Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 389 (2007) 318–329
Theoretical Computer Science
www.elsevier.com/locate/tcs

On the subword complexity of Thue–Morse polynomial extractions

Yossi Moshe*

CNRS, LRI, Université Paris Sud, Orsay, France
Einstein Institute, The Hebrew University, Jerusalem, Israel

Received 17 October 2006; received in revised form 26 September 2007; accepted 6 October 2007

Communicated by D. Perrin

Abstract

Let the (subword) complexity of a sequence $\mathbf{u} = (u_n)_{n=0}^{\infty}$ over a finite set Σ be the function $m \mapsto P_{\mathbf{u}}(m)$, where $P_{\mathbf{u}}(m)$ is the number of distinct blocks of length m in $\mathbf{u}$. Let $\mathbf{t} = (t_n)_{n=0}^{\infty}$ denote the Thue–Morse sequence. In this paper we study the complexity of the sequences $\mathbf{t}_H = (t_{H(n)})_{n=0}^{\infty}$, when $H(n) \in \mathbb{Q}[n]$ is a polynomial with $H(\mathbb{N}) \subseteq \mathbb{N}$. In particular, we solve an open problem of Allouche and Shallit regarding $(t_{n^2})_{n=0}^{\infty}$. We also study the vector space over $\mathbb{Z}/2\mathbb{Z}$, spanned by the sequences $\mathbf{t}_H$.

Keywords: Subword complexity; Automatic sequences; Thue–Morse sequence; Polynomial extractions

1. Introduction

Let s_n denote the number of 1's in the binary representation of an integer $n \geq 0$ and $t_n \in \mathbb{Z}/2\mathbb{Z}$ be the residue of s_n modulo 2. The (Prouhet)–Thue–Morse sequence, given by

$$\mathbf{t} = (t_n)_{n=0}^{\infty} = 0110100110010110100101100110100 1 \cdots,$$

was introduced by Thue [32,33] and Morse [24] and mentioned implicitly already in 1851 by Prouhet [28]. This sequence has applications in many different areas in mathematics, including differential geometry, number theory and mathematical physics (see [9] for a survey). Of particular interest is its role in the study of combinatorics on words.

One can easily see that every binary word of length ≥ 4 must contain a square (i.e., two consecutive identical blocks). Thue [32,33] provided the sequence $\mathbf{t} = (t_n)_{n=0}^{\infty}$ as an example of a cube-free infinite binary word (i.e., containing no three consecutive identical blocks). More generally, he proved that $\mathbf{t}$ is overlaps-free (i.e., containing no blocks of the form $awawa$, where $a \in \mathbb{Z}/2\mathbb{Z}$ and w is a binary block). He also proved that the sequence $(u_n)_{n=0}^{\infty} = 210201210 \cdots$, where u_n counts the number of 1's between the nth and the $(n+1)$st occurrences of

* Corresponding address: Einstein Institute of Mathematics, Edmond J. Safra Campus, Givat Ram, The Hebrew University of Jerusalem, Jerusalem, 91904, Israel. Tel.: +972 8 9431332.
E-mail address: mosheyoss@math.huji.ac.il.

0304-3975/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2007.10.015

0 in $\mathbf{t}$, is a square-free sequence over $\mathbb{Z}/3\mathbb{Z}$. This work of Thue (see also Berstel [12] and the references therein) is considered as the starting point of the study of combinatorics on words.

Let $\mathbf{u} = (u_n)_{n=0}^{\infty}$ be a sequence over a finite set Σ, and denote by $P_{\mathbf{u}}(m)$ the number of distinct blocks of length m in $\mathbf{u}$. The function $m \mapsto P_{\mathbf{u}}(m)$ is the *(subword) complexity* of $\mathbf{u}$ (see [2,13,19,20] for surveys on this function). The exact complexity of $\mathbf{t}$ was calculated independently by Brlek [14], de Luca and Varricchio [23] and Avgustinovich [10] (see [25,31] for generalizations). Since $\mathbf{t}$ is generated by the *morphism* given by $0 \mapsto 01$, $1 \mapsto 10$ (see Example 1), it is 2-*automatic* (see definition in Section 2), and thus its complexity is bounded by a linear function in m (cf. [17] and [29, Ch. V]).

In this paper we study the complexity of the sequences

$$\mathbf{t}_H = (t_{H(n)})_{n=0}^{\infty},$$

where H is a polynomial over $\mathbb{Q}$ with $H(\mathbb{N}) \subseteq \mathbb{N}$. Clearly, the complexity of $\mathbf{t}_H$ is $O(m)$ for linear polynomials $H(n) = an + b \in \mathbb{N}[n]$. This follows by the 2-automaticity of those sequences, or directly by the linear bound on the complexity of $\mathbf{t}$. (See [15] for other results on the blocks in $(t_{an+b})_{n=0}^{\infty}$.) The sequences $\mathbf{t}_H$ are, however, more complicated when $\deg H \geq 2$. The following theorem of Allouche and Salon [7] shows in particular that none of these sequences is 2-automatic.

Theorem A *([7]). Let $r \geq 0$ and $H, Q_1, \ldots, Q_r \in \mathbb{Q}[n]$ be polynomials with $H(\mathbb{N}) \subseteq \mathbb{N}$, $Q_i(\mathbb{N}) \subseteq \mathbb{N}$, $i = 1, \ldots, r$. Assume that $\deg Q_i < \deg H$ for each i and that $\deg H \geq 2$. Then the sequence $\mathbf{u} = \mathbf{t}_H + \sum_{i=1}^{r} \mathbf{t}_{Q_i} \in (\mathbb{Z}/2\mathbb{Z})^{\mathbb{N}}$ is not 2-automatic.*

In fact, Allouche and Salon [7] proved a generalization of Theorem A for the family of *quasistrongly q-additive sequences*, namely sequences $(v_n)_{n=0}^{\infty}$ over an abelian group, satisfying the condition

$$\forall r \in \mathbb{N}\ \exists k_0(r), \quad \forall k \geq k_0, \quad \forall n \in \mathbb{N}, \quad v_{q^k n + r} = v_n + v_r.$$

(See [22] for another generalization.)

In [8], Allouche and Shallit posed the following question,

Question A *([8, Open Problem 10.7])*. Is it true that the complexity of $(t_{n^2})_{n=0}^{\infty}$ is $P(m) = 2^m$?

It is interesting to note that even the much weaker property, namely the existence of arbitrarily long squares in $(t_{n^2})_{n=0}^{\infty}$ was unresolved [8, Open Problem 1.11].

In this paper, we study the sequences $(t_{H(n)})_{n=0}^{\infty}$ for non-linear polynomials H and obtain a lower bound for their complexities. This bound shows in particular that, if $\deg H = 2$, then $P_{\mathbf{t}_H}(m) = 2^m$, and thereby, provides a positive answer to Question A. We also consider the complexity of $\mathbf{t}_S = \sum_{H \in S} \mathbf{t}_H$, where $S \subseteq \mathbb{Q}[n]$ is a finite set of polynomials with $H(\mathbb{N}) \subseteq \mathbb{N}$. This enables us to provide a new generalization of Theorem A (see Corollary 5 for the precise formulation).

As a part of our proof we study the set of vectors

$$\mathcal{V}^{[0,N)} = \{(t_{cn})_{n=0}^{N-1} : c \geq 1\} \subseteq (\mathbb{Z}/2\mathbb{Z})^N,$$

which may be of independent interest. It turns out to be a vector space and we are able to construct a basis for it.

