# Periodicity and Unbordered Words 

Tero Harju<br>Dirk Nowotka

Turku Centre for Computer Science, TUCS, Department of Mathematics, University of Turku


Turku Centre for Computer Science
TUCS Technical Report No 523
April 2003
ISBN 952-12-1154-7
ISSN 1239-1891


#### Abstract

The relationship between the length of a word and the maximum length of its unbordered factors is investigated in this paper.

Consider a finite word $w$ of length $n$. Let $\mu(w)$ denote the maximum length of its unbordered factors, and let $\partial(w)$ denote the period of $w$. Clearly, $\mu(w) \leq \partial(w)$.

We establish that $\mu(w)=\partial(w)$, if $w$ has an unbordered prefix of length $\mu(w)$ and $n \geq 2 \mu(w)-1$. This bound is tight and solves a 21 year old conjecture by Duval. It follows from this result that, in general, $n \geq 3 \mu(w)-2$ implies $\mu(w)=\partial(w)$ which gives an improved bound for the question asked by Ehrenfeucht and Silberger in 1979.


Keywords: combinatorics on words, periodicity, unbordered factors, Duval's conjecture

## 1 Introduction

Periodicity and borderedness are two properties of words - the most basic data structure - which are investigated in this paper. These concepts are so foundational that they play a rôle (explicitely or implicitely) in virtually every area of computer science. Just a few of those areas are string searching algorithms $[15,3,8]$, data compression [23, 7], and codes [2], which are classical examples, but also computational biology, e.g., sequence assembly [19] or superstrings [4], and serial data communications systems [5] are areas among others where periodicity and borderedness of words (sequences) are important concepts. It is well known that these two word properties do not exist independently from each other. However, it is somewhat surprising that no clear relation has been established so far, despite the fact that this basic question has been around for more than 20 years.

Let us consider a finite word (a sequence of letters) $w$. We denote the length of $w$ by $|w|$ and call a subsequence of consecutive letters of a word factor. The period of $w$, denoted by $\partial(w)$, is the smallest positive integer $p$ such that the $i$-th letter equals the $(i+p)$-th letter for all $1 \leq i \leq|w|-p$. Let $\mu(w)$ denote the length of the longest unbordered factor of $w$. A word is bordered, if it has a proper prefix that is also a suffix, where we call a prefix proper, if it is neither empty nor contains the entire word. For the investigation of the relationship between $|w|$ and the maximality of $\mu(w)$, that is, $\mu(w)=\partial(w)$, we consider the special case where the longest unbordered prefix of a word is of the maximum length, that is, no unbordered factor is longer than that prefix. Let $w$ be an unbordered word. Then a word $w u$ is a Duval extension (of $w$ ), if every unbordered factor of $w u$ has at most length $|w|$, that is, $\mu(w u)=|w|$. We call $w u$ trivial Duval extension, if $\partial(w u)=|w|$. For example, let $w=a b a a b b$ and $u=a a b a$. Then $w u=a b a a b b a a b a$ is a nontrivial Duval extension of $w$ since $(i) w$ is unbordered, (ii) all factors of $w u$ longer than $w$ are bordered, that is, $|w|=\mu(w u)=6$, and (iii) the period of $w u$ is 7 , and hence, $\partial(w u)>|w|$. Note, that this example satisfies $|u|=|w|-2$.

In 1979 Ehrenfeucht and Silberger initiated a line of research [11, 1, 10] exploring the relationship between the length of a word $w$ and $\mu(w)$. In 1982 these efforts culminated in Duval's result: If $|w| \geq 4 \mu(w)-6$ then $\partial(w)=\mu(w)$. However, it was conjectured in [1] that $|w| \geq 3 \mu(w)$ implies $\partial(w)=\mu(w)$ which follows if Duval's conjecture [10] holds true.

Conjecture 1. Let wu be a nontrivial Duval extension of $w$. Then $|u|<|w|$.
After that, no progress was recorded, to the best of our knowledge, for 20 years. However, the topic remained popular, see for example Chapter 8
in [17]. The most recent results are by Mignosi and Zamboni [20] and the authors of this article [13]. However, not Duval's conjecture but rather its opposite is investigated in those papers, that is: Which words admit only trivial Duval extensions? It is shown in [20] that unbordered, finite factors of Sturmian words allow only trivial Duval extensions, with other words, if an unbordered, finite factor of a Sturmian word of length $\mu(w)$ is a prefix of $w$, then $\partial(w)=\mu(w)$. Sturmian words are binary infinite words of minimal complexity; see [21] and Chapter 2 in [17]. That result was improved in [13] by showing that Lyndon words [18] allow only trivial Duval extensions and the fact that every unbordered, finite factor of a Sturmian word is a Lyndon word.

The main result in this paper is an improved version of Conjecture 1.
Theorem 2. Let wu be a Duval nontrivial extension of $w$. Then $|u|<|w|-1$.
The example mentioned above shows that this bound on the length of a nontrivial Duval extension is tight. Theorem 2 implies the truth of Duval's conjecture, as well as, the following corollary (for any word $w$ ).

Corollary 3. If $|w| \geq 3 \mu(w)-2$, then $\partial(w)=\mu(w)$.
This corollary confirms the conjecture by Assous and Pouzet in [1] about a question asked by Ehrenfeucht and Silberger in [11].

Our main result, Theorem 2, is presented in Section 4, which uses the notations introduced in Section 2 and preliminary results from Section 3. We conclude with Section 5.

## 2 Notations

In this section we introduce the notations of this paper. We refer to [16, 17] for more basic and general definitions.

We consider a finite alphabet $A$ of letters. Let $A^{*}$ denote the monoid of all finite words over $A$ including the empty word, denoted by $\varepsilon$. Let $w=w_{(1)} w_{(2)} \cdots w_{(n)}$ where $w_{(i)}$ is a letter, for every $1 \leq i \leq n$. We denote the length $n$ of $w$ by $|w|$. An integer $1 \leq p \leq n$ is a period of $w$, if $w_{(i)}=w_{(i+p)}$ for all $1 \leq i \leq n-p$. The smallest period of $w$ is called the minimum period (or simply, the period) of $w$, denoted by $\partial(w)$. A nonempty word $u$ is called a border of a word $w$, if $w=u v=v^{\prime} u$ for some suitable words $v$ and $v^{\prime}$. We call $w$ bordered, if it has a border that is shorter than $w$, otherwise $w$ is called unbordered. Note, that every bordered word $w$ has a minimum border $u$ such that $w=u v u$, where $u$ is unbordered. Let $\mu(w)$ denote the maximum length of unbordered factors of $w$. Suppose $w=u v$, then $u$ is called a prefix of $w$,
denoted by $u \leq w$, and $v$ is called a suffix of $w$, denoted by $v \preccurlyeq w$. Let $u, v \neq \varepsilon$. Then we say that $u$ overlaps $v$ from the left or from the right, if there is a word $w$ such that $|w|<|u|+|v|$, and $u \leq w$ and $v \preccurlyeq w$, or $v \leq w$ and $u \preccurlyeq w$, respectively. We say that $u$ overlaps (intersects) with $v$, if either $v$ is a factor of $u$ or $u$ is a factor of $v$ or $u$ overlaps $v$ from the left or right.

