Cryptography and Communications
https://doi.org/10.1007/s12095-022-00556-9

Pseudorandom sequences derived from automatic sequences

László Mérai[1] · Arne Winterhof[1]

Received: 12 May 2021 / Accepted: 17 January 2022

Abstract
Many automatic sequences, such as the Thue-Morse sequence or the Rudin-Shapiro sequence, have some desirable features of pseudorandomness such as a large linear complexity and a small well-distribution measure. However, they also have some undesirable properties in view of certain applications. For example, the majority of possible binary patterns never appears in automatic sequences and their correlation measure of order 2 is extremely large. Certain subsequences, such as automatic sequences along squares, may keep the good properties of the original sequence but avoid the bad ones. In this survey we investigate properties of pseudorandomness and non-randomness of automatic sequences and their subsequences and present results on their behaviour under several measures of pseudorandomness including linear complexity, correlation measure of order k, expansion complexity and normality. We also mention some analogs for finite fields.

Keywords Automatic sequences · Pseudorandomness · Linear complexity · Maximum order complexity · Well-distribution measure · Correlation measure · Expansion complexity · Normality · Finite fields

Mathematics Subject Classification (2010) 11A63 · 11B85 · 11K16 · 11K31 · 11K36 · 11K45 · 11T71 · 68R15 · 94A55 · 94A60

1 Introduction

Pseudorandom sequences are sequences generated by deterministic algorithms which simulate randomness. In contrast to truly random sequences they are not random at all but guarantee certain desirable features.

This article belongs to the Topical Collection: *Surveys (invitation only)*

✉ László Mérai
laszlo.merai@oeaw.ac.at

Arne Winterhof
arne.winterhof@oeaw.ac.at

[1] Johann Radon Institute for Computational and Applied Mathematics Austrian Academy of Sciences, Altenbergerstr. 69, 4040 Linz, Austria

Published online: 19 April 2022

Automatic sequences, see Section 2 below for the definition, have some of these desirable features but also some undesirable ones.

For example, the Thue-Morse sequence $(t_n) = (t_n)_{n=0}^{\infty}$, defined by (2.2) below,

- has large Nth linear complexity, see Section 3,
- has large Nth maximum-order complexity, see Section 4,
- is balanced and has a small well-distribution measure, see Section 5.

However, the Thue-Morse sequence

- has a very large correlation measure of order 2, see Section 5,
- has a very small expansion complexity, see Section 6,
- and there are short patterns such as 000 and 111 which do not appear in the sequence and its subword complexity is only linear, see Section 7.

Hence, despite some nice features this sequence does not look random at all, see Fig. 1. The same is true for the Rudin-Shapiro sequence (r_n) defined by (2.3) below and many other related sequences.

Taking suitable subsequences may destroy the non-random structure of the original sequence but may keep the desirable features of pseudorandomness. Promising candidates for such subsequences are

- along squares, cubes, bi-squares, ... or along the values of any polynomial f of degree at least 2 with $f(\mathbb{N}_0) \subset \mathbb{N}_0$,
- along primes,
- along the Piateski-Shapiro sequence $\lfloor n^c \rfloor$, $1 < c < 2$,
- and along geometric sequences such as 3^n.

For example, the Thue-Morse sequence and the Rudin-Shapiro sequence along squares still

- have a large maximum-order complexity and thus a large linear complexity, see Section 4,
- and are asymptotically balanced, see Section 7.

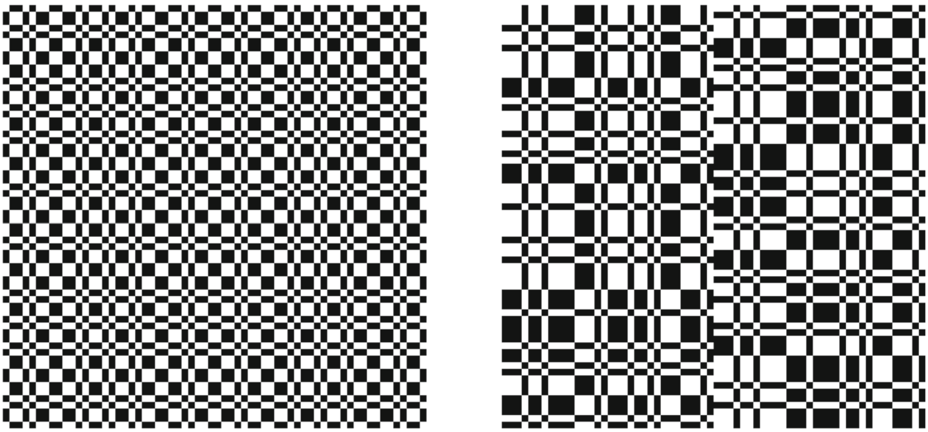

Fig. 1 The first 4096 elements of the Thue-Morse (left) and Rudin-Shapiro (right) sequence split into 64 rows of each 64 sequence elements. Zeros are represented by white, ones are represented by black

Moreover, in contrast to the original sequence they

- have unbounded expansion complexity, see Section 6,
- and are normal, that is, asymptotically each pattern appears with the right frequency in the sequence, see Section 7.

Roughly speaking, they look much more random than the original sequences, see Fig. 2.

Still some questions about these sequences remain open such as upper bounds on the correlation measure of order k and on the expansion complexity. We will state explicitly some selected open problems to motivate future research.

We also look for further directions in Section 8. In particular, we discuss analogs of the Thue-Morse and Rudin-Shapiro sequence and their subsequences in the setting of finite fields.

For general background on automatic sequences and finite automata we refer to the monograph of Allouche and Shallit [8] and also to [7, 9, 36, 37]. For surveys on pseudorandom sequences see [45, 68, 79, 88, 100].

2 Finite automata and automatic sequences

Roughly speaking, a sequence is *automatic* if it is generated by a finite automaton, see Definition 2.2 below.

Definition 2.1 Let $k \geq 2$ be an integer. A *finite k-automaton* $\mathcal{A}$ is a 6-tuple

$$\mathcal{A} = (Q, \Sigma, \delta, q_0, \varphi, \Delta),$$

where

- Q is the finite set of states,
- $\Sigma = \{0, 1, \ldots, k-1\}$ is the input alphabet,
- $\delta : Q \times \Sigma \to Q$ is the transition function,
- $q_0 \in Q$ is the initial state,

Fig. 2 The first 4096 elements of the Thue-Morse (left) and Rudin-Shapiro (right) sequence along squares split into 64 rows of each 64 sequence elements. Zeros are represented by white, ones are represented by black

- Δ is the output alphabet
- and $\varphi : Q \to \Delta$ is the output function.

For a word $w = w_0 \dots w_\ell \in \Sigma^{\ell+1}$ we define $\delta(q, w) = \delta(\delta(q, w_0 \dots w_{\ell-1}), w_\ell)$, $\ell = 0, 1, \dots$ Then the output of the automaton $\mathcal{A}$ for a given input $w \in \Sigma^*$ is $\varphi(\delta(q_0, w))$.

For example, the *Thue-Morse automaton*, see Fig. 3, is a 2-automaton with 2 states and the *Rudin-Shapiro automaton*, see Fig. 4, is a 2-automaton with 4 states, both with inputs and outputs in $\Sigma = \Delta = \{0, 1\}$.

Definition 2.2 Let Δ be a finite set. A sequence (s_n) over Δ is called a k-*automatic sequence* if there is a k-automaton $\mathcal{A}$ such that on input of the digits $n_0, n_1, \dots$ of the k-ary expansion of $n \geq 0$,

$$n = \sum_{i=0}^{\ell} n_i k^i, \quad n_i \in \{0, 1, \dots, k-1\},\ n_\ell \neq 0, \tag{2.1}$$

$\mathcal{A}$ outputs the sequence element $s_n \in \Delta$. Reading of the digits of n starting with the most significant digit n_ℓ is called *direct* whereas reading starting with the least significant digit n_0 is called *reverse*. If not stated otherwise, we use reverse reading. Finally, a sequence is called *automatic* if it is k-automatic for some k.

Note that reverse and direct reading give rise to the same notion of automatic sequences, see [8, Theorem 4.3.3].

Example 1 (Thue-Morse sequence) The *Thue-Morse sequence* (t_n) is a 2-automatic sequence generated by the Thue-Morse automaton, Fig. 3. This sequence is the *sequence of the sum of digits modulo 2*. The sequence begins with

$$011010011001\dots,$$

see also Fig. 1 for a picture of the first 4096 sequence elements. It follows from the defining automaton, see Fig. 3, that (t_n) satisfies the following recurrence relation

$$t_n = \begin{cases} t_{n/2} & \text{if } n \text{ is even,} \\ t_{(n-1)/2} + 1 \bmod 2 & \text{if } n \text{ is odd,} \end{cases} \qquad n = 1, 2, \dots \tag{2.2}$$

with initial value $t_0 = 0$.

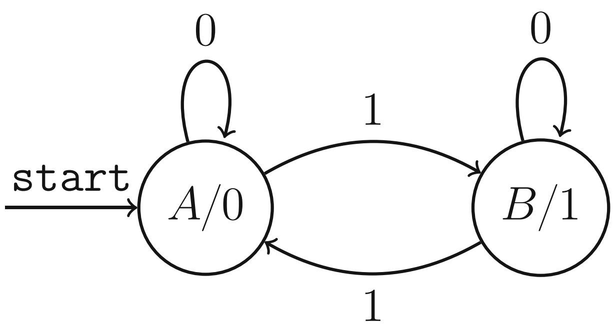

Fig. 3 Thue-Morse automaton

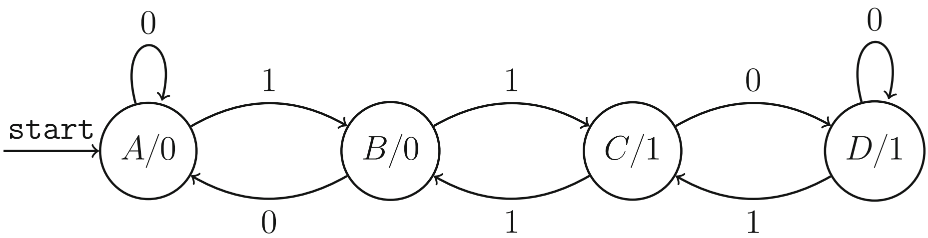

Fig. 4 Rudin-Shapiro automaton

Example 2 (Rudin-Shapiro sequence) The *Rudin-Shapiro sequence* (r_n) is a 2-automatic sequence generated by the Rudin-Shapiro automaton, see Fig. 4. The sequence begins with

$$000100100001\ldots,$$

see also Fig. 1 for a picture of the first 4096 sequence elements. It follows from the defining automaton, see Fig. 4, that (r_n) satisfies the following recurrence relation

$$r_n = \begin{cases} r_{\lfloor n/2 \rfloor} + 1 \bmod 2 & \text{if } n \equiv 3 \bmod 4, \\ r_{\lfloor n/2 \rfloor} & \text{otherwise,} \end{cases} \qquad n = 1, 2, \ldots \tag{2.3}$$

with initial value $r_0 = 0$.

The sequence $((-1)^{r_n})$ over $\{-1, +1\}$ is also called Rudin-Shapiro sequence in the literature. Here we study only the sequence (r_n) over $\{0, 1\}$.

Example 3 (Pattern sequences) For a pattern $P \in \Sigma^\ell \setminus \{(0, \ldots, 0)\}$ of length ℓ over $\Sigma = \{0, 1, \ldots, k-1\}$ define the sequence (p_n) by

$$p_n = e_P(n) \bmod m, \quad 0 \le p_n < m, \quad n = 0, 1, \ldots,$$

where $e_P(n)$ is the number of occurrences of P in the k-ary expansion of n. The sequence (p_n) over $\Delta = \{0, 1, \ldots, m-1\}$ satisfies the following recurrence relation

$$p_n = \begin{cases} p_{\lfloor n/k \rfloor} + 1 \bmod m & \text{if } n \equiv a \bmod k^\ell, \\ p_{\lfloor n/k \rfloor} & \text{otherwise,} \end{cases} \qquad n = 1, 2, \ldots \tag{2.4}$$

with initial value $p_0 = 0$, where $a = a(P)$ is the integer $0 < a < k^\ell$ such that its k-ary expansion $a = \sum_{i=0}^{\ell-1} a_i k^i$ corresponds to the pattern $P = (a_0, a_1, \ldots, a_{\ell-1})$. We focus on the case $k = m$ and call the sequence (p_n) a k-ary pattern sequence.

Classical examples for binary pattern sequences are the Thue-Morse sequence with

$$k = 2, \quad \ell = 1, \quad P = 1 \quad \text{and} \quad a = 1,$$

and the Rudin-Shapiro sequence with

$$k = 2, \quad \ell = 2, \quad P = 11 \quad \text{and} \quad a = 3.$$

In particular, if $n_0, n_1, \ldots$ are the bits of the non-negative integer taken from (2.1) with $k = 2$, then

$$t_n = \sum_{i=0}^{\infty} n_i \bmod 2 \quad \text{and} \quad r_n = \sum_{i=0}^{\infty} n_i n_{i+1} \bmod 2. \tag{2.5}$$

Example 4 (Rudin-Shapiro-like sequence) Lafrance, Rampersad and Yee [53] introduced a *Rudin-Shapiro-like sequence* (ℓ_n) which is based on the number of occurrences of the pattern 10 as a scattered subsequence in the binary representation, (2.1) with $k = 2$, of n. That is, ℓ_n is the parity of the number of pairs (i, j) with $i > j$ and $(n_i, n_j) = (1, 0)$. See Fig. 5 for its defining automaton.

