Feedback Shift Register Sequences

Hong-Yeop Song
Department of Electrical and Electronics Engineering
Yonsei University, Seoul, 120-749, Korea
E-mail: hy.song@coding.yonsei.ac.kr

Abstract

Feedback Shift Register (FSR) sequences have been successfully implemented in many communication systems for their randomness properties and ease of implementation. These include ranging and navigation systems, spread spectrum communication systems, CDMA mobile communication systems, and crypto systems such as streamciphers.

This article gives a brief overview of FSR sequences, both linear and non-linear. Two conditions on the connection logic of FSRs for better output sequences are described, which are the branchless condition and the balanced logic condition. We use mostly the state transition diagram of an FSR to describe the property of its output sequences. For linear FSR sequences, we describe the relation between the connection polynomials and the structure of the cycle decomposition in the state diagram, and hence the periodicity of the output sequences.

Various randomness properties of the maximal length linear FSR sequences, known as m-sequences, are described: balance, run-distribution, span, ideal autocorrelation, constant-on-the-coset, and cycle-and-add properties. Two properties, the ideal autocorrelation function and the smallest linear complexity, of m-sequences are described in detail. Finally, a complete analysis of 4-stage FSRs is provided.

Keywords: Feedback Shift Registers, Linear Feedback Shift Registers, Pseudorandom Sequences, M-sequences, PN Sequences, De Bruijn Sequences. Hadamard Sequences, Spread Spectrum Communications, Streamciphers.

I. Introduction

Binary random sequences are useful in many areas of engineering and science. Well known applications are digital ranging and navigation systems because of the sharp peak in their autocorrelation functions [1], spread spectrum modulation and synchronization using some of their correlation properties [2], [3], [4], [5], [6], and streamciphers in which the message bits are exclusive-ORed with key streams which must be as random as possible [7], [8]. Several randomness tests for binary sequences have been proposed in practice [8], but no universal consensus has been made yet with regard to the true randomness of binary sequences [9].

Random binary sequences can be obtained in theory by flipping an unbiased coin successively, but this is hardly possible in most practical situations. In addition, not only the random sequence itself must be produced at some time or location but also its exact replica at remote (in physical distance or time) locations must also be produced in spread spectrum modems. This forces us to consider the sequences which look random but can be easily reproduced with a set of simple rules or keys. We call these *pseudorandom* or *pseudonoise* (PN) sequences. It has been known and used for many years that *feedback shift registers* (FSR) are most useful in designing and generating such PN sequences. This is due to their simplicity of defining rules and their capability of generating sequences with much longer period [10]. Approaches using FSR sequences solve the following two basic problems in most applications: cryptographic secrecy and ease of generating the same copy over and over.

One of the basic assumptions in conventional cryptography is that the secrecy of a system does not depend on the secrecy of how it functions rather it depends on the secrecy of the key which is usually kept secret [11]. Feedback shift registers are most suitable in this situation because we do not have to keep all the terms of the sequences secret. Even though its connection is revealed and all the functionality of the system is known to the public, any unintended observer will have an hard time of locating the exact phase of the sequence in order to break the system provided that the initial condition is kept secret. The current CDMA modem (which is used in the successful second and third generation mobile telephone systems) depends heavily on this property for its privacy [12].

One of the difficulties of employing spread spectrum communication systems a few decades ago was on the effort of reproducing at the receiver the exact replica of PN sequences that were used at the transmitter [2]. Store-and-replay memory wheels to be distributed initially were once proposed, and using a secret and safe third channel to send the sequence to the receiver was also proposed. The theory and practice of FSR sequences have now been well-developed so that by simply agreeing upon the initial condition and/or connection method (which requires much smaller memory space or computing time) both ends of communicators can easily share the exact same copy.

In Section II, the very basics of feedback shift registers (FSR) and their operations are described, following the style of [10]. We will concentrate only on some basic terminologies, state transition diagrams, truth tables, cycle decompositions, and the like. In fact, the detailed proofs of claims and most of discussions and a lot more can be found in [10]. Section III covers mostly the *linear* FSRs. The linear FSR sequences have been studied in various mathematics literature under the name of *linear recurring sequences.* Lidl and Niederreiter gave a comprehensive treatment on this subject in [13]. Some other well-known textbooks on the theory of finite fields and linear recurring sequences are [14], [15], [16], [17], [18]. In this article, we will discuss the condition for their output sequences to have maximum possible period. The maximum period sequences which are known as *m-sequences*, are described in detail, including randomness properties. Two properties of m-sequences deserve special attention: m-sequences of period P have the two-level ideal autocorrelation which is the *best* over all the balanced binary sequences of the same period, and they have the linear complexity of $\log_2(P)$ which is the *worst* (or the smallest) over the same set. Some related topics on these properties will also be discussed. To give some further details of the ideal two-level autocorrelation property, we describe a larger family of balanced binary sequences which come from, so called, *cyclic Hadamard difference sets*. An m-sequence can be regarded as the characteristic sequence of a cyclic Hadamard difference set of Singer type. To give some better understanding of the linear complexity property, we describe the *Berlekamp-Massey algorithm* (BMA) that determines the shortest possible linear FSR that generates a given sequence.

In Section IV, we describe some special cases of FSRs (including non-linear FSRs) with disjoint cycles in their state diagrams. Branchless condition and balanced logic condition will be discussed. De Bruijn sequences will briefly be described. Finally, 4-stage FSRs are analyzed in detail for a complete example. Section V gives some concluding remarks. We will restrict our treatment to only the sequences over a binary alphabet $\{0, 1\}$ in this article.

II. Feedback Shift Register Sequences, Truth Tables, and State Diagrams

The operation of an FSR can best be described by its state transition diagram. Its output at one time instant depends only on the previous state. Figure 1 shows a generic block diagram of an FSR (linear or non-linear) with L stages. At every clock, the content of a stage is shifted to the left, and the connection logic or the boolean function f calculates a new value x_k,

$$x_k = f(x_{k-L}, x_{k-L+1}, \ldots, x_{k-1}), \tag{1}$$

to be fed back to the rightmost stage. The leftmost stage gives an output sequence in which the first L terms are in fact given as an initial condition.

A *state* of this FSR at one instant k can be defined simply as the vector $(x_{k-L}, x_{k-L+1}, \ldots, x_{k-1})$, and this will be changed into $(x_{k-L+1}, x_{k-L+2}, \ldots, x_k)$ at the next instant. An FSR is called *linear* if the connection logic is a linear function on $x_{k-L}, x_{k-L+1}, \ldots, x_{k-1}$, that is, if it is of the form

$$x_k = f(x_{k-L}, x_{k-L+1}, \ldots, x_{k-1}) = c_L x_{k-L} \oplus c_{L-1} x_{k-L+1} \oplus \cdots \oplus c_1 x_{k-1}, \tag{2}$$

for some fixed constants $c_1, c_2, \ldots, c_L$. Otherwise, it is called *non-linear*. Here, the values of x_i are either 0 or 1, and hence the sequence is said to be over a *binary* *alphabet*, which is usually denoted as F_2, and $c_i \in F_2$ for all i. The operation $\oplus$ can easily be implemented as *exclusive-OR* and $c_i x_{k-i}$ as *AND* operation both using digital logic gates. In the remaining discussion, we will simply use addition and multiplication (mod 2), respectively, for these operations. Over this binary alphabet, therefore, one can add and multiply two elements, and the subtraction is the same as addition. There is only one non-zero element (which is 1) and the division by 1 is the same as the multiplication by 1.

Another method of describing an FSR is to use its truth table, in which all the 2^L states are listed on the left column and the next bits calculated from the connection logic f are listed on the right. The state change can easily be illustrated by the state diagram in which every state is a node and an arrow indicates the beginning (predecessor) and ending (successor) states. Figures 2 and 3 show examples of 3-stage FSRs, their truth tables, and state diagrams. Note that there are exactly 2^L states in total, and every state has at most two predecessors and exactly one successor. Note also that there are 2^{2^L} different L-stage FSRs, corresponding to the number of choices for the next bit column in the truth table. Finally, note that Fig. 3 has two disjoint cycles while Fig. 2 has branches and some absorbing states. Therefore, the FSR in Fig. 2 will output eventually the all-zero sequence, while that in Fig. 3 will output a sequence of period 7 unless its initial condition is 000.

In order to investigate this situation more closely, observe that any state has a unique successor, but up to 2 predecessors. From this, we observe that a branch occurs at a state which has two predecessors, and this happens in a state diagram if and only if there is a state which has no predecessor. A branch in a state diagram should be avoided since it will either seriously reduce the period of the output sequences or result in an ambiguous initial behavior. The necessary and sufficient condition for a branchless state diagram is, therefore, that no two states have the same successor. Consider any pair of states $(a_0, a_1, \ldots, a_{L-1})$ and $(b_0, b_1, \ldots, b_{L-1})$. If they are different in any other position except for the first, their successors $(a_1, a_2, \ldots, a_L)$ and $(b_1, b_2, \ldots, b_L)$ will still be different, because all the components except for the first will be shifted to the left and the difference remains. The remaining case is the pair of the form $(a_0, a_1, \ldots, a_{L-1})$ and $(a_0', a_1, \ldots, a_{L-1})$, where a_0' represents the complement of a_0. Their successors will be $(a_1, a_2, \ldots, a_{L-1}, f(a_0, a_1, \ldots, a_{L-1}))$ and $(a_1, a_2, \ldots, a_{L-1}, f(a_0', a_1, \ldots, a_{L-1}))$. For these two states to be distinct, the rightmost component must be different. That is,

$$f(a_0', a_1, \ldots, a_{L-1}) = f(a_0, a_1, \ldots, a_{L-1}) \oplus 1 = f'(a_0, a_1, \ldots, a_{L-1}).$$

Let $g(a_1, a_2, \ldots, a_{L-1})$ be a boolean function on $L-1$ variables such that

$$f(0, a_1, \ldots, a_{L-1}) = g(a_1, a_2, \ldots, a_{L-1}).$$

Then, the above relation can be written as

$$f(a_0, a_1, \ldots, a_{L-1}) = a_0 \oplus g(a_1, a_2, \ldots, a_{L-1}). \tag{3}$$

This is called the *branchless condition* for an L-stage FSR. For FSRs with the branchless condition, its truth table has only 2^{L-1} independent entries, and the top half of the truth table must be the complement of the bottom half. This condition is automatically satisfied with *linear* FSRs which is the topic of the next section.

