On the Relation between Periodicity and Unbordered Factors of Finite Words

Štěpán Holub[1] and Dirk Nowotka[2]

[1] Department of Algebra, Charles University in Prague, Czech Republic
holub@karlin.mff.cuni.cz
[2] Institute for Formal Methods in Computer Science, University of Stuttgart, Germany
nowotka@fmi.uni-stuttgart.de

Abstract. Finite words and their overlap properties are considered in this paper. Let w be a finite word of length n with period p and where the maximal length of its unbordered factors equals k. A word is called unbordered if it possesses no nonempty proper prefix which is also a suffix of that word. Suppose $k < p$ in w. It is known that $n \leq 2k - 2$ if w has an unbordered prefix u of length k. We show that if $n = 2k - 2$, then u ends in ab^i, with two different letters a and b and $i \geq 1$, and b^i occurs exactly once in w. This answers a conjecture by Harju and the second author of this paper about a structural property of maximal Duval extensions. Moreover, we show here that $i < k/3$, which in turn leads us to the solution of a special case of a problem raised by Ehrenfeucht and Silberger in 1979.

1 Introduction

Overlaps are one of the central combinatorial properties of words. Despite the simplicity of this concept, its nature is very complex. For example, the problem raised by Ehrenfeucht and Silberger [5] in 1979 on the relation between the period of a word, measuring the self-overlap of a word, and the lengths of its unbordered factors, representing the absence of overlaps, has been solved only recently in [11]. The focus of this paper is on the investigation of such questions. In particular, we consider so called Duval extensions by solving a conjecture [6, 4] about the structure of maximum Duval extensions.

When repetitions in words are considered then two notions are central: the *period*, which gives the least amount by which a word has to be shifted in order to overlap with itself, and the shortest *border*, which denotes the least (nonempty) overlap of a word with itself. Both notions are related in several ways, for example, the length of the shortest border of a word w is not larger than the period of w; it follows that the period of an unbordered word is its length since the shortest border of an unbordered word is the word itself; moreover, a shortest border is always unbordered, since its border is also a border of the original word. Deeper dependencies between the period of a word and its unbordered factors have been investigated for decades; see also the references to related work below.

Let a word w be called a *Duval extension* of u if $w = uv$ such that u is unbordered and for every unbordered factor x of w the inequality $|x| \leq |u|$ holds, where $|\cdot|$ denotes the length of the word. Let $\pi(w)$ denote the smallest period of a word w. A Duval extension is called *nontrivial* if $|u| < \pi(w)$. It is known that $|v| \leq |u| - 2$ for any nontrivial Duval extension uv [8–10]. This bound is tight, that is, Duval extensions with $|v| = |u| - 2$ exist. Let those be called *maximal Duval extensions.* An example of maximal Duval extension yield the words

$$u = baababaaa, \qquad\qquad v = babaaba.$$

The following conjecture has been raised in [6]; see also [4].

Conjecture 1. Let uv be a maximal Duval extension of $u = u'ab^i$ where $i \geq 1$ and a and b are different letters. Then b^i occurs only once in uv.

This conjecture is answered positively by Theorem 4 in this paper. Moreover, we show that $i < |u|/3$ in Theorem 5 which leads us to the result that a word z with unbordered factors of length at most k and $\pi(z) > k$ that contains a maximal Duval extension uv with $|u| = k$ is of length at most $7k/3 - 2$. This gives an alternative proof for a special case of a conjecture in [5, 1], proved recently in [11].

Previous Work. In 1979 Ehrenfeucht and Silberger [5] raised the problem about the maximal length of a word w, w.r.t. the length k of its longest unbordered factor, such that k is shorter than the period $\pi(w)$ of w. They conjectured that $|w| \geq 2k$ implies $k = \pi(w)$ where $|w|$ denotes the length of w. That conjecture was falsified shortly thereafter by Assous and Pouzet [1] by the following example:

$$w = a^n ba^{n+1} ba^n ba^{n+2} ba^n ba^{n+1} ba^n$$

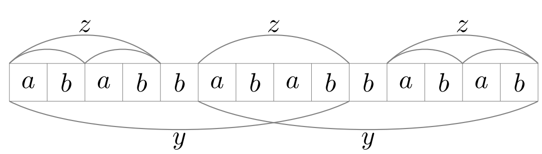
where $n \geq 1$ and $k = 3n + 6$ and $\pi(w) = 4n + 7$ and $|w| = 7n + 10$, that is, $k < \pi(w)$ and $|w| = 7k/3 - 4 > 2k$. Assous and Pouzet in turn conjectured that $3k$ is the bound on the length of w for establishing $k = \pi(w)$. Duval [3] did the next step towards solving the problem. He established that $|w| \geq 4k - 6$ implies $k = \pi(w)$ and conjectures that, if w possesses an unbordered prefix of length k, then $|w| \geq 2k$ implies $k = \pi(w)$. Note that a positive answer to Duval's conjecture yields the bound $3k$ for the general question. Despite some partial results [12, 4, 7] towards a solution, Duval's conjecture was only solved in 2004 [8, 9] with a new proof given in [10]. The proof of (the extended version of) Duval's conjecture lowered the bound for Ehrenfeucht and Silberger's problem to $3k-2$ as conjectured by Assous and Pouzet [1]. As already mentioned above, the optimal bound of $7k/3$ has been recently proved in [11].

