SEQUENCES WITH ALMOST PERFECT LINEAR COMPLEXITY PROFILE

Harald Niederreiter

Mathematical Institute, Austrian Academy of Sciences
Dr. Ignaz-Seipel-Platz 2
A-1010 Vienna, Austria

1. INTRODUCTION

Stream ciphers are based on pseudorandom key streams, i.e. on deterministically generated sequences of bits with acceptable properties of unpredictability and randomness (see [4, Ch. 9], [9]). From the cryptographic viewpoint, a useful measure for unpredictability and randomness is the linear complexity profile of a sequence (see [8], [9]). Sequences with a linear complexity profile similar to that of truly random sequences may be viewed as pseudorandom sequences. In this paper we establish connections between the linear complexity profile of a sequence and the continued fraction expansion for the generating function of the sequence and we use these connections to analyze randomness properties of the sequence.

We start with some basic definitions. If F is an arbitrary field, then a sequence $s_1, s_2, \ldots$ of elements of F is called a kth-order (linear feedback) <u>shift register sequence</u> if there exist constant coefficients $a_k, \ldots, a_0 \in F$ with $a_k \neq 0$ such that

$$a_k s_{i+k} + \ldots + a_1 s_{i+1} + a_0 s_i = 0 \quad \text{for } i = 1,2,\ldots . \tag{1}$$

The zero sequence $0,0,\ldots$ is viewed as a shift register sequence of order 0. A shift register sequence is uniquely determined by the recursion (1) and by the initial values $s_1, s_2, \ldots, s_k$.

Definition 1. Let $s_1, s_2, \ldots$ be an arbitrary sequence of elements of the field F and let n be a positive integer. Then the linear complexity $L(n)$ is defined as the least k such that $s_1, s_2, \ldots, s_n$ form the first n terms of a kth-order shift register sequence.

Definition 2. With the notation of Definition 1, the sequence $L(1), L(2), \ldots$ is called the linear complexity profile of the sequence $s_1, s_2, \ldots$.

D. Chaum and W.L. Price (Eds.): Advances in Cryptology - EUROCRYPT '87, LNCS 304, pp. 37-51, 1988.

It is clear that $0 \leq L(n) \leq n$ and $L(n) \leq L(n+1)$ for all $n \geq 1$. Therefore the linear complexity profile is a nondecreasing sequence of nonnegative integers. The linear complexity $L(n)$ can be calculated by the Berlekamp-Massey algorithm (see [3, Ch. 8], [4, Ch. 6], [5]). In the applications to stream ciphers the field F is usually the binary field, i.e. the finite field with two elements.

In Section 2 we discuss formal power series viewed as generating functions of sequences. The core of the paper is Section 3 where it is shown that the linear complexity profile of a sequence can be described in terms of the continued fraction expansion for its generating function. This yields in particular a new approach to the Berlekamp-Massey algorithm. In Section 4 we show how this approach leads to a fairly simple proof for the characterization given by Wang and Massey [11] of binary sequences with a perfect linear complexity profile. In Section 5 we discuss sequences with almost perfect linear complexity profile and the construction of sequences with a prescribed linear complexity profile.

2. GENERATING FUNCTIONS

If $s_1, s_2, \ldots$ is an arbitrary sequence of elements of the field F, then we associate with it the formal power series

$$S = \sum_{i=1}^{\infty} s_i x^{-i}$$

in x^{-1} as its generating function. We view S as an element of the field $G = F((x^{-1}))$ of formal Laurent series over F in x^{-1}. The field G consists of the elements

$$S = \sum_{i=r}^{\infty} s_i x^{-i} ,$$

where all $s_i \in F$ and r is an arbitrary integer (positive, negative, or 0); the algebraic operations in G are defined in the usual way. The field G contains the field $F(x)$ of rational functions over F as a subfield, and $F(x)$ is the quotient field of the polynomial ring $F[x]$. Elements of G that belong to $F(x)$ are called rational, all other elements of G are called irrational.

Shift register sequences can easily be characterized in this context. If $s_1, s_2, \ldots$ is a shift register sequence satisfying (1), then the polynomial

$$f(x) = a_k x^k + \ldots + a_1 x + a_0 \in F[x] \qquad (2)$$

is called a characteristic polynomial of the sequence (or of the recursion (1)). By definition, a nonzero constant polynomial is also viewed as a characteristic polynomial of the zero sequence. As usual we define $\deg(0) = -\infty$.

Lemma 1. Let $f \in F[x]$ be a nonzero polynomial. Then the sequence $s_1, s_2, \ldots$ of elements of F is a shift register sequence with characteristic polynomial f if and only if

$$\sum_{i=1}^{\infty} s_i\, x^{-i} = \frac{g(x)}{f(x)}$$

with $g \in F[x]$ and $\deg(g) < \deg(f)$.

