JOURNAL OF COMPUTATIONAL PHYSICS 71, 391–428 (1987)

Maximum-Length Sequences, Cellular Automata, and Random Numbers

A. COMPAGNER AND A. HOOGLAND

Laboratorium voor Technische Natuurkunde, Lorentzweg 1, 2628 CJ Delft, the Netherlands

Received August 21, 1986

The statistical properties of maximum-length sequences generated by linear-feedback shift registers are investigated. Two-dimensional bit patterns constructed from these sequences are found to exhibit characteristic triangular structures, in the simplest case consisting of triangles of zeros supported by a backbone of ones. The size distribution of these triangles is given, and the relation with cellular automata is pointed out. Randomness is defined in terms of the correlation functions of any order. Not only the pair correlation function, but also all higher order correlation functions are two-valued. The number of correlation functions of any order that deviate from randomness is determined exactly. All maximum-length sequences of the same length are found to be equally random in a strict sense, although their suitability as a source of random numbers may differ widely in different applications; longer sequences are more random. It is concluded that extremely long maximum-length sequences, with periods of up to $2^{9689}-1$, can in fact be used as reliable random-number generators for many purposes. Fast software or hardware programs for the production of these sequences are readily constructed. © 1987 Academic Press, Inc.

1. INTRODUCTION

Maximum-length sequences generated by linear-feedback shift registers are often used to produce the random numbers needed in simulation. In comparison with other deterministic methods of producing random numbers (undeterministic methods are unwanted because of their irreproducibility), they have two main advantages. First, the statistical properties of maximum-length sequences can be fairly well understood. Second, fast linear-feedback shift registers can be easily realized both in software and in hardware: only the logic exclusive-or operation is involved. For these reasons, maximum-length sequences generated by linear-feedback shift registers were used in the special-purpose processors built in Delft by Hoogland *et al.* [1] and in Santa Barbara by Pearson *et al.* [2] for the Monte-Carlo simulation of spin systems; these relatively inexpensive machines are very fast and can be operated days on end, swallowing in a single run typically 10^{12} random numbers of 32 bits each, at a rate of 1 MHz or more. The demands of both speed and quality put on the method of generating the random numbers then are very severe indeed.

0021-9991/87 $3.00

Copyright © 1987 by Academic Press, Inc.
All rights of reproduction in any form reserved.

Maximum-length binary sequences and their main statistical properties are described by Golomb [3], Hoffmann de Visme [4], and MacWilliams and Sloane [5]. More detailed studies of these properties, a great variety of statistical criteria to test them, and a large number of particular applications have been reported by Tausworthe [6], Lindholm [7], Tootill *et al.* [8, 9], Lewis and Payne [10], Fredricsson [11], Miller and Mars [12], Arvillias and Maritsas [13], Kirkpatrick and Stoll [14], and many others. It is generally recognized that persistent deviations from pure randomness do exist in maximum-length sequences, although both the pair correlations between bits of the sequence and the number of runs of a certain size are almost what they should be for truly random bit sequences. The extent of these deviations from randomness is not clear, but many recipes have been suggested in the literature to reduce their influence, which depends on the particular linear-feedback shift register used, on the particular way in which the bits of the maximum-length sequence are combined to form a random number, and even on the particular problem that one wants to study with these random numbers. It is therefore important to develop a general point-of-view from which the nonrandom effects can be judged.

Such a point-of-view is found here in the consideration of the complete hierarchy of correlation functions of any order. In terms of this hierarchy the concept of randomness can be fully understood; this follows from an application to the case of maximum-length sequences of a description given by Glauber [15] for the hierarchy of correlation functions for the one-dimensional Ising model. (In fact, a maximum-length sequence can be considered as the ground-state of a particular one-dimensional Ising model, with an anisotropic Hamiltonian playing the part of the production rule.) Higher-order correlation functions for maximum-length sequences were considered earlier by Lindholm [7] and Fredricsson [11], whose results, which perhaps have received too little attention in the literature, are recovered and extended in this paper.

To make this paper reasonably self-contained, in Section 2 we give a short survey of the basic properties of maximum-length sequences. Section 3 contains a large number of figures showing two-dimensional bit patterns obtained from maximum-length sequences; these bit patterns are found to exhibit triangular structures which indicate the presence of strong correlations between bits of the sequence and which are related to the triangular bit structures reported by Miller [16] for the case of a particular type of symmetric tessellations and by Wolfram [17] for the case of cellular automata. We will then study, in Section 4, the complete hierarchy of correlation functions of any order. Section 5 contains a discussion of the statistical properties of (relatively) small subsequences of a maximum-length sequence. Application to sequences with periods up to the truly maximum length of $2^{9689}-1$ bits is the subject of Section 6, which also contains a discussion. In the last section the main conclusions of this paper are drawn.

2. Survey and Definitions

In an abstract sense, a linear-feedback shift register of length n is a rule for the growth of a binary sequence $\{a_t\}$ from a seed of n bits $a_1, a_2, \ldots, a_n$. This rule is

$$a_t = a_{t-n} + \sum_{j=1}^{n-1} c_j a_{t-j} \pmod 2, \tag{1}$$

where the coefficients c_j are either 0 or 1. The set of positive integers $j < n$ for which $c_j = 1$ holds are the additional feedback positions and will be indicated by $\{j\}$; if this set has $m-1$ elements the register is called a (linear) m-bit feedback shift register. A 1-bit feedback shift register would just be the cycling register which repeats the initial seed according to $a_t = a_{t-n}$; this is excluded by assuming that the set $\{j\}$ is never empty. A linear-feedback shift register is completely specified by n and the set $\{j\}$, that is, by the production rule $R(n; \{j\})$.

Starting with $i = n+1$, one finds the complete binary sequence by iterating Eq. (1), each time using the last n bits produced as the seed for the next bit. Since Eq. (1) can equally well be used to determine bit a_{t-n} if bits $a_{t-n+1}, a_{t-n+2}, \ldots, a_t$ are given, the sequence may also be produced backwards. The sequences produced by $R(n; \{j\})$ and $R(n; \{n-j\})$ are each other's reverse, if the initial seed is also reversed; these production rules and sequences are called conjugate. All sequences grown are periodic, with period p and $a_{t+p} = a_t$, because the intial seed must ultimately be reproduced, there being only 2^n different seeds; the period p is equal to the number of different seeds contained in the sequence. All linear-feedback shift registers produce at least two sequences, one of which is the trivial all-zero sequence (with period 1) grown from the all-zero seed. If the linear-feedback shift register produces just two sequences, the other sequence is a maximum-length sequence, with period $p = N \equiv 2^n - 1$ and containing all N nonzero seeds. Maximum-length sequences can only be produced by shift registers with an odd number of additional feedback positions. The periods of all nonzero sequences produced by a linear-feedback shift register always add up to N. As examples, $R(4; 3)$ produces the maximum-length sequence 111100010011010 with period $N = 2^4 - 1 = 15$, and $R(5; 4)$ produces the sequences 111110000100011001010, 1110100, and 110 with periods 21, 7, and 3 (adding up to $N = 31$). It is often convenient to study the equivalent sequence $\{b_t\}$ with $b_t = 1 - 2a_t = \pm 1$ instead of the sequence $\{a_t\}$ with $a_t = 0, 1$. The production rule of Eq. (1) for the sequence $\{b_t\}$ reads

$$b_t = b_{t-n} \prod_{j=1}^{n-1} (1 - c_j + c_j b_{t-j}). \tag{2}$$

In statistical mechanics the sequence $\{b_t\}$ could be interpreted as the p-fold degenerate ground state of a 1-dimensional Ising system of p spins $b_t = \pm 1$ with a ferromagnetic ($\varepsilon > 0$) multiple-spin interaction given by the Hamiltonian

$$H = -\varepsilon \sum_{t=1}^{p} b_t b_{t-n} \prod_{j=1}^{n-1} (1 - c_j + c_j b_{t-j}).$$

A characteristic polynomial $f(x)$ is assigned to $R(n; \{j\})$ by means of

$$f(x) = 1 + \sum_{i=1}^{n} c_i x^i = 1 + \sum_{i \in \{j\}} x^i + x^n. \tag{3}$$

A necessary condition for the production of a maximum-length sequence by $R(n; \{j\})$ is that this polynomial of degree n is irreducible (i.e., nonfactorable mod 2). When n is such that $N = 2^n - 1$ is prime (a Mersenne prime) this condition is also sufficient. The characteristic function of a maximum-length sequence can be shown to be essentially equal to the inverse of the generating function for that sequence; see, e.g., Golomb [3]. Lists of primitive polynomials that characterize maximum-length sequences have been given by Zierler and Brillhart [18], Zierler [19], and Stahnke [20].

Golomb [3] showed that maximum-length sequences have three randomness properties, called $R1$, $R2$, and $R3$ by him, which up to terms of order $1/N$ agree with what one would expect for truly random sequences. They are: ($R1$) A period of a maximum-length sequence contains $\frac{1}{2}(N+1)$ ones and $\frac{1}{2}(N-1)$ zeros. ($R2$) The number of runs of i ones (sandwiched between two zeros) or i zeros (sandwiched between two ones) within a period is 2^{n-i-2} for $1 \leqslant i \leqslant n-2$; in addition, there is a single run of $n-1$ zeros and a single run of n ones. ($R3$) The pair correlation over the period is given, in terms of the sequence $\{b_i\}$, by

$$C_N(j) \equiv \frac{1}{N} \sum_{i=1}^{N} b_i b_{i+j} = -\frac{1}{N} + \left(1 + \frac{1}{N}\right) \delta_{0,j}, \tag{4}$$

where Kronecker's δ-symbol is used.

To complete this summary, we recall the "cycle-and-add" property of maximum-length sequences: for any $k \neq 0$, there is always an integer s such that

$$a_{i+s} = a_i \oplus a_{i+k} \qquad \text{and} \qquad b_{i+s} = b_i b_{i+k} \tag{5}$$

hold for all i; see, e.g., Golomb [3] and Hoffmann de Visme [4]. A closed expression for s as a function of k (and of the parameters n and $\{j\}$ of the shift register) does not exist; a numerical strategy for the determination of s is given by the "remote-term calculation" described by Hoffmann de Visme [4]. The cycle-and-add property is perhaps the most fundamental property of maximum-length sequences; it immediately explains randomness property $R3$ in terms of $R1$, and it is at the basis of much of our findings in this paper.

By repeatedly sampling q bits from a maximum-length sequence one can form a sequence of q-bit numbers which, usually after reduction to the interval (0, 1), can be used as the random numbers needed in the Monte-Carlo or other stochastic process one is interested in. Different sampling schemes have been suggested for the q-bit random numbers, and a variety of statistical criteria have been applied to verify the randomness of the q-bit numbers obtained or to study randomness properties of the maximum-length sequence that go beyond $R1$, $R2$, and $R3$ cited

above. Useful recipes have been given for the selection of the most suitable linear-feedback shift registers. See, for details, the references given in Section 1. Some general comments on these subjects will be made in Section 6.

3. Bit Patterns of Maximum-Length Sequences

Complete chaos is difficult to achieve. Just how difficult, in the case of maximum-length sequences, will appear when we emphasize the intrinsic regularities induced by the deterministic production rule by arranging the sequences on a quadratic lattice.

Consider a row of bits consisting of many periods of a given maximum-length sequence $a_1, a_2, \ldots, a_N$. Add identical rows below (above), each time shifting them over s positions to the left (to the right, respectively), such that a bit pattern on a square lattice results. A column contains the bits $a_t, a_{t+s}, a_{t+2s}, \ldots$, sampled from the sequence. If s and $N = 2^n - 1$ have no common divisor, the decimation runs in a cyclic manner (i.e., mod N) through the entire sequence before repetition occurs; the decimated sequence is again a maximum-length sequence (see Golomb [3]). If N is prime, the bits along any rational direction in the bit pattern again form a maximum-length sequence. Conjugate shift registers generate upside-down bit patterns.

The bit patterns thus produced have a rhomboidal symmetry, with unit cells that are N unit squares large. Instead of these rhomboidal unit cells it is more convenient to consider tiles with edges that have horizontal or vertical parts only. Tiles cannot be defined uniquely, but in a given tessellation they should all have the same shape and orientation, and they should cover the square lattice completely. Each tile should contain all the elements of $\{a_t\}$ just once; equivalent positions on different tiles are to be occupied by the same element. An obvious choice, possible for any value of s and leading to a "brickwork" tessellation, would be to take just N subsequent bits on a row as tile; for most values of s more compact shapes can be chosen.

