# Periodicity and Unbordered Words: A Proof of the Extended Duval Conjecture

TERO HARJU
University of Turku
and
DIRK NOWOTKA
University of Stuttgart

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 $n$. We call a word *bordered* if it has a proper prefix which is also a suffix of that word. Let $\mu(w)$ denote the maximum length of all unbordered factors of $w$, 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 the stronger version of an old conjecture by Duval (1983). It follows from this result that, in general, $n \geq 3\mu(w) - 3$ implies $\mu(w) = \partial(w)$ which gives an improved bound for the question raised by Ehrenfeucht and Silberger in 1979.

Categories and Subject Descriptors: F.4.m [**Mathematical Logic and Formal Languages**]: Miscellaneous

General Terms: Theory

Additional Key Words and Phrases: combinatorics on words, Duval's conjecture, periodicity, unbordered words

## 1. INTRODUCTION

Periodicity and borderedness are two properties of words which are investigated in this paper. These two fundamental notions play a rôle (explicitly or implicitly) in many areas. Just a few of those areas are string searching algorithms [Knuth et al. 1977; Boyer and Moore 1977; Crochemore and Perrin 1991], data compression [Ziv and Lempel 1977; Crochemore et al. 1999], and codes [Berstel and Perrin 1985]. These are classical examples, but also computational biology, e.g., sequence assembly [Margaritis and Skiena 1995] or superstrings [Breslauer et al. 1997], and serial data communications systems [Bylanski and Ingram 1980] are areas among others where periodicity and borderedness of words (sequences) are important concepts. It is well known that these two properties of words are not independent of 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

---

A short version of this paper appeared in the proceedings of the STACS conference 2004 (LNCS 2996:294–304, Springer-Verlag, Berlin, 2004).
Authors' addresses: T. Harju, Department of Mathematics, University of Turku, 20014 Turku, Finland; D. Nowotka, Institute for Formal Methods in Computer Science, University of Stuttgart, Universitätstr. 38, 70569 Stuttgart, Germany.

than 25 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 $w$ a *factor* of $w$. 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 maximum length of all unbordered factors 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 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 maximum length, that is, no unbordered factor is longer than that prefix. Let $w$ be an unbordered word. Then a word $wu$ is called a *Duval extension* (of $w$) if every unbordered factor of $wu$ has length at most $|w|$, that is, $\mu(wu) = |w|$. We call $wu$ a *trivial* Duval extension if $\partial(wu) = |w|$, or in other words, if $u$ is a prefix of $w^k$ for some $k \geq 1$. For example, let $w = abaabb$ and $u = aaba$. Then $wu = abaabbaaba$ is a nontrivial Duval extension of $w$ since ($i$) $w$ is unbordered, ($ii$) all factors of $wu$ longer than $w$ are bordered, that is, $|w| = \mu(wu) = 6$, and ($iii$) the period of $wu$ is 7, and hence, $\partial(wu) > |w|$. Note that this example satisfies $|u| = |w| - 2$.

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

CONJECTURE 1.1. *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 [Lothaire 2002]. The most recent results are by Mignosi and Zamboni [2002] and the authors of this article [Harju and Nowotka 2002b]. 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 [Mignosi and Zamboni 2002] that unbordered, finite factors of Sturmian words allow only trivial Duval extensions; in 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 subword complexity, that is, a Sturmian word contains exactly $n + 1$ different factors of length $n$ for every $n \geq 1$; see [Morse and Hedlund 1940] or Chapter 2 in [Lothaire 2002]. This result was later improved [Harju and Nowotka 2002b] by showing that Lyndon words [Lyndon 1954] allow only trivial Duval extensions and the fact that every unbordered, finite factor of a Sturmian word is a Lyndon word but not vice versa. A Lyndon word is a primitive word that is minimal among all its conjugates with respect to some lexicographic order.

The main result in this paper is a proof of the extended version of Conjecture 1.1.

THEOREM 1.2. *Let $wu$ be a nontrivial Duval extension of $w$. Then $|u| < |w| - 1$.*

The example mentioned above already indicates that this bound on the length of a nontrivial Duval extension is tight. An example for arbitrary lengths of $w$ is given later in Section 4. Recently, a new proof of Theorem 1.2 was given by Holub in [2005]. Theorem 1.2 implies the truth of Duval's conjecture, as well as the following corollary (for any word $w$).

COROLLARY 1.3. *If* $|w| \geq 3\mu(w) - 3$*, then* $\partial(w) = \mu(w)$.

This corollary (see Section 4) confirms the conjecture by Assous and Pouzet in [1979] about a question asked by Ehrenfeucht and Silberger in [1979].

Our main result, Theorem 1.2, is presented in Section 4 and its corollary in Section 5. Sections 4 and 5 use the notation introduced in Section 2 and preliminary results from Section 3. We conclude with Section 6.

## 2. NOTATION

In this section we introduce the notation of this paper. We refer to [Lothaire 1983; 2002] 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 $\varepsilon$. We denote the $i$-th letter of a word $w$ with $w_{(i)}$.[1] Let $w = w_{(1)}w_{(2)} \cdots w_{(n)}$. The word $w_{(n)} \cdots w_{(2)}w_{(1)}$ is called the *reversal* of $w$ denoted by $\widetilde{w}$. We denote the length $n$ of $w$ by $|w|$. If $w$ is not empty, then let $w^\bullet = w_{(1)}w_{(2)} \cdots w_{(n-1)}$. We define $\varepsilon^\bullet = \varepsilon$. 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 word $w$ is called *primitive* if $w^k$ implies $k = 1$, that is, $\partial(w)$ does not divide $|w|$. A *conjugate* of $w$ is a word $w' = uv$ such that $vu = w$. Note that every conjugate of $w$ occurs in $ww^\bullet$. A nonempty word $u$ is called a *border* of a word $w$, if $w = uv = v'u$ for some nonempty words $v$ and $v'$. We call $w$ *bordered* if it has a border, otherwise $w$ is called *unbordered*. Note that every unbordered word is primitive and every bordered word $w$ has a minimum border $u$ such that $w = uvu$, where $u$ is unbordered. Let $\mu(w)$ denote the maximum length of unbordered factors of $w$. We have that

$$\mu(w) \leq \partial(w) \ .$$

Indeed, let $u = u_{(1)}u_{(2)} \cdots u_{(\mu(w))}$ be an unbordered factor of $w$. If $\mu(w) > \partial(w)$ then $u_{(i)} = u_{(i+\partial(w))}$ for all $1 \leq i \leq \mu(w) - \partial(w)$ and $u_{(1)}u_{(2)} \cdots u_{(\mu(w)-\partial(w))}$ is a border of $u$; a contradiction.

Suppose $w = uv$, then $u$ is called a *prefix* of $w$, denoted by $u \leq_{\mathrm{p}} w$, and $v$ is called a *suffix* of $w$, denoted by $v \leq_{\mathrm{s}} w$. If $u$ and $v$ are both not the empty word, then $u$ is called *proper prefix* of $w$, denoted by $u <_{\mathrm{p}} w$, and $v$ is called *proper suffix* of $w$, denoted by $v <_{\mathrm{s}} w$. Let $u$ and $v$ be two nonempty words. We say that $u$ *overlaps* $v$ *from the left* (resp. *from the right*) if there is a word $w$ such that $|w| < |u| + |v|$, and $u <_{\mathrm{p}} w$ and $v <_{\mathrm{s}} w$, (resp. $v <_{\mathrm{p}} w$ and $u <_{\mathrm{s}} w$). We say that $u$ *overlaps* with $v$, if $u$ overlaps $v$ from the left or right. We say that $u$ *intersects* with $v$, if $u$ and $v$ overlap or one is a factor of the other.

*Example* 2.1. Let $A = \{a, b\}$ and $u, v, w \in A^*$ such that $u = abaa$ and $v = baaba$ and $w = abaaba$. Then $|w| = 6$, and 3, 5, and 6 are periods of $w$, and $\partial(w) = 3$.

[1] In general, subscripts without brackets are used for variables in $A^*$, for example $w_i \in A^*$, and subscripts with brackets for variables in $A$, for example $w_{(i)} \in A$.

We have that $a$ is the shortest border of $u$ and $w$, whereas $ba$ is the shortest border of $v$. We have $\mu(w) = 3$. We also have that $u$ and $v$ overlap since $u \leq_{\mathrm{p}} w$ and $v \leq_{\mathrm{s}} w$ and $|w| < |u| + |v|$.