Proof. This is trivial if $\deg(f) = 0$. If f is as in (2) with $\deg(f) = k \geq 1$, then consider

$$f(x) \sum_{i=1}^{\infty} s_i\, x^{-i} = (a_k\, x^k + \ldots + a_1\, x + a_0)(s_1\, x^{-1} + s_2\, x^{-2} + \ldots) .$$

The right-hand side is a polynomial of degree $< k$ if and only if the coefficient of each x^j with $j < 0$ vanishes, which means that

$$a_k\, s_{-j+k} + \ldots + a_1\, s_{-j+1} + a_0\, s_{-j} = 0 \quad \text{for all } j < 0 .$$

But this is equivalent to the validity of (1). □

Therefore shift register sequences are exactly those sequences whose generating functions are rational elements of G. The rational generating function of a shift register sequence has a reduced form g/m with $g, m \in F[x]$, $\deg(g) < \deg(m)$, and $\gcd(g,m) = 1$. The polynomial m is called a minimal polynomial of the shift register sequence. It can also be described as a characteristic polynomial of the shift register sequence of least degree. If one assumes that the minimal polynomial is monic, as is often done (see e.g. [3, Ch. 8]), then the minimal polynomial is uniquely determined; we will not make this assumption. The following lemma follows immediately from Lemma 1 and the definition of a minimal polynomial.

Lemma 2. Let $m \in F[x]$ be a nonzero polynomial. Then the sequence $s_1, s_2, \ldots$ of elements of F is a shift register sequence with minimal polynomial m if and only if

$$\sum_{i=1}^{\infty} s_i\, x^{-i} = \frac{g(x)}{m(x)}$$

with $g \in F[x]$, $\deg(g) < \deg(m)$, and $\gcd(g,m) = 1$.

If $s_1, s_2, \ldots$ is an arbitrary sequence of elements of F and n is a positive integer, then consider the least k such that $s_1, s_2, \ldots, s_n$ form the first n terms of a kth-order shift register sequence. A minimal polynomial m_n of such a kth-order shift register sequence is called an nth minimal polynomial of the sequence $s_1, s_2, \dots$. It follows from the definition of linear complexity that $L(n) = \deg(m_n)$ for all n.

3. LINEAR COMPLEXITY AND CONTINUED FRACTIONS

Let F be an arbitrary field and let the field $G = F((x^{-1}))$ be as in Section 2. Every $S \in G$ has a unique continued fraction expansion

$$S = A_0 + 1/(A_1 + 1/(A_2 + \dots)) =: [A_0, A_1, A_2, \dots] ,$$

where the A_j, $j \geqq 0$, are polynomials over F and $\deg(A_j) \geqq 1$ for $j \geqq 1$. This expansion is finite for rational S and infinite for irrational S. For $S = \sum_{i=r}^{\infty} s_i x^{-i} \in G$ we define its polynomial part by

$$\mathrm{Pol}(S) = \sum_{i=r}^{0} s_i x^{-i} .$$

Then the polynomials A_j are obtained recursively by the following algorithm:

$$A_0 = \mathrm{Pol}(S), \qquad B_0 = S - \mathrm{Pol}(S),$$
$$A_{j+1} = \mathrm{Pol}(B_j^{-1}), \qquad B_{j+1} = B_j^{-1} - \mathrm{Pol}(B_j^{-1}) \qquad \text{for } j \geqq 0 ,$$

which can be continued as long as $B_j \neq 0$. If the continued fraction expansion is broken off after the term A_j, $j \geqq 0$, we get the rational convergent P_j/Q_j. The polynomials P_j and Q_j can be calculated recursively by

$$P_{-1} = 1, \quad P_0 = A_0, \quad P_j = A_j P_{j-1} + P_{j-2} \qquad \text{for } j \geqq 1 ,$$
$$Q_{-1} = 0, \quad Q_0 = 1, \quad Q_j = A_j Q_{j-1} + Q_{j-2} \qquad \text{for } j \geqq 1 .$$

The following formulas are shown by straightforward induction on j:

$$\deg(Q_j) = \sum_{h=1}^{j} \deg(A_h) \qquad \text{for } j \geqq 1 , \tag{3}$$

$$P_{j-1} Q_j - P_j Q_{j-1} = (-1)^j \qquad \text{for } j \geqq 0 , \tag{4}$$

$$S = \frac{P_j + B_j P_{j-1}}{Q_j + B_j Q_{j-1}} \qquad \text{for } j \geqq 0 . \tag{5}$$

From (4) we get $\gcd(P_j, Q_j) = 1$ for $j \geqq 0$. For rational S we interpret $\deg(A_j) = \deg(Q_j) = \infty$ whenever A_j and Q_j do not exist.

