cryptography

MDPI

Review

A Survey on Complexity Measures for Pseudo-Random Sequences

Chunlei Li

Department of Informatics, University of Bergen, 5020 Bergen, Norway; chunlei.li@uib.no

Abstract: Since the introduction of the Kolmogorov complexity of binary sequences in the 1960s, there have been significant advancements on the topic of complexity measures for randomness assessment, which are of fundamental importance in theoretical computer science and of practical interest in cryptography. This survey reviews notable research from the past four decades on the linear, quadratic and maximum-order complexities of pseudo-random sequences, and their relations with Lempel–Ziv complexity, expansion complexity, 2-adic complexity and correlation measures.

Keywords: Kolmogorov complexity; randomness; feedback shift register; linear complexity; nonlinear complexity; Lempel–Ziv complexity; expansion complexity; 2-adic complexity; correlation

check for updates

Citation: Li, C. A Survey on Complexity Measures for Pseudo-Random Sequences. *Cryptography* **2024**, *8*, 25. https://doi.org/10.3390/cryptography8020025

Academic Editor: Josef Pieprzyk

Received: 5 May 2024
Revised: 9 June 2024
Accepted: 11 June 2024
Published: 13 June 2024

Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Cryptography and security applications make extensive use of random bits, such as keys and initialization vectors in encryption and nonces in security protocols. The generation of random bits should be designed with the security goal of *indistinguishability* from an ideal randomness source, which generates identically distributed and independent uniform random bits with full entropy (i.e., one bit of entropy per bit). However, this security goal is challenging to achieve. The list of real-world failures, where a random bit generator (RBG) is broken and the security of the reliant application crumbles with it, continues to grow [1–4]. There are two fundamentally different strategies for designing RBGs. One strategy is to produce bits non-deterministically, where every bit of output is based on a physical process that is unpredictable; the other strategy is to compute bits deterministically using an algorithm from an initial value that contains sufficient entropy to provide an assurance of randomness. This class of RBG is referred to as pseudo-random bit generators (PRBGs) or deterministic random bit generators (DRBGs) [5]. Due to their deterministic nature, PRBGs produce sequences of pseudo-random (rather than random) bits. Real-world PRBGs usually employ cryptographic primitives, such as stream ciphers, block ciphers, hash functions and elliptic curves, as their basic building blocks. For instance, the revised NIST SP 800-90A standard [5] recommends Hash-DRBG, HMAC-DRBG and CTR-DRBG based on approved hash functions and block ciphers. It is worth mentioning that, as pointed out in [2], the NIST SP 800-90A standard is not free from controversy: the disgraced algorithm DualEC-DRBG in the standard was reported to contain a back door [3]; the recommended DRBGs, when supporting a variety of optional inputs and parameters, do not fit cleanly into the usual security models of PRBGs; there is no formal competition in the standarization process; and there is a limited amount of formal analysis of the recommended DRBGs.Although not included in the NIST standard, PRBGs that use stream ciphers based on feedback shift registers (FSRs) as building blocks have several advantages, including the simple structure of operations of addition and multiplication, fast and easy hardware implementations in almost all computing devices, and good statistic characteristics (long period, balance, run properties, etc.) in output sequences [6,7]. In addition, the output sequence from other PRBGs will ultimately become periodic and therefore can be produced by certain FSRs. This provides another perspective for investigating the security strength of output sequences from PRBGs.

Consider a pseudo-random sequence $\mathbf{s} = s_0 s_1 s_2 \ldots$ generated from a PRBG. A natural question arises: how should one evaluate the *randomness* of the sequence **s**? Research on this problem can be traced back to the early 20th century. As early as 1919, Mises [8] initiated the notion of random sequences based on his frequency theory of probability. Later, Church [9] pointed out the vagueness of the second condition on randomness in [8] and proposed a less restricted but more precise definition of random sequences, as follows: An infinite binary sequence **s** is a random sequence if it satisfies the following two conditions: (1) assume $f(r, \mathbf{s})$ denotes the number of 1s among the first r terms of **s**, then the sequence $f(r, \mathbf{s})/r$ for **s** approaches a limit p as r approaches infinity; (2) for a sequence **b** with $b_1 = 1, b_{n+1} = 2b_n + s_n$ and any effectively calculable function ϕ, if the integers n such that $\phi(b_n) = 1$ form an infinite sequence $n_1, n_2, \cdots$, then the sequence $f(r, \mathbf{c})/r$ ofr $\mathbf{c} = s_{n_1} s_{n_2} s_{n_3} \ldots$ approaches the same limit p as r approaches infinity. Despite the emphasis of calculability of ϕ in Condition (2), one can see that it is hardly (if not at all) possible to test the randomness of a sequence by this definition. It could be used in randomness test by negative outcomes, namely, a binary sequence **s** is not random if one can find certain effectively calculable function ϕ, for which Condition (2) cannot hold. In order to justify a proposed definition of randomness, one has to show that the sequences, which are random in the stated sense, possess various properties of stochasticity with which we are acquainted in probability theory. Inspired by the properties obeyed by the maximum-length sequences, Golomb proposed the *randomness postulates* of binary sequences [10]: balancedness, run property (a subsequence of consecutive 1s/0s in a sequence is termed a run. A maximum-length sequence of length $2^m - 1$ contains one run of m 1s, one run of $(m-1)$ 0s and 2^k runs of $m - 2 - k$ 1s and 0s for $k = 0, 1, \ldots, m-3$) and ideal autocorrelation. Kolmogorov [11] proposed the notion of *complexity* to test the randomness of a sequence **s**, which is defined as $\min_{A(\mathbf{x})=\mathbf{s}} \text{len}(\mathbf{x})$, the length of the shortest **x** that can produce **s** by a universal Turing machine program A. This notion is later referred to as the Kolmogorov complexity in the literature.Along this line, further developments were made in [12,13] and summarized in Knuth's famous book series *The Art of Computer Programming* [14].

Rather than considering an abstract Turing machine program that generates a given sequence, the model of using FSRs to generate a given sequence has attracted considerable attention. This model has two major advantages: firstly, all sequences that are generated by a finite-state machine in a deterministic manner are ultimately periodic, and as such, can be produced by certain finite-state shift registers; secondly, it is comparatively easier and more efficient to identify the shortest FSRs producing a given sequence, either of a finite length or of infinite length with a certain period. Due to its tractability, the *linear complexity* of a specific sequence **s**, which uses linear FSRs (LFSRs) as the algorithm A in the Kolmogorov complexity, is particularly appealing and has been intensively studied. The linear complexity of a given sequence **s** can be efficiently calculated by the Berlekamp–Massey algorithm [15,16], and the stochastic behavior of linear complexity for finite-length sequences and periodic sequences can be considered to be fairly well understood [17–19]. The good understanding of linear complexity of sequences is an important factor for its adoption in the NIST randomness test suite [20]. Note that extremely long sequences with large linear complexity can usually be generated by much short FSRs with certain nonlinear feedback function. Consequently, more figures of merit to judge the randomness of sequences, such as maximum-order complexity [21], 2-adic complexity [22], quadratic complexity [23], expansion complexity [24] and their variants, have been also explored in the literature.

This paper will survey the development of complexity measures used to assess the randomness of sequences. It is important to note that this paper does not intend to offer a comprehensive survey of this broad topic. Instead, it aims to provide a preliminary overview of the topic with technical discussions to certain extent, focusing mainly on complexities within the domain of FSR that align with the author's research interests. To this end, I delved into significant findings on the study of FSR-based complexity measures, particularly on the algorithmic and algebraic methods in computation, statistical behavior,

theoretical constructions and the relations of those complexity measures. Research papers on this topic presented at flagship conferences in cryptography, such as ASIACRYPT, CRYPTO and EUROCRYPT, published at prestigious journals with IEEE, Springer, Science Direct, etc, as well as some well-recognized books since the 1970s constitute the major content of this survey. Readers may refer to relevant surveys on this topic with different focuses by Meidl and Niederreiter [19,25] and Topuzoğlus and Winterhof [26].

The remainder of this paper is organized as follows: Section 2 recalls the basics for feedback shift registers, which lays the foundation for the discussions in subsequent sections. Section 3 reviews important developments of the theory of linear complexity profile for random sequences, Section 4 discusses a mathematical tool used to efficiently calculate the quadratic complexity of sequences and Section 5 surveys computational methods, the statistical behavior for maximum-order complexity and theoretical constructions of periodic sequences with the largest maximum-order complexity. In Section 6, known relations among complexity measures of sequences are briefly summarized. Finally, conclusive remarks and discussion are given in Section 7.

2. Feedback Shift Registers

Let $\mathbb{F}_q$ be the finite field of q elements, where q is an arbitrary prime power. For a positive integer m, an m-stage feedback shift register (FSR) is a clock-controlled circuit consisting of m consecutive storage units and a feedback function f as displayed in Figure 1.

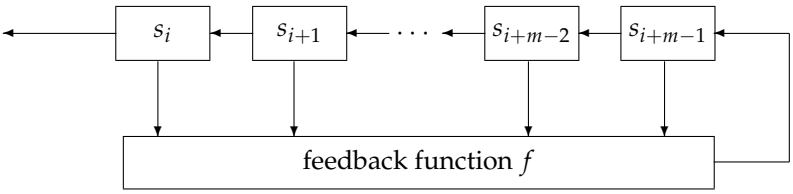


Figure 1. An m-stage FSR with feedback function f.

Starting with an initial state $s_0 s_1 \dots s_{m-1}$ over $\mathbb{F}_q$, the states in the FSR will be updated by a clock-controlled transformation as follows:

$$\mathcal{F} : s_i s_{i+1} \dots s_{i+m-1} \longmapsto s_{i+1} \dots s_{i+m-1} s_{i+m}, \; i \geq 0, \tag{1}$$

where

$$s_{i+m} = f(s_i, s_{i+1}, \dots, s_{i+m-1}),$$

and the leftmost symbol for each state will be output. In this way an FSR produces a sequence $\mathbf{s} = s_0 s_1 s_2 \dots$ based on each initial state $s_0 s_1 \dots s_{m-1}$ and its feedback function f. The output sequence from an FSR, known as an FSR sequence, can be equivalently expressed as a sequence of states $\underline{\mathbf{s}}_i = s_i s_{i+1} \dots s_{i+m-1}$, with the relation $\underline{\mathbf{s}}_i = \mathcal{F}(\underline{\mathbf{s}}_{i-1}) = \cdots = \mathcal{F}^i(\underline{\mathbf{s}}_0)$ for $i \geq 0$. When $\underline{\mathbf{s}}_p = \mathcal{F}^p(\underline{\mathbf{s}}_0) = \underline{\mathbf{s}}_0$ for the least integer $p \geq 1$, we obtain a cycle of states $\underline{\mathbf{s}}_0, \dots \underline{\mathbf{s}}_{p-1}$, or equivalently a sequence $s_0 \dots s_{p-1} \dots$ of least period p. In his influential book [10], Golomb intensively studied the properties of FSR sequences from both linear feedback functions and nonlinear feedback functions. Readers can refer to [10] for a comprehensive understanding of FSR sequences.

For an m-stage FSR with a feedback function f and an initial state $s_0 s_1 \dots s_{m-1}$, if we know n consecutive $s_i s_{i+1} \dots s_{i+n-1}$ in the output $\mathbf{s}$, then we obtain $n - m$ equations

$$\begin{aligned} f(s_i, \dots, s_{i+m-1}) &= s_{i+m} \\ f(s_{i+1}, \dots, s_{i+m}) &= s_{i+m+1} \\ &\vdots \\ f(s_{i+n-m-1}, \dots, s_{i+n-2}) &= s_{i+n-1}. \end{aligned} \tag{2}$$

When the feedback function f is linear, namely, $f(x_0, \ldots, x_{m-1}) = a_1 x_{m-1} + a_2 x_{m-2} + \cdots + a_m x_0$, and $n \geq 2m$, one can uniquely determine its m unknown coefficients a_i from the above $n - m$ linear equations. When the feedback function f is nonlinear, the number of terms in f increases significantly. For instance, for $q = 2$ a binary feedback function f of degree r has $\sum_{d=0}^{r} \binom{n}{d}$ possible terms. The above equations in Equation (2) become more difficult to analyze. The equations are not necessarily linearly independent and a lot more variables are involved. Consequently, more observations and techniques are required in the analysis of these equations. In the following, we will review some results on FSR sequences when the feedback function is linear, quadratic and arbitrarily nonlinear, as well as their relations with other complexity measures.

We will consider both finite-length sequences and infinite-length sequences with certain finite period. Throughout what follows, we will use $\mathbf{s}$ to denote a generic sequence with terms from certain alphabet, $\mathbf{s}_n = s_0 s_1 \ldots s_{n-1}$ to denote a sequence with length n that is not a repetition of any shorter sequence, $\mathbf{s}_n^k$ to denote a sequence of k repetition of $\mathbf{s}_n$ and $\mathbf{s}_n^\infty$ to denote infinite repetitions of $\mathbf{s}_n$, indicating that $\mathbf{s}_n^\infty$ has least period n.

3. Linear Complexity

Given an FSR, when its feedback function f is a linear function on the input state, namely, it is associated with the following linear recurrence:

$$s_{i+m} + a_1 s_{i+m-1} + \cdots + a_{m-1} s_{i+1} + a_m s_i = 0 \text{ for } i = 1, 2, \ldots, \tag{3}$$

where $a_1, \ldots, a_m$ are taken from $\mathbb{F}_q$, it is termed linear FSR and the output sequence $\mathbf{s}$ is called an LFSR sequence of order m. The polynomial

$$f(x) = x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0,$$

with $a_0 = 1$, is called the *feedback or characteristic polynomial* of $\mathbf{s}$. The zero sequence $00\ldots$ is viewed as an LFSR sequence of order 0.