Let us consider the following examples. Let $A=\{a, b\}$ and $u, v, w \in A^{*}$ such that $u=a b a a$ and $v=b a a b a$ and $w=a b a a b a$. Then $|w|=6$, and 3,5 , and 6 are periods of $w$, and $\partial(w)=3$. We have that $a$ is the shortest border of $u$ and $w$, whereas $b a$ is the shortest border of $v$. We have $\mu(w)=3$. We also have that $u$ and $v$ overlap since $u \leq w$ and $v \preccurlyeq w$ and $|w|<|u|+|v|$.

We continue with some more notations. Let $w$ and $u$ be nonempty words where $w$ is also unbordered. We call $w u$ a Duval extension of $w$, if every factor of $w u$ longer than $|w|$ is bordered, that is, $\mu(w u)=|w|$. A Duval extension $w u$ of $w$ is called trivial, if $\partial(w u)=\mu(w u)=|w|$. A nontrivial Duval extension $w u$ of $w$ is called minimal, if $u$ is of minimal length, that is, $u=u^{\prime} a$ and $w=u^{\prime} b w^{\prime}$ where $a, b \in A$ and $a \neq b$.

Example 4. Let $w=a b a a b b a b a a b a b b$ and $u=a a b a$. Then

$$
w \cdot u=a b a a b b a b a a b a b b \cdot a a b a
$$

(for the sake of readability, we use a dot to mark where $w$ ends) is a nontrivial Duval extension of $w$ of length $|w u|=18$, where $\mu(w u)=|w|=14$ and $\partial(w u)=15$. However, wu is not a minimal Duval extension, whereas

$$
w \cdot u^{\prime}=a b a a b b a b a a b a b b \cdot a a
$$

is minimal, with $u^{\prime}=a a \leq u$. Note, that $w u$ is not the longest nontrivial Duval extension of $w$ since

$$
w \cdot v=a b a a b b a b a a b a b b \cdot a b a a b a
$$

is longer, with $v=a b a a b a$ and $|w v|=20$ and $\partial(w v)=17$. One can check that $w v$ is a nontrivial Duval extension of $w$ of maximum length, and at the same time wv is also a minimal Duval extension of $w$.

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

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

denote the local period at point $p$ in $w$. Note, that the repetition word of length $\partial(w, p)$ at point $p$ is necessarily unbordered and $\partial(w, p) \leq \partial(w)$. A factorization $w=u v$, 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.

Example 5. The word

$$
w=a b . a a . b
$$

has the period $\partial(w)=3$ and two critical points, 2 and 4, marked by dots. The shortest repetition words at the critical points are aab and baa, respectively. Note, that the shortest repetition words at the remaining points 1 and 3 are ba and a, respectively.

## 3 Preliminary Results

We state some auxiliary and well-known results about repetitions and borders in this section which will be used to prove Theorem 2, in Section 4. The proofs of these auxiliary results are straightforward and not given in this extended abstract. Results taken from the literature are referenced to.

Lemma 6. Let $z f=g z h$ where $f, g \neq \varepsilon$. Let $a z^{\prime}$ be the maximum unbordered prefix of az. If az does not occur in $z f$, then agz' is unbordered.

Proof. Assume $a g z^{\prime}$ is bordered, and let $y$ be its shortest border. In particular, $y$ is unbordered. If $\left|z^{\prime}\right| \geq|y|$ then $y$ is a border of $a z^{\prime}$ which is a contradiction. If $\left|a z^{\prime}\right|=|y|$ or $|a z|<|y|$ then $a z$ occurs in $z f$ which is again a contradiction. If $\left|a z^{\prime}\right|<|y| \leq|a z|$ then $a z^{\prime}$ is not maximum since $y$ is unbordered; a contradiction.

The proof of the following lemma is easy.
Lemma 7. Let $w$ be an unbordered word and $u \leq w$ and $v \preccurlyeq w$. Then $u w$ and $w v$ are unbordered.

The critical factorization theorem is one of the main results about periodicity of words. A weak version of it was first conjectured by Schützenberger [22] and proved by Césari and Vincent [6]. It was developed into its current form by Duval [9]. We refer to [12] for a short proof of the CFT.

Theorem 8 (CFT). Every word $w$, with $|w| \geq 2$, has at least one critical factorization $w=u v$, with $u, v \neq \varepsilon$ and $|u|<\partial(w)$, i.e., $\partial(w,|u|)=\partial(w)$.

We have the following two lemmas about properties of critical factorizations.

Lemma 9. Let $w=u v$ be unbordered and $|u|$ be a critical point of $w$. Then $u$ and $v$ do not overlap.

Proof. Note, that $\partial(w,|u|)=\partial(w)=|w|$ since $w$ is unbordered. Let $|u| \leq|v|$ without restriction of generality. Assume that $u$ and $v$ overlap. If $u=u^{\prime} s$ and $v=s v^{\prime}$, then $\partial(w,|u|) \leq|s|<|w|$. On the other hand, if $u=s u^{\prime}$ and $v=v^{\prime} s$, then $w$ is bordered with $s$. Finally, if $v=$ sut then $\partial(w,|u|) \leq|s u|<|w|$.

The next result follows directly from Lemma 9.
Lemma 10. Let $u_{0} u_{1}$ be unbordered and $\left|u_{0}\right|$ be a critical point of $u_{0} u_{1}$. Then for any word $x$, we have $u_{i} x u_{i+1}$, where the indices are modulo 2 , is either unbordered or has a minimum border $g$ such that $|g| \geq\left|u_{0}\right|+\left|u_{1}\right|$.

The next theorem states a basic fact about minimal Duval extensions. See [14] for a proof of it.

Theorem 11. Let wu be a minimal Duval extension of $w$. Then $u$ occurs in $w$.

The following Lemmas 12, 13 and 14 and Corollary 3 are given in [10]. Let $a_{0}, a_{1} \in A$, with $a_{0} \neq a_{1}$, and $t_{0} \in A^{*}$. Let the sequences $\left(a_{i}\right),\left(s_{i}\right),\left(s_{i}^{\prime}\right)$, $\left(s_{i}^{\prime \prime}\right)$, and $\left(t_{i}\right)$, for $i \geq 1$, be defined by

- $a_{i}=a_{i(\bmod 2)}$, that is, $a_{i}=a_{0}$ or $a_{i}=a_{1}$, if $i$ is even or odd, respectively,
- $s_{i}$ such that $a_{i} s_{i}$ is the shortest border of $a_{i} t_{i-1}$,
- $s_{i}^{\prime}$ such that $a_{i+1} s_{i}^{\prime}$ is the longest unbordered prefix of $a_{i+1} s_{i}$,
- $s_{i}^{\prime \prime}$ such that $s_{i}^{\prime} s_{i}^{\prime \prime}=s_{i}$,
- $t_{i}$ such that $t_{i} s_{i}^{\prime \prime}=t_{i-1}$.

For any parameters of the above definition, the following holds.
Lemma 12. For any $a_{0}, a_{1}$, and $t_{0}$ there exists an $m \geq 1$ such that

$$
\left|s_{1}\right|<\cdots<\left|s_{m}\right|=\left|t_{m-1}\right| \leq \cdots \leq\left|t_{0}\right|
$$

and $s_{m}=t_{m-1}$ and $\left|t_{0}\right| \leq\left|s_{m}\right|+\left|s_{m-1}\right|$.
Lemma 13. Let $z \leq t_{0}$ such that $a_{0} z$ and $a_{1} z$ do not occur in $t_{0}$. Let $a_{0} z_{0}$ and $a_{1} z_{1}$ be the longest unbordered prefixes of $a_{0} z$ and $a_{1} z$, respectively. Then

1. if $m=1$ then $a_{0} t_{0}$ is unbordered,
2. if $m>1$ is odd, then $a_{1} s_{m}$ is unbordered and $\left|t_{0}\right| \leq\left|s_{m}\right|+\left|z_{0}\right|$,
3. if $m>1$ is even, then $a_{0} s_{m}$ is unbordered and $\left|t_{0}\right| \leq\left|s_{m}\right|+\left|z_{1}\right|$.

