On generalized highly potential words

Kristina Ago[a], Bojan Bašić[a], Stefan Hačko[a], Danijela Mitrović[b]

[a]*Department of Mathematics and Informatics, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia*

[b]*Mathematical Institute of the Serbian Academy of Sciences and Arts, Kneza Mihaila 36, 11000 Beograd, Serbia*

`kristina.ago@dmi.uns.ac.rs`, `bojan.basic@dmi.uns.ac.rs`, `stefan.hacko@dmi.uns.ac.rs`, `danijela.mitrovic@dmi.uns.ac.rs`

Abstract

The number of palindromic factors of a given finite word is bounded above by its length increased by 1. The difference between this upper bound and the actual number of palindromic factors of a given word is called the *palindromic defect* (or only *defect*) of a given word (by definition, the defect is always nonnegative). Though the definition of defect fundamentally relies on finiteness of a given word, it can be naturally extended to infinite words. There are many results in the literature about words of defect 0, but there are significantly less results about infinite words of finite positive defect. In this article we construct a new family of infinite words whose defect is finite, and in many cases positive (with fully characterized cases when the defect is 0). All the words from our family have the set of factors closed under reversal, and each of them is either periodic (which is a less interesting case, and explicitly characterized), or recurrent but not uniformly recurrent. The fact that they are not uniformly recurrent (unless they are periodic) is of a particular significance since: first, there are some results and examples here and there featuring uniformly recurrent words of finite defect, while next to nothing is known about aperiodic words that are not uniformly recurrent; second, it is known that any uniformly recurrent word of finite defect is a morphic image of some word of zero defect, which suggests that uniformly recurrent words are in a way pretty "tame," and that those that are not uniformly recurrent are an unexplored territory that deserves a closer look.

Mathematics Subject Classification (2010): 68R15

Keywords: palindrome, palindromic defect, word defect, full word, rich word

1 Introduction

The notion of the so-called palindromic defect of a given word was sparked by the result of Droubay, Justin and Pirillo [13], who noted that the number of palindromic factors of a given finite word is bounded above by its length increased by 1. The difference between this upper bound and the actual number of palindromic factors of a given word is called the *palindromic defect* (or only *defect*) of a given word [7] (by definition, the defect is always nonnegative). Though the definition of defect fundamentally relies on finiteness of a given word, it turns out that it can be naturally extended to infinite words (the defect of an infinite word is defined as the supremum of defects of all of its finite factors). Words of defect 0 are called *full* [7] or *rich* [14], and there are many results about them in the literature [9, 15, 17, 19, 20, 22, 24, 25].

However, infinite words of finite positive defect have been studied significantly less. One of the reasons for that is the fact that explicit constructions of such words (maybe with some additional constraints, such as aperiodicity, since periodic words are more-or-less straightforward to analyze) are somewhat deficient in the literature. For example, aperiodic words of finite positive defect, having the set of factors closed under reversal, had been deemed interesting from the point of view of some (then open) conjectures [6, 8], but examples of such words were missing. In the article [5], the second author constructed an infinite family of infinite words, called *highly potential words,* which are all aperiodic, have the set of factors closed under reversal, and are of finite positive defect (in fact, the presented construction shows a method to obtain such a word from *any* finite nonpalindromic word). As one can see in that article, those words seem to be a useful supply of examples and counterexamples for various problems of words (which explains their name). We should also say that chronologically the first example of an aperiodic infinite word of finite positive defect, whose set of factors is closed under reversal, had been given somewhat earlier: see [2, Example 3.4], where such a word has been constructed, one that is uniformly recurrent. In the article [4] an example that is not uniformly recurrent has been constructed, which was used to demonstrate a flaw in a proof from the article [3]; this word, although it has much in common with the family of highly potential words, does not belong to that family.

In this article we construct a new family of infinite words whose defect is finite, and in many cases positive (with fully characterized cases when the defect is 0). The constructed family contains, as two special cases, both the family of highly potential words (because of this, we dub them *generalized highly potential words*), as well as the mentioned word from [4]. Further, in [14, Proposition 2.10] the authors show the existence of rich infinite words that are recurrent but not uniformly recurrent, by providing three examples; it turns out that all these three words also belong to the class of generalized highly potential words. We believe that all this suggests that our construction extends the class of highly potential words in a fairly noteworthy way. All the words from our family have the set of factors closed under reversal, and each of them is either periodic (which is a less interesting case, and explicitly characterized), or recurrent but not uniformly recurrent. The fact that they are not uniformly recurrent (unless they are periodic) is of a particular significance since: first, there are some results and examples here and there featuring uniformly recurrent words of finite defect (see, e.g., [14, Proposition 4.8], or the article [2], or the counterexample to the so-called Zero defect conjecture from [11], which is defined as a fixed point of a primitive morphism, and it is known [1, Theorem 10.9.5] that fixed points of primitive morphisms are always uniformly recurrent), while next to nothing is known about aperiodic words that are not uniformly recurrent; second, it is shown in [18, Theorem 2] that any uniformly recurrent word of finite defect is a morphic image of some word of zero defect (while the result that the authors obtain without assuming uniform recurrence is weaker, and in the last section they discuss the significance of uniform recurrence and leave as an open question whether the stronger result is valid without it), everything of which suggests that uniformly recurrent words are somewhat easier to work with, and that those that are not uniformly recurrent are less explored territory that deserves a closer look.

2 On words

In this section we recall basic definitions and properties that will be needed through the article.

A *word* (respectively *infinite word*) is a finite (respectively infinite) sequence of symbols taken from a nonempty finite set Σ, which is called the *alphabet*. (We shall sometimes abuse the terminology and say only "word" when it is clear from the context that it must be infinite, or additionally emphasize "finite word" when we feel that this is appropriate.) Let Σ^* denote the set of all finite words and by Σ^∞ the set of all finite or infinite words. If $w = a_1a_2...a_n$ with $a_1, a_2, \ldots, a_n \in \Sigma$, we say that the *length* of w is n, and write $|w| = n$. The unique word of length 0, called the *empty word*, is denoted by ε. The *concatenation* (or *product*) of words u and v, $u = a_1a_2 \dots a_n$ and $v = b_1b_2 \dots b_m$, is the word $a_1a_2 \dots a_nb_1b_2 \dots b_m$, denoted by uv. The product uv for $u \in \Sigma^*$ and $v \in \Sigma^\infty \setminus \Sigma^*$ can be similarly defined. For a word w and a positive integer k we write w^k for the word $\underbrace{ww\dots w}_{k}$ and w^∞ for the infinite word $wwww\dots$; it is also convenient to define $w^0 = \varepsilon$ for any word w. A word $w \in \Sigma^*$ is *primitive* if and only if it is not of the form z^k for $z \in \Sigma^* \setminus \{\varepsilon\}$ and an integer k, $k \geqslant 2$. We define the map $\widetilde{\ }: \Sigma^* \to \Sigma^*$, called *reversal*, as follows: if $w = a_1a_2 \dots a_n$, where $a_1, a_2, \dots, a_n \in \Sigma$, then $\widetilde{w} = a_na_{n-1} \dots a_1$.

A word $u \in \Sigma^*$ is called a *factor* (respectively *prefix*, *suffix*) of a word $w \in \Sigma^\infty$ if and only if there exist words $x \in \Sigma^*$ and $y \in \Sigma^\infty$ such that $w = xuy$ (respectively $w = uy$, $w = xu$). The set of all factors (respectively prefixes, suffixes) of w is denoted by $\mathrm{Fact}(w)$ (respectively $\mathrm{Pref}(w)$, $\mathrm{Suff}(w)$).