It is convenient to introduce the (exponential) valuation v on G which extends the degree function on $F[x]$. For $S \in G$, $S \neq 0$, we put

$$v(S) = -r \quad \text{if} \quad S = \sum_{i=r}^{\infty} s_i x^{-i} \quad \text{and} \quad s_r \neq 0 .$$

For $S = 0$ we put $v(S) = -\infty$. We have the following properties for $S, T \in G$:

$$v(ST) = v(S) + v(T) ,$$
$$v(S + T) \leqq \max(v(S), v(T)),$$
$$v(S + T) = \max(v(S), v(T)) \quad \text{if} \quad v(S) \neq v(T) .$$

For $f \in F[x]$ we have $v(f) = \deg(f)$, hence

$$v(\frac{f}{g}) = \deg(f) - \deg(g) \quad \text{for} \quad f, g \in F[x], g \neq 0.$$

From (3), (4), (5), and the properties of v it follows easily that

$$v(Q_j S - P_j) = - v(Q_{j+1}) \quad \text{for} \quad j \geqq 0 . \tag{6}$$

Lemma 3. If $f, g \in F[x]$ are such that $v(fS-g) < 0$, then

$$f = \sum_{h=0}^{j} C_h Q_h \quad \text{and} \quad g = \sum_{h=0}^{j} C_h P_h$$

for some $j \geqq 0$ and $C_h \in F[x]$ with $\deg(C_h) < \deg(A_{h+1})$ for $0 \leqq h \leqq j$. If in addition $f \neq 0$, then

$$v(fS-g) = \deg(C_i) - \deg(Q_{i+1}) ,$$

where i is the least index with $C_i \neq 0$.

Proof. Using (3) we see that every $f \in F[x]$ can be represented in the indicated form. By (6) we have

$$v(C_h(Q_h S - P_h)) = v(C_h) - v(Q_{h+1}) < 0 \quad \text{for} \quad 0 \leqq h \leqq j ,$$

hence

$$v(\sum_{h=0}^{j} C_h(Q_h S - P_h)) < 0 .$$

Using $v(fS-g) < 0$ we get

$$v(\sum_{h=0}^{j} C_h P_h - g) = v(fS - g - \sum_{h=0}^{j} C_h(Q_h S - P_h)) < 0 .$$

But $\sum_{h=0}^{j} C_h P_h - g$ is a polynomial, hence it must be 0. To prove the second part, we note that if $f \neq 0$, then there exists a least index i with $C_i \neq 0$. From (6) we get

$$v(C_i(Q_i\ S - P_i)) = v(C_i) - v(Q_{i+1}) \geqq - v(Q_{i+1}) ,$$

$$v(C_h(Q_h\ S - P_h)) < \deg(A_{h+1}) - v(Q_{h+1}) = - v(Q_h) \leqq - v(Q_{i+1}) \qquad \text{for } i < h \leqq j ,$$

hence

$$v(fS-g) = v(\sum_{h=i}^{j} C_h(Q_h\ S - P_h)) = v(C_i(Q_i\ S - P_i)) = v(C_i) - v(Q_{i+1}). \ \Box$$

We can now establish the main result of this section which gives an explicit description of the linear complexity profile of a sequence in terms of the polynomials Q_j that are obtained as above from the continued fraction expansion for the generating function S. Note that for generating functions S we always have $A_0 = \mathrm{Pol}(S) = 0$.

Theorem 1. Let $s_1, s_2, \ldots$ be an arbitrary sequence of elements of F and let $S = \sum_{i=1}^{\infty} s_i\ x^{-i}$ be its generating function. Then for any $n \geqq 1$ the linear complexity $L(n)$ of $s_1, s_2, \ldots$ is given by

$$L(n) = \deg(Q_j) ,$$

where $j \geqq 0$ is uniquely determined by the condition

$$\deg(Q_{j-1}) + \deg(Q_j) \leqq n < \deg(Q_j) + \deg(Q_{j+1}) .$$

Furthermore, the nth minimal polynomials of $s_1, s_2, \ldots$ are exactly all polynomials m_n of the form

$$m_n = aQ_j + CQ_{j-1} ,$$

where $a \in F$, $a \neq 0$, and $C \in F[x]$ with $\deg(C) \leqq 2 \deg(Q_j) - n - 1$.