2 Notation and Basic Facts

Let us fix a finite set A, called alphabet, of letters. Let A^* denote the monoid of all finite words over A including the *empty word* denoted by ε. In general, we denote variables over A by a, b, c, d and e and variables over A^* are ususally denoted by f, g, h, r through z, and α, β, and γ including their subscripted and primed versions. The letters i through q are to range over the set of nonnegative integers.

Let $w = a_1 a_2 \cdots a_n$. The word $a_n a_{n-1} \cdots a_1$ is called the *reversal* of w denoted by $\overline{w}$. We denote the length n of w by $|w|$, in particular $|\varepsilon| = 0$. If w is not empty, then let ${}^\bullet w = a_2 \cdots a_{n-1} a_n$ and $w^\bullet = a_1 a_2 \cdots a_{n-1}$. We define ${}^\bullet\varepsilon = \varepsilon^\bullet = \varepsilon$. Let $0 \leq i \leq n$. Then $u = a_1 a_2 \cdots a_i$ is called a *prefix* of w, denoted by $u \leq_{\mathrm{p}} w$, and $v = a_{i+1} a_{i+2} \cdots a_n$ is called a *suffix* of w, denoted by $v \leq_{\mathrm{s}} w$. A prefix or suffix is called *proper* when $0 < i < n$. Any word u such that $w = sur$ is called a *factor* of w. We shall also say, less formally, that u occurs in w, or that w contains u. An integer $1 \leq p \leq n$ is a *period* of w if $a_i = a_{i+p}$ for all $1 \leq i \leq n - p$. The smallest period of w is called *the period* of w, denoted by $\pi(w)$. A nonempty word u is called a *border* of a word w, if $w = uy = zu$ for some words y and z. We call w *bordered*, if it has a border that is shorter than w, otherwise w is called *unbordered*.

Example 1. Consider the word $u = ababbababbabab$. One of its borders is $y = ababbabab$, which is longer than $|u|/2$. This means that the prefix y and the suffix y overlap in u. The overlap $z = abab$ is therefore a border of y, and consequently also of u. Finally, ab is a border of z. It is already unbordered, whence it is the shortest border of u (as well as of y, of z, and of itself).



Using the previous example it is not difficult to note that every bordered word w has a minimal border u such that $w = uvu$, where u is unbordered.

Let $\lhd$ be a total order on A. Then $\lhd$ extends to a *lexicographic order*, also denoted by $\lhd$, on A^* with $u \lhd v$ if either $u \leq_{\mathrm{p}} v$ or $xa \leq_{\mathrm{p}} u$ and $xb \leq_{\mathrm{p}} v$ and $a \lhd b$. Let $\overline{\lhd}$ denote a lexicographic order on the reversals, that is, $u \mathrel{\overline{\lhd}} v$ if $\overline{u} \lhd \overline{v}$. Let $\lhd^a$ and $\lhd_b$ and $\lhd_b^a$ denote lexicographic orders where the maximal letter or the minimal letter or both are fixed in the respective orders on A. We establish the following convention for the rest of this paper: in the context of a given order $\lhd$ on A, we denote the inverse order of $\lhd$ by $\blacktriangleleft$. A $\lhd$-*m*aximal prefix (*suffix*) α of a word w is defined as a prefix (suffix) of w such that $v \mathrel{\overline{\lhd}} \alpha$ ($v \lhd \alpha$) for all $v \leq_{\mathrm{p}} w$ ($v \leq_{\mathrm{s}} w$). When it causes no confusion, especially when we consider an arbitrary order, we shall simply speak about maximal suffix or maximal prefix.

The notions of maximal pre- and suffix are symmetric. It is general practice that facts involving the maximal ends of words are mostly formulated for maximal suffixes. The analogue version involving maximal prefixes is tacitly assumed.

Remark 1. Any maximal suffix of a word w is longer than $|w| - \pi(w)$ and occurs only once in w.

Indeed, if $u = x\alpha y$, then $y = \varepsilon$ since otherwise $\alpha \lhd \alpha y$ yields a contradiction with the maximality of α. Therefore α occurs only once in w. If $|\alpha| \leq |w| - \pi(w)$, then we can write $w = uv\alpha$ with $|v| = \pi(w)$. Then α is a proper prefix of $v\alpha$ and $w = u\alpha v'$ gives a contradiction again.

Let an integer q with $0 \leq q < |w|$ be called *point* in w. A nonempty word x is called a *repetition word* at point q if $w = uv$ with $|u| = q$ and there exist words y and z such that $x \leq_{\mathrm{s}} yu$ and $x \leq_{\mathrm{p}} vz$. Let $\pi(w, q)$ denote the length of the shortest repetition word at point q in w. We call $\pi(w, q)$ the *local period* at point q in w. Note that the repetition word of length $\pi(w, q)$ at point q is necessarily unbordered and $\pi(w, q) \leq \pi(w)$. A factorization $w = uv$, with $u, v \neq \varepsilon$ and $|u| = q$, is called *critical*, if $\pi(w, q) = \pi(w)$, and if this holds, then q is called a *critical point*.

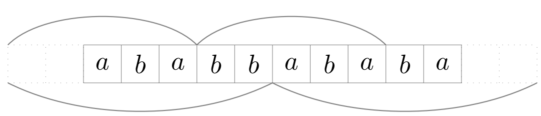
Let $\lhd$ be an order on A. Then the shorter of the $\lhd$-maximal suffix and the $\blacktriangleleft$-maximal suffix of some word w is called a *critical suffix* of w (with respect to $\lhd$). Similarly, we define a *critical prefix* of w by the shorter of the two maximal prefixes resulting from some order and its inverse. This notation is justified by the following formulation of the so called critical factorization theorem (CFT) [2] which relates maximal suffixes and critical points.

