# On Interval Linear Complexity of Binary Sequences *

Vladimir B. Balakirsky
Member, IEEE

The author is currently with the
Fakultät für Mathematik, Universität Bielefeld,
Postfach 100131, D-33501 Bielefeld 1, Germany
on leave from the Data Security Association "Confident",
193060 St.-Petersburg, Russia
E-mail addresses :
vbal@mathematik.uni-bielefeld.de
vbal@vbal.spb.su

**Abstract.** Some properties of the '$m$-interval linear complexity profile', which can be used as a criterion characterizing the level of randomness of a given binary sequence, are examined.

**Key words :**
**Sequences, Polynomials, Linear complexity**

*The work was supported in part by SFB-343, Universität Bielefeld, Germany

# 1 Introduction

A well-known measure of linear complexity of binary sequences, which is based on the linear descriptions of subsequences of a given sequence that have increasing lengths, was introduced by R.Rueppel [1]. This measure is called the linear complexity profile (LCP) of the sequence and it can be used to characterize the 'level of randomness' of that sequence. Nevertheless, there are cases when the criterions based on LCPs are not quite satisfactory [1]-[6]. In this correspondence, we develop the ideas of [7], [8] and examine the linear descriptions of the fragments of a given sequence having a fixed length $m$ and introduce a measure of linear complexity which we call the $m$-interval linear complexity profile ($m$-LCP). We assume that a comparison of the parameters of the $m$-LCP constructed for a given sequence and for fair coin-tossing sequences gives more information characterizing the closeness of the sequence to that class than the use of the LCP.

Our analysis widely uses the relationships between binary sequences and irreducible ratios of polynomials over $GF(2)$. The notation is introduced in Section 2. Some well-known properties of the LCP and the observation concerning the connections between the lengths of the linear feedback shift registers (LFSRs) and the degrees of the feedback polynomials used in the description are given in Section 3. The observation directly leads to an explicit formula for the number of irreducible ratios of polynomials over $GF(2)$ when the denominator has a fixed degree. This result should be known, but we cannot give the exact reference. A short proof of that formula and some discussion are given in Section 4. In Section 5 we define the $m$-LCP and show that it has a very regular properties for any sequence and that the formula for the number of irreducible ratios determines the statistical characteristics of the $m$-LCP of the fair coin tossing sequences.

# 2 Relationships Between Binary Sequences and Irreducible Ratios

We will denote by

$$f(z) = f_0 + f_1 \cdot z + ... + f_L \cdot z^L$$

a polynomial over $GF(2)$ having the degree $L$ or less, where $z$ is a formal variable. Throughout the paper we assume that $f_0 = 1$. A polynomial $f(z)$ is known as an irreducible polynomial if it can be divided only by 1 and $f(z)$. A ratio of two polynomials, $a(z)/f(z)$, will be referred to as an irreducible ratio of degree $n$ if

$$\deg a(z) < n, \quad \deg f(z) = n,$$

and the greatest common divisor of the polynomials $a(z)$ and $f(z)$ is equal to 1.

An equivalence between binary sequences and irreducible ratios is based on a representation of the ratio $a(z)/f(z)$ by the infinite polynomial at the right hand side of the equation

$$\frac{a(z)}{f(z)} = a(z) \cdot \sum_{t \geq 0} (1 + f(z))^t. \tag{2.1}$$

The possibility of such a representation follows from the formula

$$\begin{aligned} f(z) \cdot \sum_{t \geq 0} (1 + f(z))^t &= \sum_{t \geq 0} (1 + f(z))^t + \sum_{t \geq 0} (1 + f(z))^{t+1} \\ &= (1 + f(z))^0 + \sum_{t \geq 1} (1 + f(z))^t + \sum_{t \geq 1} (1 + f(z))^t \\ &= (1 + f(z))^0 \\ &= 1, \end{aligned}$$

where the identity $f(z) = 1 + 1 + f(z)$ was used. In particular, we may write

$$\begin{aligned} \frac{a(z)}{f(z)} + \frac{b(z)}{g(z)} &= \frac{a(z) \cdot g(z) + b(z) \cdot f(z)}{f(z) \cdot g(z)}, \\ \frac{a(z)}{f(z)} \cdot \frac{b(z)}{g(z)} &= \frac{a(z) \cdot b(z)}{f(z) \cdot g(z)}, \end{aligned}$$

where the ratios are interpreted as infinite polynomials defined by the expressions, which are similar to the expression at the right hand side of (2.1). Furthermore, if $a(z) = c(z) \cdot b(z)$ and $f(z) = c(z) \cdot g(z)$, then

$$\frac{a(z)}{f(z)} = \frac{b(z)}{g(z)}.$$

A given binary sequence $\mathbf{u}_t = (u_1, ..., u_t)$ of length $t$ can be also defined by the polynomial

$$u_t(z) = u_1 + u_2 \cdot z + ... + u_t \cdot z^{t-1},$$

and we write

$$u_t(z) = R_{z^t} \left[ \frac{a(z)}{f(z)} \right] \tag{2.2}$$

if $u_i$ is equal to the $(i-1)$-st coefficient of the polynomial at the right hand side of (2.2) for all $i = 1, ..., t$; hereafter $R_{d(z)} [\, c(z)\, ]$ denotes the remainder of the division of $c(z)$ by $d(z)$ for all $(c(z), d(z))$ such that $d(z) \neq 0$.

We will widely use the following notation.

**Definition 1** Given binary sequence $\mathbf{u}_\tau$, let

$$\mathbf{u}_t^{(k)} = (u_{t-k+1}, ..., u_t), \quad \text{for all} \quad k \geq 1 \quad \text{and} \quad t = 1, ..., \tau,$$

where $u_i = 0,\ i \leq 0$.

**Definition 2** We will write

$$\mathbf{u}_t^{(k)} \prec \mathcal{F}(L, f(z)),$$

where $f(z) = 1 + f_1 \cdot z + ... + f_L \cdot z^L$ is a polynomial of degree $L$ or less, if

$$u_j = f_1 \cdot u_{j-1} + ... + f_L \cdot u_{j-L}, \quad j = t-k+1, ..., t,$$

and the pair $(L, f(z))$ cannot be replaced with some pair $(\tilde{L}, \tilde{f}(z))$ such that $\tilde{L} < L$ and $\tilde{f}(z)$ is a polynomial of degree $\tilde{L}$ or less.

# 3 Some Properties of the Linear Complexity Profile of Binary Sequences

Given binary sequence $\mathbf{u}_\tau$, let

$$\mathbf{u}_t^{(t-L_t)} \prec \mathcal{F}(L_t, f_t(z)), \quad t = 1, ..., \tau, \tag{3.1}$$

and let

$$\mathbf{L}_\tau = (L_0, ..., L_\tau), \quad \mathbf{f}_\tau(z) = (f_0(z), ..., f_\tau(z)),$$

where $(L_0, f_0(z)) = (0, 1)$. We also denote

$$d_t = \deg f_t(z), \quad t = 0, ..., \tau. \tag{3.2}$$

The value of $L_t$ is known as a *Linear Complexity of* $\mathbf{u}_t$, and the vector $\mathbf{L}_\tau$ is known as the *LCP of* $\mathbf{u}_\tau$. Some properties of LCPs are specified in the following statement [2], [9].