Proof. Write $q_j = \deg(Q_j)$ for $j \geqq 0$ and $q_{-1} = 0$. For $j \geqq 0$ we get from (6),

$$v(S - \frac{P_j}{Q_j}) = - v(Q_j) - v(Q_{j+1}) = - q_j - q_{j+1} .$$

Hence if

$$\frac{P_j}{Q_j} = \sum_{i=1}^{\infty} t_i\ x^{-i} ,$$

then $t_i = s_i$ for $1 \leq i < q_j + q_{j+1}$. Since $\deg(P_j) < \deg(Q_j)$ and $\gcd(P_j, Q_j) = 1$, it follows from Lemma 2 that $t_1, t_2, \ldots$ is a shift register sequence with minimal polynomial Q_j. Therefore

$$L(n) \leq q_j \quad \text{for} \quad 1 \leq n < q_j + q_{j+1} \, . \tag{7}$$

Now let $q_{j-1} + q_j \leq n < q_j + q_{j+1}$. By the definition of $L(n)$ there exists a shift register sequence $u_1, u_2, \ldots$ with minimal polynomial $m_n \in F[x]$ of degree $L(n)$ and $u_i = s_i$ for $1 \leq i \leq n$. By Lemma 2 the generating function of $u_1, u_2, \ldots$ has the form g_n/m_n with $g_n \in F[x]$. Using (7) we get

$$v(m_n\, S - g_n) = L(n) + v(S - \frac{g_n}{m_n}) \leq L(n) - n - 1 < 0 \, . \tag{8}$$

Since $m_n \neq 0$, Lemma 3 yields the formulas

$$m_n = \sum_{h=i}^{j} C_h\, Q_h \quad \text{and} \quad g_n = \sum_{h=i}^{j} C_h\, P_h \quad \text{with} \quad C_i \neq 0 \, ,$$

$$v(m_n\, S - g_n) = \deg(C_i) - q_{i+1} \, . \tag{9}$$

Together with (8) this implies

$$L(n) \geq n + 1 + \deg(C_i) - q_{i+1} \geq q_{j-1} + q_j + 1 - q_{i+1} \, .$$

In view of (7) this is only possible if $i = j$ or $i = j - 1$. If $i = j$, then $m_n = aQ_j$ with $a \in F$, $a \neq 0$, and $L(n) = q_j$. If $i = j - 1$, then $m_n = C_jQ_j + C_{j-1}Q_{j-1}$ and

$$L(n) \geq n + 1 + \deg(C_{j-1}) - q_j \geq q_{j-1} + \deg(C_{j-1}) + 1 \, .$$

This shows that $C_j \neq 0$, hence $C_j = a \in F$, $a \neq 0$, and $L(n) = q_j$. Furthermore, (8) and (9) imply

$$\deg(C_{j-1}) - q_j = v(m_n\, S - g_n) \leq q_j - n - 1 \, ,$$

hence $\deg(C_{j-1}) \leq 2q_j - n - 1$.

We have now shown that an nth minimal polynomial m_n is necessarily of the form $m_n = aQ_j + CQ_{j-1}$ given in the theorem. It remains to prove that any such polynomial can serve as an nth minimal polynomial. Put $g_n = aP_j + CP_{j-1}$. Then $\deg(g_n) < \deg(m_n)$ and

$$P_{j-1}\, m_n - Q_{j-1}\, g_n = (-1)^j\, a$$

by (4), so that $\gcd(g_n, m_n) = 1$. By Lemma 2, m_n is a minimal polynomial of the shift register sequence $u_1, u_2, \ldots$ with generating function g_n/m_n. If $C \neq 0$, then by (9)

$$v(S - \frac{g_n}{m_n}) = v(m_n S - g_n) - v(m_n) = \deg(C) - 2q_j \leqq - n - 1 ,$$

and if $C = 0$, then

$$v(S - \frac{g_n}{m_n}) = v(m_n S - g_n) - v(m_n) = - q_{j+1} - q_j \leqq - n - 1 .$$

Thus $u_i = s_i$ for $1 \leqq i \leqq n$, and so m_n is an nth minimal polynomial of $s_1, s_2, \ldots$. □

We remark that in the usual situation in which the Berlekamp-Massey algorithm is applied we have a finite field F and $n = 2 \deg(m_n) = 2 \deg(Q_j)$. In this case, Theorem 1 shows that the nth minimal polynomials are exactly given by $m_n = aQ_j$ with $a \in F$, $a \neq 0$. If one wants m_n to be monic, then a is uniquely determined.

Connections between the Berlekamp-Massey algorithm, the continued fraction algorithm, and the Euclidean algorithm have already been observed earlier (see e.g. the references in [3, p. 530] and the recent work of Dai and Wan [2]), but Theorem 1 seems to give the most transparent connection between these algorithms.

4. PERFECT LINEAR COMPLEXITY PROFILE

Rueppel [8], [9, Ch. 4] has shown that for random sequences of bits (viewed as elements of the binary field F_2) the expected value of the linear complexity $L(n)$ is $\frac{n}{2} + c_n$ with $0 \leqq c_n \leqq \frac{5}{18}$. This has given rise to the notion of a sequence with a perfect linear complexity profile: this is a sequence whose linear complexity profile follows the expected behavior of random sequences as closely as possible. Extending this notion to arbitrary fields, we get the following definition. We write $\lfloor t \rfloor$ for the greatest integer $\leqq t$.

