Border Correlations of Partial Words

By: F. Blanchet-Sadri, E. Clader and O. Simpson.

F. Blanchet-Sadri, E. Clader and O. Simpson, (2010) "Border Correlations of Partial Words." *Theory of Computing Systems*, Vol. 47, No. 1, pp 179-195. DOI: 10.1007/s00224-008-9147-5

Made available courtesy of Springer: http://dx.doi.org/10.1007/s00224-008-9147-5

Abstract:
Partial words are finite sequences over a finite alphabet A that may contain a number of "do not know" symbols denoted by ◊'s. Setting $A_\diamond = A \cup \{\diamond\}$, $A_\diamond^*$ denotes the set of all partial words over A. In this paper, we investigate the border correlation function β: $A_\diamond^* \to \{a, b\}^*$ that specifies which conjugates (cyclic shifts) of a given partial word w of length n are bordered, that is, $\beta(w) = c_0c_1 \ldots c_{n-1}$ where $c_i = a$ or $c_i = b$ according to whether the ith cyclic shift $\sigma^i(w)$ of w is unbordered or bordered. A partial word w is bordered if a proper prefix x_1 of w is compatible with a proper suffix x_2 of w, in which case any partial word x containing both x_1 and x_2 is called a border of w. In addition to β, we investigate an extension β': $A_\diamond^* \to \mathbb{N}^*$ that maps a partial word w of length n to $m_0m_1 \ldots m_{n-1}$ where m_i is the length of a shortest border of $\sigma^i(w)$. Our results extend those of Harju and Nowotka.

Keywords: Combinatorics on words; Partial words; Border correlations

Article:

1 Introduction

The study of border correlations of words was begun by Harju and Nowotka, who defined the border correlation function and explored some of its properties in the case of binary full words [6]. A word is said to be bordered if one of its (nontrivial) prefixes is equal to one of its suffixes, and the border correlation function indicates which conjugates, or cyclic shifts, of a word are bordered. We extend their findings to include binary partial words, as well as full and partial words over larger alphabets.

If a partial word contains a factor of the form $c_1x_1c_2x_2c_3$, where c_1, c_2, and c_3 are pairwise-compatible letters and x_1 and x_2 are compatible partial words, then we say that it contains an overlap. In [6], Harju and Nowotka examine the relationship between the border correlation of a primitive full word and the existence of overlap-free conjugates. We discuss a set of exceptions to their result for nonprimitive full words. Furthermore, we show that there exist only finitely many partial words such that no conjugates are overlap-free, and these can all be classified as exceptional in the aforementioned sense.

An upper bound on the number of distinct border correlations of binary full words of a given length has been calculated, and we extend this bound to include binary partial words and words over larger alphabets. We also refine the border correlation so that it specifies the length of a shortest border of each conjugate as opposed to simply indicating the existence of a border, and we establish a connection between the palindromicity of the image of a word under this function and of the word itself. Finally, we examine critical factorizations of partial words as it relates to the borderedness of their conjugates.

2 Preliminaries

In this section, we recall the definitions of fundamental terms that will be used in this paper.

2.1 Words and Partial Words

We begin by presenting an overview of the basic concepts relating to words and partial words.

Let A be a finite, nonempty set of symbols, called an *alphabet*. Each element a of A is referred to as a *letter*, and a sequence of letters is called a *word* over A. Given a word w, the number of letters in the sequence represents the length of w, and is denoted by $|w|$. The empty word, denoted ε, is the unique word of length zero. We will indicate the number of occurrences of a letter $a \in A$ appearing in a word w by $|w|_a$. The set of distinct letters in w is denoted $\alpha(w)$.

For any alphabet A, the set of all words over A is denoted A^*. It is a monoid under the associative binary operation defined by concatenation of words, with ε serving as the identity element, and it is a free monoid over A. Similarly, the set of all nonempty words over A is denoted A^+. It is a semigroup under the operation of concatenation, and is a free semigroup over A. We denote by A^n the set of all words of length n over A.

The ith-power of a word w is defined recursively as

$$w^i = \begin{cases} \varepsilon & \text{if } i = 0 \\ ww^{i-1} & \text{if } i > 0 \end{cases}$$

A word of length n over the alphabet A is defined by a total function $w\colon \{0,\dots, n-1\} \to A$ and is usually represented as $w = a_0a_1 \dots a_{n-1}$ with $a_i \in A$.

A *partial word* of length n over A will be defined by a partial function $w\colon \{0,\dots, n-1\} \to A$. For each i such that $0 \le i < n$, we say that i is in the *domain* of w, denoted $D(w)$, if $w(i)$ is defined. Otherwise, we say that i is in the *hole set* of w, denoted $H(w)$. A word, or *full word*, is a partial word with an empty hole set. A *strictly partial word* is a partial word with a nonempty hole set.

Given a partial word w, the *companion* of w, denoted $w_\diamond$, is the total function $w_\diamond\colon\{0,\dots, n-1\} \to A \cup \{\diamond\}$ defined by

$$w_\diamond(i) = \begin{cases} w(i) & \text{if } i \in D(w) \\ \diamond & \text{otherwise} \end{cases}$$

where $\diamond$ is a new symbol which is not in the alphabet A, and which acts as a "do not know" symbol. The set of all partial words over A is denoted $A_\diamond^*$, where $A_\diamond = A \cup \{\diamond\}$. The set of all partial words of length n over A is denoted $A_\diamond^n$. Because the map $w \mapsto w_\diamond$ is bijective, we can extend our definitions of concatenation and powers to partial words intuitively.

A partial word u is a *factor* of a partial word w if there exist (possibly empty) partial words x, y such that $w = xuy$. We say that u is a *prefix* of w, denoted $u \le w$, if $x = \varepsilon$. Similarly, we say that u is a *suffix* of w if $y = \varepsilon$. A *factorization* of a partial word w is a sequence of partial words $w_0, w_1,\dots, w_i$ such that $w = w_0w_1\dots w_i$.

Partial words u and v are equal if $|u| = |v|$ and $D(u) = D(v)$, and $u(i) = v(i)$ for all $i \in D(u)$. If $|u| = |v|$, $D(u) \subset D(v)$, and $u(i) = v(i)$ for all $i \in D(u)$, then u is said to be *contained* in v, denoted $u \subset v$. We say u and v are *compatible*, denoted $u \uparrow\uparrow v$, if there exists a partial word w such that $u \subset w$ and $v \subset w$. We note that $u \uparrow\uparrow v$ implies $v \uparrow u$. Given partial words u and v such that $u \uparrow v$, the *least upper bound* of u and v is the partial word $u \vee v$, where $u \subset u \vee v$ and $v \subset u \vee v$, and $D(u \vee v) = D(u) \cup D(v)$.
A *strong period* of a partial word w is a positive integer p such that $w(i) = w(j)$ whenever $i, j \in D(w)$ and $i \equiv j \bmod p$. A *weak period* of w is a positive integer p such that $w(i) = w(p + i)$ whenever $i, p + i \in D(w)$. We denote the *minimal strong period* of w by $p(w)$, and the *minimal weak period* of w by $p'(w)$.

