MATHEMATICS OF COMPUTATION
VOLUME 60, NUMBER 201
JANUARY 1993, PAGES 363-374

ON THE USE OF REDUCIBLE POLYNOMIALS AS RANDOM NUMBER GENERATORS

DAKAI WANG AND AALDERT COMPAGNER

ABSTRACT. The randomness properties and the hierarchy of correlation coefficients are studied of approximate-maximum-length sequences, for which the characteristic polynomial is a product of several primitive polynomials. The randomness properties are almost the same as for maximum-length sequences characterized by a primitive polynomial with many terms and of the same degree. Reducible characteristic polynomials have acceptable figures of merit and can be of extremely high degree. Since they are also easily constructed and implemented, reducible polynomials are strong candidates for reliable random number generation, especially at the bit rates needed in large-scale Monte Carlo simulations.

1. INTRODUCTION

Methods for the generation of pseudorandom numbers have recently been reviewed by Knuth [9], Marsaglia [11], Ripley [15], James [8], and Niederreiter [13]. In this paper, the method based on shift register (or SR-) sequences is considered, because it may satisfy the most stringent requirements of very long period, uniform distribution, statistical independence of successive numbers, and high speed. Most research on SR-sequences has been focussed on the use of primitive polynomials, since the resulting maximum-length (or M-) sequences have rather favorable properties. However, the primitive polynomials of high degree n that are available usually contain three terms only (see for instance the lists provided by Zierler and Brillhart [16, 17]). As discussed in earlier papers [3–5], these trinomials lead to third-order correlations over small distances and hence to a considerable amount of intrinsic structure in the resulting M-sequences, which were called 'ill-tempered'. Indeed, for primitive trinomials the figures of merit $\rho^{(d)}$, which as defined by André, Mullen, and Niederreiter [1] are a measure for the independence of the most significant parts of d subsequent strings of n bits taken from the sequence, are not acceptable, as they are much smaller than n.

The search for so-called universally optimal primitive polynomials of a given degree, which have large figures of merit $\rho^{(d)}$ for $d = 2$ to 5 simultaneously,

Received by the editor June 11, 1991 and, in revised form, February 6, 1992.

1991 *Mathematics Subject Classification.* Primary 65C10; Secondary 11K45, 62E25.

Key words and phrases. Random number generation, Monte Carlo simulations, pseudorandom sequences, shift register sequences.

The first author is on leave of absence from the Physics Department of the Northwest University at Xian, China.

requires rather time-consuming computations. In [1], with very efficient procedures, results were obtained only up to degree $n = 32$; for higher degrees, up to $n = 127$, primitive polynomials that are optimal only with respect to $\rho^{(2)}$ have been reported [12]. However, the random numbers at bit rates $\gtrsim 1$ GHz that are needed in large-scale Monte Carlo simulations have to be derived from characteristic polynomials with many terms and of degree $n > 1000$, or even much larger, in order to be reliable (a simple reason is that only $n/32$ subsequent pseudorandom numbers of 32 bits can be guaranteed to be fully independent). The resulting problem is circumvented when reducible polynomials are used instead of primitive ones [3, 4].

In this paper, it will be shown that the sequences generated by reducible polynomials obey the randomness properties formulated by Golomb [6] in almost the same way as M-sequences. The figures of merit for reducible polynomials will be discussed and compared with those for primitive polynomials, and the hierarchy of correlation coefficients for sequences derived from reducible polynomials will be examined.

2. Approximate-maximum-length sequences

An nth-degree SR-sequence is defined as a binary sequence $\{y_i\}$ generated by a linear recursion

$$(1) \qquad y_{i+n} = a_{n-1}y_{i+n-1} + \cdots + a_0 y_i \pmod 2, \qquad i = 0, 1, \ldots,$$

where the coefficients a_j are elements of F_2. The generalization to the Galois field F_p is straightforward. The characteristic polynomial associated with the recursion (1) is

$$(2) \qquad f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 .$$

The seed, which is the set of initial values $y_0, y_1, \ldots, y_{n-1}$, and the characteristic polynomial $f(x)$ entirely determine the whole sequence $\{y_i\}$. Formally, (1) can be rewritten as

