KYBERNETIKA — VOLUME 48 (2012), NUMBER 3, PAGES 361–370

GENERALIZED THUE–MORSE WORDS AND PALINDROMIC RICHNESS

Štěpán Starosta

We prove that the generalized Thue–Morse word $\mathbf{t}_{b,m}$ defined for $b \geq 2$ and $m \geq 1$ as $\mathbf{t}_{b,m} = (s_b(n) \mod m)_{n=0}^{+\infty}$, where $s_b(n)$ denotes the sum of digits in the base-b representation of the integer n, has its language closed under all elements of a group D_m isomorphic to the dihedral group of order $2m$ consisting of morphisms and antimorphisms. Considering antimorphisms $\Theta \in D_m$, we show that $\mathbf{t}_{b,m}$ is saturated by Θ-palindromes up to the highest possible level. Using the generalisation of palindromic richness recently introduced by the author and E. Pelantová, we show that $\mathbf{t}_{b,m}$ is D_m-rich. We also calculate the factor complexity of $\mathbf{t}_{b,m}$.

Keywords: palindrome, palindromic richness, Thue–Morse, Theta-palindrome

Classification: 68R15

1. INTRODUCTION

A palindrome is a word which coincides with its reverse image, or more formally, a finite word w is a palindrome if $w = R(w)$, where the reversal (or mirror) mapping $R : \mathcal{A}^* \mapsto \mathcal{A}^*$ is defined by $R(w_1w_2\dots w_n) = w_nw_{n-1}\dots w_1$ for letters $w_i \in \mathcal{A}$, $\mathcal{A}$ being an alphabet. In [9], the authors gave an upper bound on the number of palindromic factors in a finite word: a finite word of length n contains at most $n+1$ palindromic factors. If this bound is attained, we say that a word is rich in palindromes (introduced in [11], and in [5] where such a word is called full). This definition can be naturally extended to infinite words: an infinite word is rich in palindromes if every its factor is rich in palindromes (see [11]). For infinite words with language closed under reversal, i. e., containing the reverse of every factor, there exist several equivalent characterizations of palindromic richness. Each of them can be adopted as a definition.

Let us list three of these characterizations. An infinite word $\mathbf{u}$ with language closed under reversal is rich if one of the following equivalent statements holds:

1. any factor w of $\mathbf{u}$ of length n contains exactly $n+1$ palindromic factors;
2. for any $n \in \mathbb{N}$, the equality $\Delta\mathcal{C}(n) + 2 = \mathcal{P}(n) + \mathcal{P}(n+1)$ is satisfied, where $\Delta\mathcal{C}(n) = \mathcal{C}(n+1) - \mathcal{C}(n)$ and $\mathcal{P}(n)$ denotes the palindromic complexity of $\mathbf{u}$, i. e., the number of palindromes of length n in the set of factors of $\mathbf{u}$ ([6]);
3. each complete return word of any palindrome occurring in $\mathbf{u}$ is a palindrome as well ([11]).

Let us mention that the inequality

$$\Delta\mathcal{C}(n) + 2 \geq \mathcal{P}(n) + \mathcal{P}(n+1) \tag{1}$$

is valid for any infinite word $\mathbf{u}$ with language closed under reversal and for any n (see [2]). Thus, both characterizations 1 and 2 express that a rich word $\mathbf{u}$ is saturated by palindromes up to the highest possible level. Episturmian words (see [9]) and words coding interval exchange transformations determined by a symmetric permutation (see [2]) are some of the most prominent examples of rich words.

If we replace the reversal mapping R by an antimorphism Θ, we can define Θ-palindromes as words which are fixed points of Θ, i. e., $w = \Theta(w)$. For any antimorphism Θ, the notion of Θ-palindromic richness can be introduced and analogue of characterizations 1, 2 and 3 mentioned above can be formulated, see [14]; a word is Θ-rich if it contains the maximum possible number of Θ-palindromic factors.

The famous Thue–Morse word

$$0110100110010110\ldots$$

is a binary word defined as a fixed point of the morphism defined by $0 \mapsto 01, 1 \mapsto 10$. Although the language of the Thue–Morse word is closed under two antimorphisms, it is not Θ-rich for any of these two antimorphisms. The author together with E. Pelantová in [12] explored infinite words with language closed under more antimorphisms simultaneously. For a given finite group G formed by morphisms and antimorphisms on $\mathcal{A}^*$, the words with language closed under any element of G were investigated. If, moreover, such a word $\mathbf{u}$ is uniformly recurrent, then a generalized version of the inequality (1) was obtained: there exists an integer N such that

$$\Delta\mathcal{C}(n) + \#G \quad \geq \sum_{\Theta \in G^{(2)}} \Big(\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n+1)\Big) \qquad \text{for any } n \geq N\,, \tag{2}$$

where $G^{(2)}$ denotes the set of involutive antimorphisms of G and $\mathcal{P}_\Theta$ is the Θ-palindromic complexity, i. e., $\mathcal{P}_\Theta(n)$ counts the number of Θ-palindromes of length n in set of factors of $\mathbf{u}$.