**Lemma 3.1** For any sequence $\mathbf{u}_\tau$,

(1) $0 = L_0 \leq ... \leq L_\tau$;

(2) if $L_{t^*} \neq L_{t^*-1}$, then $L_{t^*-1} \leq (t^* - 1)/2$ and $L_{t^*} = t^* - L_{t^*-1}$;

(3) if $L_t = L \leq t/2$, then there exists only one polynomial $f_t(z)$ such that (3.1) is valid;

(4) if $L_t = L_{t-1} \leq (t-1)/2$, then $f_t(z) = f_{t-1}(z)$.

The result of Lemma 3.1 shows that the LCP of any binary sequence fluctuates around the line $t/2$ (Fig.1). Only one feedback polynomial corresponds to every component of the LCP located below that line and this polynomial is the same for all components at the same horizontal line. We will pay our attention to the connections between the degree of that polynomial and the linear complexity in the points where the LCP crosses the line $t/2$.

The LCP of $\mathbf{u}_\tau$ can be found using the BM algorithm [10]. We describe that algorithm since its properties will be widely used in the analysis and give a numerical example in Table 1, where the vector $(f_{t,d_t}, ..., f_{t,0})$ consisting of the coefficients of the polynomial $f_t(z)$ is given as an integer written in the octal notation; $t = 1, ..., \tau$.

1) Set $t = 1$, $(L, f(z)) = (\lambda, \varphi(z)) = (0, 1)$, and $\delta = 0$.

2) Increase $\delta$ by 1.

3) Set

$$(L_t, f_t(z)) = \begin{cases} (L, f(z)), & \text{if } u_t = u_t^*, \\ (\max\{L, \delta + \lambda\}, f(z) + z^\delta \cdot \varphi(z)), & \text{if } u_t \neq u_t^*, \end{cases}$$

where $u_t^* = f_1 \cdot u_{t-1} + ... + f_L \cdot u_{t-L}$.

4) If $L_t > L$, then set $(\lambda, \varphi(z)) = (L, f(z))$ and $\delta = 0$.

5) Set $(L, f(z)) = (L_t, f_t(z))$.

6) Increase $t$ by 1. If $t \leq \tau$, then go to 2.

7) End.

The BM algorithm can be represented as a procedure leading to a sequence of irreducible ratios $a_t(z)/f_t(z)$, which approximates the subsequences $\mathbf{u}_t$, $t = 0, ..., \tau$; the polynomials $a_t(z)$ are defined in Lemma 3.2 below.

**Lemma 3.2** Given $t$ and polynomial $f_t(z)$ of degree $d_t$ such that (3.1) is valid, let

$$a_t(z) = 1 + a_{t,1} \cdot z + ... + a_{t,d_t-1} \cdot z^{d_t-1}, \tag{3.3}$$

where

$$a_{t,j} = \sum_{i=0}^{d_t-1} f_{t,i} \cdot u_{\Delta t+j+1-i}, \quad j = 0, ..., d_t - 1 \tag{3.4}$$

and

$$\Delta t = L_t - d_t. \tag{3.5}$$

Then

$$u_t(z) = u_{\Delta t}(z) + z^{\Delta t} \cdot R_{z^{t-\Delta t}} \left[ \frac{a_t(z)}{f_t(z)} \right]. \tag{3.6}$$

**Lemma 3.3** Given binary sequence $\mathbf{u}_\tau$, let $t_0 < t_1 < t_2$ be the elements of the set

$$\mathcal{T} = \{ \, t \in \{0, ..., \tau\} : \; L_t = t/2 \, \} \tag{3.7}$$

and let $t' \notin \mathcal{T}$ for all $t' \in \{t_0 + 1, ..., t_2 - 1\} \backslash \{t_1\}$. Then

(1) $d_{t_0} = L_{t_0}, \; d_{t_1} < L_{t_1} \implies d_{t_2} = L_{t_2}$;

(2) $d_{t_0} = L_{t_0}, \; d_{t_1} = L_{t_1} \implies d_{t_2} = L_{t_2}$ or $d_{t_2} < L_{t_2}$ depending on $u_{t_2}$;

(3) there are no binary sequences such that $d_{t_0} < L_{t_0}$ and $d_{t_1} < L_{t_1}$.

**Proof** Given $t \geq 1$, let us denote the value of $\delta$, the polynomial $\varphi(z)$, and the polynomial $f(z)$ constructed by the BM algorithm before the step 2) by $\delta_{t-1}$, $\varphi_{t-1}(z)$, and $f_{t-1}(z)$, respectively. Then the polynomial $f_t(z)$ constructed at step 3) can be expressed as follows :

$$f_t(z) = \begin{cases} f_{t-1}(z), & \text{if } u_t = u_t^*, \\ f_{t-1}(z) + z^{\delta_{t-1}+1} \cdot \varphi_{t-1}(z), & \text{if } u_t \neq u_t^*. \end{cases} \tag{3.8}$$

Hence,

$$d_{t-1} > \delta_{t-1} + 1 + \deg \varphi_{t-1}(z) \implies d_t = d_{t-1} \tag{3.9}$$

regardless of $u_t$. Let (see Fig.1)

$$t_1^* = (t_0 + t_1)/2, \;\; t_2^* = (t_1 + t_2)/2. \tag{3.10}$$

Then using the statements of Lemma 3.1 we write

$$\begin{aligned} t_0/2 &= L_{t_0} = L_{t_0+1} = ... = L_{t_1^*-1}, \\ t_1/2 &= L_{t_1^*} = L_{t_1^*+1} = ... = L_{t_2^*-1}, \\ t_2/2 &= L_{t_2^*} = L_{t_2^*+1} = ... = L_{t_2}, \end{aligned} \tag{3.11}$$

and

$$\begin{aligned} (\lambda_t, \varphi_t(z)) &= \begin{cases} (t_0/2, f_{t_0}(z)), & \text{if } t \in \{t_1^*, ..., t_2^* - 1\}, \\ (t_1/2, f_{t_1}(z)), & \text{if } t \in \{t_2^*, ..., t_2\}, \end{cases} \\ \delta_t &= \begin{cases} t - t_1^*, & \text{if } t = t_1^*, ..., t_2^* - 1, \\ t - t_2^*, & \text{if } t = t_2^*, ..., t_2. \end{cases} \end{aligned} \tag{3.12}$$