III. Linear Feedback Shift Registers and M-sequences

A. Basics

The output sequence $\{s(k)|k = 0, 1, 2, \ldots\}$ of a linear feedback shift register (LFSR) with L stages as shown in Fig. 4 satisfies a linear recursion of degree L. Given $L+1$ constants, $c_1, c_2, \ldots, c_L, b$, and the initial condition $s(0), s(1), \ldots, s(L-1)$, the terms $s(k)$ for $k \geq L$ satisfy

$$s(k) = c_1 s(k-1) + c_2 s(k-2) + \cdots + c_L s(k-L) + b, \tag{4}$$

or equivalently,

$$s(k) + c_1 s(k-1) + c_2 s(k-2) + \cdots + c_L s(k-L) + b = 0 \tag{5}$$

The recursion is called *homogeneous* linear if $b = 0$ and *inhomogeneous* linear if $b \neq 0$. We will assume that $b = 0$ in this section and mainly consider the homogeneous linear recursion.

The *characteristic polynomial* of the homogeneous linear recursion in (4) or (5) is defined as

$$f(x) = 1 + c_1x + c_2x^2 + \cdots + c_Lx^L. \tag{6}$$

It contains all the connection coefficients, and this will completely determine the operation of the LFSR provided that the initial condition is specified. Note that $c_L \neq 0$ in order for this LFSR to be genuinely with L stages. Otherwise, the recursion becomes of degree less than L, and the LFSR with less than L stages can also be used to implement the recursion.

Note that it is also the characteristic polynomial of the sequence satisfying this recursion. A given sequence may satisfy many other recursions which are different from each other. The *minimal polynomial* of a given sequence is defined as the characteristic polynomial of the minimum degree recursion satisfied by the sequence. We will return to this and more later when we discuss the linear complexity of sequences.

In the state diagram of any LFSR, every state will have a unique successor and a unique predecessor, as stated at the end of the previous section. This forces the diagram to be (possibly several) disjoint cycles of states. In Fig. 3, the state diagram has two disjoint cycles, one with length 7, and the other with length 1. From this, we can easily see that the output sequence of an LFSR is *periodic*, and the period is the length of the cycle that the initial state (or the initial condition) belongs to. We can conclude, therefore, that *the output sequence of a LFSR is periodic with some period P that depends on both the initial condition and the characteristic polynomial.*

One special initial condition is the all-zero state, and this state will always form a cycle of length 1 *for any LFSR*. For any other initial state, the cycle will have a certain length ≥ 1 and this length is the period of the output sequence with the given initial condition. Certainly, the cycle with the maximum possible length must contain every not-all-zero state exactly once, and the output sequence in this case is known as the *maximal length linear feedback shift register sequence*, or the *m-sequence*, in short. Sometimes, PN sequences are used instead of m-sequences and vice versa, but we will make a clear distinction between these two terms. PN sequences refer to (general) pseudonoise sequences that possess some or various randomness properties, and m-sequences are a specific and very special example of PN sequences. For an L-stage LFSR, this gives the period $2^L - 1$, and Fig. 3 shows an example of an m-sequence of period 7. In fact, it shows 7 different *phases* of this m-sequence depending on the 7 initial conditions. Note also that the history of any stage is the same m-sequence in different phase.

The operation of an LFSR is largely determined by its characteristic polynomial. It is a polynomial of degree L over the binary alphabet F_2. How it factors over F_2 is closely related to the properties of the output sequence. In order to discuss the relation between the characteristic polynomials and the corresponding output sequences, we define some relations between sequences of the same period.

Let $\{s(k)\}$ and $\{t(k)\}$ be *arbitrary* binary sequences of period P. Then we have the following three important relations between these two sequences. (1) One is said to be a *cyclic shift* of the other if there is a constant integer τ such that $t(k-\tau) = s(k)$ for all k. Otherwise, two sequences are said to be cyclically distinct. When one is a cyclic shift of the other with $\tau \neq 0$, they are said to be in different *phase*. Therefore, there are P distinct phases of $\{s(k)\}$ which are all cyclically equivalent. (2) One is a *complement* of the other if $t(k) = s(k) + 1$ for all k. Finally, (3) one is a *decimation* (or d-decimation) of the other if there are constants d and τ such that $t(k-\tau) = s(dk)$ for all k. If d is not relatively prime to the period P, then the d-decimation will result in a sequence with shorter period which is P/g where g is the GCD of P and d. If some combination of the above three relations applies to $\{s(k)\}$ and $\{t(k)\}$, then they are called *equivalent.* Equivalent sequences share lots of common properties, and they are essentially the same sequences even if they look much different.

The necessary and sufficient condition for an L-stage LFSR to produce an m-sequence of period $2^L - 1$ is that the characteristic polynomial of degree L is, so called, *primitive* over F_2. This means simply that $f(x)$ is irreducible and $f(x)$ divides $x^{2^L-1} - 1$, but $f(x)$ does not divide $x^j - 1$ for all j from 1 to $2^L - 2$. The elementary theory of finite fields (or, Galois fields) discusses much more about these primitive polynomials, which we will not discuss in detail here. Instead, we refer the reader to some references for further exploration in theory [13], [14], [15], [16], [17], [18]. See [19] for the list of primitive polynomials of degree up to a few hundreds, which will be mostly enough for any practical application. There are $\phi(2^L - 1)/L$ primitive polynomials of degree L over F_2. Some primitive polynomials are shown in Table I for L up to 10. Here, $\phi(n)$ is the Euler's ϕ function and it counts the number of integers from 1 to n that are relatively prime to n.

In order to describe some properties of the output sequences of LFSRs, we include all the 4-stage LFSRs: block diagrams, characteristic polynomials, truth tables, state transition diagrams, and the output sequences in Figures from 5 to 8. Note that there are 16 linear logics for 4-stage LFSRs, and the condition $c_L = 1$ reduces it into half.

Figure 5 shows two LFSRs that generate m-sequences of period 15. The detailed properties of m-sequences will be described in the next subsection. Here, we note that two characteristic polynomials $f_1(x) = x^4 + x^3 + 1$ and $f_2(x) = x^4 + x + 1$ are reciprocal to each other. That is, the coefficients are 11001 and 10011. This gives two m-sequences which are reciprocal to each other. In other words, one is a 14-decimation of the other. Little arrows with a dot under the output sequences indicate this fact. Note that the roots of $f_2(x)$ are the 14-th power of those of $f_1(x)$. In general, if $f(x)$ and $g(x)$ are primitive of the same degree and the roots of one polynomial are d-th power of the other, then the m-sequence from one polynomial is a d-decimation of that from the other. The truth table shows only the top half since its bottom half is the complement of what is shown here, due to the branchless condition.

Figure 6 shows LFSRs with the characteristic polynomials which factor into smaller degree polynomials. Note that $f_3(x) = x^4 + x^3 + x^2 + 1 = (x+1)(x^3 + x + 1)$ and $f_4(x) = x^4 + x^2 + x + 1 = (x+1)(x^3 + x^2 + 1)$. Since both of the degree-3 polynomials in their factorizations are primitive, the LFSR generates m-sequences of period 7 which could have been generated by a 3-stage LFSR. Two characteristic polynomials are reciprocal, and the output sequences are reciprocal. Figure 7 shows LFSRs with the self-reciprocal characteristic polynomials, and the output sequences are self-reciprocal also. It means that reading a sequence in the reverse direction gives a cyclically equivalent one to the original.

Figure 8 shows two special LFSRs, the Pure Summing Register (PSR) and the Pure Cycling Register (PCR). PSR has an irreducible but not primitive characteristic polynomial. Observe that all the cycles except for the cycle containing (0000) have the same length, or the same period for its output sequences except for the all-zero sequence. This happens because the characteristic polynomial is irreducible. An irreducible polynomial, therefore, corresponds to a unique period, and it is called *the period of the irreducible polynomial*. A primitive polynomial of degree L is simply an irreducible polynomial with period $2^L - 1$. Possible periods of a given irreducible polynomial of degree L are the factors of the integer $2^L - 1$, which are not of the form $2^j - 1$ for $j < L$. When $2^L - 1$ is a prime, called a Mersenne prime, then every irreducible polynomial of degree $2^L - 1$ must be primitive.

Details on the property of PCR and some other properties of FSRs will be given at the end of Section IV.

B. Properties of m-sequences

Now, we will describe some basic properties of m-sequences of period $2^L - 1$, mostly without proofs. The first three properties, namely, balance, run-distribution, and ideal autocorrelation are commonly known as Golomb's postulates on random sequences [10]. Most of the following properties can be easily checked for the examples shown in Fig. 3 and in Fig. 5.

B.1 Balance Property

In one period of an m-sequence, the number of 1's and that of 0' are nearly the same. Since the period is an odd integer, they cannot be exactly the same, but differ only by one. This is called the balance property. When a matrix of size $(2^L - 1) \times L$ is formed by listing all the states of the maximum length cycle, then the rightmost column is the m-sequence and it will contain 2^{L-1} one's and $2^{L-1} - 1$ zero's since the rows are permutation of all the vectors of length L except for the all-zero vector.