Lemma 14. Let $v$ be an unbordered factor of $w$ of length $\mu(w)$. If $v$ occurs twice in $w$, then $\mu(w)=\partial(w)$.

Corollary 15. Let wu be a Duval extension of $w$. If $w$ occurs twice in wu, then wu is a trivial Duval extension.

## 4 Main Result

The next theorem proves Duval's conjecture.
Theorem 2. Let wu be a nontrivial Duval extension of $w$. Then $|u|<|w|-1$.
Proof. Recall that every factor of $w u$ which is longer than $|w|$ is bordered since $w u$ is a Duval extension of $w$. Let $z$ be the longest suffix of $w$ that occurs twice in $z u$.

If $z=\varepsilon$ then $a \preccurlyeq w$ and $u=b^{j}$, where $a, b \in A$ and $a \neq b$ and $j \geq 1$, but now $|u|<|w|$ since $a b^{j}$ is unbordered. Moreover, $w=b^{k} a w^{\prime} a$ with $k<j$, otherwise $w u$ is a trivial Duval extension, and either $a w^{\prime} a b^{j}$ is bordered, in this case it follows $j \leq\left|w^{\prime}\right|$, or $a w^{\prime} a b^{j}$ is unbordered. In both cases it follows $|u|<|w|-1$.

So, assume $z \neq \varepsilon$. We have $z \neq w$ since $w u$ is otherwise trivial by Corollary 3. Let $a, b \in A$ be such that

$$
w=w^{\prime} a z \quad \text { and } \quad u=u^{\prime} b z r
$$

and $z$ occurs in $z r$ only once, that is, $b z$ matches the rightmost occurrence of $z$ in $u$. Note, that $b z$ does not overlap $a z$ from the right, by Lemma 7 , and therefore $u^{\prime}$ exists, although it might be empty. Naturally, $a \neq b$ by the maximality of $z$, and $w^{\prime} \neq \varepsilon$, otherwise $a z u^{\prime} b z \leq w u$ has either no border or $w$ is bordered (if $a z u^{\prime} b z$ has a border not longer than $z$ ) or $a z$ occurs in $z u$ (if $a z u^{\prime} b z$ has a border longer than $z$ ); a contradiction in any case.

Let $a z_{0}$ and $b z_{1}$ denote the longest unbordered prefix of $a z$ and $b z$, respectively. Let $a_{0}=a$ and $a_{1}=b$ and $t_{0}=z r$ and the integer $m$ be defined as in Lemma 13. We have then a word $s_{m}$, with its properties defined by Lemma 13 , such that

$$
t_{0}=s_{m} t^{\prime}
$$

Consider $a z u^{\prime} b z_{0}$. We have that $a z$ and $a z u^{\prime} b z_{0}$ are both prefixes of $a_{0} z u$, and $b z_{0}$ is a suffix of $a z u^{\prime} b z_{0}$ and $a z$ does not occur in $z u^{\prime} b z_{0}$. It follows from Lemma 6 that $a z u^{\prime} b z_{0}$ is unbordered, and hence,

$$
\begin{equation*}
\left|a z u^{\prime} b z_{0}\right| \leq|w| \tag{1}
\end{equation*}
$$

| $w$ | $u$ |  |  |
| :---: | :---: | :---: | :---: |
|  | $u^{\prime}$ | $b, z, r$ |  |
| $z_{0}$ |  | $z_{0}$ <br> $s_{m}, t^{\prime}$ |  |

Case: Suppose that $m$ is even. Then we have $2 \leq m$ and $a s_{m}\left(=a_{m} s_{m}\right)$ is unbordered and $\left|t_{0}\right| \leq\left|s_{m}\right|+\left|z_{1}\right|$ by Lemma 13.

Suppose $\left|t_{0}\right|=\left|s_{m}\right|+\left|z_{1}\right|$ and $z_{1}=z$. Then $\left|s_{m-1}\right|=|z|$ by Lemma 12. Note, that $s_{i} \leq t_{i-1} \leq t_{0}$ for all $1 \leq i \leq m$, and hence, it follows that $s_{i} \leq z$ for all $1 \leq i<m$. In particular, $s_{m-1}=z$. We have that $b z\left(=a_{1} s_{m-1}\right)$ is a border of $b t_{m-2}\left(=a_{1} t_{m-2}\right)$. But now, $b z$ occurs in $t_{0}$, and hence, in $u$, since $t_{i} \leq t_{0}$, for all $0 \leq i<m$, which is a contradiction.

So, assume that $\left|t_{0}\right|<\left|s_{m}\right|+\left|z_{1}\right|$ or $\left|z_{1}\right|<|z|$. Suppose $\left|s_{m}\right| \leq\left|z_{0}\right|$. Then $\left|a z u^{\prime} b z_{0}\right| \leq|w|$ and

$$
\begin{aligned}
|u| & =|a z u|-|z|-1 \\
& =\left|a z u^{\prime} b z_{0}\right|-\left|z_{0}\right|+\left|t_{0}\right|-|z|-1 \\
& <\left|a z u^{\prime} b z_{0}\right|-\left|z_{0}\right|+\left|s_{m}\right|+\left|z_{1}\right|-|z|-1 \\
& \leq|w|+\left|z_{1}\right|-|z|-1 \\
& \leq|w|-1
\end{aligned}
$$

if $\left|t_{0}\right|<\left|s_{m}\right|+\left|z_{1}\right|$, or

$$
\begin{aligned}
|u| & =|a z u|-|z|-1 \\
& =\left|a z u^{\prime} b z_{0}\right|-\left|z_{0}\right|+\left|t_{0}\right|-|z|-1 \\
& \leq\left|a z u^{\prime} b z_{0}\right|-\left|z_{0}\right|+\left|s_{m}\right|+\left|z_{1}\right|-|z|-1 \\
& \leq|w|+\left|z_{1}\right|-|z|-1 \\
& <|w|-1
\end{aligned}
$$

if $\left|z_{1}\right|<|z|$. We have $|u|<|w|-1$ in both cases.
Let then $\left|s_{m}\right|>\left|z_{0}\right|$. We have that $a s_{m}$ is unbordered, and since $a z_{0}$ is the longest unbordered prefix of $a z$, we have $a z \leq a s_{m}$, and hence, $|z| \leq\left|s_{m}\right|$. Now, $a z u^{\prime} b s_{m}$ is unbordered otherwise its shortest border is longer than $a z$, since no prefix of $a z$ is a suffix of $a s_{m}$, and $a z$ occurs in $u$; a contradiction. So, $\left|a z u^{\prime} b s_{m}\right| \leq|w|$ and $|u|<|w|-1$, since either $\left|z_{1}\right| \leq|z|$ or $\left|t_{0}\right|<\left|s_{m}\right|+\left|z_{1}\right|$.

Case: Suppose that $m$ is odd. Then $b s_{m}\left(=a_{m} s_{m}\right)$ is unbordered and $\left|t_{0}\right| \leq\left|s_{m}\right|+\left|z_{0}\right| ;$ see Lemma 13. Surely $s_{m} \neq \varepsilon$.