Definition 1. *Let* $\mathbf{s}$ *be an arbitrary sequence of elements of* $\mathbb{F}_q$ *and let* n *be a non-negative integer. Then the linear complexity* $L_n(\mathbf{s})$ *is defined as the length of the shortest LFSR that can generate* $\mathbf{s}_n = s_0 s_1 \ldots s_{n-1}$*. The sequence* $L_0(\mathbf{s}), L_1(\mathbf{s}), L_2(\mathbf{s}), \ldots$*, is called the linear complexity profile of the sequence* $\mathbf{s}$*.*

It is clear that $L_n(\mathbf{s}) \leq L_{n+1}(\mathbf{s})$ and $0 \leq L_n(\mathbf{s}) \leq n$ for any integer n and sequence $\mathbf{s}$. Hence, the linear complexity profile of $\mathbf{s}$ is a nondecreasing sequence of non-negative integers. Two extreme cases for $L_n(\mathbf{s})$ correspond to highly non-random sequences $\mathbf{s}$ whose first n elements $s_0 s_1 \ldots s_{n-1}$ are either $\overbrace{0 \ldots 0}^{n}$ or $\overbrace{0 \ldots 0}^{n-1} s_{n-1}$ with $s_{n-1} \neq 0$, which has $L_n(\mathbf{s}) = 0$ or $L_n(\mathbf{s}) = n$, respectively.

The linear complexity of a sequence $\mathbf{s}$ can be efficiently calculated by the well-known Berlekamp–Massey algorithm [15,16], which also returns the linear feedback function generating $\mathbf{s}$. In the Berlekamp–Massey algorithm, at each step $n - 1$ with the current linear recurrence for $s_0 s_1 \ldots s_{n-2}$, a discrepancy is calculated to assess whether the linear recurrence holds for the extended sequence $s_0 s_1 \ldots s_{n-2} s_{n-1}$: when the discrepancy is zero, indicating that the current linear recurrence indeed holds for $s_0 s_1 \ldots s_{n-2} s_{n-1}$, then move on to n; when the discrepancy is nonzero, indicating that a new (and possibly longer) linear recurrence is needed, the linear recurrence is then updated and the linear complexity is updated as

$$L_n(\mathbf{s}) = \max\{L_{n-1}(\mathbf{s}), n - L_{n-1}(\mathbf{s})\}. \tag{4}$$

Gustavson [27] showed that the above process requires $O(n^2)$ multiplication/addition operations for calculating $L_n(\mathbf{s})$. Dornstetter, in [28], pointed out the equivalence between this procedure of calculation and the Euclidean algorithm.

Rueppel [17] investigated the varying behavior of the discrepancy Δ_t, $1 \le t < n$, for a binary sequence of length n, and characterized the linear complexity profile of random sequences in the following propositions.

Proposition 1. *The number of sequences of length n and linear complexity l, denoted by $N_n(l)$, satisfies the following recursive relation*

$$N_n(l) = \begin{cases} N_{n-1}(l), & 0 \le l < \frac{n}{2}, \\ 2N_{n-1}(l), & l = \frac{n}{2}, \\ 2N_{n-1}(l) + N_{n-1}(n-l), & \frac{n}{2} < l < n, \end{cases}$$

and it, starting from $N_1(0) = N_1(1) = 1$, can be given by

$$N_n(l) = \begin{cases} 2^{\min\{2n-2l, 2l-1\}}, & 0 < l \le n, \\ 1, & 0 = l < n. \end{cases}$$

Proposition 2. *The expected linear complexity L_n of binary random sequences of length n drawn from a uniform distribution is given by*

$$E[L_n] = \frac{n}{2} + \frac{4 + R_2(n)}{18} - 2^{-n}\left(\frac{n}{3} + \frac{2}{9}\right),$$

and the variance of the linear complexity is given by

$$\mathrm{Var}[L_n] = \frac{86}{81} - 2^{-n}\frac{3n(14 - R_2(n)) - (82 - 2R_2(n))}{81} - 2^{-2n}\frac{9n^2 + 12n + 4}{81},$$

where $R_2(n)$ denotes n modulo 2.

From the above proposition, we readily see that

$$\lim_{n\to\infty} E[L_n] = \frac{n}{2} + c_n \text{ and } \lim_{n\to\infty} \mathrm{Var}[L_n] = \frac{86}{81}, \tag{5}$$

where $c_n = (4 + R_2(n))/18$. From Rueppel's characterization on the general stochastic behavior of binary random sequences, a similar and more important question arose: for a randomly chosen and then fixed sequence $\mathbf{s}$ over $\mathbb{F}_q$, what is the behavior of $L_n(\mathbf{s})$? To settle this question, Niederreiter [18] developed a probabilistic theory for linear complexity and linear complexity profiles of a given sequence $\mathbf{s}$ over any finite field $\mathbb{F}_q$. More specifically, he applied techniques from probability theory to dynamic systems and continued fractions, and deduced the probabilistic limit theorems for linear complexity by exploiting the connection between continued fractions and linear complexity of a sequence $\mathbf{s}$. Below we first recall the connection discussed in [18].

Given a sequence $\mathbf{s} = s_0 s_1 s_2 \ldots$, the fractional function $G(x) = s_0 x^{-1} + s_1 x^{-2} + s_2 x^{-3} + \cdots \in \mathbb{F}_q[x^{-1}]$ has the following continued fraction expression

$$G = 1/(A_1 + 1/(A_2 + \cdots)),$$

where the partial quotients $A_1, A_2, \ldots$ are polynomials over $\mathbb{F}_q$ with positive degrees. The n-th linear complexity of $\mathbf{s}$ can be expressed as

$$L_n(\mathbf{s}) = \sum_{i=1}^{j} A_i(\mathbf{s})$$

for the integer j satisfying

$$\sum_{i=1}^{j-1} A_i(\mathbf{s}) + \sum_{i=1}^{j} A_i(\mathbf{s}) \leq n \leq \sum_{i=1}^{j} A_i(\mathbf{s}) + \sum_{i=1}^{j+1} A_i(\mathbf{s}).$$

By the study of the above continued fraction expansion, Niederreiter obtained the following result on $L_n(\mathbf{s})$.

Theorem 1. *Suppose f is a non-negative nondecreasing function on the positive integers with $\sum_{n=1}^{\infty} q^{-f(n)} < \infty$. Then, we have*

$$\left| L_n(\mathbf{s}) - \frac{n}{2} \right| \leq \frac{1}{2} f(n) \text{ for all sufficiently large } n.$$

By taking $f(n) = (1+\epsilon)\log n/\log(q)$ for arbitrary $\epsilon > 0$, a more explicit law of the logarithm for the n-th linear complexity of a random sequence $\mathbf{s}$ can be obtained as follows.

$$\begin{aligned} \overline{\lim_{n\to\infty}} \frac{L_n(\mathbf{s}) - \frac{n}{2}}{\log n} &= \frac{1}{2\log q}, \\ \underline{\lim_{n\to\infty}} \frac{L_n(\mathbf{s}) - \frac{n}{2}}{\log n} &= -\frac{1}{2\log q}. \end{aligned} \tag{6}$$

Consequently, we have

$$L_n(\mathbf{s}) = \frac{n}{2} + O(\log(n)). \tag{7}$$

As the equalities in (6) hold for infinitely many n, there is a strong interest in sequences $\mathbf{s}$ whose n-th linear complexity has small derivations from $n/2$.

Definition 2. *Let d be a positive integer. A sequence $\mathbf{s} = s_0 s_1 s_2 \ldots$ is said to be d-perfect if*

$$\left| L_n(\mathbf{s}) - \frac{n}{2} \right| \leq \frac{d}{2} \text{ for any } n \geq 1.$$

A 1-perfect sequence is also called perfect. A sequence is called almost perfect if it is d-perfect for some integer $d > 1$.

Constructing d-perfect sequences is of significant interest. Sophisticated methods based on algebraic curves were introduced in [29], which yielded sequences with almost perfect linear complexity profiles. For cryptographic applications, especially as a keystream, a sequence $\mathbf{s}$ should have a linear complexity profile close to that of a random sequence. Moreover, this condition should be true for any starting point of the sequence. That is to say, for a sequence $\mathbf{s} = s_0 s_1 s_2 \ldots$ and any $r \geq 0$, the r-shifted sequence $s_r s_{r+1} s_{r+2} \ldots$ should have an acceptable linear complexity profile close to random sequences as well. Niederreiter and Vielhaber [30] attacked this problem with the help of continued fractions. An important fact is that the jumps in the linear complexity profile of $\mathbf{s}$ are exactly the degrees of the partial quotients $A_1, A_2, \ldots$. With such a connection, they proposed an algorithm to determine the linear complexity profile of shifted sequences $s_r s_{r+1} s_{r+2} \ldots$ by investigating the corresponding continued fractions [31]. To be more concrete, they designed a method to calculate the continued fraction expansions of $G_n(x), \ldots, G_1(x)$, where $G_r(x) = \sum_{j=r}^{n-1} s_j x^{-(j+1)}$ for $r \geq 0$. Later, in his survey paper [19], Niederreiter proposed the following open problem on d-perfect sequences.

Problem 1. *Construct a sequence $\mathbf{s} = s_0 s_1 s_2 \ldots$ over $\mathbb{F}_q$ such that the r-shifted sequences $s_r s_{r+1} s_{r+2} \ldots$ for all $r \geq 0$ are d-perfect for some integer $d \geq 1$.*

The preceding discussion is mainly concerned with generic infinite-length sequences. Note that sequences from an m-stage LFSR over $\mathbb{F}_q$ are periodic and their maximum period is $q^m - 1$, which is achieved when the characteristic polynomial $f(x)$ is a primitive polynomial in $\mathbb{F}_q[x]$. LFSR sequences of order m with period $q^m - 1$ are thus known as maximum-length sequences, or m-sequences for short. For a periodic sequence $\mathbf{s} = \mathbf{s}_n^\infty$, the values in its linear complexity profile will remain unchanged at a certain point. This reveals apparently different linear complexity profiles between an infinite-length sequence of certain period and a random infinite-length sequence, which basically can be considered to have an arbitrarily long period.

An important tool for the analysis of the linear complexity of n-periodic sequences over $\mathbb{F}_q$ is the discrete Fourier transform. Assume $\gcd(n, q) = 1$, which means that there exists an n-th primitive root β of unity in some finite extension of $\mathbb{F}_q$. Then the *discrete Fourier transform* of a time-domain n-tuple $(a_0, \ldots, a_{n-1}) \in \mathbb{F}_q^n$ is the (frequency-domain) n-tuple $(b_0, \ldots, b_{n-1})$ with $b_j = \sum_{i=0}^{n-1} a_i \beta^{ij}$ for $j = 0, \ldots, n-1$, i.e.,

$$\mathrm{DFT}((a_0, \ldots, a_{n-1})) = (a_0, \ldots, a_{n-1}) \begin{pmatrix} 1 & \beta & \ldots & \beta^{n-1} \\ 1 & \beta^2 & \ldots & \beta^{2(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \beta^{n-1} & \ldots & \beta^{n-1(n-1)} \end{pmatrix}. \tag{8}$$

In this case, the linear complexity of an n-periodic sequence can be determined via the discrete Fourier transform [32].

Proposition 3. *Let* $\mathbf{s} = s_0 s_1 s_2 \ldots$ *be an* n*-periodic sequence over* $\mathbb{F}_q$ *with* $\gcd(n, q) = 1$*. Let* β *be an* n*-th primitive root of unity in an extension of* $\mathbb{F}_q$ *and* $(b_0, \ldots, b_{n-1})$ *be the discrete Fourier transform of* $(s_0, s_1, \ldots s_{n-1})$ *as in Equation* (8)*. Then,*

$$L_n(\mathbf{s}) = \#\{j : b_j \neq 0, 0 \leq j < n\} = n - \#\left\{ j : \sum_{i=0}^{n-1} s_i \beta^{ij} = 0 \right\}.$$

Massey and Serconek in [33] further extended the above relation to the general case $\gcd(q, n) > 1$ with the generalized discrete Fourier transform. The *generalized discrete Fourier transform* of $(s_0, s_1, \ldots, s_{n-1})$, where $n = p^v w$, $\gcd(w, p) = 1$ and p is the characteristic of $\mathbb{F}_q$, is defined as the following $p^v \times n$ matrix

$$\mathrm{GDFT}((s_0, s_1, \ldots, s_{n-1})) = \begin{pmatrix} s(1) & s(\beta) & \cdots & s(\beta^{n-1}) \\ s^{[1]}(1) & s^{[1]}(\beta) & \cdots & s^{[1]}(\beta^{n-1}) \\ \vdots & & & \\ s^{[p^v-1]}(1) & s^{[p^v-1]}(\beta) & \cdots & s^{[p^v-1]}(\beta^{n-1}) \end{pmatrix}$$

where β is any n-th primitive root of unity over $\mathbb{F}_q$ and

$$s^{[t]}(x) = \sum_{i=0}^{n-1} \binom{i}{t} s_i x^{i-t}$$

is the t-th Hasse derivative of the polynomial $s_0 + s_1 x + \cdots + s_{n-1} x^{n-1}$. It is clear that when $\gcd(n, q) = 1$, the GDFT of $(s_0, s_1, \ldots, s_{n-1})$ reduced to the discrete Fourier transform of $(s_0, s_1, \ldots, s_{n-1})$. As pointed out in [33,34], the linear complexity of $\mathbf{s} = \mathbf{s}_n^\infty$ can be calculated in terms of the Günther weight of $\mathrm{GDFT}((s_0, s_1, \ldots, s_{n-1}))$, which was referred to as the Günther–Blahut theorem, which was used by Blahut implicitly in [35].