This sequence can also be defined by

$$\ell_{2n+1} = \ell_n \quad \text{and} \quad \ell_{2n} = \ell_n + t_n \bmod 2, \tag{2.6}$$

see [53, (1) and (2)], with initial value $\ell_0 = 0$ and where (t_n) is the Thue-Morse sequence.

Example 5 (Baum-Sweet sequence) The *Baum-Sweet sequence* (b_n) is a 2-automatic sequence defined by the rule $b_0 = 1$ and for $n \geq 1$

$$b_n = \begin{cases} 1 & \text{if the binary representation of } n \text{ contains no block of} \\ & \text{consecutive 0's of odd length,} \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently, we have for $n \geq 1$ of the form $n = 4^\ell m$ with $4 \nmid m$ that

$$b_n = \begin{cases} 0 & \text{if } m \text{ is even,} \\ b_{(m-1)/2} & \text{if } m \text{ is odd.} \end{cases} \tag{2.7}$$

The sequence (b_n) is generated by the Baum-Sweet automaton in Fig. 6.

Example 6 (Characteristic sequence of sums of three squares) Consider the *characteristic sequence* (c_n) *of the set of integers which are sums of three squares of an integer*, that is,

$$c_n = \begin{cases} 1 & \text{if } n = a^2 + b^2 + c^2 \text{ for some non-negative integers } a, b, c, \\ 0 & \text{otherwise.} \end{cases}$$

By Legendre's three-square theorem, we have the equivalent definition

$$c_n = \begin{cases} 1 & \text{if } n \text{ is not of the form } n = 4^\ell(8k+7), \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

See Fig. 7 for the defining automaton.

Example 7 (Regular paper-folding sequence) The *regular paper-folding sequence* (v_n) with initial value $v_0 \in \{0, 1\}$ is defined as follows. If $n = 2^k m$ with an odd m, then

$$v_n = \begin{cases} 1, & m \equiv 1 \bmod 4, \\ 0, & m \equiv 3 \bmod 4, \end{cases} \qquad n = 1, 2, \ldots \tag{2.9}$$

Its defining automaton with four states is given in Fig. 8.

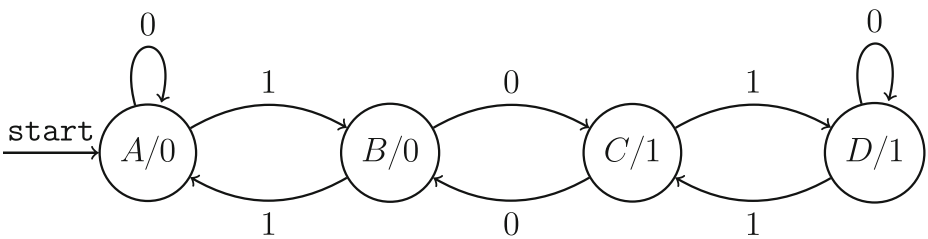

Fig. 5 Rudin-Shapiro-like automaton with direct reading

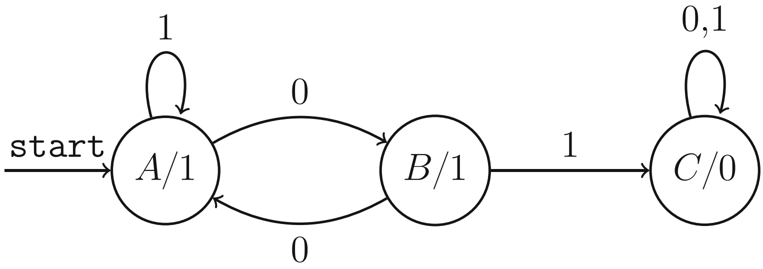

Fig. 6 Baum-Sweet automaton

The name comes from the fact that this sequence represents the left turns, represented by 0, and right turns, represented by 1, along a strip of paper that is folded repeatedly in half in the same direction.

Example 8 (An automatic apwenian sequence) Any binary sequence (a_n) satisfying $a_0 = 1$ and

$$a_{2n+2} = a_{2n+1} + a_n \bmod 2, \quad n = 0, 1, \ldots$$

is called *apwenian*,[1] see for example [4]. Apwenian sequences which are 2-automatic are characterized in [4].For example, the sequence (w_n) defined by

$$w_{2n} = 1 \quad \text{and} \quad w_{2n+1} = w_n + 1 \bmod 2, \quad n = 0, 1, \ldots \tag{2.10}$$

is apwenian and defined by the automaton in Fig. 9.

In addition to the examples above, all ultimately periodic sequences are k-automatic for all integers $k \geq 2$, see [8, Theorem 5.4.2]. Moreover, by Cobham's theorem [8, Theorem 11.2.1], if a sequence (s_n) is both k-automatic and ℓ-automatic and k and ℓ are multiplicatively independent,[2] then (s_n) is ultimately periodic. Note that k-automatic and ℓ-automatic sequences are the same if k and ℓ are multiplicatively dependent by [8, Theorem 6.6.4].

For a prime power $k = q$, k-automatic sequences (s_n) over the finite field[3] $\Delta = \mathbb{F}_q$ can be characterized by a result of Christol, see [18] for prime q and [19] for prime power q as well as [8, Theorem 12.2.5].

Theorem 2.3 *Let*[4]

$$G(x) = \sum_{n=0}^{\infty} s_n x^n \in \mathbb{F}_q[\![x]\!]$$

be the generating function of the sequence (s_n) over $\mathbb{F}_q$. Then (s_n) is q-automatic if and only if $G(x)$ is algebraic over $\mathbb{F}_q(x)$, that is, there is a polynomial $h(x, y) \in \mathbb{F}_q[x, y] \setminus \{0\}$ such that $h(x, G(x)) = 0$.

[1] *Apwenian* sequences are named after the authors of [5].

[2] Two integers k and ℓ are *multiplicatively dependent* if $k^r = \ell^s$ for some positive integers r and s. Otherwise they are *multiplicatively independent*.

[3] For a prime power q we denote the finite field of size q by $\mathbb{F}_q$.

[4] We denote by $\mathbb{F}_q[\![x]\!]$ the ring of formal power series over $\mathbb{F}_q$.

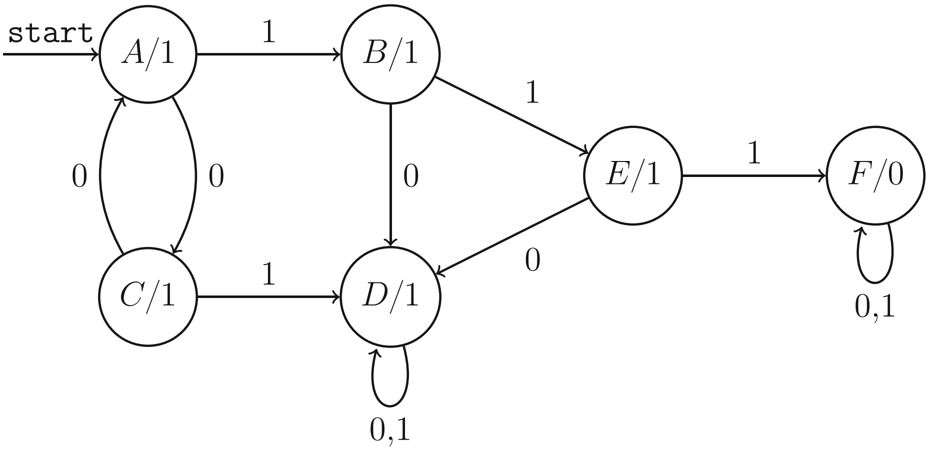

Fig. 7 Automaton of the characteristic sequence of sums of three squares with reverse reading

Example 9 The generating function $G(x)$ of the Thue-Morse sequence (t_n) over $\mathbb{F}_2$ satisfies $h(x, G(x)) = 0$ with

$$h(x, y) = (x+1)^3 y^2 + (x+1)^2 y + x. \tag{2.11}$$

The generating function $G(x)$ of the Rudin-Shapiro sequence (r_n) over $\mathbb{F}_2$ satisfies $h(x, G(x)) = 0$ with

$$h(x, y) = (x+1)^5 y^2 + (x+1)^4 y + x^3. \tag{2.12}$$

In general, for prime p the generating function $G(x)$ of the p-ary pattern sequence (p_n) over $\mathbb{F}_p$ with respect to the pattern P of length ℓ satisfies $h(x, G(x)) = 0$ with

$$h(x, y) = (x-1)^{p^\ell+p-1} y^p - (x-1)^{p^\ell} y - x^{a(P)}. \tag{2.13}$$

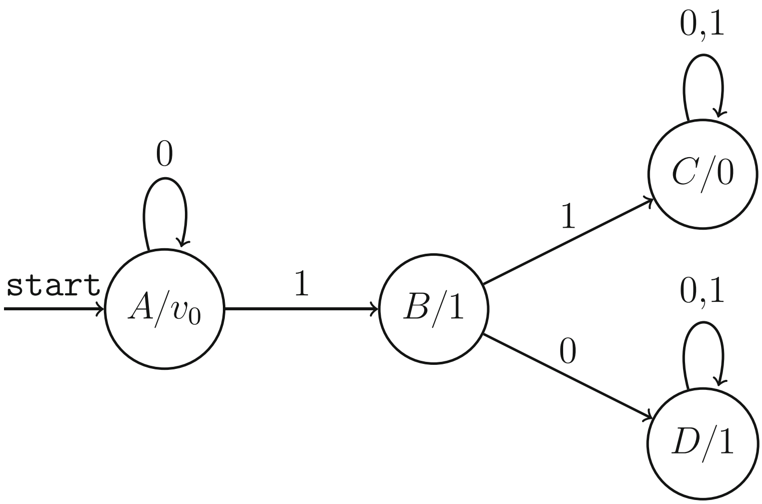

Fig. 8 Regular paper-folding automaton

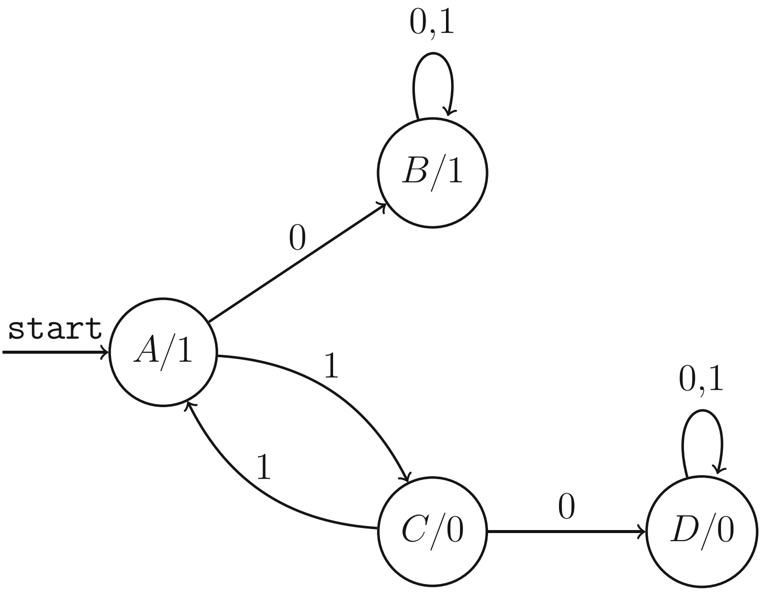

Fig. 9 Apwenian automaton

The generating function $G(x)$ of the Rudin-Shapiro-like sequence (ℓ_n) over $\mathbb{F}_2$ defined by (2.6) satisfies $h(x, G(x)) = 0$ with

$$h(x, y) = (x+1)^8 y^4 + (x^6 + x^5 + x^2 + x)y^2 + (x+1)^4 y + x^2, \tag{2.14}$$

see [96, Proof of Theorem 2].

The generating function $G(x)$ of the Baum-Sweet sequence (b_n) over $\mathbb{F}_2$ satisfies $h(x, G(x)) = 0$ with

$$h(x, y) = y^3 + xy + 1. \tag{2.15}$$

The generating function $G(x)$ of the characteristic sequence (c_n) of sums of three squares (2.8) over $\mathbb{F}_2$ satisfies $h(x, G(x)) = 0$ with

$$h(x, y) = (x+1)^8(y + y^4) + x^6 + x^5 + x^3 + x^2 + x, \tag{2.16}$$

see [47, Equation (7)].

The generating function $G(x)$ of the regular paper-folding sequence (v_n) over $\mathbb{F}_2$ satisfies $h(x, G(x)) = 0$ with

$$h(x, y) = (x+1)^4(y^2 + y) + x. \tag{2.17}$$

The generating function $G(x)$ of the apwenian sequence (w_n) over $\mathbb{F}_2$ defined by (2.10) satisfies

$$h(x, y) = (x+1)(xy^2 + y) + 1. \tag{2.18}$$

3 Linear complexity

The linear complexity is a figure of merit of pseudorandom sequences introduced to capture undesirable linear structure in a sequence. It originates in cryptography and provides a test of randomness which is a standard tool to filter sequences with non-randomness properties and is implemented in many test suites such as NIST and TestU01 [55, 85].