$$(3) \qquad 0 = y_i + a_{n-1}y_{i-1} + \cdots + a_0 y_{i-n} = D^{-n}f(D)y_i ,$$

where D^{-1} denotes the unit-delay operator obeying

$$(4) \qquad D^{-1}y_i = y_{i-1} .$$

The smallest exponent N for which $D^{-N}y_i = y_i$ holds for all i is the period of the sequence. When the polynomial $f(x)$ is primitive, N is equal to $2^n - 1$ and the SR-sequence is an M-sequence, initialized by nonzero seeds such as $(0, 0, \ldots, 1)$.

Consider m distinct primitive polynomials $f_j(x)$ for $j = 1, \ldots, m$. The number m is assumed to be small: $m \lesssim 10$ is enough for the applications we have in mind (see §4). For each $f_j(x)$ the associated M-sequences obey

$$(5) \qquad D^{-n}f_j(D)y_i^{(j)} = 0,$$

and have periods

$$(6) \qquad N_j = 2^{n_j} - 1, \qquad j = 1, 2, \ldots, m,$$

where n_j denotes the degree of $f_j(x)$. The sequence $\{y_i\}$ with elements

$$(7) \qquad y_i = y_i^{(1)} + y_i^{(2)} + \cdots + y_i^{(m)} \pmod 2,$$

where nonzero seeds are chosen for each of the partaking M-sequences, obeys

$$D^{-n}f(D)y_i = D^{-n}f(D)(y_i^{(1)} + \cdots + y_i^{(m)}) = 0. \tag{8}$$

Obviously, the reducible polynomial

$$f(x) = f_1(x)f_2(x)\cdots f_m(x) \tag{9}$$

is a characteristic polynomial for the sequence $\{y_i\}$. When the periods N_j are not only distinct but also mutually prime, the period of $\{y_i\}$ is

$$N = N_1N_2\cdots N_m = \prod_{j=1}^{m}(2^{n_j} - 1) < 2^n - 1, \tag{10}$$

with

$$n = n_1 + n_2 + \cdots + n_m. \tag{11}$$

If $n_j \gg 1$ holds for $j = 1, \dots, m$, the difference between N and 2^n is relatively small. Hence, $\{y_i\}$ in (7) may be called an approximate-maximum-length (or AM-) sequence.

3. Randomness properties of AM-sequences

It is known (see, e.g., [10, Chapters 1 and 4]) that the set $S(f(x))$ of all sequences which satisfy the linear recursion (1) and which have a characteristic polynomial that is reducible as in (9), can be divided into families according to their minimal polynomials

$$f_{J_k}(x) = \prod_{j\in J_k} f_j(x), \tag{12}$$

where the ordered set

$$J_k = \{j_1, \dots, j_k\}, \qquad 0 \le k \le m, \tag{13}$$

of k different indices is chosen from the set $J_m = \{1, 2, \dots, m\}$. The families can be specified by J_k as $F(J_k)$. A sequence of this family can be obtained from (7) by choosing all-zero seeds for the sequences $\{y_i^{(j)}\}$ with j not belonging to J_k. The period of this sequence is

$$N(J_k) = \prod_{j\in J_k} N_j, \tag{14}$$

equal to the total number of different sequences in $F(J_k)$, which are translated versions of one another. For a given k, there are $\binom{m}{k}$ different sets J_k and families $F(J_k)$, adding to 2^m families in total. The sum of $N(J_k)$ over all families gives the total number of sequences in $S(f(x))$,

$$\sum_{k=0}^{m}\sum_{J_k} N(J_k) = \prod_{j=1}^{m}(N_j + 1) = \prod_{j=1}^{m} 2^{n_j} = 2^n. \tag{15}$$

The family $F(J_0)$ consists of only one sequence, the all-zero sequence with period 1, and the family $F(J_m)$ contains the $N(J_m) = N$ translationally equivalent AM-sequences we are interested in.

