On the Density of Critical Factorizations

Tero Harju
Dirk Nowotka
Turku Centre for Computer Science, TUCS, and
Department of Mathematics, University of Turku

Turku Centre for Computer Science
TUCS Technical Report No 435
20th December 2001
ISBN 952-12-0937-2
ISSN 1239-1891

Abstract

We investigate the density of critical positions, that is, the ratio between the number of critical positions and the number of all positions of a word, in infinite sequences of words of index one, that is, the period of which is longer than half of the length of the word.

On one hand, we considered words with the lowest possible number of critical points, namely one, and show, as an example, that every Fibonacci word longer than five has exactly one critical factorization which provides a new way to prove two known facts about the periodicity of Fibonacci words.

On the other hand, sequences of words with a high density of critical points are considered. We show how to construct an infinite sequence of words in four letters where every point in every word is critical. We construct an infinite sequence of words in three letters with densities of critical points approaching one, using square-free words, and an infinite sequence of words in two letters with densities of critical points approaching two, using Thue–Morse words. It is shown that these bounds are optimal.

Furthermore, we give a short proof of the Critical Factorization Theorem and a theorem about the maximal distance between two critical points in a word. We state that only words in a binary alphabet can have just one critical factorization.

Keywords: combinatorics on words, repetitions, critical factorization theorem, density of critical factorizations, Fibonacci words, Thue–Morse words

TUCS Research Group
Automata Theory and Combinatorics on Words

1 Introduction

The Critical Factorization Theorem [3, 7] relates local periods with the global period of finite words. It states that in every word w there is a position p where the shortest repetition word z, i.e., $w = uv$ with $|u| = p$ and z is a suffix of w_1u and a prefix of vw_2 for suitable w_1 and w_2, is as long as the global period d of w, moreover, $p < d$. The position p is called critical. Actually, we have at least one critical position in every $d - 1$ consecutive positions in w. Consider the following example:

$$w = ab.aa.b$$

that has two critical positions 2 and 4 which are marked by dots. The period d of w equals 3 and w is of index 1, since $2d > |w|$. The shortest repetition word in both critical positions is aab. Note, that the shortest repetition words in the positions 1 and 3 are ba and a, respectively. The ratio of the number of critical positions and the number of all positions is called the density of critical positions. The density of w in our example is 2.

We are concerned with the density of critical positions in this paper. That is, we investigate words with exactly one and as many as possible critical positions.

After we have fixed the basic notations in Section 2, we give a technically improved version of a proof [5] of the Critical Factorization Theorem [3, 7] and state the maximal distance between two critical points in a word. In Section 3 we show that Fibonacci words, which can be defined by palindromes [6], of length greater than five have exactly one critical position in contrast to the fact that palindromes have at least two critical positions. This result also implies immediately the two well-known facts that the period of a Fibonacci word is a Fibonacci number and that the Fibonacci word is not ultimately periodic, both proven differently in the literature. Section 4 contains the constructions of infinite sequences of words in four letters with density one for every word, infinite sequences of ternary words which has a limit of their densities at one, using square-free words [12, 1, 2], and infinite sequences of binary words which has a limit of their densities at two, using Thue–Morse words [11, 10, 1, 2]. We also show that these limits are optimal.

2 Preliminaries

In this section we fix the notations for this paper. We refer to [8, 4] for more basic and general definitions.

Let A be a finite nonempty alphabet and A^* be the monoid of all finite words in A; the empty word is denoted by ε. Let A^ω denote the set of all infinite words in A. An infinite word $w \in A^\omega$ is called *ultimately periodic* if there exist two words $u, v \in A^*$ such that $w = uv^\omega$. Let $w \in A^*$ in the following. The length of w is denoted by $|w|$ and its ith letter by $w_{(i)}$. By definition $|\varepsilon| = 0$. If $w = w_1 u w_2$ then u is called a *factor* of w. If $w = uv$ then u and v are called *prefix* and *suffix*, respectively, and let $u = wv^{-1}$ and $v = u^{-1}w$. Note, that ε and w are both prefixes and suffixes of w. A word w is called *bordered* if there exists a word $v \neq \varepsilon$ such that $w = vuv$, and in this case, v is called a *border* of w. A word w is called *primitive* if $w = v^k$ implies that $k \leq 1$.