Definition 3.1 The *Nth linear complexity* $L((s_n), N)$ of a sequence (s_n) over $\mathbb{F}_q$ is the length L of a shortest linear recurrence relation satisfied by the first N elements of (s_n),

$$s_{n+L} = c_{L-1}s_{n+L-1} + \cdots + c_1 s_{n+1} + c_0 s_n, \quad 0 \le n \le N - L - 1, \tag{3.1}$$

for some $c_0, \ldots, c_{L-1} \in \mathbb{F}_q$. We use the convention that $L((s_n), N) = 0$ if the first N elements of (s_n) are all zero and $L((s_n), N) = N$ if $s_0 = \cdots = s_{N-2} = 0 \ne s_{N-1}$. The sequence $(L((s_n), N))_{N=1}^{\infty}$ is called *linear complexity profile* of (s_n) and

$$L((s_n)) = \sup_{N \ge 1} L((s_n), N)$$

is the *linear complexity* of (s_n).

Clearly, $0 \le L((s_n), N) \le N$ and $L((s_n), N) \le L((s_n), N+1)$.

We remark, that one can also consider inhomogeneous linear recursions in (3.1). However, the length of a shortest inhomogeneous linear recursion satisfied by the first N sequence elements can differ by at most one from the Nth linear complexity, see [56, page 401].

For truly random sequences (s_n) the expected value of its Nth linear complexity L is

$$\frac{N}{2} + O(\log N),$$

see for example [72, Theorem 10.4.42]. Deviations of order of magnitude $\log N$ must appear for infinitely many N. More precisely, for a prime power q consider the uniform probability measure of sequences over $\mathbb{F}_q$,

$$\mathbb{P}\Big((s_n) \in \mathbb{F}_q^{\infty} : (s_0, \ldots, s_{\ell-1}) = (c_0, \ldots, c_{\ell-1})\Big) = q^{-\ell}, \quad c_0, \ldots, c_{\ell-1} \in \mathbb{F}_q. \tag{3.2}$$

Then we have the following result on the deviation from the expected value, see[76, Theorem 10].

Theorem 3.2 *We have*

$$\limsup_{N\to\infty} \frac{L((s_n), N) - N/2}{\log N} = \frac{1}{2\log q},$$

and

$$\liminf_{N\to\infty} \frac{L((s_n), N) - N/2}{\log N} = \frac{-1}{2\log q}$$

with probability one with respect to the uniform probability measure (3.2).

It is well-known [75, Lemma 1] that $L((s_n)) < \infty$ if and only if (s_n) is ultimately periodic, that is, its generating function is rational: $G(x) = g(x)/f(x)$ with polynomials $g(x), f(x) \in \mathbb{F}_q[x]$.

The Nth linear complexity is a measure for the unpredictability of a sequence. A large Nth linear complexity, up to sufficiently large N, is necessary, but not sufficient, for cryptographic applications. Sequences of small linear complexity are also weak in view of Monte-Carlo methods, see [28–31]. For more background on linear complexity and related measures of pseudorandomness we refer to [72, Section 10.4] and [77, 100, 102].

Mérai and Winterhof [70] showed that automatic sequences which are not ultimately periodic possess large Nth linear complexity.

Theorem 3.3 *Let q be a prime power and (s_n) be a q-automatic sequence over $\mathbb{F}_q$ which is not ultimately periodic. Let $h(x, y) = h_0(x) + h_1(x)y + \cdots + h_d(x)y^d \in \mathbb{F}_q[x, y]$ be a non-zero polynomial $h(x, G(x)) = 0$ with no rational function $r(x) \in \mathbb{F}_q(x)$ satisfying $h(x, r(x)) = 0$.*

Put

$$M = \max_{0 \le i \le d} \{\deg h_i - i\}. \tag{3.3}$$

Then we have

$$\frac{N - M}{d} \le L((s_n), N) \le \frac{(d-1)N + M + 1}{d}.$$

See also [104] for the special case $d = 2$.

The idea of the proof of Theorem 3.3 is that small Nth linear complexity profile gives a good rational approximation to the generating function. However, transcendental elements over $\mathbb{F}_q(x)$ are not well-approximated.

Namely, since (s_n) is not ultimately periodic, $G(x) = \sum\limits_{n=0}^{\infty} s_n x^n \notin \mathbb{F}_q(x)$ is not rational by [75, Lemma 1].

Let $g(x)/f(x) \in \mathbb{F}_q(x)$ be a rational zero of $h(x, y)$ modulo x^N with $\deg(f) \le L((s_n), N)$ and $\deg(g) < L((s_n), N)$. More precisely, put $L = L((s_n), N)$. Then we have

$$\sum_{\ell=0}^{L} c_\ell s_{n+\ell} = 0 \quad \text{for } 0 \le n \le N - L - 1$$

for some $c_0, \ldots, c_L \in \mathbb{F}_p$ with $c_L = -1$. Take

$$f(x) = \sum_{\ell=0}^{L} c_\ell x^{L-\ell}$$

and

$$g(x) = \sum_{m=0}^{L-1} \left(\sum_{\ell=L-m}^{L} c_\ell s_{m+\ell-L} \right) x^m$$

and verify

$$f(x)G(x) \equiv g(x) \bmod x^N.$$

Then

$$h_0(x) f^d(x) + h_1(x) g(x) f^{d-1}(x) + \cdots + h_d(x) g(x)^d = K(x) x^N.$$

Here $K(x) \ne 0$ since $h(x, y)$ has no rational zero. Comparing the degrees of both sides we get

$$dL + M \ge N$$

which gives the lower bound.

The upper bound for $N = 1$ is trivial. For $N \ge 2$ the result follows from the well-known bound, see for example [31, Lemma 3],

$$L((s_n), N) \le \max \{L((s_n), N - 1), N - L((s_n), N - 1)\}$$

by induction.

The bound in Theorem 3.3 combined with (2.11)-(2.18) gives estimates for the Nth linear complexity of the automatic sequences of Section 2.

For the Thue-Morse sequence (t_n) defined by (2.2) we have

$$\left\lceil \frac{N-1}{2} \right\rceil \le L((t_n), N) \le \left\lfloor \frac{N}{2} \right\rfloor + 1. \tag{3.4}$$

For the Rudin-Shapiro sequence (r_n) defined by (2.3) and for the regular paper-folding sequence (v_n) defined by (2.9) we have

$$\left\lceil \frac{N-3}{2} \right\rceil \leq L((r_n), N), L((v_n), N) \leq \left\lfloor \frac{N}{2} \right\rfloor + 2. \tag{3.5}$$

For the p-ary pattern sequence (p_n) defined by (2.4) with any pattern P of length ℓ we have

$$\left\lceil \frac{N+1}{p} \right\rceil - p^{\ell-1} \leq L((p_n), N) \leq \left\lfloor \frac{(p-1)N}{p} \right\rfloor + p^{\ell-1}.$$

For the Rudin-Shapiro-like sequence (ℓ_n) defined by (2.6) we have

$$\left\lceil \frac{N}{4} \right\rceil - 1 \leq L((\ell_n), N) \leq \left\lfloor \frac{3N+5}{4} \right\rfloor.$$

For the Baum-Sweet sequence (b_n) defined by (2.7) we have

$$\left\lceil \frac{N}{3} \right\rceil \leq L((b_n), N) \leq \left\lfloor \frac{2N+1}{3} \right\rfloor.$$

For the characteristic sequence (c_n) of sums of three squares defined by (2.8) we have

$$\left\lceil \frac{N-7}{4} \right\rceil \leq L((c_n), N) \leq \left\lfloor \frac{3N}{4} \right\rfloor + 2.$$

Finally, for the apwenian sequence (w_n) defined by (2.10) we have

$$L((w_n), N) = \left\lfloor \frac{N+1}{2} \right\rfloor. \tag{3.6}$$

Note that the bound (3.4) is also true for the dual (t'_n) of the Thue-Morse sequence, that is, $t'_n = 1 - t_n$, and apwenian sequences are characterized by the property (3.6), see [4]. Note that not all apwenian sequences are automatic.

The bounds (3.4) for the Thue-Morse sequence and (3.5) for the Rudin-Shapiro sequence are optimal. Using the continued fraction expansions of their generating functions, Mérai and Winterhof [70] determined the exact value of the Nth linear complexity profiles of the Thue-Morse and Rudin Shapiro sequence.

Theorem 3.4 *The Nth linear complexity of the Thue-Morse sequence is*

$$L((t_n), N) = 2\left\lfloor \frac{N+2}{4} \right\rfloor, \quad N = 1, 2, \ldots$$

and the Nth linear complexity of the Rudin-Shapiro sequence is

$$L((r_n), N) = \begin{cases} 6\lfloor N/12 \rfloor + 4, & N \equiv 4, 5, 6, 7, 8, 9 \bmod 12, \\ 6\lfloor (N+2)/12 \rfloor, & \textit{otherwise}. \end{cases}$$

The result can be extended to binary pattern sequences (p_n) defined by (2.4) with the all one pattern of length $\ell \geq 3$, that is, $a = 2^\ell - 1$.

It follows from Theorem 3.3, that if an automatic sequence is not ultimately periodic and its generating function has a quadratic minimal polynomial, that is $d = 2$ in Theorem 3.2, then the deviation of the Nth linear complexity from its expected value $N/2$ is bounded by $(M+1)/2$,

$$\left| L((s_n), N) - \frac{N}{2} \right| \leq \frac{M+1}{2},$$

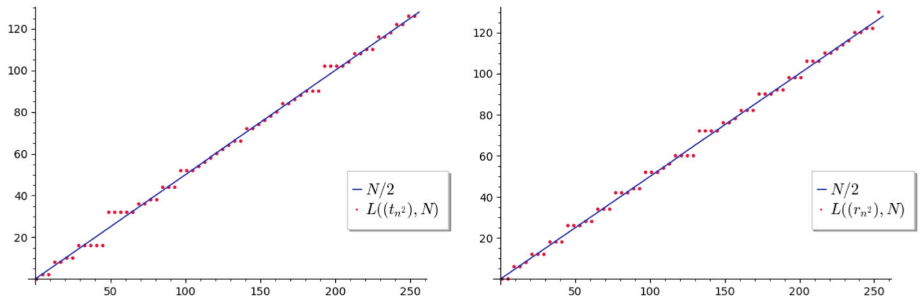

Fig. 10 The Nth linear complexity of the Thue-Morse (left) and Rudin-Shapiro (right) sequence along squares

where M is defined by (3.3). Such sequences are said to have *almost perfect* or $(M+1)$-*perfect* linear complexity profile, see [4, 75].

Apwenian sequences are those sequences having 1-perfect or just *perfect* linear complexity profile. The bounds (3.4) and (3.5) imply that the Thue-Morse sequence has 2-perfect linear complexity profile and the Rudin-Shapiro sequence and the paper-folding sequence both have 4-perfect linear complexity profile.

Although automatic sequences have some good pseudorandom properties including a desirable linear complexity profile, these sequences have also some strong non-randomness properties, see Sections 5, 6 and 7 below. Such randomness flaws may be avoided considering subsequences of automatic sequences. For example, the Thue-Morse and Rudin-Shapiro sequences along squares are not automatic, see Section 7 below, and seem to have Nth linear complexity close to $N/2$, see Fig. 10.

Problem 1 Prove that the Nth linear complexities of the Thue-Morse and Rudin-Shapiro sequences along squares satisfy [5]

$$L((t_{n^2}), N) = \frac{N}{2} + o(N) \quad \text{and} \quad L((r_{n^2}), N) = \frac{N}{2} + o(N).$$

We remark, that lower bounds on the Nth linear complexities of (t_{n^2}) and (r_{n^2}) of order of magnitude $\sqrt{N}$ follow from Theorem 4.3 and (4.2) in the next section.

In addition to these examples, the same problem is also open for other subsequences such as along other polynomial values, along primes etc.

4 Maximum order complexity

Maximum order (or nonlinear) complexity is a refinement of the linear complexity considering not only linear but *any* recurrence relation.

Definition 4.1 The *Nth maximum order complexity* $M((s_n), N)$ is the smallest positive integer M with

$$s_{n+M} = f(s_{n+M-1}, \ldots, s_n), \quad 0 \le n \le N - M - 1, \tag{4.1}$$

[5] $f(k) = o(g(k))$ is equivalent to $f(k)/g(k) \to 0$ as $k \to \infty$.

for some mapping $f : \mathbb{F}_2^M \to \mathbb{F}_2$. The sequence $(M((s_n), N))_{N=1}^{\infty}$ is called *maximum order complexity profile*.

The definition can be easily extended to sequences over any finite field $\mathbb{F}_q$. However, for large q, even a sequence (s_n) defined by $s_{n+1} = f(s_n)$, $n = 0, 1, \ldots$, with a non-constant polynomial $f(x) \in \mathbb{F}_q[x]$ of degree at most $q-1$ can have desirable features of pseudorandomness, see for example [103], but has maximum order complexity 1. Hence, we restrict ourselves to the case $q = 2$. For large q, the *nonlinear complexity of order d* is more appropriate, where the local degrees of the recurrence polynomial f in (4.1) are bounded by d, see for example [78].

Obviously, we have

$$M((s_n), N) \leq L((s_n), N) \tag{4.2}$$

and the maximum order complexity is a finer measure for the unpredictability of a binary sequence than the linear complexity. However, often the linear complexity is easier to analyze both theoretically and algorithmically.

Clearly, a sufficiently large maximum order complexity is needed for unpredictability and suitability in cryptography. However, sequences of very large maximum order complexity have also a very large autocorrelation or correlation measure of order 2, see (5.7) below, and are not suitable for many applications including cryptography, radar, sonar and wireless communications.