We write $w[i]$ for the i^{th} letter of a word $w \in \Sigma^\infty$, and for any pair (i, j) of integers such that $1 \leqslant i \leqslant j \leqslant |w|$ let $w[i, j]$ denote the factor of w whose first letter is the i^{th} letter of w and whose last letter is the j^{th} letter of w (obviously, $w[i, i] = w[i]$). By convention, this operation has precedence over concatenation; in other words, $uv[i]$ (and similarly $uv[i, j]$) will always denote $u(v[i])$, not $(uv)[i]$.

For words u and v, let $|u|_v$ denote the number of distinct occurrences of v in u, that is:

$$|u|_v = |\{i : 1 \leqslant i \leqslant |u| - |v| + 1, u[i, i + |v| - 1] = v\}|.$$

An infinite word w is *periodic* if and only if it is of the form $w = u^\infty$ for some $u \in \Sigma^*$, it is *eventually periodic* if and only if it is of the form vu^∞ for some u, $v \in \Sigma^*$, and it is *aperiodic* if and only if it is not eventually periodic. An infinite word w is *recurrent* if and only if each of its factors occurs infinitely many times in w, and it is *uniformly recurrent* if and only if for every finite factor u of w there exists an integer n such that $u \in \mathrm{Fact}(v)$ for every $v \in \mathrm{Fact}(w)$ such that $|v| = n$.

The following three theorems are well-known; see [1]: Exercise 10.50a), Example 10.9.1 and Exercise 10.37, respectively.

Theorem 2.1. *For an infinite word w, if* $\mathrm{Fact}(w)$ *is closed under reversal, then w is recurrent.*

Theorem 2.2. *Every periodic word is uniformly recurrent.*

Theorem 2.3. *Every recurrent, eventually periodic word is periodic.*

Before the next theorem (also well-known, see [16, Proposition 1.3.2]), we need the following notion: a word w' is a *conjugate* of a word w if there exist words x and y such that $w = xy$ and $w' = yx$.

Theorem 2.4. *Let $x, y \in \Sigma^*\backslash\{\varepsilon\}$. Then $xy = yx$ if and only if there exist $t \in \Sigma^*$ and positive integers p, q such that $x = t^p$, $y = t^q$. In other words, if a word is equal to one of its conjugates (different from itself), then it must be a power of exponent at least 2.*

A word u is a *palindrome* (or *palindromic*) if and only if $u = \widetilde{u}$. Let $\mathrm{Pal}(w) = \{u \in \mathrm{Fact}(w) : u = \widetilde{u}\}$. The following inequality was noted by Droubay, Justin and Pirillo [13, Proposition 2].

Theorem 2.5. *Let w be a finite word. Then:*

$$|\mathrm{Pal}(w)| \leqslant |w| + 1.$$

Inspired by this inequality, Brlek et al. [7] introduced the notion of *palindromic defect* (or only *defect*) of a word w, denoted by $D(w)$, and defined as:

$$D(w) = |w| + 1 - |\mathrm{Pal}(w)|.$$

They noticed that the defect of a word w is no smaller than the defect of any of its factors; in other words:

Theorem 2.6. *Let w be a finite word and $v \in \mathrm{Fact}(w)$. Then*

$$D(v) \leqslant D(w).$$

This motivates the following extension of the definition of defect to infinite words: for $w \in \Sigma^\infty\backslash\Sigma^*$, we define

$$D(w) = \sup_{v \in \mathrm{Fact}(w)} D(v).$$

(Of course the previous equality also holds for finite words.) Note that the defect of any finite or infinite word is always nonnegative.

For the end of this section, we present a theorem [7, Corollary 8] that gives a way to calculate the defect of an infinite word in a particular case.

Theorem 2.7. *If p is a primitive word that is a product of two palindromes (one of which can be empty), then there exists a conjugate p' of p such that*

$$D(p^\infty) = D(p').$$

3 Highly potential words

A class of infinite words called *highly potential words* has been introduced by the second author in [5]. Given a word w that is not a palindrome, let c denote any letter that does not appear in w, and let:

$$w_0 = w;$$

$$w_i = w_{i-1} c^i \widetilde{w_{i-1}},\ i \in \mathbb{N};$$

$$\mathrm{hpw}(w) = \lim_{i \to \infty} w_i.$$

(Note that $\mathbb{N}$ denotes the set of positive integers, while $\mathbb{N}_0$ denotes the set of nonnegative integers.) The infinite word $\mathrm{hpw}(w)$ is called *highly potential word generated by w*.

The main properties of highly potential words are given in the following theorem.

Theorem 3.1. *Let* $\mathrm{hpw}(w)$ *be a highly potential word. Then:*

- $\mathrm{Fact}(\mathrm{hpw}(w))$ *is closed under reversal;*
- $\mathrm{hpw}(w)$ *is recurrent;*
- $\mathrm{hpw}(w)$ *is not uniformly recurrent;*
- $\mathrm{hpw}(w)$ *is aperiodic;*
- $D(\mathrm{hpw}(w)) = D(w) + 1$.

4 Generalized highly potential words

In this section we define generalized highly potential words and investigate their properties. We give a necessary and sufficient condition for periodicity. We show that their set of factors is closed under reversal (which implies that they are recurrent), and we further show that the ones that are not periodic are not uniformly recurrent. We also prove that their defect is always finite, and give a necessary and sufficient condition for the defect to be positive. In a separate subsection at the end we analyze periodic generalized highly potential words (which is a less interesting case).

4.1 Construction

Definition 4.1. Let $w, u, v \in \Sigma^*$, where $wuv \neq \varepsilon$ and u and v are palindromes, and let $A = (a_i)_{i\in\mathbb{N}}$ be a strictly increasing sequence of positive integers. We recursively define:

$$w_0 = w;$$

$$w_i = w_{i-1}(uv)^{a_i}u\widetilde{w_{i-1}}, \;\; i \in \mathbb{N};$$

and then:

$$\mathrm{ghpw}(w, u, v, A) = \lim_{i\to\infty} w_i. \tag{1}$$

(The limit is well-defined since each w_i is a prefix of w_{i+1}.) The infinite word $\mathrm{ghpw}(w, u, v, A)$ is called *generalized highly potential word generated by w, u, v and A.*

We first note that generalized highly potential words are indeed a generalization of highly potential words: if w is a nonpalindromic word, c a letter that does not appear in w, and I the sequence $(i)_{i\in\mathbb{N}}$, then we clearly have

$$\mathrm{hpw}(w) = \mathrm{ghpw}(w, \varepsilon, c, I).$$

Also, the word from [4] mentioned in the Introduction is $\mathrm{ghpw}(1213121, 3, 2, I)$. The three words from [14] that served as a demonstration of existence of rich infinite words that are recurrent but not uniformly recurrent (also mentioned in the Introduction) are the following ones: 1) $\varphi^\infty(a)$ where $\varphi : a \mapsto aba$, $b \mapsto bb$ (an example taken from [12], where it was considered for another purpose); 2) the Cantor word (also known as the Sierpiński word), that is, $\varphi^\infty(a)$ where $\varphi : a \mapsto aba$, $b \mapsto bbb$ (a well-known word; see, for example, [21], which the authors cite); 3) $\varphi^\infty(a)$ where $\varphi : a \mapsto abab$, $b \mapsto b$ (the authors' own example). It is easy to see that they can be represented as $\mathrm{ghpw}(a, \varepsilon, b, (2^{i-1})_{i\in\mathbb{N}})$, $\mathrm{ghpw}(a, \varepsilon, b, (3^{i-1})_{i\in\mathbb{N}})$ and $\mathrm{ghpw}(a, \varepsilon, b, I)$, respectively.

4.2 Standard form

Different quadruples of parameters (w, u, v, A) can lead to the same generalized highly potential word. In the following lemma we shall prove that for each generalized highly potential word there can be chosen a quadruple with some particular properties that will be very useful.