In Section 2 we present our main results. Section 3 deals with $\mathcal{V}^{[0,N)}$. The proofs for our results regarding the sequences $\mathbf{t}_S$ are given in Section 4.

2. Notations and main results

We begin with some notations. Let Σ be a finite set (called *alphabet*). A *word* w over Σ is a concatenation of finitely many elements (*letters*) in Σ. The *length* of w is the number of letters in it, and is denoted by $|w|$. If $\Sigma = \mathbb{Z}/2\mathbb{Z}$, the word w is a *binary word*. Let wz denote the concatenation of two words w, z and w^k the concatenation of w with itself $k \geq 0$ times. Thus, for example, $10^3 1(10)^2 0 = 1000110100$ is a binary word of length 10. To avoid possible confusion between the concatenation of words w, z and the product of the integers they may represent, we will restrict the use of this notation only to cases when it is clear that the objects are letters and words. A word w is a *subword* of

z (or w *occurs* in z) if $z = w_0 w w_1$ for some words w_0, w_1. Denote by Σ^l the set of all words of length l over Σ and put $\Sigma^* = \bigcup_{l\geq 0} \Sigma^l$. Given two words $w = w_0 \cdots w_{l-1}$, $z = z_0 \cdots z_{l-1}$ in $(\mathbb{Z}/2\mathbb{Z})^l$, let

$$w + z = (w_0 + z_0)(w_1 + z_1) \cdots (w_{l-1} + z_{l-1}) \in (\mathbb{Z}/2\mathbb{Z})^l.$$

The *binary representation* of an integer $n \geq 1$ is the (unique) binary word $(n)_2 = n_{l-1} \cdots n_1 n_0$ with $n = \sum_{i=0}^{l-1} n_i 2^i$ and $n_{l-1} = 1$. Put $(0)_2 = \epsilon$, where ϵ is the empty word. Denote by $l_2(n)$ the length of $(n)_2$ (thus, $l_2(n) = \lfloor \log_2 n \rfloor + 1$ for $n \geq 1$ and $l_2(0) = 0$).

Let $\mathbf{u} = (u_n)_{n=0}^{\infty}$ be an infinite sequence over Σ. For all integers $i, j \geq 0$ with $i \leq j$, put

$$\mathbf{u}[i, j) = u_i u_{i+1} \cdots u_{j-1} \in \Sigma^{j-i}.$$

Let

$$\Omega_m(\mathbf{u}) = \{\mathbf{u}[i, i + m) : i \geq 0\}, \qquad \Omega(\mathbf{u}) = \bigcup_{m=0}^{\infty} \Omega_m(\mathbf{u}).$$

Thus, the *complexity* of $\mathbf{u}$ is the function $P_{\mathbf{u}}(m) = \#\Omega_m(\mathbf{u})$.

A function $\mu : \Sigma^* \to \Sigma^*$ is a *morphism* if

$$\mu(w_1 w_2) = \mu(w_1)\mu(w_2), \quad w_1, w_2 \in \Sigma^*.$$

Note that every function $\mu : \Sigma \to \Sigma^*$ has a unique extension to a morphism $\mu : \Sigma^* \to \Sigma^*$, given by $\mu(x_0 \cdots x_{l-1}) = \mu(x_0) \cdots \mu(x_{l-1})$, where $x_i \in \Sigma$, $i \leq l - 1$. A morphism μ is *k-uniform* if $|\mu(a)| = k$ for each $a \in \Sigma$. The morphism is *prolongable* on $a \in \Sigma$ if $\mu(a) = aw$ for some word w, and the lengths $|a|, |\mu(a)|, |\mu(\mu(a))|, \ldots$, are strictly increasing. In such a case we have $\mu^i(a) = aw\mu(w)\mu^2(w) \cdots \mu^{i-1}(w)$ for every $i \geq 0$, and thus each $\mu^i(a)$ is a prefix of

$$\mu^{\omega}(a) := aw\mu(w)\mu^2(w) \cdots \in \Sigma^{\mathbb{N}}.$$

If $\mathbf{u} = \mu^{\omega}(a)$ for some morphism μ, which is prolongable on a, then $\mathbf{u}$ is the *pure morphic sequence generated* by μ and a. If there exists a function $\tau : \Sigma \to \Gamma$ such that $\mathbf{u} = \tau(\mu^{\omega}(a))$ (i.e., $\mathbf{u}$ is the sequence over Γ which is obtained from $\mu^{\omega}(a)$ by replacing each element $x \in \Sigma$ with $\tau(x)$) then $\mathbf{u}$ is a *morphic sequence*. A morphic sequence $\mathbf{u}$ is *k-automatic* if $\mathbf{u} = \tau(\mu^{\omega}(a))$ for some k-uniform morphism μ.

Remarks. (1) There are many equivalent definitions for an automatic sequence. For example, one may define an automatic sequence as a sequence which is generated by a finite automaton with output (see [8]). The definition in terms of morphisms is convenient in this paper in order to show the connection between Theorem A and Corollary 5 *infra*.

(2) Automatic sequences have many useful closure properties. For example, if $\mathbf{u} = (u_n)_{n=0}^{\infty}$, $\mathbf{v} = (v_n)_{n=0}^{\infty}$ are k-automatic sequences over Γ, then so are $(u_n)_{n=K}^{\infty}$, $(u_{an+b})_{n=0}^{\infty}$ and $(f(u_n, v_n))_{n=0}^{\infty}$ for all integers $K, a, b \geq 0$ and function $f : \Gamma^2 \to \Gamma$. In particular, the family of k-automatic sequences over $\mathbb{Z}/q\mathbb{Z}$ is closed under addition and multiplication.

(3) Automatic sequences occur in remarkably many different areas (cf. [3–6,16,30]). The reader can refer to [8] for an extensive treatment of automatic sequences.

Example 1. Consider the Thue–Morse sequence $\mathbf{t} = (t_n)_{n=0}^{\infty}$. Let $\mu : (\mathbb{Z}/2\mathbb{Z})^* \to (\mathbb{Z}/2\mathbb{Z})^*$ be the uniform morphism given by $\mu(0) = 01$, $\mu(1) = 10$. Using the relations $t_{2n} = t_n$, $t_{2n+1} = t_n + 1$, we obtain that $t_{2n}t_{2n+1} = \mu(t_n)$ for every n. This implies that $\mathbf{t} = \mu^{\omega}(t_0) = \mu^{\omega}(0)$, and therefore (as is well-known) $\mathbf{t}$ is 2-automatic.

Let $\mathcal{P}$ denote the set of all polynomials $H \in \mathbb{Q}[n]$ with $H(\mathbb{N}) \subseteq \mathbb{N}$. Recall that for a finite set $S \subseteq \mathcal{P}$ we put

$$\mathbf{t}_S = \sum_{H\in S} \mathbf{t}_H \in (\mathbb{Z}/2\mathbb{Z})^{\mathbb{N}},$$

where $\mathbf{t}_H = (t_{H(n)})_{n=0}^{\infty}$. Let

$$\mathcal{U} = \{\mathbf{t}_S : S \subseteq \mathcal{P},\ \#S < \infty\},$$

be the vector space over $\mathbb{Z}/2\mathbb{Z}$ spanned by the sequences $\mathbf{t}_H$, $H \in \mathcal{P}$.

Define the equivalence relation $\sim$ on $\mathbb{Q}[n]$ by: $H \sim G$ if $H = 2^i G + q$ for some $i \in \mathbb{Z}, q \in \mathbb{Q}$. It will be convenient to consider the following condition on finite sets $S \subseteq \mathcal{P}$.