Theorem 1 (CFT). *Let $w \in A^*$ be a nonempty word and γ be a critical suffix of w. Then $|w| - |\gamma|$ is a critical point.*

Example 2. Consider the word $w = ababbababa$ We have $\pi(w) = 7$. The $\lhd^a$-maximal suffix of w is $\alpha = ababa$, and its $\lhd^b$-maximal suffix is $\beta = bbababa$. By CFT, the critical point is $|w| - |\alpha| = 5$. Indeed, we have the following list of local periods:

q	0	1	2	3	4	5	6	7	8	9
$\pi(w, q)$	1	2	2	5	1	7	2	2	2	2

For example, the repetition word at point 3 is $bbaba$, and at the critical point 5 it is $abababb$.



Let uv be a *Duval extension* of u if u is an unbordered word and every factor in uv longer than $|u|$ is bordered. A Duval extension uv of u is called trivial if $v \leq_{\mathrm{p}} u$. The following fact was conjectured in [3] and proven in [8–10].

Theorem 2. *Let uv be a nontrivial Duval extension of u. Then $|v| \leq |u| - 2$.*

Following Theorem 2 let a *maximal Duval extension* of u be a nontrivial Duval extension uv with $|v| = |u| - 2$. This length constraint on v will often tacitly be used in the rest of this paper.

Let wuv be an *Ehrenfeucht-Silberger extension* of u if both uv and $\overline{wu}$ are Duval extensions of u and $\overline{u}$, respectively. It is called trivial if both extensions are trivial, that is, if $\pi(wuv) = |u|$. The words uv and $\overline{wu}$ are called the Duval extensions corresponding to the Ehrenfeucht-Silberger extension of u.

Ehrenfeucht and Silberger were the first to investigate the bound on the length of a word w, w.r.t. the length k of its longest unbordered factors, such that $k < \pi(w)$. The asymptotically precise bound has been given recently in [11], where the following theorem is proven.

Theorem 3. *Let wuv be a nontrivial Ehrenfeucht-Silberger extension of u. Then $|wv| \leq 4/3\,|u| - 7/3$.*

3 Periods and Maximal Suffixes

Note the following simple but noteworthy fact.

Lemma 1. *Let u be an unbordered word, and let v be such that u does not occur in v. Let α be a maximal suffix of u. Then any prefix $w\alpha$ of uv is unbordered.*

Proof. Note that u occurs only once in uv since it is unbordered and not occurring in v. Suppose that $w\alpha$ has the shortest border $h \neq w$. Then $|h| < |u|$, otherwise $u \leq_{\mathrm{p}} h$, and u occurs twice in w. If $|h| < |\alpha|$, then $|h|$ is a border of u, a contradiction. If, on the other hand, $|h| \geq |\alpha|$, then α occurs twice in u, a contradiction with Remark 1. $\square$

This implies immediately the following version of Lemma 1 for Duval extensions which will be used frequently further below.

Lemma 2. *Let uv be a nontrivial Duval extension of u, and let α be the $\lhd$-maximal suffix of u. Then uv contains just one occurrence of α.*

The next lemma highlights an interesting fact about borders involving maximal suffixes. It will mostly be used on maximal prefixes of words, the dual to maximal suffixes, in later proofs. However, it is general practice to reason about ordered factors of words by formulating facts about suffixes rather than prefixes. Both ways are of course equivalent. We have chosen to follow general practice here despite its use on prefixes later in this paper.

Lemma 3. *Let αa be the $\lhd$-maximal suffix of a word wa where a is a letter. Let u be a word such that αa is a prefix of u and wb is a suffix of u, with $b \neq a$ and $b \lhd a$. Then u is either unbordered or its shortest border has length at least $|w| + 2$.*

Proof. Suppose that u has a shortest border hb. If $|h| < |\alpha|$ then $hb \leq_{\mathrm{p}} \alpha$ and $h \leq_{\mathrm{s}} \alpha$ and $hb \lhd ha$ contradict the maximality of αa. Note that $|h| \neq |\alpha|$ since $a \neq b$. If $|\alpha| < |h| \leq |w|$ then $\alpha a \leq_{\mathrm{p}} h$, and hence, αa occurs in w contradicting the maximality of αa again; see Remark 1. Hence, $|hb| \geq |w| + 2$. $\square$

The next lemma is taken from a result in [7] about so called minimal Duval extensions. However, the shorter argument given here (including the use of Lemma 3) gives a more concise proof than the one in [7].

Lemma 4. *Let uv be a nontrivial Duval extension of u where $u = xazb$ and $xc \leq_{\mathrm{p}} v$ and $a \neq c$. Then bxc occurs in u.*

Proof. Let ya be the $\lhd^a$-maximal suffix of xa. Consider the factor $yazbxc$ of uv which is longer than u and therefore bordered with a shortest border r. Lemma 3 implies that $|r| > |xc|$, and hence, $bxc \leq_{\mathrm{s}} r$ occurs in u. $\square$

4 Some Facts about Certain Suffixes of a Word

This section is devoted to the foundational proof technique used in the remainder of this paper. The main idea is highlighted in Lemma 5 which identifies a certain unbordered factor of a word.

