Discrete Mathematics and Theoretical Computer Science DMTCS vol. **9:1**, 2007, 293–304

On the Critical Exponent of Generalized Thue-Morse Words

Alexandre B.-Massé[†], Srečko Brlek[‡], Amy Glen[§] and Sébastien Labbé[¶]

Laboratoire de Combinatoire et d'Informatique Mathématique
Université du Québec à Montréal
Case Postale 8888, succursale Centre-ville, Montréal (QC) CANADA H3C 3P8

received 18th June 2007, revised 22nd October 2007, accepted 15th October 2007.

For certain generalized Thue-Morse words $\mathtt{t}$, we compute the *critical exponent*, i.e., the supremum of the set of rational numbers that are exponents of powers in $\mathtt{t}$, and determine exactly the occurrences of powers realizing it.

Keywords: Thue-Morse, critical exponent, occurrences.

1 Introduction

It is a well-known fact that the Norwegian mathematician Axel Thue (1863–1922) was the first to explicitly construct and study the combinatorial properties of an infinite *overlap-free word* over a 2-letter alphabet, obtained as the fixpoint of the morphism $\mu : \{a, b\}^* \to \{a, b\}^*$ defined by $\mu(a) = ab; \mu(b) = ba$:

$$\mu(\mathbf{m}) = \mathbf{m} = abbabaabbaababba \cdots .$$

For a modern account of his papers, see Berstel [3]. Rediscovered by M. Morse in 1921 in the study of symbolic dynamics, this overlap-free word is now called the *Thue-Morse word*. This "ubiquitous" sequence, already implicit in a memoir by Prouhet [17] in 1851, appears in various fields, such as combinatorics on words, symbolic dynamics, differential geometry, number theory, and mathematical physics, as surveyed by Allouche and Shallit [1]. That survey also mentions some generalizations of the Thue-Morse word, and recently other ones were considered in [2, 8]. In particular, the following result was established in [2].

Proposition 1.1 [2] *Let $m \geq 1$, $b \geq 2$ be integers and let $s_b(n)$ denote the sum of the digits in the base b representation of $n \in \mathbb{N}$. Then the infinite word $\mathbf{t}_{b,m} := (s_b(n) \bmod m)_{n\geq 0}$ over the alphabet $\Sigma_m := \{0, 1, \ldots, m-1\}$ is overlap-free if and only if $b \leq m$.* □

[†] Supported by a NSERC (Canada) scholarship.
[‡] Research supported by a NSERC (Canada) grant.
[§] Supported by CRM-ISM-LaCIM postdoctoral fellowship (Montréal, Canada).
[¶] Supported by a NSERC (Canada) scholarship.

It was also shown in [2] that the word $\mathbf{t}_{b,m}$ contains arbitrarily long squares, which extends a result previously established by Brlek [6] for $\mathbf{m}$. Moreover, it was mentioned ([2], p. 8) that,

> *It would be interesting to determine the largest (fractional) power that occurs in the sequence $\mathbf{t}_{b,m}$. For $b \leq m$, we already know that 2 is sharp.*

We solve this problem here. Specifically, we study the family $\mathcal{T}$ of generalized Thue-Morse words consisting of the words $\mathbf{t}_{b,m}$ as well as letter-renamings of them (see Section 2). For any $\mathbf{t} \in \mathcal{T}$, we compute the *critical exponent* (i.e., the supremum of the set of rational numbers that are exponents of powers in $\mathbf{t}$) and determine exactly the occurrences of powers realizing it, both in terms of b and m. A noteworthy fact is that the critical exponents of generalized Thue-Morse words are always realized, which is not necessarily true in general (see Remarks 2.1 and 2.4).

The next section contains all of the basic terminology on words, borrowed mainly from Lothaire [14], along with the generalized Thue-Morse words. Section 3 contains the technical lemmas, establishing combinatorial properties used for proving our main results. The critical exponent (Theorem 4.4) is then computed in Section 4, and Section 5 deals with the occurrences of factors realizing it (Theorem 5.5).

Let us mention that the notions of 'fractional power' and 'critical exponent' have received growing attention in recent times, especially in relation to *Sturmian* and *episturmian words*; see for instance [4, 5, 7, 10, 11, 12, 13, 15, 18].

2 Definitions and notation

Let Σ denote a finite *alphabet*, i.e., a finite set of symbols called *letters*. A *finite word* over Σ is a finite sequence $w = w_0 w_1 \cdots w_{\ell-1}$, where each $w_i \in \Sigma$. We often write $w[i]$ for w_i, when i is a complicated formula. The *length* of w is $|w| = \ell$, and the *empty word* of length 0 is denoted by ε. The set of all finite words over Σ is denoted by Σ^*.

A word u is a *factor* of w if $w = pus$ for some words p, s. Moreover, such a word u is said to be a *proper factor* of w if $u \neq w$. We also say that u is a *prefix* (resp. *suffix*) of w if $p = \varepsilon$ (resp. $s = \varepsilon$). The number $|p|$ is called an *occurrence* of u in w, i.e., $|p|$ is the beginning position of an appearance of u in w. The set of all factors of a word w is denoted by $F(w)$. A word v is a *conjugate* of w if there exists a word u such that $uv = wu$.

An *overlap* w is a word of the form $w = auaua$ where a is a letter and u is a (possibly empty) word. For example, the English word *alfalfa* is an overlap and *banana* has the overlap *anana* as a suffix. A finite word is said to be *overlap-free* if it does not contain an overlap as a factor. Factors of the form uu are called *squares*.