Condition (C1). S may be written in the form

$$S = \{H_1, H_2, \ldots, H_r, G_1, G_2, \ldots, G_r, Q_1, Q_2, \ldots, Q_k\},$$

where $r, k \geq 0$, $H_i \sim G_i$ for each $i \leq r$ and $\deg Q_i \leq 1$ for each $i \leq k$.

Note that for finite sets $S, T \subseteq \mathcal{P}$ we have $\mathbf{t}_S + \mathbf{t}_T = \mathbf{t}_{S \oplus T}$, where $S \oplus T = (S \cup T) \setminus (S \cap T)$ is the symmetric difference of S and T. Moreover, if both S and T satisfy Condition (C1), then so does $S \oplus T$. Thus,

$$\mathcal{U}_{(C1)} = \{\mathbf{t}_S : S \subseteq \mathcal{P},\ \#S < \infty,\ S \text{ satisfies Condition (C1)}\},$$

is a subspace of $\mathcal{U}$.

Our main result is

Theorem 2. *Let S be a finite subset of $\mathcal{P}$, not satisfying Condition (C1). Put $\mathbf{u} = \mathbf{t}_S$ and $d = \max\{\deg H : H \in S\}$. Then*

$$\{z_1 + \cdots + z_{2^{d-2}} : z_1, \ldots, z_{2^{d-2}} \in \Omega_m(\mathbf{u})\} = (\mathbb{Z}/2\mathbb{Z})^m, \quad m \geq 0.$$

In particular, the subword complexity of $\mathbf{u}$ grows exponentially with $P_{\mathbf{u}}(m) \geq c^m$, where $c = 2^{1/2^{d-2}}$.

Observing that $S = \{H\}$ does not satisfy Condition (C1) when $\deg H \geq 2$, we obtain

Corollary 3. *Let $H \in \mathcal{P}$ be a polynomial of degree $d \geq 2$. Then the complexity of $\mathbf{t}_H$ grows exponentially with $P_{\mathbf{t}_H}(m) \geq 2^{m/2^{d-2}}$. In particular, if $\deg H = 2$, then $P_{\mathbf{t}_H}(m) = 2^m$.*

Open Question 4. *Is it true that $P_{\mathbf{t}_H}(m) = 2^m$ for every polynomial $H \in \mathcal{P}$ of degree $d \geq 3$?*

In fact, it may be the case that $P_{\mathbf{t}_S}(m) = 2^m$ for all finite sets $S \subseteq \mathcal{P}$ that do not satisfy Condition (C1).

Another corollary of Theorem 2 is obtained by a result of Ehrenfeucht et al. [18]. Ehrenfeucht et al. proved that $P_{\mathbf{u}}(m) = O(m^2)$ for every morphic sequence $\mathbf{u}$ (see [26,27] for a more precise result). Since the complexity of the sequence $\mathbf{u}$ in Theorem 2 grows exponentially, we conclude the following.

Corollary 5. *The sequence $\mathbf{u} = \mathbf{t}_S$ is not morphic for any finite set $S \subseteq \mathcal{P}$ that does not satisfy Condition (C1).*

Note that our assumptions on the polynomials in S are weaker than the assumptions in Theorem A. Since automatic sequences are a particular type of morphic sequences, we see that Corollary 5 generalizes Theorem A.

Corollary 5, and thus also Theorem 2, fails in general if we replace the sequence $\mathbf{t}$ by any automatic sequence. For example, it is interesting to compare the behavior of $\mathbf{t}$ with that of the sequence $\mathbf{b} = (\nu_2(n) \pmod 2)_{n=1}^{\infty}$, where $\nu_p(n)$ is the p-adic valuation of n (i.e., $p^{\nu_p(n)}$ is the exact power of p dividing the integer n). Thus,

$$\mathbf{b} = (b_n)_{n=1}^{\infty} = 010001010100010001 \cdots .$$

Observing that $b_{2n+1} = 0$ and $b_{2n+2} = b_{n+1} + 1$ for all $n \geq 0$, we get that $\mathbf{b}$ is generated by the morphism given by $0 \mapsto 01$, $1 \mapsto 00$ and $0 = b_1 \in \mathbb{Z}/2\mathbb{Z}$. Hence, $\mathbf{b}$ is 2-automatic.

Since b_0 is undefined, we consider the sequences $\mathbf{b}_H = (b_{H(n)})_{n=0}^{\infty}$ only for polynomials $H \in \mathcal{P}$ with $H(\mathbb{N}) \subseteq \mathbb{N} \setminus \{0\}$. Let $\mathcal{P}_1$ denote the set of such polynomials (that is, $\mathcal{P}_1 = \{H+1 : H \in \mathcal{P}\}$). Here we have the following elegant property:

$$\mathbf{b}_{H_1 \cdot H_2} = \mathbf{b}_{H_1} + \mathbf{b}_{H_2}, \quad H_1, H_2 \in \mathcal{P}_1.$$

Using this formula and the fact that $\mathbf{b}_H$ is 2-automatic when $\deg H = 1$, we get that $\mathbf{b}_H$ is 2-automatic for every $H \in \mathcal{P}_1$ which is a product of linear polynomials H_i in $\mathcal{P}_1$ (see [11] for a more general claim). Since $S = \{H\}$ does not satisfy Condition (C1) when $\deg H \geq 2$, this already shows that Corollary 5 fails if we replace the sequence $\mathbf{t}$ by $\mathbf{b}$.

Remarks. (1) A simple connection between $\mathbf{b}$ and the Thue–Morse sequence is given by the following (cf. [31]),

$$b_{n+1} = t_{n+1} + t_n + 1 \in \mathbb{Z}/2\mathbb{Z}, \quad n \geq 0.$$

Thus, $\mathbf{b}_H = \mathbf{t}_H + \mathbf{t}_{H-1} + 1$ for every $H \in \mathcal{P}_1$. Observing that the sequence of 1's is equal to $\mathbf{t}_Q + \mathbf{t}_{2Q+1}$ (for every $Q \in \mathcal{P}$), we see that $\mathbf{b}_H = \mathbf{t}_S$ for a set S that satisfies Condition (C1).

(2) The sequence $\mathbf{b}$ is known as the *period-doubling sequence* (cf. [6, Section 1.4]).

We now describe our result regarding $\mathcal{V}^{[0,N)} = \{(t_{cn})_{n=0}^{N-1} : c \geq 1\}$.

Theorem 6. *Let $N \geq 1$. Then $\mathcal{V}^{[0,N)}$ is a subspace of $(\mathbb{Z}/2\mathbb{Z})^N$ of dimension $d_0 = \lfloor \frac{N}{2} \rfloor$. Moreover, the set $\mathcal{B} = \{v^{(1)}, v^{(3)}, v^{(5)}, \ldots, v^{(2d_0-1)}\}$, where each $v^{(k)} = (v_i^{(k)})_{i=0}^{N-1}$ is given by*

$$v_i^{(k)} = \begin{cases} 1, & i \in \{k, 2k, 4k, 8k, \ldots, \}, \\ 0, & \text{otherwise}, \end{cases}$$

forms a basis of $\mathcal{V}^{[0,N)}$.