Proposition 4. *Let* $\mathbf{s} = \mathbf{s}_n^\infty$ *be an* n*-periodic sequence over* $\mathbb{F}_q$ *with characteristic* p*, where* $n = p^v w$ *with* $\gcd(w, p) = 1$*. Then, the linear complexity of* $\mathbf{s}$ *is equal to the Günther weight of the GDFT of*

the n-tuple $(s_0, s_1, \ldots, s_{n-1})$*, where the Günther weight of a matrix M is the number of its entries that are nonzero.*

Algebraically, the linear complexity of $\mathbf{s}_n^\infty$ can be obtained via the feedback polynomial of the LFSR generating $\mathbf{s}$ [36]. Consider the generating function

$$g(x) = \sum_{i=0}^{\infty} s_i x^i = \frac{s_0 + s_1 x + \cdots + s_{n-1}x^{n-1}}{1 - x^n}.$$

Assume the characteristic polynomial of $\mathbf{s}$ is given by $f(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0$ with $a_0 = 1$. Then

$$\begin{aligned} f(x)g(x) &= s_0 + (s_1 + a_1 s_0)x + \cdots + (s_{m-1} + a_1 s_{m-2} + \cdots + a_{m-1}s_0)x^{m-1} \\ &\quad + \sum_{i \geq 0} (s_{i+m} + a_1 s_{i+m-1} + \cdots + a_m s_i)x^{i+m} \\ &= s_0 + (s_1 + a_{m-1}s_0)x + \cdots + (s_{m-1} + a_{m-1}s_{n-2} + \cdots + a_1 s_0)x^{m-1}. \end{aligned}$$

Letting $\varphi(x) = -f(x)g(x)$, we have

$$g(x) = -\frac{\varphi(x)}{f(x)} = -\frac{\varphi_0(x)}{f_0(x)},$$

where $\gcd(\varphi_0, f_0) = 1$ and f_0 is the minimal polynomial of $\mathbf{s}$. This relation implies the following equality [36]:

$$f_0(x)s_n(x) = (x^n - 1)\varphi_0(x).$$

where $s_n(x) = s_0 + s_1 x + \cdots + s_{n-1}x^{n-1}$. Since $\gcd(\varphi_0, f_0) = 1$, it follows that $f_0(x) | (x^n - 1)$ and $\varphi_0(x) | s_n(x)$. Therefore, we have the following result.

Proposition 5. *Let* $\mathbf{s} = \mathbf{s}_n^\infty$ *be an n-periodic sequence over* $\mathbb{F}_q$ *and* $s(x)$ *be the associated function of* $\mathbf{s}$ *given by* $s_n(x) = s_0 + s_1 x + \cdots + s_{n-1}x^{n-1}$*. Then the minimal polynomial of* $\mathbf{s}$ *is given by*

$$f_0(x) = \frac{x^n - 1}{\gcd(x^n - 1, s_n(x))},$$

indicating that

$$L(\mathbf{s}) = n - \deg(\gcd(x^n - 1, s_n(x))).$$

With the above neat expression of $L(\mathbf{s})$ for periodic sequences $\mathbf{s}_n^\infty$, several families of sequences with a nice algebraic structure were investigated, such as Legendre sequences [37], discrete logarithm function [38], Lempel–Cohn–Eastman sequences [39]. For more information about linear complexity of sequences with special algebraic structures, readers are referred to Shparlinski's book [40].

The above discussion is concerned with the explicit calculation of an n-periodic sequence. The statistical behavior of a random n-periodic sequence over $\mathbb{F}_q$ is of fundamental interest, particularly from the cryptographic perspective. Rueppel in [41] considered this problem: if we let $\mathbf{s} = \mathbf{s}_n^\infty$ run through all n-periodic sequences over $\mathbb{F}_q$, what is the expected linear complexity of

$$E_{q,n} = \frac{1}{q^n} \sum_{\mathbf{s}} L(\mathbf{s})?$$

For the case $q = 2$, Rueppel showed that if n is a power of 2, then $E_{2,n} = n - 1 + 2^{-n}$; and if $n = 2^m - 1$ with a prime m, then $E_{2,n} \gtrapprox e^{-\frac{1}{n}}\left(n - \frac{1}{2}\right)$. Based on observations on numerical results, he conjectured that $E_{2,n}$ is always close to n. There was little progress on this conjecture until the work by Meidl and Niederreiter [42], in which they studied $E_{q,n}$ for an arbitrary prime power q with the help of the above Günther–Blahut theorem and the

analysis of cyclotomic cosets. For an integer $0 \le j < w$, the cyclotomic coset of j modulo w is defined as $C_j = \{jq^t \pmod{w} : t \ge 0\}$.

Theorem 2. *Let $n = p^v w$, where p is the characteristic of $\mathbb{F}_q, v \ge 0$ and $\gcd(p, w) = 1$. Let $C_1, \ldots, C_h$ be the different cyclotomic cosets modulo w and $b_i = |C_i|$ for $1 \le i \le h$. Then*

$$E_{q,n} = n - \sum_{i=1}^{h} \frac{b_i\left(1 - q^{-b_i p^v}\right)}{q^{b_i} - 1}.$$

From Theorem 2, a routine inequality scaling

$$\sum_{i=1}^{h} \frac{b_i\left(1 - q^{-b_i p^v}\right)}{q^{b_i} - 1} < \sum_{i=1}^{h} \frac{b_i}{q^{b_i} - 1} < \frac{1}{q-1} \sum_{i=1}^{h} b_i = \frac{w}{q-1}$$

implies that

$$E_{q,n} \ge n - \frac{w}{q-1}.$$

In the particular case $v = 0$,

$$E_{q,n} \ge n - \frac{w}{q-1} = n(1 - \frac{1}{q-1}).$$

For small q, the above lower bounds can be further improved. For instance, for $q = 2$, as the only singleton coset is $C_0 = \{0\}$, it follows that

$$E_{2,n} \ge \begin{cases} n - \frac{w-2}{3} \text{ for any } n, \\ \frac{3n-1}{4} \text{ for odd } n. \end{cases}$$

Recall that the expected linear complexity of a random binary sequence $s_0 s_1 \ldots s_{n-1}$ is around $\frac{n}{2}$. For a periodic sequence $\mathbf{s} = \mathbf{s}_n^\infty$ from a sequence $\mathbf{s}_n$ with linear complexity l, the calculation of linear complexity of $\mathbf{s}_n^\infty$ needs to consider the sequence $s_0 s_1 \ldots s_{n-1} s_0 \ldots s_{l-1}$. This expansion sequence, when considered as a random sequence, would have expected complexity around $(n + l)/2$. This is somewhat consistent with $E_{2,n} = (3n - 1)/4$ when $l = (n - 1)/2$ for odd n. Further improved results for the case $\gcd(q, n) = 1$ were obtained by more detailed analysis and can be found in [42].

4. Quadratic Complexity

Quadratic feedback functions are the easiest nonlinear case and has been discussed by some researchers [43–45]. Chan and Games, in [43], studied the quadratic complexity of binary DeBruijn sequences of order m (in which each binary string $s_0 s_1 \ldots s_{m-1}$ appears exactly once); Youssef and Gong characterized a jump property of the quadratic complexity profile of random sequences, and Rizomiliotis et al., in [45], proposed an efficient method to calculate the quadratic complexity of binary sequences in general. We first recall the definition of quadratic complexity of a sequence in the following:

Definition 3. *Given a binary sequence $\mathbf{s} = s_0 s_1 s_2 \ldots$, its quadratic complexity is the length of the shortest FSRs with quadratic feedback functions*

$$f(x_0, x_1, \ldots, x_{m-1}) = \sum_{0 \le i,j < m} a_{i,j} x_i x_j + \sum_{0 \le i < m} a_i x_i + a, \quad a_{i,j}, a_i, a \in \mathbb{F}_2$$

that can generate $\mathbf{s}$. The quadratic complexity profile of the sequence $\mathbf{s}$ is similarly defined as $Q_0(\mathbf{s}), Q_1(\mathbf{s}), Q_2(\mathbf{s}), \ldots$, where $Q_n(\mathbf{s})$ is the quadratic complexity of the first n terms of $\mathbf{s}$.

For the first n terms $s_0 s_1 \dots s_{n-1}$ of $\mathbf{s}$, suppose it can be generated by an m-stage quadratic FSR. Following the recurrence as in Equation (2), we can obtain a system of $n-m$ equations. More concretely, denote

$$\begin{aligned} F(m) &= (a, a_0, \overbrace{a_1, a_{0,1}}, \overbrace{a_2, a_{1,2}, a_{0,2}}, \dots, \overbrace{a_{m-1}, a_{m-2,m-1}, \dots, a_{0,m-1}})^T, \\ E(n,m) &= (s_m, s_{m+1}, \dots, s_{n-1})^T, \\ M(n,m) &= (\mathbf{1}_{n-m}^T | B_0(m,n) | \dots | B_{m-1}(n,m)), \end{aligned}$$

where, for $0 \le i < m$,

$$B_i(n,m) = \begin{pmatrix} s_i & s_i s_{i-1} & \cdots & s_i\, s_0 \\ s_{i+1} & s_{i+1} s_i & \cdots & s_{i+1} s_1 \\ \vdots & \vdots & & \vdots \\ s_{i+k-1} & s_{i+k-1} s_{i+k-2} & \cdots & s_{i+k-1} s_{k-1} \end{pmatrix}.$$

with $k = n - m$. According to the ordering in $F(m)$ and $M(n,m)$, it is easy to verify that

$$M(n,m) F(m) = E(n,m). \tag{9}$$

Therefore, finding a quadratic feedback function in m variables that outputs the sequence $\mathbf{s}_n = s_0 s_1 \dots s_{n-1}$ is equivalent to solving the above system of linear equations in $1 + \frac{m(m+1)}{2}$ variables in $F(m)$. Notice that the ordering of linear and quadratic terms in $M(n,m)$ has a feature that the quadratic terms $s_i s_j$ with $j < i$ occur in $M(n,m)$ only after the term s_i occurs. This feature facilitates the calculation of $Q_n(\mathbf{s})$ term by term as in the Berlekamp–Massey algorithm.

We first give a toy example to illustrate the above linear equations. Let $n = 9$ and $m = 3$. Then we have

$$\begin{aligned} M(n,m) &= (\mathbf{1}_6^T | B_0(9,3) | B_1(9,3) | B_2(9,3)) \\ &= \begin{pmatrix} 1 & s_0 & s_1 & s_0 s_1 & s_2 & s_1 s_2 & s_0\, s_2 \\ 1 & s_1 & s_2 & s_1 s_2 & s_3 & s_2 s_3 & s_1\, s_3 \\ 1 & s_2 & s_3 & s_2 s_3 & s_4 & s_3 s_4 & s_2\, s_4 \\ 1 & s_3 & s_4 & s_3 s_4 & s_5 & s_4 s_5 & s_3\, s_5 \\ 1 & s_4 & s_5 & s_4 s_5 & s_6 & s_5 s_6 & s_4\, s_6 \\ 1 & s_5 & s_6 & s_5 s_6 & s_7 & s_6 s_7 & s_5\, s_7 \end{pmatrix}, \end{aligned}$$

and

$$F(3) = (a, a_0, a_1, a_{0,1}, a_2, a_{1,2}, a_{0,2})^T, \quad E(9,3) = (s_3, s_4, s_5, s_6, s_7, s_8)^T.$$

From the relation $M(9,3)F(3) = E(9,3)$, we obtain 6 equations with 7 variables in $F(3)$.

It is straightforward to see that the system is solvable if and only if

$$\operatorname{rank}(M(n,m)) = \operatorname{rank}(M(n,m) \mid E(n,m)).$$

Chan and Games in [43] had the following observation.

Theorem 3. *If an m-stage quadratic FSR can generate $\mathbf{s}_{n-1}$ but not $\mathbf{s}_n$, then there is no m-stage quadratic FSR that can generate $\mathbf{s}_n$ if and only if*

$$\operatorname{rank}(M(n,m)) = \operatorname{rank}(M(n-1,m)).$$

By this theorem, for each step, when $s_0 s_1 \dots s_{n-2}$ has quadratic complexity m, if $\operatorname{rank}(M(n,m)) \neq \operatorname{rank}(M(n-1,m))$, then $s_0 s_1 \dots s_{n-2} s_{n-1}$ for any s_{n-1} also has the same quadratic complexity m. Based on such an observation, they proposed a term-by-term algo-

rithm, similarly to the Berlekamp–Massey algorithm, to calculate the quadratic complexity of a sequence $\mathbf{s}_n$ [23]. Their algorithm strongly depends on the Gaussian elimination for the computation of $\text{rank}(M(n,m))$ for each new term. The special structure of the matrix $M(n,m)$ was not well taken into consideration in their algorithm, which did not reveal the precise increment of the quadratic complexity of $\mathbf{s}_n$. By looking into certain structure of the matrix $M(n,m)$, Youssef and Gong [44] showed that if the quadratic complexity of the first n terms of a sequence $\mathbf{s}$, is larger than $n/2$, then the first $n+1$ terms of $\mathbf{s}$ has the same quadratic complexity. Rizomiliotis, Kolokotronis and Kalouptsidis in [45] observed the following nesting structure of the matrix $M(n,m)$:

$$\begin{aligned} M(n,m) &= \begin{pmatrix} M(n-1,m) \\ R_{n-m-1}(n,m) \end{pmatrix} = (M(n-1,m-1) \mid B_{m-1}(n,m)) \\ &= \begin{pmatrix} R_0(n,m) \\ M'(n,m) \end{pmatrix}, \end{aligned}$$

where $R_{n-m-1}(n,m)$ is the last row in $M(n,m)$, $R_0(n,m)$ is the starting row of $M(n,m)$, and $M'(n,m)$ contains all the columns of $M(n,m+1)$ with indices of the form with indices of the form $j = \frac{b(b+1)}{2} + k$, with $0 \le k \le b < m$, including the starting all one column $\mathbf{1}^T_{n-(m+1)}$ in $M(n,m+1)$. This nesting structure played an important role in the their derivation of the following results.