An integer d, with $1 \leq d \leq |w|$, is called *a period* of w if $w_{(i)} = w_{(i+d)}$, for all $1 \leq i \leq |w| - d$. The smallest period of w is denoted by $\partial(w)$ and it is called the *minimal period* or the *global period* of w. The *index* $\operatorname{ind}(w)$ of a word w is defined by

$$\operatorname{ind}(w) = \left\lfloor \frac{|w|}{\partial(w)} \right\rfloor .$$

Let an integer p with $1 \leq p < |w|$ be called *position* or *point* in w. Intuitively, a position p denotes the place between $w_{(p)}$ and $w_{(p+1)}$ in w. A word $u \neq \varepsilon$ is called a *repetition word* at position p if $w = xy$ with $|x| = p$ and there exist x' and y' such that $u \preccurlyeq x'x$ and $u \leq yy'$. For a point p in w, let

$$\partial(w, p) = \min\bigl\{|u| \bigm| u \text{ is a repetition word at } p\bigr\}$$

denote *the local period* at point p in w. Note, the repetition word of length $\partial(w, p)$ at point p is unbordered and $\partial(w, p) \leq \partial(w)$. A factorization $w = uv$, with $u, v \neq \varepsilon$ and $|u| = p$, is called *critical* if $\partial(w, p) = \partial(w)$, and, if this holds, then p is called *critical point*, otherwise it is called *noncritical point*. Let $\eta(w)$ denote the number of critical points in a word w. We shall represent critical points of words by dots. For instance, the critical points of $w = abaaba$ are 2 and 4, and we show this by writing $w = ab.aa.ba$. In this example, $\partial(w) = 3$.

Let $\tilde{w} = w_{(n)} \cdots w_{(2)} w_{(1)}$ denote the *reverse* of $w = w_{(1)} w_{(2)} \cdots w_{(n)}$. We call a word w a *palindrome* if $w = \tilde{w}$.

Let $\lhd$ be an ordering of $A = \{a_1, a_2, \ldots, a_n\}$, say $a_1 \lhd a_2 \lhd \cdots \lhd a_n$. Then $\lhd$ induces a *lexicographic order* on A^* such that

$$u \lhd v \iff u \leq v \quad \text{or} \quad u = xau' \text{ and } v = xbu' \text{ with } a \lhd b$$

where $a, b \in A$. A suffix v (prefix u) of w is called *maximal* w.r.t. $\lhd$ if $v' \lhd v$ (and $\tilde{u}' \lhd \tilde{u}$) for any suffix v' (prefix u') of w. Let $\lhd^{-1}$ denote the *inverse* *order*, say $a_n \lhd^{-1} \cdots \lhd^{-1} a_2 \lhd^{-1} a_1$, of $\lhd$. Let $\mu_{\lhd}(w)$ and $\mu_{\rhd}(w)$ denote the maximal suffixes of w w.r.t. $\lhd$ and $\lhd^{-1}$, respectively, and let $\nu_{\lhd}(w)$ and $\nu_{\rhd}(w)$ denote the maximum prefixes of w w.r.t. $\lhd$ and $\lhd^{-1}$, respectively. If the context is clear, we may write $\mu_{\lhd}$, $\mu_{\rhd}$, $\nu_{\lhd}$, and $\nu_{\rhd}$ for $\mu_{\lhd}(w)$, $\mu_{\rhd}(w)$, $\nu_{\lhd}(w)$, and $\nu_{\rhd}(w)$, respectively. We only consider alphabets of size larger than one in the following.

The critical factorization theorem (CFT) was discovered by Césari and Vincent [3] and developed into its current form by Duval [7].

Theorem 1 (Critical Factorization Theorem). *Every word w, with $|w| \geq 2$, has at least one critical factorization $w = uv$, with $u, v \neq \varepsilon$ and $|u| < \partial(w)$, i.e., $\partial(w, |u|) = \partial(w)$.*

The following proof of the CFT is a technically improved version of the proof by Crochemore and Perrin in [5].

Proof. Let $\alpha = \mu_{\lhd}(w)$ and $\beta = \mu_{\rhd}(w)$. Suppose $|\beta| < |\alpha|$, say $\alpha = u'\beta$. Certainly $\alpha \neq \beta$, since they start with a different letter, and note, that $|w\beta^{-1}| < \partial(w)$. Let z be an unbordered repetition word at $|w\beta^{-1}|$. We show that $|z|$ is a period of w, which will prove the claim.

