

# Anti-Powers in Infinite Words

Gabriele Fici<sup>1</sup>, Antonio Restivo<sup>2</sup>, Manuel Silva<sup>3</sup>, and  
Luca Q. Zamboni<sup>4</sup>

- 1 Dipartimento di Matematica e Informatica, Università di Palermo, Palermo, Italy  
gabriele.fici@unipa.it
- 2 Dipartimento di Matematica e Informatica, Università di Palermo, Palermo, Italy  
antonio.restivo@unipa.it
- 3 Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Lisbon, Portugal  
mnsilva@gmail.com
- 4 Institut Camille Jordan, Université Claude Bernard Lyon 1, Lyon, France  
zamboni@math.univ-lyon1.fr

---

## Abstract

In combinatorics of words, a concatenation of  $k$  consecutive equal blocks is called a power of order  $k$ . In this paper we take a different point of view and define an anti-power of order  $k$  as a concatenation of  $k$  consecutive pairwise distinct blocks of the same length. As a main result, we show that every infinite word contains powers of any order or anti-powers of any order. That is, the existence of powers or anti-powers is an unavoidable regularity. Indeed, we prove a stronger result, which relates the density of anti-powers to the existence of a factor that occurs with arbitrary exponent. From these results, we derive that at every position of an aperiodic uniformly recurrent word start anti-powers of any order. We further show that any infinite word avoiding anti-powers of order 3 is ultimately periodic, and that there exist aperiodic words avoiding anti-powers of order 4. We also show that there exist aperiodic recurrent words avoiding anti-powers of order 6, and leave open the question whether there exist aperiodic recurrent words avoiding anti-powers of order  $k$  for  $k = 4, 5$ .

**1998 ACM Subject Classification** F.4.3 Formal Languages

**Keywords and phrases** infinite word, anti-power, unavoidable regularity, avoidability

**Digital Object Identifier** 10.4230/LIPIcs.ICALP.2016.124

## 1 Introduction

Suppose you are in a room with a hundred people and somebody tells you that by an incredible coincidence the people in the room have *all different* birthdays. Of course this is much less surprising than if all the people had *the same* birthday; still you remember from your first course in combinatorics (or even before) that already in a class of fifty people the probability that no two people have the same birthday is less than 3%. So actually you are in a very *special* situation!<sup>1</sup>

For a number of instances of objects taken from a fixed class using some rule, being all distinct can be viewed as a kind of regularity. This has been already considered extensively

---

<sup>1</sup> To be more precise, the probability that at least two people have the same birthday in a room of 100 people is about 0.9999996928.



in computer science. For example, the probability of collisions must be properly quantified when designing hashing functions. In the context of string processing, one usually deals with factors (substrings) of a word, and it is sometimes useful to factorize a word in blocks that are all distinct — a widely known example is the Lempel-Ziv factorization, at the basis of the eponymous compression algorithm. The problem of factoring a string in blocks that are all distinct (sometimes called equality-free factorization [7]) has practical applications also in bio-informatics, since it appears to be connected with gene synthesis [2]. Equality-free factorizations have been further considered by Fernau et al. [4], motivated by injective pattern matching with variables, which is identical to the special case of solving word equations where the left side of the equation does not contain variables, and different variables must be replaced by different words. In particular, in [4] it is proved that given a finite word  $w$  and a number  $k$ , the problem whether it is possible to factorize  $w$  into at least  $k$  distinct factors is  $\mathcal{NP}$ -complete.

In the context of infinite words, the complexity of sequences is often described by means of parameters that capture some kind of repetitiveness. To this end, one often considers as a degree of repetitiveness the maximal number of consecutive all-equal blocks occurring in the sequence, regardless of the length of the single blocks. A concatenation of  $k$  consecutive equal blocks is called a *power of order  $k$* , or simply a  *$k$ -power*. E.g., *aabaabaabaab* is a 4-power. A first classification of infinite words consists in identifying those that are  *$k$ -power-free* for some  $k \geq 2$ , meaning that they do not contain any factor that is a  $k$ -power. Words avoiding  $k$ -powers have been the object of study of combinatorics on words since the very beginning of the theory [8] (cf. also [5]).

In this paper we adopt a different point of view based on the difference rather than on the equality. We consider the problem of finding in infinite words consecutive blocks of the same length that are all distinct. Of course, in the context of infinite words it is the requirement that the blocks all have the same length that makes the problem non-trivial, since otherwise one can always take arbitrarily long concatenations of blocks of increasing length to guarantee that they are all distinct.

We define an *anti-power of order  $k$* , or simply a  *$k$ -anti-power*, as a concatenation of  $k$  consecutive pairwise distinct blocks of the same length. E.g., *aabaaabbbaba* is a 4-anti-power. A simple computation shows that there are in general much more anti-powers than powers for a fixed length and a fixed order; yet there are much less anti-powers than possible factors of the same given length.