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

*Example* 2.2. Let $w = abaabbabaababb$ and $u = aaba$. Then

$$w.u = abaabbabaababb.aaba$$

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

$$w.u' = abaabbabaababb.aa$$

is minimal, with $u' = aa \leq_{\mathrm{p}} u$. Note that $wu$ is not the longest nontrivial Duval extension of $w$ since

$$w.v = abaabbabaababb.abaaba$$

is longer, with $v = abaaba$ and $|wv| = 20$ and $\partial(wv) = 17$. One can check that $wv$ 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 $w = xy$ with $|x| = p$ and there exist words $x'$ and $y'$ such that $u \leq_{\mathrm{s}} x'x$ and $u \leq_{\mathrm{p}} yy'$. 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 that the repetition word of length $\partial(w, p)$ at point $p$ is necessarily unbordered and $\partial(w, p) \leq \partial(w)$. A factorization $w = uv$, 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*.

*Example* 2.3. The word

$$w = ab.aa.b$$

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.

Let us consider alphabets of any finite size larger than one for the rest of this paper.

## 3. PRELIMINARY RESULTS

We state some auxiliary and well-known results about repetitions and borders in this section. These results will be used to prove Theorem 1.2 and Corollary 1.3 in Section 4. The first lemma recalls a well-known fact.

LEMMA 3.1. *Let $w$ be a primitive word over a $k$-letter alphabet. Then there exist at least $k$ unbordered conjugates of $w$.*

Indeed, for every letter $a$ in an alphabet $A$ a lexicographic order $\lhd_a$ can be chosen such that $a$ is minimal in $A$. It is not hard to show that the smallest conjugate $w'$ of $w$ with respect to $\lhd_a$ is unbordered. Note that $a \leq_{\mathrm{p}} w'$, and hence, every smallest conjugate with respect to a chosen order is different for a different letter.

LEMMA 3.2. *Let $zf = gzh$ where $f, g \neq \varepsilon$. Let $az'$ be the maximum unbordered prefix of $az$ where $a$ is a letter. If $az$ does not occur in $zf$, then $agz'$ is unbordered.*

PROOF. Assume $agz'$ is bordered, and let $y$ be its shortest border. In particular, $y$ is unbordered. If $|z'| \geq |y|$ then $y$ is a border of $az'$ which is a contradiction. If $|az'| = |y|$ or $|az| < |y|$ then $az$ occurs in $zf$ which is again a contradiction. If $|az'| < |y| \leq |az|$ then $az'$ is not maximum since $y$ is unbordered; a contradiction. □

The proof of the following lemma is easy and therefore omitted.

LEMMA 3.3. *Let $w$ be an unbordered word and $u \leq_{\mathrm{p}} w$ and $v \leq_{\mathrm{s}} w$. Then $uw$ and $wv$ are unbordered.*

The critical factorization theorem (CFT) is one of the main results about periodicity of words. A weak version of it was first conjectured by Schützenberger [1979] and proved by Césari and Vincent [1978]. It was developed into its current form by Duval [1979]. We refer to [Harju and Nowotka 2002a] for a short proof of the CFT.

THEOREM 3.4 CFT. *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., $\partial(w, |u|) = \partial(w)$.*

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

LEMMA 3.5. *Let $w = uv$ be unbordered and $|u|$ be a critical point of $w$. Then $u$ and $v$ do not intersect.*

PROOF. Note that $\partial(w, |u|) = \partial(w) = |w|$ since $w$ is unbordered. Let $|u| \leq |v|$ without loss of generality. Assume that $u$ and $v$ do intersect. First, if $u = u's$ and $v = sv'$ for a nonempty $s$, then $\partial(w, |u|) \leq |s| < |w|$. On the other hand, if $u = su'$ and $v = v's$, then $s$ is a border of $w$. Finally, if $v = sut$, then $\partial(w, |u|) \leq |su| < |w|$. These contradictions prove the claim. □

The next result follows from Lemma 3.5.

LEMMA 3.6. *Let $w = u_0u_1$ be unbordered and $|u_0|$ be a critical point of $w$. Then $u_0xu_1$ (resp. $u_1xu_0$) is either unbordered or has a minimum border $g$ such that $|g| \geq |u_0| + |u_1|$ for any word $x$.*

PROOF. Indeed, since $|u_0|$ is critical for $w$ (for which $\partial(w) = |w|$), the words $u_0$ and $u_1$ are not factors of each other, and no suffix of $u_0$ can be a prefix of $u_1$. Therefore if $g$ is a border of $u_0 x u_1$, then it must be of the form $u_0 y u_1$ for some $y$. □

The next theorem states a basic fact about minimal Duval extensions; see [Harju and Nowotka 2004] for a proof of it.

THEOREM 3.7. *Let $wu$ be a minimal Duval extension of the unbordered word $w$. Then $au$ occurs in $w$ where $a$ is the last letter of $w$.*

The following Lemmas 3.8, 3.9 and 3.10 and Corollary 3.11 are given in [Duval 1982]. Let $a_0, a_1 \in A$, with $a_0 \neq a_1$, and $t_0 \in A^*$. Let the sequences $(a_i)$, $(s_i)$, $(s'_i)$, $(s''_i)$, and $(t_i)$, for $i \geq 1$, be defined by

—$a_i = a_{i \pmod 2}$, that is, $a_i = a_0$ (resp. $a_i = a_1$), if $i$ is even (resp. odd),
—$s_i$ such that $a_i s_i$ is the shortest border of $a_i t_{i-1}$,
—$s'_i$ such that $a_{i+1} s'_i$ is the longest unbordered prefix of $a_{i+1} s_i$,
—$s''_i$ such that $s'_i s''_i = s_i$,
—$t_i$ such that $t_i s''_i = t_{i-1}$.

For any parameters of the above definition, the following holds.

LEMMA 3.8. *For any $a_0$, $a_1$, and $t_0$ there exists an $m \geq 1$ such that*

$$|s_1| < \cdots < |s_m| = |t_{m-1}| \leq \cdots \leq |t_0|$$

*and $s_m = t_{m-1}$ and $|t_0| \leq |s_m| + |s_{m-1}|$.*

LEMMA 3.9. *Let $z \leq_{\mathrm{p}} t_0$ such that neither of $a_0 z$ and $a_1 z$ occurs 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 $|t_0| \leq |s_m| + |z_0|$,*
*(3) if $m > 1$ is even, then $a_0 s_m$ is unbordered and $|t_0| \leq |s_m| + |z_1|$.*

LEMMA 3.10. *Let $v$ be an unbordered factor of the unbordered word $w$ of length $\mu(w)$. If $v$ occurs twice in $w$, then $\mu(w) = \partial(w)$.*

COROLLARY 3.11. *Let $wu$ be a Duval extension of the unbordered word $w$. If $w$ occurs twice in $wu$, then $wu$ is a trivial Duval extension.*

## 4. MAIN RESULT

The extended Duval conjecture is proven in this section.

THEOREM 1.2. *Let $wu$ be a nontrivial Duval extension of the unbordered word $w$. Then $|u| < |w| - 1$.*

PROOF. Recall that every factor of $wu$ longer than $|w|$ is bordered since $wu$ is a Duval extension of $w$. Let $z$ be the longest suffix of $w$ that occurs twice in $zu$, the second occurrence possibly overlapping with the first $z$.

Assume first that $z = \varepsilon$. Then the last letter $a$ of $w$ does not occur in $u$. Let $w = u'bw''$ and $u = u'cu''$ such that $b, c \in A$ and $b \neq c$. Now $wu'c$ is a minimal Duval extension of $w$, and by Theorem 3.7, $w$ has the form $w = w'_0 a u' c w'_1$, where

$a$ is the last letter of $w$. Consider the factor $x = au'cw'_1u$. If it is unbordered then $|u| < |x| \leq |w|$ and so $|u| < |w| - 1$. Otherwise, the shortest border $g$ of $x$ satisfies $|au| \leq |g|$, since, in this case, $a$ does not occur in $u$. Since now $g$ occurs in $w$, we have $|u| < |w| - 1$ as claimed.

Assume now that $z \neq \varepsilon$. Also, $z \neq w$, since $wu$ is otherwise trivial by Corollary 3.11. Note that $bz$ does not overlap $az$ from the right, since such an overlap would give $azz' = z''bz$ where $|z'| \leq |z|$ and $wz'$ would be unbordered by Lemma 3.3. Thus there are letters $a, b \in A$ such that

$$w = w'az \qquad \text{and} \qquad u = u'bzr$$

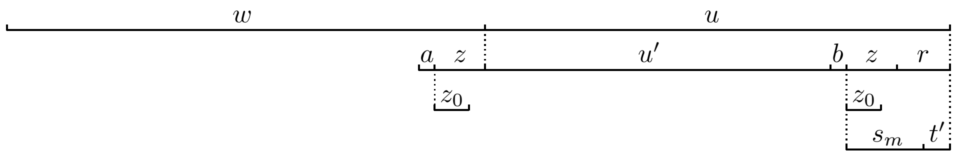
where $u' \neq \varepsilon$ and $z$ occurs in $zr$ only once, that is, $bz$ matches the rightmost occurrence of $z$ in $u$. Naturally, $a \neq b$ by maximality of $z$. Also, $w' \neq \varepsilon$, for otherwise $w = az$ and the prefix $azu'bz$ of $wu$ is bordered, say with the shortest border $g$, but then either $w$ is bordered (if $|g| \leq |z|$) or $az$ occurs in $zu$ (if $|g| > |z|$); a contradiction in both cases.