Note that (3.11) and (3.12) lead to the equations

$$\delta_{t-1} + 1 + \deg \varphi_{t-1}(z) = \begin{cases} t_2^* - t_1^* + d_{t_0}, & \text{if } t = t_2^*, \\ t - t_2^* + d_{t_1}, & \text{if } t = t_2^* + 1, ..., t_2 \end{cases} \tag{3.13}$$

and suppose that $d_{t_0} = L_{t_0}$. Since $d_{t_2^*-1} \leq t_1/2$ and $f_{t_2^*}(z) \neq f_{t_2^*-1}(z)$, using (3.8) and (3.13) we obtain

$$d_{t_2^*} = t_2^* - t_1^* + d_{t_0} = t_2/2.$$

Furthermore, using (3.10), (3.11), and the inequality $d_{t_1} \leq L_{t_1}$ we obtain

$$t_2/2 > t - (t_1 + t_2)/2 + d_{t_1}, \;\; \text{for all } \; t = t_2^* + 1, ..., t_2 - 1.$$

Thus, the inequality at the left hand side of (3.9) is valid for all $t = t_2^* + 1, ..., t_2 - 1$ and $d_{t_2-1} = t_2/2$. This inequality is also valid for $t = t_2$ if $d_{t_1} < L_{t_1}$, and the statement (1) is proved. If $d_{t_1} = L_{t_1}$ and $u_{t_2} = u_{t_2}^*$, then $f_{t_2}(z) \neq f_{t_2-1}(z)$ and $d_{t_2} = t_2/2$. Otherwise, the BM algorithm constructs a polynomial $f_{t_2}(z)$, whose degree is less than $t_2/2$, and the statement (2) is proved. To prove the statement (3) we note that $0 \in \mathcal{T}$ and $d_0 = L_0$. Hence, the algorithm either follows the rule (1), or (2). Q.E.D.

**Discussion** Given binary sequence $\mathbf{u}_\tau$, we construct the sequence $(L_t, f_t(z))$, $t = 1, 2, ...$, using the BM algorithm and note that this sequence determines the sequence of irreducible ratios $(L_t, a_t(z)/f_t(z))$, $t = 1, 2, ...$, where the polynomial $a_t(z)$ is defined in (3.3)-(3.5). For all $t \in \mathcal{T}$, let us generate the sign $'='$ if $d_t = L_t$ and the sign $'<'$ if $d_t < L_t$ (the parameter $d_t$ and the set $\mathcal{T}$ are defined in (3.2) and (3.7), respectively). As a result, we obtain a sequence $\mathbf{S}$ consisting of these signs (see Table 1). Note that, due to (3.7) and the statement (3) of Lemma 3.1, there exists only one 'minimal' pair $(L_t, f_t(z))$, which provides the relation (3.1), i.e., the sequence $\mathbf{S}$ is completely defined by $\mathbf{u}_\tau$ and the reference to the BM algorithm is not essential. Let us introduce a Markov chain whose states are the elements of $\mathbf{S}$. Lemma 3.3 guarantees that two neighboring signs $'<', <'$ will be never met for any $\mathbf{u}_\tau$, and the chain is characterized by the transition probabilities given in Fig.2, where we assume that each component of $\mathbf{u}_\tau$ is obtained by random and independent selection from $\{0, 1\}$ with probability 1/2. For example, if the 14-th bit of the sequence given in Table 1 is replaced with 0, then the feedback polynomial is equal to

$$1 + z + z^7 + z^{14-9} \cdot (1 + z) = 1 + z + z^5 + z^6 + z^7$$

and the degree coincides with the length of the LFSR as for $u_{14} = 1$. The properties of the Markov chain with the transition state diagram, given in Fig.2, lead to the result of Section 4 : the stationary probability of the state (=), which is equal to 2/3, determines the probability to get an irreducible ratio of any given degree.

# 4 An Explicit Formula for the Number of Irreducible Ratios of Polynomials over GF(2)

The well-known result concerning the number of irreducible polynomials of a given degree [11] is formulated below.

**Theorem 4.1** Let $I_n$ be the number of irreducible polynomials of degree $n$ such that $f_0 = 1$. Then

$$I_n = \frac{1}{n} \cdot \sum_{d|n} \mu(d) \cdot 2^{n/d}, \quad \text{for all } \; n \geq 2,$$

where the sum is extended over all positive divisors $d$ of $n$ and

$$\mu(d) = \begin{cases} 1, & \text{if } d = 1, \\ (-1)^k, & \text{if } d \text{ if the product of } k \text{ distinct primes}, \\ 0, & \text{if } d \text{ is divisible by a square of a prime} \end{cases}$$

is the Möbius function.

**Corollary 4.1**

$$\frac{1}{n} \cdot (1 - 2^{-n/2+1}) \cdot 2^n < I_n \leq \frac{1}{n} \cdot (1 - 2^{-n+1}) \cdot 2^n, \quad \text{for all } \; n \geq 2.$$

**Remark 4.1**

(a) We are supposed to solve the problem of factorization of a given integer to find $I_n$; this problem is very hard if $n$ is large enough [12].

(b) If each coefficient $f_i, i = 1, ..., n-1$, is obtained by random and independent selection from $\{0, 1\}$ with probability $1/2$, then the probability that the polynomial $f(z)$ is irreducible tends to 0 as $2/n$ when $n \to \infty$.

**Theorem 4.2** Given $j \in \{0, ..., n-1\}$, let $I_n^{(j)}$ be the number of irreducible ratios $a(z)/f(z)$ of degree $n$ such that $a_0 = ... = a_{j-1} = 0$ and $a_j = 1$. Then

$$I_n^{(j)} = \frac{2}{3} \cdot 2^{2n-j-2} + \frac{1}{3} \cdot 2^j, \quad \text{for all } \; n \geq 1. \tag{4.1}$$

**Corollary 4.2** Let $I_n^*$ be the number of irreducible ratios $a(z)/f(z)$ of degree $n$. Then

$$I_n^* = \frac{2}{3} \cdot 2^{2n-1} - \frac{1}{3}, \quad \text{for all } \; n \geq 1. \tag{4.2}$$

**Remark 4.2**

(a) The number of irreducible ratios can be easily found for any $n$.

(b) If each coefficient $a_i, i = 0, ..., n-1$, and $f_i, i = 1, ..., n-1$, is obtained by random and independent selection from $\{0, 1\}$ with probability $1/2$ then, for all $n$, the ratio $a(z)/f(z)$ is irreducible with the probability $\approx 2/3$.

**Remark 4.3 (references to number theory [13])** The number of primes not exceeding $n$ is asymptotic to $n/\log n$, and the number of positive integers not greater than and prime to $n$ is known as the Euler's function

$$\phi(n) = n \cdot \prod_{p|n} \left(1 - \frac{1}{p}\right),$$

where the product is extended over all positive prime divisors $p$ of $n$. It is known that

$$\Phi(n) = \phi(1) + ... + \phi(n) = \frac{3n^2}{\pi^2} + O(n \log n).$$

Therefore

$$\lim_{n\to\infty} \frac{\Phi(n)}{n(n+1)/2} = \frac{6}{\pi^2},$$

and the expression at the left hand side can be interpreted as the probability that two integers, $p$ and $q$, such that $1 \leq p \leq q \leq n$ are prime to one another when $n \to \infty$, i.e., as the probability that the ratio $p/q$ is irreducible.