Definition 3. A sequence $s_1, s_2, \ldots$ of elements of a field F is said to have a perfect linear complexity profile (PLCP) if

$$L(n) = \lfloor \frac{n+1}{2} \rfloor \qquad \text{for all } n \geqq 1 .$$

Theorem 2. The sequence $s_1, s_2, \ldots$ of elements of F has a PLCP if and only if its generating function $S = \sum_{i=1}^{\infty} s_i x^{-i}$ is irrational and has a continued fraction expansion

$$S = [0, A_1, A_2, \ldots]$$

with $\deg(A_j) = 1$ for all $j \geqq 1$.

Proof. If S is irrational and $\deg(A_j) = 1$ for all $j \geqq 1$, then by (3) we have $\deg(Q_j) = j$ for all $j \geqq 0$, and so Theorem 1 shows immediately that the sequence has a PLCP. Conversely, if the sequence has a PLCP, then $\lim\limits_{n\to\infty} L(n) = \infty$ implies that S is irrational. If we had $\deg(A_j) > 1$ for some $j \geqq 1$, then with $n = \deg(Q_{j-1}) + \deg(Q_j) = 2\deg(Q_j) - \deg(A_j)$ we get from Theorem 1

$$L(n) = \deg(Q_j) > \lfloor \tfrac{n+1}{2} \rfloor ,$$

which is a contradiction. □

Theorem 2 was shown for the binary field in Niederreiter [6] and for arbitrary finite fields in Niederreiter [7]. In the important special case where F is the binary field F_2 with two elements, Wang and Massey [11] have given a more explicit characterization of sequences with a PLCP. It was already pointed out in Niederreiter [6] that this characterization follows also from Theorem 2 and a result of Baum and Sweet [1]. We show now that the Wang-Massey characterization can be deduced from Theorem 2 by a fairly simple argument that uses an idea of Taussat [10].

Theorem 3 (Wang and Massey [11]). The sequence $s_1, s_2, \ldots$ of elements of F_2 has a PLCP if and only if it satisfies

$$s_1 = 1 \quad \text{and} \quad s_{2i+1} = s_{2i} + s_i \qquad \text{for } i = 1,2,\ldots .$$

Proof. For $T \in G = F_2((x^{-1}))$ put

$$D(T) = x^{-1} T^2 + (1 + x^{-1}) T + x^{-1} \in G .$$

We have the following properties:

$$D(T + U + V) = D(T) + D(U) + D(V) \qquad \text{for } T, U, V \in G ,$$

$$D(T^{-1}) = D(T)\, T^{-2} \qquad \text{for } T \in G,\ T \neq 0 ,$$

$$D(x) + D(c) = c + 1 \qquad \text{for } c \in F_2 .$$

If $s_1, s_2, \ldots$ has a PLCP, then by Theorem 2 its generating function S has an infinite continued fraction expansion $S = [0, A_1, A_2, \ldots]$, where $A_j = x + a_j$ with $a_j \in F_2$ for all $j \geqq 1$. For $j \geqq 0$ we have by the continued fraction algorithm $B_j^{-1} = A_{j+1} + B_{j+1}$, hence

$$D(B_j)\, B_j^{-2} = D(B_j^{-1}) = D(x + a_{j+1} + B_{j+1}) = a_{j+1} + 1 + D(B_{j+1}) .$$

By induction on j and using $S = B_0$ this yields

$$D(S) = \sum_{i=1}^{j} (a_i + 1)\, B_0^2\, B_1^2 \ldots B_{i-1}^2 + B_0^2\, B_1^2 \ldots B_{j-1}^2\, D(B_j) \qquad \text{for } j \geqq 1 .$$

We have $v(B_j) < 0$ and $v(D(B_j)) < 0$ for $j \geqq 0$, thus

$$\lim_{j\to\infty} B_0^2\, B_1^2 \ldots B_{j-1}^2\, D(B_j) = 0$$

in the topology on G induced by the valuation v. Hence

$$D(S) = \sum_{i=1}^{\infty} (a_i + 1)\, B_0^2\, B_1^2 \ldots B_{i-1}^2 = (\sum_{i=1}^{\infty} (a_i + 1)\, B_0\, B_1 \ldots B_{i-1})^2 = : U^2$$

with $U \in G$, $v(U) < 0$. By the definition of $D(S)$ we obtain

$$S^2 + (x+1)\, S + 1 = xU^2 . \tag{10}$$

Comparing constant terms we get $s_1 = 1$ and comparing the coefficients of x^{-2i} for $i \geqq 1$ we get $s_i + s_{2i+1} + s_{2i} = 0$.