One may distinguish $N(J_k)$ different states, each of which consists of n successive bits, in a sequence of family $F(J_k)$. The total number of states in all sequences of $S(f(x))$ is 2^n, each of all possible strings of n bits appearing exactly once. Specifying a bit string of length $t \leq n$, one finds that the number of states which have the specified string as their t leading bits is 2^{n-t}. Therefore, if $R_t(J_k)$ denotes the number of those states in a period of a sequence belonging to $F(J_m)$, then we have

$$\sum_{k=0}^{m} \sum_{J_k} R_t(J_k) = 2^{n-t}. \tag{16}$$

Consider the three randomness properties R1, R2, and R3 of M-sequences that were introduced by Golomb [6, Chapter III] (see also Hoffmann de Visme [7, Chapter 8]). The first property, R1, expresses that the total number of 0's in one period of any M-sequence is only one less than the total number of 1's. An equivalent result is easily found for AM-sequences. Putting $t = 1$ and taking 0 as the specified bit, one gets the number of 0's in the AM-sequence to be

$$R_1^{(0)}(J_m) = 2^{n-1} - \sum_{k=0}^{m-1} \sum_{J_k} R_1^{(0)}(J_k). \tag{17}$$

The difference between this number and $R_1^{(1)}(J_m)$, the corresponding number for the case that the specified bit is 1, obeys

$$\Delta(J_m) = -\sum_{k=0}^{m-1} \sum_{J_k} \Delta(J_k).$$

Suppose that $\Delta(J_k) = (-1)^k$ holds for $k = 0, 1, \ldots, m-1$; then one has

$$\Delta(J_m) = -\sum_{k=0}^{m-1} \binom{m}{k} (-1)^k = (-1)^m. \tag{18}$$

Since this is obviously true for $m = 0$ (all-zero sequence) and $m = 1$ (M-sequence), it is true for any m by induction. Hence, the difference between an AM-sequence and an M-sequence with respect to R1 is trivial. It may be noted that in the context of algebraic coding theory a result equivalent to (18) was earlier obtained by Niederreiter [14].

Randomness property R2 concerns the number of runs of 0's and 1's, which again can be counted by (16). For instance, taking the specified bits to be $\nu = t - 2$ ones sandwiched between two zeros, one finds the number of runs of ν ones in a period of an AM-sequence to be

$$R_{\nu+2}(J_m) = 2^{n-\nu-2} - \sum_{k=0}^{m-1} \sum_{J_k} R_{\nu+2}(J_k), \qquad 1 \leq \nu \leq n-2. \tag{19}$$

This is also the number of runs of ν zeros. It is well known that $2^{n-\nu-2}$ is the number of specified runs in a period of an M-sequence of degree n. In addition, the only possible runs with $\nu \geq n-1$ in an M-sequence or AM-sequence of degree n are the $n-1$ zeros between two ones, the number of which is one, and n ones between two zeros, the number of which is also one. Therefore, the number of any particular run in a period of an AM-sequence is the same as for an M-sequence of the same degree if the run cannot appear in any sequences belonging to families $F(J_k)$ with $k < m$, and otherwise it is slightly less. The differences between an AM-sequence and an M-sequence with respect to R2 are negligibly small for all practical cases $(n_j \gg 1)$.

Golomb's randomness property R3 concerns the pair correlation function

$$C(2, s) = \frac{1}{N} \sum_{i=0}^{N-1} b_i b_{i+s}, \tag{20}$$

where instead of the bits y_i the parities

$$b_i = b(y_i) \equiv (-1)^{y_i} \tag{21}$$

are used, which obey

$$b(y_i)b(y_j) = b(y_i + y_j). \tag{22}$$

For an M-sequence, $C(2, s)$ can only take two values: $C(2, s) = 1$ when s is a multiple of the period $2^n - 1$, and $C(2, s) = -1/(2^n - 1)$ otherwise. Within a single period, the pair correlation function of an M-sequence is single-valued and almost vanishes. This property follows from the fact that the sequence

$$\{y_{i'}\} = \{y_i + y_{i+s}\} \tag{23}$$

obeys the same recursion as $\{y_i\}$. For $s = 0 \pmod{2^n - 1}$, the sequence $\{y_{i'}\}$ is the all-zero sequence; otherwise, it is a translated version of the original M-sequence. Because of (22), the sum of parity products in (20) is just the sum of parities of $\{y_{i'}\}$ over a period.