**Proof** The proof of Theorem 4.2 follows from Lemmas 4.3, 4.4 below. To prove Corollary 4.2 we substitute the expression at the right hand side of (4.1) into the sum

$$I_n^* = \sum_{j=0}^{n-1} I_n^{(j)}$$

and obtain (4.2). Q.E.D.

**Lemma 4.3** Given $j \in \{0, ..., n-1\}$, the number of irreducible ratios $a(z)/f(z)$ such that

$$a_0 = ... = a_{j-1} = 0, \quad a_j = 1 \tag{4.3}$$

and $f(z)$ is a polynomial of degree $n$, is equal to the number of binary sequences $\mathbf{u}_{2n}$ such that

$$u_1 = ... = u_j = 0, \quad u_{j+1} = 1 \tag{4.4}$$

and

$$d_{2n} = L_{2n} = n, \tag{4.5}$$

where $d_{2n}$ is defined in (3.2) and $L_{2n}$ is the linear complexity of $\mathbf{u}_{2n}$.

**Proof** The statement directly follows from (2.2), (3.5), and (3.6). Q.E.D.

**Lemma 4.4** Given $j \in \{0, ..., n-1\}$, let $K_{2n}$ be the number of binary sequences $\mathbf{u}_{2n}$ satisfying (4.4) and (4.5). Then

$$K_{2n} = \frac{2}{3} \cdot 2^{2n-j-2} + \frac{1}{3} \cdot 2^j. \tag{4.6}$$

**Proof** Let $K'_{2n}$ be the number of binary sequences of length $2n$ satisfying (4.4) such that $L_{2n} = n$ and $d_{2n} < n$. Note that $2j+2 \in \mathcal{T}$ and $t \notin \mathcal{T}$, $t = 1, ..., 2j+1$, where the set $\mathcal{T}$ is defined in (3.7). Therefore, using Lemma 3.3 we have

$$\begin{aligned} K_{2n} &= \frac{1}{2} \cdot \sum_{l=1}^{n-j-1} K_{2n-2l} \cdot 2^l + \sum_{l=1}^{n-j-1} K'_{2n-2l} \cdot 2^l, \\ K'_{2n} &= \frac{1}{2} \cdot \sum_{l'=1}^{n-j-1} K_{2n-2l'} \cdot 2^{l'}, \quad n = j+2, j+3, ..., \end{aligned} \tag{4.7}$$

where

$$K_{2j+2} = K'_{2j+2} = 2^j. \tag{4.8}$$

Substituting the second equation in (4.7) into the first one and using (4.8) we have

$$\begin{aligned} K_{2n} &= \frac{1}{2} \cdot \sum_{l=1}^{n-j-1} K_{2n-2l} \cdot 2^l + \sum_{l=1}^{n-j-2} \frac{1}{2} \cdot \sum_{l'=1}^{n-l-j-1} K_{2n-2l-2l'} \cdot 2^{l+l'} \\ &+ K'_{2j+2} \cdot 2^{n-j-1} \end{aligned} \tag{4.9}$$

$$
\begin{aligned}
&= \frac{1}{2}\cdot\sum_{l=1}^{n-j-1} K_{2n-2l}\cdot 2^l + \sum_{l=1}^{n-j-1}\frac{1}{2}\cdot(l-1)\cdot K_{2n-2l}\cdot 2^l + 2^{n-1} \\
&= \frac{1}{2}\cdot\sum_{l=1}^{n-j-1} l\cdot K_{2n-2l}\cdot 2^l + 2^{n-1}, \quad n = j+2, j+3, ...
\end{aligned}
$$

Let us substitute the expression at the right hand side of (4.6) for $K_{2n}$. Then

$$
\frac{2}{3}\cdot 2^{2j+2-j-2} + \frac{1}{3}\cdot 2^j = 2^j
$$

and

$$
\begin{aligned}
\frac{2}{3}\cdot 2^{2n-j-2} + \frac{1}{3}\cdot 2^j \quad &= \quad \frac{1}{2}\cdot\sum_{l=1}^{n-1} l\cdot\left(\frac{2}{3}\cdot 2^{2n-2l-j-2} + \frac{1}{3}\cdot 2^j\right)\cdot 2^l + 2^{n-1}, \\
& \quad\; n = j+2, j+3, ...,
\end{aligned}
$$

where the equation follows from the identities

$$
\begin{aligned}
\sum_{l=1}^{n-j-1} l\cdot 2^{-l} \quad &= \quad 2\cdot(1-(n-j+1)\cdot 2^{-n+j}), \\
\sum_{l=1}^{n-j-1} l\cdot 2^{l} \quad &= \quad 2\cdot(1+(n-j-2)\cdot 2^{n-j-1}).
\end{aligned}
$$

Hence, if (4.6) is valid then both (4.8) and (4.9) are satisfied. Q.E.D.

# 5 $m$-Interval Linear Complexity of Binary Sequences

Let $\mathbf{u}_\tau$ be a given binary sequence such that $u_1 = 1$ and let $m > 0$ be a fixed integer. Let

$$
\mathbf{u}_t^{(m)} \prec \mathcal{F}(L_t^{(m)}, f_t^{(m)}(z)), \quad t = m+1, ..., \tau, \tag{5.1}
$$

and let

$$
\mathbf{L}_\tau^{(m)} = (L_{m+1}^{(m)}, ..., L_\tau^{(m)}), \quad \mathbf{f}_\tau^{(m)}(z) = (f_{m+1}^{(m)}(z), ..., f_\tau^{(m)}(z)).
$$

The parameter $L_t^{(m)}$ will be referred to as the *$m$-Interval Linear Complexity of* $\mathbf{u}_\tau$ at position $t$, and the sequence $\mathbf{L}_\tau^{(m)}$ will be regarded as the *$m$-LCP of* $\mathbf{u}_\tau$.

The definition of the $m$-LCP is illustrated in Fig.3. Given $\mathbf{u}_\tau$, let us fix $t \in \{m+1, ..., \tau\}$, construct a matrix consisting of the rows $\mathbf{u}_1^{(m)}, ..., \mathbf{u}_t^{(m)}$, and consider all binary vectors $\mathbf{c} = (c_1, ..., c_{t-1})$ such that

$$\mathbf{u}_t^{(m)} = \sum_{h=1}^{t-1} c_h \cdot \mathbf{u}_h^{(m)}. \tag{5.2}$$

Since $u_1 = 1$ and $u_i = 0, \ i \le 0$, the rows $\mathbf{u}_1^{(m)}, ..., \mathbf{u}_m^{(m)}$ are linearly independent and (5.2) holds for at least one $\mathbf{c}$. We are interested in a vector $\mathbf{c}^* = \mathbf{c}$ which provides (5.2) and has the maximal number of zeroes at the first positions, i.e., $k(\mathbf{c}^*) \ge k(\mathbf{c})$ for all possible $\mathbf{c}$, where $k(\mathbf{c}) = k$ if $c_1 = ... = c_k = 0$ and $c_{k+1} = 1$. Then we set $L_t^{(m)} = t - k(\mathbf{c}^*)$ and $f_{t,t-i}^{(m)} = c_i, \ \ i = k(\mathbf{c}^*) + 1, ..., t-1$.