Lemma 5. *Let α be the $\lhd$-maximal suffix and β be the $\blacktriangleleft$-maximal suffix of a word u, and let v be such that neither α nor β occur in uv more than once. Let a be the last letter of v and b be the first letter of x where $x \leq_{\mathrm{s}} \alpha v^{\bullet}$ and $|x| = \pi(\alpha v^{\bullet})$.*

If $\pi(\alpha v) > \pi(\alpha v^{\bullet})$, then αv is unbordered, in case $a \lhd b$, and βv is unbordered, in case $b \lhd a$.

Proof. Let $\alpha v^{\bullet} = \gamma x$. Since $|x| = \pi(\alpha v^{\bullet})$, we have $\gamma \leq_{\mathrm{s}} \gamma x$. From the assumption that α occurs in $\alpha v^{\bullet}$ just once now follows the inequality $|\gamma| < |\alpha|$. We have $\alpha = \gamma b \alpha'$ and $\alpha v = v' \gamma a$. Note that the inequality $\pi(\alpha v) > \pi(\alpha v^{\bullet})$ means $a \neq b$.

Suppose that $a \lhd b$. We claim that αv is unbordered in this case. Suppose the contrary, and let αv have a shortest border ha. Then $|h| < |\gamma|$ otherwise either $a = b$, if $|h| = |\gamma|$, or $|x|$ is not the smallest period of $\alpha v^{\bullet}$, if $|h| > |\gamma|$; a contradiction in both cases. But now $\alpha \lhd hb\alpha'$ since $ha \leq_{\mathrm{p}} \alpha$ and $a \lhd b$ contradicting the maximality of α because $hb\alpha' \leq_{\mathrm{s}} \alpha$.

Suppose that $b \lhd a$. In this case the word βv is unbordered. To see this suppose that βv has a shortest border ha. The assumption that uv contains just one occurrence of the maximal suffixes implies that ha is a proper prefix of β. If $|h| \geq |\gamma|$ then γa occurs in u contradicting the maximality of α since $\gamma b \leq_{\mathrm{p}} \alpha \lhd \gamma a$. But now $ha \leq_{\mathrm{p}} \beta \blacktriangleleft hb\alpha'$ (since $b \lhd a$) contradicting the maximality of β. $\square$

Proposition 1. *Let uv be a nontrivial Duval extension of u, and let α be a critical suffix w.r.t. an order $\lhd$. Then $|v| < \pi(\alpha v) \leq |u|$.*

Proof. If $|v| \geq \pi(\alpha v)$ then α occurs twice in αv contradicting Lemma 2. Suppose that $\pi(\alpha v) > |u|$, and let z be the shortest prefix of v such that already $\pi(\alpha z) > |u|$. Then $\pi(\alpha z) > \pi(\alpha z^{\bullet})$, and Lemma 5 implies that either αz or βz is unbordered, where β is the $\blacktriangleleft$-maximal suffix of u. This contradicts the assumption that uv is a Duval extension, since both the candidates are longer than u, which follows from $|\beta z| > |\alpha z| \geq \pi(\alpha z) > |u|$. $\square$

5 About Maximal Duval Extensions

In this section we consider the general results of the previous section for the special case of Duval extensions which leads to the main results, Theorem 4 and 5. Theorem 4 confirms a conjecture in [6].

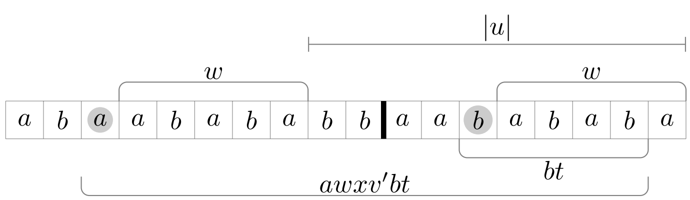
The following definition is justified by the intuition that unbordered factors are somehow connected to places where the period changes.

Definition 1. *Let uv be a Duval extension of u. The suffix s of uv is called a* trivial suffix *if $\pi(s) = |u|$ and s is of maximal length.*

Note that $s = uv$, if uv is a trivial Duval extension, and $as \leq_{\mathrm{s}} uv$ with $\pi(as) > |u|$, if uv is a nontrivial Duval extension. Moreover, Proposition 1 implies that $|s| \geq |\alpha v|$ where α is any critical suffix of u.

Let us first illustrate the main technique we shall use. Let uv be an extension of u. There is a standard way how to detect a factor suspicious of being long and unbordered. Let $u = u'awx$ and $v = v'bw$, where $|xv'bw| = |u|$. Observe that the factorizations are chosen in order to indicate the trivial suffix of uv, namely $wxv'bw$. Let bt be the $\overline{\lhd}$-maximal prefix of bw, where $a \lhd b$. The suspicious word is now $awxv'bt$. It can be bordered but its shortest border has to be relatively long. We give the following example.

Example 3. Let $u = abaabababb$ and $v = aabababa$. Then uv is a maximal Duval extension of u. Our technique detects the factor $aabababbaababab$, which is bordered and its shortest border is $aababab$.



We can begin with considerations about the periods of suffixes of maximal Duval extensions.

Lemma 6. *Let uv be a maximal Duval extension of u, and let $\lhd$ be an order such that the $\lhd$-maximal suffix α is critical. Then $\pi(\alpha v) = |u|$.*

Proof. It follows from Proposition 1 that $|u| - 1 \leq \pi(\alpha v) \leq |u|$ since $|v| = |u| - 2$. Suppose $\pi(\alpha v) = |u| - 1$. Let $w\alpha$ be the longest suffix of u such that $\pi(w\alpha v) = |u| - 1$. We have $w\alpha \neq u$ since u is unbordered. We can write $w\alpha v = w\alpha v' w\alpha^\bullet$, where v' is a prefix of v such that $|w\alpha v'| = |u| - 1$. The maximality of $w\alpha$ implies that $aw\alpha$ is a suffix of u, and $bw\alpha^\bullet$ is a suffix of αv, with $a \neq b$.