B.2 Run Distribution Property

A string of the same symbol of length l surrounded by different symbols at both ends is called a run of length l. For example, a run of 1's of length 4 looks like ...0$\underline{1111}$0.... The run distribution property of m-sequences refers to the fact that a shorter run appears more often than a longer run, and that the number of runs of 1's is the same as that of 0's. Specifically, it counts the number of runs of length l for $l \geq 1$ in one period as shown in Table II. The span property of m-sequences implies this run distribution property.

B.3 Ideal Autocorrelation Property

A periodic unnormalized autocorrelation function $R(\tau)$ of a binary sequence $\{s(k)\}$ of period P is defined as

$$R(\tau) = \sum_{k=0}^{P-1} (-1)^{s(k)+s(k-\tau)}, \quad \tau = 0, 1, 2, ...,$$

where $k - \tau$ is computed mod P. When binary phase-shift-keying is used to digitally modulate incoming bits, we are considering the incoming bits whose values are taken from the complex values $\{+1, -1\}$. The change in alphabet between $s_i \in \{0, 1\}$ and $t_i \in \{+1, -1\}$ is commonly performed by the relation $t_i = (-1)^{s_i}$. Then we have $R(\tau) = \sum_{k=0}^{P-1} t(k)t(k-\tau)$ for each τ, and this calculates *the number of agreements minus the number of disagreements* when one period of

$\{s(k)\}$ is placed on top of its (cyclically) τ-shifted version. For any integer $L \geq 2$, and for any m-sequence $\{s(k)\}$ of period $P = 2^L - 1$, the ideal autocorrelation property of m-sequences refers to the following:

$$R(\tau) = \begin{cases} 2^L - 1, & \tau \equiv 0 \pmod{2^L - 1} \\ -1, & \tau \not\equiv 0 \pmod{2^L - 1} \end{cases} \tag{7}$$

The ideal two-level autocorrelation property of an m-sequence enables one to construct a Hadamard matrix of order 2^L of, so called, *cyclic* type. A Hadamard matrix of order n is an $n \times n$ matrix with entries only of ± 1 such that any two distinct rows are orthogonal to each other [20]. When the symbols of an m-sequence are mapped onto $\{\pm 1\}$ and all the cyclic shifts are arranged in a square matrix of order $(2^L - 1)$, the relation in (7) implies that the dot product of any two distinct rows is exactly -1 over the complex numbers. Therefore, adjoining a leftmost column of all $+1$'s and a top row of all $+1$'s will give a cyclic Hadamard matrix of order 2^L.

Cyclic type Hadamard matrices can be constructed from a balanced binary sequence of period $P \equiv 3 \pmod 4$ that has the ideal two-level autocorrelation function. M-sequences is one such class of sequences. Some other well-known balanced binary sequences with period $P \equiv 3 \pmod 4$ will be described later.

B.4 Span Property

If two vectors $(s(i), s(i+1), \ldots, s(i+L-1))$ and $(s(j), s(j+1), \ldots, s(j+L-1))$ of length L are distinct whenever $i \neq j$, then the sequence $\{s(k)\}$ is said to have this property. The indices of terms are considered mod P. For an m-sequence of period P, in addition, all the not-all-zero vectors of length L appear exactly once on the windows of length L. This can easily be seen by observing that an m-sequence is the sequence of the rightmost bits of states in the maximum length cycle of the state diagram. Each window of length L can then easily be identified with a state in this state diagram.

If we insert an additional 0 right after the run of 0's of length $L - 1$, the sequence will have period 2^L and the span property becomes perfect in that every window of length L shows all the vectors of length L exactly once. This is an example of a *de Bruijn sequence* of order L. The above construction has been successfully adopted in [12] to be used in spread spectrum modulations, for the values of $L = 14$ and $L = 41$. In general, many algorithms are currently known for de Bruijn sequences of period 2^L [21], and the above modification can easily be implemented, which will be described later in Section IV.

B.5 Constant-on-the-coset Property

For any m-sequence of period $2^L - 1$, there are $2^L - 1$ cyclically equivalent sequences corresponding to the $2^L - 1$ starting points. Constant-on-the-coset property refers to the fact that there exists exactly one among all these such that it is fixed with 2-decimation. An m-sequence in this phase is said to be in the *characteristic phase*. Therefore, for the m-sequence $\{s(k)\}$ in the characteristic phase, the following relation is satisfied:

$$s(2k) = s(k), \qquad \text{for all } k. \tag{8}$$

The relation in (8) deserves some special attention. It implies that every term in the $2k$-th position is the same as the one in the k-th position. This gives a set (or several sets) of positions in which the corresponding terms are the same. For example, $\{1, 2, 4, 8, 16, \ldots\}$ is one such set so that all the terms indexed by any number in this set are the same. Since the sequence is periodic with period $2^L - 1$, the above set is a finite set and called a cyclotomic coset mod $2^L - 1$. Starting from 3 gives another such set, $\{3, 6, 12, 24, \ldots\}$, and so on. In general, the set of integers mod $2^L - 1$ can be decomposed into some number of disjoint cyclotomic cosets, and now the constant-on-the-coset property describes itself clearly.

B.6 Cycle-and-add Property

When two distinct phases of an m-sequence are added term-by-term, a sequence of the same period appears and it is a different phase of the same m-sequence.

That is, for any given constants $\tau_1 \not\equiv \tau_2 \pmod{2^L - 1}$, there exists yet another constant τ_3 such that

$$s(k-\tau_1) + s(k-\tau_2) = s(k-\tau_3), \quad k = 0, 1, 2, ... \tag{9}$$

This is the cycle-and-add property of m-sequences. On the other hand, if a balanced binary sequence of period P satisfies the cycle-and-add property, then P must be of the form $2^L - 1$ for some integer L and the sequence must be an m-sequence.

Golomb has conjectured that the span property and the ideal two-level autocorrelation property of a balanced binary sequence implies its cycle-and-add property [22]. This has been confirmed for L up to 10 by many others, but still awaits a complete solution.

B.7 Number of Cyclically Distinct M-sequences of Period $2^L - 1$

For a given L, the number of cyclically distinct m-sequences of period $2^L - 1$ is equal to the number of primitive polynomials of degree L over F_2, and this is given by $\phi(2^L - 1)/L$ where $\phi(n)$ is the Euler's ϕ function and counts the number of integers from 1 to n that are relatively prime to n. All these $\phi(2^L - 1)/L$ m-sequences of period $2^L - 1$ are equivalent, and they are related with some decimation of each other. Therefore, any given one m-sequence can be used to generate all the others of the same period by using some appropriate decimations.

B.8 Trace Function Representation of M-sequences

Let q be a prime power and let F_q be the finite field with q elements. Let $n = em > 1$ for some positive integers e and m. Then the trace function $\mathrm{tr}_m^n(\cdot)$ is a mapping from F_{2^n} to its subfield F_{2^m} given by

$$\mathrm{tr}_m^n(x) = \sum_{i=0}^{e-1} x^{2^{mi}}.$$

It is easy to check that the trace function satisfies the following: (i) $\mathrm{tr}_m^n(ax+by) = a\,\mathrm{tr}_m^n(x) + b\,\mathrm{tr}_m^n(y)$, for all $a, b \in F_{2^m}$, $x, y \in F_{2^n}$; (ii) $\mathrm{tr}_m^n(x^{2^m}) = \mathrm{tr}_m^n(x)$, for all

$x \in F_{2^n}$; and (iii) $\mathrm{tr}_1^n(x) = \mathrm{tr}_1^m(\mathrm{tr}_m^n(x))$, for all $x \in F_{2^n}$. See [13], [14], [15], [16], [17], [18] for the detailed properties of the trace function.

Let $q = 2^L$ and α be a primitive element of F_q. Then, an m-sequence $\{s(k)\}$ of period $2^L - 1$ can be represented as

$$s(k) = \mathrm{tr}_1^L(\lambda\alpha^k), \quad k = 0, 1, 2, \ldots, 2^L - 2, \tag{10}$$

where $\lambda \neq 0$ is a fixed constant in F_q. We just give a remark that λ corresponds to the initial condition and the choice of α corresponds to the choice of a primitive polynomial as a connection polynomial when this sequence is generated using an LFSR. Any such representation, on the other hand, gives an m-sequence [17]. When $\lambda = 1$, the sequence is in the characteristic phase, and the constant-on-the-coset property can easily be checked since $s(2k) = \mathrm{tr}_1^L(\alpha^{2k}) = \mathrm{tr}_1^L(\alpha^k) = s(k)$ for all k.

B.9 Crosscorrelation Properties of M-sequences

No pair of m-sequences of the same period have the ideal crosscorrelation. The best one can achieve is a three-level crosscorrelation, and the pair of m-sequences with this property is called a *preferred pair*. Since all m-sequences of a given period are some decimations or cyclic shifts of each other, and they can all be represented as a single trace function from F_{2^L} to F_2, the crosscorrelation of a pair of m-sequences can be described as

$$R_d(\tau) = \sum_{k=0}^{2^L-1} (-1)^{\mathrm{tr}_1^L(\alpha^{k+\tau}) + \mathrm{tr}_1^L(\alpha^{dk})},$$

where d represents that the second m-sequence is a d-decimation of the first, and τ represents the amount of phase offset with each other.

Lots of values of d have been identified that result in a preferred pair of m-sequences, but it is still unknown whether we have found them all. The most famous one that gives a Gold sequence family comes from the value $d = 2^i + 1$ or $d = 2^{2i} - 2^i + 1$ for some integer i when $L/(L, i)$ is odd. Some good references on this topic are [36], [37], [38], and also Chapter 5 of [2].

B.10 Linear Complexity

Given a binary periodic sequence $\{s(k)\}$ of period P, one can always construct an LFSR that outputs $\{s(k)\}$ with a suitable initial condition. One trivial solution is the *pure cycling register* as shown in Fig. 8. It has P stages, the whole period is given as its initial condition, and the characteristic polynomial (or the connection) is given by $f(x) = x^P + 1$ corresponding to $s(k) = s(k - P)$. The best one can do is to find the LFSR with the smallest number of stages, and the linear complexity of a sequence is defined as this number. Equivalently, it is the degree of the minimal polynomial of the given sequence, which is defined as the minimum degree characteristic polynomial of the sequence.