If $\left|s_{m}\right|<|z|$, then $|u|<|w|-1$ since

$$
|u|=\left|a z u^{\prime} b z_{0}\right|-\left|b z_{0}\right|+\left|b t_{0}\right|-|a z|
$$

and $\left|a z u^{\prime} b z_{0}\right| \leq|w|$, by (1), and $\left|t_{0}\right| \leq\left|s_{m}\right|+\left|z_{0}\right|$.

Assume thus that $\left|s_{m}\right| \geq|z|$, and hence, also $z \leq s_{m}$. Since $s_{m} \neq \varepsilon$, we have $\left|b s_{m}\right| \geq 2$, and therefore, by the critical factorization theorem, there exists a critical point $p$ in $b s_{m}$ such that $b s_{m}=v_{0} v_{1}$, where $\left|v_{0}\right|=p$.


In particular,

$$
\begin{equation*}
b z \leq v_{0} v_{1} \tag{2}
\end{equation*}
$$

Note, that if $s_{m}=z$ then $\left|z_{0}\right|<|z|$ since $b \preccurlyeq z_{0}$ and $b s_{m}$ does not end with $b$ because it is unbordered. We have therefore in all cases

$$
\begin{equation*}
\left|z_{0}\right|<\left|v_{0} v_{1}\right|-1 \tag{3}
\end{equation*}
$$

Let

$$
u=u_{0}^{\prime} v_{0} v_{1} u_{1}
$$

be such that $v_{0} v_{1}$ does not occur in $u_{0}^{\prime}$. Note, that $v_{0} v_{1}$ does not overlap with itself since it is unbordered, and $v_{0}$ and $v_{1}$ do not overlap by Lemma 9 . Consider the prefix $w u_{0}^{\prime} b z$ of $w u$ which is bordered and has a shortest border $g$ longer than $z$, and hence, $b z \preccurlyeq g$, otherwise $w$ is bordered since $z \preccurlyeq w$. Moreover, $g \leq w$, for otherwise $a z$ would occur in $u$, and hence, $b z$ occurs in $w$. Let

$$
w=w_{0} b z w_{1}
$$

such that $b z$ occurs in $w_{0} b z$ only once, that is, we consider the leftmost occurrence of $b z$ in $w$. Note, that

$$
\begin{equation*}
\left|w_{0} b z\right| \leq|g| \leq\left|u_{0}^{\prime} b z\right| \tag{4}
\end{equation*}
$$

where the first inequality comes from the definition of $w_{0}$ above and the second inequality from the fact that $\left|u_{0}^{\prime} b z\right|<|g|$ implies that $w$ is bordered. Let

$$
f=b z w_{1} u_{0}^{\prime} v_{0} v_{1}
$$

If $f$ is unbordered, then $|f| \leq|w|$, and hence, $\left|u_{0}^{\prime} v_{0} v_{1}\right| \leq\left|w_{0}\right|$. Now, we have $\left|u_{0}^{\prime}\right|<\left|w_{0}\right|$ which contradicts (4).

Therefore, $f$ is bordered. Let $h$ be its shortest border.


Surely, $|b z|<|h|$ otherwise $v_{0} v_{1}$ is bordered by (2). So, $b z \leq h$. Moreover, $\left|v_{0} v_{1}\right| \leq|h|$ otherwise $b z$ occurs in $s_{m}$ contradicting our assumption that $b z r$ marks the rightmost occurrence of $b z$ in $u$. So, $v_{0} v_{1} \preccurlyeq h$, and $v_{0} v_{1}$ occurs in $w$ since $w_{0} h \leq w$ by (4). Let

$$
w_{0} b z v^{\prime}=w_{0} h=w_{0}^{\prime} v_{0} v_{1} .
$$

Note, that $v_{0} v_{1}$ does not occur in $w_{0}^{\prime}$ otherwise it occurs in $u_{0}^{\prime}$ contradicting our assumption on $u_{0}^{\prime}$. Moreover, we have $h=b z v^{\prime} \preccurlyeq u_{0}^{\prime} v_{0} v_{1}$. Let $u_{0}^{\prime} v_{0} v_{1}=u_{0} h$. Consider

$$
f_{0}=w u_{0} b z
$$

which has a shortest border $h_{0}$.


Surely, $b z \preccurlyeq h_{0}$ otherwise $w$ is bordered with a suffix of $z$. Moreover, $\left|w_{0} b z\right| \leq\left|h_{0}\right| \leq\left|u_{0} b z\right|$ since $b z$ does not occur in $w_{0}$ and $w$ is unbordered. From that and $w_{0} h=w_{0}^{\prime} v_{0} v_{1}$ and $u_{0} h=u_{0}^{\prime} v_{0} v_{1}$ follows now $\left|w_{0}^{\prime}\right| \leq\left|u_{0}^{\prime}\right|$ and

$$
\begin{equation*}
u_{0}^{\prime} v_{0} v_{1}=u_{0} b z v^{\prime} \text { and } w_{0} \text { occurs in } u_{0} . \tag{5}
\end{equation*}
$$

Let now

$$
w=w_{0}^{\prime} v_{0} v_{1} w_{i}^{\prime} \cdots v_{0} v_{1} w_{2}^{\prime} v_{0} v_{1} w_{1}^{\prime} v_{0} v_{1} w_{2}
$$

for some word $w_{2}$ that does not contain $v_{0} v_{1}$, and

$$
u=u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{2}^{\prime} v_{0} v_{1} u_{1}^{\prime} v_{0} v_{1} t^{\prime}
$$

such that $v_{0} v_{1}$ does not occur in $w_{k}^{\prime}$, for all $0 \leq k \leq i$, or $v_{\ell}^{\prime}$, for all $0 \leq \ell \leq j$. Note, that these factorizations of $w$ and $u$ are unique, and, moreover, $w_{2} \neq \varepsilon$. (Indeed, if $w_{2}=\varepsilon$ then $v_{0} v_{1} \preccurlyeq w$ and $a z \preccurlyeq v_{0} v_{1}$, and $a z$ would occur in $u$; a contradiction.)

We claim that either $i=j$ and $w_{k}^{\prime}=u_{k}^{\prime}$, for all $1 \leq k \leq i$ or $|u|<|w|-1$. Assume $k=1$. We show that $w_{1}^{\prime}=u_{1}^{\prime}$. Consider

$$
f_{1}=v_{1} w_{1}^{\prime} v_{0} v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{1}^{\prime} v_{0}
$$

If $f_{1}$ is unbordered, then $|u|<|w|-1$ since $\left|f_{1}\right| \leq|w|$ and

$$
|u|=\left|f_{1}\right|-\left|v_{1} w_{1}^{\prime} v_{0} v_{1} w_{2}\right|+\left|v_{1} t^{\prime}\right|
$$

and $\left|t^{\prime}\right| \leq\left|z_{0}\right| \leq|z|<|b z| \leq\left|v_{0} v_{1}\right|$ and $w_{2} \neq \varepsilon$. Assume then that $f_{1}$ is bordered, and let $h_{1}$ be its shortest border. Clearly, $h_{1}=v_{1} g_{1} v_{0}$ for some $g_{1}$ (possibly $g_{1}=\varepsilon$ ) since $v_{0}$ and $v_{1}$ do not overlap. We show that $h_{1} \leq v_{1} w_{1}^{\prime} v_{0}$. Indeed, otherwise either

1. $a z$ occurs in $u$, in case $v_{1} w_{1}^{\prime} v_{0} v_{1} w_{2} \leq h_{1}$, a contradiction to our assumption on $a z$, or
2. $v_{0}$ and $v_{1}$ overlap, in case $\left|v_{0}\right| \leq|z|$ and

$$
\left|v_{1} w_{1}^{\prime} v_{0} v_{1} w_{2}\right|-|a z|+\left|v_{0}\right|<\left|h_{1}\right|<\left|v_{1} w_{1}^{\prime} v_{0} v_{1} w_{2}\right|
$$

and then $v_{0}$ occurs in $z$, contradicting Lemma 9 , or
3. $|u|<|w|-1$, in case $v_{0} w_{3} \preccurlyeq w_{2}$ and $|a z| \leq\left|v_{0} w_{3}\right|$, then $v_{0} w_{3} u^{\prime} v_{0} v_{1}$ is unbordered and the result follows from $\left|t^{\prime}\right|<\left|v_{0} w_{3}\right|-1$, since $|a z| \neq\left|v_{0} w_{3}\right|$ for $v_{0}$ does not begin with $a$.