Lemma 4.2. *Let $\mathrm{ghpw}(w, u, v, A)$ be a generalized highly potential word. Then there are words w^S, u^S, v^S and a sequence A^S such that w^S is a palindrome, $u^S v^S$ is primitive, and*

$$\mathrm{ghpw}(w, u, v, A) = \mathrm{ghpw}(w^S, u^S, v^S, A^S).$$

Proof. We first show that a quadruple can be chosen such that w is a palindrome. Suppose that $w \neq \widetilde{w}$. Since $w_1 = w(uv)^{a_1} u\widetilde{w}$, we see that w_1 is always a palindrome. It is not hard to see that

$$\mathrm{ghpw}(w, u, v, A) = \mathrm{ghpw}(w_1, u, v, B),$$

where $B = (b_i)_{i\in\mathbb{N}}$, $b_i = a_{i+1}$.

Suppose now that uv is not primitive, that is, $uv = t^n$ for a word t and an integer n, $n \geqslant 2$. We can assume that t is primitive. Note that we can write $t = u'v'$ where $u' \in \mathrm{Suff}(u)$ and $v' \in \mathrm{Pref}(v)$ (one of u' and v' can be ε). Then we have:

$$u = (u'v')^k u',$$
$$v = v'(u'v')^l$$

for some integers k and l that satisfy $k + l = n - 1$. Also, since u' is both the prefix and the suffix of the palindrome u, we conclude that u' is also a palindrome; in a similar manner, v' is a palindrome, too. We prove that

$$\mathrm{ghpw}(w, u, v, A) = \mathrm{ghpw}(w, u', v', C),$$

where C is an increasing sequence defined by $C = (c_i)_{i\in\mathbb{N}}$,

$$c_i = na_i + k.$$

This follows by induction, noting that

$$\begin{aligned} w_{i+1} &= w_i(uv)^{a_{i+1}} uw_i = w_i((u'v')^n)^{a_{i+1}} (u'v')^k u'w_i \\ &= w_i(u'v')^{na_{i+1}+k} u'w_i = w_i(u'v')^{c_{i+1}} u'w_i. \end{aligned}$$

The proof is completed. ■

If a quadruple (w, u, v, A) is such that w is a palindrome and uv is primitive, we shall say that $\mathrm{ghpw}(w, u, v, A)$ is in *standard form*. The previous lemma shows that each generalized highly potential word can be presented in standard form.

Remark 4.3. The assumption that uv is primitive will be used very much, most of the times in the form of the following consequence: in that case, by Theorem 2.4, uv appears as a factor of $uvuvuv\ldots$ only at the “obvious positions” (in other words, $|uvuv|_{uv} = 2$; to be more precise, by this term we shall onward refer to the appearances of uv within $uvuvuv\ldots$ that begin at a position i where $i \equiv 1 \pmod{|uv|}$); furthermore, the same also holds for each conjugate of uv (each conjugate of a primitive word is again primitive, which is also easily seen by Theorem 2.4).

Another (technical) consequence of the assumption that uv is primitive (that will also be useful) is given in the following lemma.

Lemma 4.4. *Assume that u and v are palindromes, $uv \neq \varepsilon$, such that the word uv is primitive. Let x be a palindrome such that $|x| \geqslant 2|uv| - 1$ and $x\big[1, \lfloor \frac{|x|}{2} \rfloor + |uv|\big] = (vu)^\infty\big[1, \lfloor \frac{|x|}{2} \rfloor + |uv|\big]$. Then there exists a positive integer m such that $x = (vu)^m v$.*

Proof. Let $y = x\big[\lfloor \frac{|x|}{2} \rfloor + 1, \lfloor \frac{|x|}{2} \rfloor + |uv|\big]$. By the lemma's assumption, y is a conjugate of vu, and thus we may write $y = (vuvu)[i, j]$ for some i and j, where $1 \leqslant i, j \leqslant |vuvu|$ and $j - i + 1 = |uv|$. Since x matches $(vu)^\infty$ for the first $\lfloor \frac{|x|}{2} \rfloor + |uv|$ letters, Remark 4.3 leads to

$$\left\lfloor \frac{|x|}{2} \right\rfloor + 1 \equiv i \pmod{|uv|}.$$

We also have $\widetilde{y} = (\widetilde{vuvu})[2|uv| - j + 1, 2|uv| - i + 1] = (uvuv)[2|uv| - j + 1, 2|uv| - i + 1]$ and (since x is palindromic) $\widetilde{y} = x\big[\lceil \frac{|x|}{2} \rceil - |uv| + 1, \lceil \frac{|x|}{2} \rceil\big]$; this implies (by again appealing to Remark 4.3 in a similar manner)

$$\left\lceil \frac{|x|}{2} \right\rceil \equiv |v| + (2|uv| - i + 1) \equiv |v| + 1 - i \pmod{|uv|}.$$

Adding the two congruences together gives $\lfloor \frac{|x|}{2} \rfloor + 1 + \lceil \frac{|x|}{2} \rceil \equiv |v| + 1 \pmod{|uv|}$, that is, $|x| \equiv |v| \pmod{|uv|}$. Together with the lemma's assumption and the fact that x is palindromic, this gives the required conclusion. ■

4.3 Basic properties

We first present a necessary and sufficient condition for a generalized highly potential word to be periodic.

Theorem 4.5. *Let* $\mathrm{ghpw}(w, u, v, A)$ *be given in standard form. Then* $\mathrm{ghpw}(w, u, v, A)$ *is periodic if and only if either $w = (vu)^m v$ for a nonnegative integer m, or exactly one of the words w, u and v is nonempty.*

Proof. We first assume $w = (vu)^m v$. Then

$$\mathrm{ghpw}((vu)^m v, u, v, A) = (vu)^m v (uv)^{a_1} u (vu)^m v (uv)^{a_2} u \ldots = (vu)^\infty.$$

Also, if exactly one of w, u, v is nonempty, we have

$$\mathrm{ghpw}(w, \varepsilon, \varepsilon, A) = w^\infty;$$
$$\mathrm{ghpw}(\varepsilon, u, \varepsilon, A) = u^\infty;$$
$$\mathrm{ghpw}(\varepsilon, \varepsilon, v, A) = v^\infty.$$

We conclude that in all these cases the constructed word is periodic, which completes the ($\Leftarrow$) part.

To prove the converse, we assume that $\mathrm{ghpw}(w, u, v, A)$ is periodic. Then we can write

$$\mathrm{ghpw}(w, u, v, A) = w(uv)^{a_1 - 1}uvuwuv(uv)^{a_2 - 1}uw(uv)^{a_1}uw \cdots = s^{\infty},$$

where s can be chosen (long enough) such that $vuwuv \in \mathrm{Fact}(s)$. Assume that at least one of u and v is nonempty (there is nothing to prove if $u = v = \varepsilon$). Then we can choose i large enough such that $s \in \mathrm{Fact}((uv)^{a_i})$, which implies $vuwuv \in \mathrm{Fact}((uv)^{a_i})$. Recall (by Remark 4.3) that vu and uv appear in $(uv)^{a_i}$ only at the obvious positions. If $u \neq \varepsilon$, then the above gives

$$w = (vu)^m v \quad \text{for a nonnegative integer } m, \tag{2}$$

which was to be proved (note that $w = v = \varepsilon$ is a special case of this); if $u = \varepsilon$, then we conclude $w = v^l$ for a nonnegative integer l, that is, w is again of the form (2), or $w = u = \varepsilon$. This completes the proof. ∎

We shall now prove that each generalized highly potential word is either periodic, or recurrent but not uniformly recurrent. We first need the following assertion.

