Antisquares and Critical Exponents

Aseem Baranwal, James Currie, Lucas Mol, Pascal Ochem, Narad Rampersad, Jeffrey Shallit

Discrete Mathematics and Theoretical Computer Science ....
.... **vol. 25:2 #11 (2023)**

Antisquares and Critical Exponents

Aseem Baranwal[1] James Currie[2*] Lucas Mol[3] Pascal Ochem[4]
Narad Rampersad[2†] Jeffrey Shallit[1‡]

[1] *University of Waterloo, Canada*
[2] *University of Winnipeg, Canada*
[3] *Thompson Rivers University, Canada*
[4] *LIRMM, CNRS, Université de Montpellier, France*

revisions 20th Sep. 2022, 27th July 2023; accepted 2nd Aug. 2023.

The (bitwise) complement $\overline{x}$ of a binary word x is obtained by changing each 0 in x to 1 and vice versa. An *antisquare* is a nonempty word of the form $x\,\overline{x}$. In this paper, we study infinite binary words that do not contain arbitrarily large antisquares. For example, we show that the repetition threshold for the language of infinite binary words containing exactly two distinct antisquares is $(5+\sqrt{5})/2$. We also study repetition thresholds for related classes, where "two" in the previous sentence is replaced by a larger number.

We say a binary word is *good* if the only antisquares it contains are 01 and 10. We characterize the minimal antisquares, that is, those words that are antisquares but all proper factors are good. We determine the growth rate of the number of good words of length n and determine the repetition threshold between polynomial and exponential growth for the number of good words.

Keywords: antisquare, critical exponent, binary complement, binary word, avoidability, repetition threshold, enumeration, minimal forbidden word

1 Introduction

Let x be a finite nonempty binary word. We say that x is an *antisquare* if there exists a word y such that $x = y\,\overline{y}$, where the overline denotes a morphism that maps $0 \to 1$ and $1 \to 0$. For example, 011100 is an antisquare. The *order* of an antisquare $y\,\overline{y}$ is defined to be $|y|$, where $|y|$ denotes the length of y.

Avoidance of antisquares has been studied previously in combinatorics on words. For example, Mousavi et al. (2016) proved that the infinite Fibonacci word

$$\mathbf{f} = 01001010\cdots,$$

the fixed point of the morphism $0 \to 01$, $1 \to 0$, has exactly four antisquare factors, namely, $01, 10, 1001$, and 10100101. More generally, all the antisquares in Sturmian words have recently been characterized

*The research of JC is supported by NSERC Grant 2017-03901.
†The research of NR is supported by NSERC Grant 2019-04111.
‡The research of JS is supported by NSERC Grant 2018-04118.

ISSN 1365–8050

in Hieronymi et al. (2022). On the other hand, Ng et al. (2019) classified those infinite binary words containing the minimum possible numbers of distinct squares and antisquares.

It is easy to see that no infinite binary word, except the trivial families given by $(0+\epsilon)1^\omega$ and $(1+\epsilon)0^\omega$, can contain at most one distinct antisquare. (Here the notation x^ω refers to the right-infinite word $xxx\cdots$.) However, once we move to two distinct antisquares, the situation is quite different. We have the following:

Proposition 1. *There are exponentially many finite binary words of length n having at most two distinct antisquares, and there are uncountably many infinite binary words with the same property.*

Proof: It is easy to see that every binary word in $\{1000, 10000\}^*$ has only the antisquares 01 and 10, which proves the first claim.

For the second, consider the uncountable set of infinite words $\{1000, 10000\}^\omega$. (Here, by S^ω for a set S of nonempty finite words, we mean the set of all infinite words arising from concatenations of elements of S.) $\square$

Furthermore, it is easy to see that if an infinite binary word, other than 001^ω and its complement, contains exactly two antisquares, then these antisquares must be 01 and 10. Call a binary word *good* if it contains no antisquare factors, except possibly 01 and 10. This suggests studying the following problem.

Problem 2. Find the repetition threshold for good words.

The repetition threshold for a class of (finite or infinite) words is defined as follows. First, we say that a finite word $w = w[1..n]$ has *period* $p \geq 1$ if $w[i] = w[i+p]$ for $1 \leq i \leq n-p$. The smallest period of a word w is called *the* period, and we write it as $\mathrm{per}(w)$. The *exponent* of a finite word w, written $\exp(w)$ is defined to be $|w|/\mathrm{per}(w)$. We say a word (finite or infinite) is α*-free* if the exponents of its nonempty factors are all $< \alpha$. We say a word is α^+*-free* if the exponents of its nonempty factors are all $\leq \alpha$. The *critical exponent* of a finite or infinite word x is the supremum, over all nonempty finite factors w of x, of $\exp(w)$; it is written $\mathrm{ce}(x)$. Finally, the *repetition threshold* for a language L of infinite words is defined to be the infimum, over all $x \in L$, of $\mathrm{ce}(x)$.

The critical exponent of a word can be either rational or irrational. If it is rational, then it can either be attained by a particular finite factor, or not attained. For example, the critical exponent of both

- the Thue-Morse word $\mathbf{t} = 0110100110010110100101100\cdots$, fixed point of the morphism $0 \to 01$, $1 \to 10$; and
- the variant Thue-Morse word $\mathbf{vtm} = 2102012101202102012021012\cdots$, fixed point of the morphism $2 \to 210$, $1 \to 20$, $0 \to 1$

is 2, but it is attained in the former case and not attained in the latter. If the critical exponent α is attained, we typically write it as α^+.

In 1972, Dejean (1972) wrote a classic paper on combinatorics on words, where she determined the repetition threshold for the language of all infinite words over $\{0, 1, 2\}$—it is $\frac{7}{4}^+$—and conjectured the value of the repetition threshold for the languages Σ_k^* for $k \geq 4$, where $\Sigma_k = \{0, 1, \ldots, k-1\}$. Dejean's conjecture was only completely resolved in 2011, in Rao (2011) and Currie and Rampersad (2011), independently.

The repetition threshold has been studied for many classes of words. To name a few, there are the

- Sturmian words, studied in (Carpi and Luca, 2000, Prop. 15);

- palindromes, studied in Shallit (2016);
- rich words, studied in Currie et al. (2020);
- balanced words, studied in Rampersad et al. (2019) Dvořáková et al. (2022); and
- complementary symmetric Rote words, studied in Dvořáková et al. (2020).

For variations on and generalizations of repetition threshold, see Ilie et al. (2005); Badkobeh and Crochemore (2011); Fiorenzi et al. (2011); Samsonov and Shur (2012); Mousavi and Shallit (2013).

The goal of this paper is to study the repetition threshold for two classes of infinite words:

- AO_ℓ, the binary words avoiding all antisquares of order $\geq \ell$;
- AN_n, the binary words with no more than n antisquares.

It turns out that there is an interesting and subtle hierarchy, depending on the values of ℓ and n.

Our work is very similar in flavor to that of Shallit (2004), which found a similar hierarchy concerning critical exponents and sizes of squares avoided. The hierarchy for antisquares, as we will see, however, is significantly more complex.

In this paper, in Sections 2 and 3, we solve Problem 2, and show that the repetition threshold for good words is $2 + \alpha$, where $\alpha = (1 + \sqrt{5})/2$ is the golden ratio.

Proving that the repetition threshold for a class of infinite words equals some real number β generally consists of two parts: first, an explicit construction of a word avoiding β^+ powers. This is often carried out by finding an appropriate morphism h whose infinite fixed point $\mathbf{x}$ (or an image of $\mathbf{x}$ under a second morphism) has the desired property. Second, if β is rational, then one can prove there is no infinite word avoiding β-powers by a breadth-first or depth-first search of the infinite tree of all words. If β is irrational, however, one must generally be more clever.