Moreover, $h_{1} \preccurlyeq v_{1} u_{1}^{\prime} v_{0}$ since $v_{0} v_{1}$ does not occur in $v_{1} w_{1}^{\prime} v_{0}$. So, let

$$
\begin{equation*}
w_{1}^{\prime} v_{0}=g_{1} v_{0} w_{1}^{\prime \prime} \quad \text { and } \quad v_{1} u_{1}^{\prime}=u_{1}^{\prime \prime} v_{1} g_{1} \tag{6}
\end{equation*}
$$



Consider,

$$
f_{2}=v_{0} w_{1}^{\prime \prime} v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{1}^{\prime} v_{0} v_{1}
$$

If $f_{2}$ is unbordered, then $|u|<|w|-1$ since $\left|f_{2}\right| \leq|w|$ and

$$
|u|=\left|f_{2}\right|-\left|v_{0} w_{1}^{\prime \prime} v_{1} w_{2}\right|+\left|t^{\prime}\right|
$$

and $\left|t^{\prime}\right| \leq\left|z_{0}\right| \leq|z|<|b z| \leq\left|v_{0} v_{1}\right|$ and $w_{2} \neq \varepsilon$. Assume then that $f_{2}$ is bordered, and let $h_{2}$ be its shortest border. Since $v_{0}$ and $v_{1}$ do not overlap,
$v_{0} v_{1} \preccurlyeq h_{2}$. Also $h_{2} \leq v_{0} w_{1}^{\prime \prime} v_{1}$ since $v_{0} v_{1}$ does not occur in $w_{2}$ (and $v_{0}$ and $v_{1}$ do not overlap) and $a z$ does not occur in $h_{2}$ (and so $h_{2}$ does not stretch beyond $w$ ). We have $v_{0} w_{1}^{\prime \prime} v_{1} \leq h_{2}$ since $v_{0} v_{1}$ does not occur in $v_{0} w_{1}^{\prime \prime} v_{1}$ unless $w_{1}^{\prime \prime}=\varepsilon$. Hence, we have $h_{2}=v_{0} w_{1}^{\prime \prime} v_{1}$ and


Consider,

$$
f_{3}=v_{0} v_{1} w_{1}^{\prime} v_{0} v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{2}^{\prime} v_{0} u_{1}^{\prime \prime} v_{1}
$$

If $f_{3}$ is unbordered, then $|u|<|w|-1$ since $\left|f_{3}\right| \leq|w|$ and

$$
|u|=\left|f_{3}\right|-\left|v_{0} v_{1} w_{1}^{\prime} v_{0} v_{1} w_{2}\right|+\left|g_{1} v_{0} v_{1} t^{\prime}\right|
$$

and $\left|t^{\prime}\right| \leq\left|z_{0}\right| \leq|z|<|b z| \leq\left|v_{0} v_{1}\right|$ and $\left|g_{1}\right| \leq\left|w_{1}^{\prime}\right|$ and $w_{2} \neq \varepsilon$. Assume, $f_{3}$ is bordered. Then $f_{3}$ has a shortest border $h_{3}$ such that $v_{0} v_{1} \leq h_{3}$. We have $h_{3}=v_{0} u_{1}^{\prime \prime} v_{1}$ by the arguments from the previous paragraph. Moreover,


Observe, that (7) and (8) imply that the number of occurrences of $v_{1}$ and $v_{0}$, respectively, is the same in $w_{1}^{\prime}$ and $u_{1}^{\prime}$ since $v_{0}$ and $v_{1}$ do not overlap. Now, let

$$
h_{1}=v_{1} g_{1} v_{0}=h_{1}^{\prime \prime} v_{1} h_{1}^{\prime} v_{0}=v_{1} h_{0}^{\prime} v_{0} h_{0}^{\prime \prime}
$$

where $v_{1}$ and $v_{0}$ occur only once in $v_{1} h_{1}^{\prime}$ and $h_{0}^{\prime} v_{0}$, respectively.


Now, let

$$
f_{2}^{\prime}=v_{0} h_{0}^{\prime \prime} w_{1}^{\prime \prime} v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{1}^{\prime} v_{0} v_{1}
$$

and

$$
f_{3}^{\prime}=v_{0} v_{1} w_{1}^{\prime} v_{0} v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{2}^{\prime} v_{0} u_{1}^{\prime \prime} h_{1}^{\prime \prime} v_{1}
$$

with respective shortest borders $h_{2}^{\prime}$ and $h_{3}^{\prime}$ (which are both not empty, if $|u| \geq|w|-1$; as in the case of $f_{2}$ and $f_{3}$ ) and $v_{0} v_{1} \preccurlyeq h_{2}^{\prime}$ and $v_{0} v_{1} \leq h_{3}^{\prime}$.

We have $h_{2}^{\prime} \leq v_{0} h_{0}^{\prime \prime} w_{1}^{\prime \prime} v_{1}$ since $v_{0} v_{1}$ does not occur in $w_{2}$ and $a z$ does not occur in $h_{2}^{\prime}$ (and so $h_{2}^{\prime}$ does not stretch beyond $w$ ). We have $v_{0} h_{0}^{\prime \prime} w_{1}^{\prime \prime} v_{1} \leq h_{2}^{\prime}$ since $v_{0} v_{1}$ does not occur in $w_{1}^{\prime}$. Hence, we have $h_{2}^{\prime}=v_{0} h_{0}^{\prime \prime} w_{1}^{\prime \prime} v_{1}$ and


We have $h_{3}^{\prime}=v_{0} u_{1}^{\prime \prime} h_{1}^{\prime \prime} v_{1}$ by the arguments from the previous paragraph. Moreover,


It is now straightforward to see that

$$
w_{1}^{\prime \prime}=u_{1}^{\prime \prime}=\varepsilon
$$

for otherwise $v_{1}$ and $v_{0}$ occur more than once in $v_{1} h_{1}^{\prime}$ and $h_{0}^{\prime} v_{0}$, respectively. From (6) follows now

$$
w_{1}^{\prime}=g_{1}=u_{1}^{\prime} .
$$

Assume $1<k \leq \min \{i, j\}$ and $w_{\ell}^{\prime}=u_{\ell}^{\prime}$, for all $1 \leq \ell<k$. Let us denote both $w_{\ell}^{\prime}$ and $u_{\ell}^{\prime}$ by $v_{\ell}^{\prime}$, for all $1 \leq \ell<k$.