Proposition 4.6. $\mathrm{Fact}(\mathrm{ghpw}(w, u, v, A))$ *is closed under reversal.*

Proof. Let $x \in \mathrm{Fact}(\mathrm{ghpw}(w, u, v, A))$. Choose a large enough integer i such that $x \in \mathrm{Fact}(w_i)$. Since

$$w_{i+1} = w_i(uv)^{a_{i+1}}u\widetilde{w}_i,$$

we have

$$\widetilde{x} \in \mathrm{Fact}(\widetilde{w}_i) \subseteq \mathrm{Fact}(w_{i+1}) \subseteq \mathrm{Fact}(\mathrm{ghpw}(w, u, v, A)),$$

which was to be proved. ∎

Now, in view of Theorem 2.1, we have the following corollary.

Corollary 4.7. *Each generalized highly potential word is recurrent.*

Concerning uniform recurrence, we have:

Proposition 4.8. *A generalized highly potential word is uniformly recurrent if and only if it is periodic.*

Proof. The part ($\Leftarrow$) is clear by Theorem 2.2. Let us prove the other direction. Suppose the contrary: $\mathrm{ghpw}(w,u,v,A)$ is uniformly recurrent but not periodic. Since $vuwuv$ is a factor of $\mathrm{ghpw}(w,u,v,A)$, there exists a positive integer n such that $vuwuv$ is a factor of any factor of $\mathrm{ghpw}(w,u,v,A)$ of length n. Choose x such that $|x| = n$ and $x \in \mathrm{Fact}((uv)^{a_i})$ for some i. Now in the same manner as in the proof of Theorem 4.5 we get that either $w = (vu)^m v$ for a nonnegative integer m, or that exactly one of the words w, u and v is nonempty; in other words, $\mathrm{ghpw}(w,u,v,A)$ is periodic, which is a contradiction. ■

Finally, we have the following proposition.

Proposition 4.9. *If a generalized highly potential word is not periodic, then it is aperiodic.*

Proof. Suppose the contrary: $\mathrm{ghpw}(w,u,v,A)$ is eventually periodic but not periodic. By Corollary 4.7, it is recurrent, but then Theorem 2.3 implies that it must be periodic; contradiction. ■

4.4 Defect of generalized highly potential words

In this subsection we prove that the defect of a generalized highly potential word is always finite. Before proceeding to the main theorem, we need two technical lemmas.

Lemma 4.10. *Let a nonperiodic* $\mathrm{ghpw}(w,u,v,A)$ *be given in standard form, where* $vu \notin \mathrm{Pref}(wuv)$*. Assume that there exists an integer* i *such that* $i \geqslant 3$ *and*

$$|w_i|_{(uv)^{a_i}u} = 1. \tag{3}$$

Then

$$|w_{i+1}|_{w_i} = 2 \tag{4}$$

and

$$|w_{i+1}|_{(uv)^{a_{i+1}-1}u} = 2 + 2|w_i|_{(uv)^{a_{i+1}-1}u}.$$

Proof. Let us first prove (4). Write

$$w_{i+1} = w_i (uv)^{a_{i+1}} u w_i.$$

Clearly, $|w_{i+1}|_{w_i} \geqslant 2$. Suppose that there is a third copy of w_i in w_{i+1}. Note that $(uv)^{a_i}u$ occurs in the center of the considered copy of w_i, and we now conclude that this copy of $(uv)^{a_i}u$ must (partly) overlap the central copy of $(uv)^{a_{i+1}}u$ in w_{i+1} (otherwise we would have $|w_i|_{(uv)^{a_i}u} \geqslant 2$, which is impossible). Suppose that the length of the overlapping part is greater than or equal to $|uv|$. Then the overlapping part contains the factor vu or uv, and by Remark 4.3 this factor is positioned within the central copy of $(uv)^{a_{i+1}}u$ at one of the obvious positions. But this means that the central copy of $(uv)^{a_i}u$ in w_i must be preceded by uv or followed by vu, and since it is preceded by vuw and followed by wuv, we have a contradiction with $vu \notin \mathrm{Pref}(wuv)$ (or, which is the same, $uv \notin \mathrm{Suff}(vuw)$). Therefore, the overlapping part is shorter than $|uv|$. We may assume, without loss of generality, that the overlapping part and the considered copy of $(uv)^{a_i}u$ have a common endpoint (the other possibility: that they have a common starting point, is analogous). The part of the considered copy of $(uv)^{a_i}u$ that does not overlap presents a suffix of w_i. Suppose that its length is greater than or equal to $|vuw|$. Since $vuw \in \mathrm{Suff}(w_i)$, we have that vuw is a suffix of the considered part of $(uv)^{a_i}u$. But then Remark 4.3 gives that the beginning of that suffix vuw must coincide with an obvious position of vu within the considered copy of $(uv)^{a_i}u$; since that suffix vuw is followed by uv, altogether we obtain that wuv begins with vu, which is in contradiction with the lemma's assumption. Finally, we need to check the case when the length of the non-overlapping part of the considered copy of $(uv)^{a_i}u$ is less than $|vuw|$. But then we have $(uv)^{a_i}u \in \mathrm{Fact}(vuwuv)$, and since $vuwuv \in \mathrm{Fact}(w_{i-1})$ (recall that $i \geqslant 3$), we conclude $(uv)^{a_i}u \in \mathrm{Fact}(w_{i-1})$, which contradicts (3). This proves (4).

Let us now prove

$$|w_{i+1}|_{(uv)^{a_{i+1}-1}u} = 2 + 2|w_i|_{(uv)^{a_{i+1}-1}u}.$$

The inequality ($\geqslant$) is clear (there are two copies of $(uv)^{a_{i+1}-1}u$ in the center of w_{i+1}, plus the copies in the starting and the ending w_i). We are left to show only that there are no copies of $(uv)^{a_{i+1}-1}u$ in w_{i+1} that overlap the central copy of $(uv)^{a_{i+1}}u$ but are not encompassed within it. Suppose the contrary, that there exists such a copy. Suppose first that the overlapping part is of length $|uv|$ or more. Then the overlapping part contains the factor vu or uv, and Remark 4.3 gives that this factor is positioned within the central copies of $(uv)^{a_{i+1}}u$ at one of the obvious positions. But this means that the central copy of $(uv)^{a_{i+1}}u$ in w_{i+1} must be preceded by uv or followed by vu, and since it is preceded by vuw and followed by wuv, we have a contradiction with $vu \notin \mathrm{Pref}(wuv)$ (or $uv \notin \mathrm{Suff}(vuw)$). That leaves only the case when the overlapping part is of length less than $|uv|$, but this case also leads to a contradiction in completely the same manner as in the previous paragraph. This completes the proof of the lemma. ■

Lemma 4.11. *Let a nonperiodic* $\mathrm{ghpw}(w, u, v, A)$ *be given in standard form, where* $vu \notin \mathrm{Pref}(wuv)$*. Then there exists a positive integer* i *such that:*

1) $|w_i|_{(uv)^{a_i}u} = 1$ *(where that one copy of* $(uv)^{a_i}u$ *is in the center of* w_i*);*

2) $|w_{i+1}|_{w_i} = 2$ *(where those two copies of* w_i *are at the beginning and at the end of* w_{i+1}*);*

3) $|w_{i+1}|_{(uv)^{a_{i+1}-1}u} = 2 + 2|w_i|_{(uv)^{a_{i+1}-1}u}$ *(which equals either* 2 *or* 4*, depending on whether* $a_{i+1} - 1$ *is greater than* a_i *or equal to it, respectively).*

Furthermore, if i *is any number that satisfies 1), 2) and 3), then each number* k*,* $k \geqslant i$*, has the same properties.*

Note. Since Lemmas 4.10 and 4.11 seem to somewhat overlap and might confuse the reader, before we proceed to the proof, we shall say a few words on their mutual relationship (including a sketch of the proof of Lemma 4.11, in order to let the reader know what structure of the proof to expect, and what will be the role of Lemma 4.10 there).

