A note on the critical exponent of generalized Thue-Morse words

Alexandre Blondin-Massé*, Sébastien Labbé*

LaCIM, Université du Québec à Montréal,

C.P. 8888, succ. Centre-ville, Montréal (QC), CANADA, H3C 3P8

blondin_masse.alexandre@uqam.ca, slabqc@gmail.com

July 5, 2007

1 Introduction

It is well-known 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 [2]. Rediscovered by M. Morse in 1921 in the study of symbolic dynamics, this overlap-free word is now called the *Thue-Morse word*.

Here, we study a family $\mathcal{T}$ of generalized Thue-Morse words for which we compute the *critical exponent* (i.e., the largest fractional power of a factor) and we determine the occurrences of all factors realizing it.

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 [3, 4, 6, 9, 10, 11, 12, 14, 16].

Our main results are as follows. Let $\mathbf{t} \in \mathcal{T}$ be a generalized Thue-Morse word.

Theorem 1. *The critical exponent of* $\mathbf{t}$ *is given by*

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

*Supported by a NSERC (Canada) scholarship.

Theorem 2. *Suppose* $\mathbf{t}$ *is aperiodic. If* w *is a critical factor of* $\mathbf{t}$ *of length* $\ell = Nb^i$ *such that* $b \nmid N$*, then the set of positions at which* w *occurs in* $\mathbf{t}$ *is*

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

where A*,* B_N *and* C *are defined in Section 4.*

2 Preliminaries

A word u is a *factor* of w if $u = w_i \cdots w_j$ for some i, j with $i \leq j$, in which case we say that (u, i, j) is an *occurrence of* u *in* w. In other words, i is the beginning index and j is the ending index where u appears as a factor of w. If u is a factor of w such that $u \neq w$ it is called *proper*. Furthermore, u is called a *prefix* of w if $i = 0$, and a *suffix* of w if $j = |w| - 1$. The set of all factors of a word w is denoted by $F(w)$.

A word of the form $w = (uv)^n u$ is written as $w = z^r$, where $z = uv$ and $r := n + |u|/|z|$. The rational number r is called the *exponent* of z, and w is said to be a *rational power*. For example, the word

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

contains a factor which is a $\frac{5}{2}$-power. Note that a rational power z^r is well-defined only if $r \cdot |z|$ is a non-negative integer.

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

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

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

$$E(\mathbf{x}) = \sup_{w \in F(\mathbf{x})} \textsc{Index}(w).$$

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

A *morphism* is a function $\varphi : \Sigma^* \to \Sigma^*$ such that $\varphi(uv) = \varphi(u)\varphi(v)$, and 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.

Definition 3. *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* $\bar{\alpha} \in \Sigma$*, is the infinite word given by* $\mathbf{t} = \mu^\omega(\bar{\alpha}) = \lim_{n\to\infty}\mu^n(\bar{\alpha})$*.*

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

Remark 4. The n-th letter of $\mathbf{t}$ is $\mathbf{t}[n] := \sigma^{s_b(n)}(\bar{\alpha})$ 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 $\bar{\alpha} = \diamondsuit$, we obtain the following 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 $\bar{\alpha} = 0$, Remark 4 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 [1].

Note that words in $\mathcal{T}$, up to a relabeling of the letters, are 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 $w = w_0w_1\cdots w_{\ell-1} \in F(\mathbf{t})$ is called *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.

The following technical lemmas are useful.

Lemma 5. [8, 15] *The word* $\mathbf{t}$ *is periodic if and only if* $m \mid (b-1)$*. More precisely, if* $\mathbf{t}$ *is periodic then*

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

□

Lemma 6. *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.*

Lemma 7. *Let* $w = w_0w_1\cdots w_{\ell-1}$ *be a factor of* $\mathbf{t}$ *such that* $b \nmid \ell$*.*

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.*

Lemma 8. *Suppose* $\mathbf{t}$ *is aperiodic. If* w *is a factor of* $\mathbf{t}$ *of length* $\ell \geq b$ *with* $b \nmid \ell$*, then* $\textsc{Index}(w) \leq 2$*.*

3 Critical exponent

The next three lemmas allow us to consider particular occurrences of factors of $\mathbf{t}$. Let (w, i, j) be an occurrence of w in $\mathbf{t}$. We say that (w, i, j) is *left-synchronized* if $b \mid i$. Similarly, (w, i, j) is called *right-synchronized* if $b \mid (j + 1)$. Finally, we say that (w, i, j) is *synchronized* if it is both left-synchronized and right-synchronized.