Theorem 4. *Suppose $Q_{n-1}(\mathbf{s}) < Q_n(\mathbf{s})$. Then, $Q_n(\mathbf{s})$ is the smallest integer m such that*

$$\text{rank}(M(n,m)) \neq \text{rank}(M(n-1,m)).$$

Theorems 3 and 4 laid the foundation for an algorithmic method to calculate the quadratic complexity of a sequence. More specifically, for each new term, Theorem 3 determines when the quadratic complexity will increase; and Theorem 4 characterizes how large the increment is. Combining different cases for each new term, Figure 2 provides a preliminary algorithm to recursively assess the quadratic complexity of a sequence $\mathbf{s}$. It can be seen from Figure 2 that the assessment heavily depends on the calculation of ranks of involved matrices, which becomes slower as n and m grow.

With a detailed analysis of the nesting structure of $M(n,m)$, the authors in [45] proposed a more efficient algorithm (as in [45], Figure 3) to calculate the quadratic complexity profile of $\mathbf{s}_n$. In addition, when $n < \frac{m(m+3)}{2} + 1$, the system $M(n,m)F(m) = E(n,m)$ is under-determined; and when $n \ge \frac{m(m+3)}{2} + 1$, the necessary and sufficient condition that a unique quadratic feedback function generating the sequence $\mathbf{s}_n$ can be given [45].

Theorem 5. *The quadratic feedback function of an m-stage FSR that generates the sequence $\mathbf{s}_n$ is unique if and only if*

$$\text{rank}(M(n,m)) = \frac{m(m+1)}{2} + 1.$$

Otherwise, there are $2^{1+\frac{m(m+1)}{2} - \text{rank}(M(n,m))}$ such functions.

This theorem illustrates that a binary sequence $\mathbf{s}$ with small quadratic complexity m should not be used for cryptographic applications. Otherwise, when $O(\frac{m(m+3)}{2})$ consecutive components s_i's in $\mathbf{s}$ are known, the quadratic feedback function could be (uniquely) determined, thereby producing the whole sequence and violating the requirement of unpredictability for sequences used in cryptography.

In the previous section, we saw a good theoretical understanding of the statistical behavior of linear complexity and linear complexity profile for random sequences. However, to the best of my knowledge, there is no published theoretical result on the statistical behavior for quadratic complexity and the quadratic complexity profile of random sequences

$\mathbf{s}_n$ and n-periodic sequences $\mathbf{s}_n^\infty$, except for the following two conjectures by Youssef and Gong [44] strongly indicated by numerical results.

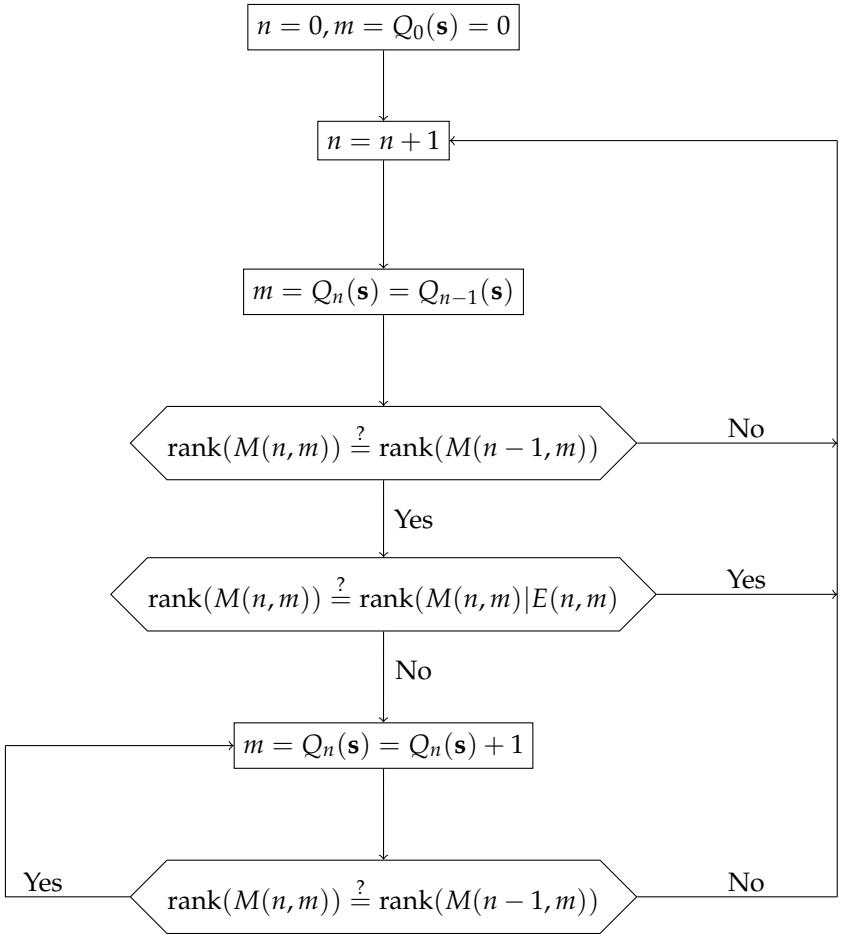


Figure 2. Recursive calculation of the quadratic complexity profile of $\mathbf{s}$.

Conjecture 1. *Let $N_n(q_c)$ be the number of binary sequences $\mathbf{s}_n$ with quadratic complexity $q_c = Q(\mathbf{s}) > n/2$. Then $N_n(q_c) = N_{n+i}(q_c + i)$ for any $i \geq 1$, indicating that $N_n(q_c)$ is a function of $n - q_c$.*

Conjecture 2. *For moderately large n, the expected value of the quadratic complexity of a random binary sequence of length n is given by*

$$E[Q_n] \approx \sqrt{2n}.$$

5. Maximum-Order Complexity

As an additional figure of merit for randomness testing, Jansen and Boekee in 1989 proposed the notion of maximum-order complexity, also later known as nonlinear complexity, of sequences [46]. We adopt the term maximum-order complexity in this survey for better distinction from the notion of quadratic complexity.

Definition 4. *The maximum-order complexity of a sequence $\mathbf{s} = s_0 s_1 s_2 \ldots$, denoted by $M(\mathbf{s})$, is the length of the shortest FSRs that can generate the sequence $\mathbf{s}$. Similarly, the maximum-order complexity profile of $\mathbf{s}$ is a sequence of $M_0(\mathbf{s}), M_1(\mathbf{s}), M_2(\mathbf{s}), \ldots$, where $M_n(\mathbf{s})$ for each $n \geq 0$ is the shortest FSRs that can generate the first n terms of $\mathbf{s}$.*

As pointed out in ref. [46], the significance of the maximum-order complexity is that it tells exactly how many terms of $\mathbf{s}$ have to be observed at least in order to be able to

generate the entire sequence by means of an FSR with length $M(\mathbf{s})$. Therefore, it has been considered as an important measure to judge the randomness of sequences. Below we recall some basic properties of maximum-order complexity of sequences [46,47].

Lemma 1. *Let* $\mathbf{s} = s_0 s_1 s_2 \cdots$ *be a sequence over* $\mathbb{F}_q$.
(i) Let l *be the length of the longest subsequence of* $\mathbf{s}$ *that occurs at least twice with different successors. Then the sequence* $\mathbf{s}$ *has maximum-order complexity* $M(\mathbf{s}) = l + 1$;
(ii) The maximum-order complexity of any n*-length sequence is at most* $n - 1$*. Moreover, the equality holds if and only if* $\mathbf{s}_n$ *has the form* $(\alpha, \cdots, \alpha, \beta)$ *with* $\alpha \neq \beta$ *in* $\mathbb{F}_q$.

Lemma 2. *Let* $\mathbf{s} = \mathbf{s}_n^\infty$ *be a sequence of period* n *over* $\mathbb{F}_q$.
(i) The maximum-order complexity of $\mathbf{s}$ *satisfies* $\lceil \log_q(n) \rceil \leq M(\mathbf{s}) \leq n - 1$;
(ii) The reciprocal sequence $\mathbf{s}^* = (s_{n-1}, \ldots, s_0)^\infty$ *has* $M(\mathbf{s}^*) = M(\mathbf{s})$.

From Lemmas 1 and 2, the difference between nonlinear complexities of finite-length sequences and periodic sequences is apparent. One typical difference is that the maximum-order complexity of a periodic sequence remains unchanged under cyclic shift operations, while that of a finite-length sequence can change dramatically. For instance, for the sequence $0 \ldots 01$ of length n, while $M(0 \ldots 01^\infty) = M(0 \ldots 01) = n - 1$, after a right cyclic shift, we have $M(10 \ldots 0^\infty) = n - 1$ but $M(10 \ldots 0) = 1$.

Recall that for the case where the linear feedback function is unique for periodic sequences, and that for the quadratic feedback function is also unique when the matrix $M(n, m)$ with $n = \frac{m(m+3)}{2} + 1$ is nonsingular, the situation for maximum-order complexity is significantly different.

Proposition 6. *Let* $\Phi_\mathbf{s}$ *denote the class of feedback functions of the FSRs that can generate a periodic sequence* $\mathbf{s} = s_n^\infty$ *with maximum-order complexity* m *over* $\mathbb{F}_q$*. Then the number of functions in* $\Phi_\mathbf{s}$ *is equal to* $|\Phi_\mathbf{s}| = q^{q^m - n}$.

By the above proposition, the class $\Phi_\mathbf{s}$ generally contains more than one feedback function that can generate the periodic sequence, which is similar for non-periodic sequences. One can search for functions exhibiting certain properties, such as the function with the least number of terms. One of the methods that minimize the number of terms and their orders in the inclusive-or sum of products of variables or their complements is to use the disjunctive normal form (DNF) representation of the feedback function. For the binary case, the DNF of f is given by

$$f(x_0, x_1, \ldots, x_{m-1}) = \sum_{b_0, \ldots, b_{m-1} \in \mathbb{F}_2} f(b_0, \ldots, b_{m-1}) x_0^{b_0} \cdots x_{m-1}^{b_{m-1}},$$

where $x_i^{b_i} = x_i + b_i + 1 = 1$ if and only if $x_i = b_i$ for $0 \leq i < m$. It is also interesting to note that a unique feedback function occurs for DeBruijn sequences of order m, which have m-tuples over $\mathbb{F}_q$ exactly once in one period q^m. For binary DeBruijn sequences of order m, Chan and Games in [23] showed their quadratic complexity are upper bounded by $2^m - 1 - \binom{m}{2}$, which can be reached by those DeBruijn sequences derived from m-sequences by inserting 0 to the all-zero subsequence of length $m - 1$ in m-sequences.

5.1. Computation and Statistical Behavior

In ref. [46] Jansen associated $M(\mathbf{s})$ with the maximum depth of a direct acyclic word graph (DAWG) from $\mathbf{s}$. Below, we recall a toy example from [46] to generate a DAWG from a binary string, which helps readers better understand relevant notions.

Example 1. *The sequence* $\mathbf{s} = 110100$ *has the following set of all subsequences*

$$\begin{aligned}\mathrm{SUB}(\mathbf{s}) = \{&\lambda, 1, 0, 11, 10, 01, 00, 110, 101, 010, 100,\\ &1101, 1010, 0100, 11010, 10100, 110100\},\end{aligned}$$

where λ *represents the empty sequence. For the sequence* $\mathbf{s}$*, the endpoints are given as in Table 1.*

Table 1. Endpoints for $\mathbf{s} = 110100$.

s		1	1	0	1	0	0
endpoints	0	1	2	3	4	5	6

For the sub-sequences of $\mathbf{s}$*, their endpoints are given as below*

$$\begin{aligned}
&E(\lambda) = \{0,1,2,3,4,5,6\}\\
&E(1) = \{1,2,4\}, \quad E(0) = \{3,5,6\},\\
&E(11) = \{2\}, \quad E(10) = \{3,5\},\\
&E(01) = \{4\}, \quad E(00) = \{6\},\\
&E(110) = \{3\}, \quad E(101) = \{4\}, \quad E(010) = \{5\}, \quad E(100) = \{6\},\\
&E(1101) = \{4\}, \quad E(1010) = \{5\}, \quad E(0100) = \{6\},\\
&E(11010) = \{5\}, \quad E(10100) = \{6\},\\
&E(110100) = \{6\}.
\end{aligned}$$

Subsequences are considered equivalent if they have the same set of endpoints. For instance, the following are three equivalent classes with endpoints 4*,* 5 *and* 6*, respectively:*

$$\begin{aligned}
&E_4 : 01, 101, 1101\\
&E_5 : 010, 1010, 11010,\\
&E_6 : 00, 100, 0100, 10100, 110100.
\end{aligned}$$