The proof of Lemma 4.11 consists of two parts: in the first part we prove that such a number i exists, and then in the second part we prove the last sentence from the lemma's statement.

We do the first part by finding a number i, $i \geqslant 3$, that has the property 1). Lemma 4.10 then automatically implies that the same value i also has the properties 2) and 3), which finishes the first part of the proof.

We then proceed to the second part of the proof, which we do by induction. We assume that a number $k-1$ is given that has all the properties 1), 2) and 3), and then prove that the number k also has the properties 1), 2) and 3), which we show one by one (the inductive assumption stands for the whole time, assuming that $k-1$ has all three properties simultaneously). In this part of the proof we have to show all the three properties one by one, that is, we cannot only show the property 1) and then refer to Lemma 4.10 for 2) and 3) (as we do in the first part of the proof), because for that we would need the condition $k \geqslant 3$, which might not hold. However, the proofs for 2) and 3) here are not just repeating the proof of Lemma 4.10 all over (although some of the steps are indeed quite similar), since the assumptions are different: in Lemma 4.10 we proved that, if k had the property 1) and $k \geqslant 3$, then k also had the properties 2) and 3), while in this proof we do not have the inequality $k \geqslant 3$ anymore, but instead have the assumption that $k-1$ has all the properties 1), 2) and 3).

We hope that this additional explanation will clear any eventual confusion of the reader. We also add that Lemma 4.11 is a key result that we shall refer to multiple times in the article, while Lemma 4.10 is a technical device that we use in the proof of Lemma 4.11 (only in the first part of the proof), and after that we shall not refer to it anymore.

Proof. We first show the existence of such an integer i. Let i be such that $i \geqslant 3$ and $|(uv)^{a_i}u| \geqslant |vuwuv|$. Recall

$$w_i = w_{i-1}(uv)^{a_i}uw_{i-1}.$$

Let us show that this choice of i satisfies the properties 1), 2) and 3) from the statement. It is enough to show only the part 1), since then the parts 2) and 3) will follow by Lemma 4.10.

We show that the factor $(uv)^{a_i}u$ occurs exactly once in w_i. We clearly have one copy of it in the center, so we need to prove that there are no other copies. Suppose the contrary, that there is another copy. Assume first that that copy (partly) overlaps the central copy, and that, without loss of generality, it is positioned to the left of the central copy. We again have, as in the proof of Lemma 4.10, that the length of the overlapping part cannot be greater than or equal to $|uv|$; therefore, the overlapping part is shorter than $|uv|$. The part of the considered copy of $(uv)^{a_i}u$ that does not overlap presents a suffix of w_{i-1} and its length is greater than $|(uv)^{a_i-1}u|$, which is, by the choice of i, at least $|vuw|$. Since $vuw \in \mathrm{Suff}(w_{i-1})$, we have that vuw is a suffix of the considered part of $(uv)^{a_i}u$. But then Remark 4.3 gives that the beginning of that suffix vuw must coincide with an obvious position of vu within the considered copy of $(uv)^{a_i}u$ (the one that is not in the center of w_i); since that suffix vuw is followed by uv, altogether we obtain that wuv begins with vu, which is in contradiction with the lemma's assumption. Therefore, we are left to analyze only the case when there is no overlap, that is, $(uv)^{a_i}u \in \mathrm{Fact}(w_{i-1})$. But this implies in the same way $(uv)^{a_i}u \in \mathrm{Fact}(w_{i-2})$, then $(uv)^{a_i}u \in \mathrm{Fact}(w_{i-3})$ etc., which is clearly a contradiction. This proves that the chosen value of i indeed satisfies the properties 1), 2) and 3).

Let us now prove the last sentence from the lemma's statement. Assume that i is given as required. We proceed by induction on k. We have the base for $k = i$. Now assume that the assertion holds for $k-1$.

We first prove

$$|w_k|_{(uv)^{a_k}u} = 1. \tag{5}$$

Recall

$$w_k = w_{k-1}(uv)^{a_k}uw_{k-1}.$$

Clearly, $|w_k|_{(uv)^{a_k}u} \geqslant 1$. Suppose that there is another copy of $(uv)^{a_k}u$ (besides the one in the center) in w_k. It cannot overlap the central copy of $(uv)^{a_{k+1}}u$ for a length of $|uv|$ nor more (as we have already seen several times); therefore, the length of the overlap is less than $|uv|$. Then there is a copy of $(uv)^{a_k-1}u$ in w_{k-1}. Note that this copy contains the factor $(uv)^{a_{k-1}}u$, and by the inductive assumption, we have that the only copy of this factor in w_{k-1} is the one in the center; however, it is followed by wuv and preceded by vuw (or only w both times, in the special case $k = 2$), and now because of the assumption $vu \notin \mathrm{Pref}(wuv)$ (then also $uv \notin \mathrm{Suff}(vuw)$) we get that this copy of $(uv)^{a_{k-1}}u$ in w_{k-1} cannot be a part of the considered copy of $(uv)^{a_k}u$ in w_k, a contradiction. This proves (5).

Let us now prove

$$|w_{k+1}|_{w_k} = 2. \tag{6}$$

(Note that we cannot simply use Lemma 4.10 here, since the inequality $k \geqslant 3$ might not hold.) Recall

$$w_{k+1} = w_k(uv)^{a_{k+1}}uw_k.$$

Clearly, $|w_{k+1}|_{w_k} \geqslant 2$. Suppose that there is a third copy of w_k in w_{k+1}. Note that $(uv)^{a_k}u$ occurs in the center of the considered copy of w_k, and we now conclude that this copy of $(uv)^{a_k}u$ must (partly) overlap the central copy of $(uv)^{a_{k+1}}u$ in w_{k+1} (otherwise we would have $|w_k|_{(uv)^{a_k}u} \geqslant 2$, which is impossible), and, by the argument that we have already seen, the length of the overlapping part cannot be greater than or equal to $|uv|$. But then, since the considered copy of $(uv)^{a_k}u$ is both preceded by and followed by w_{k-1}, we get that one of those two copies of w_{k-1} is encompassed inside w_k, neither at its beginning nor at its end. But this implies $|w_k|_{w_{k-1}} \geqslant 3$, which contradicts the inductive assumption. This proves (6).

Finally, we need to prove

$$|w_{k+1}|_{(uv)^{a_{k+1}-1}u} = 2 + 2|w_k|_{(uv)^{a_{k+1}-1}u}.$$

We again assume that there is a "superfluous" copy of $(uv)^{a_{k+1}-1}u$ in w_{k+1}. Then (again) it overlaps the central copy of $(uv)^{a_{k+1}}u$ for a length of less than $|uv|$, which means that the non-overlapping part of the considered copy of $(uv)^{a_{k+1}-1}u$ is of length at least $|(uv)^{a_{k+1}-2}u|$, which is at least $|(uv)^{a_k-1}u|$; however, this non-overlapping part is encompassed within w_k and we see that it contains a copy of $(uv)^{a_k-1}u$ that is "superfluous" in w_k, in contrary to the inductive assumption. This completes the proof of the lemma. ■

We are now ready for the main theorem of this subsection (and arguably of the whole article).

Theorem 4.12. *Let* $\mathrm{ghpw}(w, u, v, A)$ *be a generalized highly potential word. We have*

$$D(\mathrm{ghpw}(w, u, v, A)) < \infty.$$

Proof. If $\operatorname{ghpw}(w, u, v, A)$ is periodic, the proof will given in the next subsection. Therefore, assume that $\operatorname{ghpw}(w, u, v, A)$ is not periodic, and we may further assume, without loss of generality, that it is given in standard form. We shall first work under the assumption $vu \notin \operatorname{Pref}(wuv)$ (and then also $uv \notin \operatorname{Suff}(vuw)$), and then return to the general case at the end of the proof.

