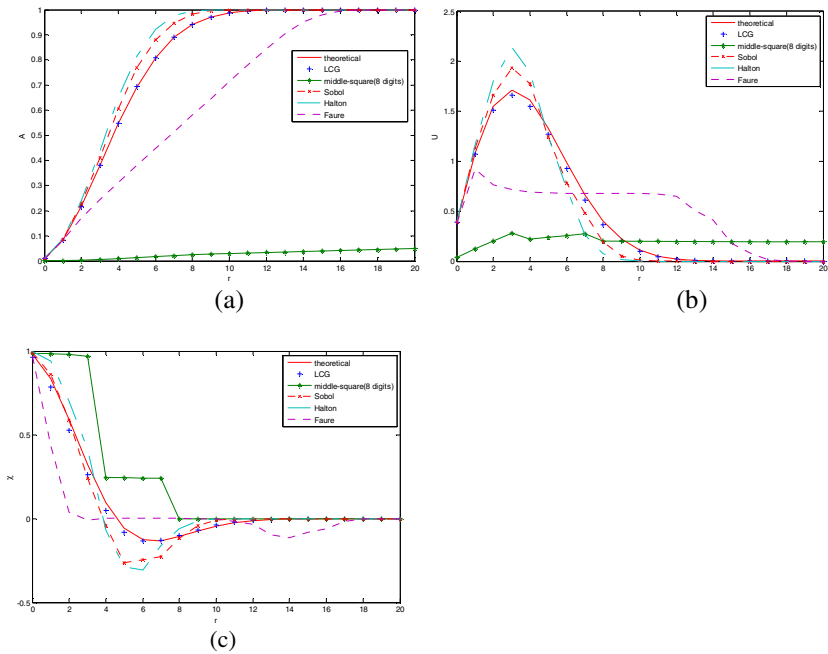


Fig. 5. Averaged Minkowski functionals as a function of the grain size

decreases and also causes a great deviation from the theoretical curve. The sequences such as Sobol and Halton with smaller gaps have their area values A grow faster and reach 1 (i.e., full coverage) at a smaller grain size, while the Faure' sequences have large gaps and reach the full coverage slowly. The degenerated patterns of middle-square reach their full coverage at a grain size greater than the cutoff radius 100.

The means and standard deviations of the distance of the centroids from the image center and the eigenvalue ratios are listed in Tables 2 and 3, respectively. The SPRNG pseudo-random number generators demonstrate similar characteristics. The quasi-random number generators yield rather different values due to their inherent patterns. The area and perimeter centroids of the quasi-random number generators are closer to the center with smaller variances, and the eigenvalue ratios of the quasi-random number sequences are smaller than pseudo-random numbers, which indicates better uniformity and isotropy.

Table 2. Distances of centroids from the image center

	dP_0		dP_1		dP_2	
	Mean	Std	Mean	Std	Mean	Std
Pseudo-random number sequences						
LCG	0.0013	0.0006	0.0907	0.0215	0.0894	0.0185
LCG64	0.0013	0.0006	0.0910	0.0223	0.0898	0.0191
LFG	0.0013	0.0006	0.0907	0.0221	0.0894	0.0189
MLFG	0.0013	0.0006	0.0910	0.0219	0.0898	0.0188
CMRG	0.0013	0.0006	0.0908	0.0221	0.0897	0.0189
Quasi-random number sequences						
Sobol	5.8681e-005	3.0018e-005	0.0589	0.0093	0.0622	0.0338
Halton	2.3473e-004	1.5721e-004	0.0753	0.0174	0.1336	0.3001
Faure	1.2252e-004	5.5186e-005	0.0465	0.0075	0.1911	0.3698

Table 3. Eigenvalue ratios of mass and perimeter tensors

	$X_{V_0^{2.0}}$		$X_{V_1^{2.0}}$	
	Mean	std	Mean	Std
Pseudo-random number sequences				
LCG	0.0018	0.0012	0.0835	0.0382
LCG64	0.0018	0.0012	0.0838	0.0383
LFG	0.0018	0.0012	0.0835	0.0388
MLFG	0.0018	0.0012	0.0837	0.0379
CMRG	0.0018	0.0012	0.0841	0.0388
Quasi-random number sequences				
Sobol	0.0002	0.0002	0.0096	0.0122
Halton	0.0004	0.0003	0.0340	0.0331
Faure	0.0001	0.0001	0.0055	0.0106

5 Correlation Identification

RANDU is an infamous pseudo-random number generator that has been used for decades on IBM mainframes [14]. If we simply generate random number doublets to build 2D binary images, it yields similar Minkowski values to those of the SPRNG generators provided in Tables 1-3. However, if we generate triplets x,y,z from a sequence and constructed a pair of random sequences in which one sequence comes from the x , and the other from $6x-9y+z$, all points fall into 15 lines in a stripe, as shown in Fig. 6(a), which indicates strong correlation [15]. In contrast, when a good random number generator is used, the points should fill in the stripe as shown in Fig. 6(b).