2.2 Border Correlations

We call a nonempty partial word w *unbordered* if no nonempty partial words x, u, v exist such that $w \subset xu$ and $w \subset vx$ (here x, u, v can be chosen to be full words). If such partial words exist, we say w is *bordered* and we call x a *border* of w. Notice that a border may be overlapping or nonoverlapping (see Figs. 1 and 2). If $w = x_1u = vx_2$ where $x_1 \subset x$ and $x_2 \subset x$, then we say that x is an *overlapping border* of w if $|x| > |u|$. Otherwise, x is called a *nonoverlapping border*. A border x of w is called *minimal* if $|x| > |y|$ implies that y is not a border of w (we also say that x is a shortest border of w).

Fig. 1 An overlapping border

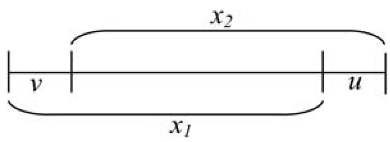
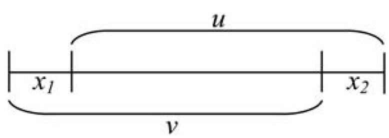


Fig. 2 A nonoverlapping border



Let $\sigma : A_\Diamond^* \to A_\Diamond^*$,where $\sigma(\varepsilon) = \varepsilon$ and $\sigma(cw) = wc$ for all $w \in A_\Diamond^*$ and $c \in A_\Diamond$, be the *shift function* of partial words. The partial words w and u are said to be *conjugates,* denoted $w \sim u$, if $w = \sigma k(u)$ for some $k \geq 0$. Two conjugates w and u are called *consecutive* if $\sigma(w) = u$ or $\sigma(u) = w$. It is easy to check that conjugacy is an equivalence relation, and we let $[w] = \{u \mid w \sim \sim u\}$ denote the *conjugacy class* of w. We define the *border correlation function* $\beta : A_\Diamond^* \to A_\Diamond^*$ as follows: Let $w \in A_\Diamond^*$be a partial word of length n, and let

$$c_i = \begin{cases} a \text{ if } \sigma^i(w) \text{ is unbordered} \\ \;b \;\text{ if } \sigma^i(w) \text{ is bordered} \end{cases}$$

Then $\beta(w) := c_0c_1 \ldots c_{n-1}$, and $\beta(\varepsilon) := \varepsilon$.

A partial word w is called *primitive* if there does not exist a partial word u such that $w \subset uk$ for $k \geq 2$ (here u can be chosen to be a full word). A word $w \in A^*$ that is primitive and minimal among its conjugates with respect to some lexicographic ordering of A is called a *Lyndon word*. It has been shown that Lyndon words are necessarily unbordered [8].

A partial word w is said to contain an *overlap* if it has a factor of the form $c_1x_1c_2x_2c_3$ where c_1, c_2, $c_3 \in A_\Diamond$ are pairwise compatible, and x_1, $x_2 \in A_\Diamond^*$ are compatible (here one must assume "pairwise," i.e. also c_1 and c_3 must be compatible; otherwise the case is trivial with $c_2 = \Diamond$). If no such factor exists, we say that w is *overlap-free*. Furthermore, w is *cyclically overlap-free* if none of its conjugates contains an overlap.

Given a partial word w, we say that w contains a *square* if it has a factor of the form u_1u_2, where u_1, $u_2 \in A_\Diamond^+$ and $u_1 \uparrow u_2$. A partial word containing no such factor is said to be *square-free*, and if none of its conjugates contains a square, it is said to be *cyclically square-free*.

We will work mostly with words over a binary alphabet, so unless otherwise specified, we will assume that $A = \{a, b\}$. Given a partial word w of length n over the binary alphabet A, the complement of w, denoted $\overline{w}$, is defined by $\overline{w} = \overline{w(0)\; w(1)} \;\ldots\; \overline{w(n-1)}$,where $\overline{a} = b$, $\overline{b} = a$, and $\overline{\Diamond} = \Diamond$.

3 Cyclical Overlap-Freeness

There exists a fundamental connection between cyclical overlap-freeness and the number of unbordered conjugates of a word. We begin this section by exploring this connection in the case of full words. It will later be shown that for partial words the situation is complicated very little, because there exist in fact only finitely-many cyclically overlap-free words that contain holes.

The results of this section are based in large part upon a reformulation of a theorem of Harju and Nowotka [6, Theorem 5]. It states that, given a primitive word w of length greater than three, every other conjugate of w is unbordered if and only if w is cyclically overlap-free. The proof of Theorem 5 of [6] assumes that w has at least one unbordered conjugate (as stated in the proof) which is equivalent to assuming that w is primitive (by considering Lyndon words). We proceed by showing the remaining case for nonprimitive words.

Example 3.1 Consider the word $w = abaaba$. It can be easily verified that all the conjugates of w are overlap-free. However, because w is nonprimitive, we have $\beta(w) = bbbbbb$, and hence it is not the case that every other conjugate of w is unbordered.

The above example suggests a way in which one might characterize the nonprimitive case.

Definition 3.1 A word $w \in A^*$ is an *exceptional cyclically overlap-free word* if there exists a primitive cyclically overlap-free word v such that $w = vv$.

Such words will also be cyclically overlap-free. Moreover, if w is nonprimitive, then all conjugates of w are also nonprimitive, so $\beta(w) = b^{|w|}$. Thus, we note that all exceptional cyclically overlap-free words will violate the stated result. Another basic observation is that if $w = vv$ is an exceptional cyclically overlap-free word, then v must be nonexceptional, since if $v = uu$ for any word u, then $w = uuuu$ contains an overlap.

In order to verify that we have entirely characterized the nonprimitive case, we re-examine the lemmas that Harju and Nowotka utilized in their proof, all of which are due to Thue [9]. All three lemmas use the Thue-Morse morphism $\tau : A^* \to\to A^*$, defined by $\tau(a) = ab$ and $\tau(b) = ba$.

Lemma 3.1 *Let $w \in A^*$. Then w has an overlap if and only if $\tau(w)$ has an overlap.*

Proof This is taken care by Thue [9] (see also Berstel [1] or Lothaire [7]).

Lemma 3.2 *Let $w \in A^*$ be a cyclically overlap-free word. Then $|w| = 2^j$ or $|w| = 3 \times 2^j$ for some $j \geq 0$.*

Proof Since the case of nonexceptional cyclically overlap-free words has been proven by Thue [9], we will only consider here the exceptional case. Let us consider w to be an exceptional cyclically overlap-free word. Thus, there exists a nonexceptional cyclically overlap-free word v such that $w = vv$, and hence, $|w| = 2 \times |v|$. The statement for the exceptional case thus follows from the nonexceptional case.