Remarks. (1) In the proof of Theorem 2 we need to consider the set of vectors $\mathcal{V}^I = \{(t_{cn})_{n\in I} : c \geq 1\}$ for some more general (finite) sets $I \subseteq \mathbb{N}$. Taking $N > \max I$, we get that $\mathcal{V}^I$ is the image of $\mathcal{V}^{[0,N)}$ under the projection $\varphi : (\mathbb{Z}/2\mathbb{Z})^N \to (\mathbb{Z}/2\mathbb{Z})^{\#I}$, given by

$$\varphi((v_i)_{i=0}^{N-1}) \mapsto (v_i)_{i\in I}.$$

Let $\mathcal{O}(I)$ denote the set of odd numbers $k \geq 1$ such that $2^l k \in I$ for some $l \in \mathbb{N}$ (for example, $\mathcal{O}(\{4, 6, 7, 12\}) = \{1, 3, 7\}$). Employing Theorem 6, we obtain that $\mathcal{V}^I$ is the vector space spanned by $\mathcal{B}_I = \{\varphi(v^{(k)}) : k \in \mathcal{O}(I)\}$. In particular, $\dim(\mathcal{V}^I) = \#\mathcal{O}(I)$.

(2) One may ask if it is possible to extend Theorem 2 to the family of quasistrongly q-additive sequences considered in [7]. It seems that the main obstacle is to provide a result analogous to Theorem 6. It is easy to prove that, if $\mathbf{u} = (u_n)_{n=0}^{\infty}$ is a quasistrongly q-additive sequence over $\mathbb{Z}/q\mathbb{Z}$, where q is a prime, then $\mathcal{V}_{\mathbf{u}}^{[0,N)} = \{(u_{cn})_{n=0}^{N-1} : c \geq 1\}$ is a vector space over $\mathbb{Z}/q\mathbb{Z}$. However, an analogous construction of $\mathcal{B}$ does not seem to follow easily.

3. The vector space $\mathcal{V}^{[0,N)}$

Lemma 7. *$\mathcal{V}^{[0,N)} = \{(t_{cn})_{n=0}^{N-1} : c \geq 1\}$ is a vector space over $\mathbb{Z}/2\mathbb{Z}$.*

Proof. Let $c_1, c_2 \geq 1$ be integers. We need to prove that $(t_{c_1 n} + t_{c_2 n})_{n=0}^{N-1}$ belongs to $\mathcal{V}^{[0,N)}$. Let $c \in \mathbb{N}$ be given by

$$(c)_2 = (c_1)_2 0^l (c_2)_2,$$

where $l = l_2((N-1)c_2)$. For every $n \leq N-1$ we have

$$(cn)_2 = (c_1 n)_2 0^a (c_2 n)_2,$$

for some $a \geq 0$, and hence $t_{cn} = t_{c_1 n} + t_{c_2 n}$. Thus,

$$(t_{c_1 n} + t_{c_2 n})_{n=0}^{N-1} = (t_{cn})_{n=0}^{N-1} \in \mathcal{V}^{[0,N)}. \quad \blacksquare$$

Proposition 8. *For every odd integer $n_0 \geq 1$ there exists an integer $c \geq 1$ with*

$$t_c = t_{2c} = \cdots = t_{c(n_0-1)} = 0, \quad t_{cn_0} = 1.$$

Proof. If $n_0 = 1$, then we may take $c = 1$. Assume therefore that $n_0 > 1$. We construct two integers N_1, N_2 with

$$t_{N_1 n_0} + t_{N_2 n_0} = 1, \qquad t_{N_1 n} + t_{N_2 n} = 0, \quad n < n_0, \tag{1}$$

(where the additions in (1) are over $\mathbb{Z}/2\mathbb{Z}$). By Lemma 7 there exists a $c \geq 1$ with $t_{cn} = t_{N_1 n} + t_{N_2 n}$, $n \leq n_0$, so that c satisfies the required property.

Since n_0 is odd, we may take an integer $R > 1$ with $Rn_0 \equiv 1 \pmod 4$. Write

$$(Rn_0)_2 = w_1 01, \tag{2}$$

for some word w_1. Take an integer $M \geq 1$ such that $l_2(Mn_0) > l_2(M(n_0-1))$. (For example, one may take $M = \lceil \frac{2^L}{n_0} \rceil$ for some $L > 2l_2(n_0)$, so that $Mn_0 \geq 2^L$, whereas $M(n_0-1) < (M-1)n_0 < 2^L$.) Write

$$(Mn_0)_2 = 1w_2. \tag{3}$$

Let N_1, N_2 be given by

$$(N_1)_2 = (R)_2 0^a (M)_2, \qquad (N_2)_2 = (R)_2 0^{a-1} (M)_2,$$

where $a = l_2(Mn_0) - l_2(M) \geq 1$. Multiplying N_1, N_2 by an integer $n < n_0$, we get

$$(N_1 n)_2 = (Rn)_2 0^{a'} (Mn)_2, \qquad (N_2 n)_2 = (Rn)_2 0^{a'-1} (Mn)_2,$$

where $a' = l_2(Mn_0) - l_2(Mn) \geq 1$. Thus, $t_{N_1 n} + t_{N_2 n} = 2t_{Rn} + 2t_{Mn} = 0$.

Using (2) and (3), and observing that $l_2(Mn_0) = a + l_2(M)$, we obtain

$$(N_1 n_0)_2 = w_1 011 w_2, \qquad (N_2 n_0)_2 = w_1 10 w_2.$$

Denoting by $|w|_1$ the number of 1's in a binary word w, we get

$$t_{N_1 n_0} + t_{N_2 n_0} = 2|w_1|_1 + 2|w_2|_1 + 3 = 1 \in \mathbb{Z}/2\mathbb{Z}. \quad \blacksquare$$

Proof of Theorem 6. Denote by $\mathcal{V}_{\mathcal{B}}$ the vector space spanned by $\mathcal{B}$. Take $u = (t_{cn})_{n=0}^{N-1} \in \mathcal{V}^{[0,N)}$ and an odd integer $k \in [0, N)$. Recall that $t_n = t_{2^i n}$ for all $n, i \geq 0$, and hence (taking $n = ck$), it follows that $t_{ck} = t_{c2^i k}$, $i \geq 0$. This implies that

$$u = \sum_{k=1,3,\ldots,2d_0-1} t_{ck} v^{(k)} \in \mathcal{V}_{\mathcal{B}},$$

and thus $\mathcal{V}^{[0,N)} \subseteq \mathcal{V}_{\mathcal{B}}$.

Proposition 8 shows that for every odd integer $k \in [0, N)$ there exists a vector $u^{(k)} = (u_i^{(k)})_{i=0}^{N-1} \in \mathcal{V}^{[0,N)}$ such that $\min\{i : u_i^{(k)} \neq 0\} = k$. Since the vectors $u^{(k)}$, $k = 1, 3, \ldots, 2d_0 - 1$, are linearly independent, we obtain

$$\dim \mathcal{V}^{[0,N)} \geq d_0 = \#\mathcal{B} = \dim \mathcal{V}_{\mathcal{B}}.$$

Thus, $\mathcal{V}^{[0,N)} = \mathcal{V}_{\mathcal{B}}$. $\blacksquare$

4. Proof of Theorem 2

For a finite set $S \subseteq \mathbb{Q}[n]$, denote

$$\deg S = \max\{\deg H : H \in S\},$$

where we put $\deg H = -\infty$ if $H = 0$ and $\deg \emptyset = -\infty$.