Infinite uniformly recurrent words with language closed under all elements of G and for which equality is attained in the inequality (2) for any n greater than some integer M are called *almost G-rich*. Note that if the group G contains (besides the identity) only the reversal mapping R, the notion of almost richness (as introduced in [11]) and the notion of almost G-richness coincide. In [12], *G-richness* is also introduced. Again, in the case of $G = \{R, \mathrm{Id}\}$ it coincides with classical palindromic richness. The definition requires further notions, thus we omit it here and restrict ourselves to the following criterion for G-richness.

Proposition 1.1. An infinite word over $\mathcal{A}$ with language closed under all elements of a finite group G consisting of morphisms and antimorphisms of $\mathcal{A}^*$ is G-rich if

- for any two antimorphisms $\Theta_1, \Theta_2 \in G$ and any non-empty factor v of $\mathbf{u}$ we have $\Theta_1 \neq \Theta_2 \Rightarrow \Theta_1(v) \neq \Theta_2(v)$, and

- for any two morphisms $\varphi_1, \varphi_2 \in G$ and any non-empty factor v of $\mathbf{u}$ we have $\varphi_1 \neq \varphi_2 \Rightarrow \varphi_1(v) \neq \varphi_2(v)$, and
- the equality (2) is attained for all $n \geq 1$.

Note that the given criterion states only a necessary condition and the converse of the proposition is not true.

In [12], the authors show that the Thue–Morse word is G-rich, where G is a group generated by the reversal mapping and the antimorphism determined by the exchange of 0 and 1. The class of so-called generalized Thue–Morse words is also partially treated in that article, but the question of their G-richness is not resolved. In this article we prove that all generalized Thue–Morse words are G-rich and we give explicitly the group G.

The generalized Thue–Morse words were already considered by E. Prouhet in 1851, see [13]. Let $s_b(n)$ denote the sum of digits in the base-b representation of the integer n, for integers $b \geq 2$ and $m \geq 1$. The generalized Thue–Morse word $\mathbf{t}_{b,m}$ is defined as

$$\mathbf{t}_{b,m} = (s_b(n) \mod m)_{n=0}^{+\infty}.$$

Using this notation, the famous Thue–Morse word equals $\mathbf{t}_{2,2}$. The word $\mathbf{t}_{b,m}$ is over the alphabet $\{0, 1, \ldots, m-1\} = \mathbb{Z}_m$. Similarly to the classical Thue–Morse word, also $\mathbf{t}_{b,m}$ is a fixed point of a primitive substitution, as already mentioned in [1]. It is easy to see that the substitution fixing the word $\mathbf{t}_{b,m}$ is defined by

$$\varphi_{b,m}(k) = k(k+1)(k+2)\ldots\big(k+(b-1)\big) \quad \text{for any } k \in \mathbb{Z}_m,$$

where the letters are expressed modulo m. As already stated in [1], it can be shown that $\mathbf{t}_{b,m}$ is periodic if and only if $b \equiv 1 \pmod{m}$.

We show that for any parameters b and m the language of the word $\mathbf{t}_{b,m}$ is closed under all elements of a group, denoted D_m, isomorphic to the dihedral group of order $2m$ (see Proposition 3.1) and that $\mathbf{t}_{b,m}$ is D_m-rich (see Theorem 4.4). In the last section, we use the results to give the formula for the factor complexity of $\mathbf{t}_{b,m}$.

For more information about generalized Thue–Morse words see [1] where authors give conditions on b and m so that $\mathbf{t}_{b,m}$ is overlap-free and also answer the question whether $\mathbf{t}_{b,m}$ contains arbitrarily long palindromes or infinitely many squares. In [3], the author describes the factor frequencies of $\mathbf{t}_{b,m}$.