Therefore, the total 17 subsequences in $\mathrm{SUB}(\mathbf{s})$ *are divided into* 9 *equivalence classes. It is customary to choose the shortest subsequence as the class representative. In this way, we have the following representatives:*

$$EQ(\mathbf{s}) = \{\lambda, 1, 0, 11, 10, 01, 00, 110, 010\}.$$

From the above discussion, we can build a DAWG as follows: each representative in $EQ(\mathbf{s})$ *is denoted by a node and* λ *is considered as the source node; there is a directed edge between one node* v_i *to another node* v_j *if and only if the equivalence class of* v_i *contains a subsequence, which when extended with the edge symbol belongs to the equivalence class of* v_j*; the edges are divided into primary and secondary edges: an edge is primary if and only if it belongs to a primary path (the longest path from the source to the node) and a depth of a node is the length of the primary path to the node, where the length of a path is the number of edges along the path. Let* $BN(\mathbf{s})$ *be a subset consisting of nodes with more than one outgoing primary edges in* $EQ(\mathbf{s})$ *and define the maximum depth* $d(\mathbf{s}) = \max_{v \in BN(\mathbf{s})} d(v)$*.*

In this way, we obtain the DAWG of the sequence $\mathbf{s} = 110100$ *as in Figure 3, where primary edges are displayed in solid arrows. From the definition of the depth of a node, we have*

$$\begin{aligned}
&d(\lambda) = 0, d(0) = d(1) = 1,\\
&d(11) = d(10) = 2, d(01) = 4, d(00) = 6,\\
&d(110) = 3, d(010) = 5.
\end{aligned}$$

Since $BN(\mathbf{s}) = \{\lambda, 1, 0, 10\}$*, we have* $d(\mathbf{s}) = d(10) = 2$*.*

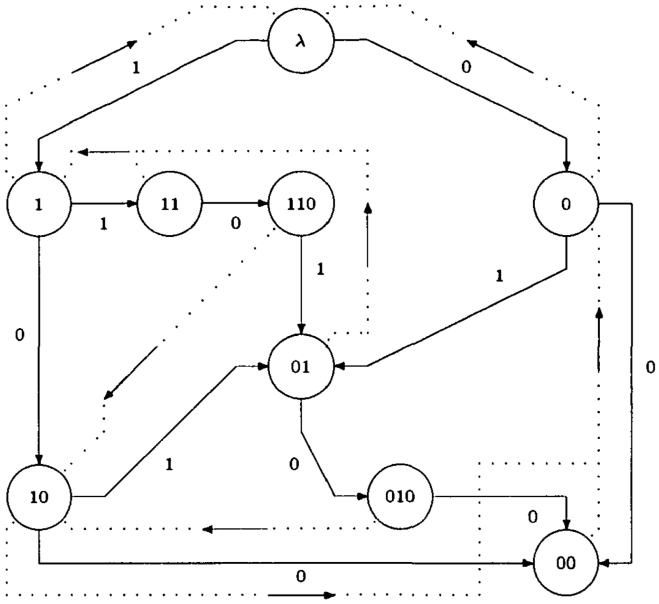


Figure 3. Direct acyclic word graph of $\mathbf{s} = 110100$ [21].

With the notions illustrated in Example 1, Jansen in [21] established the following connection between the maximum-order complexity of a sequence $\mathbf{s}$ and the depth of DAWG derived from $\mathbf{s}$, and proposed to calculate the maximum-order complexity of $\mathbf{s}$ from its DAWG.

Proposition 7. *Given a sequence* $\mathbf{s}$*, the maximum-order complexity of* $\mathbf{s}$ *and the depth* $d(\mathbf{s})$ *of its DAWG satisfy*

$$M(\mathbf{s}) = \begin{cases} 0, & \text{if } BN(\mathbf{s}) = \emptyset, \\ d(\mathbf{s}) + 1, & \text{otherwise.} \end{cases}$$

Blumer's algorithm [48] was identified as an important tool to build a DAWG of a sequence $\mathbf{s}$, thereby calculating its maximum-order complexity profile in linear time and memory with regard to the sequence length. The method worked particularly well for periodic sequences. With this algorithm, Jansen exhausted all 2^l binary l-length sequences and l-periodic sequences as l ranges from 1 up to 24 ([21], Tables 3.1–3.4), which exhibited typical statistical behaviors of maximum-order complexity profiles of random sequences: the expected maximum-order complexity of a random sequence $\mathbf{s}$ of sufficiently large length n over $\mathbb{F}_q$ is given by

$$E(M_n(\mathbf{s})) \approx 2\log_q(n).$$

Following the successful approach in ref. [45], Rizomiliotis and Kalouptsidis [49] considered the calculation of the maximum-order complexity of a sequence $\mathbf{s}$ in a similar way. From the recursive equations $f(s_i, s_{i+1}, \ldots, s_{i+m-1}) = s_{i+m}$ for $0 \leq i < n - m$, one can obtain a similar system to linear equations

$$M(n, m)F(m) = E(n, m),$$

where the coefficient matrix $M(n, m)$ from all binary terms $\prod_{i=0}^{m-1} x_i^{j_i}$, where $(j_0, j_1, \ldots, j_{m-1})$ is any element in $\mathbb{F}_2^m$ in a special ordering according to the degree of the terms. For instance, when $n = 10$ and $m = 3$, the system of linear equations is given by

$$M(10,3) = \begin{pmatrix} 1 & s_0 & s_1 & s_2 & s_0s_1 & s_0s_2 & s_1s_2 & s_0s_1\ s_2 \\ 1 & s_1 & s_2 & s_3 & s_1s_2 & s_1s_3 & s_2s_3 & s_1s_2\ s_3 \\ 1 & s_2 & s_3 & s_4 & s_2s_3 & s_2s_4 & s_3s_4 & s_2s_3\ s_4 \\ 1 & s_3 & s_4 & s_5 & s_3s_4 & s_3s_5 & s_4s_5 & s_3s_4\ s_5 \\ 1 & s_4 & s_5 & s_6 & s_4s_5 & s_4s_6 & s_5s_6 & s_4s_5\ s_6 \\ 1 & s_5 & s_6 & s_7 & s_5s_6 & s_5s_7 & s_6s_7 & s_5s_6\ s_7 \\ 1 & s_6 & s_7 & s_8 & s_6s_7 & s_6s_8 & s_7s_8 & s_6s_7s_8 \end{pmatrix}$$

with

$$F(3) = (a, a_0, a_1, a_3, a_0a_1, a_0a_2, a_1a_2, a_0a_1a_2)^T,$$
$$E(10,3) = (s_3, s_4, s_5, s_6, s_7, s_8, s_9)^T.$$

In ref. [49], the authors investigated the properties of the m-tuples $T_j(n,m) = (s_j, s_{j+1}, \ldots, s_{j+m-1})$ as j runs through the indices $0, 1, \ldots, n-m$ and proposed a term-by-term algorithm to compute the maximum-order complexity and output a feedback function for a given sequence $\mathbf{s}_n$. They also pointed out that the dominant multiplication complexity in their proposed algorithm is in the order of $O(n)$.

Attempting the above approach by analyzing the structure of $M(n,m)$ is not satisfactory enough. Later, in ref. [50], the authors made more observations that further improved the performance of calculating maximum-order complexity of sequences.

Proposition 8. *For a binary sequence* $\mathbf{s}$*, suppose* $M_{n-1}(\mathbf{s}) = m$ *and the minimal FSR of* $s_0s_1 \ldots s_{n-2}$ *does not generate* $s_0s_1 \ldots s_{n-2}s_{n-1}$*. Then* $M_n(\mathbf{s}) = m$ *if and only if the subsequence* $s_{n-m-1}s_{n-m} \ldots s_{n-2}$ *has not appeared within* $\mathbf{s}_{n-1}$*.*

The above observation is a natural extension of the property of maximum-order complexity in Lemma 1 (i). Instead, the following observation characterized the jump of maximum-order complexity profile at each term.

Proposition 9. *For a binary sequence* $\mathbf{s}$*, suppose* $M_{n-1}(\mathbf{s}) = m < M_n(\mathbf{s})$*. Let* $i \leq n - m - 1$ *be the least integer such that* $s_j \ldots s_{j+m-1} = s_i \ldots s_{i+m-1}$ *for certain* $0 \leq j < i$*. Then* $M_n(\mathbf{s}) = M_{n-1}(\mathbf{s}) + (n - m - i) = n - i$*. Moreover, if* $M_n(\mathbf{s}) = m + k$ *for* $k \geq 1$*, then the sequence* $\mathbf{s}_n||s_n \ldots s_{n+k-1}$ *by arbitrary extension* $s_n \ldots s_{n+k-1}$ *has the same maximum-order complexity* $m + k$*.*

In addition, they observed that the maximum-order complexity profile of a sequence has a close connection to its eigenvalue profile. To be more concrete, the eigenwords in a sequence $s_0s_1 \ldots s_{r-1}$ are those subsequences of $s_0s_1 \ldots s_{r-1}$ that are not contained in any proper subsequence $s_0 \ldots s_{t-1}$ of $s_0s_1 \ldots s_{r-1}$. They showed the following interesting connection.

Theorem 6. *For a binary sequence* $\mathbf{s}$*, suppose* $M_{n-1}(\mathbf{s}) = m$ *and the minimal FSR of* $\mathbf{s}_{n-1}$ *does not generate* $\mathbf{s}_n$*. Then,*

$$M_n(\mathbf{s}) = \max\{M_{n-1}(\mathbf{s}), n - \text{Eigenvalue}(\mathbf{s}_{n-1}),\}$$

where Eigenvalue$(\mathbf{s}_{n-1})$ *is the number of eigenwords in the sequence* $\mathbf{s}_{n-1}$*.*

From the observations on the update of certain feedback function when the minimal FSR of $\mathbf{s}_{n-1}$ does not generate $\mathbf{s}_n$, they proposed the following procedure to generate a minimal feedback function and maximum-order complexity of a sequence $\mathbf{s}$. Suppose $M_{n-1}(\mathbf{s}) = m$ and the minimal feedback function of $\mathbf{s}_{n-1}$ is $h_{n-1}(x_0, \ldots, x_{m-1})$. Then,

- if $s_{n-1} = h_{n-1}(s_{n-m-1}, \ldots, s_{n-2})$, then $\mathbf{s}_n$ and $\mathbf{s}_{n-1}$ share the minimal feedback function h_{n-1} and the same maximum-order complexity m

- if $s_{n-1} \neq h_{n-1}(s_{n-m-1}, \ldots, s_{n-2})$ and $M_n(\mathbf{s}) = M_{n-1}(\mathbf{s})$, then update h_{n-1} as

$$h_n(x_0, \ldots, x_{m-1}) = h_{n-1}(x_0, \ldots, x_{m-1}) + \prod_{i=n-m-1}^{n-2} (x_i + s_i + 1)$$

- if $s_{n-1} \neq h_{n-1}(s_{n-m-1}, \ldots, s_{n-2})$ and $M_n(\mathbf{s}) = m + k$, then

$$h_n(x_0, \ldots, x_{m+k-1}) = h_{n-1}(x_0, \ldots, x_{m+k-1}) + \prod_{i=n-m-k-1}^{n-2} (x_i + s_i + 1)$$

Consequently, they proposed an efficient algorithm for maximum-order complexity, which works very similarly to the Berlekamp–Massey algorithm. For ease of comparison with the Berlekamp–Massey algorithm, we recall it in Algorithm 1. The computational complexity of Algorithm 1 in the worse case is $O(n^3)$, while in the average case is $O(n^2 \log_2(n))$ since the the expected maximum-order complexity $E[M_n] \approx 2\log_2(n)$. While it has the similar complexity as the DAWG method [21], its recursive nature is an important advantage since it eliminates the need to know the entire sequence in advance. This resembles the well-known Berlekamp–Massey algorithm in the linear case: if a discrepancy occurs at certain step, then a well-determined corrective function is added to the current minimal feedback function of $\mathbf{s}_{n-1}$ to provide an updated minimal feedback of $\mathbf{s}_n$. In addition, the update of maximum-order complexity $M_n(\mathbf{s}) = \max\{M_{n-1}(\mathbf{s}), n - \text{Eigenvalue}(\mathbf{s}_{n-1})\}$ at each step is similar to $L_n(\mathbf{s}) = \max\{L_{n-1}(\mathbf{s}), n - L_{n-1}(\mathbf{s})\}$ for linear complexity of the sequence $\mathbf{s}$ in the recursive procedure.

Algorithm 1: Generation of minimal polynomial for $\mathbf{s}$.

Input: A binary sequence $\mathbf{s} = s_0 s_1 s_2 \ldots$
Output: $m = M_n(\mathbf{s})$
the minimal feedback polynomial h generating $\mathbf{s}_n$

```
// Initialization
k = 0, m = 0, h = s_0
for j ∈ [1, ..., n − 1] do
    // check discrepancy
    d = s_j − h(s_{j−m}, ..., s_{j−1})
    if d ≠ 0 then
        if m = 0 then
            k = j, m = j
        else
            if k ≤ 0 then
                t = Eigenvalue(s_n)
                if t < j + 1 − m then
                    k = j + 1 − t − m
                    m = j + 1 − t
                end
            end
            k = k − 1
        end
        f = ∏_{i=0}^{m−1} (x_i + s_i ⊕ 1)
        h = h + f
    else
        k = k − 1
    end
end
Return h
```

5.2. Approximate Statistical Behavior

Conducting a rigorous theoretical analysis of the behavior of maximum-order complexity profiles, as was done for linear complexity, appeared intractable. On the other hand, maximum-order complexity had a lot of similarities as linear complexity. In order to facilitate the randomness test with maximum-order complexity, Erdman and Murphy proposed a method to approximate the distribution of the maximum-order complexity [51]. Inspired by the property of maximum-order complexity in Lemmas 1 and 2, they investigated a function that approximates the function $P(m,n)$: the probability that the first n m-tuples are all different in a sequence.