We compute the Minkowski functionals of the 2D binary images created by RANDU and LCG and CMRG in SPRNG, for 100 sequences and plot their averages as a function of the grain size. One can easily observe the difference between RANDU and SPRNG generators in Fig. 7. The area between the Minkowski functional curve and x -axis in Fig. 7(a) is 11.26 for RANDU, and 17.0359 for LCG and CMRG.

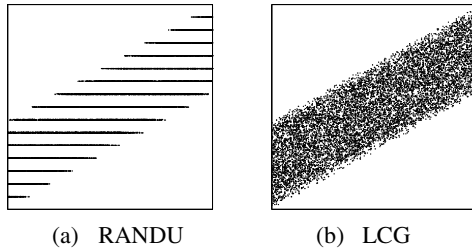


Fig. 6. Images of the RANDU and LCG sequences generated from x and $6x-9y+z$

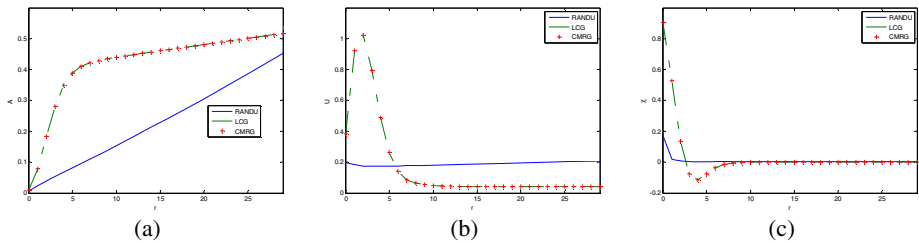


Fig. 7. Averaged Minkowski functionals as a function of grain size for RANDU, LCG, and CMRG sequences generated from x and $6x-9y+z$

6 Conclusions and Discussions

In this paper, we discussed the method of applying the Minkowski functionals and valuations to study random number sequences. Pairs of sequences are mapped to 2D lattices to form a binary image. Grains are attached to the points. We compute the Minkowski functionals and valuations as a function of the grain size. A close match

of the Minkowski functional curves with the theoretical curves of uniform uncorrelated point distribution indicates good randomness. The locations of the centroids of mass and perimeter, along with the ratios of mass and perimeter tensors give us some insight on the uniformity, symmetry, and isotropy of the patterns generated by random number sequences. Our examples given in this paper also showed that the Minkowski functionals are able to identify degenerated sequences, highly uniform sequences, and sequences with correlation. In summary, the Minkowski functionals and valuations can provide meaningful indication of the quality of random number sequences and are potential tools for testing new developed random number generators as a complementary to the existing statistical tests.

References

1. Soto, J.: Statistical Testing of Random Number Generators. In: Proceedings of the 22nd National Information Systems Security Conference (October 1999)
2. Knuth, D.: The Art of Computer Programming, vol. 2. Addison-Wesley, Reading (1969)
3. Cuccaro, S., Mascagni, M., Pryor, D.: Techniques for Testing the Quality of Parallel Pseudorandom Number Generators. In: Proceedings of the 7th SIAM Conf. on Parallel Processing for Scientific Computing, pp. 279–284. SIAM, Philadelphia (1995)
4. Coddington, P., Ko, S.: Techniques for Empirical Testing of Parallel Random Number generators. In: Proceedings of International Conference on Supercomputing, pp. 282–288 (1998)
5. Beisbart, C., Buchert, T., Wagner, H.: Morphology of spatial patterns. *Physica A* 293/3-4, 592–604 (2001)
6. Michielsen, K., Raedt, H., Hosson, J.: Aspects of Mathematical Morphology. *Advances in imaging and electron physics* 125, 119–194 (2002)
7. Michielsen, K., Raedt, H.: Integral-Geometry Morphological Image Analysis. *Physics Reports* 347, 461–538 (2001)
8. Beisbart, C.B.: Vector- and Tensor-Valued Descriptors for Spatial Patterns. In: *Morphology of Condensed Matter, Physics and Geometry of Spatial Complex Systems*, vol. 600(J), pp. 238–260. Springer, Berlin (2002)
9. Beisbart, C.: Measuring Cosmic Structure. Minkowski Valuations and Mark Correlations for Cosmological Morphometry, Dissertation, Ludwig-Maximilians-Universität, München (2001)
10. Blasquez, I., Poiradeau, J.-F.: Efficient Processing of Minkowski Functionals on a 3D Binary Image using Binary Decision Diagrams. *Journal of WSCG* 11(1) (2003)
11. Zhang, X., Watts, S., Tortorelli, D.: Computation of First- and Second-order Minkowski Valuations for Two Dimensional Binary Images. Submitted to *Applied Mathematical Modelling*
12. Mascagni, M., Srinivasan, A.: Algorithm 806: SPRNG: A Scalable Library for Pseudorandom Number Generation. *ACM Transactions on Mathematical Software* 26, 436–461 (2000)
13. Li, Y.: The Computational Measure of Uniformity, Master thesis, Florida State University (2000)
14. Zeitler, D., McKean, J., Kapenga, J.: Empirical Spectral Analysis of Random Number Generators. In: Proceedings of the 34th Symposium on the Interface. Montreal, Quebec, Canada (April 2002)
15. Marsaglia, G.: Random Numbers Fall Mainly in the Planes. *PNAS* 61, 25–28 (1968)