For AM-sequences a slightly more complicated property results. When s is not divisible by any of the periods N_j, namely

$$s \neq 0 \pmod{N_j}, \qquad j = 1, \dots, m, \tag{24}$$

then for each j the sequence

$$\{y_i^{(j)} + y_{i+s}^{(j)}\} = \{y_i'^{(j)}\} \tag{25}$$

is still an M-sequence of degree n_j. Therefore, the sequence

$$\{y_i'\} = \left\{ \sum_{j=1}^{m} y_i'^{(j)} \right\} = \left\{ \sum_{j=1}^{m} (y_i^{(j)} + y_{i+s}^{(j)}) \right\} \tag{26}$$

is still an AM-sequence belonging to family $F(J_m)$. Hence, for intervals obeying (24) the pair correlation function is

$$C(2, s) = (-1)^m / N, \qquad s \neq 0 \pmod{N_j},\ j = 1, \dots, m, \tag{27}$$

as follows from (18). However, if s is divisible by k of the periods N_j, the sum in the right-hand side of (26) consists only of $m - k$ out of the m available M-sequences, and the resulting sequence belongs to $F(\overline{J}_k)$ and has period $N(\overline{J}_k) = N/N(J_k)$, where $\overline{J}_k$ is the complement of J_k. Replacing m by $m - k$ and N by $N/N(J_k)$, one obtains from (27)

$$C(2, s) = \frac{(-1)^{m-k}}{N} \prod_{j \in J_k} N_j, \qquad s = 0 \left(\operatorname{mod} \prod_{j \in J_k} N_j \right). \tag{28}$$

Thus, there is now a whole spectrum of pair correlation values. When the periods of the partaking M-sequences are large, as is assumed here, the absolute values are all small compared with 1. Moreover, in the case of (28), the intervals are very large. In this modified form, randomness property R3 holds also for AM-sequences.

For a simple example, consider the characteristic polynomial

$$f(x) = f_1(x)f_2(x) = (1+x+x^2)(1+x^2+x^3) = 1+x+x^5,$$

with $m = 2$, $n_1 = 2$, $n_2 = 3$, $n = 5$, $N_1 = 3$, $N_2 = 7$, and $N = 21$. In this case, only three of the nine terms resulting from a product of two trinomials survive. The two M-sequences and the resulting AM-sequence are

$$\begin{aligned} \{y_i^{(1)}\} &= 101101101101101101101\ldots, \\ \{y_i^{(2)}\} &= 010011101001110100111\ldots, \\ \{y_i\} &= 111110000100011001010\ldots. \end{aligned}$$

One sees that (17) is obeyed. The runs 11111, 1111, 0000, and 000 in $\{y_i\}$, which can never appear in M-sequences of degree 2 and 3, occur precisely as often as in an M-sequence of degree 5, but the single-zero run, for instance, occurs two times less often because it occurs once in both $\{y_i^{(1)}\}$ and $\{y_i^{(2)}\}$. Transforming $\{y_i\}$ into the parity sequence

$$\{b_i\} = - - - - - + + + + - + + + - - + + - + - + \cdots,$$

and adding the products of pairs of elements 3 or 7 positions apart, one obtains $C(2, 3) = -1/7$ and $C(2, 7) = -1/3$, in agreement with (28). For intervals other than multiples of 3 and 7, the pair correlation function always takes the value $1/21$ as given by (27).

To summarize, the randomness properties R1, R2, and R3 of Golomb are somewhat less obeyed by AM-sequences than by M-sequences, but the difference is negligible when the periods of the constituting M-sequences are sufficiently large. These results show that in the search for binary sequences that are reliable random number generators there is no a priori reason to exclude AM-sequences. To proceed further, correlations of any order and size have to be considered. One approach to do so is by means of the figures of merit defined by Mullen and Niederreiter [12], in which these correlations play an implicit but important role.