The maximum order complexity was introduced by Jansen in [50, Chapter 3], see also [51]. The typical value for the Nth maximum order complexity is of order of magnitude $\log N$, see [50, 51]. An algorithm for calculating the maximum order complexity profile of linear time and memory was presented by Jansen [50, 51] using the graph algorithm introduced by Blumer et al. [11].

The maximum order complexity of the Thue-Morse sequence was determined in [95, Theorem 1].

Theorem 4.2 *For $N \geq 4$, the Nth maximum order complexity of the Thue-Morse sequence (t_n) satisfies*

$$M((t_n), N) = 2^{\ell} + 1,$$

where

$$\ell = \left\lceil \frac{\log(N/5)}{\log 2} \right\rceil.$$

It is easy to see that

$$\frac{N}{5} + 1 \leq M((t_n), N) \leq 2\frac{N-1}{5} + 1 \quad \text{for} \quad N \geq 4. \tag{4.3}$$

In Section 5 we will see that such a large maximum order complexity points to undesirable structure in a sequence.

The Nth maximum order complexity of the Rudin-Shapiro sequence and some generalizations is also of order of magnitude N, see [95, Theorem 2]. In particular we have

$$M((r_n), N) \geq \frac{N}{6} + 1, \quad N \geq 4. \tag{4.4}$$

The maximum order complexity of the subsequences of the Thue-Morse and the Rudin-Shapiro sequence along squares are still large enough, see [94].

Theorem 4.3 *The Nth maximum order complexities $M((t_{n^2}),N)$ and $M((r_{n^2}),N)$ of the subsequences (t_{n^2}) and (r_{n^2}) of the Thue-Morse and the Rudin-Shapiro sequence along squares satisfy*

$$M((t_{n^2}),N) \geq \sqrt{\frac{2N}{5}}, \quad N \geq 21, \quad \textit{and}$$
$$M((r_{n^2}),N) \geq \sqrt{\frac{N}{8}}, \quad N \geq 64.$$

We sketch the proof. First, let T be the length of the longest subsequence of (t_{n^2}) that occurs at least twice with different successors among the first N sequence elements. Then $M((t_{n^2}),N) \geq T+1$. Hence the first inequality follows from

$$t_{(i+2^{\ell+1})^2} = t_{(i+2^{\ell+2})^2}, \quad i = 0,1,\ldots,\left\lfloor \sqrt{2^{\ell+2}-1} \right\rfloor$$
$$\text{and} \quad t_{(2^\ell+2^{\ell+1})^2} \neq t_{(2^\ell+2^{\ell+2})^2},$$

which can be shown by induction over $\ell \geq 2$, where ℓ is defined by $5\cdot 2^\ell < N \leq 5\cdot 2^{\ell+1}$.

The second bound follows from

$$r_{(i+2^{\ell+3})^2} = r_{(i+2^{\ell+4})^2}, \quad i = 0,1,\ldots,\left\lfloor \sqrt{2^{\ell+3}-1} \right\rfloor,$$
$$\text{and} \quad r_{(2^{\ell+2}+2^{\ell+3})^2} \neq r_{(2^{\ell+2}+2^{\ell+4})^2},$$

where ℓ is defined by $2^{\ell+5} \leq N < 2^{\ell+6}$.

Figure 11 suggests that $\sqrt{N}$ is the right order of magnitude for the Nth maximum order complexities of (t_{n^2}) and (r_{n^2}). For $N \geq 2^{2\ell+2}$ the same lower bound $\sqrt{N/8}$ is true for binary pattern sequences along squares with the all one pattern of length ℓ, that is, $a = 2^\ell - 1$ for $\ell \geq 3$, see [94].

This result was extended by Popoli [82] to sequences along polynomial values of higher degrees d. However, the lower bounds are of order of magnitude $N^{1/d}$. Note that no better lower bounds are known for the Nth linear complexity of these subsequences of automatic sequences.

The problem for other subsequences is still open as for example for subsequences along primes.

Problem 2 Study the maximum-order complexity of the subsequences of the Thue-Morse and the Rudin-Shapiro sequence along primes.

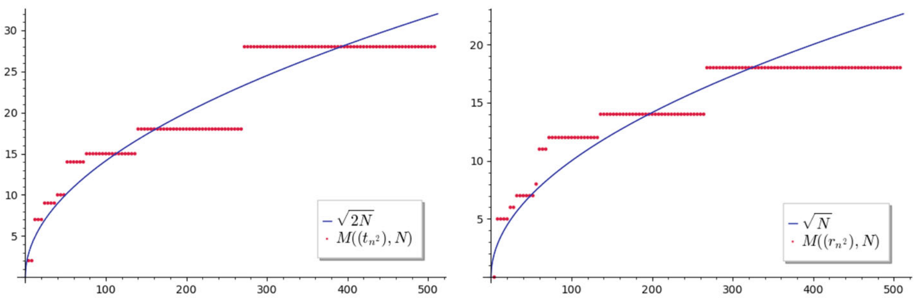

Fig. 11 The Nth maximum order complexity of the Thue-Morse (left) and Rudin-Shapiro (right) sequence along squares

The maximum order complexity of other automatic sequences has also been studied. Sun, Zeng and Lin [96] showed that the Nth maximum order complexity of the Rudin-Shapiro-like sequence (ℓ_n) defined by (2.6) is of order of magnitude N.

We remark, that in addition to automatic sequences based on the k-ary expansion (2.1) of integers, one can consider analogously sequences using other numeration systems.

In particular, consider the *Fibonacci numbers* defined by

$$F_0 = 0,\ F_1 = 1 \quad \text{and} \quad F_n = F_{n-1} + F_{n-2} \text{ for } n \geq 2.$$

Then the unique, see for example [8, Theorem 3.8.1], *Zeckendorf expansion* or *Fibonacci expansion*, of a positive integer n is

$$n = \sum_{i=0}^{\infty} e_i F_{i+2}, \quad \text{where } e_i \in \{0,1\} \text{ and } e_i e_{i+1} = 0 \text{ for } i = 0, 1, \ldots$$

Analogously to the Thue-Morse sum-of-digits sequence (t_n) and the Rudin-Shapiro sequence (r_n) which can be defined by (2.5) we can define and study the *Zeckendorf sum-of-digits sequences modulo 2* (z_n) and (u_n) defined by

$$z_n = \sum_{i=0}^{\infty} e_i \bmod 2 \quad \text{and} \quad u_n = \sum_{i=0}^{\infty} e_i e_{i+2} \bmod 2. \tag{4.5}$$

Very recently the maximum-order complexity of (z_n) and its subsequences along polynomial values has been studied by Jamet, Popoli and Stoll in [49]. A lower bound on $M((u_n), N)$ and some generalizations can be obtained along the same lines and will be contained in Popoli's thesis.

5 Well-distribution and correlation measures

Mauduit and Sárközy [64] introduced two measures of pseudorandomness for finite sequences over $\{-1, +1\}$, the well-distribution measure and the correlation measure of order k. We adjust these definitions to infinite binary sequences (s_n) over $\mathbb{F}_2$.

Definition 5.1 The *Nth well-distribution measure* of (s_n) is defined as

$$W((s_n), N) = \max_{a,b,t} \left| \sum_{j=1}^{t} (-1)^{s_{a+jb}} \right|,$$

where the maximum is taken over all integers $a, b, t \in \mathbb{Z}$ with $b \geq 1$ and $0 \leq a + b \leq a + tb \leq N - 1$.

The well-distribution measure provides information on the balance, [6] that is the distribution of zeros and ones, along arithmetic progressions. For random sequences it is expected to be small. More precisely, Alon et al. [10, Theorem 1] proved the following result on the typical value of the well-distribution measure.

[6] Note that the term *balanced* is used with a different meaning in combinatorics on words, see for example [8, Definition 10.5.4].

Theorem 5.2 *For all $\varepsilon > 0$, there are numbers $N_0 = N_0(\varepsilon)$ and $\delta = \delta(\varepsilon) > 0$ such that for $N \geq N_0$ we have*

$$\delta\sqrt{N} < W((s_n), N) < \frac{\sqrt{N}}{\delta}$$

with probability at least $1 - \varepsilon$ with respect to the uniform probability measure (3.2).

Moreover, Aistleitner [1] showed that there exists a continuous limit distribution of $\frac{W((s_n),N)}{\sqrt{N}}$. More precisely, for any $t \in \mathbb{R}$ the limit

$$F(t) = \lim_{N\to\infty} \mathbb{P}\left(\frac{W((s_n), N)}{\sqrt{N}} \leq t\right)$$

exists and satisfies

$$\lim_{t\to\infty} t(1 - F(t))e^{t^2/2} = \frac{8}{\sqrt{2\pi}}, \tag{5.1}$$

where $\mathbb{P}$ corresponds to the uniform probability measure (3.2). We remark that the distribution $F(t)$ is *not* the normal distribution. However, one can compare the tail estimate (5.1) with the corresponding asymptotic result for the tail probabilities $1 - \Phi(t)$ of a standard normal random variable for which

$$\lim_{t\to\infty} t(1 - \Phi(t))e^{t^2/2} = \frac{1}{\sqrt{2\pi}}.$$

Definition 5.3 For $k \geq 1$, the *Nth correlation measure of order k* of a binary sequence (s_n) is

$$C_k((s_n), N) = \max_{M,D} \left| \sum_{n=0}^{M-1} (-1)^{s_{n+d_1}} \cdots (-1)^{s_{n+d_k}} \right|,$$

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ with integers satisfying $0 \leq d_1 < d_2 < \cdots < d_k$ and $1 \leq M \leq N - d_k$.

The correlation measure of order k provides information about the independence of parts of the sequence and their shifts. For a random sequence this similarity and thus the correlation measure of order k is expected to be small. More precisely, Alon et al. [10, Theorem 2] proved the following result on the typical value of the correlation measure of order k.

Theorem 5.4 *For any $\varepsilon > 0$, there exist an $N_0 = N_0(\varepsilon)$ such that for all $N \geq N_0$ we have for a randomly chosen sequence (s_n) and any k with $2 \leq k \leq N/4$,*

$$\frac{2}{5}\sqrt{N \log \binom{N}{k}} < C_k((s_n), N) < \frac{7}{4}\sqrt{N \log \binom{N}{k}}$$

with probability at least $1 - \varepsilon$ with respect to the uniform probability measure (3.2).

Moreover, Schmidt [86, Theorem 1.1] showed, that for fixed k, we have

$$\lim_{N\to\infty} \frac{C_k((s_n), N)}{\sqrt{2N \log \binom{N}{k-1}}} = 1$$

with probability 1 with respect to the uniform probability measure (3.2).

A large well-distribution measure implies a large correlation measure of order 2. More precisely we have by [66, Theorem 1] [7]

$$W((s_n), N) = O\left(\sqrt{NC_2((s_n), N)}\right).$$

Mauduit and Sárközy [65] obtained bounds on the well-distribution measure and correlation measure of order 2 of Thue-Morse sequence (t_n) and Rudin-Shapiro sequence (r_n).

For example, as a consequence of the bound

$$\left|\sum_{n=0}^{N-1}(-1)^{t_n}z^n\right| \le (1+\sqrt{3})N^{\log 3/\log 4}, \quad |z|=1, \tag{5.2}$$

of Gel'fond [41, p. 262], see [38] for the explicit constant $1+\sqrt{3}$, they obtained a bound on $W((t_n), N)$.

Theorem 5.5 *We have*

$$W((t_n), N) \le 2(1+\sqrt{3})N^{\log 3/\log 4}.$$

Also, using the bound

$$\left|\sum_{n=0}^{N-1}(-1)^{r_n}z^n\right| \le (2+\sqrt{2})N^{1/2}, \quad |z|=1, \tag{5.3}$$

obtained by Rudin [84] and Shapiro [87], see also [8, Theorem 3.3.2], they proved a bound on $W((r_n), N)$.

Theorem 5.6 *We have*

$$W((r_n), N) \le 2(2+\sqrt{2})N^{1/2}.$$

In general, following the proofs of [65] we get

$$W((s_n), N) \le 2 \sup_{|z|=1, m\le N}\left|\sum_{n=0}^{m-1}(-1)^{s_n}z^n\right| \tag{5.4}$$

and thus Theorems 5.5 and 5.6 follow from (5.2) and (5.3).

However, for (t_n) and (r_n) Mauduit and Sárközy [65] detected non-randomness properties by showing that the correlation measure of order 2 of these sequences is large.

Theorem 5.7 *We have*

$$C_2((t_n), N) > \frac{N}{12}, \quad N \ge 5, \tag{5.5}$$

and

$$C_2((r_n), N) > \frac{N}{6}, \quad N \ge 4. \tag{5.6}$$

Mérai and Winterhof [69] showed that all automatic sequences share the property of having a large correlation measure of order 2. They provided the following lower bound in terms of the defining automaton.

[7] $f(k) = O(g(k))$ is equivalent to $|f(k)| \le cg(k)$ for some constant $c > 0$.

Theorem 5.8 *Let (s_n) be a k-automatic binary sequence generated by the finite automaton $(Q, \Sigma, \delta, q_0, \varphi, \{0, 1\})$. Then*

$$C_2((s_n), N) \geq \frac{N}{k(|Q|+1)} \quad \text{for } N \geq k(|Q|+1).$$

This result applied to (t_n) and (r_n) gives the following bounds

$$C_2((t_n), N) \geq \frac{N}{6}, \quad N \geq 6, \quad \text{and} \quad C_2((r_n), N) \geq \frac{N}{10}, \quad N \geq 10,$$

which improves (5.5).