Lemma 3.3 *Let $w \in A^*$ and $|w| \geq 4$. If w is cyclically overlap-free and w is not a conjugate of either abbabb or aabaab, then either w or σ (w) has a factorization in terms of ab and ba.*

Proof This lemma follows from Thue's result on primitive words. Indeed, if vn with $n \geq 2$ is cyclically overlap-free then $n = 2$ and so is the primitive word v. Assume a is a prefix of v. Thue's argument says that $v = \tau^k(ab)$, or $v = \tau^k(aab)$, or $v = \tau^k(abb)$ for some k. Therefore if v is not in $\{ab, ba\}^+$ then $k = 0$, and this leaves $v = aab$ and $v = abb$.

Harju and Nowotka's result can be reformulated as stated in the following theorem (the proof is practically the same as the proof of the original result and is therefore omitted).

Theorem 3.1 *Let* $w \in A^*$ *and* $|w| \geq 4$. *Then w is a nonexceptional cyclically overlap-free word if and only if* $\beta(w) = (ab)^k$ *or* $\beta(w) = (ba)^k$, *where* $k = \frac{|w|}{2}$.

Theorem 3.1 and the preceding lemmas consider only the case of full words. However, the next proposition shows that there are very few strictly partial words that are cyclically overlap-free.

Proposition 3.1 *The only cyclically overlap-free partial words of length at least three containing holes are the elements of [aba◊ba], [bba◊ba], [◊ba], and [◊ba◊ba], and their complements.*

Proof The notion of overlap-freeness is only well-defined for partial words of length at least 3, so we will not consider partial words of shorter length. We first note that if a hole is preceded and succeeded by the same letter, e.g. *a◊a*, then there is immediately an overlap. So, we will only consider partial words containing as a factor *a◊b* (considering partial words containing *b◊a* will simply yield the complements of these). Furthermore, we note that *ba◊ba* must be a factor of the word (if it is long enough). We will now build possible cyclically overlap-free partial words by adding letters to the left of *ba◊ba* and eliminating those that force an overlap.

We consider the three possible characters that could appear immediately to the left of *ba◊ba*. In fact, it is easily verified that none create an overlap, so we consider the possible characters that could appear immediately to the left of either *aba◊ba*, *◊ba◊ba*, or *bba◊ba*. In all three cases, any of the three possible characters that could be appended creates an overlap in the resulting partial word. Figure 3 summarizes this process (dotted connections indicate possibilities that create an overlap).

Thus, we have verified that the only overlap-free strictly partial words are the following: *aba◊*, *aba◊b*, *aba◊ba*, *◊ba◊*, *◊ba◊b*, *◊ba◊ba*, *bba◊*, *bba◊b*, *bba◊ba*, *ba◊*, *ba◊b*, *ba◊ba*, *a◊b*, *a◊ba*, and *◊ba*, as well as the complements of these (note that any shorter-length partial words are too short to be considered).

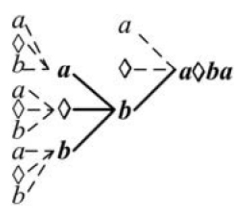


Fig. 3 Overlap-free strictly partial words

However, not all of the above partial words are cyclically overlap-free. In fact, one can check that the partial words *aba◊*, *aba◊b*, *◊ba◊*, *◊ba◊b*, *bba◊*, *bba◊b*, *ba◊ba*, and *a◊ba* all have conjugates that contain an overlap. The other six possibilities (namely *aba◊ba*, *◊ba◊ba*, *bba◊ba*, *ba◊*, *a◊b*, and *◊ba*) can all easily be verified to be cyclically overlap-free. Hence, the elements of the conjugacy classes [*aba◊ba*], [*bba◊ba*], [*◊ba*], and [*◊ba◊ba*] (and their complements) are the only cyclically overlap-free strictly partial words.

For all of the above cyclically overlap-free strictly partial words *w*, we have $\beta(w) = bn$, where $n = |w|$. Furthermore, we notice that if $|w| \geq 4$, then there exists an exceptional cyclically overlap-free full word *v* such that $w \uparrow v$. Definition 3.2 adapts the definition of exceptional cyclically overlap-free words in a way that allows us to generalize Theorem 3.1 to partial words.

Definition 3.2 A partial word $w \in A_\Diamond^*$ is an *exceptional cyclically overlap-free partial word* if *w* is cyclically overlap-free and there exists an exceptional cyclically overlap-free full word v such that $w \uparrow \uparrow v$.

Corollary 3.1 *Let* $w \in A_\Diamond^*$ *and* $|w| \geq 4$. *Then w is a nonexceptional cyclically overlap-free partial word if and only if* $\beta(w) = (ab)^k$ *or* $\beta(w) = (ba)^k$, *where* $k = \frac{|w|}{2}$.

4 Properties of β-Images

Harju and Nowotka calculated an upper bound for the number of distinct β-images of full binary words of a given length n. Using Theorem 3.1 and several observations based on Harju and Nowotka's work, we can extend this to characterize the set of β-images for partial words over binary and larger alphabets. This section is devoted to exploring this characterization.

4.1 Partial Words over a Binary Alphabet

The following lemma provides some properties of β β-images of partial words, expanding a result on full words in [6].

Lemma 4.1 *Let* $w \in A_\Diamond^*$ *of length* $n \geq 4$.

1. *If w is primitive, then* $|\beta(w)|_a \geq 1$.
2. *For each* $i \in \{0, 1, \ldots, n-1\}$, $\sigma^i(w)$ *or* $\sigma^{i+1}(w)$ *is bordered, or* $w \in [ab^{n-1}]$ *or* $w \in [ba^{n-1}]$.

Proof For claim 1, if w is primitive with $H(w) = \emptyset$, it is shown in [6] that $|\beta(w)|a \geq 2$ since, with respect to a lexicographic ordering of A and its reverse ordering, w has two Lyndon conjugates that are necessarily unbordered. If $H(w) \neq \emptyset$, however, we have words of the form $w = a^{n-4}\Diamond bab$, with $n \geq 6$, yielding $\beta(w) = ab^{n-1}$. Claim 2 follows directly from the claim in [6] for full words.

Harju and Nowotka have calculated an upper bound on the number of distinct β-images of full words of length n by characterizing the β-images that occur. The set $\beta(A^n)$ may contain β-images of the form b^n, in the case where every conjugate is bordered; those with at least two nonconsecutive a's; and the special β-image $\beta(w) = aab^{n-2}$ for $w \in [ab^{n-1}]$ or $w \in [ba^{n-1}]$. However, by Lemma 3.2, it cannot contain β-images of the form $(ab)^{\frac{n}{2}}$ or $(ba)^{\frac{n}{2}}$unless $n = 2^j$ or $3 \times \times 2^j$ for some j. For words of lengths $1 \leq \leq n \leq \leq 30$, β (A β (An) was found to contain all such β-images, and the bound

$$\|\beta(A^n)\| \leq F_n + F_{n-2} - m$$

where Fn is the nth Fibonacci number with $F_0 = 1$ and $F_1 = 1$ and $F_n = F_n{-}1 + F_{n-2}$ for all $n > 1$, and where m = 0 or 2 (specified in Theorem 4.1, below), is tight except for $n = 12$. We have extended this bound for partial words such that $\beta(A_\Diamond^n)$ also contains β-images with exactly one a, adding exactly one conjugacy class of β-images.