Certain special (though not mutually exclusive) values of s deserve attention. First consider the situation in which for some simulation q-bit random numbers (not all q bits need to be relevant in the simulation) are extracted from the maximum-length sequence by taking successive subsequences of q bits. Here, $q \ll n$ should hold in order to ensure that all q-bit random numbers occur frequently during a period of the sequence. The construction rule of the sequence implies that n subsequent bits of the sequence are independent; therefore, if m is the largest integer for which $mq \leqslant n$ holds, the closest pair in the generated list of q-bit random numbers that might show correlation is separated by $m - 1$ places in that list. An inspection of the bit pattern that results for $s = mq$ can provide a first indication about the existence of this correlation.

Second, by taking $s = n$ one obtains a bit pattern which contains successive parts of n bits of the sequence on successive rows of any n-bit wide column. A convenient

tile shape in this case would be a rectangle of n bits wide with a length of $\text{ent}(N/n)+1$ bits, from the last row of which, in the lower right-hand corner, a number of unit squares is missing. The indentation just accomodates the upper left-hand corner of a neighboring tile (see some of the figures).

Third, one can take s such that intrinsic regularities of the sequence are accentuated. The cycle-and-add property of Eqs. (5) tells us that there is always an integer $\sigma(1)$ such that $a_{t+\sigma(1)} = a_t \oplus a_{t+1}$ holds for any t (the general meaning of $\sigma(d)$ will be explained below). Therefore, if one chooses $s = \sigma(1)$ a bit pattern results in which every bit is the mod-2 sum of the bits immediately "north" and "north-east" of it.

In Fig. 1 the bit pattern resulting from the production rule $R(4; 3)$ with $s = n = \sigma(1) = 4$ is shown (for all production rules $R(n; n-1)$ the equality $\sigma(1) = n$ is valid); a possible tessellation is indicated. Each row in this figure is one period of the maximum-length sequence (ones are represented by black dots and zeros by open circles). Below the run of 4 ones a run of 3 zeros occurs, followed by runs of 2 and 1 zeros on subsequent rows. Thus, a base-3 triangle of zeros is formed; in addition, each tile contains an isolated zero, a "base-1 triangle." In Figs. 2 and 3 tiles are shown of the tessellations due to the production rules $R(7; 6)$ and $R(7; 4)$, with $s = \sigma(1) = n = 7$ and $s = \sigma(1) = 31$, respectively. The tiles are chosen for their convenient shape. The position of neighboring tiles is indicated, as is the collection of triangles of zeros. Figure 4 shows an arbitrary part of the bit pattern produced by the production rule $R(72; 66, 25, 19)$ with $s = \sigma(1)$. (All production rules used as examples here and in the sequel generate maximum-length sequences.) For the case of Fig. 4, the numerical value of $\sigma(1)$ is not known, but this does not prevent us from placing the mod-2 sum of the pair below the left bit of each pair of neighboring bits in a row. Figure 4 contains 128×128 bits and is an extremely small fraction $(1/2^{58})$ of the complete tile for this case.

Looking at these figures one would hardly believe that some maximum-length sequences are among the best pseudo-random binary sequences. The triangular bit structures observed, consisting of triangles of zeros supported by a backbone of ones, are a direct manifestation of the nonrandom higher-order correlations that exist within maximum-length sequences.

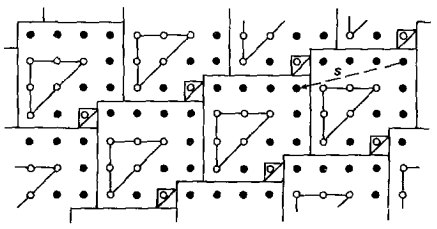

FIG. 1. The bit pattern resulting by repeatedly copying, after a shift over $s = 4$ places, the 15-bit maximum-length sequence generated by $R(4; 3)$. In this case, one has $s = \sigma(1) = n$. Black dots correspond with 1, open circles with 0. A possible tessellation is indicated, as are triangles of zeros.

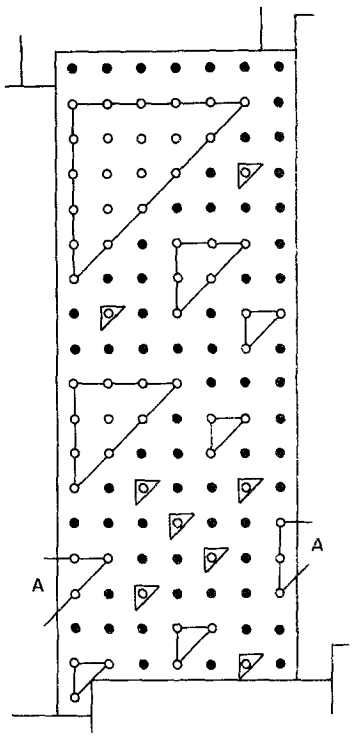

FIG. 2. A tile of a tessellation due to production rule $R(7; 6)$ with $s = \sigma(1) = 7$. The position of other tiles is indicated. At A, a triangle of zeros crossing the tile boundary is seen

The exact triangle distribution $T(n, m)$, defined as the number of base-m triangles of zeros per tile in the bit pattern with shift $s = \sigma(1)$ generated by a maximum-length sequence with period $N \equiv 2^n - 1$, is easily obtained. A base-m triangle consists of one run of m zeros (the base), followed on the next row by a run of $m - 1$ zeros, and so on; along the border of the triangle all nearest-neighboring bit positions on the square lattice are occupied by ones. The total number of zeros in a base-m triangle is $\frac{1}{2}m(m+1)$. The complete collection of triangles of zeros in a single tile must accomodate all runs of zeros that occur in a period of the

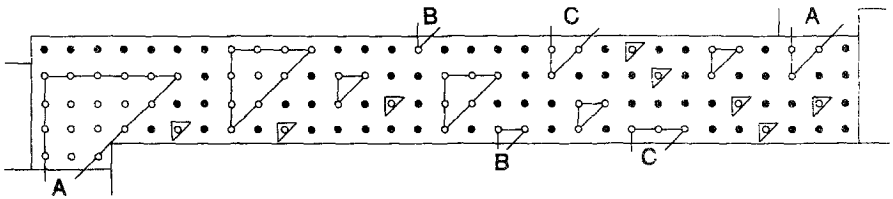

FIG. 3. The same as Fig. 2, but for $R(7; 4)$ with $s = \sigma(1) = 31$. Several triangles of zeros cross the tile boundary.

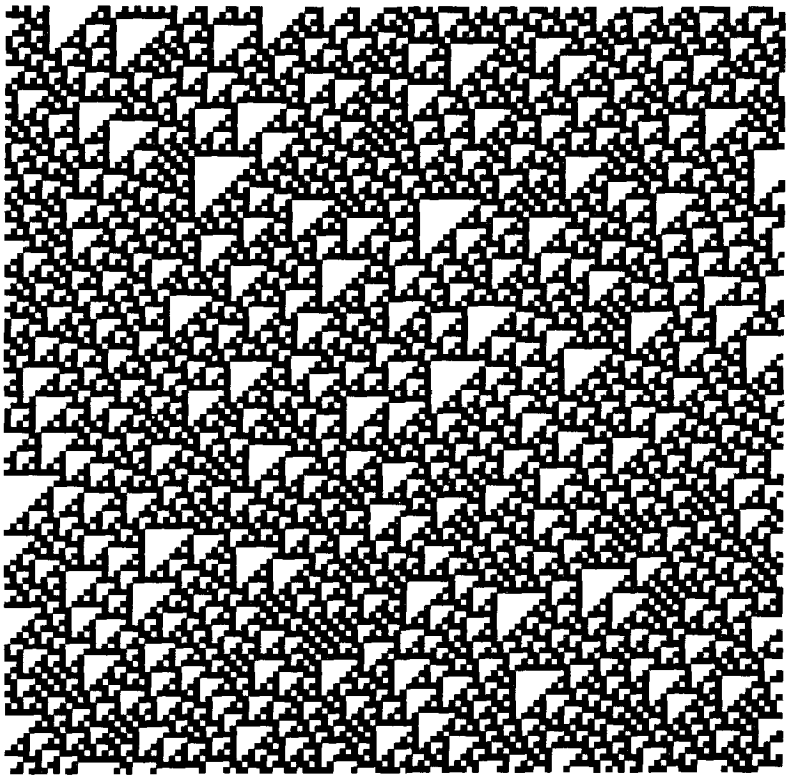

FIG. 4. An arbitrary part of 128×128 bits (i.e., $1/2^{58}$ of a complete tile) of the bit pattern generated by $R(72; 66, 25, 19)$ with $s = \sigma(1)$. Black squares correspond with 1, white squares with 0. The distribution of triangles of zeros observed in this part is within normal statistical fluctuations equivalent with the distribution for the complete tile.

maximum-length sequence; the number of these runs is given by randomness property *R2*. From these observations the triangle distribution follows immediately:

$$T(n, m) = \begin{cases} 2^{n-m-3} & \text{for} \quad 0 < m < n-2, \\ 1 & \text{for} \quad m = n-1, \\ 0 & \text{for} \quad m = n-2 \text{ and } m \geqslant n. \end{cases} \tag{6}$$

This distribution is the same for all bit patterns with $s = \sigma(1)$ obtained from any maximum-length sequence with period N. One easily verifies that the examples given in Figs. 1, 2, and 3 obey this distribution, and that the number $t(n, l)$ of runs of l zeros (sandwiched between ones) indeed is given by

$$t(n, l) = \sum_{m=l}^{n} T(n, m) = \begin{cases} 2^{n-l-2} & \text{for} \quad l \leqslant n-2, \\ 1 & \text{for} \quad l = n-1, \end{cases} \tag{7}$$

in agreement with randomness property *R*2. The actual distribution of triangles found in Fig. 4 agrees within normal statistical fluctuations with the distribution obtained by multiplying (6) with the ratio 2^{-58} between the part shown in Fig. 4 and the complete tile. This is a strong indication that properties observed over relatively small parts of the bit pattern converge in a uniform manner to the properties valid for a complete period of the maximum-length sequence, that is, for a complete tile.

The above can be generalized by noting that any mod-2 linear combination of n subsequent (and therefore linearly independent) bits of the sequence can be expressed as a single bit of the sequence:

$$a_{i+\sigma(d)} = \sum_{k=0}^{n-1} d_k a_{i+k} \pmod 2, \tag{8}$$

where the coefficients d_k are either 0 or 1, except d_0 which is always 1, and where $\sigma(d) \geqslant n$ holds. (The restriction $k < n$ is not really necessary; however, when this restriction is removed, the right-hand side of Eq. (8) will become zero for particular choices of the indices k for which $d_k = 1$, namely for the dependent subsets defined in the next section.) The quantity $\sigma(d)$ will be called the correlation distance; it depends on the coefficients d_k and on the parameters $(n; \{j\})$ of the production rule, but not on i. The integer d is a short-hand notation defined by

$$d = \sum_{k=1}^{n-1} d_k \cdot 2^{k-1}. \tag{9}$$

Alternatively, the set of coefficients will be indicated in binary notation by

$$\{d_k\} = (1 d_1 d_2 \cdots d_l), \tag{10}$$

where d_l with $l \leqslant n-1$ is the last coefficient that is not zero. For $d = 0$, one has $\{d_k\} = (1)$ and $\sigma(0) = 0$. For $d = 1$ one has $\{d_k\} = (11)$ and $a_{i+\sigma(1)} = a_i \oplus a_{i+1}$, in conformity with our earlier usage. As a last example, for $d = 5$ one has $\{d_k\} = (1101)$ and $a_{i+\sigma(5)} = a_i \oplus a_{i+1} \oplus a_{i+3}$.

Equation (8) is an immediate consequence of the cycle-and-add property, used repeatedly. The value of $\sigma(d)$ can be determined in principle by means of numerical techniques such as the remote-term calculation described by Hoffmann de Visme [4]; in practice, it can only be determined for relatively short shift registers ($N \lesssim 25$, say). In general, the behavior of $\sigma(d)$ is rather erratic.

From a given maximum-length sequence many different bit patterns exhibiting triangular structures can now be produced by taking $s = \sigma(d)$ for different values of d. Not only triangles of zeros appear, but also triangles of ones, and even stranger animals such as "checkerboard" and "zebra" triangles; of course, type and shape of the triangles depend strongly on d. We will give a few examples.