In Section 4 we determine the repetition threshold for binary words avoiding all antisquares of order $\geq \ell$, and in Section 5 we determine the repetition threshold for binary words with no more than n antisquares. In Section 6 we completely characterize the minimal antisquares; i.e., the binary words that are antisquares but have the property that all proper factors are good. This characterization is then used in Section 7, where we determine the growth rate of the number of good words of length n. In this section we also show that the repetition threshold between polynomial and exponential growth for good words avoiding α-powers is $\alpha = \frac{15}{4}$; i.e., there are exponentially many such words avoiding $\frac{15}{4}^+$-powers, but only polynomially many that avoid $\frac{15}{4}$-powers.

2 A good infinite word with critical exponent $2 + \alpha$

Consider the morphisms below:

$$\varphi: \begin{array}{l} 0 \mapsto 001 \\ 1 \mapsto 01 \end{array} \qquad g: \begin{array}{l} 0 \mapsto 01 \\ 1 \mapsto 11. \end{array}$$

Let us write $\varphi^\omega(0)$ for the (unique) infinite fixed point of φ that starts with 0. We claim that the infinite word $\mathbf{w} = g(\varphi^\omega(0))$ does not have antisquares other than 01 and 10, and has critical exponent $2 + \alpha$, where $\alpha = (1 + \sqrt{5})/2$. The infinite word $\mathbf{w}$ is Fibonacci-automatic, in the sense of Mousavi et al.

(2016), so we can apply the `Walnut` theorem-prover Mousavi (2016) to establish this claim. For more about `Walnut`, see Shallit (2022).

We start with the Fibonacci automaton for $\varphi^\omega(0)$, as displayed in Figure 1.

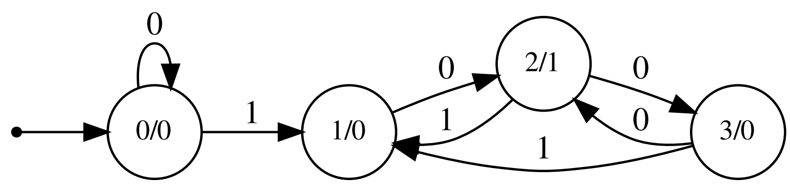

Fig. 1: Fibonacci automaton for $\varphi^\omega(0)$.

Let us name the above automaton `FF.txt`, and store it in the `Word Automata Library` of `Walnut`. We can verify the correctness of this automaton as follows. First, we claim that $\varphi^\omega(0) = 0\mathbf{f}$. To see this, let f denote the morphism that maps $0 \to 010, 1 \to 01$; i.e., the morphism f is the square of the Fibonacci morphism. One can easily verify the identities $\varphi^n(0) = 0f^n(0)0^{-1}$ and $\varphi^n(01) = 0f^n(10)0^{-1}$ by simultaneous induction, whence follows the claim. We can then use Walnut to verify the correctness of the automaton `FF` with the command

```
eval verifyFF "?msd_fib FF[0]=@0 & Ai FF[i+1]=F[i]":
```

which returns `TRUE`.

Now we can use `Walnut` to create a Fibonacci automaton for $\mathbf{w}$.

```
morphism g "0->01 1->11":
image GF g FF:
```

The resulting automaton is called `GF.txt` and is displayed in Figure 2.

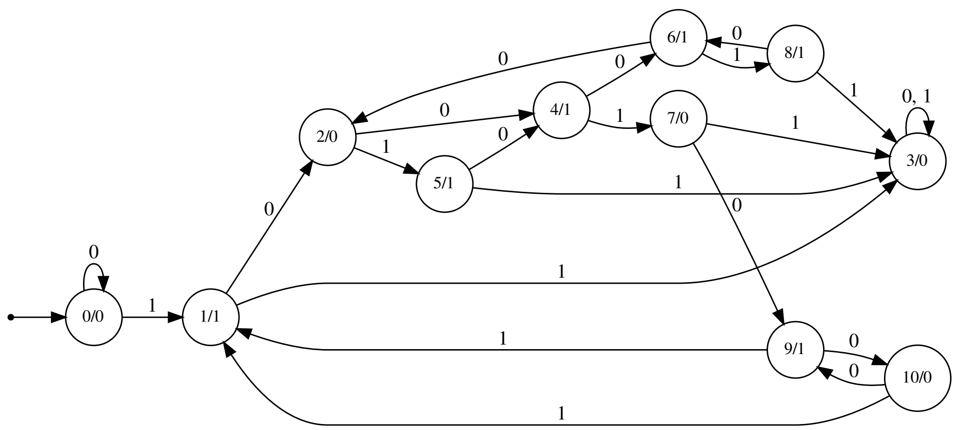

Fig. 2: Fibonacci automaton for $\mathbf{w} = g(\varphi^{\omega}(0))$.

Theorem 3. *The word* $\mathbf{w}$ *does not contain antisquares other than* 01 *and* 10*, and has critical exponent* $2 + \alpha$.

Proof: We write a Walnut formula asserting that there exists an antisquare of order ≥ 2, as follows:

```
eval antisq "?msd_fib Ei,n (n>=2) & At (t<n) => GF[i+t]!=GF[i+n+t]":
```

This returns `FALSE`, so there are no antisquares of order ≥ 2.

We now compute the periods that are associated with factors that have exponent ≥ 3.

```
eval gfper "?msd_fib Ei (p>=1) & (Aj (j<=2*p) => GF[i+j]=GF[i+j+p])":
```

The predicate above produces the automaton in Figure 3, which shows that these periods are of the form 10010^* in Fibonacci representation.

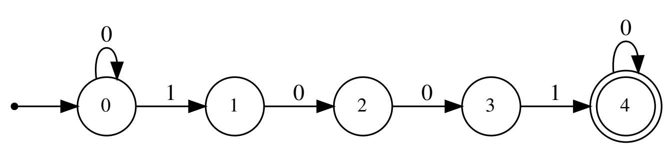

Fig. 3: Automaton for periods associated with 3^+-powers in $\mathbf{w}$.

Next, we compute the pairs (n, p) such that $\mathbf{w}$ contains a factor of length $n + p$ with period p of the form 10010^* and $n + p$ is the longest length of any factor with this period.

```
reg pows msd_fib "0*10010*";
def maximalreps "?msd_fib Ei
    (Aj (j<n) => GF[i+j] = GF[i+j+p]) & (GF[i+n] != GF[i+n+p])":
eval highestpow "?msd_fib (p>=1) & $pows(p) &
    $maximalreps(n,p) & (Am $maximalreps(m,p) => m <= n)":
```

The automaton produced by the predicate `highestpow` is given in Figure 4.

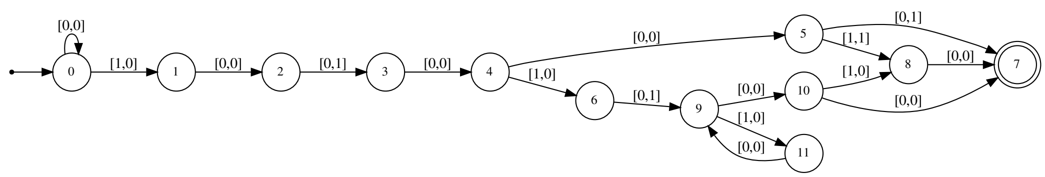

Fig. 4: Automaton for pairs (n, p) associated with highest powers in $\mathbf{w}$.

The strings accepted by this automaton, omitting the leading $[0, 0]$, are as follows:

- $[1, 0][0, 0][0, 1][0, 0][0, 0][0, 1]$
- $[1, 0][0, 0][0, 1][0, 0][0, 0][1, 1][0, 0]$
- $[1, 0][0, 0][0, 1][0, 0][1, 0][0, 1]([1, 0][0, 0])^k[0, 0][1, 0][0, 0]$, $k \geq 0$
- $[1, 0][0, 0][0, 1][0, 0][1, 0][0, 1]([1, 0][0, 0])^k[0, 0][0, 0]$, $k \geq 0$.