If w is a factor of z^2, then obviously $|z|$ is a period of w. If $w = w_1\beta w_2$ for some $w_2 \neq \varepsilon$, then $\beta \lhd^{-1} \beta w_2$ contradicts the choice of β. If $w\beta^{-1} = yz$, then, by the above, $z \leq \beta$, say $\beta = z\beta'$; but then $z^2\beta' = z\beta \lhd^{-1} \beta = z\beta'$ implies that $\beta = z\beta' \lhd^{-1} \beta'$; a contradiction. Consequently, $\beta = zw'$ and $w = z_1 z w'$ for a suffix z_1 of the unbordered word z. Therefore u' is a suffix of z, and hence, $u'w'$ is a suffix of α. Consequently, $u'w' \lhd \alpha = u'\beta$, and so $w' \lhd \beta$, which together with $w' \lhd^{-1} \beta$ implies that $w' \leq \beta$. Therefore $\beta = zw' = w'z'$, and thus $\beta = z^k z_2$ for some $z_2 \leq z$, which shows that $|z|$ is a period of w. $\square$

For a different proof of the CFT by Duval, Mignosi, and Restivo, see Chapter 8 in [9]. The next theorem justifies why we are only interested in words of index one in our investigation of the density of critical points.

Theorem 2. *Each set of $\partial(w) - 1$ consecutive points in w, where $|w| \geq 2$, has a critical point.*

Proof. If $w = u^i u_1$, where $u_1 \leq u$ and $\partial(w) = |u|$, then the maximal suffixes w.r.t. any orders of A are longer than $|u^{i-1}u_1|$. Hence w has a critical point at point p, where $p < \partial(w)$.

Let p be any critical point of $w = uv$, where $|u| = p$, and let z be the smallest repetition word at position p. So, $|z| = \partial(w)$.

We need to show that if $|v| \geq \partial(w)$, then there is critical point at $p+k$ for $1 \leq k < \partial(w)$. We have $z \leq v$ and $\partial(v) = \partial(w)$. For, if $\partial(v) < \partial(w)$, then z is bordered; a contradiction. Now, v has a critical point k such that we have $k < \partial(v) = \partial(w)$. Clearly, this point $p+k$ is critical also for w. Now, $(p+k) - p = k < \partial(w)$. $\square$

Maybe an even stronger motivation for considering only words of index one, is that in w^k, with $k \geq 3$, the critical points of the first factor w are inherited by the next $k-2$ factors w. That is, if $w^k = w_1.w_2w^{k-1}$, where $|w_1|$ is a critical point, then also $|ww_1|$ is a critical point of w^k.

3 Words with Exactly One Critical Factorization

Every word longer than one letter has at least one critical factorization. We investigate words with only one critical factorization in this section. Trivially, words of length two have no more than one critical point. We do not consider such cases but arbitrary long words. However, the following lemma limits our investigation to words in two letters.

Lemma 3. *A word w with only one critical factorization is binary, that is, it is over a two-letter alphabet.*

Proof. Assume a word w contains the letters a, b, and c and has exactly one critical factorization. Let $a \lhd b \lhd c$. By symmetry, we can assume that $|\mu_{\rhd}| < |\mu_{\lhd}|$. Then $p = |w\mu_{\rhd}^{-1}|$ is a critical point of w by the proof of Theorem 1. Let $a \blacktriangleleft c \blacktriangleleft b$. Now, either $|w\mu_{\blacktriangleleft}^{-1}|$ or $|w\mu_{\blacktriangleright}^{-1}|$ is a critical point p' of w, again by the proof of Theorem 1. But, $p \neq p'$ since $\mu_{\rhd}$ begins with c and $\mu_{\blacktriangleleft}$ and $\mu_{\blacktriangleright}$ begin with a and b, respectively. So, w has at least two critical points; a contradiction. $\square$

By Lemma 3, we will only consider words in a and b in the rest of this section. Let $a \lhd b$. Theorem 1 straightforwardly leads to the following two facts.

Lemma 4. *If a word w has exactly one critical point, then either*

$$w = \nu_{\lhd}\mu_{\rhd} \quad \textit{and} \quad \nu_{\lhd} \leq \nu_{\rhd} \quad \textit{and} \quad \mu_{\rhd} \preccurlyeq \mu_{\lhd}$$

or

$$w = \nu_{\rhd}\mu_{\lhd} \quad \textit{and} \quad \nu_{\rhd} \leq \nu_{\lhd} \quad \textit{and} \quad \mu_{\lhd} \preccurlyeq \mu_{\rhd} \ .$$

The inverse of Lemma 4 does not hold in general. Consider $w = aa.bb.abab$ which has two critical points, but we do have

$$\nu_\triangleleft = aa \leq aabb = \nu_\triangleright \qquad \text{and} \qquad \mu_\triangleright = bbabab \preccurlyeq aabbabab = \mu_\triangleleft$$

and $w = \nu_\triangleleft \mu_\triangleright$.