Let i be a number whose existence is guaranteed by Lemma 4.11. It is enough to prove that for each k, $k \geqslant i + 1$, we have

$$D(w_k) = D(w_{i+1}). \tag{7}$$

Indeed, in that case we would have, by Theorem 2.6 and the equality (1),

$$D(\operatorname{ghpw}(w, u, v, A)) = \sup_{z \in \operatorname{Fact}(\operatorname{ghpw}(w,u,v,A))} D(z) = \sup_{j \in \mathbb{N}_0} D(w_j) = D(w_{i+1}),$$

as needed.

In order to show (7), it is enough to prove only

$$D(w_{i+2}) = D(w_{i+1}); \tag{8}$$

indeed, in that case, by then choosing $i + 1$ instead of i (note that, by the second part of Lemma 4.11, $i + 1$ indeed satisfies the same requirements as i does), we would get $D(w_{i+3}) = D(w_{i+2})$ in the same way, then $D(w_{i+4}) = D(w_{i+3})$ etc., which gives (7).

Therefore, in order to prove (8), our goal is to find $|w_{i+2}| - |w_{i+1}|$ palindromes in w_{i+2} that do not occur in w_{i+1}. Since $|w_{i+2}| - |w_{i+1}| = |w_{i+1}| + a_{i+2}|uv| + |u| + |w_{i+1}| - |w_{i+1}| = |w_{i+1}| + a_{i+2}|uv| + |u|$, we need to find $|w_{i+1}| + a_{i+2}|uv| + |u|$ new palindromes. (Note: by our construction of the required number of palindromes, it will not be obvious that our list contains *all* the new palindromes. But this is not relevant: it is enough to find at least the required number of new palindromes, and Theorem 2.6 then implies that there cannot be more of them.) We distinguish four types of palindromes (after defining each type, we first explain why that type is disjoint from all the types before it; these explanations are marked by the symbol "$\lhd$").

- We first enumerate new palindromes that have the factor $(uv)^{a_{i+2}}u$ in the center; they can be obtained by "expanding" (to the left and the right side) the boxed part below:

  $$w_{i+1}\boxed{(uv)^{a_{i+2}}u}w_{i+1}.$$

  Clearly, there is a total of $|w_{i+1}|$ palindromes of this type (not counting the palindrome $(uv)^{a_{i+2}}u$ itself), and all of them must be new because $(uv)^{a_{i+2}}u$ occurs only once in w_{i+2} (and does not occur in w_{i+1}).

- We now enumerate new palindromes that have the factor w_i in the center; they can be obtained by "expanding" the boxed part below:

$$w_{i+1}(uv)^{a_{i+2}}u\boxed{w_i}(uv)^{a_{i+1}}uw_i.$$

  ◁ This type is disjoint from the first type since, by the property 1) from Lemma 4.11, there is only one copy of $(uv)^{a_{i+2}}u$ in w_{i+2}, which cannot be in the center of a palindrome of this type.

  Since there are only two occurrences of w_i in w_{i+1} (by the choice of i: Lemma 4.11, property 2)), we get that all the palindromes $\widetilde{x}w_ix$ for $x \in \mathrm{Pref}((uv)^{a_{i+1}}u) \setminus \{\varepsilon\}$ are new. But there may be even more new palindromes. Note that w_i begins with wuv. If

$$t = \max\{|p| : p \in \mathrm{Pref}(wuv) \cap \mathrm{Pref}(vu)\},$$

  then the expanding can continue further for t more new palindromes (and since $vu \notin \mathrm{Pref}(wuv)$, this is the best we can do). In total, we have

$$a_{i+1}|uv| + |u| + t$$

  new palindromes of this type.

- We now enumerate new palindromes that have the factor $(uv)^{a_{i+2}-1}u$ in the center; they can be obtained by "expanding" the boxed part below:

$$w_{i+1}uv\boxed{(uv)^{a_{i+2}-1}u}w_{i+1}.$$

  ◁ This type is disjoint from both the previous types since, by the property 3) from Lemma 4.11, there are either 2 or 4 copies of $(uv)^{a_{i+2}-1}u$ in w_{i+2}, and none of them can be in the center of a palindrome of one of the previous two types.

  Note: these palindromes are new only if $a_{i+2} > a_{i+1} + 1$, and we shall do the counting under this assumption (otherwise, all of them would be factors of w_{i+1}). Since w_{i+1} begins with wuv, we have a total of t new palindromes here.

- Finally, we enumerate new palindromes that are factors of $(uv)^{a_{i+2}}u$.

  ◁ This type is disjoint from the first type since each palindrome of the first type is longer than each palindrome of this type.

- ◁ Let us prove that this type is disjoint from the second type. Suppose the contrary: there is a palindrome p that is of both the second and the fourth type. We shall soon see that all the palindromes of the fourth type are of length strictly greater than $|(uv)^{a_{i+1}-1}u| + 2t$, and thus strictly greater than $|(uv)^{a_i}u| + 2t$. Since p is of the second type, p has $(uv)^{a_i}u$ in the center (because w_i has $(uv)^{a_i}u$ in the center). The letter at $t+1$ positions right of this copy of $(uv)^{a_i}u$ in p is $(wuv)[t+1]$, which is different from $(vu)[t+1]$ by the definition of t. However, by Remark 4.3 and the fact that $p \in \mathrm{Fact}((uv)^{a_{i+2}}u)$ (because p is of the fourth type also), we have that the observed copy of $(uv)^{a_i}u$ in p has to be followed by vu, a contradiction. This proves the claim.
- ◁ Finally, we show that this type is disjoint from the third type. As we shall see in a moment, this type will be divided into two subtypes ((9) and (10) below). The first subtype is disjoint from the third type since there are either 2 or 4 copies of $(uv)^{a_{i+2}-1}u$ in w_{i+2}, none of which is in the center of $(uv)^{a_{i+2}}u$, while all the palindromes of the first subtype are in the center of $(uv)^{a_{i+2}}u$, and therefore they cannot have a copy of $(uv)^{a_{i+2}-1}u$ in their center. The second subtype is disjoint from the third type since each palindrome of the third type is longer than each palindrome of the second subtype.

Let us first consider palindromes of the form

$$((uv)^{a_{i+2}}u)[j, |(uv)^{a_{i+2}}u| - j + 1]. \tag{9}$$

In other words, they can be obtained by removing one by one letter from both ends of $(uv)^{a_{i+2}}u$ simultaneously. At one moment, we shall arrive to $(uv)^{a_{i+1}}u$ or $(uv)^{a_{i+1}-1}u$ (depending on whether a_{i+2} and a_{i+1} are of the same parity or not, respectively). Assume, e.g., the second case (the first one is similar but even easier). At this moment, the palindrome that we arrive to belongs to $\mathrm{Fact}(w_{i+1})$, so there is no point to continue further. We shall now check how many of all those palindromes exist already in $\mathrm{Fact}(w_{i+1})$. Note that, by the choice of i (in particular, the property 3) from Lemma 4.11), we know exact positions of all the copies of $(uv)^{a_{i+1}-1}u$ within w_{i+1}: there are two copies that are parts of the central $(uv)^{a_{i+1}}u$, and additionally, if they exist (that is, if $a_{i+1}-1 = a_i$), two copies in the centers of the starting and ending w_i. Since each of these copies is either preceded by vuw or followed by wuv (or both), it is easy to see that the number of the considered palindromes that belong to $\mathrm{Fact}(w_{i+1})$ is precisely t. The same conclusion holds if a_{i+2} and a_{i+1} are of the same parity. This means that so far we have enumerated $\lceil \frac{a_{i+2}-a_{i+1}}{2} \rceil |uv| - t$ new palindromes.