These correspond, respectively, to the values

- $(n, p) = (13, 6) = (2F_6 - 3, 2F_4)$
- $(n, p) = (23, 10) = (2F_7 - 3, 2F_5)$
- $(n, p) = (F_{2k+10} + F_3 + \sum_{3 \leq i \leq k+3} F_{2i}, F_{2k+8} + F_{2k+5}) = (2F_{2k+9} - 3, 2F_{2k+7})$ for $k \geq 0$
- $(n, p) = (F_{2k+9} + \sum_{3 \leq i \leq k+3} F_{2i-1}, F_{2k+7} + F_{2k+4}) = (2F_{2k+8} - 3, 2F_{2k+6})$ for $k \geq 0$.

where we have used the well-known Fibonacci identities

$$F_2 + F_4 + F_6 + \cdots + F_{2t} = F_{2t+1} - 1$$
$$F_1 + F_3 + F_5 + \cdots + F_{2t-1} = F_{2t}.$$

Now the exponent of these finite factors is $(n+p)/p$, which is

$$\frac{2F_j + 2F_{j-2} - 3}{2F_{j-2}}$$

for $j \geq 6$. These quotients tend to $2+\alpha$ from below, and hence the critical exponent is $2+\alpha$. □

3 Optimality of the previous construction

In this section we show that the critical exponent of the word constructed in Section 2 is best possible; i.e., that every infinite good word has critical exponent at least $2+\alpha$. It is somewhat easier to work with bi-infinite words, so we begin with results concerning bi-infinite words and then explain at the end of the section how to obtain the desired result for right-infinite words.

If S is a set of nonempty finite words, then by ${}^{\omega}S^{\omega}$ we mean the set of bi-infinite words made up of concatenations of the elements of S.

Theorem 4. *Every bi-infinite binary word avoiding* 4*-powers and* $\{11, 000, 10101\}$ *has the same set of factors as* $\mathbf{f}$.

Proof: First, we check that $\mathbf{f}$ avoids 4-powers and $\{11, 000, 10101\}$.

Now consider a bi-infinite binary word $\mathbf{w}$ avoiding 4-powers and $\{11, 000, 10101\}$. Since $\mathbf{w}$ avoids 11, we have $\mathbf{w} \in {}^{\omega}\{01, 0\}^{\omega}$. Thus there exists a bi-infinite word $\mathbf{v}$ such that $\mathbf{w} = h(\mathbf{v})$, where h is the morphism $0 \to 01$, $1 \to 0$. Now it suffices to show that the pre-image $\mathbf{v}$ also avoids 4-powers and $\{11, 000, 10101\}$. Clearly, $\mathbf{v}$ avoids 4-powers, since otherwise $\mathbf{w} = h(\mathbf{v})$ would contain a 4-power. Now we show by contradiction that $\mathbf{v}$ avoids every factor in $\{11, 000, 10101\}$.

- If $\mathbf{v}$ contains 11, then $\mathbf{v}$ contains 110. So $\mathbf{w}$ contains $h(110) = 0001$—a contradiction, since $\mathbf{w}$ avoids 000.
- If $\mathbf{v}$ contains 000 then $\mathbf{w}$ contains $h(000) = 010101$—a contradiction, since $\mathbf{w}$ avoids 10101.
- If $\mathbf{v}$ contains 10101 then $\mathbf{v}$ contains 0101010, since $\mathbf{v}$ avoids 11. So $\mathbf{w}$ contains $h(0101010) = 01001001001$. Since $\mathbf{w}$ avoids 11, we see that $\mathbf{w}$ contains $010010010010 = (010)^4$—a contradiction, since $\mathbf{w}$ avoids 4-powers.

□

Lemma 5. *Every bi-infinite binary word avoiding* 4*-powers and*

$$F = \{0011, 0110, 1100, 1001, 010101, 101010, 1000101110,$$
$$0111010001, 101110111011101, 010001000100010\}$$

has the same set of factors as $g(\mathbf{f})$ *or* $\overline{g(\mathbf{f})}$.

Proof: Notice that F is closed under bitwise complement and reversal. Let $\mathbf{w}$ be a bi-infinite binary word avoiding 4-powers and F. Suppose that $\mathbf{w}$ contains 001011.

- Since $1001 \in F$ and $0110 \in F$, the word $\mathbf{w}$ contains 00010111.
- Since 0000 and 1111 are 4-powers, the word $\mathbf{w}$ contains 1000101110.

This is a contradiction since $1000101110 \in F$. So $\mathbf{w}$ avoids 001011. By considering the complement, the word $\mathbf{w}$ also avoids 110100. Now suppose that $\mathbf{w}$ contains 110111011.

- Since $0110 \in F$, the word $\mathbf{w}$ contains 11101110111.
- Since 1111 is a 4-power, the word $\mathbf{w}$ contains 0111011101110.
- Since $0011 \in F$ and $1100 \in F$, the word $\mathbf{w}$ contains 101110111011101.

This is a contradiction since $101110111011101 \in F$. So $\mathbf{w}$ avoids 110111011. By considering the complement, the word $\mathbf{w}$ also avoids 001000100. Thus, $\mathbf{w}$ avoids 4-powers and

$$F' = \{0011, 0110, 1100, 1001, 010101, 101010, 001011, 110100, 110111011, 001000100\}.$$

Notice that F' is closed under complement and reversal. By symmetry, we now suppose that $\mathbf{w}$ contains 11. Notice that 111, 11011, and 1101011 are the only possible factors of $\mathbf{w}$ that start with 11, end with two identical letters, and contain two identical letters only as a prefix and a suffix. In particular, the word $\mathbf{w}$ avoids 00. Moreover, the blocks of consecutive 1's have length 1 or 3. So $\mathbf{w} \in {}^{\omega}\{01, 11\}^{\omega}$. Thus $\mathbf{w} = g(\mathbf{v})$ for some bi-infinite binary word $\mathbf{v}$.

Since $\mathbf{w}$ avoids 4-powers, the word $\mathbf{v}$ also avoids 4-powers. Moreover,

- $g(11) = 1111$ is a 4-power,
- $g(000) = 010101$ belongs to F,
- $g(10101) = 1101110111$ contains $110111011 \in F'$.

So $\mathbf{v}$ also avoids $\{11, 000, 10101\}$. By Theorem 4, the word $\mathbf{v}$ has the same set of factors as $\mathbf{f}$. That is, the word $\mathbf{w}$ has the same set of factors as $g(\mathbf{f})$. □

Theorem 6. *Every good bi-infinite binary word has critical exponent at least* $2 + \alpha$*.*

Proof: Suppose that $\mathbf{w}$ is a good bi-infinite binary word; that is, it contain no antisquares except possibly 01 and 10. Also assume $\mathbf{w}$ has critical exponent smaller than $2+\alpha$. Consider the set F from Lemma 5 and notice that $F \setminus \{101110111011101, 010001000100010\}$ contains only antisquares. So $\mathbf{w}$ avoids 4-powers and $F \setminus \{101110111011101, 010001000100010\}$. Moreover, $\mathbf{w}$ avoids $101110111011101 = (1011)^{15/4}$ since $15/4 > 2 + \alpha$. By symmetry, $\mathbf{w}$ also avoids 010001000100010. So $\mathbf{w}$ avoids 4-powers and F.

By Lemma 5, $\mathbf{w}$ has the same set of factors as either $g(\mathbf{f})$ or $\overline{g(\mathbf{f})}$. So $\mathbf{w}$ has critical exponent $2+\alpha$. □

Corollary 7. *Every (right-) infinite good binary word has critical exponent at least* $2 + \alpha$*.*

Proof: Let $\mathbf{w}$ be an infinite good binary word, and let $\text{RecFac}(\mathbf{w})$ denote the set of its recurrent factors. That is, the set $\text{RecFac}(\mathbf{w})$ consists of the factors of $\mathbf{w}$ that appear infinitely often in $\mathbf{w}$. Then for any $y \in \text{RecFac}(\mathbf{w})$, we see that y has arbitrarily large two-sided extensions in $\text{RecFac}(\mathbf{w})$. By a 'two-sided' analogue of König's infinity lemma, there exists a bi-infinite word $\mathbf{w}'$ such that every factor of $\mathbf{w}'$ is an element of $\text{RecFac}(\mathbf{w})$. By Theorem 6, the bi-infinite word $\mathbf{w}'$ has critical exponent at least $2 + \alpha$, and thus so does the infinite word $\mathbf{w}$. □

4 The class AO_ℓ

Instead of avoiding all antisquares of order greater than one, we could consider avoiding arbitrarily large antisquares.