Conversely, if $s_1 = 1$ and $s_{2i+1} = s_{2i} + s_i$ for $i \geqq 1$, then we have (10) with a suitable $U \in G$, $v(U) < 0$. If S is either rational or $\deg(A_j) > 1$ for some $j \geqq 1$, then by (6) there is a $j \geqq 0$ with

$$v(Q_j\, S - P_j) < - v(Q_j) - 1 .$$

It follows that

$$v(Q_j^2\, S^2 - P_j^2) = v((Q_j\, S - P_j)^2) < 0 ,$$

$$v((x+1)\, Q_j\, (Q_j\, S - P_j)) = 1 + v(Q_j) + v(Q_j\, S - P_j) < 0 .$$

Therefore

$$T: = P_j^2 + (x+1)\, P_j\, Q_j + Q_j^2 + xU^2\, Q_j^2 = Q_j^2\, S^2 - P_j^2 + (x+1)\, Q_j\, (Q_j\, S - P_j)$$

satisfies $v(T) < 0$. In particular, the constant term of T is 0. With $a = P_j(0)$, $b = Q_j(0)$ this means that $a + ab + b = 0$. In F_2 this is only possible if $a = b = 0$, and this is a contradiction to $\gcd(P_j, Q_j) = 1$. Therefore S satisfies the conditions in Theorem 2 and $s_1, s_2, \ldots$ has a PLCP. □

5. ALMOST PERFECT LINEAR COMPLEXITY PROFILE

Theorem 3 shows that, at least in the field F_2, sequences with a PLCP are far from being unpredictable. Indeed this result says that in a sequence of bits with a PLCP every second term depends in a known manner on previous terms, an untenable state of affairs in a purportedly pseudorandom sequence.

The moral of this is that requiring a PLCP is too restrictive a condition. Therefore one should consider sequences for which somewhat larger deviations of the linear complexity $L(n)$ from its expected value (which is about $\frac{n}{2}$) are allowed. Such sequences may be called sequences with an almost perfect linear complexity profile. The deviation of $L(n)$ from $\frac{n}{2}$ can be controlled by the following simple consequence of Theorem 1.

Theorem 4. Let $s_1, s_2, \ldots$ be an arbitrary sequence of elements of the field F. Then, with the notation of Theorem 1, for any $n \geqq 1$ we have

$$\frac{n+1}{2} - \frac{1}{2}\deg(A_{j+1}) \leqq L(n) \leqq \frac{n}{2} + \frac{1}{2}\deg(A_j)$$

with the interpretation $\deg(A_0) = -1$, where $j \geqq 0$ is uniquely determined by the condition

$$\deg(Q_{j-1}) + \deg(Q_j) \leqq n < \deg(Q_j) + \deg(Q_{j+1}) \; .$$

Proof. By (3) the bounds for n can be rewritten in the form

$$2\deg(Q_j) - \deg(A_j) \leqq n \leqq 2\deg(Q_j) + \deg(A_{j+1}) - 1 \; ,$$

which checks also for $j = 0$. This is equivalent to

$$\frac{n+1}{2} - \frac{1}{2}\deg(A_{j+1}) \leqq \deg(Q_j) \leqq \frac{n}{2} + \frac{1}{2}\deg(A_j) \; .$$

Since $L(n) = \deg(Q_j)$ by Theorem 1, the desired result is shown. □

It is convenient to use a quantity introduced in Niederreiter [6], [7]. For an irrational generating function $S = \sum_{i=1}^{\infty} s_i x^{-i} \in F((x^{-1}))$ let

$$S = [0, A_1, A_2, \ldots]$$

be its continued fraction expansion. Then we define

$$K(S) = \sup_{j \geqq 1} \deg(A_j) \; ,$$

where we can have $K(S) = \infty$. Our interest will, however, be in the case where

$K(S) < \infty$. We note that by Theorem 2 the sequence $s_1, s_2, \ldots$ has a PLCP if and only if S is irrational and $K(S) = 1$. The following is an immediate consequence of Theorem 4.

Corollary 1. Let $s_1, s_2, \ldots$ be a sequence of elements of F which has an irrational generating function S (or, equivalently, which is not a shift register sequence). Then

$$\frac{1}{2}(n + 1 - K(S)) \leq L(n) \leq \frac{1}{2}(n + K(S)) \qquad \text{for all } n \geq 1 .$$

The result of Corollary 1 was shown for $F = F_2$ in Niederreiter [6] and for arbitrary finite fields in Niederreiter [7]. If $K(S) < \infty$, then the bounds for $L(n)$ in Corollary 1 are in general best possible. Choose $j \geq 1$ with $\deg(A_j) = K(S)$ and put $n = \deg(Q_{j-1}) + \deg(Q_j)$, then $L(n)$ equals the upper bound by Theorem 1. Choose $j \geq 1$ such that $\deg(A_j) = K(S)$ and $n = \deg(Q_{j-1}) + \deg(Q_j) - 1 \geq 1$, then $L(n)$ equals the lower bound by Theorem 1.