In Fig. 5 a bit pattern of the [illegible] shown; the numerical value of $\sigma(d)$ is not known, but the pattern is found by using

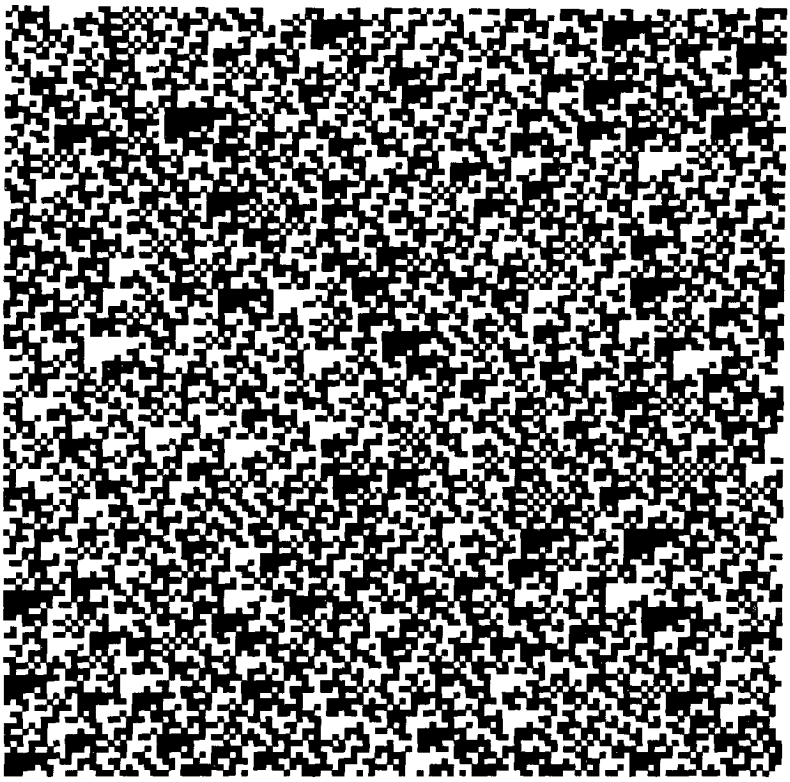

FIG. 5. An arbitrary part of the bit pattern generated by R(127; 112) with $s=\sigma(3)$. Each bit is the mod 2 sum of the bit above it and the nearest and the next-nearest neighbor to the right of that bit. Therefore, the value of s does not need to be known explicitly in order to construct the bit pattern. Notice the presence of white, black, and "checkerboard" triangles.

the relation $a_{i+\sigma(3)}=a_i \oplus a_{i+1} \oplus a_{i+2}$. In general, in patterns with $s=\sigma(3)$, every row of the triangles of zeros and of ones that appear is two bits longer than the row below; the total number of bits in a base-m triangle is $\frac{1}{4}(m+1)^2$ if m is odd, and $\frac{1}{4}m(m+2)$ if m is even. Per tile, the distribution is now

$$T(n, m)=3\times 2^{n-m-4} \qquad \text{for} \quad 0<m<n-3, \tag{11}$$

which holds both for triangles of zeros and for triangles of ones; in addition, there are two base-$(n-3)$ triangles of ones, one base-n triangle of ones, and three triangles of zeros with bases $n-3$, $n-2$, and $n-1$, respectively. The distribution follows trivially from the observation that the collection of triangles of ones and of zeros must accomodate precisely the collection of runs of ones and of zeros given by randomness property $R2$. The distribution found in Fig. 5, which samples only a small part of the complete tile for $R(127; 112)$, obeys Eq. (11), within normal

statistical fluctuations: this indicates again that properties valid for a complete tile are already observed in a small part of that tile. In Fig. 5, also "checkerboard" triangles appear. Their distribution is easily found: for $m < n-3$, the number of base-m checkerboard triangles is precisely twice as large as the number of base-m triangles of ones or zeros. In bit patterns produced with $s = \sigma(2)$ triangles of zeros are the main feature; their distribution again obeys Eq. (11).

Wolfram [17] studies bit patterns produced from a disordered first row of bits by a great variety of Boolean construction rules; his rules 90 and 150 happen to coincide with our procedures for $s = \sigma(2)$ and $s = \sigma(3)$, respectively. In both cases, he finds the numerical estimate $T'(m) \simeq c\lambda^{-m}$ with $\lambda \simeq 2$ for the density $T'(m)$ of base-m triangles of zeros. This agrees with the density $T'(m) = T(n, m)/N = c \cdot 2^{-m}$ obtained from Eq. (11); the proportionality constant is $c = 3 \cdot 2^{n-4}/N$, which for large n is almost equal to $\frac{3}{16}$. The agreement is due to the fact that unbiased disordered binary sequences must also obey randomness property $R2$, at least in average; or, the difference between the bit patterns obtained by Wolfram and by us

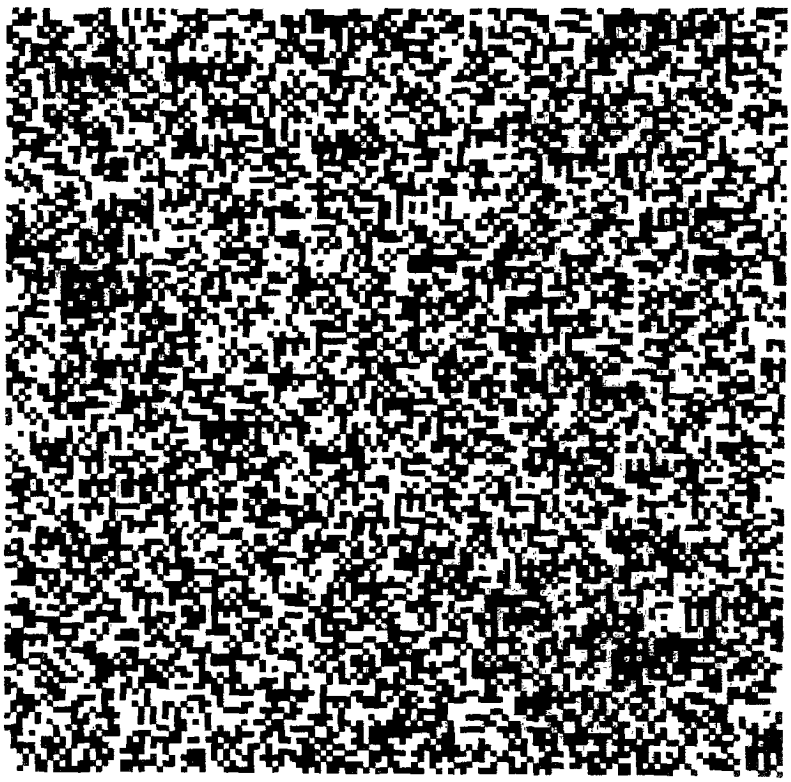

FIG. 6. An arbitrary part (containing 128×128 bits) of the bit pattern generated by $R(135, 124)$ with $s = \sigma(10)$. Bit i on row j is now the mod 2 sum of bits i, $i+2$, and $i+4$ on row $j-1$. Triangles of zeros and of ones are still present, as are "zebra" triangles, but the shape of these triangles is rather narrow now. Subsequent rows of the pattern belong to very different parts of the maximum-length sequence

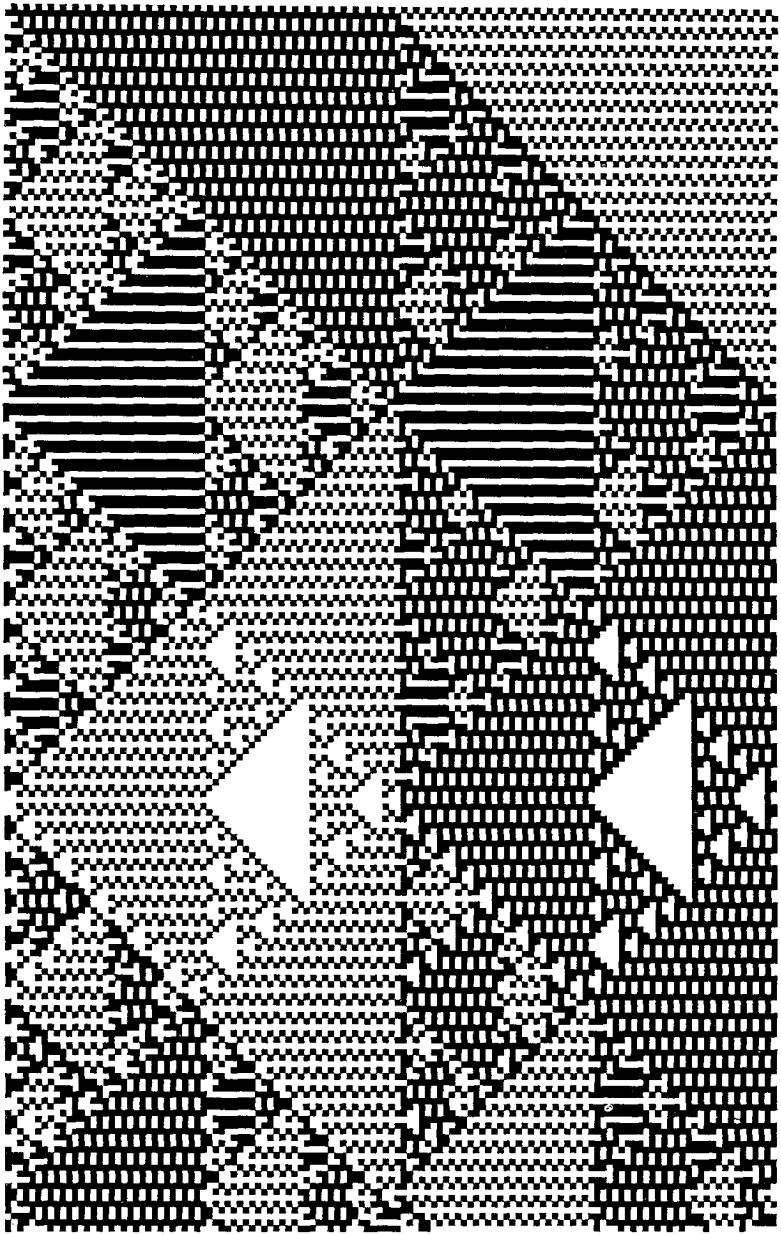

FIG. 7. An arbitrary part (127 bits wide, 200 bits long) of the bit pattern generated by $R(127; 64)$ with $s = n = 127$. Bit i on row j is the mod 2 sum of bits i and $i + 63$ on row $j - 1$. The interference between the initial condition (127 alternating bits 0101...) and the production rule leads to triangular superstructures that only decay slowly.

disappears when the shift register employed has more positions than can be accomodated within the linear dimensions of that part of the generated tile that is considered. It may be noted that maximum-length sequences are indeed a perfect example of the self-organizing and self-replicating automata considered by Wolfram.

In Fig. 6 we give a further example, now for $R(135; 124)$ with a shift $s = \sigma(d)$ pertaining to $d = 10$. For larger values of d, the triangular structures tend to become less conspicuous, but the underlying correlations are still present. Of course, when the maximum-length sequence is used to generate random numbers, correlations over distances $\sigma(d)$ are less harmful when $\sigma(d)$ is larger and/or when d indicates more complicated collections of bits. For many values of d the correlation distance $\sigma(d)$ may easily be larger than the length of the subsequence from which the random numbers are taken. Therefore, the bit patterns presented with $s = \sigma(d) \neq n$ exaggerate the importance of long-distance correlations. On the other hand, the production rule itself always gives rise to a correlation distance $\sigma(d) = n$ for a value of d which corresponds to the feedback positions. In particular, two-bit feedback production rules $R(n; k)$ with k close to n or 0 (these cases are conjugate) or close to simple fractions ($\frac{1}{2}, \frac{1}{3}, \ldots$) of n will not have satisfactory randomness properties. An example is shown in Fig. 7, which shows part of the bit pattern generated by $R(127; 64)$ with $s = \sigma(2^{61}) = n$. The triangular "superstructure" observed in that figure is caused by an interference between the intrinsic symmetry of the initial condition used (as seed the alternating sequence 0101... was taken) and the production rule; such interferences may result in a comparatively slow decay of the influence of the initial condition. For "random" seeds, the superstructure disappears and the decay to a uniform pattern of small triangular structures is much faster.

The bit patterns clearly demonstrate the existence of nonrandom correlations

sequences. How random are maximum-length sequences? A precise and rather satisfactory answer to this question can be found by considering all higher-order correlation functions.