2. PRELIMINARIES

An *alphabet* $\mathcal{A}$ is a finite set, its elements are called *letters*. A *finite word* over $\mathcal{A}$ is a finite string $w = w_1 w_2 \ldots w_n$ of letters $w_i \in \mathcal{A}$. Its *length* is $|w| = n$. The set $\mathcal{A}^*$ is formed by all finite words over $\mathcal{A}$ and it is a free monoid with *the empty word* ε as neutral element. An *infinite word* $\mathbf{u} = (u_i)_{i=0}^{+\infty}$ is an infinite sequence of letters $u_i \in \mathcal{A}$. A word $v \in \mathcal{A}^*$ is a *factor* of a word w (finite or infinite) if there exist words s, t such that $w = svt$. If $s = \varepsilon$, then v is a *prefix* of w. If $t = \varepsilon$, then v is a *suffix* of w. An integer i such that $v = w_i \ldots w_{i+|w|-1}$ is called an occurrence of v in w. We say that an infinite word is *uniformly recurrent* if each factor occurs infinitely many times and the gaps between its successive occurrences form a bounded sequence.

By $\mathcal{L}_n(\mathbf{u})$ we denote the set of factors of $\mathbf{u}$ of length n. The set of all factors of $\mathbf{u}$ is denoted by $\mathcal{L}(\mathbf{u})$ and is called the *language of* $\mathbf{u}$. The *factor complexity* $\mathcal{C}$ of an infinite word $\mathbf{u}$ is the mapping $\mathbb{N} \mapsto \mathbb{N}$ counting the number of distinct factors of given length, i. e., $\mathcal{C}(n) = \#\mathcal{L}_n(\mathbf{u})$.

A letter $a \in \mathcal{A}$ is a *left extension* of a factor $w \in \mathcal{L}(\mathbf{u})$ if $aw \in \mathcal{L}(\mathbf{u})$. If a factor has at least two distinct left extensions, then it is *left special.* The set of all left extensions of w is denoted $\mathrm{Lext}(w)$. The definitions of *right extensions*, $\mathrm{Rext}(w)$ and *right special* are analogous. A factor which is left and right special is *bispecial* (BS).

The *bilateral order* $\mathrm{b}(w)$ of a factor w is the number $\mathrm{b}(w) = \#\mathrm{Bext}(w) - \#\mathrm{Lext}(w) - \#\mathrm{Rext}(w) + 1$ where the set $\mathrm{Bext}(w) = \{awb \in \mathcal{L}(\mathbf{u}) \mid a, b \in \mathcal{A}\}$. In [7], the following relation between the second difference of factor complexity and bilateral orders is proved:

$$\mathcal{C}(n+2) - 2\mathcal{C}(n+1) + \mathcal{C}(n) = \Delta^2\mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} \mathrm{b}(w). \tag{3}$$

One can easily show that if w is not a bispecial factor of an infinite word $\mathbf{u}$, then $\mathrm{b}(w) = 0$. Thus, to enumerate the factor complexity of an infinite word $\mathbf{u}$ one needs to calculate $\mathcal{C}(0), \mathcal{C}(1)$, and the bilateral orders of all its bispecial factors.

A mapping φ on $\mathcal{A}^*$ is called a *morphism* if $\varphi(vw) = \varphi(v)\varphi(w)$ for any $v, w \in \mathcal{A}^*$ and an *antimorphism* if $\varphi(vw) = \varphi(w)\varphi(v)$ for any $v, w \in \mathcal{A}^*$. By $AM(\mathcal{A}^*)$ we denote the set of all morphisms and antimorphisms over $\mathcal{A}^*$. Let $\nu \in AM(\mathcal{A}^*)$. We say that $\mathcal{L}(\mathbf{u})$ is *closed under* ν if for all $w \in \mathcal{L}(\mathbf{u})$ we have $\nu(w) \in \mathcal{L}(\mathbf{u})$.

It is clear that the reversal mapping R is an antimorphism. Moreover, it is an involution, i. e., $R^2 = \mathrm{Id}$. A fixed point of an antimorphism Θ is a Θ-*palindrome.* If $\Theta = R$, then we say palindrome or classical palindrome instead of R-palindrome. The set of all Θ-palindromic factors of an infinite word $\mathbf{u}$ is denoted by $\mathrm{Pal}_\Theta(\mathbf{u})$. The Θ-*palindromic complexity* of $\mathbf{u}$ is the mapping $\mathcal{P}_\Theta : \mathbb{N} \mapsto \mathbb{N}$ given by $\mathcal{P}_\Theta(n) = \#\left(\mathrm{Pal}_\Theta(\mathbf{u}) \cap \mathcal{L}_n(\mathbf{u})\right)$. If $a \in \mathcal{A}$, $w \in \mathrm{Pal}_\Theta(\mathbf{u})$, and $aw\Theta(a) \in \mathcal{L}(\mathbf{u})$, then $aw\Theta(a)$ is said to be a Θ-*palindromic extension* of w in $\mathbf{u}$. The set of all Θ-palindromic extensions of w is denoted by $\mathrm{Pext}_\Theta(w)$.