The following result directly follows from the definition of the $m$-interval linear complexity.

**Lemma 5.1** Let $t$ be given. Then $\mathbf{u}_t^{(m)} \prec \mathcal{F}(n, f(z))$ iff $n$ is the minimal integer such that there exists an LFSR $(n, f'(z))$ generating the sequence $u'_1, ..., u'_{m+n}$, where

$$u'_j = u_{t-j+1}, \ \text{ for all } \ j = 1, ..., m+n, \tag{5.3}$$

whose length and the degree of the feedback polynomial

$$f'(z) = 1 + f'_1 \cdot z + ... + f'_n \cdot z^n \tag{5.4}$$

are equal to $n$. Furthermore,

$$f'_j = f_{n-j}, \ \ j = 0, ..., n, \tag{5.5}$$

and

$$u'_{m+n}(z) = R_{z^{m+n}} \left[ \frac{a'(z)}{f'(z)} \right], \tag{5.6}$$

where the coefficients of the polynomial $a'(z)$ are defined similarly to (3.4), i.e.,

$$a'_j = \sum_{i=0}^{n-1} f'_i \cdot u'_{j+1-i} = \sum_{i=0}^{n-1} f'_i \cdot u_{t-j+i}, \quad j = 0, ..., n-1. \tag{5.7}$$

For any $t \geq m+1$, the parameter $L_t^{(m)}$ can be constructed using a modified BM algorithm. Let us extend the definition (5.3) in such a way that

$$u'_j = u_{t-j+1}, \quad \text{for all } \ j = 1, 2, ... \tag{5.8}$$

and use the algorithm, whose $1^{(m)}$-$5^{(m)}$ steps coincide with the 1-5 steps of the BM algorithm, where $t'$ and $\mathbf{u}'$ are substituted for $t$ and $\mathbf{u}$, while the $6^{(m)}$-$7^{(m)}$ steps are given below.

$6^{(m)}$) If $t' - L \geq m$ and deg $f(z) = L$, then go to $7^{(m)}$. Otherwise, increase $t'$ by 1 and go to $2^{(m)}$.

$7^{(m)}$) Set $(L_t^{(m)}, f_t^{(m)}(z)) = (L, f'(z))$, where $f'(z)$ is the inverse version of $f(z)$, i.e., $f'_i = f_{L-i}, \ i = 0, ..., L$. End.

The $m$-LCP of $\mathbf{u}_\tau$ has a very regular structure, as it follows from the statement below. If the current element of the profile, $L_{t-1}^{(m)}$, is greater than $m$ then the next element, $L_t^{(m)}$, can be less than $m$, i.e., the profile 'falls into the pit'. In this case, the profile can stay in the pit for $l$ times or jumps at the level $m + l$ and stays at this level for $m - L_t^{(m)}$ times. The parameters $l$ and $m - L_t^{(m)}$ can be interpreted as the 'length' and the 'depth' of the pit, and the duality between them takes place (Figs.3,4).

**Theorem 5.2** If

$$\begin{cases} L_{t-1}^{(m)} \neq L_t^{(m)} \\ L_t^{(m)} = ... = L_{t+l-1}^{(m)} = n < m \\ L_{t+l}^{(m)} \neq L_{t+l-1}^{(m)} \end{cases} \tag{5.9}$$

then

$$\begin{cases} L_{t-1}^{(m)} \geq m \\ f_t^{(m)}(z) = ... = f_{t+l-1}^{(m)}(z) \\ L_{t+l}^{(m)} = ... = L_{t+l+m-n-1} = m + l \\ L_{t+l+m-n}^{(m)} \leq m \end{cases} \tag{5.10}$$

and the polynomial $f_t^{(m)}(z)$ is unique.

**Proof** If

$$\mathbf{u}_t^{(m)} \prec \mathcal{F}(n, f(z)),$$

where $n < m$, then the linear complexity of the sequence

$$\mathbf{u}' = (u_{t-1}, ..., u_{t-n-m+1})$$

is equal to $n$ (otherwise the statement (3) of Lemma 3.1 is violated) and this sequence is generated by the feedback polynomial $f'(z)$ (see (5.4), (5.5)). If $L_{t-1}^{(m)} \neq n$, then $\mathbf{u}'$ cannot be continued by $f'(z)$ with the component $u_{t-n-m}$ and, because of the statement (2) of Lemma 3.1, the linear complexity of the sequence $(\mathbf{u}', u_{t-n-m})$ is equal to $m$. Hence, $L_{t-1}^{(m)} \geq m$. Similar considerations lead to the conclusions that there exists only one polynomial $f(z)$ generating $\mathbf{u}_t^{(m)}$ and that $f_{t+i}^{(m)}(z) = f(z)$ for all $i = 0, ..., l-1$.

Suppose

$$\mathbf{u}_{t+l+\Delta l}^{(m)} \prec \mathcal{F}(n^*, f^*(z)), \tag{5.11}$$

where $\Delta l \in \{0, ..., m-n-1\}$. Then the vector $\mathbf{u}_{t+l+\Delta l}^{(m)}$ can be expressed as a linear combination of the vectors $\mathbf{u}_{t+l+\Delta l-j}^{(m)}$, $j = 1, ..., n^*$. However, since

$$u_{t-m-n+1}, ..., u_{t+l-1} \prec \mathcal{F}(n, f(z)),$$

the vectors $\mathbf{u}_{t-n}^{(m)}, ..., \mathbf{u}_{t+l-1-n}^{(m)}$ can be expressed as linear combinations of the vectors $\mathbf{u}_{t+l-n}^{(m)}, ..., \mathbf{u}_{t+l-1}^{(m)}$ and

$$n^* \leq n + l + \Delta l \implies n^* \leq n + \Delta l.$$

Based on the statements (2), (3) of Lemma 3.1 it is easy to see that the inequality $n^* \leq n + \Delta l$ is impossible and

$$n^* > n + l + \Delta l. \tag{5.12}$$

The statement (5.11) and the inequality (5.12) mean that the sequence

$$\mathbf{u}^0 = (u_{t-m-n+1}, ..., u_{t+l-1})$$

belongs to the sequence

$$u_{t+l+\Delta l-m-n^*+1}, ..., u_{t+l+\Delta l}$$

and $n^*$ cannot be less than the linear complexity of the sequence $(\mathbf{u}^0, u_{t+l})$, which is equal to $l+m$ (Lemma 3.1). At last, we note that we can construct the LFSR of length $L' = m + l$ using the BM algorithm.

The inequality $L^{(m)}_{t+l+m-n} \le m$ follows from the fact that the vectors $\mathbf{u}^{(m)}_{t+l+m-n-1}, ..., \mathbf{u}^{(m)}_{t+l-n-1}$ are linearly independent.
Q.E.D.