Let $az_0$ and $bz_1$ denote the longest unbordered prefix of $az$ and $bz$, respectively. Let $a_0 = a$ and $a_1 = b$ and $t_0 = zr$ and the integer $m$ be defined as in Lemma 3.9. We have then a word $s_m$, with its properties defined by Lemma 3.9, such that

$$t_0 = s_m t' \ .$$

Consider $x' = azu'bz_0$. We have $az \leq_{\mathrm{p}} a_0zu$ and $x' \leq_{\mathrm{p}} a_0zu$, and $bz_0 \leq_{\mathrm{s}} x'$. Also, $az$ occurs only as a prefix in $x'$. It follows from Lemma 3.2 that $x'$ is unbordered (where $z' = z_0$ and $f = u'bzr$ and $g = zu'b$ and $h = r$ in Lemma 3.2), and hence,

$$|x'| = |azu'bz_0| \leq |w| \ . \tag{1}$$




In the following we separately consider the two cases of even and odd parity of $m$.

Claim 4.1. *If $m$ is even then $|u| < |w| - 1$.*

Now $m \geq 2$ and $as_m$ $(= a_m s_m)$ is unbordered since $m$ is even, and $|t_0| \leq |s_m| + |z_1|$ by Lemma 3.9.

**Case:** Let $|t_0| = |s_m| + |z_1|$ with $z_1 = z$. Then $|z| \leq |s_{m-1}|$ by Lemma 3.8, and moreover, $a_{m-1}s_{m-1}$ is the shortest border of $a_{m-1}t_{m-2} = bt_{m-2} \leq_{\mathrm{p}} bt_0 = bzr$. Because $bs_{m-1}$ occurs twice in $bt_{m-2}$ and $zr$ marks the rightmost occurence of $z$ in $u$, we have that $z$ is not a proper prefix of $s_{m-1}$, and therefore, $|s_{m-1}| \leq |z|$. Hence, $|s_{m-1}| = |z|$.

Note that we have an immediate contradiction if $m = 2$ since then $|s_1| < |z|$ which contradicts $|z| \leq |s_{m-1}|$. Assume $m > 2$. But now, $bz$ occurs in $t_0$ since $bs_{m-1}$ is a border of $bt_{m-2}$ and $t_i \leq_{\mathrm{p}} t_0$, for all $0 \leq i < m$, which is a contradiction.

**Case:** Let $|t_0| < |s_m| + |z_1|$ or $|z_1| < |z|$. Then $|t'| < |z|$.

**Subcase:** Let $|s_m| \leq |z_0|$. According to (1), $|azu'bz_0| \leq |w|$, and so

$$\begin{aligned} |u| &= |azu| - |z| - 1 \\ &= |azu'bz_0| - |z_0| + |t_0| - |z| - 1 \\ &< |azu'bz_0| - |z_0| + |s_m| + |z_1| - |z| - 1 \\ &\leq |w| + |z_1| - |z| - 1 \\ &\leq |w| - 1 \end{aligned}$$

if $|t_0| < |s_m| + |z_1|$, or

$$\begin{aligned} |u| &= |azu| - |z| - 1 \\ &= |azu'bz_0| - |z_0| + |t_0| - |z| - 1 \\ &\leq |azu'bz_0| - |z_0| + |s_m| + |z_1| - |z| - 1 \\ &\leq |w| + |z_1| - |z| - 1 \\ &< |w| - 1 \end{aligned}$$

if $|z_1| < |z|$. We have $|u| < |w| - 1$ in both cases.

**Subcase:** Let $|s_m| > |z_0|$. We have that $as_m$ is unbordered, and since $az_0$ is the longest unbordered prefix of $az$, necessarily $az$ is a proper prefix of $as_m$, and hence, $|z| < |s_m|$. Now, $azu'bs_m$ is unbordered, for otherwise its shortest border is longer than $az$, since no prefix of $az$ is a suffix of $as_m$, and $az$ occurs in $u$; a contradiction. We have $|azu'bs_m| \leq |w|$ and similarly to the previous subcase, we obtain

$$\begin{aligned} |u| &= |azu| - |z| - 1 \\ &= |azu'bs_m| - |s_m| + |t_0| - |z| - 1 \\ &< |azu'bs_m| - |s_m| + |s_m| + |z_1| - |z| - 1 \\ &\leq |w| + |z_1| - |z| - 1 \\ &\leq |w| - 1 \end{aligned}$$

if $|t_0| < |s_m| + |z_1|$, or

$$\begin{aligned} |u| &= |azu| - |z| - 1 \\ &= |azu'bs_m| - |s_m| + |t_0| - |z| - 1 \\ &\leq |azu'bs_m| - |s_m| + |s_m| + |z_1| - |z| - 1 \\ &\leq |w| + |z_1| - |z| - 1 \\ &< |w| - 1 \end{aligned}$$

if $|z_1| < |z|$. We have $|u| < |w| - 1$ in both cases.

This proves Claim 4.1.

CLAIM 4.2. *If $m$ is odd then $|u| < |w| - 1$.*

The word $bs_m$ $(= a_m s_m)$ is unbordered, since $m$ is odd. We have $|t_0| \leq |s_m| + |z_0|$; see Lemma 3.9. Note that $t_0 = s_m$ and $t' = \varepsilon$ by Lemma 3.9, if $m = 1$. Surely $s_m \neq \varepsilon$. In particular, $|t'| \leq |z_0|$.

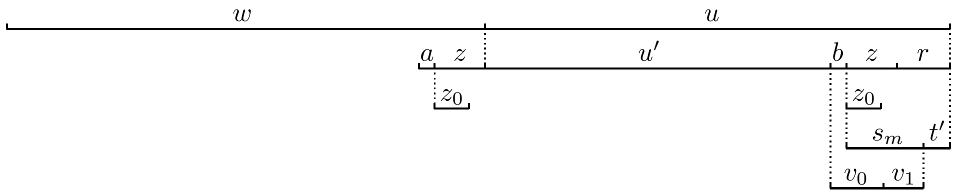
If $|s_m| < |z|$, then $|u| < |w| - 1$, since

$$|u| = |azu'bz_0| - |bz_0| + |bt_0| - |az|$$

and $|azu'bz_0| \leq |w|$, by (1), and $|t_0| \leq |s_m| + |z_0|$.

Assume thus that $|s_m| \geq |z|$, and hence, also $z \leq_\mathrm{p} s_m$. Since $s_m \neq \varepsilon$, we have $|bs_m| \geq 2$, and therefore, by the critical factorization theorem, there exists a critical point $p$ in $bs_m$ such that $bs_m = v_0v_1$, where $|v_0| = p$. In particular,

$$bz \leq_\mathrm{p} v_0v_1 \ . \tag{2}$$




Claim 4.3. *The factor $v_0v_1$ occurs in $w$.*

Let, $u'_0$ and $u_1$ be such that

$$u = u'_0v_0v_1u_1$$

where $v_0v_1$ does not occur in $u'_0$. Note that $v_0v_1$ does not overlap with itself since it is unbordered, and $v_0$ and $v_1$ do not intersect by Lemma 3.5. Consider the prefix $wu'_0bz$ of $wu$ which is bordered and has a shortest border $g$ with $|g| > |z|$. Hence, $bz \leq_\mathrm{s} g$, for otherwise $w$ would be bordered since $z \leq_\mathrm{s} w$. Moreover, $g \leq_\mathrm{p} w$, for otherwise $az$ would occur in $u$, and hence, $bz$ occurs in $w$. Let

$$w = w_0bzw_1 \tag{3}$$

such that $bz$ occurs in $w_0bz$ only once, that is, we consider the leftmost occurrence of $bz$ in $w$. Note that

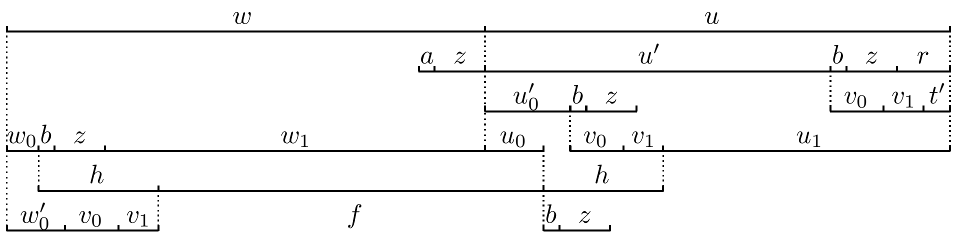
$$|w_0bz| \leq |g| \leq |u'_0bz| \tag{4}$$

where the first inequality comes (3) and the second inequality from the fact that $|u'_0bz| < |g|$ implies that $w$ is bordered. Let

$$f = bzw_1u'_0v_0v_1 \ .$$

If $f$ is unbordered, then $|f| \leq |w|$, and hence, $|u'_0v_0v_1| \leq |w_0|$. Now, we have $|u'_0| < |w_0|$, which contradicts (4).