If $\mathbf{u}$ is a fixed point of a morphism φ, then a factor $v = v_0v_1\dots v_{s-1} \in \mathcal{L}(\mathbf{u})$ is an *ancestor* of a factor $w \in \mathcal{L}(\mathbf{u})$ if w is a factor of $\varphi(v)$ and is not a factor of $\varphi(v_1 \dots v_{s-1})$ or $\varphi(v_0 \dots v_{s-2})$.

3. GENERALIZED THUE–MORSE WORDS AND DIHEDRAL GROUPS

In this section we show that $\mathcal{L}(\mathbf{t}_{b,m})$ is closed under all elements of an explicit group $G \subset AM(\mathcal{A}^*)$. Fix $b \geq 2$ and $m \geq 1$. In what follows, to ease the notation, we denote $\varphi = \varphi_{b,m}$, the alphabet is considered to be $\mathbb{Z}_m$, and letters are expressed modulo m.

For all $x \in \mathbb{Z}_m$ denote by Ψ_x the antimorphism given by

$$\Psi_x(k) = x - k \quad \text{for all } k \in \mathbb{Z}_m$$

and by Π_x the morphism given by

$$\Pi_x(k) = x + k \quad \text{for all } k \in \mathbb{Z}_m.$$

Let D_m denote the set $D_m = \{\Psi_x \bigm| x \in \mathbb{Z}_m\} \cup \{\Pi_x \bigm| x \in \mathbb{Z}_m\}$. It is easy to show that D_m is a group and can be generated by 2 elements, for instance one can choose Π_1 and

Ψ_0. Since the order of Π_1 is m, Ψ_0 is an involution, and $\Psi_0\Pi_1$ is also an involution, D_m is isomorphic to the dihedral group of order $2m$.

Let $\pi : \mathbb{Z}_m \mapsto \mathbb{Z}_m$ denote the permutation defined for all $k \in \mathbb{Z}_m$ by

$$\pi(k) = \text{the last letter of } \varphi(k) = k + b - 1 = \Pi_{b-1}(k).$$

Let q denote the order of π, i. e., the smallest positive integer q such that $q(b-1) \equiv 0 \pmod{m}$.

The following properties will help us to prove the next proposition.

Property I. For all $x \in \mathbb{Z}_m$, we have $\Pi_x\varphi = \varphi\Pi_x$ and $\Psi_x\varphi = \varphi\Psi_{x+b-1}$.

Proof. It follows directly from the definitions of φ, Π_x and Ψ_x. □

Property II. $\mathcal{L}_2(\mathbf{t}_{b,m}) = \{\pi^k(r-1)r \mid r \in \mathbb{Z}_m, 0 \le k \le q-1\}$.

Proof. Denote by $L_0 = \{(r-1)r \mid r \in \mathbb{Z}_m\}$. It is clear that $L_0 \subset \mathcal{L}_2(\mathbf{t}_{b,m})$. For all i, denote by L_{i+1} the set of factors of length 2 of the words $\varphi(w)$ for all $w \in L_i$. From the definition of π, it is clear that $L_i = \{\pi^k(r-1)r \mid r \in \mathbb{Z}_m, 0 \le k \le i\}$. The definition of q then guarantees $L_{q-1} = L_q$ and the equality $L_{q-1} = \mathcal{L}_2(\mathbf{t}_{b,m})$ follows from the construction of the sets L_i. □

Proposition 3.1. The language of $\mathbf{t}_{b,m}$ is closed under all elements of D_m.

Proof. We show the claim by induction on the length n of factors. We first verify the claim for $n = 2$. Using Property II, it is easy to show that $\mathcal{L}_2(\mathbf{t}_{b,m})$ is invariant under all elements of D_m.