The linear complexity of a PN sequence, in general, measures cryptographically how strong it is. It is well-known that the same copy (whole period) of a binary sequence can be generated whenever $2N$ consecutive terms or more are observed by a third party where N is the linear complexity of the sequence. This forces us to use those sequences with larger linear complexity in practice. The m-sequences are the worst in this sense because an m-sequence of period $2^L - 1$ has its linear complexity L and this number is the smallest possible over all the balanced binary sequences of period $2^L - 1$. In the following, we describe the famous Berlekamp-Massey algorithm for determining the linear complexity of a binary sequence [23].

C. Berlekamp-Massey Algorithm

Suppose we are given N terms of a sequence S, which we denote as $S^N = (s(0), s(1), \ldots, s(N-1))$. It does not necessarily mean that S has period N. The goal of the Berlekamp-Massey algorithm (BMA) is to find the minimum degree recursion satisfied by S. This minimum degree $L_N(S)$ is called the *linear complexity* of S^N. This recursion can be used to form an LFSR with $L_N(S)$-stages that generates N terms of S exactly, given the initial condition of $s(0), s(1), \ldots, s(L_N(S) - 1)$. We will denote this LFSR as $LFSR(f^{(N)}(x), L_N(S))$, where the characteristic polynomial after the N-th iteration is given by

$$f^{(N)}(x) = 1 + c_1^{(N)}x + c_2^{(N)}x^2 + \cdots + c_{L_N(S)}^{(N)}x^{L_N(S)}.$$

It is not difficult to check that (1) $L_N(S) = 0$ if and only if $s(0), s(1), \ldots, s(N-1)$ are all zeros, (2) $L_N(S) \leq N$, and (3) $L_N(S)$ must be monotonically nondecreasing with increasing N.

The BMA updates the degree $L_n(S)$ and the characteristic polynomial $f^{(n)}(x)$ for each $n = 1, 2, \ldots, N$. Assume that $f^{(1)}(x), f^{(2)}(x), \ldots, f^{(n)}(x)$ have been constructed, where the LFSR with connection $f^{(n)}(x)$ of degree $L_n(S)$ generates $s(0), s(1), \ldots, s(n-1)$. Let

$$f^{(n)}(x) = 1 + \sum_{i=1}^{L_n(S)} c_i^{(n)}x^i.$$

The next discrepancy, d_n, is the difference between $s(n)$ and the $(n+1)$-st bit generated by so far the minimal-length LFSR with $L_n(S)$ stages, and given as

$$d_n = s(n) + \sum_{i=1}^{L_n(S)} c_i^{(n)}s(n-i).$$

Let m be the sequence length before the last length change in the minimal-length register, i.e.,

$$L_m(S) < L_n(S), \quad \text{and} \quad L_{m+1}(S) = L_n(S).$$

The LFSR with the characteristic polynomial $f^{(m)}(x)$ and length $L_m(S)$ could not have generated $s(0), s(1), \ldots, s(m-1), s(m)$. Therefore, $d_m \neq 0$.

If $d_n = 0$, then this LFSR also generates the first $n+1$ bits $s(0), s(1), \ldots, s(n)$ and therefore, $L_{n+1}(S) = L_n(S)$ and $f^{(n+1)}(x) = f^{(n)}(x)$.

If $d_n \neq 0$, a new LFSR must be found to generate the first $n+1$ bits $s(0), s(1), \ldots, s(n)$. The connection polynomial and the length of the new LFSR are updated by the following:

$$\begin{aligned} f^{(n+1)}(x) &= f^{(n)}(x) - d_n d_m^{-1} x^{n-m} f^{(m)}(x), \\ L_{n+1}(S) &= \max\left[L_n(S),\ n+1-L_n(S)\right]. \end{aligned}$$

The complete BM algorithm for implementations is as follows.

(a) Initialization:

$$f(x)=1, \quad g(x)=1, \quad r=1, \quad L=0, \quad b=1, \quad n=0.$$

(b) If $n = N$, then stop. Otherwise compute

$$d = s(n) - \sum_{i=1}^{L} c_i s(n-i).$$

(c) If $d = 0$, then $r = r + 1$, and go to (f).

(d) If $d \neq 0$ and $2L > n$, then

$$f(x) = f(x) - db^{-1}x^r g(x), \quad r = r+1$$

and go to (f).

(e) If $d \neq 0$ and $2L \leq n$, then

$$h(x) = f(x),\ f(x) = f(x) - db^{-1}x^r g(x),\ L = n+1-L,\ g(x) = h(x),\ b = d,\ r = 1.$$

(f) Increase n by 1 and return to (b).

When $n = N$ and the algorithm is stopped in step (b), then the quantities produced by the algorithm bear the following relations.

$$\begin{aligned} f(x) &= f^{(N)}(x) \\ L &= L_N(S) \\ r &= N - m \\ d &= d_{N-1} \\ g(x) &= f^{(m)}(x) \\ b &= d_m \end{aligned}$$

An example of BM algorithm applied to a binary sequence of length 7 is shown in Table III.

D. Balanced Binary Sequences with the Ideal Two-Level Autocorrelation

In addition to m-sequences, there are some other well-known balanced binary sequences with the ideal two-level autocorrelation function. For period $P = 4n-1$ for some positive integer n, all these are equivalent to $(4n-1, 2n-1, n-1)$-cyclic difference sets [24].

In general, a (v, k, λ)-cyclic difference set (CDS) is a k-subset D of the integers mod v, Z_v, such that for each non-zero $z \in Z_v$ there are exactly λ ordered pairs (x, y), $x \in D$, $y \in D$ with $z = x - y \pmod{v}$ [24], [25], [26], [27], [28]. For example, $D = \{1, 3, 4, 5, 9\}$ is a $(11, 5, 2)$-CDS and every non-zero integer from 1 to 10 is represented by the difference $x - y \pmod{11}$, $x \in D$, $y \in D$, exactly twice.

The characteristic sequence $\{s(t)\}$ of a (v, k, λ)-cyclic difference set D is defined as

$$s(t) = 0 \text{ if and only if } t \in D.$$

For other values of t, the value 1 is assigned to $s(t)$. This completely characterizes binary sequences of period v with two-level autocorrelation function, and it is not difficult to check that the periodic unnormalized autocorrelation function is given as [29]

$$R(\tau) = \begin{cases} v, & \tau \equiv 0 \pmod{v} \\ v - 4(k - \lambda), & \tau \not\equiv 0 \pmod{v} \end{cases} \tag{11}$$

The out-of-phase value should be kept as low as possible in practice, and this could happen when $v = 4(k - \lambda) - 1$, resulting in the out-of-phase value to be independent of the period v. The CDS with this parameter is called a cyclic *Hadamard* difference set, and this has been investigated by many researchers [24], [25], [26], [27], [28].

In the following, we will simply summarize all the known constructions for

$(4n-1, 2n-1, n-1)$ cyclic *Hadamard* difference sets, or equivalently, balanced binary sequences of period $v = 4n-1$ with the two-level *ideal* autocorrelation function, which are also known as *Hadamard sequences* [30].

There are three types of periods currently known: (a) $v = 4n-1$ is a prime, (b) $v = 4n-1$ is a product of twin primes, and (c) $v = 4n-1$ is one less than a power of 2. All these sequences can be represented as a sum of some decimations of an m-sequence, and currently, the only open case in which such a representation is not known is Hall's Sextic Residue Sequences described below. Song and Golomb has conjectured that a Hadamard sequence of period v exists if and only if v is one of the above three types [31], and this has been confirmed for all the values of $v = 4n-1$ up to $v = 10000$ with 13 cases unsettled, the smallest of which is $v = 3439$ [30].

D.1 When $v = 4n-1$ is a prime

There are two methods in this case [29]. The first corresponds to all such values of v, and the resulting sequences are called *Legendre Sequences*. Here, D picks up integers mod v which are squares mod v. The second corresponds to some such values of v which can be represented as $4x^2 + 27$ for some integer x, and the resulting sequences are called *Hall's Sextic Residue Sequences*. Here, D picks up integers mod v which are sixth powers mod v and some others.

D.2 When $v = 4n-1$ is a product of twin primes

This is a generalization of the method for constructing Legendre sequences, and the resulting sequences are called *Twin Prime Sequences*. Let $v = p(p+1)$ where both p and $p+2$ are prime. Then, D picks up the integers d which are (i) squares both mod p and mod $p+2$, (ii) non-squares both mod p and mod $p+2$, and (iii) are 0 mod $p+2$.

D.3 When $v = 4n - 1 = 2^L - 1$

Currently, this is a most active area of research, and at least seven families are known. All the sequences of this case can best be described as a sum of some decimations of an m-sequence, or a sum of trace functions from F_{2^L} to F_2. M-sequences for all the positive integers L and GMW sequences for all the composite integers L [32], [33] have been known for many years. One important construction of a larger period from a given one is described in [34]. Recent discoveries are summarized in [35], most of which have now been proved by many others.

IV. Some Properties of FSR with Disjoint Cycles

We now return to the basic block diagram of an L-stage FSR as shown in Fig. 1, and its truth table, state transition diagram as shown in Figures 2 and 3. Unless otherwise stated, the feedback connection logic $f(x_{k-L}, x_{k-L+1}, \ldots, x_{k-1})$ of all FSRs in this section satisfy the branchless condition given in (3).