Figures 12 and 13 may lead to the conjecture that both well-distribution measure and correlation measure of order 2 of both (t_{n^2}) and (r_{n^2}) are at most $N^{1/2+o(1)}$.

Problem 3 For fixed $k = 2, 3, \ldots$ show that

$$C_k((t_{n^2}), N) = o(N) \quad \text{and} \quad C_k((r_{n^2}), N) = o(N).$$

Mauduit and Rivat [63] showed that

$$\left| \sum_{n=0}^{N-1} (-1)^{r_{n^2}} \cdot z^n \right| = O\left(N^{1-\eta}\right), \quad |z| = 1, \quad \text{for some } \eta > 0,$$

which, together with (5.4), gives a bound on $W((r_{n^2}), N)$ of the same order of magnitude. More precisely, [63] deals with the more general case of binary pattern sequences (p_n) defined by (2.13) with either the all one pattern of length $k \geq 2$, that is, $a = 2^k - 1$, or the patterns $10\ldots01$ of length $k \geq 3$, that is, $a = 2^{k-1} + 1$, and the constants depend on k. For the Thue-Morse sequence along squares (t_{n^2}) one can easily derive a nontrivial bound on

$$\left| \sum_{n=0}^{N-1} (-1)^{t_{n^2}} \cdot z^n \right|, \quad |z| = 1,$$

and thus on $W((t_{n^2}), N)$ since the proof of [60, Théorème 1] for $z = 1$ works also for $z \neq 1$ since after applying a variant of the van der Corput inequality, [60, Lemma 15], we get an expression which does not depend on the variable z anymore, that is, the same expression as for $z = 1$.

Theorem 4.2 in Section 4 above shows that the Thue-Morse sequence has maximum order complexity $M((t_n), N)$ of order of magnitude N. Although a large maximum order

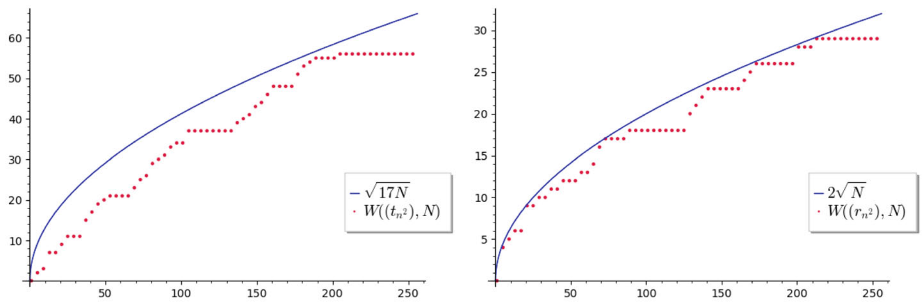

Fig. 12 The Nth well-distribution measure of the Thue-Morse (left) and Rudin-Shapiro (right) sequence along squares

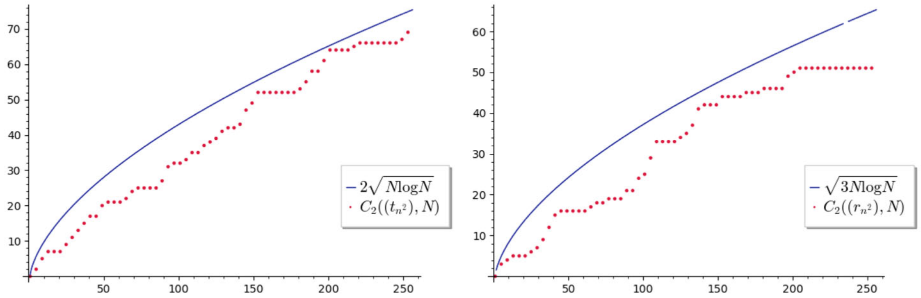

Fig. 13 The Nth second order correlation measure of the Thue-Morse (left) and Rudin-Shapiro (right) sequence along squares

complexity is desired it should be not too large since otherwise the correlation measure of order 2 is large. Namely, we have

$$C_2((s_n), N) \geq M((s_n), N) - 1 \tag{5.7}$$

since by [50, Proposition 3.1] there exist $0 \leq n_1 < n_2 \leq N - M((s_n), N) - 1$ with

$$s_{n_1+i} = s_{n_2+i}, \quad i = 0, \ldots, M((s_n), N) - 2, \quad \text{but } s_{n_1+M((s_n),N)-1} \neq s_{n_2+M((s_n),N)-1}$$

and thus

$$M((s_n), N) - 1 = \sum_{i=0}^{M((s_n),N)-2} (-1)^{s_{n_1+i}+s_{n_2+i}} \leq C_2((s_n), N).$$

Combining (4.3) and (5.7) we get for the Thue-Morse sequence

$$C_2((t_n), N) \geq \frac{N}{5}, \quad N \geq 4,$$

which further improves the constant in (5.5). Combining (4.4) and (5.7) recovers (5.6). The correlation measure of order 2 with bounded lags of some generalizations of the Rudin-Shapiro sequence has recently been studied in [58].

The Baum-Sweet sequence (b_n), the characteristic sequence (c_n) of the sums of three squares, the paper-folding sequence (v_n) and the apwenian sequence (w_n) defined by (2.10) are very unbalanced and thus have all well-distribution measure of order of magnitude N. However, it seems to be interesting to study the well-distribution measure for arbitrary apwenian sequences. For the Rudin-Shapiro like sequence (ℓ_n) defined by (2.6) Lafrance, Rampersad and Yee [53] proved

$$\liminf_{N\to\infty} \frac{\sum_{n=0}^{N-1}(-1)^{\ell_n}}{\sqrt{N}} = \frac{\sqrt{3}}{3} \quad \text{and} \quad \limsup_{N\to\infty} \frac{\sum_{n=0}^{N-1}(-1)^{\ell_n}}{\sqrt{N}} = \sqrt{2}.$$

However, a bound on $W((\ell_n), N)$ is not known and in contrast to (5.3) for the Rudin-Shapiro sequence (r_n), for (ℓ_n) the absolute values

$$\left|\sum_{n=0}^{N-1}(-1)^{\ell_n} z^n\right|$$

can be of much larger order of magnitude than $\sqrt{N}$ for some z with $|z| = 1$, see [3, Theorem 2] as well as [17].

Finally, we remark that the result of Theorem 5.8 provides an estimate on the *state complexity* of sequences in terms of the correlation measure of order 2.

Definition 5.9 Let $k \geq 2$. Then the *Nth state complexity* $SC_k((s_n), N)$ of a sequence (s_n) over $\mathbb{F}_2$ is the minimum of the number of states of any finite k-automaton which generates the first N sequence elements.

Corollary 5.10 *Let (s_n) be a binary sequence. Then for all $k \geq 2$ we have*

$$SC_k((s_n), N) \geq \frac{N}{k \cdot C_2((s_n), N)} - 1 \quad \textit{for } N \geq 3.$$

6 Expansion complexity

Theorem 3.3 indicates that automatic sequences possess good properties in terms of the linear complexity profile. However, the results of Section 5 show that these sequences have a serious lack of pseudorandomness. Diem [26] showed that these sequences are not just statistically auto-correlated, but are completely predictable from a relatively short initial segment. He introduced the notion of *expansion complexity* to turn such security flaw into a quantitative form.

Definition 6.1 Let (s_n) be a sequence over $\mathbb{F}_q$ with generating function

$$G(x) = \sum_{n=0}^{\infty} s_n x^n \in \mathbb{F}_q[\![x]\!].$$

For a positive integer N, the *Nth expansion complexity* $E((s_n), N)$ of (s_n) is $E((s_n), N) = 0$ if $s_0 = \cdots = s_{N-1} = 0$ and otherwise the least total degree of a non-zero polynomial $h(x, y) \in \mathbb{F}_q[x, y]$ such that

$$h(x, G(x)) \equiv 0 \bmod x^N. \tag{6.1}$$

The sequence $(E((s_n), N))_{N=1}^{\infty}$ is called *expansion complexity profile of* (s_n) and

$$E((s_n)) = \sup_{N \geq 1} E((s_n), N)$$

is the *expansion complexity of* (s_n).

By Christol's Theorem 2.3, a sequence is q-automatic if and only if its expansion complexity is finite. For example, we have for the Thue-Morse sequence (t_n), the Rudin-Shapiro sequence (r_n), the p-ary pattern sequence (p_n), the Baum-Sweet sequence (b_n), the Rudin-Shapiro like sequence (ℓ_n), the characteristic sequence (c_n) of sums of three squares, the regular paper-folding sequence (v_n) and the apwenian sequence (w_n) that

$$E((t_n)) = 5, \quad E((r_n)) = 7, \quad E((p_n)) \leq p^{\ell} + 2p - 1, \quad E((b_n)) = 3, \quad E((\ell_n)) \leq 12,$$
$$E((c_n)) \leq 12, \quad E((v_n)) = 6 \quad \text{and} \quad E((w_n)) = 4,$$

which follows from (2.11), (2.12), (2.13), (2.14), (2.15), (2.16), (2.17) and (2.18). The equalities follow from the fact that there is no lower degree polynomial with such property since $h(x, y)$ is irreducible in these cases, see [26, Proposition 4].

Diem showed [26] that if a sequence has small expansion complexity, then long parts of such sequences can be computed efficiently from short ones. We summarize his results.

Theorem 6.2 *Let (s_n) be a sequence over $\mathbb{F}_q$ with expansion complexity $E((s_n)) = d$. From the first d^2 elements, one can compute an irreducible polynomial $h(x, y) \in \mathbb{F}_q[x, y]$ of degree $\deg h \leq d$ with $h(x, G(x)) = 0$ in polynomial time in $d \cdot \log q$.*

Moreover, an initial segment of the sequence of length $M > N$ can be determined from h and the d^2 initial values in $O\left(d^{O(1)}(\log q)^{O(1)}M\right)$ $\mathbb{F}_q$-operations.

Theorem 6.2 shows that automatic sequences have a strong non-randomness property. The expansion complexity profile is defined to capture such non-randomness property locally, that is for initial segments of sequences.

For the Nth expansion complexity, we have the trivial bound $E((s_n), N) \leq N-1$ realized by the polynomial

$$h(x, y) = y - \sum_{n=0}^{N-1} s_n x^n.$$

Moreover, one can show the stronger upper bound

$$\binom{E((s_n), N) + 1}{2} \leq N, \tag{6.2}$$

which holds for all sequence (s_n) and all $N \geq 1$, see [43, Theorem 1].

The Nth expansion complexity of a random sequence is of the same order of magnitude $N^{1/2}$ as the upper bound (6.2), see [42, Theorem 2].

Theorem 6.3 *We have*

$$\liminf_{N\to\infty} \frac{E((s_n), N)}{\sqrt{N}} \geq \frac{\sqrt{2}}{2},$$

with probability one with respect to the uniform probability measure (3.2).

One can estimate the Nth expansion complexity $E((s_n), N)$ in terms of the Nth linear complexity $L((s_n), N)$, see [67, Theorem 3].

Theorem 6.4 *Let (s_n) be a sequence over $\mathbb{F}_q$ and let $G(x)$ be its generating function. For $N \geq 2$, assume, that*

$$G(x) \not\equiv 0 \bmod x^N.$$

Let $L_N = L((s_n), N)$ be the Nth linear complexity and let

$$\sum_{\ell=t_N}^{L_N} c_\ell s_{i+\ell} = 0, \quad 0 \leq i \leq N - L_N - 1,$$

be a shortest linear recurrence for the first N terms of (s_n), where $c_{L_N} = 1$ and $c_{t_N} \neq 0$. Then

$$E((s_n), N) \geq \begin{cases} L_N - t_N + 1 & \text{for } N > (L_N - t_N)(L_N - \min\{1, t_N - 1\}), \\ \left\lceil \frac{N}{L_N - \min\{1, t_N - 1\}} \right\rceil & \text{otherwise,} \end{cases}$$

and

$$E((s_n), N) \leq \min\{L_N + \max\{-1, -t_N + 1\}, N - L_N + 2\}.$$

This result states in a qualitative way that *very* large Nth linear complexity, that is Nth linear complexity close to N, is a non-randomness property. Moreover, it enables us to estimate the Nth expansion complexity from below if the Nth linear complexity is not too close to either 0 or N (in a logarithmic scale), say, of order of magnitude $\sqrt{N}$.

We refer to [47, 67] for applications of Theorem 6.4 for estimating the Nth expansion complexity of certain sequences.

Most subsequences of automatic sequences, say, the Thue-Morse and Rudin-Shapiro sequences along squares, are not automatic, see Section 7 below, and thus have unbounded expansion complexity profile. However, their growth rates are not known. For example, one can study further the Thue-Morse and Rudin-Shapiro sequence along squares.

Problem 4 Estimate the expansion complexity profiles of the subsequences (t_{n^2}) and (r_{n^2}) of the Thue-Morse and Rudin-Shapiro sequence along squares.

Figure 14 suggests $E((t_{n^2}), N)$ and $E((r_{n^2}), N)$ are both of order of magnitude $\sqrt{N}$.

Finally, we remark that in order to use the full strength of Theorem 6.2 for inferring sequence elements, one needs to require the irreducibly of the polynomial $h(x, y)$ in (6.1). In [42, 43], the authors studied this variant of the Nth expansion complexity and the relation between these two complexity measures.