Suppose now the claim holds for factors of length $n \ge 2$ and take $w \in \mathcal{L}_{n+1}(\mathbf{t}_{b,m})$. It is clear that there exists a factor v such that $1 \le |v| \le n$ and w is a factor of $\varphi(v)$. Let $x \in \mathbb{Z}_m$. Using Property I, one has $\Pi_x\varphi(v) = \varphi\Pi_x(v)$. Since we supposed $\Pi_x(v) \in \mathcal{L}(\mathbf{t}_{b,m})$, it is clear that $\Pi_x(w)$ is a factor of $\mathbf{t}_{b,m}$ Using again Property I for Ψ_x, we also have $\Psi_x\varphi(v) = \varphi\Psi_{x+b-1}(v)$, and thus $\Psi_x(w)$ is a factor of $\mathbf{t}_{b,m}$. □

4. D_m-RICHNESS OF $\mathbf{t}_{b,m}$

In this section we show that the word $\mathbf{t}_{b,m}$ is D_m-rich. The following properties of φ and $\mathbf{t}_{b,m}$ can be easily deduced.

Property III. φ is uniform, i. e., for all $k, \ell \in \mathbb{Z}_m$, $|\varphi(k)| = |\varphi(\ell)| = b$.

Property IV. φ is marked (see [10]), i. e., for all $k, \ell \in \mathbb{Z}_m$ such that $k \ne \ell$, the first letter of $\varphi(k)$ differs from the first letter of $\varphi(\ell)$, and the same holds for the last letters.

Property V. $\mathcal{L}_3(\mathbf{t}_{b,m}) = \{\pi^k(t-1)t(t+1) \mid t \in \mathbb{Z}_m, 0 \le k \le q-1\} \cup \{(t-1)t\pi^{-k}(t+1) \mid t \in \mathbb{Z}_m, 0 \le k \le q-1\}$.

Proof. It follows from Property II and the definition of φ. □

Property VI. For all words w of length 1 or 2 there exists exactly one x such that w is a Ψ_x-palindrome.

Proof. It follows directly from the definition of Ψ_x. $\square$

Property VII. If for $x \in \mathbb{Z}_m$ the word $w \in \mathcal{L}(\mathbf{t}_{b,m})$ is a Ψ_x-palindrome, then the factor $\varphi(w)$ is a Ψ_{x-b+1}-palindrome.

Proof. It follows directly from Property I. $\square$

Property VIII. If $w \in \mathcal{L}(\mathbf{t}_{b,m})$ is a BS factor and $\nu \in D_m$, then $\nu(w)$ is BS factor and $\mathrm{b}(w) = \mathrm{b}(\nu(w))$.

Moreover, if w is a Θ-palindrome for some antimorphism $\Theta \in D_m$, then $\nu(w)$ is a Θ'-palindrome for some $\Theta' \in D_m$.

Proof. Property IV and Proposition 3.1 guarantee the first part of the statement. The second part of the statement can be verified by setting $\Theta' = \nu\Theta\nu^{-1}$. $\square$

Property IX. If $w = w_0 \dots w_{s-1} \in \mathcal{L}(\mathbf{t}_{b,m})$ and there is an index i such that $w_{i+1} \neq w_i + 1$, then w has exactly one ancestor.

Proof. It follows directly from the definition of φ and Property III. $\square$

Property X. Let $b \not\equiv 1 \pmod{m}$ and $w \in \mathcal{L}(\mathbf{t}_{b,m})$. If $|w| > 2b$, then w has exactly one ancestor.

Proof. Take $|w| = 2b + 1$. Suppose that there is no index i such that $w_{i+1} \neq w_i + 1$, i. e., $w = k(k+1)\dots(k+2b)$ for some integer k. Since every ancestor of w is of length 3, it implies that there exists a factor $v \in \mathcal{L}_3(\mathbf{t}_{b,m})$ such that $v = \ell(\ell+b)(\ell+2b)$ for some ℓ. Since $b \not\equiv 1 \pmod{m}$, it is a contradiction with Property V. $\square$

Property XI. If $w \in \mathcal{L}(\mathbf{t}_{b,m})$ is BS and $|w| \geq b$, then there exist letters x and y such that $\varphi(x)$ is a prefix of w and $\varphi(y)$ is a suffix of w.

Proof. The claim is a direct consequence of Property IV and the definition of φ. $\square$

These properties are used to prove the next two lemmas. The first lemma summarizes the bilateral orders and Θ-palindromic extensions of longer BS factors. Since there are no non-empty BS factors in the periodic case, we need to deal with it apart.

Lemma 4.1. If $b \not\equiv 1 \pmod{m}$ and $w \in \mathcal{L}(\mathbf{t}_{b,m})$ is a BS factor of $\mathbf{t}_{b,m}$ such that $|w| \geq 2b$, then there exists a BS factor v such that $\varphi(v) = w$. Furthermore, $\mathrm{b}(w) = \mathrm{b}(v)$.