Before we prove Theorem 2 in detail, it will be instructive to sketch the proof. It goes by induction on $\deg S$. We begin by constructing finite sets $D_c(S) \subseteq \mathbb{Q}[n]$, $c = 1, 2, \ldots$, such that, under a certain assumption on c, the following properties hold:

(i) $D_c(S) \subseteq \mathcal{P}$.
(ii) $\Omega(\mathbf{t}_S) \supseteq \Omega(\mathbf{t}_{D_c(S)})$.
(iii) $\deg(D_c(S) \oplus S) < \deg S$.

Put $X_c = D_c(S) \oplus S$. The proof is split into two cases.

Case 1: There exists a c such that X_c does not satisfy Condition (C1).

In this case, by the induction hypothesis,

$$\{z_1 + \cdots + z_{2^{d'-2}} : z_1, \ldots, z_{2^{d'-2}} \in \Omega_m(\mathbf{t}_{X_c})\} = (\mathbb{Z}/2\mathbb{Z})^m, \tag{4}$$

where $d' = \deg X_c$. Now, since $\mathbf{t}_{X_c} = \mathbf{t}_{D_c(S)} + \mathbf{t}_S$, each $w \in \Omega_m(\mathbf{t}_{X_c})$ is a sum of two words, $z_1 \in \Omega_m(\mathbf{t}_{D_c(S)})$ and $z_2 \in \Omega_m(\mathbf{t}_S)$. Using (ii), we conclude that

$$\Omega_m(\mathbf{t}_{X_c}) \subseteq \{z_1 + z_2 : z_1, z_2 \in \Omega_m(\mathbf{t}_S)\}.$$

Together with (4), this implies the required property of $\Omega_m(\mathbf{t}_S)$.

Case 2: Each X_c satisfies Condition (C1).

Here we show (Lemma 13(b)) that S has some particular structure. Using this structure, the fact that $\Omega(\mathbf{t}_S) \supseteq \Omega(\mathbf{t}_{D_c(S)})$ for various values of c, and the equality $\mathcal{V}^I = (\mathbb{Z}/2\mathbb{Z})^{\#I}$ for a certain set $I \subseteq \mathbb{N}$ given by some values of polynomials in S (see the proof of Lemma 12), we prove that $\Omega(\mathbf{t}_S) = (\mathbb{Z}/2\mathbb{Z})^*$ in this case.

Remarks. (1) The induction hypothesis is used only in Case 1.

(2) If $\deg S = 2$, then $\deg X_c \le 1$, so that X_c satisfies Condition (C1). Hence only Case 2 is needed for answering Question A.

Let $L(H)$ denote the leading coefficient of a non-zero polynomial H. Put

$$L(S) = \{L(H) : H \in S,\ H \neq 0\} \subseteq \mathbb{Q}^\times.$$

Throughout the proof it will be convenient to assume the following two conditions on S:

Condition (A1). $L(S)$ does not contain a pair of numbers a, b with $a = 2^i b$ for some integer $i \ge 1$.

Condition (A2). Each $H \in S$ has non-negative coefficients.

To justify why (A1) and (A2) may be assumed without loss of generality, observe the following. For $H \in \mathcal{P}$ we have $\mathbf{t}_H = \mathbf{t}_{2^i H}$, $i \ge 1$. If S contains a pair of polynomials H, G with $G = 2^i H$ for some $i \ge 1$, then $\mathbf{t}_{S\setminus\{H,G\}} = \mathbf{t}_S$. Thus, taking some subset $S_0 \subseteq S$ with $\mathbf{t}_{S_0} = \mathbf{t}_S$, we may assume that S_0 does not contain such pairs H, G. If S_0 contains a pair H, G with $L(H) = 2^i L(G)$, $H \neq 2^i G$, $i \ge 1$, then replacing the polynomial H with $H_1 = 2^i H$ we get $L(H_1) = L(G)$. Repeating this argument we obtain a set S_1, with $\mathbf{t}_{S_1} = \mathbf{t}_S$, satisfying Condition (A1). Now take an integer $K \ge 0$ and put

$$S_2 = \{H(n + K) : H \in S_1\} \subseteq \mathbb{Q}[n].$$

Since $S_1 \subseteq \mathcal{P}$, the leading coefficient of each $H \in S_1$ is positive. Thus, for a sufficiently large K, the set S_2 satisfies Condition (A2) also. Note that S_2 satisfies Condition (C1) if and only if S does, and consequently the assumptions of Theorem 2 on S are valid for S_2. Moreover, we have $\deg S_2 \le \deg S$. Observing that $\mathbf{t}_{S_2}$ is obtained from $\mathbf{t}_{S_1}$ by omitting the first K elements, we find that

$$\Omega(\mathbf{t}_{S_2}) \subseteq \Omega(\mathbf{t}_{S_1}) = \Omega(\mathbf{t}_S).$$

Thus, if we prove that $\mathbf{t}_{S_2}$ satisfies the assertion of Theorem 2, then so does $\mathbf{t}_S$.

Let $H^{(k)} = H^{(k)}(n)$ be the kth derivative of a polynomial $H \in \mathbb{Q}[n]$. Put

$$S^{(k)} = \bigoplus\{\{H^{(k)}\} : H \in S,\ \deg H \ge k\}.$$

Thus, $S^{(k)}$ is the set of non-zero polynomials occurring an odd number of times in $(H^{(k)})_{H\in S}$. For $k = 1$ we write H', S' instead of $H^{(1)}$, $S^{(1)}$, respectively. Denote

$$D_c(H) = \left\{H, cH', \frac{c^2}{2!}H^{(2)}, \ldots, \frac{c^d}{d!}H^{(d)}\right\}, \quad c \ge 1,$$

where $d = \deg H$ and $D_c(H) = \emptyset$ if $H = 0$. Put $D_c(S) = \bigoplus_{H\in S} D_c(H)$.

Note that, even when $S \subseteq \mathcal{P}$, it may still happen that some of the polynomials in $D_c(S)$ have values which are both negative and non-integer. For example, taking $S = \{H\}$, where $H(n) = \frac{1}{2}(n - 8)(n - 9) \in \mathcal{P}$, we obtain $H'(n) = n - \frac{17}{2} \in D_1(S)$. The following lemma, which is easily proved (cf. [1, p. 284]), will enable us avoiding such cases in the proof of Theorem 2.

Lemma 9. *Let H be a polynomial with non-negative coefficients over $\mathbb{Q}$ and let $c \geq 1$ be an integer such that $cH \in \mathbb{N}[n]$. Then $\frac{c^k}{k!}H^{(k)}(n) \in \mathbb{N}[n]$ for every $k \geq 1$.*

Lemma 10. *Let S be a finite subset of $\mathcal{P}$ and $c \geq 1$ be an integer such that $D_c(H) \subseteq \mathcal{P}$ for each $H \in S$. Then*

$$\Omega(\mathbf{t}_S) \supseteq \Omega(\mathbf{t}_{D_c(S)}).$$

Proof. Let $N \geq 0$. We prove that for every sufficiently large l we have $\mathbf{t}_S[c2^l, c2^l + N) = \mathbf{t}_{D_c(S)}[0, N)$. Since every word in $\Omega(\mathbf{t}_{D_c(S)})$ occurs in $\mathbf{t}_{D_c(S)}[0, N)$ for some N, this implies the lemma.