Choose a letter c in $w\alpha^\bullet$ such that $c \neq a$. Such a letter exists for otherwise $aw\alpha^\bullet \in a^+$ and α is just a letter, different from a. But this implies $u \in a^+\alpha$ and $v \notin a^+$ for uv to be nontrivial, that is, $v'd \leq_{\mathrm{p}} v$ with $d \neq a$; a contradiction since $uv'd$ is unbordered in this case.

Consider the $\overline{\lhd}^{c}$-maximal prefix of $bw\alpha^\bullet$ denoted by bt. Note that $|t| \geq 1$. We claim that $aw\alpha v't$ is unbordered. Suppose the contrary, and let r be the shortest border of $aw\alpha v't$. By Lemma 3 applied to the reversal of $aw\alpha v't$, the border r is longer than $bw\alpha^\bullet$. Hence, r contains α contradicting Lemma 2. Since $|w\alpha v'| = |u| - 1$ and $|t| \geq 1$, the unbordered factor $aw\alpha v't$ is longer than u; a contradiction. □

Lemma 7. *Let uv be a maximal Duval extension of u, let a be the last letter of u, and let xv be the trivial suffix of uv. Then $|\alpha| \leq |x|$ for the $\lhd^a$-maximal suffix α of any order $\lhd^a$.*

Proof. Suppose on the contrary that $|\alpha| > |x|$ which implies that the $\blacktriangleleft^a$-maximal suffix β is critical and $\beta \leq_{\mathrm{s}} x$ by Lemma 6. Since uv is nontrivial, we can write $u = u'cwba$ and $v = v'dw$ where $wba = x$, where a and b are letters, not necessarily distinct.

Consider the maximal prefix t of dw with respect to any order $\overline{\lhd}^{d}$. Note that $d \leq_{\mathrm{s}} t$. The word $cwbav't$ is longer than u, therefore it is bordered. Let r be its shortest border. By Lemma 3, we have $|cw| < |r|$. Lemma 2 implies that $r = cwb$, and we have $d = b$ since $d \leq_{\mathrm{s}} t$. We deduce that $|t| < |bw|$ since otherwise $t = bw = wb$ which implies $|u| = \pi(xv) = \pi(wbav'bw) = \pi(bwav'bw) \leq |v| + 1 < |u|$; a contradiction. Hence, $te \leq_{\mathrm{p}} bw$ for some letter $e \neq b$. Moreover, $e \neq a$ since $\beta^\bullet \leq_{\mathrm{s}} r$ and β does occur only once in βv by Lemma 2.

Consider the factor $\alpha v'te$ which is longer than u, and hence, bordered. Let s be the shortest border of $\alpha v'te$. The word s is a proper suffix of α by Lemma 2. Then also $|s| < |\beta|$ otherwise $\beta^\bullet e \leq_{\mathrm{s}} s$ contradicting the maximality of β since $\beta = \beta^\bullet a \blacktriangleleft^a \beta^\bullet e$. Let $s = \beta' e$ where $\beta' \leq_{\mathrm{s}} \beta^\bullet$. But then $\beta' e \leq_{\mathrm{p}} \alpha \lhd^a \beta' a$ and $\beta' a \leq_{\mathrm{s}} u$ contradicting the maximality of α. □

Lemma 8. *Let uv be a maximal Duval extension of $u = u'ab$ where a and b are letters. Then a occurs in u'.*

Proof. Suppose on the contrary that a does not occur in u'. The letter b occurs in u' by Lemma 4. We can therefore assume that $a \neq b$. Let c be the first letter of u. Clearly, c is different from both a and b.

Let γ be the critical suffix of u with respect ot some order $\lhd_c^b$. It is easy to see that $|\gamma| > 2$.

Lemma 6 implies $\pi(\gamma v) = |u|$. Let $w\gamma v$ be the trivial suffix of uv. We have that $u \neq w\gamma$ since uv is a nontrivial Duval extension of u. Therefore, we can write $u = u'dw\gamma$ and $v = v'ew\gamma^{\bullet\bullet}$ where d and e are different letters and $|w\gamma v'e| = |u|$. We deduce that e occurs in $u^{\bullet\bullet} = u'$ since otherwise $uv'e$ is unbordered. Consider an order $\lhd^e$ and let t be the $\overline{\lhd}^{e}$-maximal prefix of $ew\gamma^{\bullet\bullet}$.

The word $dw\gamma v't$ is longer than u, therefore it is bordered. Let r be its shortest border. By Lemma 3, we have $|dw\gamma| - 2 < |r|$. Lemma 2 implies that $|r|$ is exactly $|dw\gamma| - 1$, whence $r = dw\gamma^\bullet$. Clearly, the letter e is a suffix of t, and thus also of r, which implies that $e = a$. This is a contradiction since e occurs in u'. □

The following example shows that the requirement of a maximal Duval extension cannot be omitted in Lemma 8.

Example 4. Let a, b, and c be different letters, and consider $u = c^i bc^{i+j} bab$ and $v = c^{i+j} bc^{i-1}$ with $i, j \geq 1$. Then $u \cdot v = c^i bc^{i+j} bab \cdot c^{i+j} bc^i$ is a nontrivial Duval extension of length $2|u| - 4$ such that a occurs only in the second last position of u.