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




Surely, $|bz| < |h|$, otherwise $v_0v_1$ is bordered by (2). So, $bz \leq_\mathrm{p} h$. Moreover, $|v_0v_1| \leq |h|$ otherwise $bz$ occurs in $s_m$ contradicting our assumption that $bzr$ marks the rightmost occurrence of $bz$ in $u$. So, $v_0v_1 \leq_\mathrm{s} h$, and $v_0v_1$ occurs in $w$ since $w_0h \leq_\mathrm{p} w$ by (4). Note that $|h| \leq |u'_0v_0v_1|$ otherwise $|h| > |azu'_0v_0v_1|$ (since

$bz \leq_{\rm p} h$) and $az$ occurs twice in $w$ such that $w = \bar{w}_1 az \bar{w}_2 az$, but then, $az\bar{w}_2 azu'bz_0$ is unbordered (see (1) above) and $|u| < |w| - 1$ since $|zr| < |v_0v_1t'| \leq |v_0v_1| + |z|$ and $|w| > |\bar{w}_1| + |v_0v_1| + |az| > |\bar{w}_1| + |v_0v_1| + |z| > |\bar{w}_1| + |zr| \geq |u|$.

This proves Claim 4.3.

CLAIM 4.4. *Let $u_0$ be such that $u'_0v_0v_1 = u_0h$. Then $w_0$, as defined in (3), occurs in $u_0$.*

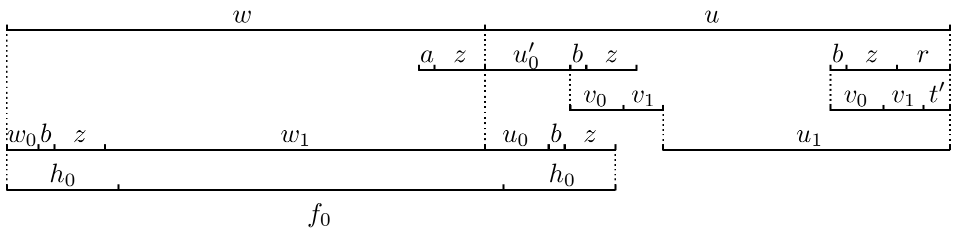
Let $v'$ and $w'_0$ be such that

$$w_0bzv' = w_0h = w'_0v_0v_1 \ .$$

Note that $v_0v_1$ does not occur in $w'_0$, otherwise $h = xv_0v_1y$ with $y \neq \varepsilon$ (since all occurrences of $bz$ in $w_0$ are also in $h$ and $bz \leq_{\rm p} v_0v_1$) and $v_0v_1$ also occurs in $u'_0$ (since $v_0v_1$ does not overlap itself) contradicting our assumption on $u'_0$. We have $h = bzv' \leq_{\rm s} u'_0v_0v_1$ (see previous figure). Consider

$$f_0 = wu_0bz$$

with the shortest border $h_0$.




Surely, $bz \leq_{\rm s} h_0$ otherwise $w$ is bordered with a suffix of $z$. Moreover, $|w_0bz| \leq |h_0|$ and $|h_0| \leq |u_0bz|$, since $bz$ does not occur in $w_0$ and $w$ is unbordered. From this and $w_0h = w'_0v_0v_1$ and $u_0h = u'_0v_0v_1$ it follows $|w'_0| \leq |u'_0|$ and

$$u'_0v_0v_1 = u_0bzv' \text{ and } w_0 \text{ occurs in } u_0. \tag{5}$$

This proves Claim 4.4.

Let now

$$w = w'_0v_0v_1w'_i \cdots v_0v_1w'_2v_0v_1w'_1v_0v_1w_2$$

for some word $w_2$ that does not contain $v_0v_1$, and

$$u = u'_0v_0v_1u'_j \cdots v_0v_1u'_2v_0v_1u'_1v_0v_1t'$$

such that $v_0v_1$ does not occur in $w'_k$, for all $0 \leq k \leq i$, or $u'_\ell$, 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 $v_0v_1 \leq_{\rm s} w$ and $az \leq_{\rm s} v_0v_1$, and $az$ would occur in $u$; a contradiction.)

We show in the following that $i = j$ and $w'_k = u'_k$ for all $1 \leq k \leq i$ if $|u| \geq |w| - 1$.

CLAIM 4.5. *It holds that $w'_k = u'_k$ for all $1 \leq k \leq \min\{i, j\}$.*

The proof goes by induction on $k$.

**Case:** First let $k = 1$. We show that $w'_1 = u'_1$. Consider

$$f_1 = v_1w'_1v_0v_1w_2u'_0v_0v_1u'_j \cdots v_0v_1u'_1v_0 \ .$$

If $f_1$ is unbordered, then $|u| < |w| - 1$ since $|f_1| \leq |w|$ and

$$|u| = |f_1| - |v_1 w_1' v_0 v_1 w_2| + |v_1 t'|$$

and $|t'| \leq |z_0| \leq |z| < |bz| \leq |v_0 v_1|$ and $w_2 \neq \varepsilon$. Assume that $f_1$ is bordered, and let $h_1$ be its shortest border. We have that $h_1 = v_1 g_1 v_0$ for some $g_1$ (possibly empty), since $v_0$ and $v_1$ do not intersect. We show that $h_1 \leq_{\mathrm{p}} v_1 w_1' v_0$. Indeed, otherwise we have one of the following cases.

(1) If $v_1 w_1' v_0 v_1 w_2 \leq_{\mathrm{p}} h_1$ then $az$ occurs in $u$; a contradiction to our assumption on $az$.

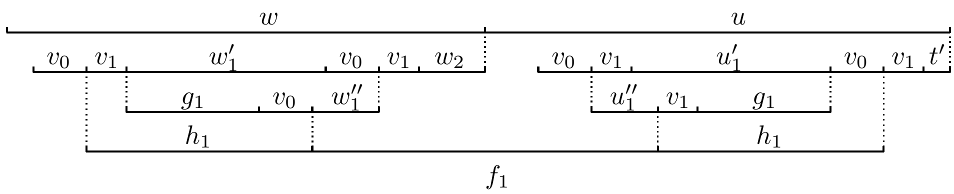
(2) If $|v_1 w_1' v_0 v_1 w_2| - |az| + |v_0| < |h_1| < |v_1 w_1' v_0 v_1 w_2|$ and $|v_0| \leq |z|$ then $v_0$ and $v_1$ intersect and $v_0$ occurs in $z$, contradicting Lemma 3.5.

(3) If $v_0$ occurs in $w_2$, then let $v_0 w_3 \leq_{\mathrm{s}} w_2$ for some $w_3$, and if $|az| \leq |v_0 w_3|$. Then we have that $v_0 w_3 u' v_0 v_1$ is unbordered (since otherwise its border is at least as long as $v_0 v_1$, because $v_0$ and $v_1$ do not intersect, that is, $v_0 v_1$ is a suffix of that border and therefore it is longer than $|az|$, but then $az$ occurs in $u$ which is a contradiction). But now $|t'| < |v_0 w_3| - 1$, since $|t'| < |az|$ and $|az| < |v_0 w_3|$, for $v_0$ does not begin with $a$, and $|u| < |w| - 1$ follows.

(4) If $|v_1 w_1' v_0| < h_1 < |v_1 w_1' v_0 v_1 v_0|$ then $v_0$ and $v_1$ intersect; a contradiction.