Proposition 8. *Every infinite binary word avoiding $\frac{7}{3}$-powers contains arbitrarily large antisquares.*

Proof: By a result of Karhumäki and Shallit (2004), we know that every infinite binary word avoiding $\frac{7}{3}$-powers can be written in the form $x_1\mu(x_2\mu(x_3\mu(\cdots))\cdots)$, where $x_i \in \{\epsilon, 0, 1, 00, 11\}$ and μ is the Thue-Morse morphism, defined by $\mu : 0 \to 01$, $1 \to 10$. It follows that every such word must contain arbitrarily large factors of the form $\mu^n(0)$. But every word $\mu^n(0)$ for $n \geq 1$ is an antisquare. $\square$

On the other hand, we can prove the following result on the class AO_ℓ of binary words avoiding antisquares of order $\geq \ell$:

Theorem 9. *There exists an infinite β^+-free binary word containing no antisquare of order $\geq \ell$ for the following pairs (ℓ, β):*

(a) $(2, 2 + \alpha)$

(b) $(3, 8/3)$

(c) $(5, 5/2)$

(d) $(6, 7/3)$

These are all optimal.

Proof: Item (a) was already proved in Section 2. For each of the remaining pairs (ℓ, β), we apply a morphism ξ_ℓ to any ternary squarefree infinite word $\mathbf{w}$ and check that it has the desired properties. The morphisms are given in Table 1. The columns are ℓ, where the word contains no antisquares of order $\geq \ell$; β, where the word avoids β^+ powers; s, the size of the uniform morphism; and the morphism.

ℓ	β	s	morphism name	morphism
3	8/3	36	ξ_3	$0 \to 001001010011001010010011001001010011$
				$1 \to 001001010010011001010011001010010011$
				$2 \to 001001010010011001001010011001010011$
5	5/2	19	ξ_5	$0 \to 0010110100101101011$
				$1 \to 0010110100101100101$
				$2 \to 0010110011001010011$
6	7/3	37	ξ_6	$0 \to 0010011010010110100110110011010011011$
				$1 \to 0010011010010110100110110010011010011$
				$2 \to 0010011010010110100110110010011001011$

Tab. 1: Morphisms generating words in AO_ℓ.

To verify the β^+-freeness of $\xi_\ell(\mathbf{w})$ we use (Mol et al., 2020, Lemma 23). In order to do so, we check that ξ_ℓ is synchronizing and that $\xi_\ell(u)$ is β^+-free for every squarefree ternary word u of length t, where

t is specified by (Mol et al., 2020, Lemma 23). In order to show that $\xi_\ell(\mathbf{w})$ contains no antisquares of order $\geq \ell$, we find a length m such that if v and its complement are both factors of $\xi_\ell(\mathbf{w})$, then $|v| \leq m$. We can then check that $\xi_\ell(\mathbf{w})$ contains no antisquares of order $\geq \ell$ by exhaustively checking all factors of length at most $2m$. The parameters t and m for each ξ_ℓ are given in Table 2.

morphism	ℓ	β	t	m
ξ_3	3	8/3	8	6
ξ_5	5	5/2	10	16
ξ_6	6	7/3	14	26

Tab. 2: Parameters for checking correctness of ξ_ℓ.

The optimality of item (a) was already proved in Section 3. The optimality of the remaining items can be established by depth-first search. For each pair (ℓ, β), a longest word containing no antisquares of order $\geq \ell$, but avoiding β-powers (instead of β^+), is given in Table 3. The columns give ℓ, β, the length L of a longest such word, and a longest such word.

ℓ	β	L	example
4	$8/3$	29	00100101001100101001100110100
5	$5/2$	32	00100101100101101001011001011011
6	$7/3$	30	001011001101001011010011001011

Tab. 3: Longest words avoiding β-powers and containing no antisquares of order $\geq \ell$.

$\square$

5 The class AN_n

In this section, we consider the class AN_n of binary words containing no more than n distinct antisquares as factors.

Theorem 10. *There exists an infinite β^+-free binary word containing no more than n antisquares for the following pairs (n, β):*

(a) $(2, 2+\alpha)$

(b) $(3, 3)$

(c) $(6, 8/3)$

(d) $(9, 38/15)$

(e) $(10, 5/2)$

(f) $(15, 17/7)$

(g) $(16, 7/3)$.

These are all optimal.

Proof: Item (a) was already proved in Section 2. For the remaining cases, the proof is similar to that of Theorem 9: for each pair (n, β), we apply a morphism ζ_n to any ternary squarefree infinite word $\mathbf{w}$ and check that it has the desired properties. The morphisms are given in Table 4. The columns are n, the largest number of allowed antisquares; β, where the word avoids β^+ powers; s, the size of the uniform morphism; and the morphism.

n	β	s	morphism name	morphism
3	3	13	ζ_3	$0 \to 0010001000101$ $1 \to 0001001000101$ $2 \to 0001000100101$
6	$8/3$	36	ζ_6	$0 \to 001001010011001010010011001001010011$ $1 \to 001001010010011001010011001010010011$ $2 \to 001001010010011001001010011001010011$
9	$38/15$	192	ζ_9	$0 \to 0010100101100110010100101100110010110010100101100110101100110010$ $1100101001011001100101100101001011001101011001100101100101001011$ $0011001010011001010010110011001011001010010110011010110011001011$ $1 \to 0010100101100110010100101100110010110010100101100110101100110010$ $1100101001011001100101001100101001011001100101100101001011001101$ $0110011001011001010010110011001011001010010110011010110011001011$ $2 \to 0010100101100110010100101100110010110010100101100110010100110010$ $1001011001100101100101001011001100101001011001100101100101001011$ $0011001010011001010010110011001011001010010110011010110011001011$
10	$5/2$	75	ζ_{10}	$0 \to 0010100101100110010100110010100101100110010110010100101100110101$ 10011001011 $1 \to 0010100101100110010100110010100101100110010110010100101100110010$ 10011001011 $2 \to 0010100101100110010100110010100101100110010100110010110010100101$ 10011001011
15	$17/7$	194	ζ_{15}	$0 \to 0010011001011001001101001011001001100101100100110100101100100110 1$ $001100100110100101100100110100110010011001011001001101001011001 00$ $1101001100100110100101100100110010110010011010010110010011010011$ $1 \to 0010011001011001001101001011001001100101100100110100101100100110 1$ $001100100110100101100100110010110010011010010110010011010011001 00$ $1100101100100110100101100100110100110010011010010110010011010011$ $2 \to 0010011001011001001101001011001001100101100100110100101100100110 1$ $001100100110010110010011010010110010011010011001001101001011001 00$ $1100101100100110100101100100110100110010011010010110010011010011$
16	$7/3$	192	ζ_{16}	$0 \to 0010011001011001001101001011001001101001100100110100101100110100$ $1011001001101001100100110100101100100110010110010011010010110010$ $0110100110010011010010110010011001011001001101001011001101001011$ $1 \to 0010011001011001001101001011001001101001100100110100101100100110$ $0101100100110100101100110100101100100110100110010011010010110010$ $0110010110010011010010110010011010011001001101001011001101001011$ $2 \to 0010011001011001001101001011001001101001100100110100101100100110$ $0101100100110100101100100110100110010011010010110011010010110010$ $0110010110010011010010110011010010110010011010011001001101001011$

Tab. 4: Morphisms generating words in AN_n.