If v is a Θ_1-palindrome for some $\Theta_1 \in D_m$, then there exists a unique $\Theta_2 \in D_m$ such that w is a Θ_2-palindrome. Moreover, $\#\mathrm{Pext}_{\Theta_2}(w) = \#\mathrm{Pext}_{\Theta_1}(v)$.

Proof. Let w be a BS factor of length $|w| \geq 2b$.

If $|w| > 2b$, then the existence of a unique ancestor v follows from Property X. If $|w| = 2b$ and there exists an ancestor $v \in \mathcal{L}_3(\mathbf{t}_{b,m})$, then we have a contradiction to

Properties XI and V. Thus, if $|w| = 2b$, then there exists a unique ancestor $v \in \mathcal{L}_2(\mathbf{t}_{b,m})$ such that $\varphi(v) = w$.

Property IV guarantees that v is BS and $\mathrm{b}(w) = \mathrm{b}(v)$.

Suppose v is a Ψ_x-palindrome for some $x \in \mathbb{Z}_m$. According to Property VII, w is a Ψ_{x-b+1}-palindrome. The fact that there is no other such antimorphism follows from Property VI.

The equality $\#\mathrm{Pext}_{\Psi_x}(v) = \#\mathrm{Pext}_{\Psi_{x-b+1}}(w)$ follows again from Property VII. $\square$

Thanks to the last lemma, we have to evaluate only the bilateral orders and the number of palindromic extensions of shorter factors. The next lemma exhibits these values for required lengths of BS factors.

Lemma 4.2. Let $b \not\equiv 1 \pmod{m}$ and let w be a BS factor of $\mathbf{t}_{b,m}$ such that $1 \leq |w| < 2b$. If $\Theta \in D_m$ is the unique antimorphism such that $w = \Theta(w)$, then the values $\mathrm{b}(w)$ and $\#\mathrm{Pext}_\Theta(w)$ are shown in Table 1 below.

$\|w\|$	$\mathrm{b}(w)$	$\#\mathrm{Pext}_\Theta(w)$
$1 \leq \|w\| \leq b-1$	0	1
$\|w\| = b$	1	2
$b+1 \leq \|w\| \leq 2b-2$	0	1
$\|w\| = 2b-1$	-1	0

Tab. 1. The bilateral order and the number of palindromic extensions of a BS factor w of $\mathbf{t}_{b,m}$ according to its length $|w|$ in the aperiodic case $b \not\equiv 1 \pmod{m}$. $\Theta \in D_m$ is the antimorphism such that $\Theta(w) = w$.

P r o o f. Let w, $0 < |w| < 2b$, be a BS factor and let s denote its length. It follows from Property XI that there exists $k \in \mathbb{Z}_m$ such that

$$w = k(k+1)\ldots(k+s-1).$$

Since $w = \Pi_k\left(01\ldots(s-1)\right)$, thanks to Property VIII we may take $w = 01\ldots(s-1)$.

Let $\Theta \in D_m$ be the unique antimorphism such that $w = \Theta(w)$. From the form of w it follows that $\Theta = \Psi_{s-1}$.

We discuss the following cases distinguished by the length s of w.

a) $s = 2b-1$.
One can see that w has exactly 2 ancestors: the words $0b$ and $(m-1)(b-1)$. (Both $0b$ and $(m-1)(b-1)$ belong to $\mathcal{L}_2(\mathbf{t}_{b,m})$ since $\pi^{q-1}(b-1)b = 0b$ and $\pi^{q-1}(b-2)(b-1) = (m-1)(b-1)$.) The only pairs of letters x and y such that $xwy \in \mathcal{L}(\mathbf{t}_{b,m})$ are $m-1+b-1$ and $2b-1$, and $m-1$ and b. Therefore, $\mathrm{b}(w) = 2-2-2+1 = -1$. Since $\Theta = \Psi_{2b-2}$, one can see that no extension xwy is a Θ-palindrome, i.e., $\#\mathrm{Pext}_\Theta(w) = 0$.

b) $b+1 \leq s \leq 2b-2$.
One can deduce that w has $2b-s+1$ ancestors, namely $(s-2b+i)(s-b+i)$ for $0 \leq i < 2b-s$. (Again, all these words are factors of $\mathbf{t}_{b,m}$ since $\pi^{q-1}(i-1) = i-b$ for all i.) The only extensions xwy appearing in $\mathcal{L}(\mathbf{t}_{b,m})$ are $(m-1)w(s-b+1)$, $(m-1)ws$, and $(m+b-2)ws$. Thus we have $\mathrm{b}(w) = 3-2-2+1 = 0$ and $\#\mathrm{Pext}_\Theta(w) = 1$ ($(m-1)ws$ is a Θ-palindrome).