Let us focus on an example. The Thue-Morse word

$$t = 0110100110010110100101100110100110010110010110011010 \dots$$

is perhaps the most prominent example in combinatorics of words [1]. It is defined as the word whose  $n$ -th digit is the parity of the number of 1s in the binary expansion of  $n - 1$  (so the first digit is the parity of 1s in 0, the second digit is the parity of 1s in 1, the third digit is the parity of 1s in 10, etc.). The Thue-Morse word does not contain *overlaps*, i.e., factors of the form *awawa* for a letter  $a$  and a word  $w$ . In particular, the Thue-Morse word does not contain 3-powers (note that, on the other hand, every infinite binary word must contain 2-powers).

The shortest prefix of the Thue-Morse word that is a 2-antipower is 01. The shortest prefix that is a 3-anti-power is 01101 · 00110 · 01011, of length 15. One can verify that the shortest 4-anti-power prefix has length 20. The first few lengths of the shortest prefixes of  $t$  that are  $k$ -anti-powers for different values of  $k$  are displayed in Table 1. A natural question is therefore the following: Given an integer  $k > 1$ , is it always possible to find a prefix of  $t$

■ **Table 1** The first few values of the sequence of lengths of the shortest prefixes of the Thue-Morse word that are  $k$ -anti-powers.

$k$	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	30	50	100
length	15	20	25	30	77	88	99	110	121	132	143	154	195	208	221	234	247	260	870	2450	9700
ratio	5	5	5	5	11	11	11	11	11	11	11	11	13	13	13	13	13	13	29	49	97

that is a  $k$ -anti-power? In this paper we answer this question in the affirmative. Actually, we prove a much stronger result. Indeed, we prove that the existence of powers of any order or anti-powers of any order is an unavoidable regularity:

► **Theorem 1.** *Every infinite word contains powers of any order or anti-powers of any order.*

Formally, an *unavoidable regularity* is a property  $P$  such that it is not possible to construct arbitrarily long words not satisfying  $P$  (cf. [3]). Some important theorems in combinatorics on words concern unavoidable regularities. Most of them follow from results originally stated in other areas of combinatorics, e.g., the Ramsey's, van der Waerden's and Shirshov's theorems (see [3, 5, 6] for further details).

Actually, we prove a stronger result, from which Theorem 1 follows. Given an infinite word  $x$ , we prove that if for some integer  $k$  the lower density of the set of lengths  $n$  for which the prefix of  $x$  of length  $kn$  is a  $k$ -anti-power is smaller than one, then there exists a word (whose length depends on  $k$ ) that occurs in  $x$  with arbitrary exponent (Theorem 4). This implies that if an infinite word  $x$  has the property that each of its factors appears with bounded exponent (in the terminology of combinatorics on words we say that  $x$  is  $\omega$ -power-free), then in  $x$  must start anti-powers of any order at every position. In particular, since a uniformly recurrent word is either purely periodic or  $\omega$ -power-free, this property holds for every aperiodic uniformly recurrent word, as for example the Thue-Morse word or any Sturmian word<sup>2</sup>.

In the second part of the paper, we focus on the avoidability of anti-powers. We show that any infinite word avoiding 3-anti-powers is ultimately periodic, and that there exist aperiodic words avoiding 4-anti-powers. We also show that there exist aperiodic *recurrent* words avoiding 6-anti-powers. We leave it as an open question to determine whether there exist aperiodic recurrent words avoiding 4-anti-powers or 5-anti-powers.

We conclude with final considerations and discuss open problems and further possible directions of investigation.

## 2 Preliminaries

Let  $\mathbb{N} = \{1, 2, 3, \dots\}$ . Let  $\mathbb{A}$  be a (possibly infinite) non-empty set, called the *alphabet*, whose elements are called *letters*. A *word* over  $\mathbb{A}$  is a finite or infinite sequence of letters from  $\mathbb{A}$ . The *length*  $|u|$  of a finite word  $u$  is the number of its letters. We let  $\mathbb{A}^+$  denote the set of all finite words of positive length over  $\mathbb{A}$ , and  $\mathbb{A}^{\mathbb{N}}$  the set of all infinite words over  $\mathbb{A}$ , that is, the

<sup>2</sup> Sturmian words are aperiodic words of minimal factor complexity. They are very well studied objects in combinatorics on words (see for instance [6]).

set of all maps from  $\mathbb{N}$  to  $\mathbb{A}$ . Given a finite word  $u$ , we write  $u^\omega$  the infinite word  $uuu\cdots$  obtained by concatenating an infinite number of copies of  $u$ .

Given a finite or infinite word  $x$ , we say that a word  $u$  is a *factor* of  $x$  if  $x = vuy$  for some words  $v$  and  $y$ . We say that  $u$  is a *prefix* (resp. *suffix*) of  $x$  if  $x = uy$  (resp.  $x = yu$ ) for some word  $y$ . We say that a word  $u \neq x$  is a *border* of  $x$  if  $u$  is both a prefix and a suffix of  $x$ .

An infinite word  $x$  is *purely periodic* if there exists a positive integer  $p$  such that the letters occurring at positions  $i$  and  $j$  coincide whenever  $i = j \pmod p$ . Equivalently,  $x$  is purely periodic if and only if  $x = u^\omega$  for some word  $u$  of length  $p$ . An infinite word  $x$  is *ultimately periodic* if  $x = uy$  for a finite word  $u$  and a purely periodic word  $y$ . An infinite word is *aperiodic* if it is not ultimately periodic.