To verify the β^+-freeness of $\zeta_n(\mathbf{w})$ we use (Mol et al., 2020, Lemma 23). In order to do so, we check that ζ_n is synchronizing and that $\zeta_n(u)$ is β^+-free for every squarefree ternary word u of length t, where t is specified by (Mol et al., 2020, Lemma 23). In order to show that $\zeta_n(\mathbf{w})$ contains at most n distinct antisquares, we find a length m such that if v and its complement are both factors of $\zeta_n(\mathbf{w})$, then $|v| \leq m$. We can then enumerate the antisquares appearing in $\zeta_n(\mathbf{w})$ and check that there are at most n of them. The parameters t and ℓ for each ζ_n are given in Table 5.

morphism	n	β	t	m
ζ_3	3	3	6	4
ζ_6	6	8/3	8	6
ζ_9	9	38/15	9	17
ζ_{10}	10	5/2	10	17
ζ_{15}	15	17/7	11	12
ζ_{16}	16	7/3	14	13

Tab. 5: Parameters for checking correctness of ζ_n.

The optimality of item (a) was already proved in Section 3. The optimality of the remaining items can be established by depth-first search. For each pair (n, β), a longest word containing at most n antisquares, but avoiding β-powers (instead of β^+), is given in Table 6. The columns give n, β, the length L of a longest such word, and a longest such word.

n	β	L	example
5	3	17	00101001010010011
8	8/3	52	0010010100110010100110011010011001101011001101011011
9	38/15	407	00100101001101001010011010011001101001010011001010011 00110100101001101001100110101100110100101001101001100 11010010100110010100110011010010100110100110011010110 01101001010011010011001101001010011010011001101011001 10100101001101001100110100101001100101001100110100101 00110100110011010110011010010100110100110011010010100 11001010011001101001010011010011001101001010011001010 011001101001010011001101001101011011
14	5/2	92	001101001011001101100110100101100110110011010011 01100110100101100110110011010011011001101100
15	17/7	156	0010110011010010110010011010011001001101001011001001 1001011001001101001011001001101001100100110100101100 1001100101100100110100101100100110010110010011001001
16	7/3	38	00101100101101001011001101001011001011

Tab. 6: Longest words with at most n antisquares and avoiding β-powers.

□

6 Minimal antisquares

Consider the language L of all finite good words. In this section we determine the *minimal antisquares* or *minimal forbidden factors* for L Mignosi et al. (2002). These are the words w such that w is an antisquare, but w properly contains no antisquare factors, except possibly 01 and 10. This characterization will be useful for enumerating the number of length-n words in L.

The goal is to prove the following theorem:

Theorem 11. *The minimal antisquares are, organized by order* n*, as follows:*

- $n = 1$*:* $\{01, 10\}$
- $n = 2$*:* $\{0011, 0110, 1001, 1100\}$
- $n = 3$*:* $\{010101, 101010\}$
- $n = 4$*:* $\emptyset$
- $n \geq 5$*: all the* $2n$ *conjugates (cyclic shifts) of* $0^{n-2}101^{n-2}01$.

We start with some basic results about antisquares.

Lemma 12. *If* x *is an antisquare, so is every conjugate of* x*.*

Proof: Write x as $ay\overline{ay}$ for some (possibly empty) word y. Then a cyclic shift by one symbol gives $y\overline{ay}a$, which is clearly an antisquare. Repeating this argument $|x|$ times gives the result. □

Lemma 13. *A word* w *is a minimal antisquare if and only if all conjugates of* w *are minimal antisquares.*

Proof: Let $w = u\overline{u}$. Let us prove the forward direction first. Assume, contrary to what we want to prove, that w is a minimal antisquare, but some rotation of w contains a shorter antisquare, other than 01 and 10. Let rs be an antisquare of minimum length with $|rs| > 2$ that is not a factor of w, but is a factor of some rotation of w. Then we can write $w = sxr$ with $r, s, x \neq \epsilon$. There are three cases to consider:

$|r| \geq |\overline{u}|$: Let t be defined by $r = t\overline{u}$. Then $w = u\overline{u} = sxt\overline{sx}\overline{t}$, where $rs = t\overline{sx}\overline{t}s$ is an antisquare. If $t = \epsilon$, then $rs = \overline{sx}s$; hence $w = sx\overline{sx}$ contains the antisquare $sx\overline{s}$. This is a contradiction, since we assumed w does not have a shorter antisquare, but $sx\overline{s}$ is an antisquare in w with $s, x \neq \epsilon$. Therefore $t \neq \epsilon$.

Since w and rs are both antisquares, they have even length. Therefore $|x|$ is even. Consider the factorization $x = x_1x_2$ with $|x_1| = |x_2|$. Then $rs = t\overline{sx}\overline{t}s = t\overline{s}\overline{x}_1\overline{x}_2\overline{t}s$ is an antisquare, where $|t\overline{s}\overline{x}_1| = |\overline{x}_2\overline{t}s|$. Therefore, $t\overline{s}\overline{x}_1$ must end with $t\overline{s}$ and $\overline{x}_2\overline{t}s$ must begin with $\overline{t}s$, giving the antisquare $t\overline{s}\overline{t}s$, which is not 01 or 10 (since $t, s \neq \epsilon$), and is shorter than rs. This contradicts our assumption that rs is the smallest such antisquare.

$|s| \geq |u|$: Let s be defined by $s = ut$. Then $w = u\overline{u} = \overline{txr}txr$. Now $rs = r\overline{txr}t$ is an antisquare, and hence the complement $\overline{rs} = \overline{r}txr\overline{t}$ is also an antisquare. Let $r' = \overline{r}txr$ and $s' = \overline{t}$. So we can write $w = \overline{txr}txr = s'\overline{x}r'$ where $r's'$ is an antisquare of the same length as rs, and $|r'| > |\overline{u}|$. This reduces the problem to the previous case.

$|s| < |u|, |r| < |\overline{u}|$: In this case we can write $u = sy$ and $\overline{u} = zr$ where $y, z \neq \epsilon$. This means that u ends in $\overline{r}$ and $\overline{u}$ begins with $\overline{s}$, implying that w contains the antisquare $\overline{rs}$. This contradicts the assumption that w does not contain a shorter antisquare.

The reverse direction is easy. If w contains a shorter antisquare, other than 01 and 10, then at least one rotation of w also has the same antisquare. □

We introduce some terminology. A *run* in a word is a maximal block of consecutive identical symbols. The *run-length encoding* $r : \Sigma^* \to \mathbb{N}^*$, where $\mathbb{N} = \{1, 2, 3, \ldots\}$ is a map sending a word x to the list of the lengths of the consecutive runs in x. For example, $r(\texttt{access}) = 1212$.

Lemma 14. *If a nonempty word x is an antisquare, then the number of runs it contains must be congruent to* 2 *or* $3 \pmod 4$.

Proof: Write $x = u\overline{u}$ and consider the runs in u. If u has $2k + 1$ runs, then so does $\overline{u}$. Furthermore, u ends in a different letter than the start of $\overline{u}$. Hence x has $4k + 2$ runs.

Otherwise u has $2k$ runs. Then the last letter of u is the same as the first letter of $\overline{u}$. Hence x has $4k - 1$ runs. □

Lemma 15. *The word $0^k 10\, 1^k 01$ is a minimal antisquare for $k \geq 3$.*

Proof: It is easy to see that $x = 0^k 10\, 1^k 01$ is an antisquare. Suppose $k \geq 3$ and suppose x has an antisquare proper factor w other than 01 and 10. Now x has six runs, so by Lemma 14 we know that w has either two, three, or six runs.

If w has two runs, it must be of the form $a^i \overline{a}^i$ for $i \geq 2$, but inspection shows that x has no factor of that form.

If w has three runs, it must be of the form $a^i \overline{a}^{i+j} a^j$ for some $i, j \geq 1$. The only possibility is $i = 1$, $i + j = k$, $j = 1$, which forces $k = 2$, a contradiction.

Finally, if w has six runs, then $w = 0^\ell 101^\ell 01$ for some $\ell < k$, which would not be a factor of x. □

We are now ready to complete the proof of Theorem 11. We say x has an *interior occurrence* of y if we can write $x = wyz$ for nonempty words w, z.