4. Figures of merit for reducible polynomials

From tables of primitive polynomials [16, 17] it is easy to construct a reducible polynomial of a desired degree. To ensure that the m periods N_j are mutually prime, the factorizations of $2^n - 1$ by Brillhart and Selfridge [2] can be used. Next to the degree of the polynomial, also its number of terms q is of interest, as it is the order of the first-correlated set discussed in [3, 4]. When q is sufficiently large, say in the range of $\frac{1}{2}n \pm n^{1/2}$, deviations from randomness are most likely to be due only to correlations that either are of large order or cover a large distance, so that the associated sequence may be called well-tempered.

The number of terms of the resulting reducible polynomial (9) obeys

$$q \leq \prod_{j=1}^{m} q_j,$$

TABLE 1. Figures of merit $\rho^{(d)}$ and number of terms q for a few reducible polynomials $f(x) = f_1 f_2 f_3$ of degree 32. For instance, the first three numbers, 0, 1, 4, indicate the polynomial $f_1(x) = 1 + x + x^4$.

f_1	f_2	f_3	q	$\rho^{(2)}$	$\rho^{(3)}$	$\rho^{(4)}$	$\rho^{(5)}$
0, 1, 4	0, 2, 11	0, 3, 17	14	31	28	23	22
0, 2, 5	0, 3, 10	0, 5, 17	11	29	26	26	23
0, 1, 7	0, 1, 5, 6, 8	0, 3, 17	15	31	26	26	22
0, 1, 4	0, 1, 3, 4, 13	0, 14, 15	16	28	26	24	23
0, 2, 5	0, 1, 3, 4, 13	0, 1, 11, 12, 14	19	30	22	22	22
0, 3, 7	0, 2, 11	0, 1, 11, 12, 14	19	30	26	26	21
0, 1, 5, 6, 8	0, 2, 11	0, 1, 3, 4, 13	21	30	22	22	21
0, 4, 9	0, 3, 10	0, 9, 10, 12, 13	19	31	25	25	24
0,1,2,3,4,5,6,7,8,9,11,12,21,22,24,26,27,28,30,31,32			21	32	26	26	25

where q_j is the number of terms of $f_j(x)$. The equality sign is valid when no cancellations due to the restriction to the Galois field F_2 occur. When the product is larger than $\frac{1}{2}n$, cancellations will almost always cause $q \approx \frac{1}{2}n$. In the above-mentioned lists of primitive polynomials, q_j is 3 or 5. If m of those polynomials are chosen, with $3^m \approx \frac{1}{2}n$, the resulting q will very likely be in the favorable range.

Because of the time-consuming calculations involved, the practical use of the figures of merit $\rho^{(d)}$ is restricted to polynomials of rather low degree, below $n \approx 100$. Roughly speaking, if the figure of merit $\rho^{(d)}$ of a characteristic polynomial is not much less than n, the pseudorandom numbers produced from its SR-sequence will pass the d-dimensional serial test. For the so-called universally optimal primitive polynomials of degree $n \leq 32$ that have good figures of merit $\rho^{(d)}$ for $d = 2$ to 5, as given in [1], the number of terms q is indeed always in the desirable range. These optimal polynomials were found by an exhaustive search, in which the computational task is so heavy that the search had to stop at $n = 32$. Conversely, when q is large, as is usually the case for reducible polynomials, the figures of merit $\rho^{(d)}$ may in general be expected to be reasonably large a priori.

To demonstrate this, several reducible polynomials of degree 32 and their figures of merit $\rho^{(d)}$ for $d = 2$ to 5 are listed in Table 1. The selection of the factors $f_j(x)$ was rather arbitrary, though occasionally somewhat lower figures of merit were encountered. The last line of Table 1 is the universally optimal primitive polynomial of degree 32 given in [1].

For degrees $n \leq 127$, primitive polynomials that are optimal in terms of $\rho^{(2)}$ only (then, $\rho^{(2)}$ should be equal to n or $n + 1$), were determined by Mullen and Niederreiter [12]. In Table 2, several examples are shown of reducible polynomials with n varying from 40 to 110. The results obtained for $\rho^{(2)}$ are only slightly less than n.