The *rational power* of a word w is defined by $w^r = w^{\lfloor r \rfloor} p$ where r is a rational such that $r|w| \in \mathbb{N}$ and p is the prefix of w of length $(r - \lfloor r \rfloor)|w|$. For example, the word

$$cab\ \underline{abbacc}\ \underline{abbacc}\ \underline{abb}\ bca$$

contains a factor which is a $\frac{5}{2}$-power.

A (right) *infinite word* (or simply a *sequence*) $\mathbf{x}$ over Σ is a sequence indexed by $\mathbb{N}$ with values in Σ, i.e., $\mathbf{x} = x_0 x_1 x_2 \cdots$ where each $x_i \in \Sigma$. All of the terminology above naturally extends to infinite words.

An *ultimately periodic* infinite word can be written as $uv^\omega = uvvv\cdots$, for some $u, v \in \Sigma^*$, $v \neq \varepsilon$. If $u = \varepsilon$, then such a word is *periodic*. An infinite word that is not ultimately periodic is said to be *aperiodic*.

For any factor w of an infinite word $\mathbf{x}$, the *index* of w in $\mathbf{x}$ is given by the number

$$\textsc{Index}(w) = \max\{r \in \mathbb{Q} \mid w^r \in F(\mathbf{x})\},$$

if such a number exists; otherwise, w is said to have infinite index in $\mathbf{x}$. The *critical exponent* $E(\mathbf{x})$ of an infinite word is given by

$$E(\mathbf{x}) = \sup\{\textsc{Index}(w) \mid w \in F(\mathbf{x})\backslash\{\varepsilon\}\}.$$

It may be finite or infinite. A factor w of $\mathbf{x}$ is said to be a *critical factor* if its index realizes the critical exponent of $\mathbf{x}$, that is, when $\textsc{Index}(w) = E(\mathbf{x})$.

Remark 2.1 For any factor w of an infinite word $\mathbf{x}$, $\textsc{Index}(w)$ is always realized when finite, whereas the critical exponent $E(\mathbf{x})$, even when finite, is not always realized; in particular, an infinite word may contain no critical factors. For example, the critical exponent of the well-known *Fibonacci word* $\mathbf{f}$ is $E(\mathbf{f}) = 2 + \phi$, where $\phi = \frac{\sqrt{5}+1}{2}$ is the *golden ratio* (see [15]), but none of the factors of $\mathbf{f}$ realize $E(\mathbf{f})$.

A *morphism* is a function $\varphi : \Sigma^* \to \Sigma^*$ such that $\varphi(uv) = \varphi(u)\varphi(v)$ for all $u, v \in \Sigma^*$. For each letter $a \in \Sigma$, $\varphi(a)$ is called a *block*.

Generalized Thue-Morse words

There exist many generalizations of the Thue-Morse word. Here, we introduce a morphism based formulation which is more convenient for our purposes.

Definition 2.2 *Let $b \geq 2$, $m \geq 1$ be integers, Σ an alphabet of m letters, $\sigma : \Sigma \to \Sigma$ a cyclic permutation and $\mu : \Sigma^* \to \Sigma^*$ the morphism given by*

$$\mu(\alpha) = \prod_{i=0}^{b-1} \sigma^i(\alpha) = \sigma^0(\alpha)\sigma^1(\alpha)\sigma^2(\alpha)\cdots\sigma^{b-1}(\alpha).$$

Then the generalized Thue-Morse word $\mathbf{t}$, beginning with $\overline{\alpha} \in \Sigma$, is the infinite word given by $\mathbf{t} = \mu^\omega(\overline{\alpha}) = \lim_{n\to\infty}\mu^n(\overline{\alpha})$.

Hereafter, we study the family $\mathcal{T}$ of generalized Thue-Morse words $\mathbf{t}$, for all $b \geq 2$, $m \geq 1$ and $\overline{\alpha} \in \Sigma$.

Remark 2.3 The n-th letter of $\mathbf{t}$ is $\mathbf{t}[n] := \sigma^{s_b(n)}(\overline{\alpha})$ where $s_b(n)$ denotes the sum of the digits in the base b representation of $n \in \mathbb{N}$.

Example 1 Let $b = 5$, $m = 3$, $\Sigma = \{\triangle, \diamondsuit, \heartsuit\}$ and $\sigma : \triangle \mapsto \diamondsuit \mapsto \heartsuit \mapsto \triangle$. This gives the following morphism : $\mu(\triangle) = \triangle\diamondsuit\heartsuit\triangle\diamondsuit$, $\mu(\diamondsuit) = \diamondsuit\heartsuit\triangle\diamondsuit\heartsuit$ and $\mu(\heartsuit) = \heartsuit\triangle\diamondsuit\heartsuit\triangle$. By fixing $\overline{\alpha} = \diamondsuit$, we obtain the generalized Thue-Morse word $\mathbf{t} = \mu^\omega(\diamondsuit) = \diamondsuit\heartsuit\triangle\diamondsuit\heartsuit\heartsuit\triangle\diamondsuit\heartsuit\triangle\triangle\diamondsuit\heartsuit\triangle\diamondsuit\diamondsuit\heartsuit\triangle\diamondsuit\heartsuit\cdots$.

Example 2 With $\Sigma = \mathbb{Z}_m$, $\sigma(i) = (i+1) \bmod m$ and $\overline{\alpha} = 0$, Remark 2.3 implies that

$$\mathbf{t}[n] = \sigma^{s_b(n)}(0) = s_b(n) \bmod m, \tag{1}$$

which is the case proposed by Allouche and Shallit [2].