Theorem 4.1 *For all* $n > 2$

$$\beta(A_\Diamond^n) \subseteq \{b^n\} \cup \{w \| w|_a \geq 1, a^2 \textit{ not in } ww\} \cup [aab^{n-2}] \setminus \{(ab)^{\frac{n}{2}}, (ba)^{\frac{n}{2}} | n \neq 2^j \textit{or } 3 \times 2^j\} \quad (1)$$

$$\|\beta(A_\Diamond^n)\| \leq F_n + F_{n-2} - m + n \quad (2)$$

where $m = 2$ *if* n *is even and* $n \notin \{2^j, 3 \times 2^j \mid j \geq 0\}$, *and* $m = 0$ *otherwise.*

Proof (1) is extended from the equation for full words presented in [6]. Using this with claim 1 of Lemma 4.1, we may include β-images with exactly one unbordered conjugate.

For (2), we again extend the bound for full words in [6] by n, the size of the conjugacy class of β-images where $|\beta(w)|_a = 1$ for words of length n.

Table 1 illustrates how $\|\beta(A_\Diamond^n)\|$ and the bounds are related. For partial words of length n, where $2 \leq n \leq 5$, $\|\beta(A_\Diamond^n)\| = \|\beta(A^n)\|$, so we observe that the original bound in [6] holds for partial words of a small enough length, since there does not exist any partial word w with n < 6 such that $\beta(w) \in [ab^{n-1}]$. However, the extended bound for partial words is again tight for words of length $5 < n < 19$, excluding length 12. In the case where $n = 12$, there does not exist a word w such that $\beta(w) \in [abababbababb]$, so the bound exceeds the actual size of $\beta(A_\Diamond^{12})$ by the size of this conjugacy class. The reason for this is unclear.

Table 1 The number of distinct β-images for lengths $1 < n < 19$

Length (n)	$\|\beta(A^n)\|$	$F_n + F_{n-2} - m$	$\|\beta(A_\Diamond^n)\|$	$F_n + F_{n-2} - m + n$
2	2	–	2	–
3	4	4	4	7
4	7	7	7	11
5	11	11	11	16
6	18	18	24	24
7	29	29	36	36
8	47	47	55	55
9	76	76	85	85
10	121	121	131	131
11	199	199	210	210
12*	310	322	322	334
13	521	521	534	534
14	841	841	855	855
15	1364	1364	1379	1379
16	2207	2207	2223	2223
17	3571	3571	3588	3588
18	5776	5776	5794	5794

Harju and Nowotka prove that full words may have at most $\lfloor\frac{|w|}{2}\rfloor$ unbordered conjugates. For partial words with $\|H(w)\| \geq |1$ of even length greater than 5, it is impossible to reach this maximum.

Remark 4.1 Let w $\in A_\Diamond^*$ with $|w| = 2k$ for $k \geq 3$ and $\|H(w)\| \geq 1$. Then w can have at most $k - 1$ unbordered conjugates.

Indeed, consider the conjugate $\sigma^i(w)$ such that $a_0 = \Diamond$ where $\sigma^i(w) = a_0a_1 \ldots a_{n-1}$. Here, both $\sigma^i(w)$ and σ^{i+1} (w) are bordered, and by statement 2 of Lemma 4.1 we know that no consecutive conjugates can be unbordered, so w can have at most $k - 1$ unbordered conjugates. This remark follows from Corollary 3.1, since no partial words with at least one hole are nonexceptionally cyclically overlap-free.

However, if w is of odd length, the maximum number of unbordered conjugates of w is not $\lfloor\frac{|w|}{2}\rfloor - 1$, as one might expect. For example, given the partial word w = $\Diamond abbaababbabaabba$, $\beta(w) = b(ba)^{\lfloor\frac{|w|}{2}\rfloor}$, so w has $\lfloor\frac{|w|}{2}\rfloor$ unbordered conjugates. This is the shortest length for which a β-image of this form exists for strictly partial words.

4.2 Partial Words over Arbitrary Alphabets

We begin this section with a study of full words.

Let A be an alphabet of arbitrary size. We can use the notion of Lyndon conjugates to find a lower bound for the number of unbordered conjugates a word may have.

Remark 4.2 If w is a primitive full word such that $\|\alpha(w)\| \geq n$, then $|\beta(w)|_a \geq \geq n$.

With this remark we can show that no full word may contain exactly one unbordered conjugate, generalizing Harju and Nowotka's result for words over a binary alphabet.

Remark 4.3 Let w be a primitive full word of length greater than one over an arbitrary alphabet. Then $|\beta(w)|_a > 1$.

As with a binary alphabet, by allowing holes we observe that one conjugacy class of β-images, those with exactly one a, is added to the set $\beta(A^n)$ for an alphabet of arbitrary size. Furthermore, by extending our binary alphabet by just one letter, we find that our β-population grows significantly. Words over larger alphabets may have many consecutive unbordered conjugates. Remarkably, we observe that some words are cyclically unbordered. To avoid borders in an entire conjugacy class, it is necessary to construct a word that is cyclically square-free. In the following lemma and theorem, we describe the lengths and alphabets for which these words are possible.

Lemma 4.2 *Let w be a partial word of length n over an alphabet A. Then* $\beta(w) = a^n$ *(and hence w is a full word) if and only if w is cyclically square-free.*

Proof To prove the forward implication, we assume w is not cyclically square-free. Then there exists a conjugate $\sigma^i(w) = u_1u_2v$ for some partial words u_1, u_2, v such that $u_1 \uparrow u_2$. If $j = i + |u_1|$, then σ $\sigma^j(w) = u_2\, vu_1$. Since $\sigma^j(w)$ is bordered, $|\beta(w)|_b > 0$.

We prove the reverse implication by contradiction. Assume w is a cyclically square-free partial word such that $\beta(w) \neq a^n$. Hence, one of its conjugates, say $\sigma^i(w)$, is bordered. Since a hole creates a square in a word, we also assume w has an empty hole set, so the minimal border of $\sigma^i(w)$ must be nonoverlapping. Let $\sigma^i(w) = uvu$ for words u, v where v is possibly empty. Let $j = i + |uv|$. Then $\sigma^j(w) = uuv$, which contains a square and therefore leads to a contradiction.