Let $l \geq \max\{l_2(\frac{c^k}{k!}H^{(k)}(n)) : H \in S,\ k \leq \deg H,\ n < N\}$. Take $H \in S$, $d = \deg H$, $r_0 \in \mathbb{N}$, and let $H(r_0 + n) = \sum_{k=0}^{d} H^{(k)}(r_0)\frac{n^k}{k!}$ be the Taylor expansion of $H(r_0 + n) \in \mathbb{Q}[n]$. Substituting $n = c2^l$, we get

$$H(c2^l + r_0) = \sum_{k=0}^{d} 2^{lk}\frac{c^k}{k!}H^{(k)}(r_0).$$

Take $r_0 \in [0, N)$. Note that $l \geq l_2(\frac{c^k}{k!}H^{(k)}(r_0))$ for all $k \leq d$, and therefore

$$(H(c2^l + r_0))_2 = \left(\frac{c^d}{d!}H^{(d)}(r_0)\right)_2 0^{a_{d-1}} \cdots \big(cH'(r_0)\big)_2 0^{a_0}\big(H(r_0)\big)_2$$

for some $a_0, \ldots, a_{d-1} \geq 0$. Thus,

$$t_{H(c2^l+r_0)} = \sum_{k=0}^{d} t_{\frac{c^k}{k!}H^{(k)}(r_0)}, \quad r_0 = 0, \ldots, N-1.$$

This yields, $\mathbf{t}_H[c2^l, c2^l + N) = \mathbf{t}_{D_c(H)}[0, N)$. Since $D_c(S) = \bigoplus_{H\in S} D_c(H)$, we get $\mathbf{t}_S[c2^l, c2^l + N) = \mathbf{t}_{D_c(S)}[0, N)$. ■

Recall that the *density* of a set $\mathcal{Y} \subseteq \mathbb{N}$ is given by $D(\mathcal{Y}) = \lim_{N\to\infty} \frac{\#([0,N)\cap\mathcal{Y})}{N}$, if the limit exists (cf. [21]).

Lemma 11. *Let $S \neq \emptyset$ be a finite set of non-zero polynomials over $\mathbb{Q}$, that do not contain any pair H, G of polynomials with $H = 2^i G$ for an integer $i \neq 0$. Then the set*

$$\mathcal{Y} = \mathcal{Y}(S) = \left\{a \in \mathbb{N} : \exists H, G \in S;\ H \neq G,\ \frac{H(a)}{G(a)} \in \{2^i : i \in \mathbb{Z}\}\right\},$$

is of density 0.

Proof. Since $\mathcal{Y}(S) = \bigcup_{H,G\in S} \mathcal{Y}(\{H, G\})$, it suffices to prove the lemma for a set $S = \{H, G\}$ of size 2. Take $d = \max(\deg H, \deg G)$. Clearly, $\max(\frac{H(n)}{G(n)}, \frac{G(n)}{H(n)}) = O(n^d) = O(2^{d\log_2 n})$. Thus, for sufficiently large N,

$$\left\{\frac{H(a)}{G(a)} : a \in [0, N)\right\} \cap \{2^i : i \in \mathbb{Z}\} \subseteq \{2^i : |i| < (d+1)\log_2 N\}.$$

Observing that, for any fixed $i = i_0$, the equation $\frac{H(n)}{G(n)} = 2^i$ has at most d solutions, we conclude that for every sufficiently large N,

$$\begin{aligned}\#(\mathcal{Y} \cap [0, N)) &= \#\left\{a \in [0, N) : \frac{H(a)}{G(a)} = 2^i,\ i \in \mathbb{Z},\ |i| < (d+1)\log_2 N\right\}\\ &< d(2(d+1)\log_2 N + 1).\end{aligned}$$

Thus $D(\mathcal{Y}) = 0$. ■

Remark. One can show that, if the polynomials in S are of the same degree, then $\mathcal{Y}(S)$ must be finite. However, in general, $\mathcal{Y}(S)$ may be infinite. For example, taking $S = \{H, G\}$, where $G(n) = nH(n)$, we obtain $\mathcal{Y}(S) = \{2^i : i \geq 0\}$.

Let

$$cS = \{cH : H \in S\}, \quad c \geq 1.$$

Lemma 12. *Let $S \neq \emptyset$ be a finite set of non-constant polynomials in $\mathcal{P}$. Take an integer $m \geq 1$. Then*

$$\{\mathbf{t}_{cS}[a, a+m) : c \geq 1\} = (\mathbb{Z}/2\mathbb{Z})^m \tag{5}$$

for almost every $a \in \mathbb{N}$ (i.e., for a set of a's of density 1).

Proof. Without loss of generality we may assume that S satisfies Condition (A1). Thus, for every $H_1, H_2 \in S$ we have $H_1 \sim H_2$ if and only if $H_1(n) = H_2(n + q_0)$ for some integer q_0. Consider the decomposition $S = S_1 \cup S_2 \cup \cdots \cup S_r$ of S into equivalence classes.

We will prove that, for almost every $a \in \mathbb{N}$, there exists a solution $c \geq 1$ to the system of equations

$$\mathbf{t}_{cS_1}[a, a+m) = w_1, \qquad \mathbf{t}_{cS_2}[a, a+m) = w_2, \ldots, \qquad \mathbf{t}_{cS_r}[a, a+m) = w_r,$$

for every r words $w_1, \ldots, w_r \in (\mathbb{Z}/2\mathbb{Z})^m$. Since

$$\mathbf{t}_{cS}[a, a+m) = \sum_{j=1}^{r} \mathbf{t}_{cS_j}[a, a+m),$$

we see that (5) holds for every such a, and thus we will obtain the lemma.

For each j write $S_j = \{H_j(n+k) : k \in X_j\}$ for some $H_j \in S_j$ and a finite set $X_j \subseteq \mathbb{N}$. Put $K = \max \bigcup \{X_j : j \leq r\}$. Consider the set of $r(K+m)$ polynomials

$$E = \{H_j(n+k) : j \leq r,\ k \in [0, K+m)\} \subseteq \mathcal{P}.$$

Since we assume that S satisfies Condition (A1) we get that E satisfies the conditions of Lemma 11. Thus, $D(\mathcal{Y}(E)) = 0$, and therefore $D(\mathbb{N} \setminus \mathcal{Y}(E)) = 1$. Take $a \in \mathbb{N} \setminus \mathcal{Y}(E)$, and denote

$$I = I_{E,a} = \{H(a) : H \in E\} \subseteq \mathbb{N}.$$

Since $a \notin \mathcal{Y}(E)$, the numbers $H(a)$, $H \in E$, are distinct, and so $\#I = r(K+m)$. Moreover, for every $k_1, k_2 \in I$ we have $\frac{k_1}{k_2} \neq 2^i$ for all $i \neq 0$. By Theorem 6, this property of I yields

$$\{(t_{cq})_{q \in I} : c \geq 1\} = (\mathbb{Z}/2\mathbb{Z})^I. \tag{6}$$

Let $w_1, \ldots, w_r$ be binary words of length m, say, $w_j = b_0^{(j)} \cdots b_{m-1}^{(j)}$. For each $j \leq r$, let $x^{(j)} = (x_k^{(j)})_{k=0}^{K+m-1}$ be a solution to the system of linear equations

$$\sum_{k \in X_j} x_{k+n}^{(j)} = b_n^{(j)} \in \mathbb{Z}/2\mathbb{Z}, \quad n = 0, \ldots, m-1.$$

(Note that this system is in echelon form, so that we easily obtain the existence of a solution.) By (6), there exists a $c \geq 1$ such that $t_{cH_j(a+k)} = x_k^{(j)}$ for all $j \leq r$, $k \leq K+m-1$. Thus, for every $j \leq r$ we have

$$\begin{aligned}
\mathbf{t}_{cS_j}[a, a+m) &= \sum_{H \in S_j} \mathbf{t}_{cH}[a, a+m) \\
&= \sum_{k \in X_j} t_{cH_j(a+k)} t_{cH_j(a+k+1)} \cdots t_{cH_j(a+k+m-1)} \\
&= \sum_{k \in X_j} x_k^{(j)} x_{k+1}^{(j)} \cdots x_{k+m-1}^{(j)} \\
&= b_0^{(j)} \cdots b_{m-1}^{(j)} = w_j.
\end{aligned}$$

Since $w_1, \ldots, w_r \in (\mathbb{Z}/2\mathbb{Z})^m$ are arbitrary and $D(\mathbb{N} \setminus \mathcal{Y}(E)) = 1$, this completes the proof. ■

Remark. In the proof of Theorem 2, we only need the existence of an a $(=a(m))$ as in Lemma 12. However, since we are using Lemma 11, we actually conclude that most a's would work.