**Proposition 5.3** Suppose that each component $u_2, ..., u_t$, where $t > 2m$, is obtained by random and independent selection from $\{0, 1\}$ with probability $1/2$. Then

$$\begin{aligned} Pr\{\ L^{(m)}_t = n\ \} &= \frac{1}{2^{m+1-n}} \cdot I^*_n \cdot 2^{-2n+1} \\ &= \frac{1}{3 \cdot 2^{m-n}} - \frac{1}{6 \cdot 2^{m-n}} \cdot 2^{-2n+1}, \quad n = 1, ..., m, \end{aligned} \tag{5.13}$$

where $I^*_n$ is the number of irreducible ratios of degree $n$.

**Proof** Let the sequence $\mathbf{u}'_{m+n}$ be connected with $\mathbf{u}_t$ by (5.3). If $L^{(m)}_t = n$ then there exists an irreducible ratio $a'(z)/f'(z)$ such that (5.6) is valid. Hence,

$$Pr\{\ L^{(m)}_t = n\ \} = I^*_n \cdot 2^{-(m+n)},$$

and (5.13) follows from (4.2). Q.E.D.

**Remark 5.3** In further considerations, we will use the formula (5.13) for all $t = 1, ..., \tau$ and assume that the ensemble of fair coin-tossing sequences contains all sequences of length $\tau$ whose components are obtained by random and independent selection from $\{0, 1\}$ with probability $1/2$. Furthermore, summarizing over the lengths of a pit, we will write the sum up to $\infty$ instead of $\tau$. These assumptions seem to be reasonable since the value of $\tau$ is much greater than $m$, while a special analysis of the initial part and the tails of the sequences essentially complicates formalization.

Let

$$\begin{aligned} J_\tau &= \left\{\ t:\ L^{(m)}_{t-1} \ge m,\ L^{(m)}_t < m\ \right\}, \\ J'_\tau &= \left\{\ t:\ f^{(m)}_{t-1}(z) \ne f^{(m)}_t(z),\ L^{(m)}_t = m,\ f^{(m)}_t(z) = f^{(m)}_{t+1}(z)\ \right\}, \\ J^{(m)}_\tau &= J_\tau \bigcup J'_\tau. \end{aligned} \tag{5.14}$$

For all $t \in J_\tau^{(m)}$, let us define the 'length' $l_t^{(m)}$ setting

$$l_t^{(m)} = l \iff f_t^{(m)}(z) = ... = f_{t+l-1}^{(m)}(z),\ f_t^{(m)}(z) \neq f_{t+l}^{(m)}(z) \tag{5.15}$$

and the 'redundancy'

$$r_t^{(m)} = m + l_t^{(m)} - 1 - L_t^{(m)}. \tag{5.16}$$

If $(L_t^{(m)}, f_t^{(m)}(z)) = (n, f(z))$ and $l_t^{(m)} = l$ for some $t \in J_\tau^{(m)}$, then

$$\mathbf{u}_t^{(m)}, ..., \mathbf{u}_{t+l-1}^{(m)} \prec \mathcal{F}(n, f(z))$$

and the sequence $u_{t-m+1}, ..., u_{t+l-1}$ is generated at the output of the LFSR of length $n$ having the feedback polynomial $f(z)$ if we load $u_{t-m-n+1}, ..., u_{t-m}$ into the register. This sequence can be defined by the $n$ coefficients of the polynomial $f(z)$ instead of $m + l - 1$ bits provided that we know $t$ and the initial contents of the shift register. Hence, the parameter $r_t^{(m)}$ defined in (5.16) gives the number of bits which can be compressed if we consider only the $t$-th pit. Note that the definitions (5.14)-(5.16) guarantee the inequality $r_t^{(m)} > 0$ and that the fragments defined by the neighboring pits can overlap. Thus, the sum of $r_t^{(m)}$ taken over all $t \in J_\tau^{(m)}$ derives a 'potential redundancy' of $\mathbf{u}_\tau$, which can be unrealizable meaning the data compression. Nevertheless, a comparison between the statistical characteristics of the parameters obtained from the $m$-LCP of a given sequence and the average characteristics over the ensemble of fair coin-tossing sequences can be used to specify the closeness of the sequence to that ensemble. In proposition below we calculate the expectations of the parameters

$$\pi^{(m)} = \frac{1}{\tau} \cdot |J_\tau^{(m)}|,\ \ r^{(m)} = \frac{1}{\tau} \cdot \sum_{t \in J_\tau^{(m)}} r_t^{(m)}. \tag{5.17}$$

**Proposition 5.4** Let $\overline{\pi^{(m)}}$ and $\overline{r^{(m)}}$ denote the average values of $\pi^{(m)}$ and $r^{(m)}$ taken over the ensemble of fair coin-tossing sequences. Then

$$\begin{aligned} \overline{\pi^{(m)}} &= \frac{1}{4} \cdot \left(1 + 2^{-2m}\right), \\ \overline{r^{(m)}} &= \frac{2}{3} + \frac{3m-1}{6} \cdot 2^{-m} + \frac{1}{6} \cdot 2^{-2m}. \end{aligned} \tag{5.18}$$

**Proof** Using (5.14)-(5.17) we write

$$
\begin{aligned}
\pi^{(m)} &= \frac{1}{\tau}\cdot\sum_{t=1}^{\tau}\left[\sum_{n=0}^{m-1}\chi_{tn}^{(m)}+\chi_{tm}^{(m)\prime}\right],\\
r^{(m)} &= \frac{1}{\tau}\cdot\sum_{t=1}^{\tau}\left[\sum_{n=0}^{m-1}\chi_{tn}^{(m)}\cdot\sum_{l\geq 1}(m+l-1-n)\cdot\chi\left\{\, l_t^{(m)}=l\,\right\}\right.\\
&\qquad\left.+\ \chi_{tm}^{(m)\prime}\cdot\sum_{l\geq 2}(l-1)\cdot\chi\left\{\, l_t^{(m)}=l\,\right\}\right]\\
&= \frac{1}{\tau}\cdot\sum_{t=1}^{\tau}\left[\sum_{n=0}^{m}\chi_{tn}^{(m)}\cdot\sum_{l\geq 1}(m+l-1-n)\cdot\chi\left\{\, l_t^{(m)}=l\,\right\}\right]
\end{aligned}
$$

where

$$
\begin{aligned}
\chi_{tn}^{(m)} &= \chi\left\{\, L_t^{(m)}=n,\ \mathbf{u}_{t-1}^{(m)}\not\prec\mathcal{F}(n,f_t^{(m)}(z))\,\right\},\\
\chi_{tm}^{(m)\prime} &= \chi\left\{\, L_t^{(m)}=m,\ \mathbf{u}_{t-1}^{(m)}\not\prec\mathcal{F}(m,f_t^{(m)}(z)),\ l_t^{(m)}\geq 2\,\right\}
\end{aligned}
$$