Lemma 5. *Every palindrome w has at least two critical factorizations.*

Proof. Assume w has exactly one critical point. By symmetry, we can assume that $\mu_\triangleright \preccurlyeq \mu_\triangleleft$. By the definition of maximal prefix and suffix and since w is a palindrome

$$\mu_\triangleleft(w) = \tilde{\nu}_\triangleleft(\tilde{w}) = \tilde{\nu}_\triangleleft(w) \qquad \text{and} \qquad \mu_\triangleright(w) = \tilde{\nu}_\triangleright(\tilde{w}) = \tilde{\nu}_\triangleright(w)$$

where $\tilde{\nu}_\triangleright(w)$ and $\tilde{\nu}_\triangleleft(w)$ denotes the inverse of $\nu_\triangleright(w)$ and $\nu_\triangleleft(w)$, respectively. Now, $\tilde{\nu}_\triangleright \preccurlyeq \tilde{\nu}_\triangleleft$, and hence, $\nu_\triangleright \leq \nu_\triangleleft$, which contradicts Lemma 4. □

Let us now consider the critical points of Fibonacci words. Fibonacci numbers are defined by

$$f_0 = 1 \ , \qquad f_1 = 1 \ , \qquad f_{k+2} = f_{k+1} + f_k \ .$$

Fibonacci words are defined by

$$F_1 = a \ , \qquad F_2 = ab \ , \qquad F_{k+2} = F_{k+1} F_k \ .$$

Obviously, $|F_i| = f_i$. Let $F = \lim_{n\to\infty} F_n$ be *the* Fibonacci word. Observe that $F_i \leq F_n$, with $1 \leq i \leq n$. It is also clear that all Fibonacci words are primitive. The following lemma will be used to estimate the number of critical points in Fibonacci words.

Lemma 6. *We have that $f_{n-2} < \partial(F_n) \leq f_{n-1}$ for all $n > 2$.*

Proof. Indeed, $\partial(F_n) \leq f_{n-1}$, and if $\partial(F_n) < f_{n-2}$, then F_{n-2} is not primitive since

$$F_{n-2}F_{n-2} < F_{n-2}F_{n-3}F_{n-2} = F_n \ ,$$

a contradiction, whereas if $\partial(F_n) = f_{n-2}$, then $F_n \leq F_{n-2}^{+}$, and $F_{n-2} \preccurlyeq F_n$ implies that F_{n-1} or F_{n-2} is not primitive; a contradiction. □

Remark 7. Fibonacci words have a close connection to palindromes as the following properties show. Firstly, $F_n = \alpha_n d_n$ where $n \geq 3$ and α_n is a palindrome and $d_n = ab$ if n is even and $d_n = ba$ if n is odd. This result is credited to Berstel in [6]. Secondly, $F_n = \beta_n \gamma_n$, where $n \geq 5$ and β_n and γ_n are palindromes of length $f_{n-1} - 2$ and $f_{n-2} + 2$, respectively, by de Luca [6]. Moreover, de Luca shows that these two properties define the set of Fibonacci words.

Given Remark 7 and Lemma 5, every palindrome has at least two critical factorizations, the following Theorem 9 is rather surprising.

Example 8 (Fibonacci words). *We have*

$$F_2 = a.b \;, \qquad\qquad F_3 = a.b.a \;, \qquad\qquad F_4 = ab.aa.b \;.$$

By the following Theorem, however, all Fibonacci words F_n, with $n > 4$ has exactly one critical point, and that critical point is at position $f_{n-1} - 1$.