Lemma 13. *Let S be a finite subset of $\mathcal{P}$ and $c \geq 1$ be an integer. Assume that $D_c(H) \subseteq \mathcal{P}$ for each $H \in S$, and put $X = D_c(S) \oplus S$.*

(a) *If $S \neq \emptyset, \{0\}$, then* $\deg X < \deg S$.

(b) *Assume that S satisfies Condition (A1). Then X satisfies Condition (C1) if and only if S is of the form*

$$S = \{H_1, H_2, \ldots, H_r, G_1, G_2, \ldots, G_r, Q_1, Q_2, \ldots, Q_k\}, \tag{7}$$

where $\deg(H_i - G_i) \leq 1$ *for $i \leq r$ and* $\deg Q_i \leq 2$ *for $i \leq k$. Moreover, if this condition holds, then*

$$X = cS' \oplus \bigoplus \left\{ \frac{c^2}{2} H^{(2)} : H \in S,\ \deg H = 2 \right\}. \tag{8}$$

Proof. (a) Since

$$X = \bigoplus_{H \in S} (D_c(H) \oplus \{H\}) = \bigoplus_{H \in S} \left\{ cH', \frac{c^2}{2} H^{(2)}, \ldots, \frac{c^{\deg H}}{(\deg H)!} H^{(\deg H)} \right\}, \tag{9}$$

we have $\deg X < \deg S$.

(b) Assume that S can be ordered as in (7). Note that, if H, G are polynomials with $\deg(H - G) \leq 1$, then $H^{(k)} = G^{(k)}$ for each $k \geq 2$, so that $D_c(\{H, G\}) \oplus \{H, G\} = \{cH'\} \oplus \{cG'\}$. Thus we easily obtain (8). Observing that for each such pair H, G we have $H' \sim G'$, we conclude that X satisfies Condition (C1).

Conversely, assume that X satisfies Condition (C1). We prove by induction on $d = \deg S$ that S can be ordered as in (7). The cases $d = 0, 1, 2$ are trivial. Assume therefore that $d \geq 3$. Given a set of polynomials E, put $E(k) = \{H \in E : \deg H = k\}$, $k \geq 0$. From (9) it follows that $\deg X \leq d - 1$ and

$$X(d-1) = cS(d)' = \bigoplus \{\{cH'\} : H \in S(d)\}. \tag{10}$$

Note that $L(X(d-1)) \subseteq cL(S)$, and therefore $X(d-1)$ satisfies Condition (A1). Since X satisfies Condition (C1), so does $X(d-1)$. Thus, $X(d-1)$ must be of the form

$$X(d-1) = \{H_1, \ldots, H_l, G_1, \ldots, G_l\},$$

where $H_i - G_i \in \mathbb{Z}$ is a constant for each $i \leq l$. Since $X(d-1) = cS(d)'$, this implies that the polynomials in $S(d)$ can be ordered in pairs (H, G) with $\deg(H - G) \leq 1$. Thus, by the proof of the first direction,

$$D_c(S(d)) \oplus S(d) = cS(d)'. \tag{11}$$

Now put $\hat{S} = S \oplus S(d)$ $(=S \setminus S(d))$ and $\hat{X} = D_c(\hat{S}) \oplus \hat{S}$. Then

$$\hat{X} = D_c(S) \oplus D_c(S(d)) \oplus S \oplus S(d) = X \oplus X(d-1),$$

where the second equality follows from (10) and (11). Since $X, X(d-1)$ satisfy Condition (C1)), so does $\hat{X}$. Using the induction hypothesis on $\hat{S}$, we get that $\hat{S}$ can be ordered as in (7). Since we already provided a similar ordering of $S(d)$, this completes the proof. ■

Proof of Theorem 2. We prove the theorem by induction on $d = \deg S$. Let $\mathbf{u} = \mathbf{t}_S$, where S is a finite subset of $\mathcal{P}$ that does not satisfy Condition (C1). Without loss of generality, assume that S satisfies Conditions (A1) and (A2). Take an integer $q \geq 1$ such that $qH \in \mathbb{N}[n]$ for each $H \in S$. By Lemma 9, $D_{cq}(H) \subseteq \mathcal{P}$ for all $H \in S$, $c \geq 1$. Let $X_c = D_c(S) \oplus S$, $c \geq 1$. We have $X_{cq} \subseteq \mathcal{P}$, $c \geq 1$.

Since S does not satisfy Condition (C1), d must be at least 2. We begin by proving the theorem for cases where X_{cq} satisfies Condition (C1) for each c. This, in particular, proves the theorem for the case $d = 2$ (see the remarks at the beginning of this section).

By Eq. (8) in Lemma 13(b) we have $D_{cq}(S) = S \oplus cqS' \oplus c^2 T$, where

$$T = \bigoplus \left\{ \frac{q^2}{2} H^{(2)} : H \in S,\ \deg H = 2 \right\}.$$

Write $qS' = T_0 \cup T_1$, where $T_0 = \{H \in qS' : \deg H = 0\}$ and $T_1 = qS' \setminus T_0$. Thus,

$$\mathbf{t}_{D_{cq}(S)} = \mathbf{t}_S + \mathbf{t}_{cT_0} + \mathbf{t}_{cT_1} + \mathbf{t}_{c^2T}, \quad c \geq 1. \tag{12}$$

Since S does not satisfy Condition (C1), we have $T_1 \neq \emptyset$. Now fix an integer $m \geq 1$. Lemma 12 proves the existence of an $a \geq 0$ with

$$\{\mathbf{t}_{cqT_1}[a, a+m) : c \geq 1\} = (\mathbb{Z}/2\mathbb{Z})^m.$$

Observing that $\mathbf{t}_S[a, a+m)$ does not depend on c, we find that

$$\{\mathbf{t}_S[a, a+m) + \mathbf{t}_{cT_1}[a, a+m) : c \geq 1\} = (\mathbb{Z}/2\mathbb{Z})^m. \tag{13}$$

Since every polynomial in $T_0 \cup T$ is of degree 0, the sequences $\mathbf{t}_{cT_0}$ and $\mathbf{t}_{c^2T}$ are constant, so that

$$\mathbf{t}_{cT_0}[a, a+m) + \mathbf{t}_{c^2T}[a, a+m) \in \{0^m, 1^m\}, \quad c \geq 1. \tag{14}$$

Using (12)–(14), we conclude that for each $w \in (\mathbb{Z}/2\mathbb{Z})^m$ there exists a $c \geq 1$ such that either $\mathbf{t}_{D_{cq}(S)}[a, a+m) = w$ or $\mathbf{t}_{D_{cq}(S)}[a, a+m) = \overline{w}$ (where we denote $\overline{w} = w + 1^{|w|}$). Thus, at least one of the words $w, \overline{w}$ belongs to $\Omega(\mathbf{t}_{D_{cq}(S)})$. By Lemma 10, we have $\Omega(\mathbf{t}_{D_{cq}(S)}) \subseteq \Omega(\mathbf{t}_S)$, and consequently one of $w, \overline{w}$ belongs to $\Omega(\mathbf{t}_S)$. Since m is arbitrary, $\Omega(\mathbf{t}_S)$ must contain one of the two words $w' = w\overline{w}$ and $\overline{w'} = \overline{w}w \in (\mathbb{Z}/2\mathbb{Z})^{2|w|}$. This shows that every binary word w occurs in $\mathbf{t}_S$, i.e., that $\Omega_m(\mathbf{t}_S) = (\mathbb{Z}/2\mathbb{Z})^m$, $m \geq 0$, in this case.