Moreover, $h_1 \leq_{\mathrm{s}} v_1 u_1' v_0$ since otherwise $v_0 v_1 \leq_{\mathrm{p}} h_1$ and $v_0 v_1$ occurs in $v_1 w_1' v_0$; a contradiction. Let $w_1''$ and $u_1''$ be such that

$$w_1' v_0 = g_1 v_0 w_1'' \qquad \text{and} \qquad v_1 u_1' = u_1'' v_1 g_1 \,. \tag{6}$$




Consider,

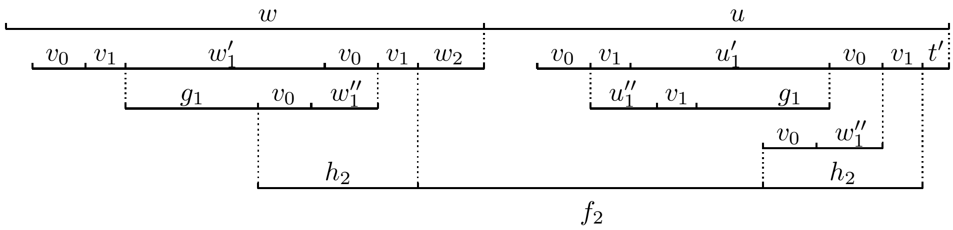
$$f_2 = v_0 w_1'' v_1 w_2 u_0' v_0 v_1 u_j' \cdots v_0 v_1 u_1' v_0 v_1 \ .$$

If $f_2$ is unbordered, then $|u| < |w| - 1$ since $|f_2| \leq |w|$ and

$$|u| = |f_2| - |v_0 w_1'' v_1 w_2| + |t'|$$

and $|t'| \leq |z_0| \leq |z| < |bz| \leq |v_0 v_1|$ and $w_2 \neq \varepsilon$. Assume that $f_2$ is bordered, and let $h_2$ be its shortest border. Since $v_0$ and $v_1$ do not intersect, $v_0 v_1 \leq_{\mathrm{s}} h_2$. Also $h_2 \leq_{\mathrm{p}} v_0 w_1'' v_1$ since $v_0 v_1$ does not occur in $w_2$ (and $v_0$ and $v_1$ do not intersect) and $az$ does not occur in $h_2$ (and so $h_2$ does not stretch beyond $w$). We have $v_0 w_1'' v_1 \leq_{\mathrm{p}} h_2$ since $v_0 v_1$ occurs in $v_0 w_1'' v_1$ only as a suffix. Hence, $h_2 = v_0 w_1'' v_1$. Note that $|h_2| \leq |u_1' v_0 v_1|$ since otherwise $|h_2| \geq |v_0 v_1 u_1' v_0 v_1|$ (because $v_0$ and $v_1$ do not intersect) and $v_0 v_1$ occurs twice in $h_2$, but $v_0 v_1$ occurs only once in $h_2$ since it occurs only once in $w_1' v_0 v_1 w_2$ and $az$ does not occur in $h_2$. Hence,

$$w_1' v_0 v_1 = g_1 h_2 \qquad \text{and} \qquad h_2 \leq_{\mathrm{s}} u_1' v_0 v_1 \ . \tag{7}$$



Consider,

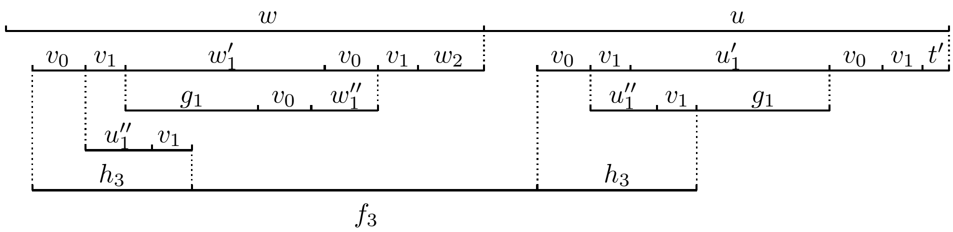
$$f_3 = v_0 v_1 w_1' v_0 v_1 w_2 u_0' v_0 v_1 u_j' \cdots v_0 v_1 u_2' v_0 u_1'' v_1 \ .$$

If $f_3$ is unbordered, then $|u| < |w| - 1$ since $|f_3| \le |w|$ and

$$|u| = |f_3| - |v_0 v_1 w_1' v_0 v_1 w_2| + |g_1 v_0 v_1 t'|$$

and $|t'| \le |z_0| \le |z| < |bz| \le |v_0 v_1|$ and $|g_1| \le |w_1'|$ and $w_2 \ne \varepsilon$. Assume $f_3$ is bordered. Then $f_3$ has a shortest border $h_3$ such that $v_0 v_1 \le_{\mathrm{p}} h_3$ since $v_0$ and $v_1$ do not intersect. We have $h_3 = v_0 u_1'' v_1$ by the arguments of the previous paragraph. Moreover,

$$v_0 v_1 u_1' = h_3 g_1 \qquad \text{and} \qquad h_3 \le_{\mathrm{p}} v_0 v_1 w_1' \ . \tag{8}$$

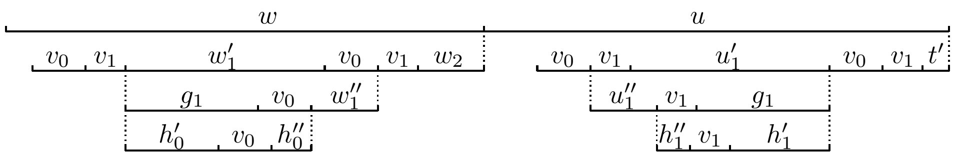


Observe, that (7) and (8) imply that the number of occurrences of $v_0$ and $v_1$, respectively, is the same in $w_1'$ and $u_1'$ since $v_0$ and $v_1$ do not intersect. Let

$$h_1 = v_1 g_1 v_0 = h_1'' v_1 h_1' v_0 = v_1 h_0' v_0 h_0''$$

where

$$v_1 \text{ and } v_0 \text{ occur only once in } v_1 h_1' \text{ and } h_0' v_0 \text{, respectively.} \tag{9}$$



Let

$$f_2' = v_0 h_0'' w_1'' v_1 w_2 u_0' v_0 v_1 u_j' \cdots v_0 v_1 u_1' v_0 v_1$$

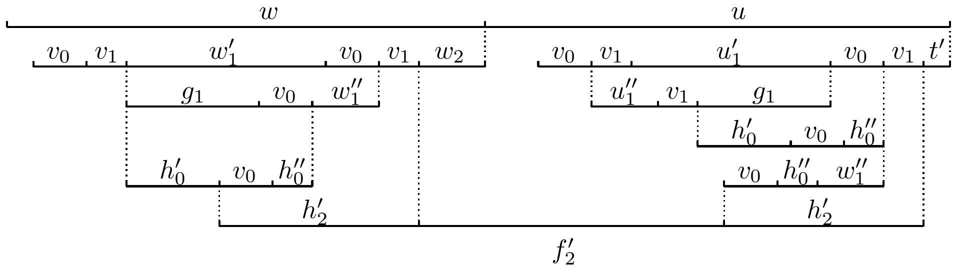
and

$$f_3' = v_0 v_1 w_1' v_0 v_1 w_2 u_0' v_0 v_1 u_j' \cdots v_0 v_1 u_2' v_0 u_1'' h_1'' v_1$$

with the respective shortest borders $h_2'$ and $h_3'$ (which do both exist if $|u| \ge |w| - 1$; as in the case of $f_2$ and $f_3$) and $v_0 v_1 \le_{\mathrm{s}} h_2'$ and $v_0 v_1 \le_{\mathrm{p}} h_3'$.

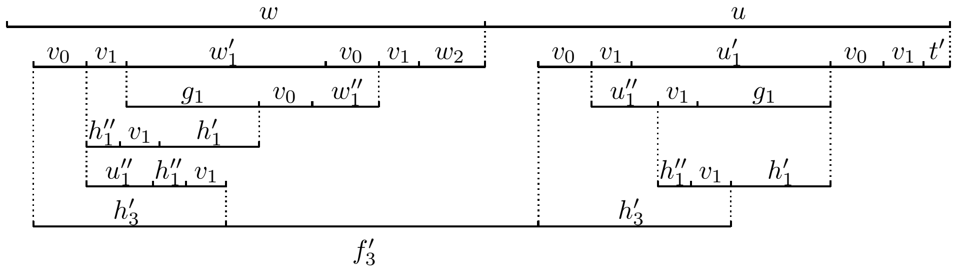
We have $h'_2 \leq_{\rm p} v_0 h''_0 w''_1 v_1$ since $v_0 v_1$ does not occur in $w_2$ and $az$ does not occur in $h'_2$ (and so $h'_2$ does not stretch beyond $w$). We have $v_0 h''_0 w''_1 v_1 \leq_{\rm p} h'_2$ since $v_0 v_1$ does not occur in $w'_1$. Hence, $h'_2 = v_0 h''_0 w''_1 v_1$ and

$$w'_1 v_0 v_1 = h'_0 v_0 h''_0 w''_1 v_1 = h'_0 h'_2 \qquad \text{and} \qquad h'_2 \leq_{\rm s} u'_1 v_0 v_1 \ .$$




We have $h'_3 = v_0 u''_1 h''_1 v_1$ by the arguments of the previous paragraph. Moreover,

$$v_0 v_1 u'_1 = v_0 u''_1 h''_1 v_1 h'_1 = h'_3 h'_1 \qquad \text{and} \qquad h'_3 \leq_{\rm p} v_0 v_1 w'_1 \ .$$




It is now straightforward to see that

$$w''_1 = u''_1 = \varepsilon$$

for otherwise $v_1$ and $v_0$ occur more than once in $v_1 h'_1$ and $h'_0 v_0$, respectively, contradicting (9). From (6) it follows

$$w'_1 = g_1 = u'_1 \ .$$

**Case:** Assume $1 < k \leq \min\{i, j\}$ and $w'_\ell = u'_\ell$, for all $1 \leq \ell < k$. Let us denote both $w'_\ell$ and $u'_\ell$ by $v'_\ell$, for all $1 \leq \ell < k$.