Theorem 4.2 *Let A be an alphabet.*

- If ||A|| = 3, then for all n $\notin$ {5, 7, 9,10,14,17}, $a^n \in \beta(A^n)$.
- If ||A|| > 3, then for all n ≥ 1, $a^n \in \beta(A^n)$.

Proof From [5], we know that over a ternary alphabet, cyclically square-free words exist for all lengths except 5, 7, 9, 10, 14 and 17. For all other lengths, there exist cyclically square-free ternary words. By Lemma 4.2, the β-image of any such word is a^n.

For larger alphabets, if $n \notin \{5, 7, 9, 10, 14, 17\}$, the statement follows from above. If $n \in \{5, 7, 9, 14, 17\}$, there exists a ternary word w of length $n - 1$ that is cyclically square-free. For any letter d that does not appear in w, wd is a cyclically square-free word of length n over a quaternary alphabet. If $n = 10$, we can explicitly show the existence of cyclically square-free words over a quaternary alphabet (for example, $w = abacabcacd$). Any alphabets of larger size contain all words over a quaternary alphabet, hence the statement hold for larger alphabets as well.

Since β-images are full words over a binary alphabet, a clear upper bound on the number of distinct β-images of words of length n is 2^n. However, by Theorem 4.2, we can tighten this bound.

Proposition 4.1 *Let A be an alphabet of size greater than 2. Then*

$$\|\beta(A^n)\| \leq 2^n - n - m \qquad (3)$$

$$\|\beta(A^n_\diamond)\| \leq 2^n - m \qquad (4)$$

where $m = 1$ *if* $||A|| = 3$ *and* $n \in \{5, 7, 9, 10, 14, 17\}$*, and* $m = 0$ *otherwise.*

Proof By Remark 4.3, we know that full words cannot contain exactly one unbordered conjugate, so we subtract the size of the conjugacy class of β-images with exactly one a from the bound for full words. Furthermore, by Theorem 4.2, neither full words nor partial words of length n $\in$ {5, 7, 9, 10, 14, 17} over a ternary alphabet may have β-images of the form a^n, so we subtract 1 from both bounds for these cases.

Table 2 below gives $||\beta(A_\lozenge^n$n~$)||$ for words over alphabets of different sizes and lengths $1 < n < 13$. We observe that the bound given in Proposition 4.1 is tight for lengths $5 < n \leq 12$ and conjecture that it is tight for all greater lengths.

Table 2 The number of distinct β-images for lengths $1 < n < 13$ over alphabets of varying size

n	Ternary				Quaternary			
	Full		Partial		Full		Partial	
2	2	2^2-2	2	2^2-2	2	2^2-2	2	2^2-2
3	5	2^3-3	5	2^3-3	5	2^3-3	5	2^3-3
4	12	2^4-4	12	2^4-4	12	2^4-4	12	2^4-4
5	26	2^5-5-1	26	2^5-5-1	27	2^5-5	27	2^5-5
6	58	2^6-6	64	2^6	58	2^6-6	64	2^6
7	120	2^7-7-1	127	2^7-1	121	2^7-7	128	2^7
8	248	2^8-8	256	2^8	248	2^8-8	256	2^8
9	502	2^9-9-1	511	2^9-1	503	2^9-9	512	2^9
10	1013	$2^{10}-10-1$	1023	$2^{10}-1$	1014	$2^{10}-10$	1024	2^{10}
11	2037	$2^{11}-11$	2048	2^{11}	2037	$2^{11}-11$	2048	2^{11}
12	4084	$2^{12}-12$	4096	2^{12}	4084	$2^{12}-12$	4096	2^{12}

4.3 The Refined Border Correlation Function

We extend the border correlation function for partial words to $\beta' : A_\lozenge^* \to \mathbb{N}^*$ over a binary alphabet A, defined such that for all $0 < i < n$, a word w of length n is mapped to $m_0m_1...m_{n-1}$ where m_i is the length of a shortest border of σ^i (w) if it is bordered and m_i is 0 if it is unbordered. The function β' for full words is introduced in [6], and it is shown to be injective up to complementary words; that is, a single β'-image is shared only by a word and its complement.

Lemma 4.3 Let u and $\bar{v}$ be two full words such that $\beta'(u) = \beta'(v)$. Then either $u = v$ *or* $u = \bar{v}$.

Since w and $\bar{w}$ map to one distinct β'-image, $||\beta'(A^n)|| = 2^{n-1}$, which is significantly larger than the upper bound of $||\beta\ (An)||11$, which was calculated in Sect. 4.1. There also exists a relationship between $\beta'(w)$ and $\beta'(\text{rev}(w))$, the image of the reverse of w.

Lemma 4.4 *Let* $\beta'(w) = m_0m_1 \ldots m_{n-1}$ *fora word w of length n. Then* $\beta'(rev(w)) = m_0\ rev(m_1...m_{n-1})$.

Proof Let $\beta'(rev(w)) = m'_0m'_1 \ldots m'_{n-1}$. We begin by noting that for all $w \in A^*$, the length of the shortest border of w is equal to the length of the shortest border of $rev(w)$. Thus, $m_0 = m'_0$. A forward cyclic shift of $rev(w)$ and a backward cyclic shift of w yield words that are reverses of each other, i.e., $\sigma^i(\text{rev}(w)) = rev(\sigma^{-i}(w))$. Since these words are reverses, the lengths of their respective shortest borders are equal, so $m_i = m'_{n-i}$ for all i such that $0 < i < n$.

Interesting properties arise when we consider palindromes, words w such that $w = rev(w)$.

Proposition 4.2 *Let w be a full word of odd length n. Then w is a palindrome if and only if* $\beta'(\sigma^{|\frac{n}{2}|}(w))$ *is a palindrome.*

Proof For the forward implication, let w be a palindrome. So $w = rev(w)$, which implies that β $\beta'(w) = \beta'(\mathrm{rev}(w))$. Let $\beta'(w) = m_0m_1 \ldots m_{n\text{-}1}$. By Lemma 4.4, we know that $m_i = m_{n\text{-}i}$ for all i such that $0 < i < n$. Thus $\beta'(\mathrm{rev}(w)) = \beta'(w) = m_0u\ \mathrm{rev}(u)$ for some $u \in \mathbb{N}^*$. So $\sigma^{|\frac{n}{2}|}(\beta'(w)) = um_0\mathrm{rev}(u) = \beta'(\sigma^{|\frac{n}{2}|}(w))$, which is a palindrome.