Example 3 With $\Sigma = \mathbb{Z}_m$, $\sigma(i) = (i + c) \bmod m$ for some integer $c > 0$ and $\overline{\alpha} = 0$, we have $\mu(0) = 0\,\overline{c}\,\overline{2c} \cdots \overline{(b-1)c}$ and

$$\mu(i) = i\,\overline{(i+c)}\,\overline{(i+2c)} \cdots \overline{i+(b-1)c} = i^b + \mu(0)$$

where $\overline{i} \equiv i \bmod m$. This is a subclass of the family of symmetric morphisms, defined by A. Frid [8]. That paper also contains an extension of Proposition 1.1 to more general words.

It is important to note that words in $\mathcal{T}$, up to letter renaming, are exactly those given in Equation (1). Our definition avoids modular arithmetic on integers, and simplifies proofs by using the combinatorial properties of $\mathcal{T}$ instead. For that purpose, we say that a word $w = w_0 w_1 \cdots w_{\ell-1} \in F(\mathbf{t})$ is *σ-cyclic* if $w_i = \sigma(w_{i-1})$ for $1 \leq i \leq \ell - 1$, or equivalently, if $w_i = \sigma^i(w_0)$ for $0 \leq i \leq \ell - 1$. As a consequence, blocks of $\mathbf{t}$ are σ-cyclic.

Example 4 With $\Sigma = \{a, b\}$, $\sigma : a \mapsto b \mapsto a$ and $\overline{\alpha} = a$, we get $\mu(a) = ab$, $\mu(b) = ba$ and

$$\mathbf{t} = \mu^\omega(a) = abbabaabbaababba \cdots$$

which is the original Thue-Morse word $\mathbf{m}$.

Remark 2.4 Since $\mathbf{t} \in \mathcal{T}$ is a fixpoint of a *uniform non-erasing* morphism, it follows immediately from Krieger's results in [13] that the critical exponent of $\mathbf{t}$ is either infinite (if $\mathbf{t}$ is periodic) or rational. Moreover, since $\mu(\alpha)$ and $\mu(\beta)$ neither begin nor end with a common word when $\alpha \neq \beta$, it also follows from [13] that the critical exponent, when finite, is reached (i.e., $\mathbf{t}$ contains critical factors).

3 Preliminary results

In this section, $\mathbf{t} \in \mathcal{T}$ denotes an infinite generalized Thue-Morse word, as given in Definition 2.2.

Lemma 3.1 [9, 16] *The word $\mathbf{t}$ is periodic if and only if $m \mid (b-1)$. More precisely, if $\mathbf{t}$ is periodic then*

$$\mathbf{t} = \left[\prod_{i=0}^{m-1} \sigma^i(\overline{\alpha}) \right]^\omega .$$

□

Lemma 3.2 *Let w be a σ-cyclic factor of $\mathbf{t}$ of length ℓ. If there exists an occurrence of w overlapping three consecutive blocks, then $\mathbf{t}$ is periodic.*

Proof: Suppose that there exists such an occurrence of w in $\mathbf{t}$. Since w is σ-cyclic and since every block is σ-cyclic, those three consecutive blocks, say β_1, β_2 and β_3, satisfy the fact that $\beta_1\beta_2\beta_3$ is σ-cyclic.

Let α_1, α_2 and α_3 be the first letters of β_1, β_2 and β_3. Then $\alpha_2 = \sigma^b(\alpha_1)$ and $\alpha_3 = \sigma^b(\alpha_2)$, and $\alpha_1\alpha_2\alpha_3 = \mu^{-1}(\beta_1\beta_2\beta_3)$ are three consecutive letters occurring in $\mathbf{t}$. This means, in particular, that either $\alpha_1\alpha_2$ or $\alpha_2\alpha_3$ occur in the same block. Hence, either $\alpha_2 = \sigma(\alpha_1)$ or $\alpha_3 = \sigma(\alpha_2)$, which implies that $\alpha_1 = \sigma^{b-1}(\alpha_1)$ or $\alpha_2 = \sigma^{b-1}(\alpha_2)$. Therefore, $m \mid (b-1)$ and by Lemma 3.1, $\mathbf{t}$ is periodic. □

Lemma 3.3 *Let $w = w_0w_1 \cdots w_{\ell-1}$ be a factor of $\mathbf{t}$ such that $b \not| \ell$.*

i) *If wp occurs in $\mathbf{t}$ where p is any prefix of w, then p is σ-cyclic.*

ii) *If w^e occurs in $\mathbf{t}$ for some rational $e > 2$, then w^e is σ-cyclic.*

Proof: Let p be any prefix of length ℓ' of w and assume $v = wp$ occurs in $\mathbf{t}$. Let v_i be the i-th letter of v, with $0 \leq i \leq \ell + \ell' - 1$. For $0 \leq i \leq \ell' - 2$, either v_iv_{i+1} or $v_{i+\ell}v_{i+\ell+1}$ is contained in the same block: if this was not the case, then b would divide $(i + \ell) - i = \ell$, contradicting our assumption. Since $v_iv_{i+1} = w_iw_{i+1} = v_{i+\ell}v_{i+\ell+1}$, we have $w_{i+1} = \sigma(w_i)$ so that the first result follows.

Since $w^e = www^{e-2}$, w is σ-cyclic from i). Let x be the conjugate of w such that $w_0x = ww_0$. Then, $w^e = w_0xxx^{e-2-1/\ell}$ and from i) x is σ-cyclic and so too is its factor $w_{\ell-1}w_0$. Hence, w^e is σ-cyclic, which ends the second part. □

Lemma 3.4 *Suppose $\mathbf{t}$ is aperiodic. If w is a factor of $\mathbf{t}$ of length $\ell \geq b$ with $b \not| \ell$, then $\text{INDEX}(w) \leq 2$.*

Proof: By contradiction, assume $v = www_0$ occurs in $\mathbf{t}$, where w_0 is the first letter of w. Lemma 3.3 implies that v is σ-cyclic. Moreover, $|v| = 2\ell + 1 > 2b$, which means that v overlaps at least three consecutive blocks of $\mathbf{t}$. Thus, by Lemma 3.2, $\mathbf{t}$ is periodic which contradicts our assumption. □

4 Critical exponent

In this section, we use exactly the same notation as previously. Before proving Theorem 4.4, we need a few additional facts.