We show that $w'_k = u'_k$. Consider

$$f_4 = v_1 w'_k v_0 v_1 v'_{k-1} v_0 v_1 \cdots v'_1 v_0 v_1 w_2 u'_0 v_0 v_1 u'_j \cdots v_0 v_1 u'_k v_0 \ .$$

If $f_4$ is unbordered, then $|u| < |w| - 1$ since $|f_4| \leq |w|$ and

$$|u| = |f_4| - |v_1 w'_k v_0 v_1 v'_{k-1} v_0 v_1 \cdots v'_1 v_0 v_1 w_2| + |v_1 v'_{k-1} v_0 v_1 \cdots v'_1 v_0 v_1 t'|$$

and $|t'| \leq |z_0| \leq |z| < |bz| \leq |v_0 v_1|$ and $w_2 \neq \varepsilon$. Assume $f_4$ is bordered. Then $f_4$ has a shortest border $h_4$ such that $|v_0 v_1| \leq |h_4|$. Let $h_4 = v_1 g_4 v_0$.

**Subcase:** Let $|v_1 w'_k v_0| < |h_4|$. Then there exists an $\ell < k$ such that

$$h_4 = v_1 w'_k v_0 v_1 v'_{k-1} v_0 v_1 \cdots v'_{\ell+1} v_0 v_1 v''_\ell v_0$$

where $v''_\ell \leq_\mathrm{p} v'_\ell$. That implies $u'_k = v''_\ell$, since $v_0v_1$ does neither occur in $v''_\ell$ nor in $u'_k$. Now, consider

$$f_5 = v_1w'_kv_0v_1v'_{k-1}v_0v_1 \cdots v'_1v_0v_1w_2u'_0v_0v_1u'_j \cdots v_0v_1u'_kv_0v_1v'_{k-1}v_0v_1 \cdots v''_\ell v_0 \ .$$

If $f_5$ is unbordered, then $|u| < |w| - 1$ since $|f_4| < |f_5|$, see above. Assume, $f_5$ is bordered. Then $f_5$ has a shortest border $h_5$ such that $|h_4| < |h_5|$, for otherwise $h_4$ is not the shortest border of $f_4$, since either $h_4 \leq_\mathrm{p} h_5$ or $h_5 \leq_\mathrm{p} h_4$, and the latter implies that $h_4$ is bordered, and hence, not minimal. There exists an $\ell' < \ell$ such that

$$h_5 = v_1w'_kv_0v_1v'_{k-1}v_0v_1 \cdots v'_{\ell'+1}v_0v_1v''_{\ell'}v_0$$

where $v''_{\ell'} \leq_\mathrm{p} v'_{\ell'}$. We have $|f_4| < |f_5| < |f_6|$ where

$$f_6 = v_1w'_kv_0v_1v'_{k-1}v_0v_1 \cdots v'_1v_0v_1w_2u'_0v_0v_1u'_j \cdots v_0v_1u'_kv_0v_1v'_{k-1}v_0v_1 \cdots v''_{\ell'}v_0 \ ,$$

which is either unbordered and $|u| < |w| - 1$ since $|f_4| < |f_5|$, or it is bordered with a shortest border $h_6$, and we have $|h_4| < |h_5| < |h_6|$ and a factor $f_7$, such that $|f_4| < |f_5| < |f_6| < |f_7|$, and so on, until eventually an unbordered factor is reached proving that $|u| < |w| - 1$.

**Subcase:** Let $h_4 \leq_\mathrm{p} v_1w'_kv_0$. We also have that $h_4 \leq_\mathrm{s} v_1u'_kv_0$ since $v_0v_1$ does not occur in $w'_k$. Let $w'_kv_0 = g_4v_0w''_k$ and $v_1u'_k = u''_kv_1g_4$.

Consider,

$$f_8 = v_0w''_kv_1v'_{k-1}v_0v_1 \cdots v'_1v_0v_1w_2u'_0v_0v_1u'_jv_0v_1 \cdots u'_kv_0v_1 \ .$$

If $f_8$ is unbordered, then $|u| < |w| - 1$ since $|f_8| \leq |w|$ and

$$|u| = |f_8| - |v_0w''_kv_1v'_{k-1}v_0v_1 \cdots v'_1v_0v_1w_2| + |v'_{k-1}v_0v_1 \cdots v'_1v_0v_1t'|$$

and $|t'| \leq |z_0| \leq |z| < |bz| \leq |v_0v_1|$ and $w_2 \neq \varepsilon$. Assume $f_8$ is bordered. Then $f_8$ has a shortest border $h_8$ such that $v_0v_1 \leq_\mathrm{s} h_8$.

If $|h_8| > |v_0w''_kv_1|$ then the same argument as in the case $|v_1w'_kv_0| < |h_4|$ above shows that $|u| < |w| - 1$. If $|h_8| < |v_0w''_kv_1|$ then $v_0v_1$ occurs in $w'_k$; a contradiction. Hence, we have $h_8 = v_0w''_kv_1$ and

$$w'_kv_0v_1 = g_4h_8 \qquad \text{and} \qquad h_8 \leq_\mathrm{s} u'_kv_0v_1 \ . \tag{10}$$

Consider,

$$f_9 = v_0v_1w'_kv_0v_1v'_{k-1}v_0v_1 \cdots v'_1v_0v_1w_2u'_0v_0v_1u'_jv_0v_1 \cdots u'_{k+1}v_0u''_kv_1 \ .$$

If $f_9$ is unbordered, then $|u| < |w| - 1$ since $|f_9| \leq |w|$ and

$$|u| = |f_9| - |v_0v_1w'_kv_0v_1v'_{k-1}v_0v_1 \cdots v'_1v_0v_1w_2| + |g_4v_0v_1v'_{k-1}v_0v_1 \cdots v'_1v_0v_1t'|$$

and $|t'| \leq |z_0| \leq |z| < |bz| \leq |v_0v_1|$ and $|g_4| \leq |w'_k|$ and $w_2 \neq \varepsilon$. Assume $f_9$ is bordered. Then $f_9$ has a shortest border $h_9$ such that $v_0v_1 \leq_\mathrm{p} h_9$. We have $h_9 = v_0u''_kv_1$ by the arguments from the previous paragraph. Moreover,

$$v_0v_1u'_k = h_9g_1 \qquad \text{and} \qquad h_9 \leq_\mathrm{p} v_0v_1w'_k \ . \tag{11}$$

Observe, that (10) and (11) imply that the number of occurrences of $v_1$ and $v_0$, respectively, is the same in $w'_k$ and $u'_k$ since $v_0$ and $v_1$ do not intersect. Let

$$h_4 = v_1g_4v_0 = h''_1v_1h'_1v_0 = v_1h'_0v_0h''_0$$

where $v_1$ and $v_0$ occur only once in $v_1h'_1$ and $h'_0v_0$, respectively, by (9).

Let

$$f'_8 = v_0h''_0w''_kv_1v'_{k-1}\cdots v_0v_1v'_1v_0v_1w_2.u'_0v_0v_1u'_j\cdots v_0v_1u'_kv_0v_1$$

and

$$f'_9 = v_0v_1w'_kv_0v_1v'_{k-1}\cdots v_0v_1v'_1v_0v_1w_2.u'_0v_0v_1u'_j\cdots v_0v_1u'_{k+1}v_0u''_1h''_1v_1$$

with the respective shortest borders $h'_8$ and $h'_9$ (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

$$w'_kv_0v_1 = h'_0h'_8 \qquad \text{and} \qquad v_0v_1u'_k = h'_9h'_1 \ .$$

It is now straightforward to see that

$$h'_8 = h'_9 = v_0v_1$$

and

$$h_4 = v_0w'_kv_1 = v_0u'_kv_1$$

and hence, $w'_k = u'_k$.