For the reverse implication, let $v = \sigma^{|\frac{n}{2}|}(w)$ and suppose β'(v) is a palindrome so that $\beta'(\mathrm{v}) = \mathrm{umrev(u)}$. Therefore, $\sigma^{|\frac{n}{2}|}(\beta'(v)) = \beta'(\sigma^{|\frac{n}{2}|}(v)) = m\ \mathrm{rev}(u)u$. Since $v = \sigma^{|\frac{n}{2}|}(w)$, we have $\sigma^{|\frac{n}{2}|}(\mathrm{v}) = \sigma^{|\frac{n}{2}|}(\sigma^{|\frac{n}{2}|}(w)) = w$. So $\beta'(w) = m\ \mathrm{rev}(u)u$. By Lemma 4.4, $\beta'(\mathrm{rev(w)}) = m\ \underline{\mathrm{rev}(\mathrm{rev}(u)}u) = m\ \mathrm{rev}(u)u$. Thus $\beta'(w) = \beta'(\mathrm{rev}(w))$, and by Lemma 4.3 this implies that either $w = \mathrm{rev}(w)$ or $w = \overline{\mathrm{rev(w)}}$. However, the latter leads to a contradiction since $|w|$ is odd. Thus we have $w = \mathrm{rev}(w)$, so w is a palindrome.

Proposition 4.3 *Let w be a full word of even length n.*

1. *If* $\beta'(w)$ *is a palindrome, then* $\sigma^{\frac{n}{2}-1}(w)\sigma^{\frac{n}{2}}(w)$ *is a palindrome.*
2. *If w is a palindrome, then both* $\beta'(\sigma^{\frac{n}{2}}(w))\ \beta'(\sigma^{\frac{n}{2}+1}(w))$ *and* $\beta'(\sigma^{\frac{n}{2}}(w)\ \sigma^{\frac{n}{2}+1}(w))$ *are palindromes.*

Proof For the first statement, suppose that $\beta'(w)$ is a palindrome; that is, $\beta'(w) = u\ \mathrm{rev}(u)$ for some $u \in \mathbb{N}^*$. Then $\sigma^{\frac{n}{2}}\ (\beta'(w)) = \beta'(\sigma^{\frac{n}{2}}(w)) = \mathrm{rev}(u)u$. Let $u = m_0m_1 \ldots m_k$, so $\beta'(\sigma^{\frac{n}{2}}(w)) = m_km_{k-1}\ldots m_0m_0\ldots m_{k-1}m_k$. By Lemma 4.4, this implies $\beta'(\mathrm{rev}(\sigma^{\frac{n}{2}}(w))) = m_km_km_{k-1}\ldots m_0m_0\ldots m_{k-1}$, so $\beta'(\sigma^{\frac{n}{2}}(w)) = \sigma(\beta'(\mathrm{rev}(\sigma^{\frac{n}{2}}(w))))$. This is equivalent to $\beta'(\sigma(\mathrm{rev}(\sigma^{\frac{n}{2}}(w))))$. From our proof of Lemma 4.4 we know that $\sigma^i(\mathrm{rev}(w)) = \mathrm{rev}(\sigma^{-i}(w))$, so

$$\beta'(\sigma(\mathrm{rev}(\sigma^{\frac{n}{2}}(w)))) = \beta'(\mathrm{rev}(\sigma^{-1}(\sigma^{\frac{n}{2}}(w)))) = \beta'(\mathrm{rev}(\sigma^{\frac{n}{2}-1}(w)))$$

Therefore, by Lemma 4.3, $\beta'(\sigma^{\frac{n}{2}}(\mathrm{w})) = \beta'(\mathrm{rev}(\sigma^{\frac{n}{2}-1}(\mathrm{w})))$ implies that either $\sigma^{\frac{n}{2}}(\mathrm{w}) = \mathrm{rev}\ (\sigma^{\frac{n}{2}-1}\ (\mathrm{w}))$ or $\sigma^{\frac{n}{2}}(\mathrm{w}) = \overline{\text{rev(}\sigma\text{\^{}(n/2 -1) (w))}}$. However, the latter case presents a contradiction. To see this, let $\sigma^{\frac{n}{2}}(\mathrm{w}) = c_0c_1\ldots c_{n-1}$. Then $\sigma^{\frac{n}{2}-1}(w) = c_{n-1}c_0\ldots c_{n-2}$, so $\mathrm{rev}(\sigma^{\frac{n}{2}-1}\ (w)) = c_{n-2}\ldots c_0c_{n-1}$. Hence, if $\sigma^{\frac{n}{2}}\ (w) = \overline{\text{rev (}\sigma\text{\^{}(n/2 -1) (w))}}$, we have that the final letters of $\sigma^{\frac{n}{2}}(w)$ and $\overline{\text{rev (}\sigma\text{\^{}(n/2 -1) (w))}}$ are equal, so $c_{n-1} = \overline{c_{n-1}}$, which is a contradiction. Thus, we may conclude that $\sigma^{\frac{n}{2}}(w) = \mathrm{rev}(\sigma^{\frac{n}{2}-1}(w))$, and $\sigma^{\frac{n}{2}-1}(\mathrm{w})\ \sigma^{\frac{n}{2}}(w)$ is a palindrome.

For the first half of the second statement, suppose that w is a palindrome. Then $w = \mathrm{rev}(w)$, so $\beta'(w) = \beta'(\mathrm{rev}(w))$. By the same reasoning as the corresponding case for odd lengths, we let $\beta'(w) = m_0m_1\ldots m_{n-1}$ so $m_i = m_{n-i}$ for all i such that $0 < i < n$. In this case, however, the word $m_1m_2\ldots m_{n-1}$ is of odd length, so we can let $\beta'(w) = \beta'(\mathrm{rev}(w)) = m_0um_{\frac{n}{2}}\ \mathrm{rev}(u)$ for some $u \in \mathbb{N}^*$. Consequently, $\sigma^{\frac{n}{2}}(\beta'(w)) = \beta'(\sigma^{\frac{n}{2}}(w)) = m_{\frac{n}{2}}\ \mathrm{rev}(u)m_0u$ and $\sigma^{\frac{n}{2}+1}(\beta'(w)) = \beta'(\sigma^{\frac{n}{2}+1}(w)) = \mathrm{rev}(u)m_0um_{\frac{n}{2}}$. Since $m_{\frac{n}{2}}\ \mathrm{rev}(u)m_0u = \mathrm{rev}(\mathrm{rev}(u)m_0um_{\frac{n}{2}})$, this allows us to conclude that the concatenation $\beta'(\sigma^{\frac{n}{2}}(w))\beta'(\sigma^{\frac{n}{2}+1}(w))$ is a palindrome.