The simplest FSR with L stages is the *pure cycling register* (PCR), as shown in Fig. 8 for $L = 4$. It is linear and has the characteristic polynomial $f(x) = x^L + 1$, or the feedback connection logic $x_k = x_{k-L}$. This obviously satisfies the branchless condition (3), and the state diagram consists only of disjoint cycles. In fact, one can prove that the number $Z(L)$ of disjoint cycles of an L-stage PCR is given as

$$Z(L) = \frac{1}{L} \sum_{d|L} \phi(d) 2^{L/d}, \tag{12}$$

where $\phi(d)$ is the Euler's ϕ-function and the summation is over all the divisors of L. It is not very surprising that this number is the same as the number of irreducible polynomials of degree L over the binary alphabet F_2. Golomb had conjectured that the number of disjoint cycles from an L-stage FSR with branchless condition satisfied by its connection logic is *at most* $Z(L)$ given in (12), and this was confirmed by Mykkeltveit in 1972 [39].

On the other hand, the minimum number of disjoint cycles is 1, and this corresponds to de Bruijn sequences of period 2^L. Inserting a single 0 into any m-sequence right after the run of 0's of length $L-1$ gives a de Bruijn sequence of period 2^L. This can be done by the following modification to the linear logic f_{old} that generates the m-sequence:

$$f_{\text{new}} = f_{\text{old}} \oplus x'_{k-1}x'_{k-2}\cdots x'_{k-L+1}, \tag{13}$$

where x' represents the complement of x. A de Bruijn sequence can best be described using Good's diagram. It is shown in Fig. 9 for $L = 2$ and $L = 3$. Note that any node in a Good's diagram has two incoming edges as well as two outgoing edges. A de Bruijn sequence of period 2^L corresponds to a closed path (or a cycle) on the Good's diagram of order L, which visits every node exactly once. It was shown earlier in 1946 by de Bruijn that the number of such cycles is given as $2^{2^{L-1}-L}$ [10].

The number of disjoint cycles of an L-stage FSR with (possibly) non-linear logic connections is not completely determined. Toward this direction, we simply state a condition for the parity of this number for FSRs with branchless condition in (3). *The number of disjoint cycles of an FSR with the branchless condition is even (or odd, respectively) if and only if the number of 1's in its truth table of* $g(x_{k-1}, x_{k-2}, \ldots, x_{k-L+1})$ *is even (or odd, respectively)* [10]. In other words, the parity of the top half of the truth table is the parity of the number of disjoint cycles. This implies that PCR has an even number of disjoint cycles for all L except for $L = 2$.

In addition to PCR, there are three more degenerate cases: Complemented Cycling Registers (CCR), Pure Summing Registers (PSR), and Complemented Summing Registers (CSR). Note that PCR and PSR are homogeneous linear, but CCR and CSR are inhomogeneous linear.

$$\begin{aligned} x_k &= x_{k-L}, \quad \text{PCR}, \\ x_k &= 1 + x_{k-L}, \quad \text{CCR}, \\ x_k &= x_{k-1} + x_{k-2} + \cdots + x_{k-L} \quad \text{PSR}, \end{aligned}$$

$$x_k \;=\; 1 + x_{k-1} + x_{k-2} + \cdots + x_{k-L} \;\; \text{CSR.}$$

All these satisfy the branchless condition, and the output sequences from CCR and CSR for $L = 4$ are listed in Table IV.

The number of L-stage FSRs with branchless condition is given as $2^{2^{L-1}}$. Another condition for the truth table of an FSR to be practically useful is the so called *balanced logic* condition. A truth table of f for an L-stage FSR is said to have the balanced logic condition if f has equally many 1's and 0's in its truth table. Both the branchless condition and the balanced logic condition together imply that f has equally many 1's and 0's in its top half of the truth table. The balanced logic condition guarantees that the autocorrelation of the output sequence of period approaching 2^L tends to zero for $\tau = 1, 2, \ldots, L$.

Finally, we give a detailed analysis on the branchless 4-stage FSRs with all possible connections. With the branchless condition on it, there are $2^{2^3} = 2^8 = 256$ such FSRs. Among these, 8 FSRs are linear as shown in Figures from 5 to 8. These are collected in Table V. Except for PCR, all the other 7 LFSRs are balanced logic FSRs. Among the 248 nonlinear FSRs, there are $2^{2^3-4} = 16$ FSRs which generate de Bruijn sequences of period 16, as shown in Table VI. This table also shows all the 16 output sequences in 4 equivalent classes, denoted as A, B, C, and D. The linear complexity of each de Bruijn sequence is also shown here. The fourth and fourteenth sequences are modified versions from the sixth and eighth m-sequences in Table V, respectively. The modification is the insertion of a single 0 into the m-sequence right after the run of 0's of length 3. Note that none of the connection logic satisfy the balanced logic condition. Among the 256 FSRs with the branchless condition, there are $\binom{8}{4} = 70$ FSRs that satisfy the balanced logic condition. Table VII shows all of them. In this table, $*$ represents that it is linear and also shown in Table V.

V. Concluding Remarks

There is, in fact, a large amount of literature on FSRs, on FSR sequences, and their variations, generalizations, and applications.

Analysis of LFSR and LFSR sequences can also be done using at least two other standard methods than described in this article: the generating function approach and the matrix representation approach. See [10], [13] for the details.

There are generalizations of LFSR over a non-binary alphabet. For this, at least two operations, addition and multiplication, must be well-defined over the alphabet. The well known example of such an alphabet is a finite field with q elements. A finite commutative ring sometimes serves as an appropriate alphabet over which an LFSR is operating. Integers mod 4 is the best known in this regard due to the application of the output sequences into QPSK modulation. See [13], [38] for the details.

There are other directions to which FSR may be generalized. For example, one can consider LFSR with inputs. One application is to use the LFSR with input as a polynomial division circuit. These are used in the decoding/encoding of Hamming codes or other channel (block) codes. Another example is to use multiple (non-linear) FSRs on a stack so that the stages of an FSR in one layer are used to produce inputs to upper layer FSRs on top of it. These find some important applications to generating PN sequences with larger linear complexity in streamcipher systems.

All these topics are currently under active research, and one should look at journal transactions for the most recent results and applications.

References

[1] S. W. Golomb, *Digital Communications with Space Applications*, Prentice-Hall, Englewood Cliffs, 1964.

[2] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook*, Computer Science Press, Rockville, MD, 1985; revised edition, McGraw-Hill, 1994.

[3] R. L. Perterson, R. E. Ziemer, and D. E. Borth, *Introduction to Spread Spectrum Communications*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[4] J. G. Proakis and M Salehi, *Communication Systems Engineering*, Prentice-Hall, Englewood Cliffs, 1994.

[5] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Addison-Wesley, Reading, 1995.

[6] J. G. Proakis, *Digital Communications*, Fourth Edition, McGraw-Hill Book Company, N. Y., 2001.

[7] R. A. Rueppel, *Analysis and Design of Stream Ciphers*, Springer-Verlag, Berlin, 1986.

[8] A. J. Menezes, P. C. van Oorschot and S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996.

[9] S. B. Volchan, "What is a Random Sequence?," *Amer. Math. Monthly*, 109:46-63, 2002.

[10] S. W. Golomb, *Shift Register Sequences*, Holden-Day, San Francisco, CA, 1967; Revised Edition, Aegean Park Press, Laguna Hills, CA, 1982.

[11] D. R. Stinson, *Cryptography: Theory and Practice*, CRC Press, Boca Raton, Florida, 1995.

[12] TIA/EIA/IS-95, *Mobile Station – Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System*, published by Telecommunications Industry Association as a North American 1.5 MHz Cellular CDMA Air-Interface Standard, July 1993.

[13] R. Lidl and H. Niederreiter, *Finite Fields*, Encyclopedia of Mathematics and Its Applications, vol. 20, Addison-Wesley, Reading, MA, 1983.

[14] E. R. Berlekamp, *Algebraic Coding Theory*, McGraw-Hills, Inc., New York, 1968

[15] W. W. Peterson and E. J. Weldon, *Error-Correcting Codes*, 2nd Edition, MIT Press, Cambridge, Mass., 1972.

[16] F. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes.* North-Holland Publishing Company, 1977.

[17] R. J. McEliece, *Finite Fields for Computer Scientists and Engineers,* Kluwer Academic Publishers, 1987.

[18] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, Springer-Verlag, New York, Second Edition, 1991.

[19] T. Hansen and G. L. Mullen, "Supplement to Primitive Polynomials over Finite Fields," *Math. Comp.*, 59:S47-S50, Oct., 1992.

[20] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, Cambridge University Press, New York, 1992.

[21] H. Fredricksen, "A Survey of Full Length Nonlinear Shift Register Cycle Algorithms," *SIAM Review*, 24:195-221, 1982.

[22] S. W. Golomb, "On the Classification of Balanced Binary Sequences of Period $2^n - 1$," *IEEE Trans. Info. Theory*, 26:730-732, 1980.

[23] J. L. Massey, "Shift-Register Synthesis and BCH decoding," *IEEE Trans. Info. Theory*, 15:122-127, 1969.

[24] L. D. Baumert, *Cylic Difference Sets*, Lecture notes in mathematics, vol. 182, Springer-Verlag, New York, 1971.

[25] M. Hall Jr., “A Survey of Difference Sets,” *Proc. Amer. Math. Soc.*, 7:975-986, 1956.

[26] H. J. Ryser, *Combinatorial Mathematics.* The Carus Mathematical Monographs No. 14, Mathematical Association of America, 1963.

[27] D. Jungnickel, “Difference Sets,” in *Contemporary Design Theory*, J. H. Dinitz and D. R. Stinson, Eds., John Wiley and Sons, Inc., pp. 241-324, 1992.

[28] C. J. Colbourn and J. H. Dinitz, *The CRC Handbook of Combinatorial Designs*, CRC Press, New York, 1996.

[29] S. W. Golomb, “Construction of Signals with Favourable Correlation Properties,” in *Survey in Combinatorics*, A. D. Keedwell, Editor; LMS Lecture Note Series 166, Cambridge University Press, pp. 1-40, 1991.

[30] J.-H. Kim, *On the Hadamard Sequences*, Ph.D thesis, Yonsei University, Korea, 2001.