Proof Proof of Theorem 11.: By combining Lemmas 12, 13, and 15, and verifying the listed cases for $n \leq 4$, we see that all the words given in the statement of the theorem are minimal antisquares.

It now remains to see that there are no other minimal antisquares. The idea is to classify antisquares x by the number of runs. In what follows, we assume, without loss of generality, that x begins with 0.

Two runs: then $x = 0^i 1^i$ for some $i \geq 1$. If $i \geq 3$ then x contains the antisquare 0011. So the only minimal antisquares are 01 and 0011.

Three runs: then $x = 0^i 1^{i+j} 0^j$. If either i or j is at least 2, then x contains the antisquare 0011 or 1100. So the only minimal antisquare is 0110.

Six runs: then $x = 0^i 1^j 0^k 1^i 0^j 1^k$. If $i, j \geq 2$ then x contains the antisquare 0011, and similarly if $j, k \geq 2$ and $k, i \geq 2$. It follows that $(i, j, k) \in \{(1, 1, 1), (1, 1, n), (1, n, 1), (n, 1, 1)\}$ for $n \geq 2$. The cases $(1, 1, 2), (1, 2, 1), (2, 1, 1)$ are ruled out by an antisquare of the form 1001 or 0110. So $(i, j, k) \in \{(1, 1, 1), (1, 1, n), (1, n, 1), (n, 1, 1)\}$ for $n \geq 3$. The case $(1, 1, 1)$ corresponds to the word 010101, and the remaining cases correspond to certain conjugates of 0^n101^n01 for $n \geq 3$, already listed in the statement of the theorem.

Seven runs: then $x = 0^i1^j0^k1^{i+l}0^j1^k0^l$. Again, if $i, j \geq 2$, or $j, k \geq 2$, or $k, l \geq 2$, then x contains a shorter antisquare 0011 or 1100. Since $i + l \geq 2$, the same argument rules out $k \geq 2$ and $j \geq 2$. So the only cases remaining are

$$(i, j, k, l) \in \{(1, 1, 1, 1), (1, 1, 1, n), (n, 1, 1, 1), (n, 1, 1, n)\}$$

for $n \geq 2$. The first case $(1, 1, 1, 1)$ corresponds to 010110101, which has the antisquare 0110, and it is easy to verify that the remaining cases are certain conjugates of $0^{n+1}101^{n+1}01$ for $n \geq 3$, already listed in the statement of the theorem.

It now remains to handle the case of more than 7 runs. By Lemma 14 x has at least 10 runs. This involves a rather tedious examination of cases, based on the following three simple observations:

(a) if $r(x)$ contains two consecutive terms, both ≥ 2, then x contains the shorter antisquare 0011 or 1100;

(b) if $r(x)$ contains six consecutive terms $a1bc1d$ with $a \geq c$ and $b \leq d$, then x contains the shorter antisquare $0^c10^b1^c01^b$ or its complement.

(c) if $r(x)$ contains an interior occurrence of 2, then x contains the antisquare 0110 or 1001.

Suppose $x = u\overline{u}$. If $z = r(u)$ is of odd length, then $zz = r(x)$. If $z = r(u)$ is of even length, then writing $z = ayb$ with a, b single numbers, we have $r(x) = ay(a + b)yb$. When we speak of a maximal 1-block in what follows, we mean one that cannot be extended by additional 1's to the left or right.

It now suffices to prove the following two lemmas:

Lemma 16. *Let $z \in \mathbb{N}^*$, and suppose $|z| \geq 5$ is odd. Then zz contains either*

(a) two consecutive terms that are ≥ 2, or

(b) six consecutive terms $a1bc1d$ with $a \geq c$ and $b \leq d$.

Proof: If condition (a) is not satisfied, then z consists of isolated occurrences of numbers ≥ 2, separated by blocks of consecutive 1's. We assume this in what follows.

If z both begins and ends with a number ≥ 2, then zz satisfies (a). Thus we may assume that z either begins or ends with 1 (or both).

Suppose z contains the block 1111. Then zz contains the block $b1111c$ for $b, c \geq 1$, and hence satisfies (b). Thus we may assume that the maximal 1-blocks in z are of length 1, 2, or 3.

Suppose all the maximal 1-blocks of z are of length 1 or 3. If z begins with 1 and ends with $b \geq 2$, then z cannot be of odd length, and similarly if z ends with 1 and begins with $b \geq 2$. So z must begin and end with 1. Since $|z| \geq 5$, we know z has a prefix of the form $1c1d$ and a suffix of the form $a1b1$, where $a, b, c, d \geq 1$. Hence zz contains the block $a1b11c1d$. If $b \leq c$, then the block $a1b11c$ fulfills condition (b); if $b \geq c$, then the block $b11c1d$ fulfills condition (b).

Thus there must be a maximal 1-block of length 2 in z. Then zz contains two blocks, one of the form $a1b11c$ and one of the form $b11c1d$, where $a, d \geq 1$ and $b, c \geq 2$. If $b \leq c$, then the block $a1b11c$ fulfills condition (b); if $b \geq c$, then the block $b11c1d$ fulfills condition (b). □

Lemma 17. *Let $z \in \mathbb{N}^*$, and suppose $|z| \geq 6$ is even, and write $z = ayb$. Define $z' = ay(a+b)yb$. Then z' contains either*

(a) two consecutive terms that are ≥ 2, or

(b) six consecutive terms $a1bc1d$ with $a \geq c$ and $b \leq d$, or

(c) an interior occurrence of 2.

Proof: If condition (a) is not satisfied, then z consists of isolated occurrences of numbers ≥ 2, separated by blocks of consecutive 1's. We assume this in what follows.

If z begins and ends with 1, then z' has an interior occurrence of 2, so (c) is satisfied. So assume this is not the case.

If z begins $1b$ with $b \geq 2$, then by the previous paragraph it must end in $c \geq 2$. Then z' has an occurrence of $(c+1)b$, so (a) is satisfied. Exactly the same argument works if z ends with $b1$ with $b \geq 2$. So assume neither of these hold.

If z has an interior occurrence of 11111, then z' has an occurrence of $c11111$, fulfilling condition (b). If z begins $11111c$ for $c \geq 1$, it must end with $d \geq 2$, so z' has an occurrence of $(d+1)1111c$, fulfilling (b). The analogous argument holds if z ends $c11111$. So all maximal 1-blocks in z are of length ≤ 4.

Now we consider the case that z has a maximal block of the form 1111. If this occurrence is interior in z, then z' contains the block $b1111c$ for $b, c \geq 1$, and hence satisfies (b). If z has the prefix 1111, then z cannot end in 1 by above. Hence, since z has even length, it must contain another maximal 1-block of even length, which must be interior. Since we have already ruled out the possibility of an interior occurrence of 1111, it must be an interior 1-block of size 2. But then z' has a block of the form $a1b11c1d$ where $a, d \geq 1$ and $b, c \geq 2$. As in the previous lemma, if $b \leq c$, then the block $a1b11c$ fulfills condition (b); if $b \geq c$, then the block $b11c1d$ fulfills condition (b). The analogous argument holds if the occurrence of 1111 is a suffix. Hence z contains no maximal 1-block of size 4.

Thus we may assume that the maximal 1-blocks in z are of length 1, 2, or 3.

If z has a maximal 1-block of size 2, then since $|z|$ is even, it must have a second maximal 1-block of size 2. Then z' has a factor of the form $a1b11c1d$ where $a, d \geq 1$ and $b, c \geq 2$, and by the argument above, this satisfies condition (b).