Sequences with an almost perfect linear complexity profile can be constructed on the basis of Corollary 1, by considering irrational generating functions S with a relatively small $K(S) > 1$. The precise information on the linear complexity profile given in Theorem 1 can be used to produce a desired pattern in the linear complexity profile by an appropriate choice of the polynomials $A_1, A_2, \ldots$. By a formula of Rueppel [8], [9, Ch. 4], the standard deviation of $L(n)$ for random sequences of bits is asymptotically equal to $\frac{1}{9}\sqrt{86} = 1.03\ldots$. Hence by Corollary 1, choosing an irrational generating function $S \in F_2((x^{-1}))$ with $K(S) = 2$ or 3 would be roughly in accordance with the asymptotic behavior of the standard deviation of $L(n)$.

Sequences of elements of an arbitrary field F with a prescribed linear complexity profile can be constructed by an appropriate choice of the polynomials $A_1, A_2, \ldots$ determining the continued fraction expansion for the generating function $S = \sum_{i=1}^{\infty} s_i x^{-i}$. If we have chosen the j polynomials $A_1, A_2, \ldots, A_j$ for some $j \geq 1$, then it is important to know to what extent this determines the sequence $s_1, s_2, \ldots$. We note that by (6) we have

$$v(S - \frac{P_j}{Q_j}) = - \deg(Q_j) - \deg(Q_{j+1}) . \tag{11}$$

Therefore all the terms s_i with

$$1 \leq i \leq \deg(Q_j) + \deg(Q_{j+1}) - 1 ,$$

and so at least all the terms with $1 \leq i \leq 2 \deg(Q_j)$, are determined by $A_1, A_2, \ldots, A_j$ and they are not affected by later choices of $A_{j+1}, A_{j+2}, \ldots$. These terms agree with the corresponding terms in the power series expansion of P_j/Q_j. The polynomials P_j and Q_j are easily obtained by the recursions in Section 3 with $A_0 = 0$. The calculation of the coefficients in the expansion of P_j/Q_j is facilitated by the fact that these coefficients form a shift register sequence with minimal polynomial Q_j (by Lemma 2). Therefore we only need the $\deg(Q_j)$ initial values of this shift register sequence, the remaining terms can be calculated quickly by the recursion of the form (1) with characteristic polynomial Q_j. But since

$$\deg(Q_j) \leqq \deg(Q_{j-1}) + \deg(Q_j) - 1 \qquad \text{for} \quad j \geqq 2 ,$$

the $\deg(Q_j)$ initial values can be obtained from the expansion of P_{j-1}/Q_{j-1}.

This leads to the following <u>algorithm</u> for calculating the terms s_i with $1 \leqq i \leqq 2\deg(Q_j)$, given the polynomials $A_1, A_2, \ldots, A_j$. We assume that the polynomials $Q_1, Q_2, \ldots, Q_j$ have already been calculated by the recursion in Section 3 and we put $q_h = \deg(Q_h)$ for $1 \leqq h \leqq j$.

<u>Step 1:</u> We have

$$\frac{P_1}{Q_1} = \frac{1}{Q_1} = c^{-1} x^{-q_1} + \text{smaller powers of } x ,$$

where c is the leading coefficient of Q_1. Therefore we can calculate the terms s_i with $1 \leqq i \leqq q_1 + q_2 - 1$ by the recursion (1) with characteristic polynomial Q_1 and initial values

$$s_i = 0 \qquad \text{for} \quad 1 \leqq i \leqq q_1 - 1, \qquad s_i = c^{-1} \qquad \text{for} \quad i = q_1 .$$

<u>Step h</u> $(2 \leqq h \leqq j)$: Suppose the terms s_i with $1 \leqq i \leqq q_{h-1} + q_h - 1$ have already been calculated. Then we calculate the terms s_i with $q_{h-1} + q_h \leqq i \leqq q_h + q_{h+1} - 1$ from the previously calculated terms by the recursion (1) with characteristic polynomial Q_h. If $h = j$, then we replace $q_j + q_{j+1} - 1$ by $2\,q_j$.

If a sequence $A_1, A_2, \ldots$ of polynomials is given, then the algorithm above can be continued indefinitely. The third sentence in Step h can then be deleted. A special situation occurs in the case where F is the binary field F_2. It is due to the trivial fact that if $a \in F_2$ is given and we know that $b \in F_2$ satisfies $b \neq a$, then b is uniquely determined as the binary complement of a, i.e. $b = a + 1$. It follows from this fact that for $F = F_2$ the relation (11) also determines the term s_i with $i = \deg(Q_j) + \deg(Q_{j+1})$. Consequently, in the algorithm above the numbers $q_h + q_{h+1} - 1$ can be replaced by $q_h + q_{h+1}$ for $1 \leqq h \leqq j - 1$ whenever the underlying field is F_2, provided the term s_i with $i = q_h + q_{h+1}$ is defined as the binary complement of the corresponding term obtained from the recursion (1).