7 Subword complexity and normality

The results of Section 5 show that many automatic sequences, including Thue-Morse and Rudin-Shapiro sequence, are balanced, that is, the frequencies of the symbols are close to the expected values. However, the frequencies of longer patterns are far from uniform and most patterns do not appear at all. This phenomenon can be made precise by the notion of *subword complexity*.

Definition 7.1 For a sequence (s_n) over the alphabet Δ the *subword complexity* $p((s_n), k)$ is the number of distinct subsequences of length k.

Trivially we have $1 \leq p((s_n), k) \leq |\Delta|^k$ and for ultimately periodic sequences we have $p((s_n), k) = O(1)$.

By [8, Corollary 10.3.2] the subword complexity $p((s_n), k)$ of automatic sequences (s_n) is of order of magnitude k.

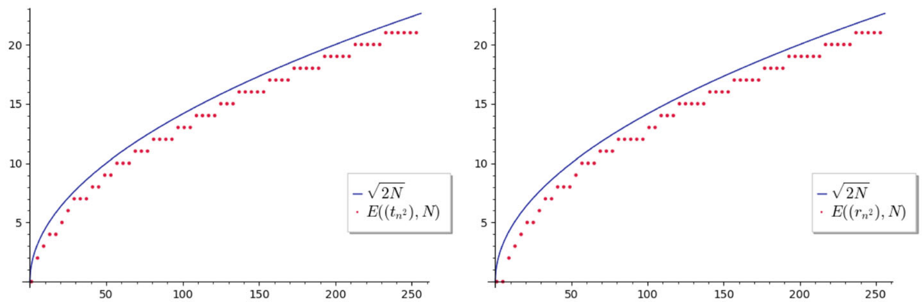

Fig. 14 The Nth expansion complexity of the Thue-Morse (left) and Rudin-Shapiro (right) sequence along squares

Theorem 7.2 *If (s_n) is an automatic sequence that is not ultimately periodic, then we have* [8]

$$p((s_n), k) = \Theta(k).$$

For the Thue-Morse sequence (t_n), the exact value of its subword complexity $p((t_n), k)$ was independently determined by Brlek [14, Proposition 4.4] and by de Luca and Varricchio [25, Proposition 4.4], see also [8, Exercise 10.11.10]. De Luca and Varricchio [25, Property 3.3] also showed that patterns such as 000 and 111 do not appear in the Thue-Morse sequence and more general the following result.

Theorem 7.3 *The Thue-Morse sequence is cube-free, that is, no pattern of the form www with $w \in \{0, 1\}^k$ for some $k \geq 1$ appears in the sequence.*

The papers [14, 25] contain also several other results on the non-existence of certain patterns in the Thue-Morse sequence.

The subword complexity and the correlation measure of order ℓ are related by the following result of Cassaigne et al. [16, Theorem 6].

Theorem 7.4 *If for some positive integers k and N*

$$C_\ell((s_n), N) \leq \frac{N}{2^{2k+1}}, \quad \ell = 1, 2, \ldots, k,$$

then

$$p((s_n), k) = 2^k.$$

For automatic sequences we can have $p((s_n), k) = 2^k$ only for finitely many k since $p((s_n), k) = \Theta(k)$. However, certain subsequences of automatic squences are *normal*, that is, all patterns appear in the sequence with the expected frequencies. More formally, a sequence (s_n) is called *normal* if for any fixed length k and any pattern $\mathbf{e} \in \Delta^k$

$$N_k((s_n), \mathbf{e}, N) = \frac{\#\{0 \leq n < N : (s_n, s_{n+1}, \ldots, s_{n+k-1}) = \mathbf{e}\}}{N} \to \frac{1}{|\Delta|^k} \quad \text{as } N \to \infty.$$

Drmota et al. [33] and Müllner [73] proved the normality of the Thue-Morse and the Rudin-Shapiro sequences along squares, that is

$$\lim_{N\to\infty} N_k((t_{n^2}), \mathbf{e}, N) = 2^{-k} \quad \text{and} \quad \lim_{N\to\infty} N_k((r_{n^2}), \mathbf{e}, N) = 2^{-k} \tag{7.1}$$

for any $\mathbf{e} \in \{0, 1\}^k$. The main tool to obtain the results (7.1) is to prove estimates on the sums

$$\sum_{n<N} (-1)^{e_0 t_{n^2} + \cdots + e_{k-1} t_{(n+k-1)^2}} \quad \text{and} \quad \sum_{n<N} (-1)^{e_0 r_{n^2} + \cdots + e_{k-1} r_{(n+k-1)^2}}$$

for any $e_0, \ldots, e_{k-1} \in \{0, 1\}$. These sums can be estimated via a Fourier analytic method of Mauduit and Rivat which has its origin in [60, 61]. For more details we refer to the survey [32] of Drmota and the original papers [33, 73].

In particular, the normality results (7.1) yield the subword complexities

$$p((t_{n^2}), k) = p((r_{n^2}), k) = 2^k. \tag{7.2}$$

[8] $f(k) = \Theta(g(k))$ is equivalent to $c_1 g(k) \leq f(k) \leq c_2 g(k)$ for some constants $c_2 \geq c_1 > 0$.

It is conjectured but not proved yet that the subsequences of the Thue-Morse sequence $(t_{f(n)})$ and Rudin-Shapiro sequence $(r_{f(n)})$ along any polynomial f of degree $d \geq 3$ are normal, see [33, Conjecture 1]. Even the weaker problem of determining the frequency of 0 and 1 in the subsequences $(t_{f(n)})$ and $(r_{f(n)})$ along any polynomial $f(x)$ of degree $d \geq 3$ with $f(\mathbb{N}_0) \subset \mathbb{N}_0$ seems to be very intricate, see [33, above Conjecture 1].

Problem 5 Show that the subsequences of Thue-Morse and Rudin-Shapiro sequence along cubes, bi-squares, ..., any polynomial values for a polynomial of degree at least 3 are normal.

However, Moshe [71] proved the following lower bound on the subword complexity of $(t_{f(n)})$,

$$p((t_{f(n)}), k) \geq 2^{k/2^{d-2}}. \tag{7.3}$$

Stoll [91, 93] showed that the number of zeros (resp. ones) among the first N sequence elements of both, $(t_{f(n)})$ and $(r_{f(n)})$, is at least of order of magnitude $N^{4/(3d+1)}$, $d \geq 3$. For subsequences $(z_{f(n)})$ of the Zeckendorf sum of digits sequence (z_n) defined by (4.5) the numbers of zeros and ones among the first N sequence elements are both lower bounded by $N^{4/(6d+1)}$, see Stoll [92].

Müllner and Spiegelhofer [74, 89] addressed the normality problem for the Thue-Morse sequence along the Piateski-Shapiro sequence $\lfloor n^c \rfloor$ for $1 < c < 3/2$. Moreover, it is asymptotically balanced (or simply normal)[90, Theorem 1.2] for $1 < c < 2$. For results on the Thue-Morse and Rudin-Shapiro sequence along primes see [12, 13, 61, 62] and references therein. In particular, the Thue-Morse sequence (t_p) along primes is balanced, see Mauduit and Rivat [61]. However, it is not known whether $(t_{f(p)})_p$ is normal for any non-constant polynomial f. For very recent results on the Zeckendorf-sum-of-digits sequence along primes see the masterpiece of Drmota, Müllner and Spiegelhofer [34].

From Theorem 7.2 and (7.2) we know that (t_{n^2}) and (r_{n^2}) are not automatic and by Theorem 2.3 these subsequences are, in contrast to the original sequence, not of bounded expansion complexity, that is,

$$\lim_{N\to\infty} E((t_{n^2}), N) = \lim_{N\to\infty} E((r_{n^2}), N) = \infty.$$

Theorem 7.2 combined with (7.3) implies that $(t_{f(n)})$ is not automatic and

$$\lim_{N\to\infty} E((t_{f(n)}), N) = \infty$$

for any polynomial of degree at least 2 with $f(\mathbb{N}_0) \subset \mathbb{N}_0$. Note that it was shown in [2] that $(t_{f(n)})$ is not 2-automatic and in [6] that $(r_{f(n)})$ is not 2-automatic and thus we also have

$$\lim_{N\to\infty} E((r_{f(n)}), N) = \infty.$$

Subsequences of the Thue-Morse sequence along geometric sequences such as (t_{3^n}) seem to be even more difficult to analyze. For example, Lagarias [54, Conjecture 1.12] conjectured that each pattern appears at least once in (t_{3^n}). For other related results see[35, 52].

For more details on the normality of automatic sequences and their subsequences we refer to [32].

8 Analogs for finite fields

An analog for finite fields of the problem on the distribution of automatic sequences and their subsequences was introduced by Dartyge and Sárközy [23]. It has been further investigated in [22, 57, 59, 97, 98], see also [21, 27, 39, 80, 99].

In the finite field setting some problems can be solved, although the analog for integers seems to be out of reach, including the normality problem for the analog of the Thue-Morse sequence and the frequency problem for the analog of the Rudin-Shapiro sequence both along polynomials. Hence, these analogs for finite fields are further attractive sources of pseudorandomness.

For a prime p and $q = p^r$ with $r \geq 2$ let $(\beta_1, \dots, \beta_r)$ be an ordered basis of $\mathbb{F}_q$ over $\mathbb{F}_p$. Then one can write all elements $\xi \in \mathbb{F}_q$ as

$$\xi = \sum_{i=1}^{r} x_i \beta_i, \quad x_1, \dots, x_r \in \mathbb{F}_p. \tag{8.1}$$

It is natural to consider the coefficients $x_1, \dots, x_r$ as digits with respect to the basis $(\beta_1, \dots, \beta_r)$. Then, in analogy to the Thue-Morse and Rudin-Shapiro sequence satisfying (2.5) we define the *Thue-Morse function*

$$T\left(\sum_{i=1}^{r} x_i \beta_i\right) = \sum_{i=1}^{r} x_i, \quad x_1, \dots, x_r \in \mathbb{F}_p,$$

and *Rudin-Shapiro function*

$$R\left(\sum_{i=1}^{r} x_i \beta_i\right) = \sum_{i=1}^{r-1} x_i x_{i+1}, \quad x_1, \dots, x_r \in \mathbb{F}_p,$$

on $\mathbb{F}_q$.

Dartyge and Sárközy [23] studied the balance of the Thue-Morse function along polynomial values. They derived results using the Weil bound [56, Theorem 5.38] on additive character sums:

Lemma 8.1 *Let $f \in \mathbb{F}_q[x]$ be of degree $d \geq 1$ with $\gcd(d, q) = 1$ and ψ be a nontrivial additive character of $\mathbb{F}_q$. Then*

$$\left| \sum_{\xi \in \mathbb{F}_q} \psi(f(\xi)) \right| \leq (d-1)\sqrt{q}.$$

Put

$$e(\alpha) = \exp(2\pi i \alpha), \quad \alpha \in \mathbb{R},$$

and note that $\psi(x) = e(T(x)/p)$ is a nontrivial additive character of $\mathbb{F}_q$. Then from

$$\sum_{h=0}^{p-1} e\left(\frac{ha}{p}\right) = \begin{cases} 0, & a \neq 0, \\ p, & a = 0, \end{cases} \quad a \in \mathbb{F}_p,$$

we get

$$\#\{\xi \in \mathbb{F}_q : T(f(\xi)) = c\} = \frac{1}{p} \sum_{h=0}^{p-1} \sum_{\xi \in \mathbb{F}_q} \psi(hf(\xi))\, e\left(\frac{-hc}{p}\right).$$

The contribution of $h = 0$ is trivially p^{r-1}, which is the expected number of solutions. The other terms for $h \neq 0$ contribute to the error term and can be bounded by Lemma 8.1. We immediately get [23, Theorem 1.2]:

Theorem 8.2 *Let $f \in \mathbb{F}_q[x]$ be of degree d with $\gcd(d, q) = 1$. Then for all $c \in \mathbb{F}_p$, we have*

$$\left|\#\{\xi \in \mathbb{F}_q : T(f(\xi)) = c\} - p^{r-1}\right| \leq (d-1)p^{r/2}.$$

Later Dartyge, Mérai and Winterhof [22] investigated this problem for the Rudin-Shapiro function. The main difference between the two problems is that the Rudin-Shapiro function is not a linear map in contradistinction to the Thue-Morse function. Standard character sum techniques fail in this situation. Namely, consider $R(f(\xi))$ with ξ having the form (8.1) as a polynomial in the r variables $x_1, \ldots, x_r$. Then using Lemma 8.1 for one coordinate x_i one gets an error term larger than the main term. Stronger results in higher dimension such as the Deligne bound [24, Théorème 8.4] also cannot be applied as it needs some more technically intricate conditions which are not satisfied in our situation. However, sacrificing the explicit dependence of the degree d, one can use an affine version of the Hooley-Katz Theorem, see [48] or [72, Theorem 7.1.14].

First recall that the *(affine) singular locus* $\mathcal{L}(F)$ of a polynomial $F \in \mathbb{F}_p[x_1, \ldots, x_r]$ is the set of common zeros in $\overline{\mathbb{F}_p}^r$ of the polynomials[9]

$$F, \frac{\partial F}{\partial x_1}, \ldots, \frac{\partial F}{\partial x_r}.$$

We also recall that the dimension of $\mathcal{L}(F)$ is the largest d for which there exist $1 \leq i_1 < i_2 < \ldots < i_d \leq r$ such that there is no nonzero polynomial P in d variables with $P(y_{i_1}, \ldots, y_{i_d}) = 0$ for all $(y_1, \ldots, y_r) \in \mathcal{L}(F)$, see [20, Corollary 9.5.4].