c) $s = b$.
The ancestors of w are the factors 0 and $(i-b)i$ for $0 < i < b$. The extensions xwy are $(m-1)wb$, $\pi^\ell(-1)w1$, and $(m+b-2)w\pi^{-\ell}(1)$, where $0 \leq \ell < q$. Therefore, $\mathrm{b}(w) = 2q-q-q+1 = 1$. Since the only Θ-palindromes are $(m-1)wb$ and $(m+b-2)w1$, we have $\#\mathrm{Pext}_\Theta(w) = 2$.

d) $2 \leq s \leq b-1$.
The ancestors are the factors $m-1, \ldots, m+s-b$, the factor 0, and the factors $(i-b)i$ for $0 < i < s$. The extensions xwy are $\pi^\ell(-1)ws$ and $(-1)w\pi^{-\ell}(s)$ where $0 \leq \ell < q$. Thus, $\mathrm{b}(w) = (2q-1)-q-q+1 = 0$ and since $(m-1)ws$ is the only Θ-palindromic extension, we have $\#\mathrm{Pext}_\Theta(w) = 1$.

e) $s = 1$.
One can see that the extensions are $\pi^\ell(-1)w1$ and $(-1)w\pi^{-\ell}(1)$, where $0 \leq \ell < q$. Thus, $\mathrm{b}(w) = 0$ and $\#\mathrm{Pext}_\Theta(w) = 1$ as in the previous case.

□

Corollary 4.3. If $b \not\equiv 1 \pmod{m}$ and w is a non-empty BS factor of $\mathbf{t}_{b,m}$, then

1. there exists a unique antimorphism $\Theta \in D_m$ such that $\Theta(w) = w$;
2. $\mathrm{b}(w) = \#\mathrm{Pext}_\Theta(w) - 1$.

Theorem 4.4. The word $\mathbf{t}_{b,m}$ is D_m-rich.

P r o o f. First, let $b \not\equiv 1 \pmod{m}$. We show that

$$\Delta\mathcal{C}(n) + 2m = \sum_{\substack{\Theta \in D_m \\ \Theta \text{ antimorphism}}} \Big(\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n+1)\Big) \qquad \text{for all } n \geq 1. \tag{4}$$

Note that $\#D_m = 2m$.

First, we show the relation (4) for $n = 1$. It is clear that $\mathcal{C}(1) = m$ and $\mathcal{C}(2) = qm$. Thus, the left side equals $qm - m + 2m = qm + m$. According to Properties II and IV, it is clear that $\sum_{\substack{\Theta \in D_m \\ \Theta \text{ antimorphism}}} \mathcal{P}_\Theta(1) = m$ and $\sum_{\substack{\Theta \in D_m \\ \Theta \text{ antimorphism}}} \mathcal{P}_\Theta(2) = qm$. Therefore, the right side equals $qm + m$.

To show the relation (4), we are going to verify that for all $n \geq 1$ the difference of the left sides for indices $n+1$ and n equals the difference of the right sides for the same indices. In other words, we are going to show that

$$\Delta\mathcal{C}(n+1) - \Delta\mathcal{C}(n) = \sum_{\substack{\Theta \in D_m \\ \Theta \text{ antimorphism}}} \Big(\mathcal{P}_\Theta(n+2) - \mathcal{P}_\Theta(n)\Big) \tag{5}$$

for all n.

According to the equation (3), the left side can be written as

$$\Delta^2\mathcal{C}(n) = \sum_{\substack{w\in\mathcal{L}_n(\mathbf{t}_{b,m})\\ w \text{ BS}}} \mathrm{b}(w)$$

and for the right side we can use

$$\mathcal{P}_\Theta(n+2) - \mathcal{P}_\Theta(n) = \sum_{\substack{w\in\mathcal{L}_n(\mathbf{t}_{b,m})\\ w=\Theta(w)}} \big(\#\mathrm{Pext}_\Theta(w) - 1\big).$$

Using the fact that a non-bispecial Θ-palindrome has exactly one Θ-palindromic extension and Corollary 4.3, the relation (5) holds.

If $\mathbf{t}_{b,m}$ is periodic, i. e., $b \equiv 1 \pmod m$, the proof of the relation (5) can be done in a very similar way and is left to the reader.

Finally, according to Proposition 1.1 and Property VI, $\mathbf{t}_{b,m}$ is D_m-rich. □

5. FACTOR COMPLEXITY

To our knowledge, the factor complexity of the Thue–Morse word $\mathbf{t}_{2,2}$ was described in 1989 independently in [4] and [8] and of $\mathbf{t}_{2,m}$ in [15].