TABLE 2. Examples of reducible polynomials up to degree $n = 110$, together with the number of terms q in $f(x) = f_1 f_2 f_2$ and the figure of merit $\rho^{(2)}$. A comparison of n, q, and $Q \equiv q_1 q_2 q_3$ shows the effect of cancellations.

n	f_1	f_2	f_3	Q	q	$\rho^{(2)}$
40	0, 7, 10	0, 9, 11	0, 13, 14, 18, 19	45	21	39
50	0, 1, 2, 13, 14	0, 11, 17	0, 13, 14, 18, 19	75	27	47
60	0, 11, 17	0, 17, 20	0, 22, 23	27	19	57
70	0, 13, 14, 18, 19	0, 21, 22	0, 27, 29	45	25	67
80	0, 13, 14, 18, 19	0, 14, 15, 29, 30	0, 18, 31	75	29	77
90	0, 7, 18	0, 27, 29	0, 37, 38, 52, 53	45	41	87
100	0, 11, 13, 14, 16	0, 18, 31	0, 37, 38, 52, 53	75	39	97
110	0, 1, 22	0, 38, 41	0, 42, 47	27	25	107

The data in Tables 1 and 2 show that the choices for m and for the factors in (9) are not critical indeed; below $n \approx 100$, already for $m = 3$ the number of terms q takes care of itself and the resulting figures of merit have acceptable values. Occasionally, particular choices may occur for which some figures of merit would be too small, but in general this is not very likely to happen. For very high degrees, $n > 1000$ as required in large-scale Monte Carlo work, primitive polynomials with many terms, for which one could expect good figures of merit if one could calculate them, are difficult to find, and their implementation in soft- or hardware would not be easy. However, characteristic polynomials with $\approx 3^m$ terms for the production of AM-sequences are easily found by multiplying m primitive trinomials, for instance selected from those with Mersenne exponent degrees given by Zierler [16]. Since m does not need to be large, they can also be easily implemented.

5. THE CORRELATION COEFFICIENTS OF AM-SEQUENCES

By means of ensemble theory, a complete hierarchy of correlation coefficients for binary sequences can be defined [3–5]. To discuss the randomness of a given sequence, the scanning ensemble, consisting of that sequence and all its translated versions with equal weights, is most suitable. The correlation coefficient in that ensemble for a fixed set

$$I(q, s) = \{i_1, i_2, \dots, i_q\}, \qquad 0 \le i_1 < i_2 < \cdots < i_q \le N - 1, \tag{29}$$

is defined by

$$C_{I(q,s)} = \frac{1}{N} \sum_{j=0}^{N-1} \prod_{i \in I} b_{j+i}. \tag{30}$$

The pair correlation function $C(2, s)$ discussed in §3 is equal to $C_{I(2,s)}$. The set $I(q, s)$ is mainly characterized by its order, the number q of elements in I, and its size

$$s = i_q - i_1. \tag{31}$$

Notice that q is used here in a slightly more general sense than in the last section, where it was only the order of the first-correlated set (which is equal to the number of terms of the characteristic polynomial; see [3]).

For each set $I(q, s)$ a polynomial

$$g(x) = x^{i_1} + x^{i_2} + \cdots + x^{i_q} \tag{32}$$

can be constructed. Because of

$$\left\{\sum_{i\in I} y_{i+j}\right\} = g(D)\{y_j\} = \{0\} \leftrightarrow g(x) = 0 \pmod{f(x)}, \tag{33}$$

the necessary and sufficient condition for a set to be completely correlated is that its polynomial $g(x)$ can be divided by the minimal polynomial $f(x)$ of the SR-sequence. For the AM-sequences of $F(J_m)$ we thus have

$$C_{I(q,s)} = 1, \qquad \gcd(g(x), f(x)) = f(x), \tag{34}$$

where "gcd" stands for greatest common divisor.