Lemma 8.3 *Let $Q \in \mathbb{F}_p[x_1, \ldots, x_r]$ be of degree $d \geq 1$ such that the dimensions of the singular loci of Q and its homogeneous part Q_d of degree d satisfy*

$$\max\{\dim(\mathcal{L}(Q)), \dim(\mathcal{L}(Q_d)) - 1\} \leq s.$$

Then the number N of zeros of Q in $\mathbb{F}_p^r$ satisfies

$$\left|N - p^{r-1}\right| \leq C_{d,r}\, p^{(r+s)/2},$$

where $C_{d,r}$ is a constant depending only on d and r.

Then using Lemma 8.3, one can show that the Rudin-Shapiro function is also asymptotically balanced on polynomial values, see [22, Theorem 1].

Theorem 8.4 *Let $f \in \mathbb{F}_q[x]$ be of degree d with $\gcd(d, q) = 1$. Then for all $c \in \mathbb{F}_p$, we have*

$$\left|\#\{\xi \in \mathbb{F}_q : R(f(\xi)) = c\} - p^{r-1}\right| \leq C_{d,r}\, p^{(3r+1)/4},$$

where the constant $C_{d,r}$ depends only on the degree d of f and r.

Theorem 8.4 is nontrivial if r is fixed and $p \to \infty$. Contrary to Theorem 8.2, nothing is known for the dual situation.

[9] $\overline{\mathbb{F}_p} = \bigcup_{n=1}^{\infty} \mathbb{F}_{p^n}$ denotes the algebraic closure of $\mathbb{F}_p$.

Problem 6 For fixed prime p show that if r is large enough, then the Rudin-Shapiro function along polynomial values is balanced possibly under some natural restrictions on the polynomial.

Analogously to the normality results of Section 7, Makhul and Winterhof [57] obtained results on the normality of the Thue-Morse function along polynomial values. For sake of simplicity we state the case when the polynomial f has degree d smaller than the characteristic p, [57, Corollary 1].

Theorem 8.5 *Assume $1 \le d < p$ and $s \le d$. For any polynomial $f \in \mathbb{F}_q[x]$ of degree d and any pairwise distinct $\alpha_1, \dots, \alpha_s \in \mathbb{F}_q$ and any $c_1, \dots, c_s \in \mathbb{F}_p$ we have*

$$\left|\#\{\xi \in \mathbb{F}_q : T(f(\xi + \alpha_i)) = c_i, 1 \le i \le s\} - p^{r-s}\right| \le (d-1)p^{r/2}.$$

Note that the restriction $s \le d$ is natural and counterexamples for $s > d$ are easy to construct.

The case of the Rudin-Shapiro function is much more intricate.

Problem 7 Study the normality of the Rudin-Shapiro function at $f(x)$. Namely, show that

$$\frac{\#\{\xi \in \mathbb{F}_q : R(f(\xi + \alpha_i)) = c_i, 1 \le i \le s\}}{p^{r-s}} \to 1 \quad \text{as } p \to \infty$$

for some $s \ge 2$ and any $f \in \mathbb{F}_q[x]$ of fixed degree.

Of course, this problem is also open for fixed p and $r \to \infty$ even in the simplest case $s = 1$, see Problem 6.

It is natural to define the *Rudin-Shapiro function* on the polynomial ring $\mathbb{F}_p[t]$ by assigning the coefficients of the polynomial $f(t) \in \mathbb{F}_p[t]$ to $(x_1, \dots, x_r)$, that is,

$$R(t^r + x_1 t^{r-1} + \cdots + x_r) = \sum_{i=1}^{r-1} x_i x_{i+1},$$

for $x_1, \dots, x_r \in \mathbb{F}_p$.

Analogously to the result of Mauduit and Rivat [62] on the Rudin-Shapiro sequence along prime numbers, it is natural to investigate the balance and the normality of the Rudin-Shapiro function along irreducible polynomials. As the number of monic irreducible polynomials of degree r is $p^r/r + o(p^r)$, see for example [56, Theorem 3.25], we expect that the frequency of each element c is $\frac{p^{r-1}}{r} + o(p^{r-1})$. For $r = 2$ and fixed $c \in \mathbb{F}_p$ we have to count the number of $x_2 \in \mathbb{F}_p^*$ such that $t^2 + x_2^{-1}ct + x_2$ is irreducible over $\mathbb{F}_p$ or equivalently the discriminant $x_2^{-2}c^2 - 4x_2$ is a quadratic non-residue modulo p. This number is

$$\frac{1}{2} \sum_{\substack{x_2 \in \mathbb{F}_p \\ 4x_2^3 \ne c^2}} \left(1 - \left(\frac{c^2 - 4x_2^3}{p}\right)\right) = \begin{cases} \frac{p-1}{2}, & c = 0, \\ \frac{p-1}{2} + O(p^{1/2}), & c \ne 0, \end{cases}$$

by the Weil bound for multiplicative character sums [56, Theorem 5.41], where $\left(\frac{\cdot}{\cdot}\right)$ is the Legendre symbol.

Problem 8 Prove that for all $c \in \mathbb{F}_p$ and $r \geq 3$ we have

$$\lim_{p\to\infty} \frac{\#\{f \in \mathbb{F}_p[t] : \deg f = r, f \text{ monic and irreducible over} \mathbb{F}_p, R(f) = c\}}{p^{r-1}} = \frac{1}{r}.$$

We remark that one can define the *Thue-Morse function* by

$$T(f) = T(t^r + x_1 t^{r-1} + \ldots + x_r) = x_1 + \ldots + x_r = f(1) - 1.$$

Note that for irreducible polynomials $f(x)$ we have $T(f) \neq -1$ and for $c \neq -1$ the number of monic irreducible polynomials of degree $r = 2$ with $T(f) = c$ is

$$\frac{1}{2} \sum_{\substack{u\in\mathbb{F}_p \\ u^2\neq c+1}} \left(1 - \left(\frac{u^2 - c - 1}{p}\right)\right) = \frac{p - \left(\frac{c+1}{p}\right)}{2},$$

where we used a well-known result on sums of Legendre symbols of quadratic polynomials, see for example [56, Theorem 5.48]. In general, since $f(x)$ is irreducible whenever $f(x-1)$ is irreducible we have to estimate the number I_c of monic irreducible polynomials with fixed constant term $c \neq 0$ which satisfies

$$\frac{1}{r}\left(\frac{p^r - 1}{p - 1} - 2p^{r/2}\right) \leq I_c \leq \frac{p^r - 1}{r(p-1)},$$

see [15] or [72, Theorem 3.5.9], and we get the desired

$$I_c = \frac{p^{r-1}}{r} + o(p^{r-1})$$

for $r \geq 3$ as well.

Moreover, the corresponding normality problem is trivial since for any polynomial $g(x)$ of degree at most $r - 1$ the value $T(f + g) = f(1) + g(1) - 1$ is uniquely defined by $T(f) = f(1) - 1$ and $g(1)$.

For other results on 'digits' along irreducible polynomials see for example [72, Chapter 3] and [40, 44, 46, 81, 83, 101].

Acknowledgment The authors were supported by the Austrian Science Fund FWF grants P 30405 and P 31762. They wish to thank Jean-Paul Allouche, Harald Niederreiter, Igor Shparlinski, Cathy Swaenepoel, Thomas Stoll and Steven Wang for very useful discussions as well as the anonymous referees and the associate editor Daniel Katz for their concise and very helpful reports.

References

1. Aistleitner, C.: On the limit distribution of the well-distribution measure of random binary sequences. J. Thé,or. Nombres Bordeaux **25**(2), 245–259 (2013)
2. Allouche, J.-P.: Somme des chiffres et transcendance. Bull. Soc. Math France 110 **3**, 279–285 (1982)
3. Allouche, J.-P.: On a Golay-Shapiro-like sequence. Unif. Distrib. Theory **11**(2), 205–210 (2016)
4. Allouche, J.-P., Han, G.-N., Niederreiter, H.: Perfect linear complexity profile and apwenian sequences. Finite Fields Appl. **68**(101761), 13 (2020)
5. Allouche, J.-P., Peyriére, J., Wen, Z.-X., Wen, Z.-Y.: Hankel determinants of the Thue-Morse sequence. Ann. Inst. Fourier (Grenoble) **48**(1), 1–27 (1998)
6. Allouche, J.-P., Salon, O.: Sous-suites polynomiales de certaines suites automatiques. J. Thé,or. Nombres Bordeaux **5**(1), 111–121 (1993)
7. Allouche, J.-P., Shallit, J.: The Ubiquitous Prouhet-Thue-Morse Sequence. Sequences and Their Applications (Singapore, 1998), 1–16, Springer Ser. Discrete Math. Theor. Comput. Sci. Springer, London (1999)

8. Allouche, J.-P., Shallit, J.: Automatic Sequences. Theory, Applications, Generalizations. Cambridge University Press, Cambridge (2003)
9. Allouche, J.-P., Shallit, J., Yassawi, R.: How to prove that a sequence is not automatic. To appear in Expositiones Mathematicae, online available https://doi.org/10.1016/j.exmath.2021.08.001 (2021)
10. Alon, N., Kohayakawa, Y., Mauduit, C., Moreira, C.G., Rödl, V.: Measures of pseudorandomness for finite sequences: typical values. Proc. Lond. Math. Soc. (3) **95**(3), 778–812 (2007)
11. Blumer, A., Blumer, J., Ehrenfeucht, A., Haussler, D., McConnell, R.: Linear size finite automata for the set of all subwords of a word: an outline of results. Bul. Eur. Assoc. Theor. Comp. Sci. **21**, 12–20 (1983)
12. Bourgain, J.: Prescribing the binary digits of primes. Israel J. Math. **194**(2), 935–955 (2013)
13. Bourgain, J.: Prescribing the binary digits of primes, II. Israel J. Math. **206**(1), 165–182 (2015)
14. Brlek, S.: Enumeration of factors in the Thue-Morse word. First Montreal Conference on Combinatorics and Computer Science, 1987. Discrete Appl Math. **24**(1-3), 83–96 (1989)
15. Car, M.: Distribution des polynômes irréductibles dans $f_q[t]$. Acta Arith. **88**(2), 141–153 (1999)
16. Cassaigne, J., Ferenczi, S., Mauduit, C., Rivat, J., Sárközy, A.: On finite pseudorandom binary sequences. III. The Liouville function. I. Acta Arith. **87**(4), 367–390 (1999)
17. Chan, L., Grimm, U.: Spectrum of a Rudin-Shapiro-like sequence. Adv. in Appl. Math. **87**, 16–23 (2017)
18. Christol, G., presque, E.nsembles.: Périodiques k-reconnaissables. Theoret. Comput. Sci. **9**(1), 141–145 (1979)
19. Christol, G., Kamae, T., Mendès France, M., Rauzy, G.: Suites algébriques, automates et substitutions. Bull. Soc. Math. France **108**(4), 401–419 (1980)
20. Cox, D.A., Little, J., O'Shea, D.: Ideals, varieties and Algorithms. Undergraduate Texts in Mathematics. Springer, Cham, fourth edition. An introduction to computational algebraic geometry and commutative algebra (2015)
21. Dartyge, C., Mauduit, C., Sárközy, A.: Polynomial values and generators with missing digits in finite fields. Funct. Approx Comment. Math. **52**(1), 65–74 (2015)
22. Dartyge, C., Mérai, L., Winterhof, A.: On the distribution of the Rudin-Shapiro function for finite fields. Proc. Amer. Math. Soc. **149**(12), 5013–5023 (2021)
23. Dartyge, C., Sárközy, A.: The sum of digits function in finite fields. Proc. Amer. Math. Soc. **141**(12), 4119–4124 (2013)
24. Deligne, P.: La conjecture de Weil. Inst. Hautes Études Sci. Publ. Math. **43**, 273–307 (1974)
25. de Luca, A., Varricchio, S.: Some combinatorial properties of the Thue-Morse sequence and a problem in semigroups. Theoret Comput. Sci. **63**(3), 333–348 (1989)
26. Diem, C.: On the use of expansion series for stream ciphers. LMS. J. Comput. Math. **15**, 326–340 (2012)
27. Dietmann, R., Elsholtz, C., Shparlinski, I.E.: Prescribing the binary digits of squarefree numbers and quadratic residues. Trans. Amer. Math. Soc. **369**(12), 8369–8388 (2017)
28. Dorfer, G.: Lattice profile and linear complexity profile of pseudorandom number sequences. Finite fields and applications, 69–78, Lecture Notes in Comput. Sci Springer, Berlin (2948)
29. Dorfer, G., Meidl, W., Winterhof, A.: Counting functions and expected values for the lattice profile at n. Finite Fields Appl. **10**(4), 636–652 (2004)
30. Dorfer, G., Winterhof, A.: Lattice structure and linear complexity profile of nonlinear pseudorandom number generators. Appl. Algebra Engrg. Comm Comput. **13**(6), 499–508 (2003)
31. Dorfer, G., Winterhof, A.: Lattice Structure of Nonlinear Pseudorandom Number Generators in Parts of the Period. Monte Carlo and quasi-Monte-Methods 2002, 199–211. Springer, Berlin (2004)
32. Drmota, M.: Subsequences of Automatic Sequences and Uniform Distribution. Uniform Distribution and quasi-Monte Carlo Methods, 87–104, Radon Ser. Comput. Appl Math., vol. 15. De Gruyter, Berlin (2014)
33. Drmota, M., Mauduit, C., Rivat, J.: Normality along squares. J. Eur. Math. Soc. **21**(2), 507–548 (2019)
34. Drmota, M., Müllner, C., Spiegelhofer, L.: Primes as sums of Fibonacci numbers. Preprint 2021.2109.04068 (2021)
35. Dupuy, T., Weirich, D.E.: Bits of 3^n in binary, Wieferich primes and a conjecture of Erdös. J. Number Theory **158**, 268–280 (2016)
36. Everest, G., van der Poorten, A., Shparlinski, I., Ward, T.: Recurrence Sequences Mathematical Surveys and Monographs, 104. American Mathematical Society, Providence, RI (2003)
37. Fogg, N.P.: Substitutions in Dynamics, Arithmetics and Combinatorics. Edited by V. BerthÉ, S. Ferenczi, C. Mauduit and A. Siegel Lecture Notes in Mathematics, vol. 1794. Springer-Verlag, Berlin (2002)
38. Fouvry, E., Mauduit, C.: Sommes des chiffres et nombres presque premiers. Math. Ann. **305**, 571–599 (1996)
39. Gabdullin, M.R.: On the squares in the set of elements of a finite field with constraints on the coefficients of its basis expansion. (Russian) Mat. Zametki 100 (2016), no. 6, 807–824; translation in Math. Notes 101, no. 1–2, 234–249 (2017)