Proposition 10. *Let $R(m,n-1)$ be the probability that the n-th m-tuple is unique given that the previous $(n-1)$ m-tuples were unique. Then,*

$$P(m,n) = \prod_{i=1}^{n} R(m,i) \approx \prod_{i=1}^{n}\left(1-\frac{i+1}{2^{m+1}}\right) = p(m,n).$$

Simulations on random sequences for m from 4 to 24 ([51], Table 2) indicated that the above approximation was accurate, particularly when $m \geq 17$.

Recall that a purely periodic sequence has maximum-order complexity m if the first n m-tuples are unique but at least one of the first n $(m-1)$-tuples is repeated. Thus, for calculating a periodic sequence $\mathbf{s} = \mathbf{s}_n^{\infty}$, it suffices to only look at the first $n+m-1$ bits to see if the n m-tuples are unique and the first $n+m-2$ bits to see if the n $(m-1)$-tuples are not unique. Denote by $Q(m,n)$ the probability that the first n m-tuples are unique while the n $(m-1)$-tuples are not. This probability function can be well approximated by $q(m,n)$, which is given by

$$q(m,n) = \begin{cases} p(m,n) - p(m-1,n), & \text{if } n \leq 2^{m-1}, \\ p(m,n) & \text{otherwise.} \end{cases}$$

Based on this approximation (of which the accuracy was shown in ([51], Table 3)), the expected maximum-order complexity was approximated as follows:

Theorem 7. *Let M_n be the maximum-order complexity of a random periodic binary sequence $\mathbf{s}_n^{\infty}$. Then,*

$$\begin{aligned} E[M_n] &= \sum_{m=\log_2(n)}^{n-1} mQ(m,n) = (n-1) - \sum_{m=\log_2(n)}^{n-2} P(m,n) \\ &\approx (n-1) - \sum_{m=\log_2(n)}^{n-2} p(m,n) \\ &= (n-1) - \sum_{m=\log_2(n)}^{n-2} \prod_{i=1}^{n}\left(1-\frac{i+1}{2^{m+1}}\right). \end{aligned}$$

Similarly, for random binary sequences of length n, the expected maximum-order complexity was approximated below.

Theorem 8. *Let $N(m,n)$ be the number of binary sequences of length n with maximum-order complexity m. Then, it can be approximated as*

$$N(m,n) \approx 2^n q(m,n-m+1),$$

and the expected maximum-order complexity $\widehat{M}_n$ *of random binary sequence of length* n *is given by*

$$\begin{aligned} E[\widehat{M}_n] &= \sum_{m=0}^{n-1} m \frac{N(m,n)}{\sum_{i=0}^{n-1} N(i,n)} \\ &\approx \sum_{m=0}^{n-1} m \frac{2^n q(m, n-m+1)}{\sum_{i=0}^{n-1} 2^n q(i, n-i+1)} \\ &\approx 2\log_2(n) \quad (\textit{for sufficiently large } n). \end{aligned}$$

5.3. Sequences with High Maximum-Order Complexity

Different complexity measures were proposed in the literature to evaluate the randomness of sequences. A sequence with low complexity, including the linear, quadratic and maximum-order complexity, allows for short FSRs to generate the whole period of the sequence. That is to say, all remaining unknown terms in this sequence can be efficiently uncovered when the feedback function and the initial state of the short FSRs are determined. It is clear that this kind of sequences should be avoided for any cryptographic applications. On the other hand, the relation between high complexity of a sequence and good randomness of a sequence is not yet well understood.

For the aforementioned complexity measures, the sequences of the form $(0, \ldots, 0, 1)$ have the largest complexity, but have clearly poor randomness. According to the exhaustive search results on maximum-order complexity in Tables 1–3 in [21], while sequences $(0, \ldots, 0, 1)$ are the only instances of n-length sequences that have the highest possible complexity $n-1$, there are multiple n-periodic binary sequences having the highest maximum-order complexity $n-1$.

Researchers have been interested in the study of sequences with high maximum-order complexity [52,53], particularly on sequences having highest possible maximum-order complexity [54,55]. For the latter case, Rizomiliotis [49] in 2005 proposed the following necessary and sufficient condition for an n-periodic binary sequence to have maximum-order complexity the same as its linear complexity.

Proposition 11. *For a periodic binary sequence* $\mathbf{s} = \mathbf{s}_n^\infty$*, let* $\mathbf{s}_n(x) = s_0 + s_1 x + \cdots + s_{n-1}x^{n-1}$ *be the polynomial representation of* $\mathbf{s}_n$*, and let* $h(x) = \gcd(x^n + 1, \mathbf{s}_n(x))$*,* $c(x) = \mathbf{s}_n(x)/h(x)$ *and* $f(x) = x^n + 1/\gcd(x^n + 1, \mathbf{s}_n(x))$*. Then,*

$$M(\mathbf{s}) = \min\{n-1, L(\mathbf{s})\} \tag{10}$$

if and only if there exists integers $0 \le p_1 < p_2 \le n-1$ *satisfying*

$$f(x) | 1 + (x^{p_1} + x^{p_2}(x))c(x)$$

when $LC(\mathbf{s}) \le n-1$*.*

Based on the above condition, different families of n-periodic binary sequences that have the same linear complexity and maximum-order complexity were proposed in ref. [49]. Moreover, an algorithm based on Lagrange interpolation and an algorithm based on the relative shift of the component sequences were proposed to generate binary sequences with highest possible maximum-order complexity. These two ideas were further developed for constructing such periodic sequences with period of the form $2^m - 1$ [54]. Roughly 10 years later, n-periodic sequences with highest maximum-order complexity were revisited in ref. [55], which provided a complete picture of sequences with maximum-order complexity. The authors of ref. [55] gave the necessary and sufficient condition for n-periodic sequences over any alphabet to have maximum-order complexity $n-1$.

Theorem 9. *Let* $\mathbf{s} = \mathbf{s}_n^\infty$ *be an* n*-periodic sequence over any field* $\mathbb{F}$*. Then* $M(\mathbf{s}) = n-1$ *if and only if there exists an integer* p *with* $1 \le p < n$ *such that* $s_i = s_{i+p \pmod n}$ *for* $0 \le i < n-2$ *and*

$s_i \neq s_{i+p \pmod n}$ *for* $i = n-2, n-1$. *Furthermore, such a sequence can, up to shift equivalence, be represented as one of the following forms*

(1) $\mathbf{s}_n = (\alpha^{n-1}\beta) = (\overbrace{\alpha, \ldots, \alpha}^{n-1}, \beta)$ *where* $\alpha \neq \beta \in \mathbb{F}$;
(2) $\mathbf{s}_n = \mathbf{s}_{r_0} = (\mathbf{s}_{r_1})^{m_1}\mathbf{s}_{r_2}$ *for certain integer* $r_1 \in \mathbb{Z}_n^*$ *with*

$$\mathbf{s}_{r_{i-1}} = (\mathbf{s}_{r_i})^{m_i}\mathbf{s}_{r_{i+1}}, i = 1, 2, \cdots, k,$$

and

$$\mathbf{s}_{r_k} = \left((\alpha)^{r_k - 1}\beta\right), \mathbf{s}_{r_{k+1}} = (\alpha),$$

where the integers m_i, r_{i+1} *for* $i = 1, 2, \cdots, k$ *are derived from* $r_{i+1} = r_{i-1} - m_i r_i$ *with* $r_1 > r_2 > \cdots > r_{k+1} = 1$,

With the full characterization in Theorem 9, the distribution of random n-periodic binary sequences having maximum-order complexity $n-1$ can be derived as follows.

Proposition 12. *The probability that a randomly generated sequence* $\mathbf{s} = \mathbf{s}_n^\infty$ *has the highest* $M(\mathbf{s}) = n-1$ *is given by*

$$P_{n-1} = \text{Prob}(M(\mathbf{s}) = n-1) = \frac{q(q-1)n\varphi(n)}{2\sum_{d|n}\mu(d)q^{n/d}},$$

where $\varphi(\cdot)$ *is the Euler's totient function and* $\mu(\cdot)$ *is the Möbius function given by* $\mu(n) = \sum_{k \in \mathbb{Z}_n^*} e^{2\pi i k/n}$. *In particular, when* $q = 2$, *the probability*

$$P_{n-1} = \frac{\varphi(n)}{\sum_{d|n}\mu(d)2^{\frac{n}{d}}}.$$

Interestingly, the sequences $\mathbf{s}$ characterized in Theorem 9 exhibit a strong recursive structure, which can be derived by applying the Euclidean algorithm on n and the integer p, a smaller period of a subsequence of $\mathbf{s}_n$. This strong recursive structure, on the other hand, implies that $\mathbf{s}$ is far from being random. As discussed in ref. [55], the balancedness, stability and scalability of such sequences are not good, indicating that they should be avoided for cryptographic applications. Very recently, binary sequences with periodic n were further studied in ref. [56], where the authors further investigated the structure of $\mathbf{s}_n^\infty$ and proposed an algorithmic method to determine all n-periodic binary sequences with maximum-order complexity $\geq n/2$.

In addition, Liang et al. recently investigated the structure of n-length binary sequences with high maximum-order complexity $\geq n/2$ and proposed an algorithm that can completely generate all those binary sequences [53]. Based on the completeness, they managed to provide an explicit expression of the number of n-length binary sequences with any maximum-order complexity between $n/2$ and n.

Proposition 13. *The number of n-length binary sequences with maximum-order complexity m with* $n/2 \leq m \leq n-1$ *is given by*

$$N(m, n) = \sum_{d=1}^{n-m}(n-m-d+2)2^{n-m-d-1}N(m, m+d, d)$$

where

$$N(m, m+d, d) = 2^{n-m} - \sum_{t=1}^{r}(-1)^{t-1}\sum_{1 \leq j_1 < \cdots < j_t} 2^{\frac{n-m}{p_1^{j_1}\cdots p_t^{j_t}}}$$

when $n-m$ *is factorized as* $n-m = p_1^{j_1}\cdots p_r^{j_r}$.

6. Relations between Complexity Measures

Besides the complexity measures in the context of FSRs in previous sections, some other measures have been discussed by researchers, such as Lempel–Ziv complexity [57], 2-adic complexity for FSRs with carry operation [22], k-error complexity, expansion complexity [58] and correlation measure [59,60]. This section will briefly review the relations among these measures. Below we start with the relation between the complexity measures in the previous sections, and we will be mainly focusing on extreme cases.

6.1. Relations between FSR-Based Complexities

From their definitions, for a sequence $\mathbf{s}$, it is easily seen that

$$M_n(\mathbf{s}) \leq Q_n(\mathbf{s}) \leq L_n(\mathbf{s}) \leq n.$$

We know that the special sequence of the form $\mathbf{s}_n = \alpha \dots \alpha\beta$, where $\alpha \neq \beta$, has the largest linear complexity n, quadratic and maximum-order complexity $n-1$. As for n-periodic sequences $\mathbf{s} = \mathbf{s}_n^\infty$ over $\mathbb{F}_q$, the largest linear complexity is also n, which can be seen from the expression $L(\mathbf{s}) = n - \deg(\gcd(x^n - 1, s_n(x)))$, as recalled in Proposition 5. This can be used to characterize the periodic sequences with largest linear complexity. Assume $n = p^v w$ with $\gcd(p, w) = 1$, where p is the characteristic of $\mathbb{F}_q$. Let $C_1 = \{0\}, C_2, \dots, C_h$ be the cyclotomic cosets modulo w relative to the power q. Let β be a primitive n-th root of unity in an extension of $\mathbb{F}_q$. Then, we have the following factorization of $x^n - 1$ over $\mathbb{F}_q$:

$$x^n - 1 = \left(\prod_{1 \leq i \leq h} f_j(x) \right) = \left(\prod_{1 \leq i \leq h} \prod_{j \in C_i} (x - \beta^i) \right)^{p^u},$$

which indicates that an n-periodic sequences over $\mathbb{F}_q$ has the largest linear complexity n if and only if

$$\gcd(s_n(x), f_j) = 1 \text{ for } j = 1, 2, \dots, h.$$

Niederreiter, in ref. [61], proved the existence of n-periodic sequences with the largest linear complexity n and large k-error linear complexity, which is defined as

$$L_k(\mathbf{s}_n^\infty) = \min\{L(\mathbf{t}_n^\infty) \mid \mathrm{wt}(\mathbf{s}_n - \mathbf{t}_n) \leq k\}.$$

This result was obtained by finding integers $n = p^v w$ satisfying a condition, which was derived from the following observation [61]: let $l = \min_{2 \leq i \leq h} |C_i|$ and suppose k is an integer such that $\sum_{j=0}^{k} \binom{n}{j} (q-1)^j < q^l$, then there exists an n-periodic sequence $\mathbf{s}$ satisfies

$$L(\mathbf{s}) = n \text{ and } L_k(\mathbf{s}) \geq n - p^v.$$

As shown in the examples in ref. [61], there are infinitely many primes n allowing for the existence of binary n-periodic sequences of maximum possible linear complexity n.

Finding n-periodic sequences with the largest maximum-order complexity seems more challenging, especially when k-error complexity is also considered. As recalled in Section 5, Sun et al. [55] revealed n-periodic sequences with the largest maximum-order complexity $n-1$, which have surprisingly strong recursive structure. Besides this, little results about k-error maximum-order complexity and quadratic complexity have been reported. Here, we propose two open problems for interested readers.