Assume now that X_{cq} does not satisfy Condition (C1) for some $c \geq 1$. (In fact, by Lemma 13(b), this means that X_{cq} does not satisfy Condition (C1) for every $c \geq 1$, but this will be of no consequence.) Denote $d_0 = \deg X_{cq}$. Lemma 13(a) yields $d_0 < d$, and thus, by the induction hypothesis, every word $w \in (\mathbb{Z}/2\mathbb{Z})^*$ is a sum of 2^{d_0-2} words in $\Omega(\mathbf{t}_{X_{cq}})$. By the arguments at the beginning of this section, each word in $\Omega(\mathbf{t}_{X_{cq}})$ is a sum of 2 words in $\Omega(\mathbf{t}_S)$. It follows that each word $w \in (\mathbb{Z}/2\mathbb{Z})^*$ is a sum of $2 \cdot 2^{d_0-2} = 2^{d_0-1}$ words in $\Omega(\mathbf{t}_S)$. Since $2^{d_0-1} \mid 2^{d-2}$, this completes the proof. ■

Acknowledgments

The author is grateful to J.-P. Allouche, D. Berend, X. Le Breton and the anonymous referee for their comments and many useful suggestions regarding the paper.

The author was supported in part by Bourse Chateaubriand and by Edmund Landau Center.

References

[1] J.-P. Allouche, Somme des chiffres et transcendance, Bull. Soc. Math. France 110 (1982) 279–285.
[2] J.-P. Allouche, Sur la complexité des suites infinies, Bull. Belg. Math. Soc. 1 (1994) 133–143.
[3] J.-P. Allouche, Transcendence of formal power series with rational coefficients, Theoret. Comput. Sci. 218 (1999) 143–160.
[4] J.-P. Allouche, V. Berthé, Triangle de Pascal, complexité et automates, Bull. Belg. Math. Soc. 4 (1997) 1–23.
[5] J.-P. Allouche, T. Johnson, Finite automata and morphisms in assisted musical composition, J. New Music Res. 24 (1995) 97–108.
[6] J.-P. Allouche, M. Mendès France, Automata and automatic sequences, in: Beyond Quasicrystals, Springer/Les Éditions de Physique, 1995, pp. 293–367.
[7] J.-P. Allouche, O. Salon, Sous-suites polynomiales de certaines suites automatiques, J. Théor. Nombres Bordeaux 5 (1993) 111–121.
[8] J.-P. Allouche, J.O. Shallit, Automatic Sequences. Theory, Applications, Generalizations, Cambridge University Press, 2003.
[9] J.-P. Allouche, J.O. Shallit, The ubiquitous Prouhet–Thue–Morse sequence, sequences and their applications, in: Proceedings of SETA'98, Springer, 1999, pp. 1–16.
[10] S.V. Avgustinovich, The number of different subwords of given length in the Morse–Hedlund sequence, Sibirsk. Zh. Issled. Oper. 1 (1994) 3–7.
[11] J. Bell, p-adic valuations and k-regular sequences, Discrete Math. (in press).
[12] J. Berstel, Axel Thue's papers on repetitions in words: A translation, Publications du Laboratoire de Combinatoire et d'Informatique Mathématique, vol. 20, Université du Québec à Montréal, 1995.
[13] V. Berthé, Sequences of low complexity: Automatic and Sturmian sequences, in: London Math. Soc. Lecture Note Series, vol. 279, Cambridge University Press, 2000, pp. 1–34.
[14] S. Brlek, Enumeration of factors in the Thue–Morse word, Discrete Appl. Math. 24 (1989) 83–96.
[15] S. Brown, N. Rampersad, J. Shallit, T. Vasiga, Squares and overlaps in the Thue–Morse sequence and some variants, RAIRO- Info. Theor. Appl. 40 (2006) 473–484.

[16] G. Christol, T. Kamae, M. Mendès France, G. Rauzy, Suites algébriques, automates et substitutions, Bull. Soc. Math. France 108 (1980) 401–419.
[17] A. Cobham, Uniform tag sequences, Math. Systems Theory 6 (1972) 164–192.
[18] A. Ehrenfeucht, K.P. Lee, G. Rozenberg, Subword complexities of various classes of deterministic developmental languages without interactions, Theoret. Comput. Sci. 1 (1975) 59–75.
[19] S. Ferenczi, Complexity of sequences and dynamical systems, Discrete Math. 206 (1999) 145–154.
[20] S. Ferenczi, Z. Kása, Complexity for finite factors of infinite sequences, Theoret. Comput. Sci. 218 (1999) 177–195.
[21] H. Furstenberg, Recurrence in Ergodic Theory and Combinatorial Number Theory, Princeton University Press, 1981.
[22] X. Le Breton, Linear independence of automatic formal power series, Discrete Math. 306 (2006) 1776–1780.
[23] A. de Luca, S. Varricchio, Some combinatorial properties of the Thue–Morse sequence and a problem in semigroups, Theoret. Comput. Sci. 63 (1989) 333–348.
[24] M. Morse, Recurrent geodesics on a surface of negative curvature, Trans. Amer. Math. Soc. 22 (1921) 84–100.
[25] B. Mossé, Reconnaissabilité des substitutions et complexité des suites automatiques, Bull. Soc. Math. France 124 (1996) 329–346.
[26] J.-J. Pansiot, Bornes inférieures sur la complexité des facteurs des mots infinis engendrés par morphismes itérés, in: Lecture Notes in Computer Science, vol. 166, Springer, 1984, pp. 230–240.
[27] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, in: Lecture Notes in Computer Science, vol. 172, Springer, 1984, pp. 380–389.
[28] E. Prouhet, Mémoire sur quelques relations entre les puissances des nombres, C. R. Acad. Sci. Paris Sér. I 33 (1851) 225.
[29] M. Queffélec, Substitution Dynamical Systems—Spectral Analysis, in: Lecture Notes in Mathematics, vol. 1294, Springer, 1987.
[30] J.O. Shallit, Number theory and formal languages, in: Emerging Applications of Number Theory, in: IMA Vol. Math. Appl., vol. 109, Springer, 1999, pp. 547–570.
[31] J.O. Shallit, J. Tromp, Subword complexity of a generalized Thue–Morse word, Inform. Process. Lett. 54 (1995) 313–316.
[32] A. Thue, Über unendliche Zeichenreihen, Norske vid. Selsk. Skr. I. Mat. Kl. Christiana 7 (1906) 1–22.
[33] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Norske vid. Selsk. Skr. I. Mat. Kl. Christiana 1 (1912) 1–67.