The next two lemmas allow us to consider particular occurrences of factors of $\mathbf{t}$. We say that an occurrence i of w in $\mathbf{t}$ is *synchronized* if $b \mid i$ and $b \mid (i + |w|)$.

Lemma 4.1 *Suppose $w = w_0w_1 \cdots w_{\ell-1}$ is a factor of $\mathbf{t}$ such that $b \mid \ell$. Let p be a non-empty prefix of w such that $bq + r$ is an occurrence of w^np in $\mathbf{t}$, where $q, r \in \mathbb{N}$, $0 \leq r < b$ and $n \geq 1$. Let s be the possibly empty suffix of w such that $w = ps$. Moreover, let $bq' + r' = bq + r + |w^np|$, where $q', r' \in \mathbb{N}$, $0 \leq r' < b$. Then,*

i) *$v = \mathbf{t}[bq] \cdots \mathbf{t}[bq + r - 1]$ is a suffix of w,*

ii) *$u = \mathbf{t}[bq' + r'] \cdots \mathbf{t}[bq' + b - 1]$ is a prefix of sp,*

iii) *bq is a synchronized occurrence of vw^npu in $\mathbf{t}$.*

Proof: Let $v = v_0v_1 \cdots v_{r-1} = \mathbf{t}[bq] \cdots \mathbf{t}[bq + r - 1]$ and let $s = s_0s_1 \cdots s_{r-1}$ be the suffix of w of length r. We show that $v = s$. First note that vw_0 is contained in the block starting at position bq. Also, since $b \mid \ell$, sw_0 is contained in one block. Both remarks imply that vw_0 and sw_0 are σ-cyclic. Hence,

$$v_i = \sigma^{i-r}(w_0) = s_i \quad \text{for } 0 \leq i \leq r - 1,$$

which gives the result. The proof of ii) is symmetric to the proof of i), and iii) follows from i) and ii). □

Lemma 4.2 *Let w be a word of length ℓ such that $b \mid \ell$ and suppose w^e occurs in $\mathbf{t}$ for some rational $e > 1$. Then there exists a conjugate x of w and a rational $f \geq e$ such that both x and x^f have a synchronized occurrence in $\mathbf{t}$.*

Proof: Let p be the non-empty prefix of w such that $w^e = w^n p$ where $n \in \mathbb{N}^+$ and let s be the suffix of w such that $w = ps$. Also, let $bq + r$ be an occurrence of w^e in $\mathbf{t}$, where $q, r \in \mathbb{N}$ and $0 \leq r < b$. By Lemma 4.1, there exist a suffix v of w and a prefix u of sp such that bq is a synchronized occurrence of $vw^e u$ in $\mathbf{t}$.

Now let y be the prefix of w such that $w = yv$, and define $x = vy$. We have $vw^e u = v(yv)^n pu = (vy)^n vpu = x^n vpu$. Moreover, vpu is a prefix of $vpsps = vw^2 = v(yv)^2 = x^2 v$ so that $vw^e u = x^n vpu = x^f$ for some rational f. Finally,

$$f = \frac{\ell n + |vpu|}{\ell} \geq \frac{\ell n + |p|}{\ell} = e,$$

which ends the proof, since x and x^f both satisfy the required conditions. □

The next lemma deals with factors w of $\mathbf{t}$ of length ℓ not divisible by b.

Lemma 4.3 *Suppose $\mathbf{t}$ is aperiodic and let w be a factor of $\mathbf{t}$ of length ℓ such that $b \nmid \ell$. Then*

$$\textsc{Index}(w) \leq \begin{cases} 2b/m & \text{if } b > m, \\ 2 & \text{if } b \leq m. \end{cases}$$

Proof: We distinguish three cases according to ℓ, b and m.

Case $1 : \ell > b$. By Lemma 3.4, we have $\textsc{Index}(w) \leq 2$.

Case $2 : \ell < b$ *and* $b > m$. Let w_0 be the first letter of w and $e = \textsc{Index}(w)$. If $e > 2$, then by Lemma 3.3 w^e is σ-cyclic. In particular, $w_0 = \sigma^\ell(w_0)$, i.e., $m \mid \ell$ and $\ell \geq m$. However, by Lemma 3.2, we know that $|w^e| \leq 2b$. Therefore,

$$e = \frac{|w^e|}{\ell} \leq \frac{2b}{\ell} \leq \frac{2b}{m} \tag{2}$$

which ends this part.

Case $3 : \ell < b$ *and* $b \leq m$. As above, let $e = \textsc{Index}(w)$ and suppose $e > 2$, which means w^e is σ-cyclic and $m \mid \ell$. The latter statement is not possible since $1 \leq \ell < b \leq m$. We conclude in this case that $\textsc{Index}(w) \leq 2$. □

We are now ready to prove the main theorem of this section, which gives the critical exponent of $\mathbf{t}$.