Lemma 9. *Let $w = w_0 w_1 \cdots w_{\ell-1}$ be a factor of $\mathbf{t}$ such that $b \mid \ell$. Moreover, let $(ww_0, bq + r, j)$ be an occurrence of ww_0 in $\mathbf{t}$, where $q, r \in \mathbb{N}$, $0 \leq r < b$. Then there exists a suffix s of w such that (sww_0, bq, j) is left-synchronized.*

Lemma 10. *Let $w = w_0 w_1 \cdots w_{\ell-1}$ be a factor of $\mathbf{t}$ such that $b \mid \ell$ and let $(w_{\ell-1}w, i, j)$ be an occurrence of $w_{\ell-1}w$ in $\mathbf{t}$. Then there exists a prefix p of w such that $(w_{\ell-1}wp, i, j + |p|)$ is right-synchronized.*

Lemma 11. *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 word x of length ℓ and a rational $f \geq e$ such that both x and x^f have a synchronized occurrence in $\mathbf{t}$.*

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

Lemma 12. *Suppose $\mathbf{t}$ is aperiodic and let w be a factor of $\mathbf{t}$ of length ℓ such that $b \nmid \ell$. Then*

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

Note that since $\mathbf{t}$ is a fixpoint of a *uniform non-erasing* morphism, it follows immediately from Krieger's results in [12] that the critical exponent of $\mathbf{t}$ is either infinite (if $\mathbf{t}$ is periodic) or rational. The proof of Theorem 1 is then obtained by induction and uses the previous lemmas.

4 Occurrences of critical factors

Here, we assume that $\mathbf{t}$ is *aperiodic*, i.e., $m \nmid (b-1)$, and denote by e the critical exponent of $\mathbf{t}$. We say that $b > m$ is the *overlap case* and that $b \leq m$ is the *square case* . We describe now the *positions* and the *lengths* of all critical factors of $\mathbf{t}$.

Lemma 13. *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 $\mathbf{t}$.*

In view of Lemma 13, it is enough to consider only the case $b \nmid \ell$ when describing the positions of critical factors in $\mathbf{t}$.

Lemma 14. *Let w be a critical factor of $\mathbf{t}$ of length ℓ such that $b \nmid \ell$.*

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.*

It has already been noticed that squares of certain factors of length 3 appear in the Thue-Morse word. See [5] for example. The following lemma proves the uniqueness of this fact.

Lemma 15. *The Thue-Morse word $\mathbf{m}$ is the unique word in $\mathcal{T}$ containing a critical factor w of length $\ell > b$ such that $b \nmid \ell$. Moreover, $\ell = 3$.*

We now state the following lemma which is a more general result than Lemma 5 in [1].

Lemma 16. *Let $k, N \in \mathbb{N}^+$ be such that $b \nmid k$ and $1 \leq N < b$. Then,*

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

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. We suppose $\gcd(x, y) \mid p$ and we define

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

Moreover, let us define the following three sets :

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

Lemma 17. *Let w be a critical factor of $\mathbf{t}$ of length ℓ such that $b \nmid \ell$. Then, the set of positions at which w occurs in $\mathbf{t}$ is*

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

The proof of Theorem 2 follows easily from Lemmas 13 and 17. From Lemma 17, we obtain the following easy fact that generalizes Theorem 7(a) in [1], stating that $\mathbf{t}$ contains the square of a single letter if and only if $\gcd(b-1, m) = 1$.

Corollary 18. *In the square case, there exists a critical factor of $\mathbf{t}$ of length ℓ with $b \nmid \ell$ if and only if $\gcd(b-1, m) \mid \ell$.*

Proofs of all results are available upon request to the authors and will appear in an extended version. The authors wish to thank S. Brlek and A. Glen for the fruitful discussions that arose during the seminar course held in the Winter semester 2007, in which they enthusiastically motivated students to study combinatorics on words.

References

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

[2] J. Berstel, Axel Thue's work on repetitions in words, *Proceedings of the* 4*th Conference on Formal Power Series and Algebraic Combinatorics, Publications du LaCIM* **11** (1992), 65–80.

[3] J. Berstel, On the index of Sturmian words, in: *Jewels Are Forever*, Springer-Verlag, Berlin, 1999, pp. 287–294.

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

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

[6] D. Damanik, D. Lenz, Powers in Sturmian sequences, *European J. Combin.* **24** (2003), 377–390.

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

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

[9] A. Glen, Powers in a class of $\mathcal{A}$-strict standard episturmian words, *Theoret. Comput. Sci.* **380** (3) (2007), 330–354.

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

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

[12] D. Krieger, On critical exponents in fixed points of non-erasing morphisms, in: *Developments in Language Theory* 2006*, Lecture Notes in Computer Science*, Vol. 4036, *Springer-Verlag*, Berlin, 2006, pp. 280–291.

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

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

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

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