We show now that if the sequence $s_1, s_2, \ldots$ of elements of F is constructed by the algorithm above with polynomials $A_1, A_2, \ldots,$ then its generating function S has the continued fraction expansion

$$S = [0, A_1, A_2, \ldots] .$$

For suppose we had a different expansion $S = [0, A'_1, A'_2, \ldots]$ with corresponding convergents P'_h/Q'_h for $h \geqq 0$. Then there exists a least $k \geqq 1$ with $A'_k \neq A_k$. By the construction we have

$$v(S - \frac{P_h}{Q_h}) \leqq - \deg(Q_h) - \deg(Q_{h+1}) \qquad \text{for all} \quad h \geqq 0 . \tag{12}$$

In particular $v(Q_k\ S - P_k) < 0$, hence by Lemma 3 we get

$$Q_k = \sum_{h=i}^{j} C_h\ Q'_h \qquad \text{and} \qquad P_k = \sum_{h=i}^{j} C_h\ P'_h$$

with $C_i \neq 0$, $C_j \neq 0$, and $\deg(C_h) < \deg(A'_{h+1})$ for $i \leqq h \leqq j$. By the second part of Lemma 3 and (12),

$$- \deg(Q'_{i+1}) \leqq v(Q_k\ S - P_k) \leqq - \deg(Q_{k+1}) < - \deg(Q_k) \leqq - \deg(Q'_j) ,$$

hence we must have $i = j$. From $Q_k = C_j\ Q'_j$, $P_k = C_j\ P'_j$, and $\gcd(P_k, Q_k) = 1$ it follows that $C_j = c \in F$. Since $Q'_h = Q_h$ and $P'_h = P_h$ for $0 \leqq h < k$, we must have $j \geqq k$. On the other hand, (6) and (12) yield

$$- \deg(Q'_k) = v(Q'_{k-1}\ S - P'_{k-1}) = v(Q_{k-1}\ S - P_{k-1}) \leqq - \deg(Q_k) = - \deg(Q'_j) ,$$

hence $j \leqq k$, and so $j = k$ and $Q_k = cQ'_k$, $P_k = cP'_k$. Then by (4),

$$(-1)^k = P_{k-1}\ Q_k - P_k\ Q_{k-1} = cP'_{k-1}\ Q'_k - cP'_k\ Q'_{k-1} = (-1)^k\ c ,$$

thus $c = 1$. From $Q_k = Q'_k$ it follows that

$$A_k\ Q_{k-1} + Q_{k-2} = A'_k\ Q'_{k-1} + Q'_{k-2} = A'_k\ Q_{k-1} + Q_{k-2} ,$$

hence $A_k = A'_k$, which is a contradiction.

REFERENCES

[1] L. E. Baum and M. M. Sweet: Badly approximable power series in characteristic 2, Ann. of Math. 105, 573-580 (1977).

[2] Z.-D. Dai and Z.-X. Wan: A relationship between the Berlekamp-Massey and the Euclidean algorithms for linear feedback shift register synthesis, Preprint, Academia Sinica, Beijing, 1986.

[3] R. Lidl and H. Niederreiter: Finite Fields, Addison-Wesley, Reading, Mass., 1983.

[4] R. Lidl and H. Niederreiter: Introduction to Finite Fields and Their Applications, Cambridge Univ. Press, Cambridge, 1986.

[5] J. L. Massey: Shift-register synthesis and BCH decoding, IEEE Trans. Information Theory 15, 122-127 (1969).

[6] H. Niederreiter: Continued fractions for formal power series, pseudorandom numbers, and linear complexity of sequences, Contributions to General Algebra 5 (Proc. Conf. Salzburg, 1986), Teubner, Stuttgart, to appear.

[7] H. Niederreiter: Cryptology - The mathematical theory of data security, Proc. Internat. Symp. on Prospects of Math. Science (Tokyo, 1986), to appear.

[8] R. A. Rueppel: Linear complexity and random sequences, Advances in Cryptology - EUROCRYPT '85 (F. Pichler, ed.), Lecture Notes in Computer Science, Vol. 219, pp. 167-188, Springer-Verlag, Berlin, 1986.

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

[10] Y. Taussat: Approximation diophantienne dans un corps de séries formelles, Thèse, Université de Bordeaux, 1986.

[11] M.-Z. Wang and J. L. Massey: The characterization of all binary sequences with a perfect linear complexity profile, Paper presented at EUROCRYPT '86, Linköping, 1986.

The author gratefully acknowledges support for this research project by the Austrian Ministry for Science and Research.