Theorem 4.4 *The critical exponent of $\mathbf{t}$ is given by*

$$E(\mathbf{t}) = \begin{cases} \infty & \text{if } m \mid (b-1), \\ 2b/m & \text{if } m \nmid (b-1) \text{ and } b > m, \\ 2 & \text{if } b \leq m. \end{cases}$$

Proof: If $m \mid (b-1)$, then $E(\mathbf{t}) = \infty$ since $\mathbf{t}$ is periodic by Lemma 3.1. Now suppose $m \nmid (b-1)$ and let

$$E_{b,m} = \begin{cases} 2b/m & \text{if } b > m, \\ 2 & \text{if } b \le m. \end{cases}$$

First we show that $E(\mathbf{t}) \ge E_{b,m}$. If $b \le m$, it is easy to see that there is a square in the first two blocks of $\mathbf{t}$, as noticed in [2]. On the other hand, if $b > m$, there exists a $2b/m$-power in $\mathbf{t}$. Indeed, let β_1 and β_2 be the blocks starting at positions $b^m - b$ and b^m respectively. From Remark 2.3, the last letter of β_1 is $\mathbf{t}[b^m-1] = \sigma^{s_b(b^m-1)}(\overline{\alpha}) = \sigma^{m(b-1)}(\overline{\alpha}) = \overline{\alpha}$ and the first letter of β_2 is $\mathbf{t}[b^m] = \sigma^{s_b(b^m)}(\overline{\alpha}) = \sigma(\overline{\alpha})$. Therefore, the whole factor $\beta_1\beta_2$ of length $2b$ is σ-cyclic, and hence $\beta_1\beta_2 = w^{2b/m}$ where w is the prefix of length m of β_1. Thus $E(\mathbf{t}) \ge E_{b,m}$.

We now prove that $E(\mathbf{t}) = E_{b,m}$ by showing that $\textsc{Index}(w) \le E_{b,m}$ for any factor w of $\mathbf{t}$. Suppose w is a factor of $\mathbf{t}$ of length $\ell = b^i N$, where $b \nmid N$ for some $i, N \in \mathbb{N}$. The proof proceeds by induction on i.

BASIS. If $i = 0$ then $b \nmid \ell$. This case is proved in Lemma 4.3.

HYPOTHESIS. We assume $\textsc{Index}(w) \le E_{b,m}$ for all factors w of $\mathbf{t}$ of length $\ell = b^i N$.

INDUCTION. Let w be a factor of $\mathbf{t}$ of length $\ell = b^{i+1}N$. Assume $e = \textsc{Index}(w) > 1$. Since $b \mid \ell$, from Lemma 4.2 we know that there exists a factor x of length ℓ and a rational $f \ge e$ such that both x and x^f are synchronized with the blocks. Let p be the proper prefix of x such that $x^f = x^n p$ where $n \in \mathbb{N}^+$. Then, $\mu^{-1}(x^f) = \mu^{-1}(x^n p) = \left(\mu^{-1}(x)\right)^n \mu^{-1}(p)$ is a factor of $\mathbf{t}$. Since $\mu^{-1}(p)$ is a prefix of $\mu^{-1}(x)$, we obtain

$$\textsc{Index}\left(\mu^{-1}(x)\right) \ge n + \frac{|\mu^{-1}(p)|}{|\mu^{-1}(x)|} = n + \frac{|p|/b}{|x|/b} = n + \frac{|p|}{|x|} = \textsc{Index}(x). \tag{3}$$

By the induction hypothesis, we have

$$\textsc{Index}(w) \le f \le \textsc{Index}(x) \le \textsc{Index}\left(\mu^{-1}(x)\right) \le E_{b,m}.$$

□

5 Occurrences of critical factors

In this section, we assume that $\mathbf{t}$ is *aperiodic*, i.e., $m \nmid (b-1)$. We say that $b > m$ is the *overlap case* and that $b \le m$ is the *square case* and denote by e the critical exponent of $\mathbf{t}$.

Here, we describe the *occurrences* and the *lengths* of the critical factors of $\mathbf{t}$.

Lemma 5.1 *Let w be a critical factor of $\mathbf{t}$ of length ℓ. Then the following properties hold.*

i) *$\mu(w)$ is a critical factor of $\mathbf{t}$.*

ii) *If $b \mid \ell$, then $\mu^{-1}(w)$ is a critical factor of $\mathbf{t}$.*

Proof: Property i) is trivial. For ii), suppose $b \mid \ell$. If w^e is not synchronized with the blocks, then Lemma 4.1 contradicts the maximality of e. Therefore, both w and w^e are synchronized and their preimages under μ are well-defined. Finally, from inequality (3), we obtain the index of the preimage. □

In view of Lemma 5.1, it is enough to consider only the case $b \nmid \ell$ when describing the occurences of critical factors of length ℓ in $\mathbf{t}$.

Lemma 5.2 *Let w be a critical factor of $\mathbf{t}$ of length ℓ such that $b \not| \ell$.*

i) *In the overlap case, $\ell = m$ and $w^e = \beta_1\beta_2$, where β_1, β_2 are two consecutive blocks.*

ii) *In the square case, write $w^2 = w^{(1)}w^{(2)}$. Then $w^{(2)}$ occurs at the beginning of a block.*

Proof: *Overlap case* ($b > m$). We consider separately the cases $\ell < b$ and $\ell > b$. If $\ell < b$, then we have $\ell = m$ and $|w^e| = 2b$ by inequality (2). It follows that w^e is composed of two consecutive blocks. Otherwise, if w^e overlaps three consecutive blocks, then from Lemma 3.2 $\mathbf{t}$ is periodic, a contradiction.

On the other hand, if $\ell > b$, then from the last section we know that $\text{INDEX}(w) = 2b/m > 2$. This contradicts Lemma 3.4; hence there is no such critical factor.