We now consider palindromes of the form

$$((uv)^{a_{i+2}-1}u)[j, |(uv)^{a_{i+2}-1}u| - j + 1]. \tag{10}$$

We again remove one by one letter from both ends of $(uv)^{a_{i+2}-1}u$ simultaneously until we reach $(uv)^{a_{i+1}}u$ or $(uv)^{a_{i+1}-1}u$. The same argument as in the previous paragraph shows that there are $\lfloor \frac{a_{i+2}-a_{i+1}}{2} \rfloor |uv| - t$ new palindromes here, but there is one exceptional case: namely, if $a_{i+2} = a_{i+1} + 1$, then already the starting palindrome $(uv)^{a_{i+2}-1}u$ belongs to $\mathrm{Fact}(w_{i+1})$, and thus then we get 0 new palindromes (the formula above would give a senseless value of $-t$, the explanation of which is that the subtracted t palindromes in this exceptional case are not factors of $(uv)^{a_{i+2}-1}u$, and thus we do not need to subtract them).

Since $\lceil \frac{a_{i+2}-a_{i+1}}{2} \rceil + \lfloor \frac{a_{i+2}-a_{i+1}}{2} \rfloor = a_{i+2} - a_{i+1}$, we may conclude that, altogether, there is a total of

$$(a_{i+2} - a_{i+1})|uv| - 2t$$

new palindromes of this type if $a_{i+2} > a_{i+1} + 1$, and

$$|uv| - t$$

if $a_{i+2} = a_{i+1} + 1$.

Finally, let us sum all the numbers. If $a_{i+2} = a_{i+1} + 1$, then we have found

$$\begin{aligned} |w_{i+1}| + (a_{i+1}|uv| + |u| + t) + (|uv| - t) &= |w_{i+1}| + (a_{i+1} + 1)|uv| + |u| \\ &= |w_{i+1}| + a_{i+2}|uv| + |u| \end{aligned}$$

new palindromes (recall that we ignore the third bullet here); if $a_{i+2} > a_{i+1} + 1$, then we have found

$$|w_{i+1}| + (a_{i+1}|uv| + |u| + t) + t + ((a_{i+2} - a_{i+1})|uv| - 2t) = |w_{i+1}| + |u| + a_{i+2}|uv|$$

new palindromes. In both cases, we get what was to be proved.

We now need to address the case $vu \in \mathrm{Pref}(wuv)$. Let

$$s = \min\{j : (wuv)[j] \neq (vu)^{\infty}[j]\} - |uv|$$

(such a number s must exist since otherwise Remark 4.3 would imply that w is of the form $vuvu \ldots vuv$ and thus $\mathrm{ghpw}(w, u, v, A)$ would be periodic).

Note that the assumption $vu \in \mathrm{Pref}(wuv)$ implies that s is positive. We also show that $s \leqslant \left\lfloor \frac{|w|}{2} \right\rfloor$: indeed, if this were not the case, then the word $vuwuv$ would be a palindromic word that would match $(vu)^\infty$ for the first $|vu| + \left\lfloor \frac{|w|}{2} \right\rfloor + |uv|$ (which is $\left\lfloor \frac{|vuwuv|}{2} \right\rfloor + |uv|$) letters, and then Lemma 4.4 would imply $vuwuv = (vu)^m v$; therefore, w would also be of the form $vuvu \ldots vuv$, which would contradict the fact that $\mathrm{ghpw}(w, u, v, A)$ is not periodic. Now, let:

$$\begin{aligned} e &= (|u| + 2s) \bmod |uv|; \\ l &= (-s) \bmod |uv|; \\ w' &= w[s+1, |w| - s]; \\ u' &= (uvuv)[l+1, l+e]; \\ v' &= (uvuv)[l+e+1, l+|uv|]; \\ A' &= \left(a_i + \frac{|u| + 2s - e}{|uv|} \right)_{i=1}^{\infty} = (a_i')_{i=1}^{\infty}. \end{aligned} \tag{11}$$

Notice that $u'v'$ is a conjugate of uv, and we have

$$(u'v')^\infty = (uv)[l+1, |uv|](uv)^\infty = (u'v')^\infty[1, |uv| - l](uv)^\infty = (u'v')^\infty[1, s](uv)^\infty, \tag{12}$$

and also

$$\begin{aligned} \widetilde{(u'v')^\infty[1, s]} &= \widetilde{(uv)[l+1, |uv|](uv)^{\left\lfloor \frac{s}{|uv|} \right\rfloor}} \\ &= (vu)^{\left\lfloor \frac{s}{|uv|} \right\rfloor}(vu)[1, |uv| - l] = (vu)^\infty[1, s]. \end{aligned}$$

Further, by the definition of s, we have

$$w[1, s] = (vu)^\infty[1, s] = \widetilde{(u'v')^\infty[1, s]}.$$

We claim:

$$\mathrm{ghpw}(w, u, v, A) = w[1, s]\, \mathrm{ghpw}(w', u', v', A'). \tag{13}$$

It is enough to prove that for each i we have

$$w[1, s] w_i' = w_i[1, |w_i| - s]. \tag{14}$$

Before we proceed, we shall first check what we get of the word $(u'v')^{a_i'}u'$ when we erase the prefix and the suffix of the length s. By (12), we notice that, after erasing the prefix, there remains a word of the form $uvuvuv \ldots$. Therefore, since $(u'v')^{a_i'}u'$ is a palidrome, erasing both the prefix and the suffix leaves a word of the form $(uv)^k u$ for a nonnegative integer k. We have $k|uv| + |u| + 2s = a_i'|uv| + |u'|$, which reduces to

$$k = a_i' + \frac{|u'| - |u| - 2s}{|uv|} = a_i + \frac{|u| + 2s - e}{|uv|} + \frac{e - |u| - 2s}{|uv|} = a_i.$$

The proof of (14) is now a straightforward induction: the base (for $i = 0$) is clear, and for the induction step we have:

$$\begin{aligned} w[1,s]w_{i+1}' &= w[1,s]w_i'(u'v')^{a_{i+1}'}u'w_i' \\ &= w_i[1,|w_i|-s](u'v')^{\infty}[1,s](uv)^{a_{i+1}}\overbrace{u(u'v')^{\infty}[1,s]}w_i' \\ &= w_i(uv)^{a_{i+1}}uw[1,s]w_i' = w_i(uv)^{a_{i+1}}uw_i[1,|w_i|-s] \\ &= w_{i+1}[1,|w_{i+1}|-s], \end{aligned}$$

which was to be proved.

Now notice the following: w' is a palindrome (by its definition), $u'v'$ is primitive (since it is a conjugate of uv, which is primitive), and $v'u' \notin \mathrm{Pref}(w'u'v')$ (because of $(v'u')[|v'u'|] = (u'v')[1] = (uv)[l+1] = (vu)^{\infty}[s] = (vu)^{\infty}[s + |uv|]$ and $(w'u'v')[|v'u'|] = (wuv)[s + |uv|]$, and these two letters are different by the choice of s). Therefore, the word $\mathrm{ghpw}(w', u', v', A')$ satisfies all the assumptions of the first part of the proof, and we conclude that its defect is finite. Now, since $\mathrm{ghpw}(w, u, v, A)$ is recurrent, by (13) we conclude that each its factor is also a factor of $\mathrm{ghpw}(w', u', v', A')$ (and the other direction obviously holds, too), which finally implies:

$$D(\mathrm{ghpw}(w, u, v, A)) = D(\mathrm{ghpw}(w', u', v', A')) < \infty.$$

The proof is completed. ■

Since, as mentioned in the Introduction, infinite words of defect 0 have been studied significantly more than infinite words of finite nonzero defect, it makes sense to give a characterization of generalized highly potential words of (non)zero defect. Such a characterization can easily be inferred from the proof of Theorem 4.12. We give it in the following corollary (the corollary assumes that a word is given in standard form, but if it is not, we can always rechoose the defining parameters as in the proof of Lemma 4.2 and make it in standard form).