4. The Complete Hierarchy of Correlation Functions

Consider an ensemble of binary sequences $\{b_t\} = b_1, b_2, \ldots, b_N$ with $b_t = \pm 1$. For the moment, these sequences need not be maximum-length sequences, but they are supposed to be periodic: $b_{t+N} = b_t$. The ensemble is defined by the probability distribution $p(\{b_t\}) \geq 0$ with

$$\sum_{\{b_t\}} p(\{b_t\}) = 1. \tag{12}$$

Let $\{k_j\} = k_1, k_2, ..., k_m$ be a subset of order m of the indices $i = 1, 2, ..., N$ (this is tantamount to a subset of the sequence) with

$$1 \leqslant k_1 < k_2 < \cdots < k_m \leqslant N. \tag{13}$$

This subset defines a product of elements from the sequence,

$$P(\{k_j\}; \{b_i\}) \equiv \prod_{j=1}^{m} b_{k_j}. \tag{14}$$

The expectation value of this product over the ensemble is

$$C_N(\{k_j\}) \equiv \sum_{\{b_i\}} P(\{k_j\}; \{b_i\})\, p(\{b_i\}). \tag{15}$$

Summing these expressions over all possible choices (including the empty set) for $\{k_j\}$ one finds

$$\sum_{\{k_j\}} P(\{k_j\}; \{b_i\}) = \prod_{i=1}^{N} (1 + b_i), \tag{16}$$

$$\sum_{\{k_j\}} C_N(\{k_j\}) = 2^N p(\{1\}), \tag{17}$$

where $p(\{1\})$ is the probability of the sequence with $b_i = 1$ for all i (for all other sequences, at least one of the factors $1 + b_i$ that occur in Eq. (16) is zero). We will assume that the ensemble is translationally invariant, such that

$$p(b_i, b_{i+1}, ..., b_{i+N-1}) = p(b_1, b_2, ..., b_N) \tag{18}$$

holds for $i = 1, 2, ..., N$. Then, $C_N(\{k_j\})$ depends only on the $m - 1$ differences $k_j - k_1$ with $j = 2, 3, ..., m$, and is identical with the usual mth-order correlation function.

If the sequences $\{b_i\}$ are the result of tossing an ideal coin N times (head $= +1$, tail $= -1$), all sequences are equally probable:

$$p(\{b_i\}) = \frac{1}{2^N}. \tag{19}$$

The correlation functions then obey

$$C_N(\{k_j\}) = \frac{1}{2^N} \left(\sum_{b_{k_1} = \pm 1} b_{k_1} \right) \sum_{b_{k_j} = \pm 1}^{(j \neq 1)} \prod_{j=2}^{m} b_{k_j} = 0, \tag{20}$$

except for the empty subset $\{k_j\} = 0$ with $m = 0$, for which $C_N(0) = P(0; \{b_i\}) = 1$ holds, in agreement with Eq. (17).

Equation (20) is not only a necessary condition for the validity of Eq. (19), it is

also sufficient. To show this, we repeat a derivation given by Glauber [15] for the case of the time-dependent 1-dimensional Ising model; if in this model we suppress the time-dependence by taking the temperature equal to zero, it becomes identical with our description of binary sequences. We start from a particular representation of Kronecker's δ-symbol in this case, i.e., from the relation

$$\tfrac{1}{2}(1+b_t b_t') = \begin{cases} 0 & \text{if} \quad b_t \neq b_t', \\ 1 & \text{if} \quad b_t = b_t', \end{cases} \tag{21}$$

which always holds for $b_t, b_t' = \pm 1$. We may therefore write

$$p(\{b_t\}) = \frac{1}{2^N} \sum_{\{b_t'\}} \left[\prod_{t=1}^{N} (1 + b_t b_t') \right] p(\{b_t'\}). \tag{22}$$

Expanding the product we get

$$p(\{b_t\}) = \frac{1}{2^N} \sum_{\{b_t'\}} \sum_{\{k_j\}} \left(\prod_{j=1}^{m} b_{k_j} b_{k_j}' \right) p(\{b_t'\}). \tag{23}$$

Use of Eqs. (14) and (15) leads to

$$p(\{b_t\}) = \frac{1}{2^N} \sum_{\{k_j\}} P(\{k_j\}; \{b_t\})\, C_N(\{k_j\}). \tag{24}$$

which equation is generally valid; it expresses the probabilities in terms of the correlation functions and thus is the inverse of Eq. (15).

For truly random sequences, all correlation functions except $C_N(0)$ must vanish identically; in the summation over $\{k_j\}$ in Eq. (24) only the empty set survives, and one is led back to the coin-tossing ensemble of Eq. (19). Therefore, Eqs. (19) and (20) are completely equivalent and provide a precise definition of randomness for binary sequences. Randomness can never be a property of a single sequence but can only be attributed to the coin-tossing ensemble.

We now concentrate on maximum-length sequences $\{b_t\}$, developed from a seed $b_1, b_2, \ldots, b_n = \pm 1$ (not all equal to 1) by means of a production rule given by Eq. (2). As before, we have $N = 2^n - 1$. The ensemble will now consist of all N sequences grown by the same production rule from the N different initial seeds that are possible; that is, the ensemble contains, in addition to a given maximum-length sequence, just all translated versions of that sequence. The probability distribution is

$$p(\{b_t\}) = \frac{1}{N} \sum_{j=1}^{N} \prod_{t=1}^{N} \delta_{b_{t+j}, b_t'}. \tag{25}$$

where $\{b'_i\}$ is a particular sequence of the ensemble, say the one developed from the seed $b'_i = -1$ for $i = 1, 2, \ldots, n$. The correlation functions are now given by

$$C_N(\{k_j\}) = \frac{1}{N} \sum_{i=1}^{N} \prod_{j=1}^{m} b_{k_j+i}, \tag{26}$$

where the prime is dropped again.

Because of the production rule of Eq. (2), any general element of the sequence can be written as a product of preceding elements of the sequence. By iterating the process, every element of the sequence can ultimately be expressed as a product of factors chosen from $b_1, b_2, \ldots, b_n$. When, during the process, the same element appears twice as a factor it disappears altogether ($b_i^2 = 1$). Different elements are expressed as different products, of which there are precisely $N = 2^n - 1$ (the empty product does not appear). We may write

$$b_i = \prod_{j=1}^{n} b_j^{\alpha_j}, \tag{27}$$

where the exponents α_j are either 0 or 1 (not all zero). The exponents α_j together form an n-bit binary number; there is a one-to-one correspondence between this number and i.

To give an example, we show in Fig. 8 a single tile of the bit pattern produced by $R(5; 3)$ with $s = \sigma(1) = 18$, in its general form; for the purpose of the figure we put $b_1 = a$, $b_2 = b$, $b_3 = c$, $b_4 = d$, and $b_5 = e$. Because $N = 31$ is prime, not only the horizontal but also every other rational direction in the bit pattern leads to a maximum-length sequence; for two of these directions the production rules are

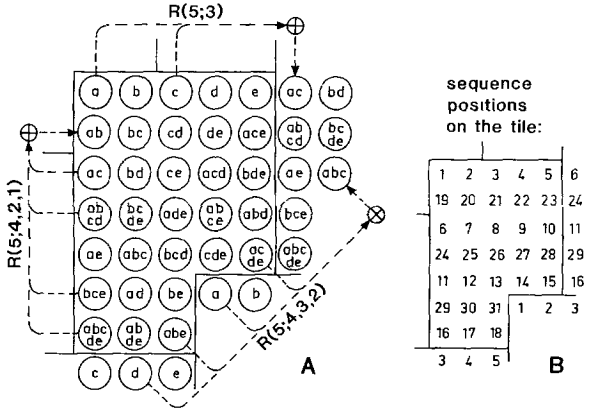

FIG. 8. A single tile of the bit pattern produced by $R(5; 3)$ with $s = \sigma(1) = 18$. The 5 seed bits $b_1, \ldots, b_5 = \pm 1$ are indicated by a, b, c, d, e. Part B of the figure indicates where the elements of the sequence are located on the tile. In this case, N is prime; therefore, any rational direction in the pattern corresponds with a maximum-length sequence. A few alternative production rules are indicated.

indicated. For clarity, part B of the figure indicates in what order the bits of the sequence cover the tile. In part A one immediately recognizes that due to the choice $s = \sigma(1)$, every element is the product of the elements "north" and "north-east" of it, taking into account that $b_l^2 = 1$. This, of course, is equivalent to the mod-2 summation in terms of the sequence $\{a_l\}$ with $a_l = 0, 1$. All different nonempty products of factors chosen from a, b, c, d, and e appear. It is evident how, in terms of $\{a_l\}$, triangles of zeros are produced in this bit pattern. It is also evident that many other combinations of bits can be selected which are highly correlated although this is not conspicious in the bit pattern. For instance, the product of $b_1 = a$, $b_7 = bd$, and $b_{28} = abd$ is always 1, but their geometrical relation in the bit pattern is weak

Consider now the product defined in Eq. (14). Because of Eq. (27) we have

$$P(\{k_j\}; \{b_l\}) = \prod_{l=1}^{n} b_l^{\beta_l}, \tag{28}$$

where the exponents β_l can be 0, 1, 2,..., but again because of $b_l^2 = 1$ we may assume that β_l is either 0 or 1. A subset $\{k_j\}$ for which $\beta_l = 0$ holds for all l will be called a dependent subset. For a dependent subset one has $P(\{k_j\}; \{b_l\}) = 1$. Because of the translational invariance of the production rule, a subset $\{k_j + i\}$ which is a translation over i positions of a dependent subset $\{k_j\}$ is again a dependent subset. This follows also from the fact that we did not specify the seed $b_1, b_2, \ldots, b_n$. As a consequence, the correlation function of a dependent subset is always equal to 1.

When one or several exponents β_l are equal to 1, the subset $\{k_j\}$ is called independent. Now the product of Eq. (28) can take the values ± 1. In fact, the product of Eq. (28) is equal to a particular element b_l of the sequence, according to Eq. (27); the value of i depends on which exponents β_l are equal to 1. This follows also directly from a repeated use of the cycle-and-add property. The correlation function pertaining to an independent subset is therefore the average value of b_l over the ensemble, i.e., the average value of the elements of the sequence. Surprisingly enough, we thus obtain the result

$$C_N(\{k_j\}) = \begin{cases} 1 & \text{if } \{k_j\} \text{ is dependent,} \\ -1/N & \text{if } \{k_j\} \text{ is independent.} \end{cases} \tag{29}$$

The correlation functions of third and higher order are two-valued, just like the pair correlation given by Eq. (4). However, in our present terminology the pair correlation is single-valued; the case $j = 0$ in Eq. (4) refers to the empty set rather than to a subset of size 2. There are no dependent subsets of size 1 or 2. The result of Eq. (29) was obtained before by Fredricsson [11] in a more formal manner.

A natural question to ask now is, how many subsets $\{k_j\}$ are dependent. We define $A_N(m)$ as the number of independent subsets $\{k_j\}$ of order m that can be taken from the elements of the maximum-length sequence and $B_N(m)$ as the number of dependent subsets of order m. One easily verifies that the following relations hold:

$$A_N(m) + B_N(m) = \binom{N}{m},$$

$$\begin{aligned} A_N(m) &= A_N(N-m), & B_N(m) &= B_N(N-m), \\ A_N(0) &= A_N(N) = 0, & B_N(0) &= B_N(N) = 1, \\ A_N(1) &= A_N(N-1) = N, & B_N(1) &= B_N(N-1) = 0. \end{aligned} \tag{30}$$

these subsets consist of a dependent subset of order $m-1$ plus one extra element. This number is the product of $B_N(m-1)$, the number of dependent subsets of order $m-1$, and the number of different choices, $N-m+1$, that one can make for the additional element. Therefore, the number of independent subsets of order m that do not contain a dependent subset of order $m-1$ is $A_N(m) - (N-m+1)\, B_N(m-1)$. Given such a "doubly independent" subset of order m, one can always find precisely one element of the sequence which makes the doubly independent subset of order m into a dependent subset of order $m+1$. This element is not already contained in the subset of order m, since a dependent subset of order $m+1$ that would contain two identical elements would necessarily contain a dependent subset of order $m-1$, contrary to the fact that the subset of order m is doubly independent. Finally, every dependent subset of order $m+1$ can be formed in $m+1$ ways from an independent subset of order m. Thus we find

$$B_N(m+1) = \frac{1}{m+1}\,[A_N(m) - (N-m+1)\, B_N(m-1)]. \tag{31}$$

Using the first of Eqs. (30) we get the recurrence relation

$$\begin{aligned} &(m+1)\, A_N(m+1) + A_N(m) + (N-m+1)\, A_N(m-1) \\ &\quad = (m+1)\binom{N}{m+1} + (N-m+1)\binom{N}{m-1}. \end{aligned} \tag{32}$$