This proves Claim 4.5. If $w'_k = u'_k$ we denote both $w'_k$ and $u'_k$ by $v'_k$. We set

$$\begin{aligned}\bar{v} &= v_0v_1w'_\iota\cdots v_0v_1w'_2v_0v_1w'_1\\ &= v_0v_1u'_\iota\cdots v_0v_1u'_2v_0v_1u'_1\end{aligned}$$

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

CLAIM 4.6. *If $i < j$ then $|u| < |w| - 1$.*

We have that

$$|w'_0| < |u'_0v_0v_1u'_j\cdots v_0v_1u'_{i+1}| \tag{12}$$

since $|w'_0| \leq |u'_0|$ by (5). Let

$$f_{11} = v_1w_2u'_0v_0v_1u'_j\cdots v_0v_1u'_{i+1}\bar{v}v_0 \ .$$

Note that $w = w'_0\bar{v}v_0v_1w_2$. Then $|w| < |f_{11}|$ by (12), and hence, $f_{11}$ is bordered. Let $h_{11} = v_1g_{11}v_0$ be the shortest border of $f_{11}$. Recall, that $w_2 \neq \varepsilon$ and either $az \leq_{\mathrm{s}} v_1w_2$ or $v_1w_2 \leq_{\mathrm{s}} az$. If $|v_1w_2| < |az|$ then $v_1$ necessarily occurs in $z$, and hence, it intersects with $v_0$ (since $bz \leq_{\mathrm{p}} v_0v_1$); a contradiction. We have $az \leq_{\mathrm{s}} v_1w_2$. Surely, $|h_{11}| < |v_1w_2|$ (and so $h_{11} \leq_{\mathrm{p}} v_1w_2$) for otherwise $az$ occurs in $u$ which contradicts our assumption that $z$ is of maximum length. Let $w_2 = g_{11}v_0w_5$. Note that $|v_0w_5| \neq |az|$ since $az$ and $v_0$ begin with different letters. We have $|az| < |v_0w_5|$ since otherwise $v_0$ occurs in $z$, and hence, intersects with $v_1$ which is a contradiction. Consider,

$$f_{12} = v_0w_5u'_0v_0v_1u'_j\cdots v_0v_1u'_{i+1}\bar{v}v_0v_1 \ .$$

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

$$|u| = |f_{12}| - |v_0w_5| + |t'|$$

and $|az| < |v_0w_5|$ and $|t'| \leq |z_0| \leq |z| < |bz| < |v_0w_5|$. Assume $f_{12}$ is bordered. Then $f_{12}$ has a shortest border $h_{12} = g_{12}v_0v_1$ (we have $|h_{12} \geq |v_0v_1|$ since $v_0$ and

$v_1$ cannot intersect) with $|az| < |h_{12}|$, because $h_{12} \leq_{\mathrm{p}} v_0w_5$ (otherwise $az$ occurs in $h_{12}$ and also in $u$; a contradiction) and $|az| < |v_0v_1|$. Let $v_0w_5 = g_{12}v_0v_1w_6$. But, now

$$w = w_0'\bar{v}v_0v_1g_{11}g_{12}v_0v_1w_6$$

where $v_0v_1w_6 \leq_{\mathrm{s}} w_2$, contradicting our assumption that $v_0v_1$ does not occur in $w_2$.

This proves Claim 4.6.

Claim 4.7. *If $i > j$ then $|u| < |w| - 1$.*

We have that

$$w = w_0'v_0v_1w_i' \cdots v_0v_1w_{j+1}'\bar{v}v_0v_1w_2 \qquad \text{and} \qquad u = u_0'\bar{v}v_0v_1t'$$

and $|w| \geq |u| - |t'| + |v_0v_1|$. We have $|u| < |w| - 1$ since $w_2 \neq \varepsilon$ and $|t'| \leq |z_0|$ and $|z_0| < |v_0v_1| - 1$.

This proves Claim 4.7.

Claim 4.8. *If $i = j$ then $|u| < |w| - 1$.*

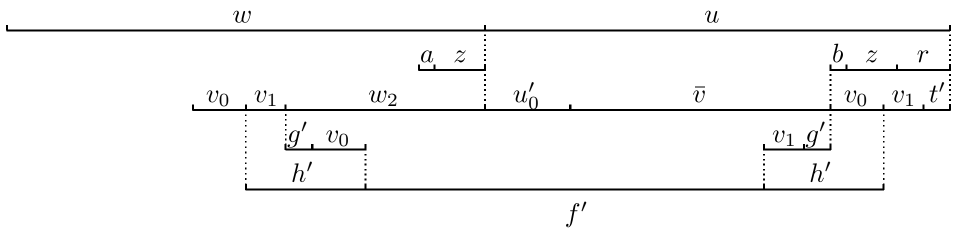
We have

$$w = w_0'\bar{v}v_0v_1w_2 \qquad \text{and} \qquad u = u_0'\bar{v}v_0v_1t' \ .$$

Consider

$$f' = v_1w_2u_0'\bar{v}v_0 \ .$$

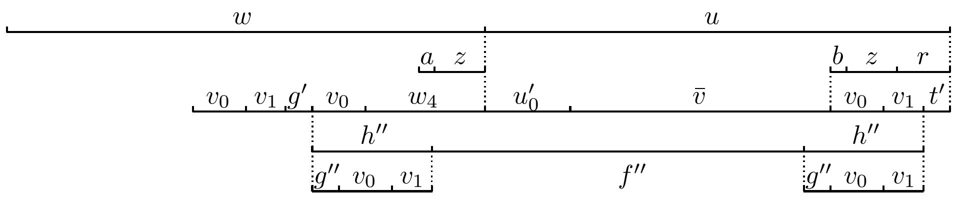
**Case:** Assume that $f'$ is bordered. Then $f'$ has a shortest border $h' = v_1g'v_0$.




Recall, that $w_2 \neq \varepsilon$ and either $az \leq_{\mathrm{s}} v_1w_2$ or $v_1w_2 \leq_{\mathrm{s}} az$. If $|v_1w_2| < |az|$ then $v_1$ occurs in $z$, and hence, intersects with $v_0$ since $bz \leq_{\mathrm{p}} v_0v_1$; a contradiction. We have $az \leq_{\mathrm{s}} v_1w_2$. Surely, $|h'| < |v_1w_2|$ for otherwise $az$ occurs in $u$ which contradicts our assumption. Let $w_2 = g'v_0w_4$. Note that $|v_0w_4| \neq |az|$ since $az$ and $v_0$ begin with different letters. We have $|az| < |v_0w_4|$ since otherwise $v_0$ occurs in $z$, and hence, intersects with $v_1$ which is a contradiction. Consider now,

$$f'' = v_0w_4u_0'\bar{v}v_0v_1 \ .$$

**Subcase:** Assume that $f''$ is unbordered. Then it easily follows that $|u| < |w| - 1$ since we have $|t'| < |az|$ and $|az| < |v_0w_4|$.



**Subcase:** Assume then that $f''$ is bordered. Then it has a shortest border $h'' = g''v_0v_1$ with $|az| < |h''|$, for otherwise $az$ occurs in $u$. Let $v_0w_4 = g''v_0v_1w_5$. But, now

$$w = w_0'\bar{v}v_0v_1g'g''v_0v_1w_5 = w_0'\bar{v}v_0v_1w_2$$

which contradicts our assumption that $v_0v_1$ does not occur in $w_2$.

**Case:** Assume that $f'$ is unbordered. Then $|f'| \leq |w|$, and hence, $|w_0'| \geq |u_0'|$. But, we also have $|w_0'| \leq |u_0'|$; see (5). That implies $|w_0'| = |u_0'|$. Moreover, the factors $w_0$ and $bzv'$ have both nonoverlapping occurrences in $u_0'v_0v_1$ by (5). Therefore, $w_0' = u_0'$. Let, $w = xaw_7$ and $u = xbt''$, where $w_0'\bar{v}v_0v_1 \leq_{\mathrm{p}} x$ and $a, b \in A$ and $a \neq b$ and $w_7 \leq_{\mathrm{s}} w_2$ and $t'' \leq_{\mathrm{s}} t'$. We have that $xb$ occurs in $w$ by Theorem 3.7. Since $xb$ is not a prefix of $w$ and $v_0v_1$ does not overlap with itself, we have $|xb| + |v_0v_1| \leq |w|$. From $|t'| \leq |z_0| < |v_0v_1|$ and $|t''| < |t'|$, we obtain $|u| < |w| - 1$.