Corollary 4.13. *Given a nonperiodic* $\mathrm{ghpw}(w, u, v, A)$ *in standard form, we have:*

1° *If* $vu \notin \mathrm{Pref}(wuv)$*, we choose the smallest integer* i *that satisfies 1), 2) and 3) from the statement of Lemma 4.11, and then* $D(\mathrm{ghpw}(w, u, v, A)) = D(w_{i+1})$.

2° *If* $vu \in \mathrm{Pref}(wuv)$*, we choose* w'*,* u'*,* v' *and* A' *as in (11), and then* $D(\mathrm{ghpw}(w, u, v, A)) = D(\mathrm{ghpw}(w', u', v', A'))$*, which is evaluated as in* 1° *above.*

In particular, in this way we can determine whether $D(\mathrm{ghpw}(w, u, v, A))$ *is* 0 *or positive, which gives a characterization of generalized highly potential words of (non)zero defect.*

Also note, we can easily produce many examples of generalized highly potential words of nonzero defect. The simplest way is just to take any of the words w, u or v to have nonzero defect, and then $\mathrm{ghpw}(w, u, v, A)$ also has nonzero defect. This is a sufficient, but not necessary condition: for example, if w, u and v are all rich words, but such that one of the words wu or uv has positive defect, then $\mathrm{ghpw}(w, u, v, A)$ again has positive defect.

In fact, if any two of the words w, u and v are such that they cannot be factors of the same rich word, then $\mathrm{ghpw}(w, u, v, A)$ has positive defect. It was an open question posed in [20] if it is decidable whether two rich words can be factors of the same rich word; this question has been settled (in the affirmative) very recently [23], though the deciding algorithm is not really practical. An elegant necessary condition for two rich words to be factors of the same rich words is as follows [10, Theorem 6]: no two different factors of a rich word can have the same longest palindromic prefix as well as the same longest palindromic suffix (therefore, if we want $\mathrm{ghpw}(w, u, v, A)$ to have positive defect, it is sufficient to have the stated condition violated for any two of the words w, u, v). It has been asked in [24, Open problem 6.2] whether the stated condition is also sufficient (that is, whether any two rich words that have different longest palindromic prefix or longest palindromic suffix must be factors of the same rich word); if true, this would greatly simplify the mentioned algorithm from [23], but up to the present authors' knowledge, this problem is still open.

4.5 Periodic case

Finally, we show that periodic generalized highly potential words also have finite defect.

Theorem 4.14. *The defect of a periodic generalized highly potential word is finite.*

Proof. Let $\mathrm{ghpw}(w, u, v, A)$ be given in standard form, and let $\mathrm{ghpw}(w, u, v, A) = p^\infty$. We claim that we may assume that p is a primitive word that is a product of two palindromes (where one of them is possibly ε). Indeed, Theorem

4.5 implies that we may assume either $p = vu$ or $p \in \{w, u, v\}$, but in the latter case, if p is not primitive but, say, $p = t^n$, we may take t in place of p (t is then a palindrome since it is both a prefix and a suffix of a palindrome). Now Theorem 2.7 gives $D(\mathrm{ghpw}(w, u, v, A)) < \infty$. ■

Acknowledgments

The authors would like to thank the anonymous reviewers for their devoted time and a number of suggestions that helped to improve the content of the article.

The first, the third and the fourth author were supported by the project 174018 of the Ministry of Education, Science and Technological Development of Serbia, and the second author by the project 174006 of the same Ministry.

References

[1] J.-P. Allouche & J. Shallit, *Automatic Sequences. Theory, Applications, Generalizations*, Cambridge University Press, 2003.

[2] L'. Balková & E. Pelantová & Š. Starosta, Infinite words with finite defect, *Adv. in Appl. Math.* **47** (2011), 562–574.

[3] L'. Balková & E. Pelantová & Š. Starosta, On Brlek-Reutenauer conjecture, *Theoret. Comput. Sci.* **412** (2011), 5649–5655; Corrigendum: *Theoret. Comput. Sci.* **465** (2012), 73–74.

[4] B. Bašić, A note on the paper "On Brlek-Reutenauer conjecture," *Theoret. Comput. Sci.* **160** (2012), 181–186.

[5] B. Bašić, On highly potential words, *European J. Combin.* **34** (2013), 1028–1039.

[6] A. Blondin-Massé & S. Brlek & A. Garon & S. Labbé, Combinatorial properties of f-palindromes in the Thue-Morse sequence, *Pure Math. Appl.* **19 (2-3)** (2008), 39–52.

[7] S. Brlek & S. Hamel & M. Nivat & C. Reutenauer, On the palindromic complexity of infinite words, *Internat. J. Found. Comput. Sci.* **15** (2004), 293–306.

[8] S. Brlek & C. Reutenauer, Complexity and palindromic defect of infinite words, *Theoret. Comput. Sci.* **412** (2011), 493–497.

[9] M. Bucci & A. De Luca & A. Glen & L. Q. Zamboni, A connection between palindromic and factor complexity using return words, *Adv. in Appl. Math.* **42** (2009), 60–74.

[10] M. Bucci & A. De Luca & A. Glen & L. Q. Zamboni, A new characteristic property of rich words, *Theoret. Comput. Sci.* **410** (2009), 2860–2863.

[11] M. Bucci & E. Vaslet, Palindromic defect of pure morphic aperiodic words, in: *14th Mons Days of Theoretical Computer Science*, Louvain-la-Neuve, Belgium, 2012.

[12] J. Cassaigne, Complexité et facteurs spéciaux, *Bull. Belg. Math. Soc. Simon Stevin* **4** (1997), 67–88.

[13] X. Droubay & J. Justin & G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* **255** (2001), 539–553.

[14] A. Glen & J. Justin & S. Widmer & L. Q. Zamboni, Palindromic richness, *European J. Combin.* **30** (2009), 510–531.

[15] C. Guo & J. Shallit & A. M. Shur, Palindromic rich words and run-length encodings, *Information Processing Letters* **116** (2016), 735–738.

[16] M. Lothaire, *Combinatrics on words*, Addison-Wesley, 1983.

[17] A. de Luca & A. Glen & L. Zamboni, Rich, Sturmian and trapezoidal words, *Theoret. Comput. Sci.* **407** (2008), 569–573.

[18] E. Pelantová & Š. Starosta, Almost rich words as morphic images of rich words, *Internat. J. Found. Comput. Sci.* **23** (2012), 1067–1083.

[19] E. Pelantová & Š. Starosta, Languages invariant under more symmetries: overlapping factors versus palindromic richness, *Discrete Math.* **313** (2013), 2432–2445.

[20] E. Pelantová & Š. Starosta, On words with the zero palindromic defect, in: *Combinatorics on Words, 11th International Conference (WORDS 2017)*, Montréal, QC, Canada, 2017, pp. 203–213.

[21] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics,* Springer-Verlag, Berlin, 2002.

[22] J. Rukavicka, On the number of rich words, in: *Developments in Language Theory, 21st International Conference (DLT 2017)*, Liège, Belgium, 2017, pp. 345–352.

[23] J. Rukavicka, Rich words containing two given factors, in: *Combinatorics on Words, 12th International Conference (WORDS 2019)*, Loughborough, UK, 2019, pp. 286–298.

[24] J. Vesti, Extensions of rich words, *Theoret. Comput. Sci.* **548** (2014), 14–24.

[25] J. Vesti, Rich square-free words, *Theoret. Comput. Sci.* **687** (2017), 48–61.