This equation can be solved by means of the generating function

$$G(x) \equiv \sum_{n=0}^{N} A_N(m)\, x^m. \tag{33}$$

Multiplying Eq. (32) with x^m, summing over m, and taking some extra care at the boundaries of this summation, one finds that $G(x)$ must obey the following first-order linear differential equation:

$$(1-x^2)\frac{d}{dx}G(x) + (1+Nx)\, G(x) = N(1+x)^N. \tag{34}$$

The solution is

$$G(x)=\frac{N}{N+1}(1+x)^{(1/2)(N-1)}[(1+x)^{(1/2)(N+1)}-(1-x)^{(1/2)(N+1)}], \tag{35}$$

in which a constant of integration was determined from $G(0)=A_N(0)=0$. Using a binomial expansion for the difference in square brackets (odd powers of x vanish) and multiplying with the expansion for $(1+x)^{(1/2)(N-1)}$ one obtains

$$A_N(m)=\frac{2N}{N+1}\sum_{j=1}^{(1/4)(N+1)}\binom{\frac{1}{2}(N+1)}{2j-1}\binom{\frac{1}{2}(N-1)}{m+1-2j}. \tag{36}$$

Rewriting Eq. (35) in the form

$$G(x)=\frac{N}{N+1}[(1+x)^N-(1-x)(1-x^2)^{(1/2)(N-1)}], \tag{37}$$

and using the binomial expansions for the two terms one obtains

$$\begin{aligned}
A_N(4j)&=\frac{N}{N+1}\left[\binom{N}{4j}-\binom{\frac{1}{2}(N-1)}{2j}\right],\\
A_N(4j+1)&=\frac{N}{N+1}\left[\binom{N}{4j+1}+\binom{\frac{1}{2}(N-1)}{2j}\right],\\
A_N(4j+2)&=\frac{N}{N+1}\left[\binom{N}{4j+2}+\binom{\frac{1}{2}(N-1)}{2j+1}\right],\\
A_N(4j+3)&=\frac{N}{N+1}\left[\binom{N}{4j+3}-\binom{\frac{1}{2}(N-1)}{2j+1}\right],
\end{aligned} \tag{38}$$

for $j=0, 1, 2,\ldots, \frac{1}{4}(N-3)$. These expressions are equivalent with Eq. (36), which is more compact but calculationally less convenient. Closed expressions for $B_N(m)$ follow straightforwardly:

$$\frac{N+1}{N}B_N(m)=\frac{1}{N}\binom{N}{m}+\begin{cases}(-1)^{(1/2)m}\binom{\frac{1}{2}(N-1)}{\frac{1}{2}m}, & m \text{ even},\\ (-1)^{(1/2)(m+1)}\binom{\frac{1}{2}(N-1)}{\frac{1}{2}(m-1)}, & m \text{ odd}.\end{cases} \tag{39}$$

Both $A_N(m)$ and $B_N(m)$ contain a certain amount of redundancy, because translated subsets are counted separately although they lead to identical correlation functions. When N is prime, a subset of order m and its translations form a group of N elements, and the number of different mth-order correlation functions that take the value $-1/N$ is $A_N(m)/N$, whereas the number of those that take the value 1 is $B_N(m)/N$; the only exceptions are $m=0$ and $m=N$. This is an application

of Fermat's so-called little theorem. When N is not prime, there are for any m that divides N always certain symmetric subsets of order m which are invariant under translations over only N/m positions. The number of different correlation functions could then be counted by means of Polya's enumeration theorem (see, e.g., Harary [21]), but this counting problem is already quite complicated if m

The total numbers of independent and dependent subsets of any order are given by

$$\begin{aligned} \sum_{m=0}^{N} A_N(m) &= G(1) = \frac{N}{N+1} 2^N, \\ \sum_{m=0}^{N} B_N(m) &= 2^N - G(1) = \frac{1}{N+1} 2^N. \end{aligned} \tag{40}$$

For large values of N, the second terms on the right-hand side in Eqs. (38) can be neglected; one has

$$A_N(m) \simeq \frac{N}{N+1} \binom{N}{m}, \qquad B_N(m) \simeq \frac{1}{N+1} \binom{N}{m}. \tag{41}$$

The total number of independent subsets is therefore always exactly N times larger than the total number of dependent subsets; asymptotically, this ratio also holds for every value of m separately.

For the sum of all correlation functions we find

$$\sum_{\{k_j\}} C_N(\{k_j\}) = \sum_{m=0}^{N} \left[A_N(m) \cdot \left(-\frac{1}{N} \right) + B_N(m) \cdot 1 \right] = 0, \tag{42}$$

in agreement with Eq. (17), because $p(\{1\}) = 0$ holds for maximum-length sequences. The average absolute value

$$\Gamma \equiv \frac{1}{2^N} \sum_{\{k_j\}} |C_N(\{k_j\})| \tag{43}$$

of all correlation functions can be used as a measure for the total amount of nonrandomness. For a completely ordered sequence (either $b_i = 1$ or $b_i = -1$ for all i) one has $\Gamma = 1$; in fact, $\Gamma = 1$ holds for all ensembles that contain a single binary sequence only. For an ensemble consisting of the two sequences of alternating $+1$'s and -1's one finds $\Gamma = \frac{1}{2}$. For a truly random sequence (described by the coin-tossing ensemble) one has $\Gamma = 2^{-N}$, a small but nonzero value which is solely due to the contribution of the trivial empty set $\{k_j\} = 0$. For the translationally invariant

ensemble consisting of a maximum-length sequence and all its translated versions one finds

$$\Gamma = \frac{1}{2^N} \sum_{m=0}^{N} \left[A_N(m) \cdot \frac{1}{N} + B_N(m) \right] = \frac{2}{N+1} = 2^{1-n}, \tag{44}$$

which is already quite small for maximum-length sequences produced by shift registers of moderate length; however, this should not be taken to imply that such sequences are already sufficiently random. The fact that according to Eq. (44) dependent and independent subsets contribute equally to Γ is somewhat misleading also. In fact, the dependent subsets are much more detrimental to the randomness properties of the sequence than the independent ones.

The quantity Γ can be compared with the more conventional measure for the amount of randomness, the entropy

$$S \equiv -\sum_{\{b_i\}} p(\{b_i\}) \ln p(\{b_i\}). \tag{45}$$

One easily finds that the relation $S = -\ln \Gamma$ holds for an ensemble consisting of a single binary sequence, for the ensemble consisting of the two alternating sequences, and for the coin-tossing ensemble. For the ensemble consisting of a maximum-length sequence and all its translated versions one finds $S = \ln N = \ln(2^n - 1)$, which differs slightly from $-\ln \Gamma$ given by Eq. (44). Neither Γ nor S does justice to the details of maximum-length sequences and their randomness properties.

Nevertheless, Eqs. (38)–(44) provide part of an answer to the question posed at the end of Section 3. In terms of the hierarchy of correlation functions, all maximum-length sequences with the same period N are equally random. The well-known and rather trivial statement that longer maximum-length sequences are more random is now seen to reduce to two precise properties: (1) all correlation functions of independent subsets have $-1/N$ as value; (2) all correlation functions of dependent subsets are equal to 1, but their relative number $\sum B_N(m)/2^n$ vanishes as $1/N$ (asymptotically, this holds for each subset order m separately).

5. Subsequence Averages

All maximum-length sequences of the same period may be equally random, but some are more random than others when only part of the sequence is considered. To explain this we introduce the partial correlation functions

$$C_M(\{k_j\}) = \frac{1}{M} \sum_{i=0}^{M-1} \prod_{j=1}^{m} b_{k_j+i} \tag{46}$$

for $M \leqslant N$. There is no restriction on the subset $\{k_j\}$, though of course those with a small distance between k_1 and k_m, the first and the last element of the subset, are

most relevant, in particular when m is also small; correlations between the elements of small-distance low-order subsets can be most detrimental to the randomness properties.

For dependent subsets $\{k_j\}$ one has immediately $C_M(\{k_j\}) = 1$. We have seen that the total number $B_N(m)$ of dependent subsets of order m is relatively small compared with $\binom{N}{m}$, but is the same true for the number $B_M(m)$ of dependent subsets of order m taken from a subsequence of length M in comparison with $\binom{M}{m}$? Because of translational invariance, $B_M(m)$ depends on the length of the subsequence, but not on its location in the full sequence; in contrast with $B_N(m)$, it does depend on the details of the production rule.

The general structure of $B_M(m)$ is best described by giving a simple example: in Table I its nonzero values obtained by numerical examination for the case of production rule $R(4; 3)$ are shown ($B_M(0) = 1$ and $B_{15}(15) = 1$ were omitted). The last row in this table obeys Eq. (39). The following simple relations, valid for all production rules, are easily proven:

$$\begin{aligned} &B_M(0) = 1 && \text{for all } M, \\ &B_M(1) = B_M(2) = 0 && \text{for all } M, \\ &B_M(m) = 0 && \text{for } 0 \leqslant M \leqslant n \text{ and } m \neq 0, \\ &B_M(m) = 0 && \text{for } m \geqslant M \quad \text{and } M \neq 0, N. \end{aligned} \tag{47}$$

For instance, $B_M(M) = 1$ would imply that any M subsequent elements from a dependent subset; a shift of this subset over one position yields $b_{M+i} = b_i$ for all i, which means that either $M = 0$ or $M = N$ holds.

TABLE I

The Nonzero Values of $B_M(m)$ for Production Rule $R(4; 3)$

M \ m	3	4	5	6	7	8	9	10	11	12
5	1									
6	2	1						In addition,		
7	3	2	1	1				$B_M(0) = 1$ and		
8	4	5	4	2				$B_{15}(15) = 1$		
9	6	10	8	4	2	1				
10	9	16	15	12	7	3	1			
11	13	25	25	27	23	10	3	1		
12	17	38	44	52	54	33	12	4	1	
13	22	55	72	96	116	87	40	16	6	1
14	28	77	112	168	232	203	112	56	28	7
15	35	105	168	280	435	435	280	168	105	35

Inspection of the last rows of Table I (and of a similar table for production rule $R(5;3)$) suggest that the following recursion relation holds for $M \geqslant N-n+1$:

$$B_M(m) = B_{M-1}(m-1) + B_{M-1}(m) - \begin{cases} 0 & \text{for even } m. \\ (-1)^{(1/2)(m-1)} \binom{M-2^{n-1}}{\frac{1}{2}(m-1)} & \text{for odd } m. \end{cases} \tag{48}$$

A general proof of this relation can be found starting from the observation that the complement of a dependent subset is itself a dependent subset; however, the proof is rather lengthy and will be omitted here. Using the generating function

$$H_M(x) \equiv \sum_{m=0}^{M} B_M(m)\, x^m \tag{49}$$

one may find from the recursion relation the following expression, valid for $0 \leqslant i \leqslant n$:

$$H_{N-i}(x) = \frac{(1+x)^{(1/2)(N-1)-i}}{N+1} \left[(1+x)^{(1/2)(N+1)} - (1-x)^{(1/2)(N+1)}\right] + (1-x)(1-x^2)^{(1/2)(N-1)-i}. \tag{50}$$

where Eqs. (30), (33), and (37) were employed to obtain $H_N(x)$. For large N, the leading term is

$$H_{N-i}(x) \approx \frac{1}{N+1}(1+x)^{N-i}. \tag{51}$$

For $n < M < N-n$ the numbers $B_M(m)$ depend on the production rule. The first nonzero value with $m \neq 0$ appears when $M = n+1$. Then, b_1 and the elements indicated by the additional feedback positions of the production rule $R(n; \{j\})$ from together with element b_{n+1} of the sequence the first nonempty dependent subset. If the production rule has k additional feedback positions, one has

$$B_{n+1}(k+2) = 1. \tag{52}$$

In general, for $M > n$ any of the 2^{M-n} subsets taken from elements $b_{n+1}, b_{n+2}, \ldots, b_M$ will form a dependent subset when combined with an appropriate, possibly empty, selection taken from the first n elements of sequence. This leads to

$$\sum_{m=0}^{M} B_M(m) = 2^{M-n} \qquad \text{for} \qquad M \geqslant n. \tag{53}$$

Consider now a collection of subsets of size m, taken from a subsequence of

length M and consisting of at least one dependent subset together with all its translated versions that fit into the subsequence. When M is increased by 1, at least one other translated version becomes available; in addition, an altogether new dependent subset (containing b_1 and b_{M+1} together with other elements) may appear. Therefore, if $B_M(m)$ for $m \neq 0$ differs from zero it increases at least linearly with M:

$$B_{M+1}(m) \geqslant B_M(m) + 1. \tag{54}$$

At which values of m the new dependent subsets appear is completely determined by the production rule, but in a rather haphazard manner: in this sense, the non-random contributions are still random. For $M \gg n$, this probabilistic argument leads to the conclusion that $B_M(m)$ can be approximated by the binomial distribution

$$B'_M(m) \equiv \frac{1}{N+1}\binom{M}{m} + \delta_{m,0}. \tag{55}$$

Here, the Kronecker δ takes the empty-set contribution into account; apart from this term, Eq. (55) agrees with Eqs. (51) and (53). It also agrees semi-quantitatively with the numbers in Table I and in similar tables for larger n. However, especially for small values of m, relatively small deviations between $B_M(m)$ and $B'_M(m)$ may be important, as we will see.