We show that $w_{k}^{\prime}=u_{k}^{\prime}$. Consider

$$
f_{4}=v_{1} w_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{k}^{\prime} v_{0}
$$

If $f_{4}$ is unbordered, then $|u|<|w|-1$ since $\left|f_{4}\right| \leq|w|$ and

$$
|u|=\left|f_{4}\right|-\left|v_{1} w_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} w_{2}\right|+\left|v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} t^{\prime}\right|
$$

and $\left|t^{\prime}\right| \leq\left|z_{0}\right| \leq|z|<|b z| \leq\left|v_{0} v_{1}\right|$ and $w_{2} \neq \varepsilon$. Assume, $f_{4}$ is bordered. Then $f_{4}$ has a shortest border $h_{4}$ such that $\left|v_{0} v_{1}\right| \leq\left|h_{4}\right|$. Let $h_{4}=v_{1} g_{4} v_{0}$.

If $\left|v_{1} w_{k}^{\prime} v_{0}\right|<\left|h_{4}\right|$ then there exists an $\ell<k$ such that

$$
h_{4}=v_{1} w_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{\ell+1}^{\prime} v_{0} v_{1} v_{\ell}^{\prime \prime} v_{0}
$$

where $v_{\ell}^{\prime \prime} \leq v_{\ell}^{\prime}$. That implies

$$
u_{k}^{\prime}=v_{\ell}^{\prime \prime}
$$

since $v_{0} v_{1}$ does neither occur in $v_{\ell}^{\prime \prime}$ nor in $u_{k}^{\prime}$. Now, consider

$$
f_{5}=v_{1} w_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{\ell}^{\prime \prime} v_{0}
$$

If $f_{5}$ is unbordered, then $|u|<|w|-1$ since $\left|f_{4}\right|<\left|f_{5}\right|$, see above. Assume, $f_{5}$ is bordered. Then $f_{5}$ has a shortest border $h_{5}$ such that

$$
\left|h_{4}\right|<\left|h_{5}\right|
$$

for otherwise $h_{4}$ is not the shortest border of $f_{4}$, since either $h_{4} \leq h_{5}$ or $h_{5} \leq h_{4}$, and the latter implies that $h_{4}$ is bordered, and hence, not minimal. But now, we have a $\ell^{\prime}<\ell$ such that

$$
h_{5}=v_{1} w_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{\ell^{\prime}+1}^{\prime} v_{0} v_{1} v_{\ell^{\prime}}^{\prime \prime} v_{0}
$$

where $v_{\ell^{\prime}}^{\prime \prime} \leq v_{\ell^{\prime}}^{\prime}$. We have $\left|f_{4}\right|<\left|f_{5}\right|<\left|f_{6}\right|$ where

$$
f_{6}=v_{1} w_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{\ell^{\prime}}^{\prime \prime} v_{0}
$$

which is either unbordered and $|u|<|w|-1$ since $\left|f_{4}\right|<\left|f_{5}\right|$, or it is bordered with a shortest border $h_{6}$, and we have $\left|h_{4}\right|<\left|h_{5}\right|<\left|h_{6}\right|$ and a factor $f_{7}$, such that $\left|f_{4}\right|<\left|f_{5}\right|<\left|f_{6}\right|<\left|f_{7}\right|$, and so on, until eventually an unbordered factor is reached proving that $|u|<|w|-1$.

Assume then that $h_{4} \leq v_{1} w_{k}^{\prime} v_{0}$. We also have that $h_{4} \preccurlyeq v_{1} u_{k}^{\prime} v_{0}$ since $v_{0} v_{1}$ does not occur in $w_{k}^{\prime}$. So, let $w_{k}^{\prime} v_{0}=g_{4} v_{0} w_{k}^{\prime \prime}$ and $v_{1} u_{k}^{\prime}=u_{k}^{\prime \prime} v_{1} g_{4}$.

Consider,

$$
f_{8}=v_{0} w_{k}^{\prime \prime} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} v_{0} v_{1} \cdots u_{k}^{\prime} v_{0} v_{1} .
$$

If $f_{8}$ is unbordered, then $|u|<|w|-1$ since $\left|f_{8}\right| \leq|w|$ and

$$
|u|=\left|f_{8}\right|-\left|v_{0} w_{k}^{\prime \prime} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} w_{2}\right|+\left|v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} t^{\prime}\right|
$$

and $\left|t^{\prime}\right| \leq\left|z_{0}\right| \leq|z|<|b z| \leq\left|v_{0} v_{1}\right|$ and $w_{2} \neq \varepsilon$. Assume, $f_{8}$ is bordered. Then $f_{8}$ has a shortest border $h_{8}$ such that $v_{0} v_{1} \preccurlyeq h_{8}$.

If $\left|h_{8}\right|>\left|v_{0} w_{k}^{\prime \prime} v_{1}\right|$ then the same argument as in the case $\left|v_{1} w_{k}^{\prime} v_{0}\right|<\left|h_{4}\right|$ above shows that $|u|<|w|-1$. If $\left|h_{8}\right|<\left|v_{0} w_{k}^{\prime \prime} v_{1}\right|$ then $v_{0} v_{1}$ occurs in $w_{k}^{\prime}$; a contradiction. Hence, we have $h_{8}=v_{0} w_{k}^{\prime \prime} v_{1}$ and

$$
\begin{equation*}
w_{k}^{\prime} v_{0} v_{1}=g_{1} h_{8} \quad \text { and } \quad h_{8} \preccurlyeq u_{k}^{\prime} v_{0} v_{1} \tag{9}
\end{equation*}
$$

Consider,

$$
f_{9}=v_{0} v_{1} w_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} v_{0} v_{1} \cdots u_{k+1}^{\prime} v_{0} u_{k}^{\prime \prime} v_{1}
$$

If $f_{9}$ is unbordered, then $|u|<|w|-1$ since $\left|f_{9}\right| \leq|w|$ and

$$
|u|=\left|f_{9}\right|-\left|v_{0} v_{1} w_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} w_{2}\right|+\left|g_{4} v_{0} v_{1} v_{k-1}^{\prime} v_{0} v_{1} \cdots v_{1}^{\prime} v_{0} v_{1} t^{\prime}\right|
$$

and $\left|t^{\prime}\right| \leq\left|z_{0}\right| \leq|z|<|b z| \leq\left|v_{0} v_{1}\right|$ and $\left|g_{4}\right| \leq\left|w_{k}^{\prime}\right|$ and $w_{2} \neq \varepsilon$. Assume, $f_{9}$ is bordered. Then $f_{9}$ has a shortest border $h_{9}$ such that $v_{0} v_{1} \leq h_{9}$. We have $h_{9}=v_{0} u_{k}^{\prime \prime} v_{1}$ by the arguments from the previous paragraph. Moreover,

$$
\begin{equation*}
v_{0} v_{1} u_{k}^{\prime}=h_{9} g_{1} \quad \text { and } \quad h_{9} \leq v_{0} v_{1} w_{k}^{\prime} \tag{10}
\end{equation*}
$$