Hence all the maximal 1-blocks of z are of size 1 or 3. Hence all the maximal 1-blocks of z are size 1 or 3. Since z has even length, exactly one of its first and last symbols must be 1. We know from above that z cannot begin $1b$ or end $b1$ with $b \geq 2$. So z must either begin 111 or end 111. If z begins 111, then z ends in $b \geq 2$, and z' contains the block $a1(b+1)11c1d$ for some $a, d \geq 1$ and $b, c \geq 2$. If $c \leq b+1$, then the block $a1(b+1)11c$ fulfills condition (b); if $b+1 \geq c$, then the block $(b+1)11c1d$ fulfills (b). If z ends 111, then an analogous argument holds. □

Applying the two lemmas to the case where $x = u\,\overline{u}$ and $z = r(u)$ completes the proof of Theorem 11. □

7 Enumerating words with only two distinct antisquares

In this section we obtain some enumeration results for good words. Here the notation ${}^{\omega}x$ refers to the left-infinite word $\cdots xxx$.

Proposition 18. *If a bi-infinite good word w contains the factor 001011, then $w = {}^{\omega}0101^{\omega}$.*

Proof: Consider a maximal factor of w of the form 0^k101^k, with $2 \le k < \infty$. By maximality and by symmetry, we can assume that w contains 0^k101^k0. This factor is not extendable to the right:

- 0^k101^k00 contains the antisquare 1100 as a suffix.
- 0^k101^k01 is an antisquare.

This is a contradiction to $k < \infty$, so $w = {}^{\omega}0101^{\omega}$. □

Theorem 19. *There are $\Theta(\psi^n)$ good words of length n, where $\psi \doteq 1.465571231876768$ is the super-golden ratio, root of the equation $X^3 = X^2 + 1$.*

Proof: Let $F = \{0011, 1100, 0110, 1001, 010101, 101010, 001011, 110100\}$. By Theorem 11, F contains the minimal antisquares of order 2 and 3, and the minimal antisquares of order at least 4 contain 001011 or 110100. Thus, the binary words avoiding F are exactly the good words that avoid $\{001011, 110100\}$.

By Proposition 18, there are not enough good words containing 001011 to contribute to the growth rate of good words. This also holds for good words containing the symmetric factor 110100. Thus, good words and binary words avoiding F have the same growth rate.

To enumerate binary words avoiding F, we instead enumerate the 'Pansiot codes' of these words. If $x = x_1x_2 \cdots x_n$ is a binary word, then the *Pansiot code* of x is the binary word $p_1p_2 \cdots p_{n-1}$ such that for $i = 1, \ldots, n-1$

$$x_{i+1} = \begin{cases} x_i & \text{if } p_i = 0; \\ \overline{x_i} & \text{if } p_i = 1. \end{cases}$$

For example, the binary word 010 is the Pansiot code for the two binary words 0011 and 1100.

The Pansiot codes of binary words avoiding F are the binary words avoiding $\{010, 101, 11111, 01110\}$. These words consist of blocks of 0's of length at least 2 and blocks of 1's of length 2 or 4. Consider the number C_n of such words ending with 00. They are obtained from shorter words by adding a suffix 0, 1100, or 111100. From the relation $C_n = C_{n-1} + C_{n-4} + C_{n-6}$, the growth rate is the positive real root of $X^6 = X^5 + X^2 + 1$. Since $X^6 - X^5 - X^2 - 1 = (X+1)(X^2 - X + 1)(X^3 - X^2 - 1)$, this is the root ψ of $X^3 = X^2 + 1$. □

Next we show that the threshold exponent at which the number of good words becomes exponential is $\frac{15}{4}$. (For overlap-free words, the threshold is $\frac{7}{3}$; see Karhumäki and Shallit (2004).)

Theorem 20. *Let w be any squarefree word over the alphabet $\{0, 1, 2\}$. Apply the map h that sends*

$$\begin{aligned} 0 &\to 010001 \\ 1 &\to 0100010001 \\ 2 &\to 01000100010001. \end{aligned}$$

The resulting word is good, and has exponent at most $\frac{15}{4}$, and it is exactly $\frac{15}{4}$ if $|w| \ge 5$.

Proof: The goodness of $h(w)$ can be seen by inspection. Regarding the exponent $\frac{15}{4}$, suppose that $h(w)$ contains a $\frac{15}{4}$ power $zzzz'$, where z' is a prefix of z. Note that in $h(w)$ the factor 101 can only occur at the 'boundary' between $h(a)$ and $h(b)$, where $a, b \in \{0, 1, 2\}$. So we have two cases:

Case 1: z contains 101. Write $z = x101y$. Then $h(w)$ contains the square $01yx101yx1$, where $01yx1 = h(Z)$ for some factor Z of w. Then w contains the square ZZ, which is a contradiction.

Case 2: z does not contain 101. Clearly $z = h(a)$ or $z = h(a)0$ for some $a \in \{0, 1, 2\}$ is not possible, so z is contained within some $h(a)$, where $a \in \{0, 1, 2\}$. The only such $\frac{15}{4}$-power is $h(2)0 = (0100)^3010$, which establishes the claim. □

Corollary 21. *There are exponentially many length-n $\frac{15}{4}^+$-free good words.*

To show that there are only polynomially many length-n good words avoiding $\frac{15}{4}$-powers we need a version of the results in Section 3 for finite words rather than bi-infinite words. Let g and φ be defined as in Section 2 and let g' be the morphism that maps $0 \mapsto 01$ and $1 \mapsto 00$.

Lemma 22. *Let w be a 4-free word of length ≥ 15 that contains no antisquares other than 01 and 10. Then w can be written as either $w = w_1 g(v) w_2$ or $w = w_1 g'(v) w_2$ for some v, where $|w_1|, |w_2| \leq 5$.*

Proof: By a finite search, one verifies that any w satisfying the hypotheses of the lemma has a prefix of length ≤ 9 that contains either 0001 or 0111. Suppose it contains 0111. Since w avoids the antisquares 0110 and 1001 and the 4-powers 0000 and 1111 we can write $w = w_1 0111 z$, where z is a prefix of a word in $\{0001, 01, 0111\}^*$ and $|w_1| \leq 5$. We claim that z does not contain 0001. Note that $|z| \geq 6$.

If z has 0001 as a prefix, then w contains the antisquare 111000. Suppose z has 01 as a prefix. If z has 010001 as a prefix, then w contains the antisquare 0111010001. If z has 01010 as a prefix, then w contains the antisquare 101010, so necessarily z has 010111 as a prefix. Finally, it may be the case that z has 0111 as a prefix. Applying this argument repeatedly to the suffix of w following this new occurrence of 0111, until we no longer have such a suffix of length at least 6, we see that z does not contain 0001. It follows that w can be written as $w = w_1 x w_2$, where $x \in \{01, 0111\}^*$ (and hence $x \in \{01, 11\}^*$), and $|w_2| \leq 5$. Thus $w = w_1 g(v) w_2$ for some v, as required.

A similar argument shows that if w contains 0001 then $w = w_1 g'(v) w_2$ for some v, where $|w_1| \leq 5$ and $|w_2| \leq 5$. □

In what follows we will consider words of the form $g(v)$; the analysis for $g'(v)$ is similar.

Lemma 23. *Let $n \geq 1$ and $y = g(\varphi^n(x))$ for some binary word x with $|x| \geq 5$. If y is $\frac{15}{4}$-free, then x is 4-free and can be written in the form $x = px's$ where $|p| \leq 2$, $|s| \leq 1$ and x' has no 000 or 11.*

Proof: Clearly x is 4-free. Suppose x contains an occurrence of 000 that is neither a prefix nor a suffix of x. Then $\varphi(x)$ contains the 4-power $01(001)^3 0 = (010)^4$ and hence y contains a 4-power, which is a contradiction. Suppose x contains an occurrence of 11 of the form $x = u11v$, where $|u| \geq 2$ and $|v| \geq 1$. Then $\varphi(x)$ contains 0101 and hence, extending this occurrence of 0101 to the left and right with blocks 001 and 01 and avoiding 4-powers, we see that $\varphi(x)$ contains 00101010. Indeed, extending two blocks to the left and one block to the right suffices. If $n = 1$ then $g(00101010)$ contains the $\frac{15}{4}$-power $(1011)^3 101$, which is a contradiction. If $n > 1$, then $\varphi(00101010)$ contains the 4-power $(01001)^4$ and hence y contains a 4-power, which is a contradiction. □

Lemma 24. *Let w be a $\frac{15}{4}$-free word of length ≥ 33 that contains no antisquares other than* 01 *and* 10*. Then w can be written in the form*

$$w = w_1 G(u_1 \varphi(u_2 \cdots \varphi(u_r \varphi(V) v_r) \cdots v_2) v_1) w_2$$

for some r, where $G \in \{g, g'\}$, $|w_1|, |w_2| \leq 5$, $|u_i| \leq 4$, $|v_i| \leq 3$, for $i = 1, \ldots, r$, and $|V| \leq 4$.