From Eq. (53) it follows that, when M is larger than n, the total number of dependent subsets is always a factor $1/(N+1)$ less than the total number of independent subsets. It is also very likely that this holds for each subset order m separately, except perhaps when M is only slightly larger than n or when m is small.

Let us now turn to the independent subsets. For those, $C_M(\{k_j\})$ differs for different subsequences, but the product in Eqs. (46) is always equal to an element b_{k+l} of the sequence at a given distance $k - k_1$ from the first index appearing in the product. The partial correlation functions for independent subsets are therefore equal to the average bit value of the subsequence of M bits starting at b_k:

$$C_M(\{k_j\}) = \frac{1}{M}\sum_{l=0}^{M-1} b_{k+l}. \tag{56}$$

Usually, the value of k is unknown; it may be very large, such that the subsequence starting at b_k is far away from the original subsequence, the correlations of which are studied. Translationally different independent subsets only lead to different values of k, i.e., to the same value of $k - k_1$. Two questions arise: how is the right-hand side of Eq. (56) distributed for different values of k, and how does the distribution behave for increasing values of M? The answers will be found to depend again on $B_M(m)$.

TABLE II

The Distribution $Nf_M^{(1)}(\Delta)$ for Production Rule $R(5;3)$

Δ \ M	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
4			1	3	5	5	6	5	3	2	1	1		1		
2		7	8	6	4	5	4	5	9	9	10	8	9	5	4	
0	31	16	12	10	7	6	4	6	5	8	8	10	9	10	12	16
−2		8	8	9	12	10	12	9	6	4	4	5	6	10	12	15
−4			2	3	3	5	5	5	7	6	6	5	6	5	3	
−6								1	1	2	2	2	1			

Note. For odd values of M the distribution follows from Eq. (50).

By means of the formal expansion

$$\Delta^l = \left(\sum_{i=1}^{M} b_i \right)^l \equiv \sum_{m=0}^{l} F(m, l, M) \sum_{\{k_j\}} b_{k_1} \cdots b_{k_m}, \tag{64}$$

coefficients $F(m, l, M)$ were defined by Lindholm [7], who also calculated their first few nonzero values. The sum over $\{k_j\}$ is to be taken over all subsets of order m. Comparing this equation with the equivalent relation for Δ^{l+1} and noting that each term b_i of the extra factor $\sum b_i$ either does or does not annihilate one of the factors selected by the subset $\{k_j\}$, one finds the recurrence relation

$$F(m, l+1, M) = mF(m-1, l, M) + (M-m)\, F(m+1, l, M). \tag{65}$$

This must be solved for $l \geqslant 0$, with the initial condition $F(m, 0, M) = \delta_{m,0}$. Although Eq. (64) defines the Lindholm coefficients (as they will be called here) only in the range $0 \leqslant m \leqslant \min(l, M)$, the recurrence relation (65) also generates coefficients outside that range. However, coefficients with $m < 0$ or $m > M$ never influence coefficients in the relevant range because of the factors m and $M-m$ in Eq. (65). To get rid of the irrelevant coefficients outside the range $0 \leqslant m \leqslant M$ we rewrite the recurrence relation in the form

$$\binom{M}{m} F(m, l+1, M)$$

$$= (m+1)\binom{M}{m+1} F(m+1, l, M) + (M-m+1)\binom{M}{m-1} F(m-1, l, M). \tag{66}$$

This recurrence relation for the quantity $\binom{M}{m} F(m, l, M)$ turns out to be equivalent with the recurrence relation for the Ehrenfest urn model, which was solved explicitly by Kac [22]. Adapting that solution to the present case and using the

initial condition, one finds the following generating function for the Lindholm coefficients:

$$\sum_{m=0}^{l} F(m, l, M)\binom{M}{m} z^m = \frac{1}{2^M}\sum_{k=0}^{M}(M-2k)^l\binom{M}{k}(1-z)^k(1+z)^{M-k}. \tag{67}$$

Equating like powers of z one obtains after some manipulations

$$F(m, l, M) = \frac{1}{2^M}\sum_{i=0}^{m}(-1)^i\binom{m}{i}\sum_{j=0}^{M-m}(M-2i-2j)^l\binom{M-m}{j}. \tag{68}$$

Multiplication with $y^l/l!$ and summation over l yields yet another generating function:

$$\sum_{l=0}^{\infty} F(m, l, M)\frac{y^l}{l!} = (\cosh y)^{M-m}(\sinh y)^m. \tag{69}$$

It is, however, a sobering thought to realize that explicit expressions for the Lindholm coefficients are most easily obtained from Eq. (65), starting from $F(m, 0, M) = \delta_{m,0}$. The nonzero Lindholm coefficients with $0 \leqslant m \leqslant l \leqslant 7$ are:

$$\begin{aligned}
&F(0, 0, M) = 1, && F(1, 1, M) = 1,\\
&F(0, 2, M) = M, && F(2, 2, M) = 2,\\
&F(1, 3, M) = 3M - 2, && F(3, 3, M) = 3!,\\
&F(0, 4, M) = 3M^2 - 2M, && F(2, 4, M) = 12M - 16, \quad F(4, 4, M) = 4!,\\
&F(1, 5, M) = 15M^2 - 30M + 16, && F(3, 5, M) = 60M - 120, \quad F(5, 5, M) = 5!,\\
&F(0, 6, M) = 15M^3 - 30M^2 + 16M, && F(2, 6, M) = 90M^2 - 300M + 272,\\
&F(4, 6, M) = 360M - 960, && F(6, 6, M) = 6!,\\
&F(1, 7, M) = 105M^3 - 420M^2 + 588M - 272, && F(3, 7, M) = 630M^2 - 2940M + 3696,\\
&F(5, 7, M) = 2520M - 8400, && F(7, 7, M) = 7!
\end{aligned} \tag{70}$$

There is no restriction on the integers M, although of course only the coefficients for which $0 \leqslant m \leqslant \min(l, M)$ holds occur in Eq. (64) and hence contribute to the lth moment.

Returning now to the moments $\mu_l^{(q)}(M)$ we write, averaging over the complete sequence of N bits,

$$\begin{aligned}
\mu_l^{(q)}(M) &= \sum_{\Delta=-M}^{M}\Delta^l f_M^{(q)}(\Delta) = \frac{1}{N}\sum_{k=1}^{N}\left(\sum_{i=0}^{M-1} b_{k+i}\right)^l\\
&= \frac{1}{N}\sum_{m=0}^{l} F(m, l, M)\sum_{k=1}^{N}\sum_{\{k_j\}} b_{k+k_1}\cdots b_{k+k_m}.
\end{aligned} \tag{71}$$

The index k just numbers the N different M-bit sequences. For $q=2$, the average of the products of bits in the subset $\{k_j\}$ always vanishes, apart from the empty subset with $m=0$. That is, the lth moment of the coin-tossing ensemble is

$$\mu_l^{(2)}(M)=F(0, l, M). \tag{72}$$

Indeed, Eq. (68) for $m=0$ just gives the moments of the binomial distribution.

For $q=1$, we consider the numbers $A_M(m)$ and $B_M(m)$ of independent respectively dependent subsets $\{k_j\}$ or order m that can be taken from a subsequence of length M. These numbers obey

$$A_M(m)+B_M(m)=\binom{M}{m}. \tag{73}$$

Since the product of bits selected by an independent subset when summed over all translated versions of that subset is -1, whereas the product of bits selected by a dependent subset is 1 for each translated version, we may write

$$\sum_{k=1}^{N}\sum_{\{k_j\}} b_{k+k_1}\cdots b_{k+k_m}=-A_M(m)+NB_M(m). \tag{74}$$

Thus we obtain

$$\mu_l^{(1)}(M)=\frac{1}{N}\sum_{m=0}^{l} F(m, l, M)\left[(N+1)\,B_M(m)-\binom{M}{m}\right], \tag{75}$$

an equation first derived by Lindholm [7].

When instead of $B_M(m)$ the approximation $B'_M(m)$ of Eq. (55) is inserted into Eq. (74) the result is

$$\mu_l^{(1)}(M)\approx\frac{N+1}{N}F(0, l, M)=\frac{(N+1)}{N}\mu_l^{(2)}(M). \tag{76}$$

Unfortunately, this is too good to be true. For instance, for $N-n\leqslant M\leqslant N$ the distribution $f_M^{(1)}(\Delta)$ is indeed binomial (Eq. (61)), but equal to $f_{N-M}^{(2)}(\Delta)$ instead of $f_M^{(2)}(\Delta)$. The reason is that the Lindholm coefficients $F(m, l, M)$ are large for large M, such that relatively small deviations from zero of the quantity in square brackets in Eq. (75) cannot be neglected.

Exact expressions for the moments are readily obtained from Eqs. (47), (70), and (75), in combination with the fact that the quantity expressed by Eq. (67) is equal to M^l for $z=1$. The results are

$$\mu_0^{(1)}(M)=1, \qquad \mu_1^{(1)}(M)=-\frac{M}{N}, \qquad \mu_2^{(1)}(M)=M-\frac{M(M-1)}{N},$$

$$\mu_3^{(1)}(M)=-\frac{M^3}{N}+6\frac{N+1}{N}B_M(3),$$

$$\mu_4^{(1)}(M)=3M^2-2M-\frac{1}{N}M(M-1)^2(M+2)+24\frac{N+1}{N}B_M(4), \tag{77}$$

$$\mu_5^{(1)}(M)=-\frac{M^5}{N}+60\frac{N+1}{N}\{(M-2)B_M(3)+2B_M(5)\},$$

and so on. Most of these results were already obtained by Lindholm [7], but, in the derivation, a certain amount of both simplification and generalization has been achieved. For comparison, the even binomial moments are

$$\mu_0^{(2)}(M)=1, \qquad \mu_2^{(2)}(M)=M, \qquad \mu_4^{(2)}(M)=3M^2-2M, \tag{78}$$

whereas the odd ones vanish. For $M \ll N$, the first three moments (with $l=0, 1, 2$) of $f_M^{(1)}(\Delta)$ are almost equal to those of the binomial distribution; the condition $M \ll N$ is easily obeyed in practice. For the higher moments, the comparison can best be made in terms of the skewness γ_1 and the excess γ_2, given by

$$\gamma_1^{(1)} \equiv \frac{\mu_3^{(1)}}{(\mu_2^{(1)})^{3/2}}, \qquad \gamma_2^{(1)} \equiv \frac{\mu_4^{(1)}}{(\mu_2^{(1)})^2}-3, \tag{79}$$

with $\gamma_1^{(2)}=0$ and $\gamma_2^{(2)}=-2/M$ for the corresponding binomial quantities. For $M \ll N$ one has

$$\gamma_1^{(1)} \approx \frac{6B_M(3)}{M^{3/2}}, \qquad \gamma_2^{(1)} \approx -\frac{2}{M}+\frac{24B_M(4)}{M^2}, \tag{80}$$

with $B_M(3)=B_M(4)=0$ when $M \leqslant n$.

6. Discussion and Application

Lindholm [7] and Fredricsson [11] studied the distribution $f_M^{(1)}(\Delta)$ for shift registers with $n=23$, and $n=17$, respectively, for a variety of production rules and with M up to ~ 500. Only for some of these production rules they found a satisfactory behavior. Their results suggest that production rules with more than two feedback bits, at irregular positions, are better.