[31] H.-Y. Song and S. W. Golomb, “On the existence of cyclic hadamard difference sets,” *IEEE Trans. Info. Theory*, 40:1266-1268, 1994.

[32] B. Gordon, W. H. Mills, and L. R. Welch, “Some new difference sets,” *Can. J. Math.*, 14:614-625, 1962.

[33] R. A. Scholtz and L. R. Welch, “GMW Sequences,” *IEEE Trans. Info. Theory*, 30:548-553, 1984.

[34] J. S. No, H. Chung, K. Yang, and H. Y. Song, “On the Construction of Binary Sequences with Ideal Autocorrelation Property,” *IEEE International Symposium on Information Theory and Its Application*, Proceedings, pp. 837-840, held in Victoria, B.C., Canada, Sept., 1996.

[35] J.-S. No, S. W. Golomb, Guang Gong, H.-K. Lee, and Peter Gaal, “Binary pseudorandom sequences of period $2^n - 1$ with ideal autocorrelation,” *IEEE Trans. Info. Theory*, 44:814-817, 1998.

[36] D. V. Sarwate and M. B. Pursley, “Crosscorrelation Properties of Pseudorandom and Related Sequences,” *Proc. IEEE*, 68:593-619, 1980.

[37] P. Fan and M. Darnell, *Sequence Design for Communications Applications*, Research Studies Press LTD, Taunton, Somerset, England, 1995.

[38] T. Helleseth and P. V. Kumar, “Sequences with Low Correlation,” Chapter 21, *Handbook of Coding Theory*, V. S. Pless and W. C. Huffman, editors, Elsevier Science B.V. 1998.

[39] J. Mykkeltveit, “A Proof of Golomb’s Conjecture on the De Bruijn Graph,” *J. Comb. Theory, Ser. B*, 13:40-45, 1972.

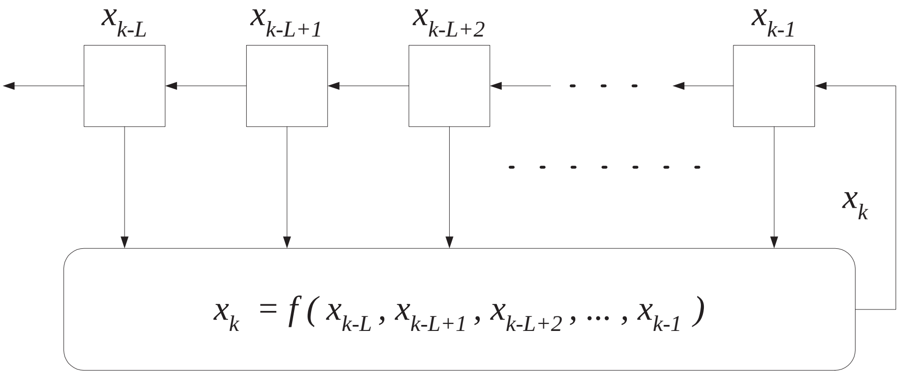
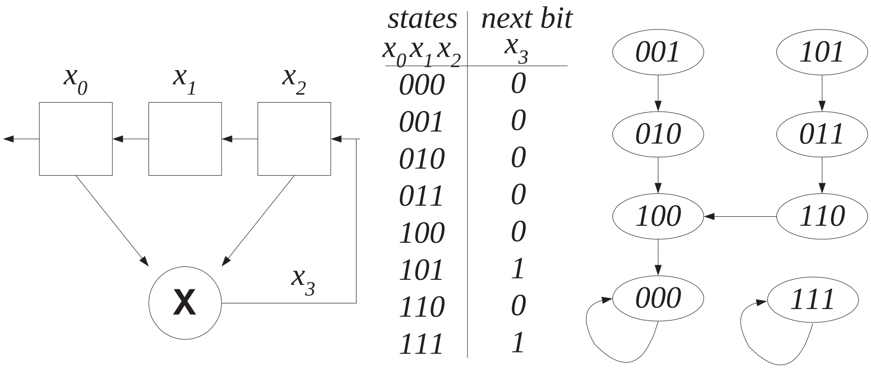
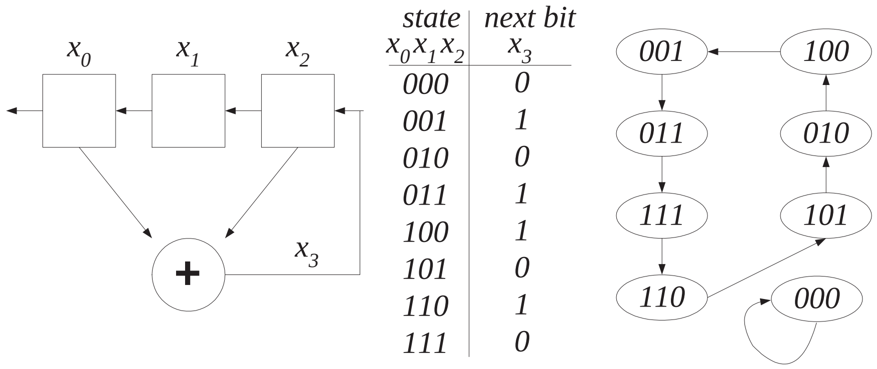


Fig. 1. An L-stage FSR with a feedback (boolean logic) function f.



states $x_0x_1x_2$	next bit x_3
000	0
001	0
010	0
011	0
100	0
101	1
110	0
111	1

Fig. 2. A 3-stage non-linear FSR with a feedback function $x_k = x_{k-1}x_{k-3}$, its truth table and state diagram.



state $x_0x_1x_2$	next bit x_3
000	0
001	1
010	0
011	1
100	1
101	0
110	1
111	0

Fig. 3. A 3-stage linear FSR with a feedback function $x_k = x_{k-1} \oplus x_{k-3}$, its truth table and state diagram.

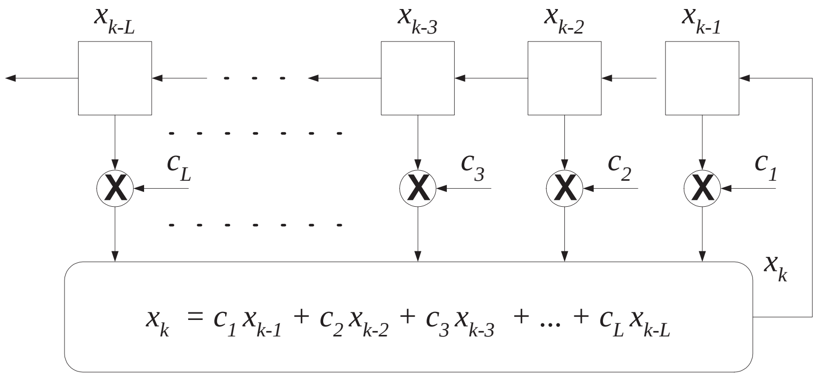
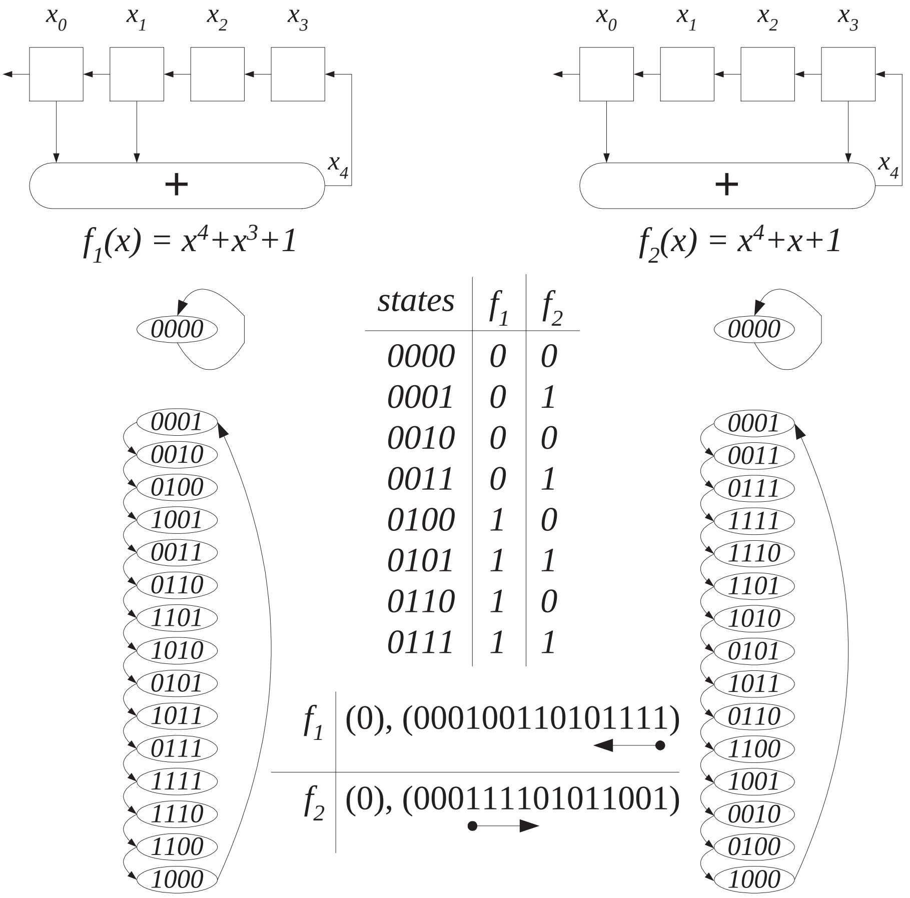

x_{k-L}
x_{k-3}
x_{k-2}
x_{k-1}
c_L
c_3
c_2
c_1
X
x_k
$x_k = c_1 x_{k-1} + c_2 x_{k-2} + c_3 x_{k-3} + \ldots + c_L x_{k-L}$