The next lemma highlights a relation between the trivial suffix of a maximal Duval extension uv and the set $\mathrm{alph}(u)$ of all letters occuring in u.

Lemma 9. *Let uv be a maximal Duval extension of u and wxw be the trivial suffix of uv where $|wx| = |u|$. Then either $\mathrm{alph}(w) = \mathrm{alph}(u)$ or there exists a letter b such that $\mathrm{alph}(w) = \mathrm{alph}(u) \setminus \{b\}$ and $u = u'bb$ and bb does not occur in u'.*

Proof. Suppose that $|\mathrm{alph}(w)| < |\mathrm{alph}(u)|$ and $b \in \mathrm{alph}(u) \setminus \mathrm{alph}(w)$.

Let $btwac \leq_{\mathrm{s}} u$ where $a, b, c \in \mathrm{alph}(u)$ and b does not occur in tw. Consider $btwxw$ which is longer than u and therefore has to be bordered. Let r be the shortest border of $btwxw$. Certainly, $|w| < |r|$ since $b \leq_{\mathrm{p}} r$ and $b \notin \mathrm{alph}(w)$. Moreover, $btw \leq_{\mathrm{p}} r$ implies $\pi(btwxw) \leq |u|$ contradicting the maximality of wxw. We conclude that $|w| < |r| < |btw|$.

Suppose $a \neq b$. Let $v = v'r$ and consider the factor $twacv'b$ which has to be bordered since $|twacv'b| = |twacv| - |r| + 1 > |acv| = |u|$. Let s be the shortest border of $twacv'b$. We have $|s| > |twa|$ because b is a suffix of s and does not occur in tw and $a \neq b$ by assumption. But now, $twac \leq_{\mathrm{p}} s$ contradicting Lemma 2 since wac contains a maximal suffix of u by Lemma 6.

Suppose next that $a = b$ and $c \neq b$. Consider an order $\lhd_c^b$ and let β be the $\lhd_c^b$-maximal suffix of u. We deduce that $|\beta| > |wbc|$ since w does not contain b and $bc \lhd_c^b btwac$. By Lemma 6, the $\blacktriangleleft_c^b$-maximal suffix α of u is a suffix of wbc. Moreover, $|\alpha| > 2$ since c occurs in $u^\bullet$ by Lemma 4. We have that $\alpha^{\bullet\bullet} \leq_{\mathrm{s}} w \leq_{\mathrm{s}} r$. From $|r| < |btw|$ and $r \leq_{\mathrm{p}} btw$ follows that $\alpha^{\bullet\bullet} d$ occurs in btw, where d is a letter in tw, and therefore $d \neq b$. But this contradicts the maximality of α since $b \blacktriangleleft_c^b d$ implies $\alpha = \alpha^{\bullet\bullet} bc \blacktriangleleft_c^b \alpha^{\bullet\bullet} d$.

Hence bb is a suffix of u. Suppose that bb occurs in u and consider an order $\lhd^b$. Certainly, the $\lhd^b$-maximal suffix of u is longer than wbb and therefore the $\blacktriangleleft^b$-maximal suffix α of u is critical. By Lemma 6, the word α is a suffix of wbb and $|\alpha| > 2$. As above, we deduce that $\alpha^{\bullet\bullet} d$ is a factor of btw, with $b \blacktriangleleft^b d$; a contradiction with the maximality of α. $\square$

The next two results, Lemma 10 and 11, constitute a case split of the proof of Theorem 4. Namely, the cases when exactly two or more than two letters occur in a maximal Duval extension.

Lemma 10. *Let uv be a maximal Duval extension of $u = u'ab^i$ where $i \geq 1$ and $|\mathrm{alph}(u)| > 2$ and $a \neq b$. Then u' does not contain the factor b^i.*

Proof. Suppose, contrary to the claim, that b^i occurs in u'. Consider the trivial suffix $wcbv'dw$ of uv where $|cbv'dw| = |u|$ and $c \in \{a, b\}$. Since $|u| > |wcb|$, we can write $u = u'ewcb$, where $d \neq e$. Lemma 9 yields $\mathrm{alph}(w) = \mathrm{alph}(u)$. Choose a letter f in dw such that $f \neq e$ and $f \neq c$. Let dt be the $\overline{\lhd}_e^f$-maximal prefix of dw

for some order $\lhd_e^f$. The word $ewcbv'dt$ is longer than u, therefore it is bordered. Let r be its shortest border. By Lemma 3, we have $|dw| < |r|$. Lemma 2 implies that $|r|$ is exactly $|dwc|$, and hence, $r = ewc$. Clearly, the letter f is a suffix of t, and thus also of r, which implies that $f = c$; a contradiction. $\square$

Lemma 11. *Let uv be a maximal Duval extension of $u = u'ab^i$ over a binary alphabet where $i \geq 1$ and $a \neq b$. Then u' does not contain the factor b^i and $awbb \leq_{\mathrm{s}} u$ and $v = v'bw$ where $wbbv$ is the trivial suffix of uv.*

Proof. Let s be the trivial suffix of uv, and let $u = u_0cwdb$ and $v = v'ew$ where $wdbv'ew = s$ and $c \neq e$. Let $\lhd$ be the order defined by $a \lhd b$.