In the aperiodic case, to calculate the factor complexity, one can use the equality (3) and Lemmas 4.1 and 4.2. Table 2 shows the result: $\Delta\mathcal{C}(n)$ and $\mathcal{C}(n)$. In the periodic case, the factor complexity is trivial: $\mathcal{C}(n) = m$ for all $n > 0$.

n	$\Delta\mathcal{C}(n)$	$\mathcal{C}(n)$
0	$m-1$	1
1	$qm-m$	m
$2 \le n \le b$	$qm-m$	$qm(n-1)-m(n-2)$
$b^k+1+\ell$ $k\ge 1,\ 0\le \ell < b^k-b^{k-1}$	qm	$qm(n-1)-m(b^k-b^{k-1})$
$(2b-1)b^{k-1}+1+\ell$ $k\ge 1,\ 0\le\ell< b^{k+1}-2b^k+b^{k-1}$	$qm-m$	$qm(n-1)-m(b^k-b^{k-1}+\ell)$

Tab. 2. Values of $\Delta\mathcal{C}(n)$ and $\mathcal{C}(n)$ of the generalized Thue–Morse word $\mathbf{t}_{b,m}$ for the aperiodic case $b \not\equiv 1 \pmod m$.

ACKNOWLEDGMENTS

This work was supported by the Czech Science Foundation grant GA ČR 201/09/0584, by the grants MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic, and by the grant of the Grant Agency of the Czech Technical University in Prague grant No. SGS11/162/OHK4/3T/14.

(Received July 11, 2011)

REFERENCES

[1] J.-P. Allouche and J. Shallit: Sums of digits, overlaps, and palindromes. Discrete Math. Theoret. Comput. Sci. *4* (2000), 1–10.

[2] P. Baláži, Z. Masáková, and E. Pelantová: Factor versus palindromic complexity of uniformly recurrent infinite words. Theoret. Comput. Sci. *380* (2007), 3, 266–275.

[3] L. Balková: Factor frequencies in generalized Thue–Morse words. Kybernetika *48* (2012), 3, 371–385.

[4] S. Brlek: Enumeration of factors in the Thue–Morse word. Discrete Appl. Math. *24* (1989), 1–3, 83–96.

[5] S. Brlek, S. Hamel, M. Nivat, and C. Reutenauer: On the palindromic complexity of infinite words. Internat. J. Found. Comput. *15* (2004), 2, 293–306.

[6] M. Bucci, A. De Luca, A. Glen, and L. Q. Zamboni: A connection between palindromic and factor complexity using return words. Adv. Appl. Math. *42* (2009), no. 1, 60–74.

[7] J. Cassaigne: Complexity and special factors. Bull. Belg. Math. Soc. Simon Stevin *4* (1997), 1, 67–88.

[8] A. de Luca and S. Varricchio: Some combinatorial properties of the Thue–Morse sequence and a problem in semigroups. Theoret. Comput. Sci. *63* (1989), 3, 333–348.

[9] X. Droubay, J. Justin, and G. Pirillo: Episturmian words and some constructions of de Luca and Rauzy. Theoret. Comput. Sci. *255* (2001), 1–2, 539–553.

[10] A. Frid: Applying a uniform marked morphism to a word. Discrete Math. Theoret. Comput. Sci. *3* (1999), 125–140.

[11] A. Glen, J. Justin, S. Widmer, and L. Q. Zamboni: Palindromic richness. European J. Combin. *30* (2009), 2, 510–531.

[12] E. Pelantová and Š. Starosta: Languages invariant under more symmetries: overlapping factors versus palindromic richness. To appear in Discrete Math., preprint available at `http://arxiv.org/abs/1103.4051` (2011).

[13] E. Prouhet: Mémoire sur quelques relations entre les puissances des nombres. C. R. Acad. Sci. Paris *33* (1851), 225.

[14] Š. Starosta: On theta-palindromic richness. Theoret. Comput. Sci. *412* (2011), 12–14, 1111–1121.

[15] J. Tromp and J. Shallit: Subword complexity of a generalized Thue–Morse word. Inf. Process. Lett. (1995), 313–316.

Štěpán Starosta, FIT, Czech Technical University in Prague, Thákurova 9 160 00 Praha 6 and FNSPE, Czech Technical University in Prague, Trojanova 13 120 00 Praha 2. Czech Republic.
e-mail: stepan.starosta@fit.cvut.cz