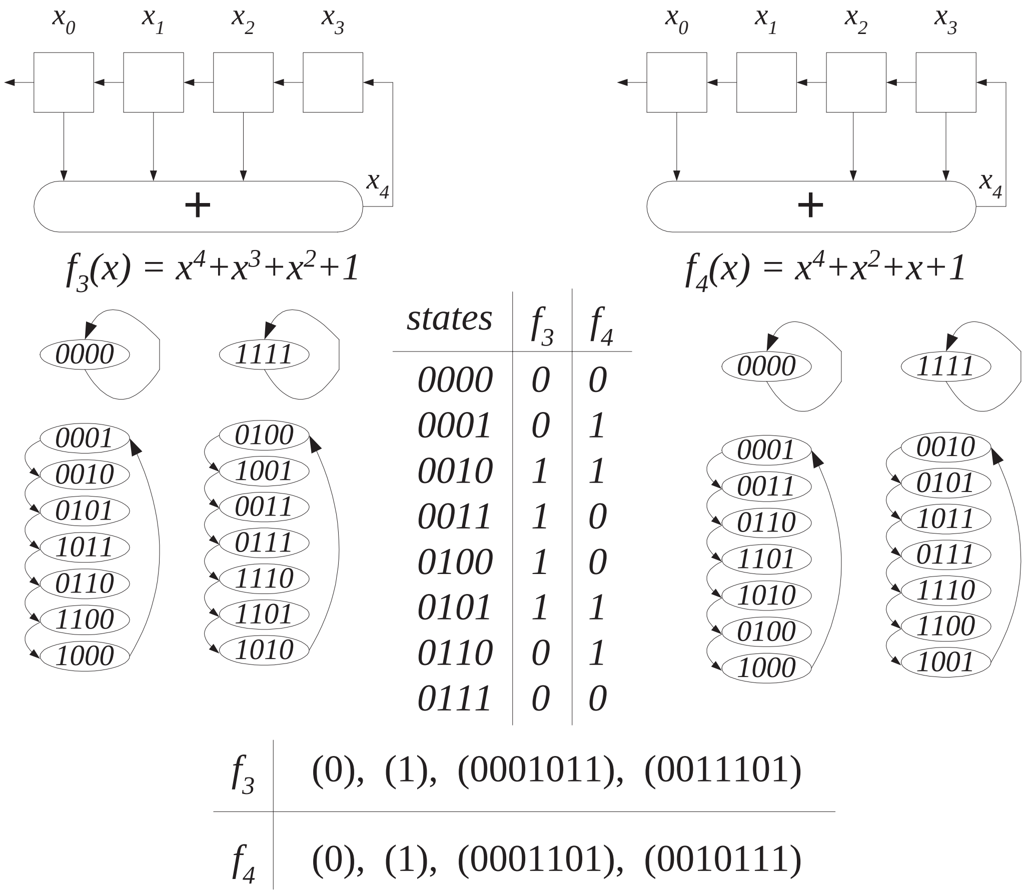
Fig. 4. An L-stage LFSR with connection coefficients $c_1, c_2, \ldots, c_L$. Note that $c_L = 1$ for LFSR to have genuinely L stages.



states	f_1	f_2
0000	0	0
0001	0	1
0010	0	0
0011	0	1
0100	1	0
0101	1	1
0110	1	0
0111	1	1

f_1	(0), (000100110101111)
f_2	(0), (000111101011001)

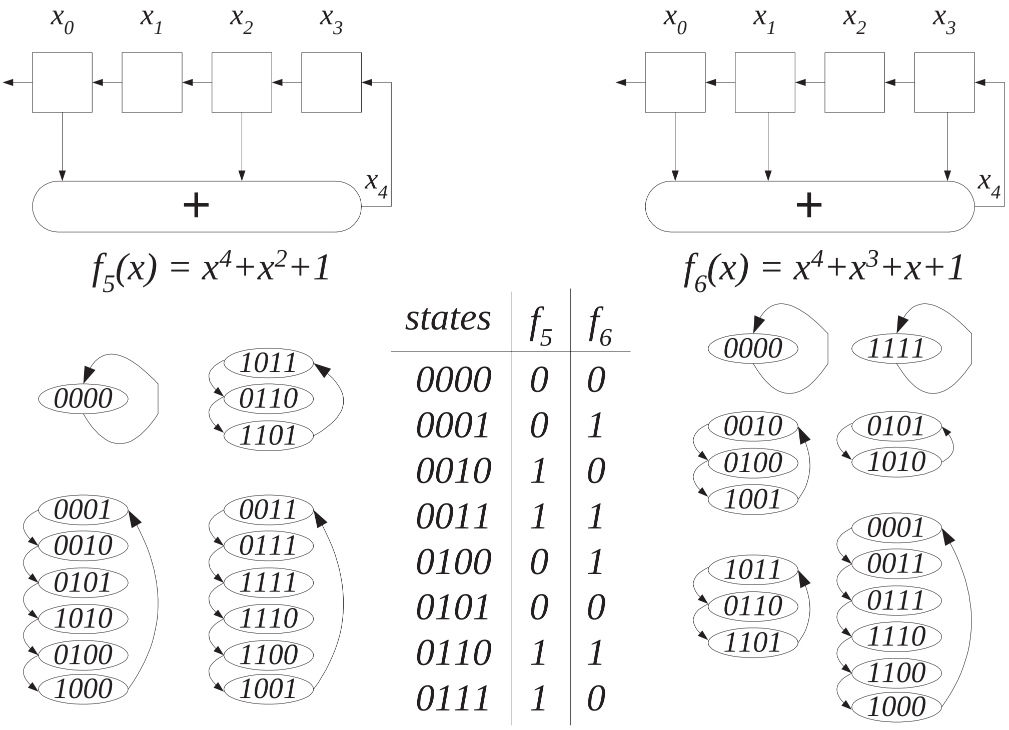
Fig. 5. Block diagrams, state diagrams, truth tables, and output sequences of 4-stage LFSRs that generate m-sequences: $f_1(x) = x^4 + x^3 + 1$ and $f_2(x) = x^4 + x + 1$.



states	f_3	f_4
0000	0	0
0001	0	1
0010	1	1
0011	1	0
0100	1	0
0101	1	1
0110	0	1
0111	0	0

f_3	(0), (1), (0001011), (0011101)
f_4	(0), (1), (0001101), (0010111)

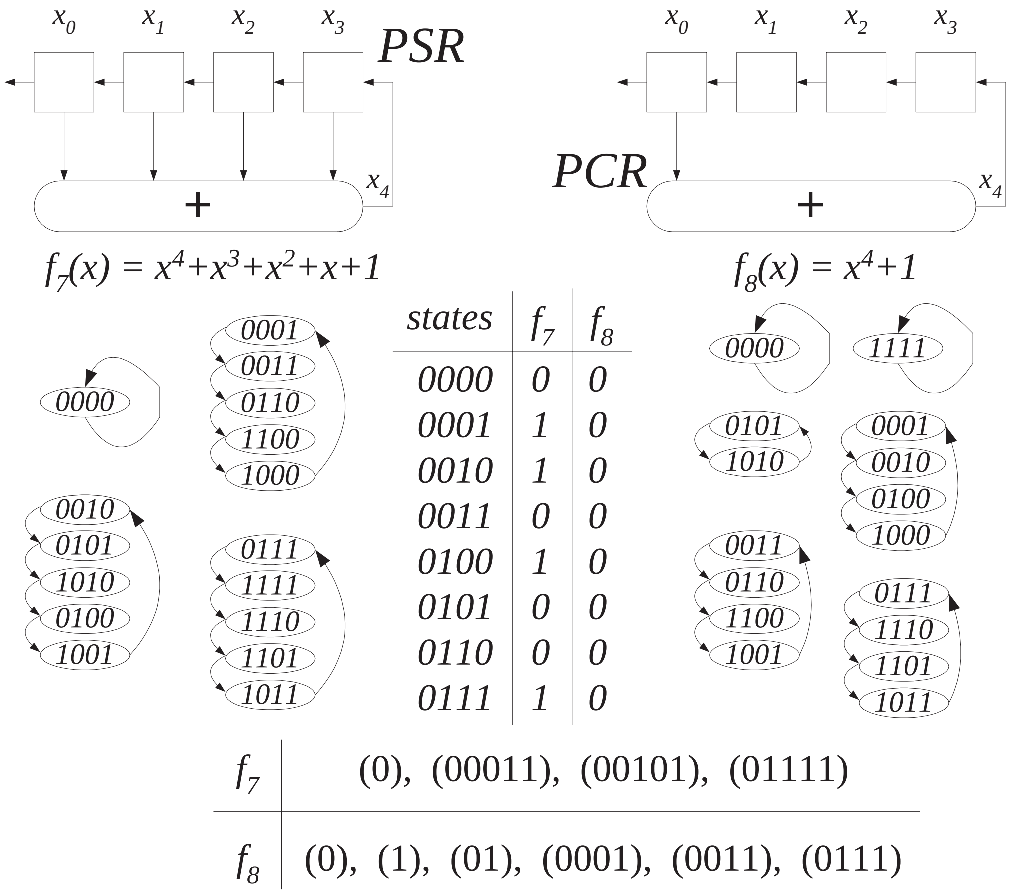
Fig. 6. Block diagrams, state diagrams, truth tables, and output sequences of 4-stage LFSRs with characteristic polynomials $f_3(x) = x^4 + x^3 + x^2 + 1$ and $f_4(x) = x^4 + x^2 + x + 1$.



states	f_5	f_6
0000	0	0
0001	0	1
0010	1	0
0011	1	1
0100	0	1
0101	0	0
0110	1	1
0111	1	0

f_5	(0), (011), (000101), (001111)
f_6	(0), (1), (01), (001), (011), (000111)

Fig. 7. Block diagrams, state diagrams, truth tables, and output sequences of 4-stage LFSRs with characteristic polynomials $f_5(x) = x^4 + x^2 + 1$ and $f_6(x) = x^4 + x^3 + x + 1$.



states	f_7	f_8
0000	0	0
0001	1	0
0010	1	0
0011	0	0
0100	1	0
0101	0	0
0110	0	0
0111	1	0

f_7	(0), (00011), (00101), (01111)
f_8	(0), (1), (01), (0001), (0011), (0111)

Fig. 8. Block diagrams, state diagrams, truth tables, and output sequences of 4-stage LFSRs with characteristic polynomials $f_7(x) = x^4 + x^3 + x^2 + x + 1$ (PSR) and $f_8(x) = x^4 + 1$ (PCR).