When $g(x)$ is divisible by some of the factors of $f(x)$, the resulting sequence in the left-hand side of (33) belongs to the family $F(\overline{J}_k)$. Then one has

$$C_{I(q,s)} = \frac{(-1)^{m-k}}{N} \prod_{j\in J_k} N_j, \qquad \gcd(g(x), f(x)) = \prod_{j\in J_k} f_j(x). \tag{35}$$

In fact, this is a general expression for C_I, valid for all possible k. The case $k = m$ refers to the completely correlated sets in (34), and the case $k = 0$ refers to the almost uncorrelated sets with $\gcd(g(x), f(x)) = 1$, for which C_I is equal to $(-1)/N$. The different cases of $I(q, s)$ can be referred to as k-correlated sets.

Now consider the two conservation laws,

$$\langle C_I\rangle = \frac{1}{2^N}\sum_I C_I = 0, \tag{36}$$

$$\langle C_I^2\rangle = \frac{1}{2^N}\sum_I C_I^2 = \frac{1}{N}, \tag{37}$$

which in [3] were shown to be valid for any periodic binary sequence. Especially the strong conservation law (37) tells us that if a periodic sequence has small absolute correlation values for some sets, it must have large values for other sets. Tests for randomness that do not take this into account tend to have an ambiguous character.

To find the contributions from each of the k-correlated sets to the sum of squared correlation coefficients, the number $A(J_k)$ of k-correlated sets within a period is needed. If i_1 in (29) is equal to 0, the set $I(s, q)$ is called a basic set. According to (30), all translated versions of a basic set have the same correlation coefficient. For a given J_k, the first basic k-correlated set always corresponds to the minimal polynomial of the sequences that belong to $F(J_k)$. Multiplication by any polynomial $m(x)$ yields a new basic k-correlated set of size s, provided that $m(0) \neq 0$ and that the degree of $m(x)$ is equal to $s - n(J_k)$, with

$$n(J_k) = \sum_{j\in J_k} n_j. \tag{38}$$

Therefore, the number of basic k-correlated sets of size s is

$$G_s(J_k) = \begin{cases} 1, & s = n(J_k), \\ 2^{s-n(J_k)-1}, & s \geq n(J_k) + 1. \end{cases} \tag{39}$$

When all the translated versions of these basic k-correlated sets within a period N are counted, irrespective of their size, one obtains

$$B(J_k) = 2^{N-n(J_k)} - 1, \tag{40}$$

which differs from $A(J_k)$ because some of the k-correlated sets produced so far will also be k'-correlated sets with $k' = k+1, \dots, m$, if the polynomial $m(x)$ can be divided by factors $f_j(x)$ with j not belonging to J_k. Thus, (40) does not measure the exact number of k-correlated sets except for $k = m$, in which case

$$A(J_m) = B(J_m) = 2^{N-n} - 1 \tag{41}$$

is the number of completely correlated sets within one period. However, the numbers $A(J_k)$ for $k \neq m$ can be derived from $B(J_k)$ by iteration:

$$A(J_k) = B(J_k) - \sum_{i=k+1}^{m} \sum_{J_i \supset J_k} A(J_i), \tag{42}$$

starting from (40) with $k = m - 1$. The result is

$$A(J_k) = 2^{N-n} N / \prod_{j \in J_k} N_j, \qquad k = 0, 1, \dots, m-1. \tag{43}$$

This relation is even valid for $k = m$, if the empty set $I = \varnothing$ with $C_\varnothing = 1$ is included into the class of m-correlated sets (increasing $A(J_m)$ by 1).

Using (35), (41), and (43), one may verify

$$\sum_{k=0}^{m} \sum_{J_k} A(J_k) C_{I(J_k)} + 1 = 0, \tag{44}$$

$$\sum_{k=0}^{m} \sum_{J_k} A(J_k) C_{I(J_k)}^2 + 1 = \frac{2^N}{N}, \tag{45}$$

in agreement with the conservation laws. From (35), (41), and (43) one also finds the following relations:

$$A(J_0) \gg \sum_{J_1} A(J_1) \gg \cdots \gg A(J_m), \tag{46}$$

$$A(J_0) C_{I(J_0)}^2 \ll \sum_{J_1} A(J_1) C_{I(J_1)}^2 \ll \cdots \ll A(J_m) C_{I(J_m)}^2, \tag{47}$$

where $I(J_k)$ denotes a k-correlated set $I(q, s)$. In fact, the contributions to the sum of squared correlation coefficients are entirely dominated by the completely correlated sets $I(J_m)$ because of

$$\sum_{k=0}^{m-1} \sum_{J_k} A(J_k) C_{I(J_k)}^2 = 2^{N-n}(2^n - N)/N \ll 2^{N-n}. \tag{48}$$