Problem 2. *Characterize all the n-periodic sequences over $\mathbb{F}_q$ with the largest quadratic complexity $n-1$.*

Problem 3. *Characterize certain lower bound on the k-error maximum-order complexity of n-periodic sequences with maximum-order complexity $n-1$.*

Another extreme case for sequences is that they have the least possible complexity values. For finite-length sequences, we know that the linear, quadratic and maximum-order complexity can be as small as zero, which do not appear to be interesting to explore. When $\mathbf{s}$ is a periodic sequence over $\mathbb{F}_q$, we know that m-sequences of length $n = q^m - 1$ have the least complexities $M(\mathbf{s}) = Q(\mathbf{s}) = L(\mathbf{s}) = m$ and DeBruijn sequences of length $n = q^m$ have the least maximum-order complexity $M(\mathbf{s}) = m$. In addition to the aforementioned observations, readers may wonder what else we know about their relations so far?

Due to the limited understanding of DeBruijn sequences, there have been only partial results for the above problem. It was well known [62] that the linear complexity of binary DeBruijn sequences of order m is between $2^{m-1} + m$ and $2^m - 1$, where both the lower and upper bound are attainable. However, there exists no binary DeBruijn sequences with linear complexity $2^{m-1} + m + 1$ [63]. Based on the evidence of existing results, Etzion, in his survey, made the following conjecture:

Conjecture 3. *For any integer $2^{m-1} + m + 2 \leq d \leq 2^m - 1$, there exists a binary DeBruijn sequence with minimal polynomial $(x + 1)^d$.*

There is a close connection between m-sequence over $\mathbb{F}_q$ of order m and DeBruijn sequences of order m. There have been also some works on the linear complexity of modified DeBruijn sequences, which were obtained by removing one zero from the longest run of zeros in a DeBruijn sequence. Mayhew and Golomb [64] showed that the minimal polynomial of binary modified DeBruijn sequences of order m is a product of distinct irreducible polynomials with degrees not equal to 1 and dividing m. Further results on the restrictions of the degrees were developed in [65–67].

When considering the fact for a binary sequence $\mathbf{s}$ with low quadratic complexity m, any subsequence of length slightly larger than $\frac{(m+3)m}{2} + 1$ can be sufficient to recover the whole sequence $\mathbf{s}$. However, the topic of quadratic complexity was significantly less explored. There has been little research progress on quadratic complexity, even though it was already studied in 1989. To the best of my knowledge, the only results about the quadratic complexity of DeBruijn sequences were reported by Chan and Games [23].

Proposition 14. *Given a binary DeBruijn sequence $\mathbf{s}$ of order $m \geq 3$, then*

$$Q(\mathbf{s}) \leq 2^m - \binom{m}{2} - 1.$$

In particular, this upper bound can be attained by the DeBruijn sequence derived from an m-sequence of order by including the all zero m-tuple.

In addition, numerical results for $m = 3, 4, 5, 6$ led to a conjecture in [23]: the quadratic complexity of binary DeBruijn sequences of order $m \geq 3$ is lower bounded by $m + 2$. Three years later, this conjecture was later confirmed by Khachatrian in [68].

6.2. FSR-Based Complexities, Lempel–Ziv Complexity, Expansion Complexity and Autocorrelation

This subsection will briefly review recent works on the relations among the linear, maximum-order complexity and other significant metrics for assessing pseudo-random sequences.

6.2.1. 2-Adic Complexity

The 2-adic complexity introduced by Goresky and Klapper [22,69] is closely related to the length of a shortest feedback with carry shift register (FCSR) that generates the sequence. The theory of 2-adic complexity of periodic sequences has been well developed.

Given a sequence $\mathbf{s} = \mathbf{s}_n^\infty$ of period n, one has

$$\sum_{i=0}^{\infty} s_i 2^i = -\frac{\sum_{i=0}^{n-1} s_i 2^i}{2^n - 1} = \frac{A}{q},$$

where $0 \le A \le q, \gcd(A, q) = 1$ and

$$q = \frac{2^n - 1}{\gcd\left(2^n - 1, \sum_{i=0}^{n-1} s_i 2^i\right)}.$$

The 2-adic complexity of $\mathbf{s}$ is defined as $\Phi(\mathbf{s}) = \log_2(q)$. In the aperiodic case, the n-th 2-adic complexity, denoted by $\Phi_{2,n}(\mathbf{s})$, is the binary logarithm of

$$\min\left\{\max\{|f|, |q|\} : f, q \in \mathbb{Z}, q \text{ odd }, q \sum_{i=0}^{n-1} s_i 2^i \equiv f \pmod{2^n}\right\}.$$

There was very little work on exploring the relation between linear, quadratic and maximum-order complexities of a binary sequence and its 2-adic complexity. Until recently, it was shown [70] that the maximum-order complexity of a binary sequence $\mathbf{s}$ is upper bounded by its 2-adic complexity.

6.2.2. Lempel–Ziv Complexity

The Lempel–Ziv complexity of a sequence is one classical complexity measure based on pattern counting, introduced by Lempel and Ziv in 1976 [57] and later named after them. The parsing procedure in the calculation of Lempel–Ziv complexity laid the basis for the prominent Lempel–Ziv compression algorithms [71,72]. The Lempel–Ziv complexity reflects the compressibility of a sequence and thus has significant cryptographic interest.

The Lempel–Ziv complexity measures the rate at which new patterns emerge as we move along a given sequence. For a binary sequence $\mathbf{s}_n = s_0 s_1 \dots s_{n-1}$, we split up $\mathbf{s}_n$ into adjacent blocks. By definition, the first block consists of s_0. If $s_0 \dots s_{m-1}$ is a union of blocks, or in other words if s_{m-1} is the last bit in a block, then the next block $s_m \dots s_{m+k-1}$ is uniquely determined by two properties: (i) the sequence $s_m \dots s_{m+k-2}$ occurs as a subsequence in $s_1 \dots s_{m+k-3}$; (ii) the sequence $s_m \dots s_{m+k-1}$ does not occur as a subsequence in $s_1 \dots s_{m+k-2}$. The Lempel–Ziv complexity is then the number of blocks into which $\mathbf{s}_n$ is split up by this procedure.

Motivated by Niederreiter's statement [31], Limniotis, Kolokotronis and Kalouptsidis explored the relation between the maximum-order complexity and Lempel–Ziv complexity of a sequence, which are both closely connected to the eigenvalue profile of the sequence. They pointed out that although the exhaustive history (which uniquely determines the Lempel–Ziv complexity) cannot fully estimate the maximum-order complexity profile and vise versa, there still exists a relationship between them in the sense that the eigenvalue profile of a sequence of given Lempel–Ziv complexity, which determines the maximum-order complexity profile, is restricted rather than arbitrary. They also established a lower bound on the compression ratio of sequences with a prescribed maximum-order complexity.

6.2.3. Expansion Complexity

In 2012, Diem introduced the notion of expansion complexity of sequences [24], in which he showed that the sequences over finite fields with optimal linear complexity proposed by Xing and Lam via function expansion [29] can be efficiently computed from a relatively short subsequence.

Let $\mathbf{s} = s_0 s_1 s_2 \dots$ be a sequence over $\mathbb{F}_q$. For a positive integer n, the n th expansion complexity $E_n(\mathbf{s})$ is 0 if $s_0 = \dots = s_{n-1} = 0$ and otherwise the least total degree of a nonzero polynomial $h(x, y) = \sum_{i,j} h_{ij} x^i y^j \in \mathbb{F}_q[x, y]$ with

$$h(x, G(x)) \equiv 0 (\mathrm{mod}\, x^n),$$

where the total degree of $h(x,y)$ is given by

$$\deg(h(x,y)) = \max_{i,j\geq 0}\left\{ i+j \mid x^i y^j \text{ is a monomial of } h(x,y), h_{ij} \neq 0\right\}.$$

The expansion complexity of the sequence $\mathbf{s}$ is defined as

$$E(\mathbf{s}) = \sup_{n\geq 1} E_n(\mathbf{s}).$$

It can be verified that $E_n(\mathbf{s}) \leq E_{n+1}(\mathbf{s}) \leq E_n(\mathbf{s}) + 1$. In particular, if $h(x,y)$ is restricted to be irreducible, then the n-th irreducible expansion complexity and irreducible expansion complexity of $\mathbf{s}$ can be defined accordingly. The relations between the linear complexity, maximum-order complexity and expansion complexity were discussed recently in [58,73].

Proposition 15. *Let* $\mathbf{s}$ *be an ultimately periodic sequence with pre-period* u*, period* n *and linear complexity* $L(\mathbf{s})$*. Then, the irreducible expansion complexity*

$$E^*(\mathbf{s}) = \begin{cases} L(\mathbf{s})+1 & \text{if } u = 0 \\ L(\mathbf{s}) & \text{if } u = 1 \\ L(\mathbf{s})-1 & \text{if } u = 2 \text{ or } u > 2, s_{u-1} \neq s_{u+n-1} \end{cases}$$

and $E^*(\mathbf{s}) < L(\mathbf{s}) - 1$ *otherwise.*

Proposition 16. *For any infinite sequence* $\mathbf{s}$ *over* $\mathbb{F}_q$*, if* $M_n(\mathbf{s}) = n - k$ *with* $1 \leq k < \sqrt{2n} - 2$ *and* $n > 8$*, then its* n *th expansion complexity* $E_n(\mathbf{s})$ *is no more than* $k + 2$*.*

6.2.4. k-Order Correlation Measure

Let k be a positive integer. The n-th correlation measure of order k of a binary sequence $\mathbf{s} = s_0 s_1 s_2 \ldots$ is given by [74]

$$C_n(\mathbf{s}, k) = \max_{u,\mathbf{d}} \left| \sum_{i=0}^{u-1} (-1)^{s_{i+d_1} + s_{i+d_2} + \ldots + s_{i+d_k}} \right|,$$

where the maximum is taken over all $\mathbf{d} = (d_1, d_2, \ldots, d_k)$ with non-negative integers $0 \leq d_1 < d_2 < \cdots < d_k$ and u such that $u + d_k < n$.

The relations between the linear complexity, maximum-order complexity and correlation measure of order k were explored in [59,60,74].

Proposition 17. *Given a binary sequence* $\mathbf{s} = s_0 s_1 s_2 \ldots$*, the* n*-th linear complexity and* n*-th correlation measure of* $\mathbf{s}$ *satisfy*

$$L_n(\mathbf{s}) \geq n - \max_{1\leq k\leq L_n(\mathbf{s})+1} C_n(\mathbf{s}, k), \quad n \geq 1,$$

and the n*-th maximum-order complexity and* n*-th correlation measure of* $\mathbf{s}$ *satisfy*

$$M_n(\mathbf{s}) \geq N - 2^{M_n(\mathbf{s})+1} \max_{1\leq k\leq M_n(\mathbf{s})+1} C_n(\mathbf{s}, k), \quad n \geq 1.$$

The above bounds were recently improved in [60].

Proposition 18. *Given a binary sequence* $\mathbf{s} = s_0 s_1 s_2 \ldots$*, if for some integer* t*, the* n*-th linear complexity of* $\mathbf{s}$ *satisfies*

$$2^{L_n(\mathbf{s})} < \binom{\lfloor \frac{n}{2} \rfloor}{t},$$

then for some k with $1 < k \leq 2t$*, one has*

$$C_n(\mathbf{s}, k) \geq \lceil \frac{n}{2} \rceil.$$

In addition, if the n-th maximum-order complexity of $\mathbf{s}$ *satisfies* $M_n(\mathbf{s}) \leq M$*, then one has*

$$C_n(\mathbf{s}, 2) \geq n - 2^M + 1.$$

7. Conclusions

This survey primarily focused on the complexity measures of sequences within the domain of feedback shift registers, which are efficiently computable and hold significant interest in cryptographic applications. Notable works on the computation, stochastic behavior and theoretical results of these complexities were examined to a certain technical extent. Several conjectures were revisited and new open problems were proposed. To offer a sounding overview of the research development on complexity measures, this survey also briefly reviewed the established relation between these complexities as well as their connection with other important metrics for pseudo-random sequences, including the well-known 2-adic complexity, Lempel–Ziv complexity, expansion complexity and k-order correlation measure. This survey indicates that although the study of complexity measures of randomness traces back to the 1960s, only linear complexity and maximum-order complexity are relatively well explored. Other complexity measures and their interrelations appear more intractable and require new tools, techniques and theoretical studies in future research.