Given $(L_t^{(m)},f_t^{(m)}(z))$, the condition on $\mathbf{u}_{t-1}^{(m)}$ in $\chi_{tn}^{(m)}$ means that the component $u_{t-m-n-1}$ is fixed, while the conditions on $\mathbf{u}_{t-1}^{(m)}$ and $l_t^{(m)}$ in $\chi_{tm}^{(m)\prime}$ mean that the components $u_{t-m-n-1}$ and $u_{t+1}$ are fixed. Hence, taking the average value of $\pi^{(m)}$ and $r^{(m)}$ over the ensemble of fair coin-tossing sequences, we obtain

$$
\begin{aligned}
\overline{\pi^{(m)}} &= \frac{1}{\tau}\cdot\sum_{t=1}^{\tau}\left[\frac{1}{2}\cdot\sum_{n=0}^{m-1}Pr\{\, L_t^{(m)}=n\,\}+\frac{1}{4}\cdot Pr\{\, L_t^{(m)}=m\,\}\right]. \qquad (5.19)\\
\overline{r^{(m)}} &= \frac{1}{\tau}\cdot\sum_{t=1}^{\tau}\frac{1}{2}\cdot\sum_{n=0}^{m}Pr\{\, L_t^{(m)}=n\,\}\cdot\sum_{l\geq 1}(m+l-1-n)\cdot Pr\{\, l_t^{(m)}=l\,\}.
\end{aligned}
$$

Noticing that

$$Pr\{\, l_t^{(m)}=l\,\}=2^{-l}$$

and taking the sum over $l$ up to $\infty$ (see Remark 5.3) we write

$$\sum_{l\geq 1}(m+l-1-n)\cdot Pr\{\, l_t^{(m)}=l\,\}=m-n+1. \qquad (5.20)$$

Since

$$Pr\{\, L_t^{(m)}=0\,\}=2^{-m}$$

Thus, using (5.13), (5.19), and (5.20) we obtain

$$\overline{\pi^{(m)}} = \frac{1}{2}\cdot 2^{-m} + \frac{1}{2}\cdot\sum_{n=1}^{m-1}\left(\frac{1}{3\cdot 2^{m-n}} - \frac{1}{6\cdot 2^{m-n}}\cdot 2^{-2n+1}\right) + \frac{1}{4}\cdot\left(\frac{1}{3} - \frac{1}{6}\cdot 2^{-2m+1}\right)$$

$$\overline{r^{(m)}} = \frac{m+1}{2}\cdot 2^{-m} + \frac{1}{2}\cdot\sum_{n=1}^{m}\left(\frac{1}{3\cdot 2^{m-n}} - \frac{1}{6\cdot 2^{m-n}}\cdot 2^{-2n+1}\right)\cdot(m-n+1)$$

and simple calculations lead to (5.18). Q.E.D.

# 6 Acknowledgment

The author is grateful to Prof. Rudolf Ahlswede, Prof. Gerhard Schiffels, and Prof. Natalia Shekhunova for interesting discussions.

Table 1: Constructing the LCP of the sequence $\mathbf{u}_{15} = 011111110101010$ and the vector $\mathbf{S}$.

| $t$ | $u_t$ | $f_t(z)$ | $f_t$ | $d_t$ | $L_t$ | $S_j$ |
|---|---|---|---|---|---|---|
| 0 | | $1$ | 1 | 0 | 0 | $=$ |
| 1 | 0 | $1$ | 1 | 0 | 0 | |
| 2 | 1 | $1 + z^2 \cdot 1$ | 5 | 2 | 2 | |
| 3 | 1 | $1 + z^2 + z^{3-2} \cdot 1$ | 7 | 2 | 2 | |
| 4 | 1 | $1 + z + z^2 + z^{4-2} \cdot 1$ | 3 | 1 | 2 | $<$ |
| 5 | 1 | $1 + z$ | 3 | 1 | 2 | |
| 6 | 1 | $1 + z$ | 3 | 1 | 2 | |
| 7 | 1 | $1 + z$ | 3 | 1 | 2 | |
| 8 | 1 | $1 + z$ | 3 | 1 | 2 | |
| 9 | 0 | $1 + z + z^{9-2} \cdot 1$ | 203 | 7 | 7 | |
| 10 | 1 | $1 + z + z^7$ | 203 | 7 | 7 | |
| 11 | 0 | $1 + z + z^7$ | 203 | 7 | 7 | |
| 12 | 1 | $1 + z + z^7$ | 203 | 7 | 7 | |
| 13 | 0 | $1 + z + z^7$ | 203 | 7 | 7 | |
| 14 | 1 | $1 + z + z^7$ | 203 | 7 | 7 | $=$ |
| 15 | 0 | $1 + z + z^7$ | 203 | 7 | 7 | |

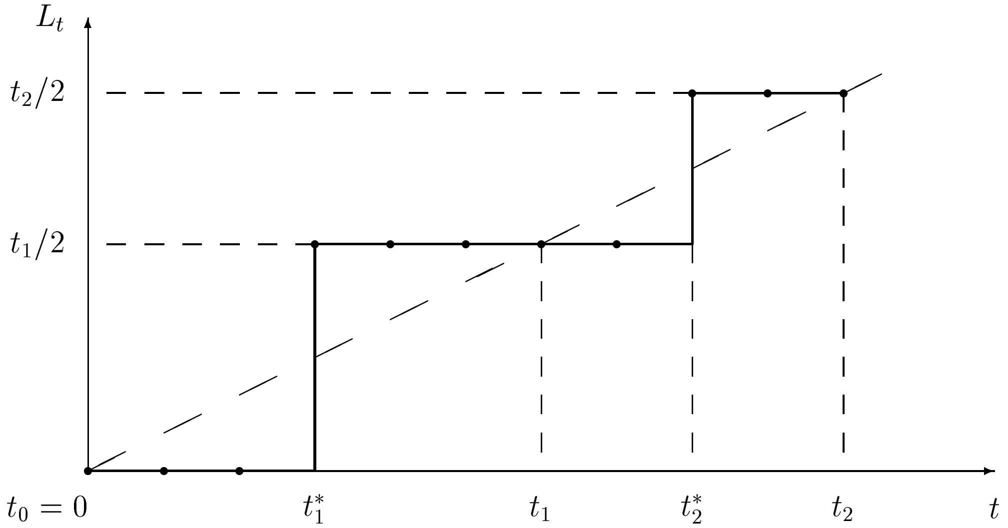


Figure 1: A possible LCP of binary sequences.

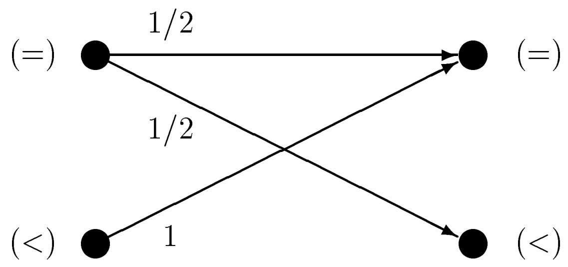


Figure 2: The transition state diagram of a Markov chain, which describes the relations between the length and the degree of the LFSRs in the points where the LCP crosses the line $t/2$.