Suppose $c = b$ and $e = a$. Let at be the $\overline{\blacktriangleleft}$-maximal prefix of aw. Consider the factor $bwdbv'at$ which is longer than $|u|$ and hence bordered. Let r be its shortest border. Lema 3 implies that $|bw| < |r|$. Lemma 2 implies that $r = bwd$, in fact, $r = bwa$ since $a \leq_{\mathrm{s}} t$. We deduce $|t| < |w|$, otherwise $t = w$ and $r = bwa = baw = ba^{|w|+1}$ contradicting Lemma 9. Therefore $atb \leq_{\mathrm{p}} aw$ by the maximality of at. But now $rb = bwab$ is a sufix of $v'atb$, and hence, the critical suffix of u occurs in v by Lemma 6 contradicting Lemma 2.

It remains that $c = a$ and $e = b$. Consider the $\lhd$-maximal suffix β of u. Suppose contrary to the claim that b^i occurs in u'. Then $b^j a \leq_{\mathrm{p}} \beta$ for some $j \geq i$.

Let bt be the $\overline{\lhd}$-maximal prefix of bw. Similarly to the reasoning above, we consider the factor $awdbv'bt$ and conlude that it has the border $r = awb$ and $d = b$ and $bta \leq_{\mathrm{p}} bw$. Lemma 7 implies that $\beta \leq_{\mathrm{s}} wbb$. Note that b^j is a power of b in u of maximal length and occurs in w by assumption, and hence, $b^j \leq_{\mathrm{s}} t$. But now, $b^j \leq_{\mathrm{s}} r$ and $b^{j+1} \leq_{\mathrm{s}} u$; a contradiction. $\square$

The main result follows directly from the previous two lemmas.

Theorem 4. *Let uv be a maximal Duval extension of $u = u'ab^i$ where $i \geq 1$ and $a \neq b$. Then b^i occurs only once in uv.*

Indeed, b^i does not occur in u' by Lemma 10 and 11. If b^i occurs in $b^{i-1}v$, that is, $b^{i-1}v = v''b^iv'$, then $u'abv''b^i$ is unbordered; a contradiction.

In the rest of the paper we investigate relation between our result and Ehrenfeucht-Silberger extensions.

Theorem 5. *Let uv be a maximal Duval extension of $u = u'ab^i$ where $i \geq 1$ and $a \neq b$. Then $3i \leq |u|$.*

Proof. It is an exercise to show that unbordered words of length at most 5 have no maximal Duval extensions. The shortest possible maximal Duval extension of a word u is of the form uv with $u = abaabb$ and $v = aaba$. This proves the claim for $i \leq 2$. Assume $i > 2$ in the following.

Let $cb^k \leq_{\mathrm{s}} v$ with $c \neq b$. First, note that $i - 2 \leq k \leq i - 1$; the first inequality follows from Lemma 2, the second from Lemma 6. Consider the shortest border h of uv. Then $|h| < |u| - 2$ since u is unbordered and uv is a nontrivial extension.

Let $h = gb^k$, and let j be the maximal integer such that $gb^j \leq_{\mathrm{p}} u$. Clearly, $i-2 \leq j \leq i-1$ since b^i occurs only as a suffix of u. Let $u = gb^j fb^i$. Note that

$$b \notin \{\mathrm{pref}_1(g), \mathrm{pref}_1(f), \mathrm{suff}_1(g), \mathrm{suff}_1(f)\} \ . \tag{1}$$

We show that b^k occurs in g or f. Suppose the contrary, that is, neither g nor f contains b^k. Consider the shortest border x of fb^iv. We have $|x| < |fb^i|$, since b^i does not occur in v. Property (1) and the assumption that b^k does not occur in f imply that $x = fb^k$. Let $v = v'fb^k$. Consider the shortest border y of $b^jfb^iv'f$. Again, we have $|y| < |b^jfb^i|$ since b^i does not occur in v, and property (1) implies that $y = b^jf$. Since $|b^jfb^k| \leq |gb^jfb^i| - 2$, we can write $v = v''b^jfb^k$. Finally, consider the shortest border z of $uv''b^j$. Property (1) and the assumption that b^k does not occur in g or f impliy that either $z = gb^j$ or $z = gb^jfb^j$. The former implies that $uv = gb^jfb^igb^jfb^k$ is a trivial Duval extension, and the latter implies that $|u| < |v|$; a contradiction in both cases.

We conclude that b^k occurs in g or f. Let $u = u_1b^mu_2b^nu_3b^i$ where u_1, u_2, and u_3 are not empty and neither begin nor end with b and $i-2 \leq k \leq m, n \leq i-1$. The claim is proven if $|u_1u_2u_3| > 3$ or $m = i-1$ or $n = i-1$. Suppose the contrary, that is, let $u = c_1b^{i-2}c_2b^{i-2}c_3b^i$, where c_1, c_2 and c_3 are letters. From $i-2 \leq k \leq j \leq m$ we deduce $k = i-2$.

Let us now consider the shape of v. Let $v'cb^\ell d$ be a prefix of v where $c \neq b$ and $d \neq b$. Since the factor $b^{i-2}c_2b^{i-2}c_3b^iv'cb^\ell d$ of uv has to be bordered, we conclude that $\ell \geq i-2$, in particular $\ell \neq 0$. Lemma 2 implies $\ell \leq i-1$. From $|v| = |u| - 2$ we now deduce that $v = d_1b^{i-2}d_2b^{i-2}d_3b^{i-2}$ for some letters d_1, d_2 and d_3 distinct from b. Since $c_1b^{i-2}c_2b^{i-2}c_3b^id_1$ has to be bordered, we obtain $c_1 = d_1$. Similarly, from $c_3b^id_1b^{i-2}d_2b^{i-2}d_3$ we deduce $c_3 = d_3$. Since uv is a nontrivial extension, we have $c_2 \neq d_2$. Considering possible borders of the word $c_1b^{i-2}c_2b^{i-2}c_3b^id_1b^{i-2}d_2$ we obtain $c_1 = d_2$, and similarly, from $c_2b^{i-2}c_3b^id_1b^{i-2}d_2^{i-2}d_3$ we deduce $c_2 = d_3$. The word $c_2b^{i-2}c_3b^id_1b^{i-2}d_2$ is now unbordered; the final contradiction. □