Observe, that (9) and (10) imply that the number of occurrences of $v_{1}$ and $v_{0}$, respectively, is the same in $w_{k}^{\prime}$ and $u_{k}^{\prime}$ since $v_{0}$ and $v_{1}$ do not overlap. Now, let

$$
h_{4}=v_{1} g_{4} v_{0}=h_{1}^{\prime \prime} v_{1} h_{1}^{\prime} v_{0}=v_{1} h_{0}^{\prime} v_{0} h_{0}^{\prime \prime}
$$

where $v_{1}$ and $v_{0}$ occur only once in $v_{1} h_{1}^{\prime}$ and $h_{0}^{\prime} v_{0}$, respectively.
Now, let

$$
f_{8}^{\prime}=v_{0} h_{0}^{\prime \prime} w_{k}^{\prime \prime} v_{1} v_{k-1}^{\prime} \cdots v_{0} v_{1} v_{1}^{\prime} v_{0} v_{1} w_{2} \cdot u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{k}^{\prime} v_{0} v_{1}
$$

and

$$
f_{9}^{\prime}=v_{0} v_{1} w_{k}^{\prime} v_{0} v_{1} v_{k-1}^{\prime} \cdots v_{0} v_{1} v_{1}^{\prime} v_{0} v_{1} w_{2} \cdot u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{k+1}^{\prime} v_{0} u_{1}^{\prime \prime} h_{1}^{\prime \prime} v_{1}
$$

with respective shortest borders $h_{8}^{\prime}$ and $h_{9}^{\prime}$ (which are both not empty, if $|u| \geq|w|-1$; as in the case of $f_{8}$ and $f_{9}$ ). Analogously to the cases of $f_{8}$ and $f_{9}$, we have

$$
w_{k}^{\prime} v_{0} v_{1}=h_{0}^{\prime} h_{8}^{\prime} \quad \text { and } \quad v_{0} v_{1} u_{k}^{\prime}=h_{9}^{\prime} h_{1}^{\prime}
$$

It is now straightforward to see that

$$
h_{8}^{\prime}=h_{9}^{\prime}=v_{0} v_{1}
$$

and

$$
h_{4}=v_{0} w_{k}^{\prime} v_{1}=v_{0} u_{k}^{\prime} v_{1}
$$

and hence, $w_{k}^{\prime}=u_{k}^{\prime}$. In this case, we denote both $w_{k}^{\prime}$ and $u_{k}^{\prime}$ by $v_{k}^{\prime}$.
Now, we have

$$
\begin{aligned}
\bar{v} & =v_{0} v_{1} w_{\iota}^{\prime} \cdots v_{0} v_{1} w_{2}^{\prime} v_{0} v_{1} w_{1}^{\prime} \\
& =v_{0} v_{1} u_{\iota}^{\prime} \cdots v_{0} v_{1} u_{2}^{\prime} v_{0} v_{1} u_{1}^{\prime}
\end{aligned}
$$

where $\iota=\min \{i, j\}$.
If $i<j$ then

$$
\begin{equation*}
\left|w_{0}^{\prime}\right|<\left|u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{i+1}^{\prime}\right| \tag{11}
\end{equation*}
$$

since $\left|w_{0}^{\prime}\right| \leq\left|u_{0}^{\prime}\right|$ by (5). Let

$$
f_{11}=v_{1} w_{2} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{i+1}^{\prime} \bar{v} v_{0}
$$

Then $|w|<\left|f_{11}\right|$ by (11), and hence, $f_{11}$ is bordered. Let $h_{11}=v_{1} g_{11} v_{0}$ be the shortest border of $f_{11}$. Recall, that $w_{2} \neq \varepsilon$ and either $a z \preccurlyeq v_{1} w_{2}$ or $v_{1} w_{2} \preccurlyeq a z$. If $\left|v_{1} w_{2}\right|<|a z|$ then $v_{1}$ necessarily occurs in $z$, and hence, it overlaps with $v_{0}$ (since $b z \leq v_{0} v_{1}$ ); a contradiction. So, we have $a z \preccurlyeq v_{1} w_{2}$. Surely, $\left|h_{11}\right|<\left|v_{1} w_{2}\right|$ (and so $h_{11} \leq v_{1} w_{2}$ ) for otherwise $a z$ occurs in $u$ which contradicts our assumption that $z$ is of maximum length. Let $w_{2}=g_{11} v_{0} w_{5}$. Note, that $\left|v_{0} w_{5}\right| \neq|a z|$ since $a z$ and $v_{0}$ begin with different letters. We have $|a z|<\left|v_{0} w_{5}\right|$ since otherwise $v_{0}$ occurs in $z$, and hence, overlaps with $v_{1}$ which is a contradiction. Consider,

$$
f_{12}=v_{0} w_{5} u_{0}^{\prime} v_{0} v_{1} u_{j}^{\prime} \cdots v_{0} v_{1} u_{i+1}^{\prime} \bar{v} v_{0} v_{1}
$$

If $f_{12}$ is unbordered, then $|u|<|w|-1$ since $\left|f_{12}\right| \leq|w|$ and

$$
|u|=\left|f_{12}\right|-\left|v_{0} w_{5}\right|+\left|t^{\prime}\right|
$$

and $|a z|<\left|v_{0} w_{5}\right|$ and $\left|t^{\prime}\right| \leq\left|z_{0}\right| \leq|z|<|b z|<\left|v_{0} w_{5}\right|$. Assume, $f_{12}$ is bordered. Then $f_{12}$ has a shortest border $h_{12}=g_{12} v_{0} v_{1}$ with $|a z|<\left|h_{12}\right|$, for otherwise $a z$ occurs in $u$. Let $v_{0} w_{5}=g_{12} v_{0} v_{1} w_{6}$. But, now

$$
w=w_{0}^{\prime} \bar{v} v_{0} v_{1} g_{12} v_{0} v_{1} w_{6}
$$

where $v_{0} v_{1} w_{6} \preccurlyeq w_{2}$, contradicting our assumption that $v_{0} v_{1}$ does not occur in $w_{2}$.

If $i>j$ then

$$
w=w_{0}^{\prime} v_{0} v_{1} w_{i}^{\prime} \cdots v_{0} v_{1} w_{j+1}^{\prime} \bar{v} v_{0} v_{1} w_{2} \quad \text { and } \quad u=u_{0}^{\prime} \bar{v} v_{0} v_{1} t^{\prime}
$$

and $|w| \geq|u|-\left|t^{\prime}\right|+\left|v_{0} v_{1}\right|$. We have $|u|<|w|-1$ since $\left|t^{\prime}\right| \leq\left|z_{0}\right|<\left|v_{0} v_{1}\right|-1$ by (3).

Assume $i=j$. Then

$$
w=w_{0}^{\prime} \bar{v} v_{0} v_{1} w_{2} \quad \text { and } \quad u=u_{0}^{\prime} \bar{v} v_{0} v_{1} t^{\prime}
$$

Consider

$$
f^{\prime}=v_{1} w_{2} u_{0}^{\prime} \bar{v} v_{0}
$$

If $f^{\prime}$ is bordered, then it has a shortest border $h^{\prime}=v_{1} g^{\prime} v_{0}$.


Recall, that $w_{2} \neq \varepsilon$ and either $a z \preccurlyeq v_{1} w_{2}$ or $v_{1} w_{2} \preccurlyeq a z$. If $\left|v_{1} w_{2}\right|<|a z|$ then $v_{1}$ occurs in $z$, and hence, overlaps with $v_{0}$ since $b z \leq v_{0} v_{1}$; a contradiction. So, we have $a z \preccurlyeq v_{1} w_{2}$. Surely, $\left|h^{\prime}\right|<\left|v_{1} w_{2}\right|$ for otherwise $a z$ occurs in $u$ which contradicts our assumption. Let $w_{2}=g^{\prime} v_{0} w_{4}$. Note, that $\left|v_{0} w_{4}\right| \neq|a z|$ since $a z$ and $v_{0}$ begin with different letters. We have $|a z|<\left|v_{0} w_{4}\right|$ since otherwise $v_{0}$ occurs in $z$, and hence, overlaps with $v_{1}$ which is a contradiction. Consider now,

$$
f^{\prime \prime}=v_{0} w_{4} u_{0}^{\prime} \bar{v} v_{0} v_{1}
$$

If $f^{\prime \prime}$ is unbordered, then it easily follows that $|u|<|w|-1$ since we have $\left|t^{\prime}\right|<|a z|<\left|v_{0} w_{4}\right|$.