Theorem 9. *A Fibonacci word F_n, with $n > 4$, has exactly one critical point p. Moreover, p is at position $f_{n-1} - 1$.*

Proof. Let $n \geq 7$, and let p be a critical point of F_n. Then $p > f_{n-2}$, because

$$F_{n-2}F_{n-2} < F_{n-2}F_{n-3}F_{n-2} = F_n$$

and by Lemma 6. Consider the factorization

$$F_n = F_{n-2}F_{n-3}F_{n-2} = F_{n-2}F_{n-4}F_{n-5}F_{n-2} \;.$$

Then $p > f_{n-2} + f_{n-4}$, since positions i, where $f_{n-2} \leq i \leq f_{n-2} + f_{n-4}$, are not critical

$$F_{n-3}F_{n-4}F_{n-4}F_{n-4} = F_{n-2}F_{n-4}F_{n-4} < F_{n-2}F_{n-4}F_{n-5}F_{n-2} = F_n \;,$$

since these positions occur inside the second factor F_{n-4} in the power F_{n-4}^3. By induction we obtain

$$\begin{aligned} & F_{n-2}\ F_{n-4} \cdots\ F_{n-2i+1}\ F_{n-2i}\ F_{n-2i}\ F_{n-2i} \\ = {} & F_{n-2}\ F_{n-4} \cdots \quad F_{n-2i+2} \quad F_{n-2i}\ F_{n-2i} \\ < {} & F_{n-2}\ F_{n-4} \cdots \quad F_{n-2i+2} \quad F_{n-2i}\ F_{n-2i-1}F_{n-2} \\ = {} & F_n \end{aligned}$$

where $1 \leq i \leq \lceil \frac{n}{2} \rceil - 2$, and

$$p > \sum_{i=1}^{\lceil \frac{n}{2} \rceil - 2} f_{n-2i} \;.$$

So, we have

$$F_n = F_{n-2}F_{n-4}\cdots F_3F_2F_{n-2} \qquad \text{or} \qquad F_n = F_{n-2}F_{n-4}\cdots F_4F_3F_{n-2}$$

where $p > f_{n-1} - 2$ and $p > f_{n-1} - 3$, respectively, and $f_{n-1} > p$ by Lemma 6 and Theorem 2. So, $p = f_{n-1} - 1$ or $p = f_{n-1} - 2$, that is, a critical point has to exist in the suffix

$$F_2F_{n-2} \preccurlyeq F_n \qquad \text{or} \qquad F_3F_{n-2} \preccurlyeq F_n$$

where the former case gives the result. The latter case leaves the possibilities $a.b.aF_{n-2} \preccurlyeq F_n$. But since $b \preccurlyeq F_4$, we have $bab.aF_{n-2} \preccurlyeq F_n$ and only the marked position is critical which proves the claim. □

The following well known facts follow immediately from Theorem 9.

Corollary 10. *A Fibonacci word F_n has the period f_{n-1}, and the Fibonacci word F is not ultimately periodic.*

The Fibonacci words are certainly not the only words with exactly one critical factorization.

Example 11. *Let $w = a^iba^j$, with $i \neq j$ and $i + j > 0$, then $\eta(w) = 1$. If $i > j$, then $\partial(w) = |a^ib|$ and i is the only critical point of w. Similarly for $i < j$, where $i + 1$ is the only critical point of w. See Lemma 13 for the case when $i = j$.*

4 Words with a High Density of Critical Factorizations

We investigate the densities of the critical points in words. The *density* $\delta(w)$ of a word w is defined by

$$\delta(w) = \frac{\eta(w)}{|w| - 1} \ .$$

Notice that in the above $|w| - 1$ is the number of all positions in w. Recall that we require all words to be of index one, otherwise, for any given alphabet $\{a_1, a_2, \ldots, a_k\}$ and $n > 0$, we have

$$\delta\left((a_1a_2\cdots a_k)^n\right) = 1$$

that is, every position is critical. Moreover, there exists a sequence of words of index one in the alphabet $A = \{a, b, c\}$ such that the limit of their densities is one.

Example 12 (Square-free words). *Let us consider the endomorphism* $\vartheta\colon A^* \to A^*$ *with*

$$a \mapsto abc \qquad\qquad b \mapsto ac \qquad\qquad c \mapsto b$$

by Thue [12], cf. [1, 2], and let

$$T_{2k+1} = a\ \vartheta^{2k+1}(a)\ c \qquad \text{and} \qquad T_{2k} = a\ \vartheta^{2k}(a)\ b,$$

for all $k > 0$*, then*

$$\lim_{n\to\infty} \delta(T_n) = 1$$

because every word T_n *has a square prefix and suffix and* $\vartheta^n(a)$ *is square-free, so,* $\eta(T_n) = |T_n| - 3$ *and* $\delta(T_n) = 1 - 2/(|T_n| - 1)$*.*

Of course, any square-free word with suitable borders can be used in Example 12. It is also clear that with an alphabet with at least four letters, say a, b, c, and d, the sequence $\{T'_n\}$, with $n \geq 1$ and $T'_n = d\ \vartheta^n(a)\ d$, consists of words with density one only. Words in two letters, however, cannot be square-free, if they are longer than three. So, the question arises: What is the highest density for words in $A = \{a, b\}$? The following lemma implies that ab, ba, aba, and bab are the only words in A which have density one.