Production rules of the type $R(n; k)$ lead to $B_M(3)=1$ already for $M=n+1$ and to $B_M(3) \geqslant M-n$ for $M>n+1$, according to Eq. (54). Hence, undesirable contributions to the third moment occur already for relatively small values of M. Par-

ticularly awkward results can be expected when k is small or close to n (the [illegible] tion rules $R(n,k)$ and $R(n,n-k)$ are conjugate). The case considered by Knuth [23] as a warning against the use of maximum-length sequences for random-number generation is an example (with $n=35$ and $k=2$). For k or $n-k$ small also $B_M(4)$ will already differ from zero when M is only slightly larger than n. In terms of the shift s and the correlation distance $\sigma(d)$ used in Section 3 to discuss the bit patterns generated from maximum-length sequences, the cases $k=1$ and $k=2$ lead to $s=\sigma(1)=n$ and $s=\sigma(2)=n$, respectively. In such cases, subsequent parts of n bits from the sequence scan through adjacent rows of the triangular bit structures observed (see the figures in Section 3): when in terms of the sequence $\{a_i=0,1\}$ an excess of zeros is observed in a row of n bits, the next n bits will most likely show again an excess of zeros. This results in a certain skewness of the distribution $f_M^{(1)}(A)$. When neither k nor $n-k$ are small, adjacent rows of the triangular bit structures that result from choosing $s=\sigma(1)$ or $s=\sigma(2)$ belong to widely different parts of the maximum-length sequence, s now being very different from n. Thus a much better mixing of highly correlated parts over the whole sequence and a greater amount of uniformity are achieved.

Extending these arguments, one may also show that n should not be close to a multiple of k as, e.g., in $R(127;63)$, which generates a maximum-length sequence but which also generates quite remarkable triangular bit structures for $s=n$ as shown in Fig. 7. Recipes like k not small, nor close to n, $n/2$, $n/3$,..., etc. have been discussed in the literature (see, e.g., Refs. [8, 10, 13]); they can be understood in terms of the bit patterns and higher-order correlations studied in the present paper.

Production rules with a larger number of feedback positions tend to give rise to improved randomness properties, at least when these positions are sufficiently irregular to exclude possible mutual interferences between their effects. With a larger number of feedback positions the first value of M for which, say, $B_M(3)$ and $B_M(4)$ differ from zero, can be large compared with n. The scattering of correlated parts of the sequence that would result from an increased number of feedback positions could indeed be a very efficient means to improve uniformity.

Multiple-bit feedback production rules are not known for larger values of n, but can be found in principle from two-bit feedback production rules in the following manner, based on properties of maximum-length sequences described by Golomb [3] and Hoffmann de Visme [4]. Restricting the attention to the case where $N=2^n-1$ is prime, we note that decimating a maximum-length sequence $\{a_i\}$, i.e., replacing it by $\{a_{iq}\}$ with q an arbitrary integer, leads to a new sequence which is again of maximum length. When q is a power of 2, the new sequence is just a translated version of the old one, but for all other values of q the new sequence is really different and governed by a different production rule, which can be found by solving a set of n linear equations. An example of the situation can be found in Fig. 8. Since the number of different r-bit feedback production rules that generate maximum-length sequences with a Mersenne prime 2^n-1 as period is equal to $(2/n)\cdot\binom{n}{r}$, where r ($\neq 0$) is even, the new production rule is likely to have a large number of feedback positions. Thus, the two-bit feedback production rules given by

Zierler [19] for cases where $2^n - 1$ is prime can in principle be transformed into multiple-bit feedback production rules.

In our opinion, future research in the generation of random numbers by means of maximum-length sequences should be mainly directed towards these multiple-bit feedback production rules. However, at present they have two disadvantages: these rules are not readily available, and, when known, they are more cumbersome in

A straightforward manner to improve the randomness properties of maximum-length sequences is to increase the shift register length n. This trivial statement has been sharpened in this paper by showing that the relative number of correlation functions of any order that show nonrandom behavior decreases as 2^{-n}. The same holds true for the partial correlation functions of dependent subsets. Although usually only shift registers with $n \lesssim 100$ are considered for the generation of random numbers, two-bit feedback production rules for maximum-length sequences are known up to $n = 9689$ (see Zierler [19], from whose list also the examples to follow are drawn).

What can be expected for the behavior of the quantities expressed in Eqs. (77) and (80) when production rules such as $R(607; 147)$, $R(1279; 261)$, $R(4423; 271)$, or $R(9689; 471)$, to mention just a few, are used? Indeed, they all have $B_{n+1}(3) = 1$ and therefore $B_M(3) \geqslant M - n$, according to Eq. (54). Only when more than two feedback bits are involved can the first undesirable contribution to the third moment be postponed until larger M. However, $B_{n+1}(3) = 1$ does not need to be really detrimental, due to the denominator in Eq. (80) for the skewness.

By studying the first part of a general maximum-length sequence generated by a two-bit feedback production rule one may verify that, after the appearance of the first dependent triple (b_1, b_k, b_{n+1}) at $M = n + 1$, only the translated versions of this triple are found until $M = 2n + 1$; then, a new dependent triple occurs, which in turn will have its own translated versions when M increased further. If this behavior would continue, the following expression would hold:

$$B_M(3) = \sum_{i=1}^{\text{ent}[M/n]} (M - in). \tag{81}$$

In fact, the increase of $B_M(3)$ with M will at least initially and in average be less than indicated by Eq. (81). Ultimately, according to Eq. (55), $B_M(3)$ should behave as $M^3/6(N+1)$ which is much smaller than the right-hand side of Eq. (81) until M reaches a value $M^* \approx N/n$. Up to that value, Eq. (81) takes more dependent triples into account than correspond with an appropriate share of the total number of dependent subsets given by Eq. (53). For $M < M^*$ it seems that Eq. (81) provides an upper bound, which can be replaced by its asymptotic behavior:

$$B_M(3) \lesssim \tfrac{1}{2}M^2/n. \tag{82}$$

To pursue this argument further we need to go into still more detail. We first

refer to Eq. (27) and note that each element of a particular maximum-length sequence can be represented in a general form (i.e., without specifying its value) by the n-bit binary number formed by the exponents α_j. Element b_i of the seed ($1 \leqslant i \leqslant n$) has a one at position i and zeros elsewhere. The binary representation of subsequent elements of the sequence is found by adding without carry the binary representations of the elements involved in the production rule; no confusion will arise when we use the notation $b_i \oplus b_j$ to indicate this operation.

Consider now the probability p_j that a new dependent triple can be formed when the element b_j is added to the preceding elements of the sequence. This new dependent triple must contain b_1 and b_j in order to exclude translated versions of new dependent triples found in shorter parts of the sequence; of course, the third element of the triple, b_k, must lie in between $1 < k < j$. To see whether with b_j a new dependent triple is formed one only has to compare the binary representation of $b_1 \oplus b_j$ with that of each of the elements in between: $b_1 \oplus b_k \oplus b_j = 0$ must hold. Using a computer one can in principle find the set $\{J\}$ of indices $\leqslant M$ at which new dependent triples occur; for two-bit feedback production rules $J' = n + 1$ and $J'' = 2n + 1$ are the first elements of $\{J\}$. The probability p_j can now be written as

$$p_j = \sum_{J' \in \{J\}} \delta_{j,J'}. \tag{83}$$

When the sequence proceeds further up to b_M, each new dependent triple found at $j < M$ produces more and more translated versions; thus we get

$$B_M(3) = \sum_{j=1}^{M} (M - j + 1)\, p_j. \tag{84}$$

The density of new dependent triples is defined by

$$D_L(M) \equiv \sum_{j=M-L+1}^{M} p_j. \tag{85}$$

The algorithm sketched above consists of a number of operations that is proportional to M^2, and hence can only be carried out for rather small values of M. The shift-register length n cannot be too large either, because otherwise the density of new dependent triples becomes small, spoiling the statistics, and also because the range of M values considered should cover an appreciable part of the complete period. Also, in order to compare their effects, different two-bit feedback production rules should exist at the same value of n. These requirements are met by $n = 17$, for which case Zierler and Brillhart [18] give $R(17; 11)$, $R(17; 12)$, and $R(17; 14)$ as production rules for maximum-length sequences. We determined p_j, that is, the set $\{J\}$ of Eq. (83), for $j \leqslant M = 15000$. Figure 9 shows the results in terms of $D_L(M)$, with $L = 1000$ as interval size. As could be expected, there is an appreciable amount of scatter even between the points belonging to a single production rule.

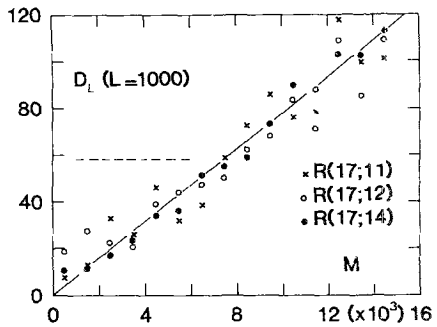

FIG 9. The density $D_L\langle M\rangle$ of the new dependent triples for three different maximum-length sequences with period $N=2^{17}-1$. The full line corresponds with $p_j=j/N$, the dashed line with $p_j=1/n$ (see Eq. (85)).

However, there is also a general trend visible, indicated by the diagonal straight line. This line is not drawn arbitrarily, but results in Eq. (85) when the choice

$$p_j=j/N \tag{86}$$

is made for the probability, which now starts to deserve its name. This choice is completely logical: if b_1 and b_j are given, there is within the complete period of the sequence always one element b_k such that $b_1\oplus b_j$ is equal to b_k; the probability that b_k lies between b_1 and b_j, thus forming a new dependent triple at b_j, is indeed j/N. When this choice for p_j is inserted in Eq. (84) then in fact the binomial approximation to $B_M(3)$ of Eq. (55) is found. The upper bound (82) is obtained by putting $p_j=1/n$ in Eq. (84); the dashed line is the corresponding behavior of $D_L(M)$.

It seems that a rather consistent picture emerges. We have obtained similar results for a number of production rules with $n=22$ and $n=31$, but the statistics gets poorer because of the factor $1/N$ in Eq. (86), which requires much larger values for M and L when n increases. The relative scatter of the points representing $D_M(L)$ diminishes, but the computer effort needed grows steeply because of the M^2 character of the algorithm for finding new dependent subsets.

When the upper bound of Eq. (82) is inserted into Eq. (80) one finds for the skewness

$$\gamma_1^{(1)}\lesssim 3M^{1.2}/n. \tag{87}$$

For $n\approx 20$ this implies a rather unsatisfactory behavior already for $M\gtrsim 50$, even when it is taken into account that the upper bound is rather weak, as is the case in Fig. 9. This agrees with the results obtained by Lindholm [7] and Fredricsson [11] for such values of n with $50\lesssim M\lesssim 500$. Neither of the production rules featured in Fig. 9 have very good randomness properties; the first contributions to $D_L(M)$ arrive too early, in particular for $R(17;12)$ where the position of the additional

feedback bit is almost equal to $2n/3$. Because of the cumulative effects on $B_M(3)$ such early contributions are real culprits. However, for larger n the situation improves considerably: the upper bound goes down, and M can become larger, which smoothes out the influence of regions with high and with low densities of new dependent triples. (In passing by we remark that the present arguments can in principle be extended to cover also $B_M(4)$ and the fourth moment.)

To test the above ideas we have sampled the distribution $f_M^{(1)}(\Delta)$ for a number of production rules $R(n; k)$ with $n < 100$ and k not close to 0, n, or a simple fraction of n. Starting from a given initial configuration of n bits we first generated M further elements of the sequence and determined (and stored) Δ, the difference between the numbers of zeros and ones in these M bits. Then the sequence was generated for a further n bits, but also the n times retarded sequence was generated; this saves memory space (only $2n$ bits need to be kept instead of $M+n$), while it enables one to correct Δ for the new bits that enter and the old bits that leave the subsequence. The value of Δ reached after this procedure was again stored. Repeating this process t times, and thereby generating a total number of $tn+M$ bits of the maximum-length sequence, we determined the histogram $f(\Delta)$ which is a sample of size t drawn from the true distribution $f_M^{(1)}(\Delta)$. Since subsequent values of Δ are taken from partly overlapping subsequences they are not independent; the effective number of independent samples is $t' = tn/M$ (even these are not truly independent, of course). The first four moments of the histogram obtained were calculated. In order to facilitate the comparison with the normal distribution to which the binomial distribution $f_M^{(2)}(\Delta)$ is equivalent for large M, the scales of the histogram were manipulated: the variable $\Delta' = \Delta/M^{1/2}$ was introduced and a number of intervals of the original histogram were collected in a single interval of the scaled histogram in such a way that the range $-2 < \Delta' < 2$ would cover ≈ 40 intervals.