This proves the Claims 4.8 and 4.2 and finishes the proof of Theorem 1.2. □

## 5. COROLLARY

Note that the bound $|u| < |w| - 1$ on the length of a nontrivial Duval extension $wu$ of $w$ is tight, as the following example shows.

*Example* 5.1. Let $w = a^nba^{n+m}bb$ and $u = a^{n+m}ba^n$ with $n, m \geq 1$. Then

$$w.u = a^nba^{n+m}bb.a^{n+m}ba^n$$

is a nontrivial Duval extension of $w$ and $|u| = |w| - 2$.

In general, Duval [1982] proved that we have $\partial(w) = \mu(w)$, for any word $w$, if $|w| \geq 4\mu(w) - 6$. Duval also noted that already $|w| \geq 3\mu(w)$ implies $\partial(w) = \mu(w)$, provided his conjecture holds. Corollary 1.3 follows from Theorem 1.2.

Corollary 1.3. *If* $|w| \geq 3\mu(w) - 3$ *then* $\partial(w) = \mu(w)$.

Proof. Assume $\partial(w) \neq \mu(w)$ and $|w| \geq 3\mu(w) - 3$. Let $w = xvy$ such that $v$ is the leftmost unbordered factor of $w$ of maximum length, that is, $|v| = \mu(w)$ and $\mu(xv^\bullet) < \mu(xv)$. Then $\widetilde{xv}$ and $vy$ are Duval extensions of $\widetilde{v}$ and $v$, respectively. We have by our assumption that $|x|$ or $|y|$ is larger than $\mu(w) - 2$.

If $|x| \geq \mu(w) - 1$ then $\widetilde{xv}$ is a trivial Duval extension by Theorem 1.2 and all conjugates of $v$ occur in $xv$. Since $v$ is primitive and only alphabets with at least two letters are considered there occurs an unbordered conjugate $u$ of $v$ in $xv^\bullet$ by Lemma 3.1 contradicting our assumption that $v$ is the leftmost unbordered factor of $w$ of maximum length.

If $|y| \geq \mu(w) - 1$ then $vy$ is a trivial Duval extension by Theorem 1.2 and all conjugates of $v$ occur in $vy$. Moreover, all conjugates of $v$ occur in the suffix $y'$ of $vy$ of length $2|v| - 1$. Let $u$ be an unbordered conjugate of $v$, with $u \neq v$ (which exists since we consider only words with at least two different letters), occurring in $y'$, that is $w = sut$ with $|t| \leq |v| - 2$. Consider the Duval extension $\widetilde{su}$. If $\widetilde{su}$ is trivial than $\partial(w) = \mu(w)$ contradicting our assumption. So, $\widetilde{su}$ is a nontrivial Duval extension, and hence, $|s| < |v| - 1$ by Theorem 1.2. Now, $|w| < 3|v| - 3$ which is again a contradiction. □

However, this bound is unlikely to be tight. The best example for a large bound that we could find is taken from [Assous and Pouzet 1979].

*Example* 5.2. Let

$$w = a^n b a^{n+1} b a^n b a^{n+2} b a^n b a^{n+1} b a^n \ .$$

We have $|w| = 7n + 10$ and $\mu(w) = 3n + 6$ and $\partial(w) = 4n + 7$.

Example 5.2 shows that the precise bound for the length of a word that implies $\partial(w) = \mu(w)$ is larger than $(7/3)\mu(w) - 4$ and not larger than $3\mu(w) - 3$ (by Corollary 1.3). The characterization of the precise bound of the length of a word as a function of its longest unbordered factor is still an open problem.

## 6. CONCLUSIONS

In this paper we have given an affirmative answer to a long standing conjecture [Duval 1982] by proving that a Duval extension $wu$ of $w$ longer than $2|w| - 2$ is trivial. This bound is tight 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 $|w| \geq 3\mu(w) - 3$ implies $\partial(w) = \mu(w)$ as conjectured (more weakly) in [Assous and Pouzet 1979]. However, the best known example of a word $w$ satisfying $\partial(w) > \mu(w)$ gives $|w| = (7/3)\mu(w) - 4$. We believe that the actual bound of $|w|$ is indeed close to $(7/3)\mu(w)$ rather than $3\mu(w)$. We pose the following conjecture.

CONJECTURE 6.1. *If* $|w| \geq \frac{7}{3}\mu(w) - 3$ *then* $\partial(w) = \mu(w)$.

Certainly, more information about the structure of nontrivial Duval extensions, like the one described in [Harju and Nowotka 2002b], would be useful for solving Conjecture 6.1.


### ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for the time and effort that they put into the review of this manuscript and their detailed comments and suggestions which greatly helped to improve this paper.


### REFERENCES


ASSOUS, R. AND POUZET, M. 1979. Une caractérisation des mots périodiques. *Discrete Math. 25,* 1, 1–5.

BERSTEL, J. AND PERRIN, D. 1985. *Theory of codes.* Pure and Applied Mathematics, vol. 117. Academic Press Inc., Orlando, FL.

BOYER, R. S. AND MOORE, J. S. 1977. A fast string searching algorithm. *Commun. ACM 20,* 10 (Oct.), 762–772.

BRESLAUER, D., JIANG, T., AND JIANG, Z. 1997. Rotations of periodic strings and short superstrings. *J. Algorithms 24,* 2, 340–353.

BYLANSKI, P. AND INGRAM, D. G. W. 1980. *Digital transmission systems.* IEE.

CÉSARI, Y. AND VINCENT, M. 1978. Une caractérisation des mots périodiques. *C. R. Acad. Sci. Paris Sér. A 286*, 1175–1177.

CROCHEMORE, M., MIGNOSI, F., RESTIVO, A., AND SALEMI, S. 1999. Text compression using antidictionaries. In *26th Internationale Colloquium on Automata, Languages and Programming (ICALP), Prague.* Lecture Notes in Comput. Sci., vol. 1644. Springer, Berlin, 261–270.

CROCHEMORE, M. AND PERRIN, D. 1991. Two-way string-matching. *J. ACM 38,* 3, 651–675.

DUVAL, J.-P. 1979. Périodes et répétitions des mots de monoïde libre. *Theoret. Comput. Sci. 9,* 1, 17–26.

DUVAL, J.-P. 1982. Relationship between the period of a finite word and the length of its unbordered segments. *Discrete Math. 40,* 1, 31–44.

EHRENFEUCHT, A. AND SILBERGER, D. M. 1979. Periodicity and unbordered segments of words. *Discrete Math. 26,* 2, 101–109.

HARJU, T. AND NOWOTKA, D. 2002a. Density of critical factorizations. *Theor. Inform. Appl. 36,* 3, 315–327.

HARJU, T. AND NOWOTKA, D. 2002b. Duval's conjecture and Lyndon words. technical report 479, Turku Centre of Computer Science (TUCS), Turku, Finland. (submitted).

HARJU, T. AND NOWOTKA, D. 2004. Minimal Duval extensions. *Internat. J. Found. Comput. Sci. 15,* 2, 349–354.

HOLUB, S. 2005. A proof of the extended Duval's conjecture. *Theoret. Comput. Sci. 339,* 1, 61–67.

KNUTH, D. E., MORRIS, J. H., AND PRATT, V. R. 1977. Fast pattern matching in strings. *SIAM J. Comput. 6,* 2, 323–350.

LOTHAIRE, M. 1983. *Combinatorics on Words*. Encyclopedia of Mathematics, vol. 17. Addison-Wesley, Reading, MA.

LOTHAIRE, M. 2002. *Algebraic Combinatorics on Words*. Encyclopedia of Mathematics and its Applications, vol. 90. Cambridge University Press, Cambridge, United Kingdom.

LYNDON, R. C. 1954. On Burnside's problem. *Trans. Amer. Math. Soc. 77*, 202–215.

MARGARITIS, D. AND SKIENA, S. 1995. Reconstructing strings from substrings in rounds. In *36th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, Milwaukee, WI, 613–620.

MIGNOSI, F. AND ZAMBONI, L. Q. 2002. A note on a conjecture of Duval and Sturmian words. *Theor. Inform. Appl. 36,* 1, 1–3.

MORSE, M. AND HEDLUND, G. A. 1940. Symbolic dynamics II: Sturmian trajectories. *Amer. J. Math. 61*, 1–42.

SCHÜTZENBERGER, M.-P. 1979. A property of finitely generated submonoids of free monoids. In *Algebraic theory of semigroups (Proc. Sixth Algebraic Conf., Szeged, 1976)*. Colloq. Math. Soc. János Bolyai, vol. 20. North-Holland, Amsterdam, 545–576.

ZIV, J. AND LEMPEL, A. 1977. A universal algorithm for sequential data compression. *IEEE Trans. Information Theory 23,* 3, 337–343.