Lemma 13. *If* w *has two consecutive critical points, then either* $w = a^iba^i$ *or* $w = b^iab^i$*, with* $i > 0$*.*

Proof. Assume, two consecutive critical points in w that are around b. Then clearly $w = w_1a^i.b.a^jw_2$, for some $i, j > 0$. If $w_1 = \varepsilon = w_2$, then necessarily $i = j$, see Example 11. Assume, $i \neq j$. By symmetry, we can assume that $w_1 \neq \varepsilon$, that is, $w = vba^iba^jw_2$. If $j \geq i$, then w has the repetition ba^i at the first critical point: $w = v[ba^i][ba^i]a^{j-i}w_2$, where $\partial(w) = |ba^i|$. But then w has index greater than one; a contradiction. Therefore, $j < i$, and in this case, $w_2 = \varepsilon$ in order to avoid repetition inside w at the second critical point. Also, $\partial(w) = |a^jb|$, which implies that $i = j$; again a contradiction. $\square$

By Lemma 13 we have $\limsup_{n\to\infty} \delta(w_n) \geq 1/2$, for any infinite sequence w_1, w_2, w_3, ... of words in A, and this bound is tight by the following example.

Example 14 (Thue–Morse words). *We consider the Thue–Morse endomorphism* $\varphi\colon A^* \to A^*$ *with*

$$a \mapsto ab \qquad\qquad \text{and} \qquad\qquad b \mapsto ba$$

see [11, 10, 1, 2]. For the sake of brevity, let φ_n denote $\varphi^n(a)$ and $\bar{\varphi}_n$ denote $\varphi^n(b)$. Let

$$M_{2k+1} = a^2\varphi_{2k+1}b^2 \qquad \textit{and} \qquad M_{2k} = a^2\varphi_{2k}a^2,$$

for all $k \geq 0$, then we show in Theorem 15 that $\eta(M_n) = 2^{n-1} + 1$ and

$$\delta(M_n) = \frac{1}{2} - \frac{1}{2^{n+1} + 6}$$

and hence,

$$\lim_{n\to\infty} \delta(M_n) = \frac{1}{2} .$$

Note, that φ_n equals $\bar{\varphi}_n$ up to renaming of a and b. Moreover, φ_n does not contain overlapping factors, that is, factors of the form $cucuc$ where $c \in A$. Note also that $|\varphi_n| = 2^{n-1}$.

Theorem 15. *Every odd position in M_n, with $n \geq 1$, except position 1 and $2^n + 3$, is critical.*

We consider the following lemma before proving Theorem 15.

Lemma 16. *The repetition words at every noncritical position in M_n, for all $n \geq 1$, are of length one or two.*

Proof. In M_n, for any $n > 2$, the positions 1, 2, $2^n + 2$, and $2^n + 3$ are noncritical, with repetition words of length one, and the positions 3 and 2^n+1 are critical.

Clearly, the repetition word at every noncritical position in $M_1 = aaabbb$ and $M_2 = aaabbaaa$ is of length one.

Assume, the repetition word at every noncritical position in M_k, with $k > 2$, is of length one or two. By induction, the repetition word at every noncritical position in M_{k+1} is at most of length two because $M_{k+1} = a^2\varphi_k\bar{\varphi}_k a^2$ and $M_{k+1} = a^2\varphi_k\bar{\varphi}_k b^2$ for odd end even k, respectively. Note, that, by induction hypothesis, the repetition word at every noncritical position in M_k is at most of length two. Clearly, the repetition words of length less or equal than two at positions 2 to $2^k - 2$ in φ_k are not changed by preceding and succeeding words (a^2 or b^2). The repetition word at position $|M_{k+1}|/2 = 2^k + 2$ in M_{k+1} is either b or ba since $ab \preccurlyeq \varphi_k$ or $ba \preccurlyeq \varphi_k$ and $ba \leq \bar{\varphi}_k$.

It remains to show that the positions 2^k+1 and 2^k+3 are critical in M_{k+1}. Assume position $2^k + 1$ or $2^k + 3$ is not critical.