Square case ($b \leq m$). Again, we distinguish two cases. First assume $\ell < b$. Note that each block of $\mathbf{t}$ contains distinct letters. Also, w^2 is a factor of $\mathbf{t}$ and, from Lemma 3.3, w is σ-cyclic. If α_i denotes the i-th letter of w^2 with $0 \leq i < 2\ell$, then $\alpha_i = \alpha_{i+\ell}$ for all $0 \leq i < \ell$. This implies that α_0 and α_ℓ are not in the same block so that there exist two consecutive letters α_{i-1} and α_i, where $1 \leq i \leq \ell$, which are not contained in the same block. Suppose $1 \leq i < \ell$. Then, α_i and $\alpha_{i+\ell}$ belong to the same block since $\ell < b$. This is a contradiction because $\alpha_i = \alpha_{i+\ell}$. We conclude that the pair $\alpha_{\ell-1}$ and α_ℓ is not contained in the same block. In other words, $w^{(2)}$ occurs at the beginning of a block.

Lastly, consider the case $\ell > b$. From Lemma 3.3, w is σ-cyclic. Suppose $w^{(2)}$ does not occur at the beginning of a block. Then the pair $w_{\ell-1}w_0$ formed by the last and the first letter of w is in the same block and is σ-cyclic. Hence, the whole factor w^2 of length $2\ell > 2b$ is σ-cyclic. This factor overlaps three consecutive blocks so $\mathbf{t}$ is periodic from Lemma 3.2, a contradiction. □

It has already been noticed that squares of certain factors of length 3 appear in the Thue-Morse word (see [6] for example). The following lemma proves the uniqueness of this fact.

Lemma 5.3 *The Thue-Morse word $\mathbf{m}$ is the unique word in $\mathcal{T}$ containing a critical factor w of length $\ell > b$ such that $b \not| \ell$. Moreover, $\ell = 3$.*

Proof: Suppose that w is critical factor of $\mathbf{t}$ of length $\ell > b$ such that $b \not| \ell$. In the proof of Lemma 5.2, we saw that in the overlap case there is no such critical factor. In the square case, we know also from Lemma 5.2 that if $w^2 = w^{(1)}w^{(2)}$ then $w^{(2)}$ occurs at the beginning of a block. Moreover, Lemma 3.3 implies that w is σ-cyclic. Since $\ell > 2b$ implies that w overlaps three consecutive blocks, it follows from Lemma 3.2 that $b < \ell < 2b$.

Hence w^2 overlaps exactly four consecutive blocks, say β_1, β_2, β_3 and β_4. Let α_i be the first letter of the block β_i, $i \in \{1, 2, 3, 4\}$, and let $r = 2b - \ell$ be the distance between α_1 and the first letter of w^2. Then $\alpha_3 = w_0 = \sigma^r(\alpha_1)$. Also, since w is σ-cyclic, we have $\alpha_2 = \sigma^b(\alpha_1)$ and $\alpha_4 = \sigma^b(\alpha_3) = \sigma^{b+r}(\alpha_1)$. Moreover, it follows that $\alpha_1\alpha_2\alpha_3\alpha_4 = \mu^{-1}(\beta_1\beta_2\beta_3\beta_4)$ is a factor of $\mathbf{t}$. Assume $\alpha_1\alpha_2$ is in the same block, then $\alpha_2 = \sigma(\alpha_1)$; but then, since $\alpha_2 = \sigma^b(\alpha_1)$, we get that $\alpha_1 = \sigma^{b-1}\alpha_1$, and so $m \mid (b-1)$, which is in contradiction with $b \leq m$. Assuming $\alpha_3\alpha_4$ is in the same block gives the same contradiction. Hence, neither $\alpha_1\alpha_2$ nor $\alpha_3\alpha_4$ are in the same block. We conclude from the last observation that $b = 2$ since $\alpha_2\alpha_3 = \sigma^b(\alpha_1)\sigma^r(\alpha_1)$ must form a block. Therefore $\ell = 3$ and $r = 1$. Also, m divides $b+1-r = b$ and $m = b = 2$ because the block $\alpha_2\alpha_3$ is σ-cyclic, which ends the proof.

Note that $\alpha_2 = \alpha_1$, $\alpha_4 = \alpha_3$, $\alpha_3 = \sigma(\alpha_1)$ and $w = \sigma(\alpha_1)\alpha_1\sigma(\alpha_1)$, which has already been noticed in [6]. □

We now prove the following lemma which is a more general result than Lemma 5 in [2].

Lemma 5.4 *Let $k, N \in \mathbb{N}^+$ be such that $b \nmid k$ and $1 \leq N < b$. Then,*

$$s_b(kb^q - N) - s_b(kb^q) = q(b-1) - N \quad \text{for any } q \in \mathbb{N}^+.$$

Proof: By direct computation,

$$\begin{aligned} s_b(kb^q - N) &= s_b\left((k-1)b^q + (b^q - b) + (b - N)\right) \\ &= s_b\left[(k-1)b^q + (b-1)\sum_{i=1}^{q-1} b^i + \ (b-N)\right] \\ &= s_b(k-1) + (q-1)(b-1) + (b-N) \\ &= s_b(k) - 1 + q(b-1) - N + 1 \\ &= s_b(kb^q) + q(b-1) - N, \end{aligned}$$

as desired. □

From *Bezout's Identity*, we know that for any $x, y \in \mathbb{Z}$ there exist $s, t \in \mathbb{Z}$ such that $\gcd(x, y) = sx + ty$ where s can be chosen positive. Let $g \in \mathbb{Z}$ and

$$S^g_{x,y} = \{s \in \mathbb{N}^+ \mid g = sx + ty, t \in \mathbb{Z}\}.$$

The set $S^g_{x,y}$ is non-empty when $\gcd(x, y) \mid g$. Moreover, let us define the following three sets:

$$\begin{aligned} A &= \left\{ kb^q - b \ \mid \ k \in \mathbb{N}^+, b \nmid k \text{ and } q \in S^m_{b-1,m} \right\}, \\ B_N &= \left\{ kb^q - N \ \mid \ k \in \mathbb{N}^+, b \nmid k \text{ and } q \in S^N_{b-1,m} \right\}, \\ C &= (8 \cdot B_1 + 3) \cup (8 \cdot B_1 + 7). \end{aligned}$$