Proof: By Lemma 22, we can write $w = w_1 G(v) w_2$, where $|w_1|, |w_2| \leq 5$. Without loss of generality, suppose $G = g$. Clearly v must be 4-free. Furthermore, v does not contain 000 or 11, since otherwise $g(v)$ would contain either the antisquare 010101 or the 4-power 1111. Thus $v = u_1 \varphi(v') v_1$, where $|u_1|, |v_1| \leq 2$ and $|v'| \geq 5$. We can then apply Lemma 23 to $g(\varphi(v'))$ to find that v' is 4-free and can be written in the form $v' = px's$, where $|p| \leq 2$, $|s| \leq 1$ and x' has no 000 or 11. Then we can write $v' = u_2 \varphi(v'') v_2$, where $|u_2| \leq 4$ and $|v_2| \leq 3$, and repeat the process to obtain the desired decomposition. □

Theorem 25. *There are polynomially many length-n good words avoiding $\frac{15}{4}$-powers.*

Proof: Let w be such a word of length n, where $n \geq 33$. By Lemma 24 w can be written in the form

$$w = w_1 G(u_1 \varphi(u_2 \cdots \varphi(u_r \varphi(V) v_r) \cdots v_2) v_1) w_2$$

for some r, where $G \in \{g, g'\}$, $|w_1|, |w_2| \leq 5$, $|u_i| \leq 4$, $|v_i| \leq 3$, for $i = 1, \ldots, r$, and $|V| \leq 4$. Suppose $G = g$ and under this assumption let A (resp. B, C, D, E) be the maximum number of possible choices for w_1 (resp. w_2, u_i, v_i, V). Then the number of words w is at most $AB(CD)^r E$. There is a constant ρ such that $r \leq \rho \log n$, so the number of words w is at most $ABEn^{\rho \log(CD)}$. A similar calculation applies when $G = g'$. □

8 Further work

In this paper we have studied antisquares. This situation has an obvious generalization to patterns with morphic and antimorphic permutations, as studied in Currie et al. (2015). This could be the subject of a future study.

A companion paper to this one is Currie et al. (2023), which investigates complement avoidance in binary words.

References

G. Badkobeh and M. Crochemore. Finite-repetition threshold for infinite ternary words. In P. Ambrož, Š. Holub, and Z. Masáková, editors, *WORDS 2011*, volume 63 of *Elec. Proc. Theor. Comput. Sci.*, pages 37–43. Open Publishing Association, 2011.

A. Carpi and A. d. Luca. Special factors, periodicity, and an application to Sturmian words. *Acta Informatica*, 36:983–1006, 2000.

J. Currie and N. Rampersad. A proof of Dejean's conjecture. *Math. Comp.*, 80:1063–1070, 2011.

J. Currie, F. Manea, and D. Nowotka. Unary patterns with permutations. In I. Potapov, editor, *DLT 2015*, volume 9168 of *Lecture Notes in Computer Science*, pages 191–202. Springer-Verlag, 2015.

J. Currie, L. Dvořaková, D. Opočenská, N. Rampersad, and J. Shallit. Complement avoidance in binary words. Arxiv preprint arXiv:2209.09598 [math.CO]. Available at `https://arxiv.org/abs/2209.09598`, 2023.

J. D. Currie, L. Mol, and N. Rampersad. The repetition threshold for binary rich words. *Discrete Math. & Theoret. Comput. Sci.*, 22(1):DMTCS–22–1–6, 2020. URL `https://dmtcs.episciences.org/6082`.

F. Dejean. Sur un théorème de Thue. *J. Combin. Theory. Ser. A*, 13:90–99, 1972.

L. Dvořáková, K. Medková, and E. Pelantová. Complementary symmetric Rote sequences: the critical exponent and the recurrence function. *Discrete Math. & Theoret. Comput. Sci.*, 22(1):DMTCS–22–1–20, 2020. URL `https://dmtcs.episciences.org/6519/`.

L. Dvořáková, D. Opočenská, E. Pelantová, and A. M. Shur. On minimal critical exponent of balanced sequences. Preprint, available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4011686`, 2022.

F. Fiorenzi, P. Ochem, and E. Vaslet. Bounds for the generalized repetition threshold. *Theoret. Comput. Sci.*, 412:2955–2963, 2011.

P. Hieronymi, D. Ma, R. Oei, L. Schaeffer, C. Schulz, and J. Shallit. Decidability for Sturmian words. In F. Manea and A. Simpson, editors, *30th EACSL Annual Conference on Computer Science Logic (CSL 2022)*, Leibniz International Proceedings in Informatics, pages 24:1–24:23. Schloss Dagstuhl—Leibniz-Zentrum für Informatik, 2022.

L. Ilie, P. Ochem, and J. Shallit. A generalization of repetition threshold. *Theoret. Comput. Sci.*, 345: 359–369, 2005.

J. Karhumäki and J. Shallit. Polynomial versus exponential growth in repetition-free binary words. *J. Combin. Theory. Ser. A*, 105(2):335–347, 2004.

F. Mignosi, A. Restivo, and M. Sciortino. Words and forbidden factors. *Theoret. Comput. Sci.*, 273: 99–117, 2002.

L. Mol, N. Rampersad, and J. Shallit. Extremal overlap-free and extremal β-free binary words. *Electronic J. Combinatorics*, 27(4):#P4.42, 2020.

H. Mousavi. Automatic theorem proving in `Walnut`. Arxiv preprint arXiv:1603.06017 [cs.FL], available at `http://arxiv.org/abs/1603.06017`, 2016.

H. Mousavi and J. Shallit. Repetition avoidance in circular factors. In M.-P. Béal and O. Carton, editors, *DLT 2013*, volume 7907 of *Lecture Notes in Computer Science*, pages 384–395. Springer-Verlag, 2013.

H. Mousavi, L. Schaeffer, and J. Shallit. Decision algorithms for Fibonacci-automatic words, I: Basic results. *RAIRO Inform. Théor. App.*, 50:39–66, 2016.

T. Ng, P. Ochem, N. Rampersad, and J. Shallit. New results on pseudosquare avoidance. In R. Mercas and D. Reidenbach, editors, *WORDS 2019*, volume 11682 of *Lecture Notes in Computer Science*, pages 264–274. Springer-Verlag, 2019.

N. Rampersad, J. Shallit, and É. Vandomme. Critical exponents of infinite balanced words. *Theoret. Comput. Sci.*, 777:454–463, 2019.

M. Rao. Last cases of Dejean's conjecture. *Theoret. Comput. Sci.*, 412:3010–3018, 2011.

A. V. Samsonov and A. M. Shur. On Abelian repetition threshold. *RAIRO Inform. Théor. App.*, 46: 147–163, 2012.

J. Shallit. Simultaneous avoidance of large squares and fractional powers in infinite binary words. *Internat. J. Found. Comp. Sci.*, 15:317–327, 2004.

J. Shallit. Minimal critical exponents for palindromes. ArXiv preprint, arXiv:1612.05320 [cs.FL]. Available at `https://arxiv.org/abs/1612.05320`, 2016.

J. Shallit. *The Logical Approach To Automatic Sequences: Exploring Combinatorics on Words with* `Walnut`, volume 482 of *London Math. Soc. Lecture Note Series*. Cambridge University Press, 2022.