If k is even, then the repetition word u at position $2^k + 1$ is of the form $abavb$ and $|u| < 2^k + 2$, otherwise position $2^k + 1$ is critical. The factor uu is followed by b in M_{k+1}, otherwise $abavbabavba$ is a factor of φ_{k+1}; a contradiction. But, now we have $a \preccurlyeq v$, otherwise b^3 is a factor of φ_{k+1}, a contradiction, and $ababa$ is a factor of φ_{k+1}; again a contradiction.

If k is odd, then the repetition word u' at position $2^k + 1$ is of the form $bbava$ and $|u| < 2^k - 1$, otherwise position $2^k + 1$ is critical. Certainly, the factor uu must be preceded by a in M_{k+1}, otherwise b^3 is a factor of φ_{k+1}; a contradiction. But, now $abbavabbava$ is a factor of φ_{k+1}; again a contradiction.

Position $2^k + 3$ is shown to be critical by similar arguments. □

Proof of Theorem 15. We show that there are no two consecutive noncritical positions in M_n except 1 and 2, and $2^n + 2$ and $2^n + 3$.

By Lemma 16, the words a, b, ab, and ba are the only repetition words at noncritical positions in M_n. We need to consider only positions from 4 to 2^n, since positions 3 and $2^n + 1$ are certainly critical.

Assume a is the repetition word at some position p and position $p - 1$ is noncritical. Now, a must be the repetition word at position $p - 1$ and a^3 is a factor of φ_n; a contradiction. The same argument holds if b is the repetition word at position p.

Assume ab is the repetition word at some position p and position $p - 1$ is noncritical. Now, ba must be the repetition word at position $p - 1$ and $babab$ is a factor of φ_n; a contradiction. The same argument holds if ba is the repetition word at position p.

The claim follows now from Lemma 13. □

Remark 17. Is there a sequence with a higher density of critical points than $\{M_n\}$, with $n > 0$? Certainly, there is no sequence with a limit larger than 1/2 by Lemma 13. Actually, there is no binary word larger than 5 with a density equal to 1/2 by the following Lemma 18. A word M_n is basically an overlap-free word with cubic prefix and suffix. In any case, Lemma 20 will show that any infinite sequence with a limit 1/2 of densities must include infinitely many words where the first and the last two positions are noncritical. However, could we use other words than Thue–Morse words to construct $\{M_n\}$, with $n > 0$? If we choose a word with an overlapping factor, say $w = w_1 a u_\circ a_\circ u a w_2$, then w has two consecutive positions, marked by $\circ$, that are not critical. Lemma 13 implies that w would not be a good choice. So, what about other overlap-free words? Any infinite set of finite overlap-free binary words would certainly do for $\{M_n\}$, with $n > 0$. However, φ is the smallest morphism that takes an overlap-free word to a longer overlap-free word. So, $\{M_n\}$, with $n > 0$, is optimal from that point of view.

Lemma 18. *Every binary word w, with* $\operatorname{ind}(w) = 1$ *and* $|w| > 5$*, implies* $\delta(w) < \frac{1}{2}$.

Proof. Assume $|w| > 5$ and $\delta(w) = 1/2$. Then there are no two consecutive critical points in w by Lemma 13. Certainly $\partial(w) > 3$ since $\operatorname{ind}(w) = 1$.

The first and the last position of w is not critical. Otherwise, let $a.b \leq w$ and point 1 is critical, then aba and $abba$ are not prefixes of w since the repetition word in position 1 are then ba and bba, respectively; contradicting $\partial(w) > 3$. Also, $abbbb$ is not a prefix of w since then $\delta(w) < 1/2$ by Lemma 13. Hence, $abbba \leq w$ and $\partial(w) = 4$ since the smallest repetition word in position 1 is now $bbba$. So, w equals $a.bbb.aa$ or $a.bbb.aab$ and has just two critical points; a contradiction. The last position is a symetric case.

The claim follows now from Lemma 13. □

Remark 19. The largest binary words of index one and density two are given by Lemma 13:

$$aa.b.aa\ , \qquad\qquad bb.a.bb\ ,$$

and by the Fibonacci word F_4 and its reverse $\tilde{F}_4$:

$$ab.aa.b\ , \qquad ba.bb.a\ , \qquad b.aa.ba\ , \qquad a.bb.ab\ .$$

Lemma 20. *Let* $\{w_n\}$*, with* $|w_n| > 5$ *for all* $n > 0$*, be an infinite sequence of binary words such that*

$$\limsup_{n\to\infty} \delta(w_n) = \frac{1}{2}\ .$$

Then there is an infinite set I of natural numbers such that the first and the last two positions of w_i, for all $i \in I$, are noncritical.