40. Gao, Z., Kuttner, S., Wang, Q.: On enumeration of irreducible polynomials and related objects over a finite field with respect to their trace and norm. Finite Fields Appl. **69**(101770), 25 (2021)
41. Gel'fond, A.O.: Sur les nombres qui ont des propriétés additives et multiplicatives données. (French) Acta Arith. 13 , 259–265 (1967)
42. Gómez-Pérez, D., Mérai, L.: Algebraic dependence in generating functions and expansion complexity. Adv. Math. Commun. **14**(2), 307–318 (2020)
43. Gómez-Pérez, D., Mérai, L., Niederreiter, H.: On the expansion complexity of sequences over finite fields. IEEE Trans. Inform. Theory **64**(6), 4228–4232 (2018)
44. Granger, R.: On the enumeration of irreducible polynomials over $GF(q)$ with prescribed coefficients. Finite Fields Appl. **57**, 156–229 (2019)
45. Gyarmati, K.: Measures of Pseudorandomness. Finite Fields and Their Applications. 43–64, Radon Ser. Comput. Appl Math., vol. 11. De Gruyter, Berlin (2013)
46. Ha, J.: Irreducible polynomials with several prescribed coefficients. Finite Fields Appl. **40**, 10–25 (2016)
47. Hofer, R., Winterhof, A., complexity, L.inear.: Expansion Complexity of Some Number Theoretic Sequences. Arithmetic of Finite Fields, 67–74 Lecture Notes in Comput Sci., vol. 10064. Springer, Cham (2016)
48. Hooley, C.: On the number of points on a complete intersection over a finite field. With an appendix by Nicholas M. Katz. J. Number Theory **38**(3), 338–358 (1991)
49. Jamet, D., Popoli, P., Stoll, T.: Maximum order complexity of the sum of digits function in Zeckendorf base and polynomial subsequences. Cryptogr. Commun. **13**(5), 791–814 (2021)
50. Jansen, C.J.A.: Investigations on nonlinear streamcipher systems: Construction and evaluation methods. Thesis (Dr.)–Technische Universiteit Delft (The Netherlands). Proquest LLC, Ann Arbor, MI, p 195 (1989)
51. Jansen, C.J.A., Boekee, D.E.: The Shortest Feedback Shift Register that Can Generate a Given Sequence. Advances in Cryptology—CRYPTO '89 (Santa Barbara, CA, 1989), 90–99, Lecture Notes in Comput Sci., vol. 435. Springer, New York (1990)
52. Kaneko, H., Stoll, T.: On subwords in the base-q expansion of polynomial and exponential functions. Integers 18A Paper **A11**, 11 (2018)
53. Lafrance, P., Rampersad, N., Yee, R.: Some properties of a Rudin-Shapiro-like sequence. Adv. in Appl. Math. **63**, 19–40 (2015)
54. Lagarias, J.: Ternary expansions of powers of 2. J. Lond. Math. Soc. (2) **79**(3), 562–588 (2009)
55. L'Ecuyer, P., Simard, R.: Testu01: A C Library for Empirical Testing of Random Number Generators. ACM Transactions on Mathematical Software. Vol. 33, article 22 (2007)
56. Lidl, R., Niederreiter, H.: Finite Fields. Second Edition Encyclopedia of Mathematics and Its Applications, vol. 20. Cambridge University Press, Cambridge (1997)
57. Makhul, M., Winterhof, A.: Normality of the Thue-Morse function for finite fields along polynomial values. Preprint 2021. 2106.12218 (2021)
58. Marcovici, I., Stoll, T., Tahay, P.-A.: Discrete correlations of order 2 of generalized Golay-Shapiro sequences: A combinatorial approach. Integers 21, Paper No. A45, 21 pp (2021)
59. Mattheus, S.: Trace of products in finite fields from a combinatorial point of view. SIAM J. Discrete Math. **33**(4), 2126–2139 (2019)
60. Mauduit, C., Rivat, J.: La somme des chiffres des carrés. Acta Math. **203**(1), 107–148 (2009)
61. Mauduit, C., Rivat, J.: Sur un probléme de gelfond: la somme des chiffres des nombres premiers. Ann. of Math. (2) **171**(3), 1591–1646 (2010)
62. Mauduit, C., Rivat, J.: Prime numbers along Rudin-Shapiro sequences. J. Eur. Math. Soc. (JEMS) **17**(10), 2595–2642 (2015)
63. Mauduit, C., Rivat, J.: Rudin-shapiro sequences along squares. Trans. Amer. Math. Soc. **370**(11), 7899–7921 (2018)
64. Mauduit, C., Sárközy, A.: On finite pseudorandom binary sequences. I. Measure of pseudorandomness, the Legendre symbol. Acta Arith. **82**(4), 365–377 (1997)
65. Mauduit, C., Sárközy, A.: On finite pseudorandom binary sequences. II. The Champernowne, Rudin-Shapiro, and Thue-Morse sequences, a further construction. J. Number Theory **73**(2), 256–276 (1998)
66. Mauduit, C., Sárközy, A.: On the measures of pseudorandomness of binary sequences. Discrete Math. 271, no. 1–3, 195–207 (2003)
67. Mérai, L., Niederreiter, H., Winterhof, A.: Expansion complexity and linear complexity of sequences over finite fields Cryptogr. Commun. **9**(4), 501–509 (2017)
68. Mérai, L., Rivat, J., Sárközy, A.: The Measures of Pseudorandomness and the NIST Tests. Number-theoretic Methods in Cryptology, 197–216, Lecture Notes in Comput Sci., 10737, Springer, Cham (2018)
69. Mérai, L., Winterhof, A.: On the pseudorandomness of automatic sequences. Cryptogr. Commun. **10**(6), 1013–1022 (2018)

70. Mérai, L., Winterhof, A.: On the Nth linear complexity of automatic sequences. J. Number Theory **187**, 415–429 (2018)
71. Moshe, Y.: On the subword complexity of Thue-Morse polynomial extractions. Theoret. Comput. Sci. **389**(1–2), 318–329 (2007)
72. Mullen, G.L., Panario, D. (eds.): Handbook of Finite Fields. Discrete Mathematics and Its Applications (Boca Raton). CRC Press, Boca Raton, FL (2013)
73. Müllner, C.: The Rudin-Shapiro sequence and similar sequences are normal along squares. Canad. J. Math. **70**(5), 1096–1129 (2018)
74. Müllner, C., Spiegelhofer, L.: Normality of the Thue-Morse sequence along Piatetski-Shapiro sequences, II. Israel. J. Math. **220**(2), 691–738 (2017)
75. Niederreiter, H.: Sequences with Almost Perfect Linear Complexity Profile. Advances in Cryptology-EUROCRYPT '87 (D. Chaum and W. L. Price, Eds.), Lecture Notes in Computer Science, Vol. 304, Pp. 37–51. Springer-Verlag, Berlin/Heidelberg/New York (1988)
76. Niederreiter, H.: The Probabilistic Theory of Linear Complexity. Advances in Cryptology – EUROCRYPT '88 (C. G. Günther, Ed.) Lecture Notes in Computer Science, Vol. 330, Pp. 191–209. Springer, Berlin (1988)
77. Niederreiter, H.: Linear complexity and related complexity measures for sequences. Progress in Cryptology—INDOCRYPT 2003, 1–17, Lecture Notes in Comput. Sci. 2904, Springer, Berlin (2003)
78. Niederreiter, H., Xing, C.: Sequences with high nonlinear complexity. IEEE Trans. Inform. Theory **60**(10), 6696–6701 (2014)
79. Niederreiter, H., Winterhof, A.: Applied number theory. Springer Cham (2015)
80. Ostafe, A.: Polynomial values in affine subspaces of finite fields. J. Anal. Math. **138**(1), 49–81 (2019)
81. Pollack, P.: Irreducible polynomials with several prescribed coefficients. Finite Fields Appl. **22**, 70–78 (2013)
82. Popoli, P.: On the maximum order complexity of Thue-Morse and Rudin-Shapiro sequences along polynomial values. Unif. Distrib. Theory **15**(2), 9–22 (2020)
83. Porritt, S.: Irreducible polynomials over a finite field with restricted coefficients. Canad. Math. Bull. **2**, 429–439 (2019)
84. Rudin, W.: Some theorems on Fourier coefficients. Proc. Amer. Math. Soc. **10**, 855–859 (1959)
85. Rukhin, A., et al.: NIST Special Publication 800-22, Revision 1.a, A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, https://www.nist.gov/publications/statistical-test-suite-random-and-pseudorandom-number-generators-cryptographic (2021)
86. Schmidt, K.-U.: The correlation measures of finite sequences: limiting distributions and minimum values. Trans. Amer. Math. Soc. **369**(1), 429–446 (2017)
87. Shapiro, H.S.: Extremal Problems for Polynomials and Power Series. Master's thesis, MIT (1952)
88. Shparlinski, I.: Cryptographic Applications of Analytic Number Theory. Complexity Lower Bounds and Pseudorandomness Progress in Computer Science and Applied Logic, vol. 22. Basel, Birkhäuser Verlag (2003)
89. Spiegelhofer, L.: Normality of the Thue-Morse sequence along Piatetski-Shapiro sequences. Q. J. Math. **66**(4), 1127–1138 (2015)
90. Spiegelhofer, L.: The level of distribution of the Thue-Morse sequence. Compos. Math. **156**(12), 2560–2587 (2020)
91. Stoll, T.: The sum of digits of polynomial values in arithmetic progressions. Funct. Approx. Comment. Math. **47**(2), 233–239 (2012)
92. Stoll, T.: Combinatorial constructions for the Zeckendorf sum of digits of polynomial values. Ramanujan J. **32**(2), 227–243 (2013)
93. Stoll, T.: On digital blocks of polynomial values and extractions in the Rudin-Shapiro sequence. RAIRO Theor. Inform. Appl. **50**(1), 93–99 (2016)
94. Sun, Z., Winterhof, A.: On the maximum order complexity of subsequences of the Thue-Morse and Rudin-Shapiro sequence along squares. Int. J. Comput. Math. Comput. Syst. Theory **4**(1), 30–36 (2019)
95. Sun, Z., Winterhof, A.: On the maximum order complexity of the Thue-Morse and Rudin-Shapiro sequence. Unif. Distrib. Theory **14**(2), 33–42 (2019)
96. Sun, Z., Zeng, X., Lin, D.: On the Nth maximum order complexity and the expansion complexity of a Rudin-Shapiro-like sequence. Cryptogr. Commun. **12**(3), 415–426 (2020)
97. Swaenepoel, C.: Trace of products in finite fields. Finite Fields Appl. **51**, 93–129 (2018)
98. Swaenepoel, C.: On the sum of digits of special sequences in finite fields. Monatsh. Math. **187**(4), 705–728 (2018)
99. Swaenepoel, C.: Prescribing digits in finite fields. J. Number Theory **189**, 97–114 (2018)
100. Topuzoğlu, A., Winterhof, A.: Pseudorandom sequences. Topics in geometry, coding theory and cryptography, 135–166, Algebr Appl., vol. 6. Springer, Dordrecht (2007)

101. Tuxanidy, A., Wang, Q.: Irreducible polynomials with prescribed sums of coefficients. Preprint 2016. 1605.00351 (2016)
102. Winterhof, A.: Linear Complexity and Related Complexity Measures. Selected Topics in Information and Coding Theory, 3–40, Ser Coding Theory Cryptol., vol. 7. World Sci. Publ., Hackensack, NJ (2010)
103. Winterhof, A.: Recent Results on Recursive Nonlinear Pseudorandom Number Generators (Invited Paper). Sequences and Their Applications-SETA 2010, 113–124, Lecture Notes in Comput Sci., vol. 6338. Springer, Berlin (2010)
104. Xing, C., Lam, K.: Sequences with almost perfect linear complexity profiles and curves over finite fields. IEEE Trans. Inform. Theory, pp 1267–1270 (1999)