Funding: This research was funded by the Research Council of Norway under Grant No. 311646.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Ristenpart, T.; Yilek, S. When good randomness goes bad: Virtual machine reset vulnerabilities and hedging deployed cryptography. In Proceedings of the Network and Distributed System Security (NDSS) Symposium, San Diego, CA, USA, 28 February–3 March 2010.
2. Woodage, J.; Shumow, D. An analysis of NIST SP 800-90A. In *Advances in Cryptology—EUROCRYPT 2019*; Ishai, Y., Rijmen, V., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 151–180. [CrossRef]
3. Bernstein, D.J.; Lange, T.; Niederhagen, R. Dual EC: A standardized back door. *Cryptol. ePrint Arch.* **2015**, 767. Available online: https://eprint.iacr.org/2015/767 (accessed on 7 June 2024).
4. Davis, H.; Green, M.D.; Heninger, N.; Ryan, K.; Suhl, A. On the possibility of a backdoor in the Micali-Schnorr generator. In *Public-Key Cryptography—PKC 2024*; Tang, Q., Teague, V., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 352–386. [CrossRef]
5. Barker, E.; Kelsey, J. *Recommendation for Random Number Generation Using Deterministic Random Bit Generators*; Special Publication; NIST (National Institute of Standards and Technology): Washington, DC, USA, 2015. [CrossRef]
6. Mukhopadhyay, S.; Sarkar, P. Application of LFSRsfor parallel sequence generation in cryptologic algorithms. In *Computational Science and Its Applications—ICCSA 2006*; Gavrilova, M., Gervasi, O., Kumar, V., Tan, C.J.K., Taniar, D., Laganá, A., Mun, Y., Choo, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 436–445. [CrossRef]
7. Kuznetsov, A.A.; Potii, O.V.; Poluyanenko, N.A.; Gorbenko, Y.I.; Kryvinska, N. Areas of Application for Nonlinear Shift Registers in PRS Generators. In *Stream Ciphers in Modern Real-Time IT Systems: Analysis, Design and Comparative Studies*; Springer International Publishing: Cham, Switzerland, 2022; pp. 295–318. [CrossRef]
8. Mises, R.V. Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Z.* **1919**, *5*, 52–99. [CrossRef]
9. Church, A. On the concept of a random sequence. *Bull. Am. Math. Soc.* **1940**, *46*, 130–135. [CrossRef]
10. Golomb, S.W. *Shift Register Sequences: Secure and Limited-Access Code Generators, Efficiency Code Generators, Prescribed Property Generators, Mathematical Models*, 3rd Revised ed.; World Scientific Publishing Company: Singapore, 2017.
11. Kolmogorov, A.N. On tables of random numbers. *Theor. Comput. Sci.* **1998**, *207*, 387–395. [CrossRef]
12. Martin-Löf, P. The definition of random sequences. *Inf. Control* **1966**, *9*, 602–619. [CrossRef]
13. Schnorr, C. Process complexity and effective random tests. *J. Comput. Syst. Sci.* **1973**, *7*, 376–388. [CrossRef]
14. Knuth, D.E. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, 3rd ed.; Addison-Wesley: Boston, MA, USA, 1997.
15. Berlekamp, E.R. *Algebraic Coding Theory*; World Scientific Publishing Company: Singapore, 1968. [CrossRef]
16. Massey, J. Shift-register synthesis and BCH decoding. *IEEE Trans. Inf. Theory* **1969**, *15*, 122–127. [CrossRef]
17. Rueppel, R.A. Linear complexity and random sequences. In *Advances in Cryptology—EUROCRYPT'85*; Pichler, F., Ed.; Springer: Berlin/Heidelberg, Germany, 1986; pp. 167–188. [CrossRef]

18. Niederreiter, H. The probabilistic theory of linear complexity. In *Advances in Cryptology— EUROCRYPT'88*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 1988; pp. 191–209. [CrossRef]
19. Niederreiter, H. Linear complexity and related complexity measures for sequences. In *Progress in Cryptology—INDOCRYPT 2003*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2003; pp. 1–17. [CrossRef]
20. Bassham, L.E.; Rukhin, A.L.; Soto, J.; Nechvatal, J.R.; Smid, M.E.; Leigh, S.D.; Levenson, M.; Vangel, M.; Heckert, N.A.; Banks, D.L. *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*; NIST: Washington, DC, USA, 2010.
21. Jansen, C.J.A. Investigations on Nonlinear Streamcipher Systems: Construction and Evaluation Methods. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 1989.
22. Klapper, A.; Goresky, M. Feedback shift registers, 2-adic span, and combiners with memory. *J. Cryptol.* **1997**, *10*, 111–147. [CrossRef]
23. Chan, A.H.; Games, R.A. On the quadratic spans of DeBruijn sequences. *IEEE Trans. Inf. Theory* **1990**, *36*, 822–829. [CrossRef]
24. Diem, C. On the use of expansion series for stream ciphers. *LMS J. Comput. Math.* **2012**, *15*, 326–340. [CrossRef]
25. Meidl, W.; Niederreiter, H. Linear complexity, *k*-error linear complexity, and the discrete Fourier transform. *J. Complex.* **2002**, *18*, 87–103. [CrossRef]
26. Topuzoğlus, A.; Winterhof, A. Pseudorandom Sequences. In *Topics in Geometry, Coding Theory and Cryptography*; Garcia, A., Stichtenoth, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 135–166. [CrossRef]
27. Gustavson, F.G. Analysis of the Berlekamp-Massey linear feedback shift-register synthesis algorithm. *IBM J. Res. Dev.* **1976**, *20*, 204–212. [CrossRef]
28. Dornstetter, J. On the equivalence between Berlekamp's and Euclid's algorithms. *IEEE Trans. Inf. Theory* **1987**, *33*, 428–431. [CrossRef]
29. Xing, C.; Lam, K.Y. Sequences with almost perfect linear complexity profiles and curves over finite fields. *IEEE Trans. Inf. Theory* **1999**, *45*, 1267–1270. [CrossRef]
30. Niederreiter, H.; Vielhaber, M. An algorithm for shifted continued fraction expansions in parallel linear time. *Theor. Comput. Sci.* **1999**, *226*, 93–104. [CrossRef]
31. Niederreiter, H. Some computable complexity measures for binary sequences. In *SEquences and Their Applications (SETA)*; Springer: London, UK, 1999; pp. 67–78. [CrossRef]
32. Jungnickel, D. *Finite Fields: Structure and Arithmetics*; B.I. Wissenschaftsverlag: Berlin, Germany, 1993.
33. Massey, J.L.; Serconek, S. Linear complexity of periodic sequences: A general theory. In *Advances in Cryptology—CRYPTO'96*; Koblitz, N., Ed.; Springer: Berlin/Heidelberg, Germany, 1996; pp. 358–371.
34. Massey, J.L.; Serconek, S. A Fourier transform approach to the linear complexity of nonlinearly filtered sequences. In *Advances in Cryptology—CRYPTO'94*; Springer: Berlin/Heidelberg, Germany, 1994; pp. 332–340. [CrossRef]
35. Blahut, R.E. *Theory and Practice of Error Control Codes*; Addison-Wesley Publishing Company: Boston, MA, USA, 1983.
36. Selmer, E.S. *Linear Recurrence Relations over Finite Fields*; Department of Informatics, University of Bergen: Bergen, Norway, 1966.
37. Ding, C.; Hesseseth, T.; Shan, W. On the linear complexity of Legendre sequences. *IEEE Trans. Inf. Theory* **1998**, *44*, 1276–1278. [CrossRef]
38. Meidl, W.; Winterhof, A. Lower bounds on the linear complexity of the discrete logarithm in finite fields. *IEEE Trans. Inf. Theory* **2001**, *47*, 2807–2811. [CrossRef]
39. Helleseth, T.; Kim, S.H.; No, J.S. Linear complexity over $GF(p)$ and trace representation of Lempel-Cohn-Eastman sequences. In Proceedings of the IEEE International Symposium on Information Theory, Lausanne, Switzerland, 30 June–5 July 2002; IEEE: Piscataway, NJ, USA, 2003. [CrossRef]
40. Shparlinski, I. *Cryptographic Applications of Analytic Number Theory*; Birkhäuser: Basel, Switzerland, 2003.
41. Rueppel, R.A. *Analysis and Design of Stream Ciphers*; Springer: Berlin/Heidelberg, Germany, 1986. [CrossRef]
42. Meidl, W.; Niederreiter, H. On the expected value of the linear complexity and the *k*-error linear complexity of periodic sequences. *IEEE Trans. Inf. Theory* **2002**, *48*, 2817–2825. [CrossRef]
43. Chan, A.H.; Games, R.A. On the quadratic spans of periodic sequences. In *Advances in Cryptology — CRYPTO'89*; Lecture Notes in Computer Science, Proceedings; Springer: Berlin/Heidelberg, Germany, 1990; pp. 82–89. [CrossRef]
44. Youssef, A.; Gong, G. On the quadratic span of binary sequences. In Proceedings of the 20th Biennial Symposium on Communications, Kingston, ON, Canada, 28–31 May 2000; pp. 159–163.
45. Rizomiliotis, P.; Kolokotronis, N.; Kalouptsidis, N. On the quadratic span of binary sequences. *IEEE Trans. Inf. Theory* **2005**, *51*, 1840–1848. [CrossRef]
46. Jansen, C.; Boekee, D. The shortest feedback shift register that can generate a given sequence. In *Advances in Cryptology—CRYPTO'89*; Lecture Notes in Computer Science, Proceedings; Springer: Berlin/Heidelberg, Germany, 1990; pp. 90–99. [CrossRef]
47. Jansen, C. The maximum order complexity of sequence ensembles. In *Advances in Cryptology—EUROCRYPT '91*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 1991; pp. 153–159. [CrossRef]
48. Blumer, A.; Blumer, J.; Ehrenfeucht, A.; Haussler, D.; McConnell, R.M. Linear size finite automata for the set of all subwords of a word—An outline of results. *Bull. EATCS* **1983**, *21*, 12–20.
49. Rizomiliotis, P.; Kalouptsidis, N. Results on the nonlinear span of binary sequences. *IEEE Trans. Inf. Theory* **2005**, *51*, 1555–1563. [CrossRef]

50. Limniotis, K.; Kolokotronis, N.; Kalouptsidis, N. On the nonlinear complexity and Lempel–Ziv complexity of finite length sequences. *IEEE Trans. Inf. Theory* **2007**, *53*, 4293–4302. [CrossRef]
51. Erdmann, D.; Murphy, S. An approximate distribution for the maximum order complexity. *Des. Codes Cryptogr.* **1997**, *10*, 325–339. [CrossRef]
52. Luo, Y.; Xing, C.; You, L. Construction of sequences with high nonlinear complexity from function fields. *IEEE Trans. Inf. Theory* **2017**, *63*, 7646–7650. [CrossRef]
53. Liang, S.; Zeng, X.; Xiao, Z.; Sun, Z. Binary sequences with length n and nonlinear complexity not less than $n/2$. *IEEE Trans. Inf. Theory* **2023**, *69*, 8116–8125. [CrossRef]
54. Rizomiliotis, P. Constructing periodic binary sequences with maximum nonlinear span. *IEEE Trans. Inf. Theory* **2006**, *52*, 4257–4261. [CrossRef]
55. Sun, Z.; Zeng, X.; Li, C.; Helleseth, T. Investigations on periodic sequences with maximum nonlinear complexity. *IEEE Trans. Inf. Theory* **2017**, *63*, 6188–6198. [CrossRef]
56. Yuan, Q.; Li, C.; Zeng, X.; Helleseth, T.; He, D. Further investigations on nonlinear complexity of periodic binary sequences. *IEEE Trans. Inf. Theory* **2024**, 1. [CrossRef]
57. Lempel, A.; Ziv, J. On the complexity of finite sequences. *IEEE Trans. Inf. Theory* **1976**, *22*, 75–81. [CrossRef]
58. Merai, L.; Niederreiter, H.; Winterhof, A. Expansion complexity and linear complexity of sequences over finite fields. *Cryptogr. Commun.* **2017**, *9*, 501–509. [CrossRef]
59. Işık, L.; Winterhof, A. Maximum-order complexity and correlation measures. *Cryptography* **2017**, *1*, 7. [CrossRef]
60. Chen, Z.; Gómez, A.I.; Gómez-Pérez, D.; Tirkel, A. Correlation measure, linear complexity and maximum order complexity for families of binary sequences. *Finite Fields Their Appl.* **2022**, *78*, 101977. [CrossRef]
61. Niederreiter, H. Periodic sequences with large k-error linear complexity. *IEEE Trans. Inf. Theory* **2003**, *49*, 501–505. [CrossRef]
62. Chan, A.H.; Games, R.A.; Key, E.L. On the complexities of DeBruijn sequences. *J. Comb. Theory A* **1982**, *33*, 233–246. [CrossRef]
63. Games, R.A. There are no DeBruijn sequences of span n with complexity $2^{n-1}+n+1$. *J. Comb. Theory A* **1983**, *34*, 248–251. [CrossRef]
64. Mayhew, G.; Golomb, S. Linear spans of modified DeBruijn sequences. *IEEE Trans. Inf. Theory* **1990**, *36*, 1166–1167. [CrossRef]
65. Kyureghyan, G.M. Minimal polynomials of the modified DeBruijn sequences. *Discret. Appl. Math.* **2008**, *156*, 1549–1553. [CrossRef]
66. Dong, Y.J.; Tian, T.; Qi, W.F.; Wang, Z.X. New results on the minimal polynomials of modified DeBruijn sequences. *Finite Fields Their Appl.* **2019**, *60*, 101583. [CrossRef]
67. Wang, H.Y.; Zheng, Q.X.; Wang, Z.X.; Qi, W.F. The minimal polynomials of modified DeBruijn sequences revisited. *Finite Fields Their Appl.* **2020**, *68*, 101735. [CrossRef]
68. Khachatrian, L.H. The lower bound of the quadratic spans of DeBruijn sequences. *Des. Codes Cryptogr.* **1993**, *3*, 29–32. [CrossRef]
69. Klapper, A.; Goresky, M. Cryptanalysis based on 2-adic rational approximation. In *Advances in Cryptology—CRYPTO'95*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 262–273. [CrossRef]
70. Chen, Z.; Chen, Z.; Obrovsky, J.; Winterhof, A. Maximum-order complexity and 2-adic complexity. *arXiv* **2023**, arXiv:2309.12769.
71. Ziv, J.; Lempel, A. A universal algorithm for sequential data compression. *IEEE Trans. Inf. Theory* **1977**, *23*, 337–343. [CrossRef]
72. Ziv, J.; Lempel, A. Compression of individual sequences via variable-rate coding. *IEEE Trans. Inf. Theory* **1978**, *24*, 530–536. [CrossRef]
73. Sun, Z.; Zeng, X.; Li, C.; Zhang, Y.; Yi, L. The expansion complexity of ultimately periodic sequences over finite fields. *IEEE Trans. Inf. Theory* **2021**, *67*, 7550–7560. [CrossRef]
74. Brandstätter, N.; Winterhof, A. Linear complexity profile of binary sequences with small correlation measure. *Period. Math. Hung.* **2006**, *52*, 1–8. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.