For the second half of the second statement, again w is a palindrome, so $w = u\ \mathrm{rev}(u)$ for some $u \in A^*$. Then $\sigma^{\frac{n}{2}}(w) = \mathrm{rev}(u)u$. Let $u = c_0c_1\ldots c_k$, so $\sigma^{\frac{n}{2}}(w)\ \sigma^{\frac{n}{2}+1}(w) = c_k\ldots c_0c_0\ldots c_kc_{k-1}\ldots c_0c_0\ldots c_kc_k$. We have that the nth cyclic shift of this word is $c_{k-1}\ldots c_0c_0\ldots c_kc_kc_k\ldots c_0c_0\ldots c_k$, and the $(n-1)$th cyclic shift is $c_kc_{k-1}\ldots c_0c_0\ldots c_kc_kc_k\ldots c_0c_0\ldots c_{k-1}$. These words are reverses of each other, so the lengths of their shortest borders are equal. Let v be the $(n-1)$th cyclic shift, so rev(v) is the nth cyclic shift. Recall, again, that σ^{i} (rev(v)) = rev(σ^{-i} (v)), implying that the lengths of their shortest borders are equal. Notice that σ^{i} (rev(v)) = σ^{n+i} ($\sigma^{\frac{n}{2}}$ (w)$\sigma^{\frac{n}{2}+1}$(w)), so the above equivalence implies that $\beta'(\sigma^{\frac{n}{2}}(w)\sigma^{\frac{n}{2}+1}(w))$ is a palindrome.

In the partial word case, we use a slightly modified definition of a palindrome, referring to a partial word w as a compatible-palindrome, denoted ↑-palindrome, if $w \uparrow$ rev(w). Because Lemma 4.4 holds for partial words as well, we are able to generalize our propositions as follows:

Proposition 4.4 *Let a partial word w be a ↑-palindrome, and let v = w* $\vee$ rev(w).

1. *If w is of odd length, then* $\beta'(\sigma^{\lceil\frac{n}{2}\rceil}(v))$ *is a palindrome.*
2. *If w is of even length, then both* $\beta'(\sigma^{\frac{n}{2}}(v))\beta'(\sigma^{\frac{n}{2}+1}(v))$ *and* $\beta'(\sigma^{\frac{n}{2}}(v)\sigma^{\frac{n}{2}+1}(v))$ *are palindromes.*

Proof If there are any holes in v, then they are located in symmetric positions. Hence, v = rev(v), so v is a palindrome in the same sense as full words. The proof of the statement is thus the same as that of the full word case.

5 Critical Factorization and Borderedness of Partial Words

Harju and Nowotka have presented a characterization of the relationship between borderedness and critical factorization of full words. We expand these to include partial words. We begin by recalling the definition of a critical factorization and the statement of the critical factorization theorem.

A nonnegative integer $q < |w| - 1$ is called a *point* and refers to the space between letters $w(q)$ and $w(q + 1)$. A *local period* of w at point q is a positive integer p such that there exist nonempty partial words u, v, x, y where $w = uv$, $|u| = q + 1$, $|x| = |y| = p$, $x \uparrow y$ and there exist partial words r and s such that one of the following conditions holds:

1. $u = rx$ and $v = ys$ (internal square),
2. $x = ru$ and $v = ys$ (left-external square),
3. $u = rx$ and $y = vs$ (right-external square),
4. $x = ru$ and $y = vs$ (left- and right-external square).

In this case, we call x and y *repetition words*. The *minimal local period* of w at point q is denoted $p(w, q)$.

A factorization of a partial word w such that $w = uv$ is called *critical* if the minimal local period of w at point $i = | u | -| - 1$ is equal to the minimal weak period of w, i.e. $p(w, |u| - 1) = p'(w)$. In this case, the point i is called a *critical point*. This definition is adapted from that of full words, where a factorization is critical if the minimal local period at point i is equal to the minimal strong period of a full word, w. The critical factorization theorem states that for all full words of length greater than one, there exists at least one critical factorization. Blanchet-Sadri et al. have extended the critical factorization theorem to partial words, proving that there exists at least one critical factorization for all partial words meeting a specified set of conditions [4–6].

A point p, with $0 \leq p < |w|$ is called an *internal critical point* of w if $p + |w|$ is a critical point of www.

Lemma 5.1 *Let w be a partial word such that* $|w| = n$ *and let* $u = \sigma^{i}(w)$ *with* $0 \leq i < n$. *The point p is an internal critical point of w if and only if the point*

$$q = \begin{cases} p - i & \text{if } p \geq i \\ p + n - i & \text{otherwise} \end{cases}$$

is an internal critical point of u.

Proof We prove only the forward implication, as the reverse follows directly. Harju and Nowotka [6] have shown that for a full word w, if $v = \sigma^j(w)$ for some $0 \leq j < n$, then there exist full words x and z such that $www = xvvz$ and $|x| = j$. Moreover, $uuu = x'vvz'$ where $|x'| = j - i$ if $j \geq i$ and $|x'| = j + n - i$ otherwise. These results hold for the partial word case as well.

Suppose p is an internal critical point of w. Let $p'(w) = m$, and let r_0 and r_1 be the shortest repetition words at point $p + n$ in www. Clearly, www has a weak period of length n. If this is the minimal weak period, then $|r_0| = |r_1| = n$. In this case, as Harju and Nowotka have proven, we have that $r_0 = r_1$ is a conjugate of w, say $\sigma^s(w)$. In fact, by writing $www = xr_0r_1z = x\sigma^s(w)\sigma(w)\sigma^s(w)z$, then by the above we see that $|x| = s$, and hence, $|x\sigma^s(w)| = s + n$. In order to have r_0 and r_1 be repetition words at point $p + n$, then, it must be the case that $s = p + 1$. Then $uuu = x'\sigma^s(w)\sigma^s(w)z'$, where $|x'| =s - i = p - i + 1$ if $p \geq i$ and $|x'| =s + n - i = p + n - i + 1$ otherwise. This implies that $|x'| = q + 1$, and hence the point q is a critical point of uuu. So q is an internal critical point of u, completing the proof.

However, unlike the case of full words studied by Harju and Nowotka, the partial word www may have a minimal weak period of length shorter than n. Let $p'(w) = m < n$ and let r_0 and r_1 be the shortest repetition words at point $p + n$ in www. By definition, $|r_0| = |r_1| = m$. Suppose that we form a complete conjugate of w, say $\sigma^s(w)$, by looking at the letters that appear to the left of r_0 and to the right of r_1 in www. Then we will have $\sigma^s(w) = ur_0 = r_1y$, and from above we see that $www = xur_0r_1yz = x\sigma^s(w)\sigma^s(w)z$, where $|x| = s$. Again, this forces that $s = p + 1$. Then $uuu =x'ur_0r_1yz'$, where $|x'| =s - i = p - i + 1$ if $p \geq i$ and $|x'| = s + n - i = p + n - i + 1$ otherwise. Hence, by the same reasoning as above, r_0 and r_1 are the shortest repetition words at point $q + n$ in uuu, so q is an internal critical point of u.

The following theorem clarifies the relationship between internal critical points and borderedness. Note that if $p'(www) = n$, then the statement "the length of a shortest border of $\sigma\ \sigma^{p+1}(w)$ is q" means that $\sigma\ \sigma^{p+1}(w)$ is unbordered.