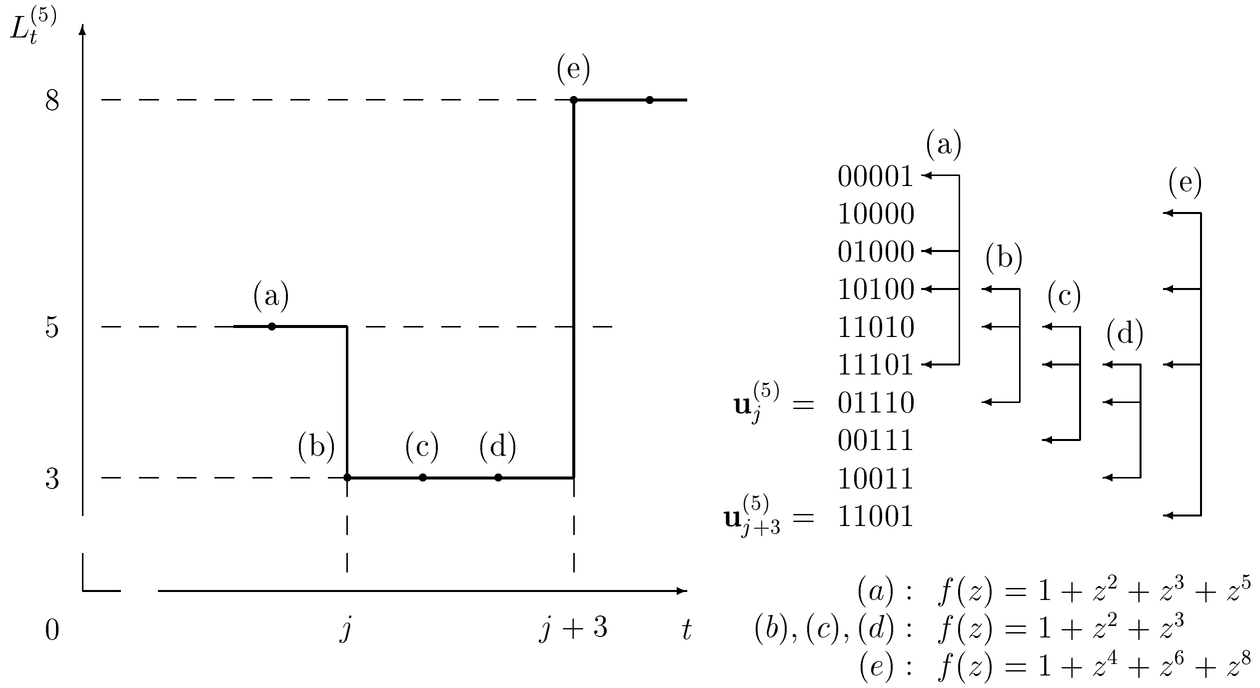


Figure 3: A fragment of the $m$-LCP of the sequence ...10000101110011...; $m = 5$.

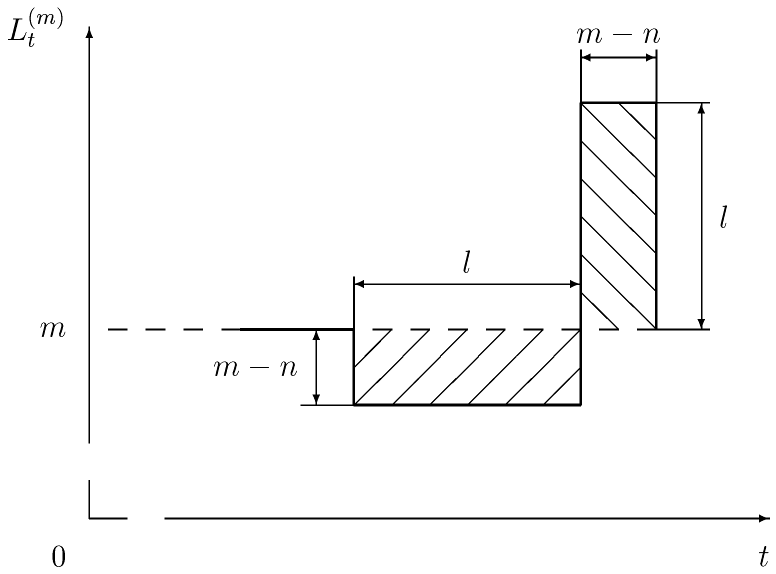


Figure 4: A fragment of the $m$-LCP of any binary sequence.

# References

[1] R.A.Rueppel, “New approaches to stream ciphers,” Ph.D.dissertation, Swiss Federal Institute of Technology, 1984.

[2] R.A.Rueppel, ”Linear complexity and random sequences,” In: *Lecture Notes in Computer Science, 219; Advances in Cryptology; Proc. Eurocrypt'85,* F.Pilcher, Ed., Linz, Austria, April 1985, pp.167-188, Berlin: Springer-Verlag, 1986.

[3] R.A.Rueppel, ”Stream ciphers,” In: *Contemporary Cryptology,* G.J.Simmons, Ed., New York : IEEE Press, pp.65-134, 1991.

[4] Z.Dai, ”Proof of Rueppel's linear complexity conjecture,” *IEEE Trans.Inform.Theory,* vol.32, pp.440-443, May 1986.

[5] H.Niederreiter, ”Keystream sequences with a good linear complexity profile for every starting point,” In: *Lecture Notes in Computer Science, 434; Advances in Cryptology; Proc. Eurocrypt'89,* J.-J.Quisqauter and J.Vandewalle, Eds., Hauthalen, Belgium, April 1989, pp.523-532, Berlin: Springer-Verlag, 1990.

[6] G.Carter, ”Enumeration results on linear complexity profiles,” In: *Cryptography and Coding II,* C.Mitchell, Ed., Oxford : Clarendon Press, pp.24-34, 1992.

[7] V.B.Balakirsky, ”Asymptotic lower bound on the free distance of constant linear convolutional codes with rate 1/n,” *Problemy Peredachi Informatsii,* vol.26, no.3, pp.3-11, 1990.
English translation - *Problems of Information Transmission,* vol.26, pp.181-188, New York : Plenum Publishing Corp., 1991.

[8] V.B.Balakirsky, ”On interval linear complexity of binary sequences”, In: *Proc. 1995 IEEE International Symposium on Information Theory,* p.490, Whistler, British Columbia, Canada, Sep.17-22 1995.

[9] R.E.Blahut, *Theory and Practice of Error Control Codes,* Section 7.4, New York : Addison-Wesley, 1984.

[10] J.L.Massey, "Shift register synthesis and BCH decoding," *IEEE Trans.Inform.Theory,* vol.15, pp.122-127, Jan. 1969.

[11] E.R.Berlekamp, *Algebraic Coding Theory,* New York : McGraw-Hill, 1968.

[12] R.L.Rivest, A.Shamir, and L.Adleman, "A method for obtaining digital signitures and public-key cryptosystems," *Commun.ACM,* vol.21, pp.120-126, 1978.

[13] G.H.Hardy and E.M.Wright, *An Introduction to the Theory of Numbers,* Chap.18, 4-th Ed., Oxford : Clarendon Press, 1960.