Corollary 1. *Let w be a nontrivial Ehrenfeucht-Silberger extension of u such that one of its corresponding Duval extensions is of maximal length. Then $|w| < 7/3\,|u| - 2$.*

Proof. Let $w = xuv$ and suppose, by symmetry, that uv is a maximal Duval extension of u. Suppose that $|w| \geq 7/3\,|u| - 2$, which is equivalent to $|x| \geq |u|/3 \geq i$, where $ab^i \leq_{\mathrm{s}} u$ with $a \neq b$.

If $eb^j \leq_{\mathrm{s}} x$ with $j < i$ and $e \neq b$, then eb^ju is unbordered; a contradiction. Therefore $b^i \leq_{\mathrm{s}} x$, and Theorem 4 implies that the word b^iub^{-i} is unbordered. Since w is an Ehrenfeucht-Silberger extension of u, the word b^iuv is a Duval extension of b^iub^{-i}. The extension must be trivial, since it is too long.

If $i = 1$, then $u = a^{|u|-1}b$, by Theorem 4, and it is an easy exercise to show that such a word has only trivial Ehrenfeucht-Silberger extensions.

Let $i \geq 2$. By Theorem 4, the word b^i is a critical suffix of u, which implies that $z = b^ivb^{-i+2}$ is unborderd. The word $\overline{z} \cdot \overline{xub^{-1}}$ is a Duval extension of $\overline{z}$, and it must be trivial, again because it is too long.

Therefore xuv has period $|u|$ and it is a trivial Ehrenfeucht-Silberger extension of u. □

We conclude with the following example, taken from [1].

Example 5. Consider the following word xuv where we separate the factors x, u, and v for better readability

$$x \cdot u \cdot v = b^{i-2} \cdot ab^{i-1}ab^{i-2}ab^{i} \cdot ab^{i-2}ab^{i-1}ab^{i-2}$$

where $i > 2$. The largest unbordered factors of xuv are of length $3i$, namely the factors $u = ab^{i-1}ab^{i-2}ab^{i}$ and $b^{i}ab^{i-2}ab^{i-1}a$, and $\pi(xuv) = 4i - 1$, and hence, xuv is a nontrivial Ehrenfeucht-Silberger extension of u. Note that uv is a maximal Duval extension. We have $|xuv| = 7i - 4 = 7/3\,|u| - 4$.

6 Conclusions

At the outset, the goal of our investigations was to find the solution of Conjecture 1. On the way of getting there we had to investigate a number of properties of unbordered factors which, we think, are of interest on their own. Although the Ehrenfeucht-Silberger problem has been solved in full generality in [11], the insights gained on the work on Conjecture 1 yield an alternative and straightforward proof for an interesting subcase.

Since all apparent conjectures around unbordered factors and their relation to periodicity have been solved, there are no clear questions left for now. However, that does not diminish the importance of and the interest in the investigation of the word structure of large Duval and Ehrenfeucht-Silberger extensions. Substantial progress in this field would certainly lead to new insights in periodicity questions of words in general.

Acknowledgements

The authors would like to thank the anonymous referees for their detailed comments which helped to further improve the presentation.

References

1. R. Assous and M. Pouzet. Une caractérisation des mots périodiques. *Discrete Math.*, 25(1):1–5, 1979.
2. M. Crochemore and D. Perrin. Two-way string-matching. *J. ACM*, 38(3):651–675, 1991.
3. 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.
4. J.-P. Duval, T. Harju, and D. Nowotka. Unbordered factors and Lyndon words. *Discrete Math.*, 308(11):2261–2264, 2008.
5. A. Ehrenfeucht and D. M. Silberger. Periodicity and unbordered segments of words. *Discrete Math.*, 26(2):101–109, 1979.
6. T. Harju and D. Nowotka. Duval's conjecture and Lyndon words. TUCS Tech. Rep. 479, Turku Centre of Computer Science, Finland, 2002.

7. T. Harju and D. Nowotka. Minimal Duval extensions. *Internat. J. Found. Comput. Sci.*, 15(2):349–354, 2004.
8. T. Harju and D. Nowotka. Periodicity and unbordered words. In *STACS 2004 (Montpellier)*, volume 2996 of *Lecture Notes in Comput. Sci.*, pages 294–304, Berlin, 2004. Springer-Verlag.
9. T. Harju and D. Nowotka. Periodicity and unbordered words: A proof of the extended Duval conjecture. *J. ACM*, 54(4), 2007.
10. Š. Holub. A proof of the extended Duval's conjecture. *Theoret. Comput. Sci.*, 339(1):61–67, 2005.
11. Š. Holub and D. Nowotka. The Ehrenfeucht-Silberger problem. In *ICALP '09: Proceedings of the 36th International Colloquium on Automata, Languages and Programming*, pages 537–548, Berlin, Heidelberg, 2009. Springer-Verlag.
12. F. Mignosi and L. Q. Zamboni. A note on a conjecture of Duval and Sturmian words. *Theor. Inform. Appl.*, 36(1):1–3, 2002.