Theorem 5.1 *Let w be a partial word of length n such that* $p'(www) = q$ *and let* $0 \leq p < n$. *Then p is an internal critical point of w if and only if the length of a shortest border of* $\sigma^{p+1}(w)$ *is q.*

Proof Assume p is an internal critical point of w. Then $www = xur_0r_1yz$, where $|x| = p + 1$, $\sigma^{p+1}(w) = ur_0 = r_1y$, and r_0 and r_1 are the shortest repetition words at point $p + n$ of www. Because $\sigma^{p+1}(w) = ur_0 = r_1y$ and $r_0 \uparrow r_1$ (there exists r such that $r_0 \subset r$ and $r_1 \subset r$), it is clear that $\sigma^{p+1}(w)$ has a border r of length $|r_0| = |r_1| = q$. Furthermore, r is necessarily unbordered, so there can be no shorter border of $\sigma^{p+1}(w)$.

Now assume that the length of a shortest border of $v = \sigma^{p+1}(w)$ is q. Then $v = ur_0 = r_1y$ for some partial words u, r_0, r_1, y such that $|r_0| = |r_1| = q$, $r_0 \uparrow r_1$, and $r_0 \subset r$ and $r_1 \subset r$ for some r. We have $www = xvvz = xur_0r_1yz$ with $|x| = p + 1$ and r_0 and r_1 are repetition words of www at point $p + n$. Moreover, since r is unbordered, r_0 and r_1 are the shortest repetition words at this point. Thus, $|r_0| = |r_1| = q = p'(www)$, so p is an internal critical point of *www*.

In calculating the minimal weak period of www, it is not necessary to check all possible lengths. The following algorithm gives a decision procedure for determining $p'(www)$.

Algorithm 5.1 Given a partial word w of length n:
Step 1: Find $\beta(w)$. If $\beta(w) \neq b^n$, then output $p'(www) = n$. Otherwise, go to Step 2.
Step 2: For each i such that $0 \leq i < n$, determine the lengths of all borders of $\sigma^i(w)$.
Step 3: Find the smallest m such that every conjugate of w has a border of length m. Output $p'(www) = m$.

Proposition 5.1 *Given input partial word w of length n, Algorithm* 5.1 *outputs* $p'(www)$.

Proof In order to prove the correctness of Algorithm 5.1, let us suppose that *www* has a weak period of length m.

We prove by contradiction that every conjugate of w must have a border of length m. Let i be an integer such that $0 \leq i < n$, and assume that $\sigma^i(w)$ does not have a border of length m. This implies that $\sigma^i(w) = u_0vu_1$, where $|u_0| = |u_1| = m$ and $u_0 \not\uparrow u_1$. However, recall that $www = x\sigma^i(w)\sigma(w)\sigma^i(w)z = xu_0vu_1u_0vu_1z$, which contains the factor u_1u_0. Since $|u_1| = |u_0| = m$ and *www* has a weak period of length m, it must be the case that $u_1 \uparrow u_0$. Since $u_0 \not\uparrow u_1$, this is a contradiction.

We show, in addition, that it is impossible for every conjugate of w to have a border of length $k < m$. Again, for the sake of a contradiction, we assume that there exists an integer $k < m$ such that every conjugate of w has a border of length k. Consider the conjugate $\sigma^k(w) = u_1vu_0$, where $|u_0| = |u_1| = k$ and $u_0 \uparrow u_1$. This gives us that $w = u_0u_1v$ and $u_0 \uparrow u_1$. Since every conjugate of w is a factor of *www*, we can conclude by the same reasoning that every length-k factor of *www* is compatible with the next length-k factor, and hence, *www* has a weak period of length k. However, this contradicts our assumption that m is the minimal weak period of *www*.

Example 5.1 Let $w = bba\Diamond b\Diamond a$. We first determine that $\beta(w)\,\beta(w) = bbbbbbb$. In the following table, we list all conjugates of w and the lengths of all borders of each conjugate.

Conjugate	Lengths of borders
$bba\diamond b\diamond a$	3, 4
$ba\diamond b\diamond ab$	1, 3, 4
$a\diamond b\diamond abb$	3, 4
$\diamond b\diamond abba$	1, 3, 4
$b\diamond abba\diamond$	1, 3, 4
$\diamond abba\diamond b$	1, 3, 4
$abba\diamond b\diamond$	1, 3, 4

Every conjugate of w has a border of length 3 and one of length 4. Choosing the smallest of these numbers, we find that $p'(www) = 3$.

6 Conclusion

There appears to be a relationship between borderedness and periodicity that remains largely unexplored. The ternary correlation of a partial word w is a ternary vector v such that | v | = | w |, v0 = 1, and for all i such that 0 < i < n:

$$v_i = \begin{cases} 1 & \text{if } w \text{ has strong period } i \\ 2 & \text{if } w \text{ has weak period } i \\ 0 & \text{otherwise} \end{cases}$$

We expect a correspondence between this vector and the image of a word under the refined border correlation function.

Although we have investigated properties of $\beta\,\beta'$-images of binary words, we note that the function will no longer be injective up to complements when the alphabet size is greater than two, so many of the properties that we have discussed will not hold. Also, while we have given an upper bound on the number of distinct $\beta\,\beta$- and β'-images of binary partial words of a given length, one could also explore the population distribution of these images, i.e., the number of words that share a given border correlation.

References:

1. Berstel, J.: Axel Thue's papers on repetitions in words: a translation. Publications du LaCIM 20
2. Blanchet-Sadri, F.: Algorithmic Combinatorics on Partial Words. Chapman & Hall, Boca Raton (2007)
3. Blanchet-Sadri, F., Duncan, S.: Partial words and the critical factorization theorem. J. Comb. Theory Ser. A **109**, 221–245 (2005). www.uncg.edu/mat/cft
4. Blanchet-Sadri, F., Wetzler, N.D.: Partial words and the critical factorization theorem revisited. Theor. Comput. Sci. **385**, 179–192 (2007). www.uncg.edu/mat/research/cft2
5. Currie, J.D.: There are ternary circular square-free words of length n for $n \geq \geq 18$. Electron. J. Comb. **9**, N10 (2002)
6. Harju, T., Nowotka, D.: Border correlation of binary words. J. Comb. Theory Ser. A **108**, 331–341 (2004)
7. Lothaire, M.: Combinatorics on Words. Cambridge University Press, Cambridge (1997)
8. Lothaire, M.: Algebraic Combinatorics on Words, Encyclopedia of Mathematics and its Applications, vol. 90. Cambridge University Press, Cambridge (2002)
9. Thue, A.: Ober die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. Det Kongelige Norske Videnskabersselskabs Skrifter, I Mat.-Nat. K1. Christiana **1**, 1–67 (1912)

Notes:

This material is based upon work supported by the National Science Foundation under Grant No. DMS–0452020. We thank the referees of a preliminary version of this paper for their very valuable comments and suggestions.

A World Wide Web server interface has been established at www.uncg.edu/cmp/research/bordercorrelation for automated use of the program.