Therefore, the k-correlated sets with $k = 0, \dots, m-1$ have negligible effects on the randomness properties of AM-sequences.

An analysis of the m-correlated sets with $C_{I(q,s)} = 1$ for the most relevant region of q and s will be the subject of a following paper.

6. Conclusions

For characteristic polynomials of a given degree and with a similar number of terms, the randomness properties of AM-sequences resemble those of M-sequences. Reducible polynomials are very promising as pseudorandom number generators, and have great practical advantages.

Acknowledgments

The expert assistance of Yves Fonk with the computer facilities was highly appreciated. The first author is grateful to Professor Ir. B. P. Th. Veltman for the hospitality offered during his stay in the Faculty of Applied Physics in Delft, which stay was supported by the Dutch Technology Foundation. The second author acknowledges the partial support of a NATO grant (ref. CRG 900661).

Note added in proof. The list of primitive trinomials up to degree $n = 9689$ that was given by Zierler [16] has recently been extended up to degree $n = 132049$; see J. R. Heringa, H. W. J. Blöte and A. Compagner, *New primitive trinomials of Mersenne-exponent degrees for random-number generation*, Internat. J. Modern Phys. C3, No. 3 (June 1992), 561–564.

Bibliography

1. D. A. André, G. L. Mullen, and H. Niederreiter, *Figures of merit for digital multistep pseudorandom numbers*, Math. Comp. **54** (1990), 737–748.
2. J. Brillhart and J. L. Selfridge, *Some factorizations of* $2^n \pm 1$ *and related results*, Math. Comp. **21** (1967), 87–96.
3. A. Compagner, *The hierarchy of correlation coefficients for random binary sequences*, J. Statist. Phys. **63** (1991), 883–896.
4. ——, *Definitions of randomness*, Amer. J. Phys. **59** (1991), 700–705.
5. A. Compagner and A. Hoogland, *Maximum-length sequences, cellular automata and random numbers*, J. Comput. Phys. **71** (1986), 391–428.
6. S. W. Golomb, *Shift register sequences*, Holden-Day, San Francisco, 1967.
7. G. Hoffmann de Visme, *Binary sequences*, English Univ. Press, London, 1971.
8. F. James, *A review of pseudorandom number generators*, Comput. Phys. Comm. **60** (1990), 329–344.
9. D. E. Knuth, *The art of computer programming*, Vol. 2, Chapter 3, Addison-Wesley, Reading, MA, 1981.
10. R. Lidl and H. Niederreiter, *Introduction to finite fields and their applications*, Cambridge Univ. Press, 1986.
11. G. Marsaglia, *A current view of random number generators*, Computer Science and Statistics (L. Billard, ed.), North-Holland, Amsterdam, 1985, pp. 3–10.
12. G. L. Mullen and H. Niederreiter, *Optimal characteristic polynomials for digital multistep pseudorandom numbers*, Computing **39** (1987), 155–163.
13. H. Niederreiter, *Recent trends in random number and random vector generation*, Ann. Oper. Res. **31** (1991), 323–345.
14. ——, *Weights of cyclic codes*, Inform. and Control **34** (1977), 130–140.

→ B. D. Ripley, *Thoughts on pseudorandom number generators*, J. Comput. Appl. Math. **31** (1990), 153–163.

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

17. N. Zierler and J. Brillhart, *On primitive trinomials* (mod 2), Inform. and Control **13** (1968), 541–554; **14** (1968), 566–569.

Laboratory of Applied Physics, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, the Netherlands
E-mail address: ac@dutncp8.tn.tudelft.nl