We are now ready to state the main theorem of this section, which gives the set of occurrences where critical factors in an aperiodic generalized Thue-Morse word $\mathtt{t}$ realize the critical exponent $e = E(\mathtt{t})$.

Theorem 5.5 *If w is a critical factor of $\mathtt{t}$ of length $\ell = Nb^i$ such that $b \nmid N$, then the set of occurrences of w^e in $\mathtt{t}$ is*

$$\begin{cases} b^i A & \text{if } b > m, \\ b^i B_N & \text{if } b \leq m \text{ and } (b,m) \neq (2,2), \\ b^i(B_1 \cup C) & \text{if } b = m = 2. \end{cases}$$

The proof of Theorem 5.5 follows easily from Lemma 5.1 and the next lemma.

Lemma 5.6 *Let w be a critical factor of $\mathtt{t}$ of length ℓ such that $b \nmid \ell$. Then, the set of occurrences of w^e in $\mathtt{t}$ is*

$$\begin{cases} A & \text{if } b > m, \\ B_\ell & \text{if } b \leq m \text{ and } (b,m) \neq (2,2), \\ B_1 \cup C & \text{if } b = m = 2. \end{cases}$$

Proof: *Overlap case* ($b > m$). From Lemma 5.2, we know that $w^e = \beta_1\beta_2$, where β_1 and β_2 are blocks. Suppose p is an occurence of w^e and let kb^q be the starting position of β_2 where $b \not| k$ and $k, q \in \mathbb{N}^+$. Then $p = kb^q - b$. Now, let $\gamma_1 = \mathbf{t}[kb^q - 1]$ be the last letter of β_1 and $\alpha_2 = \mathbf{t}[kb^q]$ be the first letter of β_2. Since $w^e = \beta_1\beta_2$ is σ-cyclic from Lemma 3.3, we have

$$\sigma(\sigma^{s_b(kb^q-1)}(\overline{\alpha})) = \sigma(\gamma_1) = \alpha_2 = \sigma^{s_b(kb^q)}(\overline{\alpha}).$$

That is,

$$\begin{aligned}\overline{\alpha} &= \sigma^{s_b(kb^q-1)-s_b(kb^q)+1}(\overline{\alpha})\\ &= \sigma^{q(b-1)-1+1}(\overline{\alpha})\\ &= \sigma^{q(b-1)}(\overline{\alpha}),\end{aligned}$$

and hence $m \mid q(b-1)$. Therefore, there exists $t \in \mathbb{Z}$ such that $mt = q(b-1)$. In particular, we have $m = q(b-1) - (t-1)m$ where $q \in S^m_{b-1,m}$ which ends the first part.

Square case ($b \leq m$). If we write $w^2 = w^{(1)}w^{(2)}$, then from Lemma 5.2 we know that $w^{(2)}$ occurs at the beginning of a block. We distinguish two cases.

First consider the case $\ell < b$. Suppose p is an occurrence of w^2 and let kb^q be the starting position of $w^{(2)}$ (i.e., the second block) where $b \not| k$ and $k, q \in \mathbb{N}^+$. Then $p = kb^q - \ell$, and hence

$$\sigma^{s_b(kb^q-\ell)}(\overline{\alpha}) = \sigma^{s_b(kb^q)}(\overline{\alpha}),$$

that is

$$\begin{aligned}\overline{\alpha} &= \sigma^{s_b(kb^q-\ell)-s_b(kb^q)}(\overline{\alpha})\\ &= \sigma^{q(b-1)-\ell}(\overline{\alpha}).\end{aligned}$$

The last equation holds if and only if m divides $q(b-1) - \ell$, in which case there exists $t \in \mathbb{Z}$ such that $\ell = q(b-1) - tm$, that is $q \in S^\ell_{b-1,m}$.

Now consider $\ell > b$. From Lemma 5.3, we know that such critical factors occur only in the Thue-Morse word $\mathbf{m}$ and that $w^2 \in \{abaaba, babbab\}$. By the fixpoint property,

$$\mathbf{m} = \mu^3(\mathbf{m}) = \mu^3(a)\mu^3(b)\mu^3(b)\mu^3(a)\mu^3(b)\cdots,$$

so that it can be factorized into blocks of length 8. Since $|w^2| = 6$, w^2 is a factor of two consecutive blocks of length 8. Those two blocks are either $\mu^3(\alpha\beta)$ or $\mu^3(\alpha\alpha)$, where $\alpha, \beta \in \{a, b\}$, $\alpha \neq \beta$. We observe that

$$\mu^3(\alpha\beta) = \alpha\beta\beta\alpha\beta\alpha\alpha\beta\beta\alpha\alpha\beta\alpha\beta\beta\alpha$$

contains no square of length 6. On the other hand, w^2 occurs in

$$\mu^3(\alpha\alpha) = \alpha\beta\beta\underline{\alpha}\beta\alpha\alpha\underline{\beta}\alpha\beta\beta\alpha\beta\alpha\alpha\beta$$

at position 3 or 7. Thus, the set of occurrences of squares of length 6 is exactly $(8 \cdot B_1 + 3) \cup (8 \cdot B_1 + 7)$, since B_1 enumerates the occurrences of squares of single letters. □