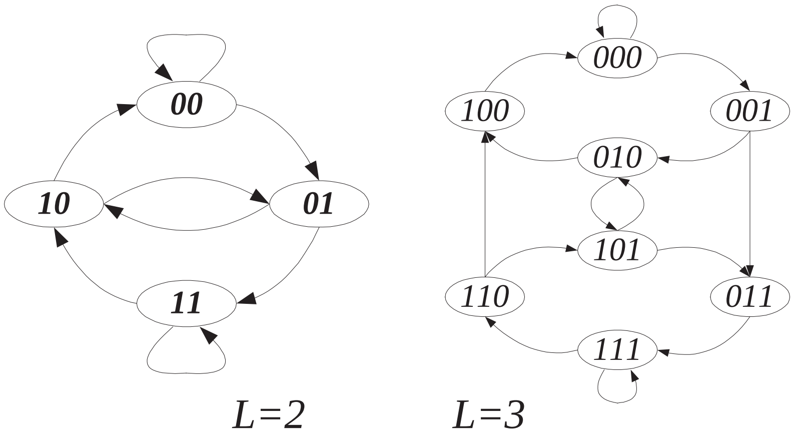


Fig. 9. Good's Diagrams for $L = 2$ and $L = 3$.

TABLE I

THE NUMBER OF PRIMITIVE IRREDUCIBLE POLYNOMIALS OF DEGREE L AND SOME EXAMPLES. THE BINARY VECTOR 1011 FOR $L = 3$ REPRESENTS EITHER $x^3 + x + 1$ OR $1 + x^2 + x^3$.

degree L	$\phi(2^L - 1)/L$	primitive polynomial
1	1	11
2	1	111
3	2	1011
4	2	10011
5	6	100101
6	6	1000011
7	18	10000011
8	16	100011101
9	48	1000010001
10	60	10000001001

TABLE II

Run distribution property of m-sequences of period $2^L - 1$.

length	number of runs of 1's	number of runs of 0's
L	1	0
$L-1$	0	1
$L-2$	1	1
$L-3$	2	2
$L-4$	4	4
$\vdots$	$\vdots$	$\vdots$
2	2^{L-4}	2^{L-4}
1	2^{L-3}	2^{L-3}
total	2^{L-3}	2^{L-3}

TABLE III

Example of BM algorithm to the sequence $(s_0, s_1, s_2, s_3, s_4, s_5, s_6) = (1,0,1,0,0,1,1)$ over F_2.

n	L	$f(x)$	r	$g(x)$	b	s_n	d
0	0	1	1	1	1	1	1
1	1	$1+x$	1	1	1	0	1
2	1	1	2	1	1	1	1
3	2	$1+x^2$	1	1	1	0	0
4	2	$1+x^2$	2	1	1	0	1
5	3	1	1	$1+x^2$	1	1	1
6	3	$1+x+x^3$	2	$1+x^2$	1	1	0
7	3	$1+x+x^3$	3	$1+x^2$	1		

TABLE IV

Output sequences from two degenerated FSRs with $L = 4$.

CCR	period	CSR	period
(00001111)	8	(1)	1
		(00001)	5
(01011010)	8	(00111)	5
		(01011)	5

TABLE V

Truth tables of all the 4-stage LFSRs, their output sequences and characteristic polynomials.

	truth table	output sequences	ch. poly.	Figures
1	00000000 11111111	0, 1, 01, 0001, 0011, 0111	$x^4 + 1$	Fig.8
2	00111100 11000011	0, 1, 0001011, 0011101	$x^4 + x^3 + x^2 + 1$	Fig.6
3	01011010 10100101	0, 1, 01, 001, 011, 000111	$x^4 + x^3 + x + 1$	Fig.7
4	01100110 10011001	0, 1, 0001101, 0010111	$x^4 + x^2 + x + 1$	Fig.6
5	01101001 10010110	0, 00011, 00101, 01111	$x^4 + x^3 + x^2 + x + 1$	Fig.8
6	01010101 10101010	0, 000111101011001	$x^4 + x + 1$	Fig.5
7	00110011 11001100	0, 011, 000101, 001111	$x^4 + x^2 + 1$	Fig.7
8	00001111 11110000	0, 000100110101111	$x^4 + x^3 + 1$	Fig.5

TABLE VI

The truth tables of all the 4-stage FSRs that generate de Bruijn sequences in 4 equivalent classes. The linear complexity of each de Bruijn sequence is also shown. The fourth and fourteenth sequences are modified versions from sixth and eighth m-sequences in Table V, respectively.

	truth table	output sequence	eq. class	LC
1	11110001 00001110	0000111101100101	A	15
2	10111001 01000110	0000101001111011	C	15
3	11100101 00011010	0000110010111101	D	14
4*	11010101 00101010	0000111101011001	A	15
5	10110101 01001010	0000101100111101	D	14
6	10101101 01010010	0000101111010011	D	14
7	10011101 01100010	0000100111101011	C	15
8	11111101 00000010	0000111101001011	B	12
9	11100011 00011100	0000110111100101	C	15
10	10101011 01010100	0000101001101111	A	15
11	11000111 00111000	0000110101111001	C	15
12	10100111 01011000	0000101111001101	D	14
13	11110111 00001000	0000111100101101	B	12
14*	10001111 01110000	0000100110101111	A	15
15	11101111 00010000	0000110100101111	B	12
16	10111111 01000000	0000101101001111	B	12

TABLE VII

Truth tables and cycle length distributions of all the 4-stage FSRs that satisfy both the branchless and balanced logic conditions. Here, $a : b$ means that there are b cycles of length a, and $*$ implies that the FSR is linear.

	truth table	cycle length distribution		truth table	cycle length distribution
1	1111000000001111	1:1, 15:1	36	1110000100011110	5:1, 11:1
2	1110100000010111	1:1, 4:1, 5:1, 6:1	37	1101000100101110	2:1, 14:1
3	1101100000100111	1:1, 2:1, 3:1, 10:1	38	1011000101001110	7:1, 9:1
4	1011100001000111	1:1, 15:1	39	0111000110001110	1:1, 15:1
5	0111100010000111	1:2, 5:1, 9:1	40	1100100100110110	2:1, 3:1, 5:1, 6:1
6	1110010000011011	1:1, 15:1	41	1010100101010110	5:1, 11:1
7	1101010000101011	1:1, 15:1	42*	0110100110010110	1:1, 5:3
8	1011010001001011	1:1, 15:1	43	1001100101100110	2:1, 14:1
9	0111010010001011	1:2, 6:1, 8:1	44	0101100110100110	1:1, 2:1, 3:1, 10:1
10	1100110000110011	1:1, 3:1, 6:2	45	0011100111000110	1:1, 15:1
11	1010110001010011	1:1, 15:1	46	1100010100111010	7:1, 9:1
12	0110110010010011	1:2, 5:1, 9:1	47	1010010101011010	4:1, 12:1
13	1001110001100011	1:1, 15:1	48	0110010110011010	1:1, 15:1
14	0101110010100011	1:2, 3:1, 11:1	49	1001010101101010	5:1, 11:1
15*	0011110011000011	1:2, 7:2	50*	0101010110101010	1:1, 15:1
16	1110001000011101	1:1, 15:1	51	0011010111001010	1:1, 15:1
17	1101001000101101	1:1, 2:1, 3:1, 10:1	52	1000110101110010	7:1, 9:1
18	1011001001001101	1:1, 3:1, 5:1, 7:1	53	0100110110110010	1:1, 3:1, 5:1, 7:1
19	0111001010001101	1:2, 3:1, 11:1	54	0010110111010010	1:1, 15:1
20	1100101000110101	1:1, 2:1, 3:1, 10:1	55	0001110111100010	1:1, 15:1
21	1010101001010101	1:1, 15:1	56	1100001100111100	2:1, 14:1
22	0110101010010101	1:2, 5:1, 9:1	57	1010001101011100	7:1, 9:1
23	1001101001100101	1:1, 2:1, 3:1, 10:1	58	0110001110011100	1:1, 15:1
24*	0101101010100101	1:2, 2:1, 3:2, 6:1	59	1001001101101100	2:1, 3:1, 5:1, 6:1
25	0011101011000101	1:2, 3:1, 11:1	60	0101001110101100	1:1, 2:1, 3:1, 10:1
26	1100011000111001	1:1, 15:1	61*	0011001111001100	1:1, 3:1, 6:2
27	1010011001011001	1:1, 15:1	62	1000101101110100	2:1, 14:1
28*	0110011010011001	1:2, 7:2	63	0100101110110100	1:1, 2:1, 3:1, 10:1
29	1001011001101001	1:1, 5:3	64	0010101111010100	1:1, 15:1
30	0101011010101001	1:2, 5:1, 9:1	65	0001101111100100	1:1, 2:1, 3:1, 10:1
31	0011011011001001	1:2, 5:1, 9:1	66	1000011101111000	5:1, 11:1
32	1000111001110001	1:1, 15:1	67	0100011110111000	1:1, 15:1
33	0100111010110001	1:2, 3:1, 11:1	68	0010011111011000	1:1, 15:1
34	0010111011010001	1:2, 6:1, 8:1	69	0001011111101000	1:1, 4:1, 5:1, 6:1
35	0001111011100001	1:2, 5:1, 9:1	70*	0000111111110000	1:1, 15:1