If $f^{\prime \prime}$ is bordered, then it has a shortest border $h^{\prime \prime}=g^{\prime \prime} v_{0} v_{1}$ with $|a z|<\left|h^{\prime \prime}\right|$, for otherwise $a z$ occurs in $u$. Let $v_{0} w_{4}=g^{\prime \prime} v_{0} v_{1} w_{5}$. But, now

$$
w=w_{0}^{\prime} \bar{v} v_{0} v_{1} g^{\prime} g^{\prime \prime} v_{0} v_{1} w_{5}
$$

which contradicts our assumption that $w=w_{0}^{\prime} \bar{v} v_{0} v_{1} w_{2}$ and $v_{0} v_{1}$ does not occur in $w_{2}$.

If $f^{\prime}$ is unbordered, then $\left|f^{\prime}\right| \leq|w|$, and hence, $\left|w_{0}^{\prime}\right| \geq\left|u_{0}^{\prime}\right|$. But, we also have $\left|w_{0}^{\prime}\right| \leq\left|u_{0}^{\prime}\right|$; see (5). That implies $\left|w_{0}^{\prime}\right|=\left|u_{0}^{\prime}\right|$. Moreover, the factors $w_{0}$ and $b z v^{\prime}$ have both nonoverlaping occurrences in $u_{0}^{\prime} v_{0} v_{1}$ by (5). Therefore, $w_{0}^{\prime}=u_{0}^{\prime}$. Now,

$$
w=x a w_{7} \quad \text { and } \quad u=x b t^{\prime \prime}
$$

where $w_{0}^{\prime} \bar{v} v_{0} v_{1} \leq x$ and $a, b \in A$ and $a \neq b$ and $w_{7} \preccurlyeq w_{2}$ and $t^{\prime \prime} \preccurlyeq t^{\prime}$. We have that $x b$ occurs in $w$ by Theorem 11. Since $x b$ is not a prefix of $w$ and $v_{0} v_{1}$ does not overlap with itself, we have $|x b|+\left|v_{0} v_{1}\right| \leq|w|$. From $\left|t^{\prime}\right| \leq\left|z_{0}\right|<\left|v_{0} v_{1}\right|-1$ we get $|u|<|w|-1$ and the claim follows.

Note, that the bound $|u|<|w|-1$ on the length of a nontrivial Duval extension $w u$ of $w$ is tight, as the example given in the introduction shows. Theorem 2 also implies a new bound on the length of any word $w$ such that $\partial(w)=\mu(w)$ must hold.

Corollary 3. If $|w| \geq 3 \mu(w)-2$ then $\partial(w)=\mu(w)$.

## 5 Conclusions

In this paper we have given a confirmative answer to a long standing conjecture [10] by proving that a Duval extension $w u$ of $w$ longer than $2|w|-2$ is trivial. This bound is thight and also gives a new bound on the relation between the length of an arbitrary word $w$ and its longest unbordered factors $\mu(w)$, namely that $|w| \geq 3 \mu(w)-2$ implies $\partial(w)=\mu(w)$ as conjectured (more weakly) in [1]. We believe that the precise bound can be achieved with methods similar to those presented in this paper.

## References

[1] R. Assous and M. Pouzet. Une caractérisation des mots périodiques. Discrete Math., 25(1):1-5, 1979.
[2] J. Berstel and D. Perrin. Theory of codes, volume 117 of Pure and Applied Mathematics. Academic Press Inc., Orlando, FL, 1985.
[3] R. S. Boyer and J. S. Moore. A fast string searching algorithm. Commun. ACM, 20(10):762-772, October 1977.
[4] D. Breslauer, T. Jiang, and Z. Jiang. Rotations of periodic strings and short superstrings. J. Algorithms, 24(2), 1997.
[5] P. Bylanski and D. G. W. Ingram. Digital transmission systems. IEE, 1980.
[6] Y. Césari and M. Vincent. Une caractérisation des mots périodiques. C. R. Acad. Sci. Paris Sér. A, 286:1175-1177, 1978.
[7] M. Crochemore, F. Mignosi, A. Restivo, and S. Salemi. Text compression using antidictionaries. In 26th Internationale Colloquium on Automata, Languages and Programming (ICALP), Prague, volume 1644 of Lecture Notes in Comput. Sci., pages 261-270. Springer, Berlin, 1999.
[8] M. Crochemore and D. Perrin. Two-way string-matching. J. ACM, 38(3):651-675, 1991.
[9] J.-P. Duval. Périodes et répétitions des mots de monoïde libre. Theoret. Comput. Sci., 9(1):17-26, 1979.
[10] J.-P. Duval. Relationship between the period of a finite word and the length of its unbordered segments. Discrete Math., 40(1):31-44, 1982.
[11] A. Ehrenfeucht and D. M. Silberger. Periodicity and unbordered segments of words. Discrete Math., 26(2):101-109, 1979.
[12] T. Harju and D. Nowotka. Density of critical factorizations. Theor. Inform. Appl., 36(3):315-327, 2002.
[13] T. Harju and D. Nowotka. Duval's conjecture and Lyndon words. technical report 479, Turku Centre of Computer Science (TUCS), Turku, Finland, October 2002. submitted.
[14] T. Harju and D. Nowotka. Minimal Duval extensions. technical report 520, Turku Centre of Computer Science (TUCS), Turku, Finland, April 2003. submitted.
[15] D. E. Knuth, J. H. Morris, and V. R. Pratt. Fast pattern matching in strings. SIAM J. Comput., 6(2):323-350, 1977.
[16] M. Lothaire. Combinatorics on Words, volume 17 of Encyclopedia of Mathematics. Addison-Wesley, Reading, MA, 1983.
[17] M. Lothaire. Algebraic Combinatorics on Words, volume 90 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, United Kingdom, 2002.
[18] R. C. Lyndon. On Burnside's problem. Trans. Amer. Math. Soc., 77:202-215, 1954.
[19] D. Margaritis and S. Skiena. Reconstructing strings from substrings in rounds. In 36th Annual Symposium on Foundations of Computer Science (FOCS), pages 613-620, Milwaukee, WI, 1995. IEEE Computer Society.
[20] F. Mignosi and L. Q. Zamboni. A note on a conjecture of Duval and Sturmian words. Theor. Inform. Appl., 36(1):1-3, 2002.
[21] M. Morse and G. A. Hedlund. Symbolic dynamics II: Sturmian trajectories. Amer. J. Math., 61:1-42, 1940.
[22] M.-P. Schützenberger. A property of finitely generated submonoids of free monoids. In Algebraic theory of semigroups (Proc. Sixth Algebraic Conf., Szeged, 1976), volume 20 of Colloq. Math. Soc. János Bolyai, pages 545-576. North-Holland, Amsterdam, 1979.
[23] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. IEEE Trans. Information Theory, 23(3):337-343, 1977.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland
http://www.tucs.fi


University of Turku

- Department of Information Technology
- Department of Mathematics


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