Example 5 Recall the generalized Thue-Morse word $\mathbf{t} = \diamondsuit\heartsuit\triangle\diamondsuit\heartsuit\heartsuit\triangle\cdots$ that was defined on $\Sigma = \{\triangle, \diamondsuit, \heartsuit\}$ in Example 1. From Theorem 4.4, $E(\mathbf{t}) = 2b/m = 10/3$. From Theorem 5.5, we compute $S^3_{5-1,3} = \{3, 6, 9, 12, 15, \ldots\}$ and obtain the set of occurrences $A = \{120, 245, 370, 495, 745, \ldots\}$ of critical factors of length 3. Here are the first two critical factors of $\mathbf{t}$:

$$\mathbf{t} = \overset{\overset{0}{\downarrow}}{\diamondsuit}\overset{\overset{1}{\downarrow}}{\heartsuit}\overset{\overset{2}{\downarrow}}{\triangle}\overset{\overset{3}{\downarrow}}{\diamondsuit}\ \heartsuit\heartsuit\triangle\cdots\diamondsuit\heartsuit\triangle\ \underbrace{\overset{\overset{120}{\downarrow}}{\triangle}\ \diamondsuit\heartsuit\triangle\diamondsuit\heartsuit\triangle\diamondsuit\heartsuit\triangle}_{(\triangle\diamondsuit\heartsuit)^{10/3}}\triangle\diamondsuit\heartsuit\cdots\heartsuit\triangle\diamondsuit\ \underbrace{\overset{\overset{245}{\downarrow}}{\diamondsuit}\ \heartsuit\triangle\diamondsuit\heartsuit\triangle\diamondsuit\heartsuit\triangle\diamondsuit}_{(\diamondsuit\heartsuit\triangle)^{10/3}}\diamondsuit\heartsuit\triangle\cdots$$

From Lemma 5.6, we obtain the following easy fact generalizing Theorem 7(a) in [2], which states that $\mathbf{t}$ contains the square of a single letter if and only if $\gcd(b-1, m) = 1$.

Corollary 5.7 *In the square case, there exists a critical factor of* $\mathbf{t}$ *of length* ℓ *with* $b \nmid \ell$ *if and only if* $\gcd(b-1, m) \mid \ell$.

Proof: If $\ell < b$, then there exists such a critical factor if and only if $B_\ell \neq \emptyset$, that is, if and only if $S^\ell_{b-1,m} \neq \emptyset$, in which case $\gcd(b-1, m) \mid \ell$. Otherwise, if $\ell > b$, then we know that $b = m = 2$ and $\ell = 3$, so the result follows. □

Acknowledgements. The authors wish to thank the anonymous referees for carefully reading the proofs and providing helpful comments that improved the presentation.

References

[1] J.-P. Allouche and J. Shallit. The ubiquitous Prouhet-Thue-Morse sequence. In *Sequences and their applications (Singapore,* 1998*)*, Discrete Math. Theor. Comput. Sci., pages 1–16. Springer, London, 1999.

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

[3] J. Berstel. Axel Thue's work on repetitions in words. In *Proceedings of the* 4*th Conference on Formal Power Series and Algebraic Combinatorics, Montréal*, volume 11, pages 65–80. Publications du LaCIM, Université du Québec à Montréal, 1992.

[4] J. Berstel. On the index of Sturmian words. In *Jewels Are Forever*, pages 287–294. Springer-Verlag, Berlin, 1999.

[5] V. Berthe, C. Holton, and L. Q. Zamboni. Initial powers of Sturmian sequences. *Acta Arith.*, 122(4):315–347, 2006.

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

[7] D. Damanik and D. Lenz. The index of Sturmian sequences. *European J. Combin.*, 23(1):23–29, 2002.

[8] A. E. Frid. Overlap-free symmetric D0L words. *Discrete Math. Theor. Comput. Sci.*, 4(2):357–362, 2001.

[9] A. E. Frid. Arithmetical complexity of symmetric D0L words. *Theoret. Comput. Sci.*, 306:535–542, 2003.

[10] A. Glen. Powers in a class of $\mathcal{A}$-strict standard episturmian words. In *Words* 2005*, 5th International Conference on Words*, volume 36, pages 249–263. Publications du LaCIM, Université du Québec à Montréal, 2005. *Theoret. Comput. Sci.* (in press).

[11] J. Justin and G. Pirillo. Fractional powers in Sturmian words. *Theoret. Comput. Sci.*, 255(1–2):363–376, 2001.

[12] J. Justin and G. Pirillo. Episturmian words and episturmian morphisms. *Theoret. Comput. Sci.*, 276(1–2):281–313, 2002.

[13] D. Krieger. On critical exponents in fixed points of non-erasing morphisms. In *Developments in Language Theory* 2006, volume 4036 of *Lecture Notes in Computer Science*, pages 280–291, Springer, Berlin, 2006.

[14] M. Lothaire. *Combinatorics On Words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, Reading, Massachusetts, 1983.

[15] F. Mignosi and G. Pirillo. Repetitions in the Fibonacci infinite word. *Theor. Inform. Appl.*, 26(3):199–204, 1992.

[16] P. Morton and W. J. Mourant. Digit patterns and transcendental numbers. *J. Austral. Math. Soc. Ser. A*, 51:216–236, 1991.

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

[18] D. Vandeth. Sturmian words and words with a critical exponent. *Theoret. Comput. Sci.*, 242:283–300, 2000.