In this manner a large number of production rules with $n \gtrsim 100$ were studied. In general, the distributions obtained were quite satisfactory, although for large M the

Figure 10 shows the distribution found for $R(127; 1)$, which is conjugate to $(127; 126)$, with $M = 2000$ and $t = 20{,}000$. By scanning through subsequent rows of triangles of zeros this production rule tends to overconcentrate zeros, thereby lifting the wing of the distribution for positive Δ. That wing extends also beyond $\Delta = 2M^{1/2}$; the opposite is true for the other wing (where no dots are shown, no values of Δ were found). The values obtained for skewness and excess are $\gamma_1 = 0.523$ and $\gamma_2 = 0.655$, both higher than is acceptable, but lower than we observed for production rules of the type $R(n; n-1)$ with smaller n. In Fig. 11 the distribution for $R(1279; 216)$ is depicted, with $M = 32{,}000$ and $t = 20{,}000$. Skewness and excess are now $\gamma_1 = 0.175$ and $\gamma_2 = 0.199$.

Similar results as for $R(1279; 216)$ were obtained for $R(127; 112)$, $R(607; 147)$, $R(9689; 9218)$, and a number of other production rules, but in particular for the larger n values the statistics of the data is not sufficient. Also, larger values of M should be considered, in order to be sure that no regions with an exceptionally high

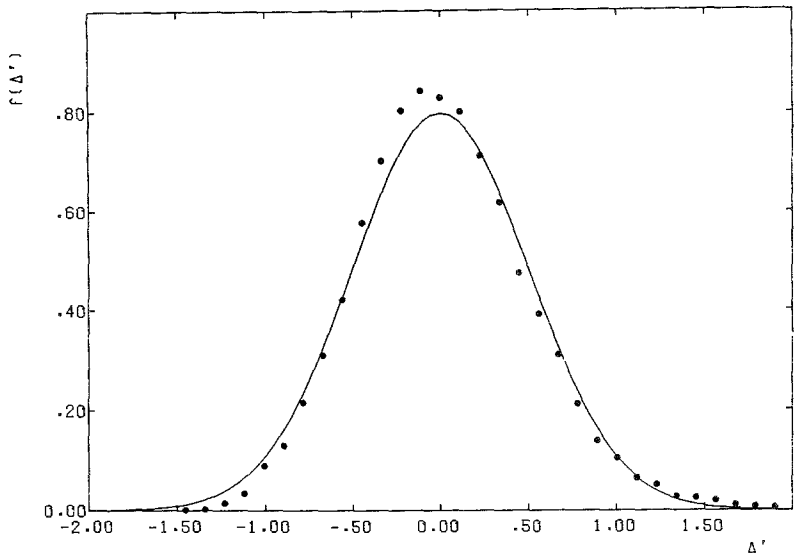

FIG. 10 The scaled histogram $f(\Delta')$ obtained from production rule $R(127; 126)$ for subsequences of length $M = 2000$. Units are such that the area under the full curve, which represents the normal distribution, is equal to 1. The first four moments of the unscaled histogram are: $\mu_1 = 1.404$, $\mu_2 = 1\,919 \times 10^3$, $\mu_3 = 4.395 \times 10^4$, and $\mu_4 = 1.346 \times 10^7$. The effective sample size is $t' = tn/M = 1270$.

density of new dependent triples occur. However, above some value of M the cumulative effects on $B_M(3)$ of those regions should be cancelled by the effects due to low-density regions.

We close this section with three remarks. The first one concerns the second moment of $f_M^{(1)}(\Delta)$. This distribution is (apart from the Kronecker-δ contribution) exactly binomial for $M \leqslant n$; but beyond that value the binomial distribution just proceeds to spread its wings without restriction while $f_M^{(1)}(\Delta)$ cannot follow suit completely. This is most conspicuous for small values of n; see, e.g., Table II. For larger n, the bulk of the distributions is concentrated at small values of Δ/M, and their second moments remain almost equal for a larger range of M. However, the second moment of $f_M^{(1)}(\Delta)$ does not increase indefinitely but reaches a maximum of $\frac{1}{4}N$ at $M = \frac{1}{2}N$ (the correction terms M'/N in Eqs. (77) now become important); when M approaches N the other binomial distribution, that of Eq. (61), must be met. When $M \approx \frac{1}{2}N$, the distribution $f_M^{(1)}(\Delta)$ is indeed far from binomial but behaves rather wildly, as one may discover for $n \lesssim 30$. Somewhere in the range $n < M < \frac{1}{2}N$ the distribution $f_M^{(1)}(\Delta)$ must definitively lose its ability to imitate the proper binomial distribution, but when n is larger than about 100 the second moments start to deviate only appreciably when M attains values of order N/n, far beyond reach.

Second, in the above little was said about the consequences of correlations in

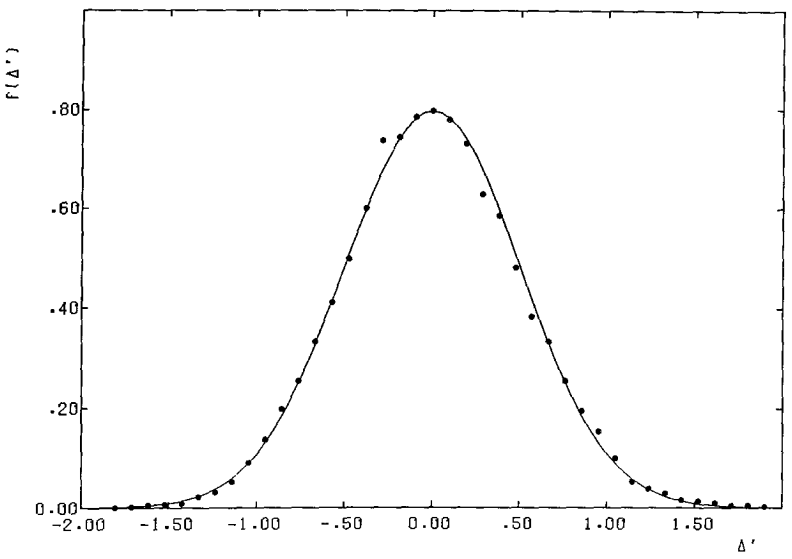

Fig. 11. The same as Fig. 10, but now for $R(1297; 216)$ with $M = 32{,}000$. The moments of the unscaled histogram are $\mu_1 = 1.490$, $\mu_2 = 3.295 \times 10^4$, $\mu_3 = 1.049 \times 10^6$, and $\mu_4 = 3.473 \times 10^9$. The effective sample size is $t' = tn/M = 811$.

maximum-length sequences for the properties of the q-bit random numbers extracted from them. We will restrict our attention to the case in which subsequent parts of q bits are taken from the sequence to form subsequent random numbers. Dependent subsets of size m in the sequence that would coincide with the most-significant bits of more or less neighboring random numbers are of course undesirable, but it follows from our considerations that the probability for this happening becomes extremely small even when 2-bit feedback production rules are used, at least when the precaution is taken to choose q not equal or close to a simple fraction of n. If r is the largest number for which rq is less than n, a sequence of r random numbers is completely independent. The most significant bit of the random numbers is taken from a decimated version of the original sequence, which is again a maximum-length sequence; if q is not a power of 2, the decimated sequence will probably belong to a multiple-bit feedback production rule, with (again probably) improved randomness properties. In any case, the most significant bits of n subsequent random numbers then are completely independent. Of vital importance, we found strong indications that much longer subsequences behave "normally" when n increases. Care must be taken though, as we learned when we inadvertently used $R(127; 15)$ to generate 32-bit random numbers in an Ising model simulation where the sites of a quadratic lattice with dimensions that are a power of 2 were visited sequentially (see Ref. [1]). The coincidence of so many powers or close powers of 2 was too much, and we found small but persistent deviations between our results and known exact data. Preliminary and somewhat superficial statistical tests of the

usual type gave no reason to mistrust $R(127; 15)$. This is a reminder to test the random-number generator one uses against known data for the particular simulation problem that one wants to study; only then are the correlations inherent in the problem present and can a possible interference with correlations in the random-number generator be detected.

Third, a few words about the actual random-number generator now in use in the DISP, the Delft Ising System Processor [1]. When we discovered the small disagreement with exact data, we changed over to visiting the lattice sites randomly; to have this possibility was one of our design requirements, but using it reduced the speed of the machine since the same random-number generator now had to provide two random numbers at each Monte Carlo step, one for the position and one for comparison with a Boltzmann probability. Although in principle new undesirable correlations could in this way have entered our problem the deviations vanished. In the DISP it is now standard practice to select lattice sites randomly, for which the old random-number generator is used. For the Monte-Carlo process proper a more powerful and programmable random-number generator was built and installed, based again on two-bit feedback production rules. The length n of the shift register and the position of the additional feedback bit can be chosen freely up to $n = 9689$ and even larger; these numbers are fed into the machine in the initialization phase, together with a varying collection of n seed bits. Since the installation of this random-number generator, three years ago, no reason to mistrust its properties arose. Usually, the production rule $R(9689; 471)$ is adopted, but occasionally runs are repeated with a different production rule for $n > 1000$ taken from Zierler's list [19].

7. Conclusions

A general description of the properties of maximum-length sequences in terms of the complete hierarchy of correlation functions is possible. New and detailed arguments have been found to support the somewhat trivial statement that production rules which generate sequences of extreme length in general have much better randomness properties. There are strong indications that such maximum-length sequences can be used as a reliable source of random numbers, if care is taken to avoid sequences in which correlated parts are not uniformly distributed.

For future research in the generation of random numbers by means of linear-feedback production rules two directions seem to be particularly interesting. The first one is to study multiple-bit feedback production rules: can they be made available for large shift registers, and do they indeed have the improved randomness properties that compensate their obvious disadvantages? The second one is to improve the statistics of the data found for sequences of extreme length, by means of efficient programming and possibly even the construction of dedicated hardware.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. J. Groeneveld for reminding one of the authors (A.C.) of Glauber's paper at precisely the right moment. We are indebted to Bart Selman and Arnoud van Willigen, who as students contributed in many memorable ways to various parts of this research.

REFERENCES

1. A. HOOGLAND, J. SPAA, B. SELMAN, AND A. COMPAGNER, *J. Comput. Phys.* **51**, 250 (1983); A. HOOGLAND, A. COMPAGNER, AND H. W. J. BLÖTE, in *Architecture and Performance of Specialized Computer Systems*, edited by B. Alder (Academic Press, San Diego), in press.
2. R. B. PEARSON, J. L. RICHARDSON, AND D. TOUSSAINT, *J. Comput. Phys.* **51**, 241 (1983).
3. S. W. GOLOMB, *Shift Register Sequences* (Holden–Day, San Francisco, 1967).
4. G. HOFFMANN DE VISME, *Binary Sequences* (The English Univ. Press, London, 1971).
5. F. J. MACWILLIAMS AND N. J. A. SLOANE, *The Theory of Error-Correcting Codes*, Part II, North-Holland, Amsterdam, 1977).
6. R. C. TAUSWORTHE, *Math. Comput.* **19**, 201 (1965).
7. J. H LINDHOLM, *IEEE Trans. Inform. Theory* **14**, 569 (1968).
8. J. P. R. TOOTILL, W. D. ROBINSON, AND A. G. ADAMS, *J. Assoc. Comput. Mach.* **18**, 381 (1971).
9. J P. R. TOOTILL, W. D. ROBINSON, AND D. J. EAGLE, *J. Assoc. Comput. Mach.* **20**, 469 (1973).
10. T. G. LEWIS AND W. PH. PAYNE, *J. Assoc. Comput. Mach.* **20**, 256 (1973).
11. S. A. FREDRICSSON, *IEEE Trans. Inform. Theory* **21**, 113 (1975).
12. A. J. MILLER AND P. MARS, *Math. Comput. Simulation* **19**, 198 (1977).
13. A. C. ARVILLIAS AND D. G.MARITSAS, *J. Assoc. Comput. Mach.* **25**, 675 (1978).
14. S. KIRKPATRICK AND E. P. STOLL, *J. Comput. Phys.* **40**, 517 (1981).
15. R. J. GLAUBER, *J. Math. Phys.* **4**, 294 (1963).
16. J. C. P. MILLER, *Philos. Trans. R. Soc. London Ser. A* **293**, 599 (1980).
17. S. WOLFRAM, *Rev. Mod. Phys.* **55**, 601 (1983).
18. N. ZIERLER AND J. BRILLHART, *Inform. and Control* **13**, 541 (1968); **14**, 566 (1969).
19. N. ZIERLER, *Inform. and Control* **15**, 67 (1969); **16**, 502 (1970).
20. W. STAHNKE, *Math. Comput.* **27**, 977 (1973).
21. F. HARARY, *Graph. Theory* (Addison–Wesley, Reading, MA, 1969).
22. M KAC, *Amer. Math. Monthly* **54**, 369 (1947); reprinted in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954), 295.
23. D. E. KNUTH, *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms*, 2nd ed. (Addison–Wesley, Reading, MA, 1981), p. 30.