Proof. Let w_k be such that $\delta(w_k) > 1/2 - \epsilon$ for some positive real number $\epsilon < 1/4$. The first and the last position of w_k are noncritical by the proof of Lemma 18. We have $|w_k| > 1/(2\epsilon)$ since

$$\frac{\eta(w_k)}{2\eta(w_k)+1} \geq \delta(w_k) > \frac{1}{2} - \epsilon \qquad \text{and} \qquad \eta(w_k) > \frac{1}{4\epsilon} - \frac{1}{2}$$

by the proof of Lemma 18, and

$$|w_k| \geq 2(\eta(w_k)+1) > \frac{1}{2\epsilon}$$

again by the proof of Lemma 18. Assume the second position of w_k is critical.

Let $aa.b \leq au \leq w_k$ where $|u| = \partial(w_k) - 1$. The factor aa does not appear in u. Actually, $u = ab^{k_1}ab^{k_2} \cdots ab^{k_t}$, where $k_j \geq 1$ for all $1 \leq j \leq t$, and since $|w_k| > 1/\epsilon$ and $\mathrm{ind}(w) = 1$, we have $|u| > 1/(4\epsilon)$ and $t > 1/(8\epsilon)$. Since $C = \{ab^i \mid 1 \leq i \leq t\}$ is a code, we can consider u to be encoded in an alphabet X of size $|C|$, let u' be the encoded u. Certainly, u' is square-free. However, by the assumption that $\delta(w_k) > 1/2 - \epsilon$, we must have a factor v in u, with $|v| > 1/(4\epsilon)$, where critical and noncritical positions alternate, so, bbb is not a factor of v. Let v' be the encoding in X of the smallest factor that contains v. Now, v' must be a square-free word in at most two letters, namely the once that encode ab and abb. But the longest square-free word in two letters is xyx, with $x, y \in X$, and hence, $|v'| < 9$; a contradiction.

Let $ab.a \leq w = uv$ where $|u| = \partial(w)$. Then $u = aba^{\partial(w)-2}$; a contradiction.

Similar arguments hold for the last but one position when u ends in aa or ba. □

References

[1] Jean Berstel. Axel Thue's work on repetitions in words. In Pierre Leroux and Christophe Reutenauer, editors, *Séries formelles et combinatoire algébrique*, volume 11 of *Publications du LaCIM*, pages 65–80. Université du Québec à Montréal, June 1992.

[2] Jean Berstel. *Axel Thue's papers on repetitions in words: a translation*, volume 20 of *Publications du LaCIM*. Université du Québec à Montréal, 1995.

[3] Yves Césari and Max Vincent. Une caractérisation des mots périodiques. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 286(A):1175–1177, 1978.

[4] Christian Choffrut and Juhani Karhumäki. Combinatorics of words. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 329–438. Springer-Verlag, Berlin, 1997.

[5] Maxime Crochemore and Dominique Perrin. Two-way string-matching. *Journal of the Association for Computing Machinery*, 38(3):651–675, 1991.

[6] Aldo de Luca. A combinatorial property of the fibonacci words. *Information Processing Letters*, 12(4):193–195, August 1981.

[7] Jean-Pierre Duval. Périodes et répétitions des mots de monoïde libre. *Theoretical Computer Science*, 9(1):17–26, 1979.

[8] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics*. Addison-Wesley, Reading, Massachusetts, 1983.

[9] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, Cambridge, United Kingdom, 2002.

[10] Harold Marston Morse. Recurrent geodesics on a surface of negative curvature. *Transactions of the American Mathematical Society*, 22(1):84–100, 1921.

[11] Axel Thue. Über unendliche Zeichenreihen. *Det Kongelige Norske Videnskabersselskabs Skrifter, I Mat.-nat. Kl. Christiania*, 7:1–22, 1906.

[12] Axel Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Det Kongelige Norske Videnskabersselskabs Skrifter, I Mat.-nat. Kl. Christiania*, 1:1–67, 1912.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.abo.fi

University of Turku
- **Department of Mathematical Sciences**

Åbo Akademi University
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

Turku School of Economics and Business Administration
- **Institute of Information Systems Science**