An infinite word  $x$  is said to be *recurrent* if every finite factor of  $x$  occurs in  $x$  infinitely often. Equivalently,  $x$  is recurrent if and only if every finite prefix of  $x$  has a second occurrence as a factor. An infinite word  $x$  is said to be *uniformly recurrent* if every finite factor of  $x$  occurs syndetically (that is, it occurs infinitely often and with bounded gaps). Equivalently,  $x$  is uniformly recurrent if and only if for every finite factor  $u$  of  $x$  there exists an integer  $m$  such that  $u$  occurs in every factor of  $x$  of length  $m$ .

An infinite word  $x$  is said to be  *$k$ -power-free* for some integer  $k > 1$  if for every finite factor  $u$  of  $x$ , one has that  $u^k$  is not a factor of  $x$ . An infinite word  $x$  is said to be  *$\omega$ -power-free* if for every finite factor  $u$  of  $x$  there exists a positive integer  $l$  such that  $u^l$  is not a factor of  $x$ . Of course, if a word is  $k$ -power-free for some integer  $k$ , then it is  $\omega$ -power-free, but the converse is not always true.

An important relationship between uniformly recurrent and  $\omega$ -power-free words is the following (see for instance [3]):

► **Theorem 2.** *Every uniformly recurrent word is either purely periodic or  $\omega$ -power-free.*

### 3 Unavoidability of powers or anti-powers

In order to state our main result, we need to introduce some definitions.

Let  $x$  be an infinite word and  $k \in \mathbb{N}$ . We set

$$P(x, k) = \{m \in \mathbb{N} \mid \text{the prefix of } x \text{ of length } km \text{ is a } k\text{-power}\}.$$

Analogously, we set

$$AP(x, k) = \{m \in \mathbb{N} \mid \text{the prefix of } x \text{ of length } km \text{ is a } k\text{-anti-power}\}.$$

Note that  $P(x, 1) = AP(x, 1) = \mathbb{N}$  and that  $P(x, k) \cap AP(x, k) = \emptyset$  for every  $k \geq 2$ . For example, if  $x = 01^\omega$ , we have  $P(x, k) = AP(x, k) = \emptyset$  for every  $k \geq 3$ .

Recall that for any subset  $X \subseteq \mathbb{N}$ , the *lower density* of  $X$  is defined by

$$\underline{d}(X) = \liminf_{n \rightarrow \infty} \frac{|X \cap \{1, 2, \dots, n\}|}{n}.$$

Note that if  $X$  is finite, then  $\underline{d}(X) = 0$ . Moreover, if  $\underline{d}(X) = 1/t$  for some integer  $t > 0$ , then that there exist infinitely many integers  $m$  such that  $\{m, m+1, \dots, m+t-1\} \subset \mathbb{N} \setminus X$ .

We are now going to prove our main result (Theorem 4). We premise a technical lemma.

► **Lemma 3.** *Let  $v$  be a border of a word  $w$  and  $u$  the word such that  $w = uv$ . If  $l$  is an integer such that  $|w| > l|u|$ , then  $u^l$  is a prefix of  $w$ .*

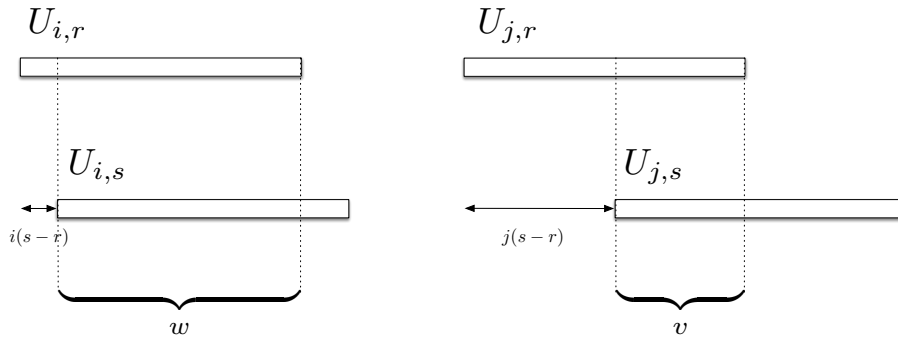


Figure 1 The proof of Theorem 4.

**Proof.** By induction on  $l$ . For  $l = 1$  the statement trivially holds. Suppose  $l > 1$ . Since  $u$  is shorter than  $v$  and both are prefixes of  $w$ , we have that  $u$  is a prefix of  $v$ . Let us write  $v = uv'$ . Then  $w = uvv'$  and  $v'$  is a border of  $v$ . Since  $|v'| = |w| - |u| > (l - 1)|u|$ , we can apply the induction hypothesis and derive that  $u^{l-1}$  is a prefix of  $v$ , whence  $u^l$  is a prefix of  $w$ . ◀

► **Theorem 4.** Let  $x$  be an infinite word. Suppose that

$$\underline{d}(AP(x, k)) < \left(1 + \binom{k}{2}\right)^{-1} = \frac{2}{2 + k(k - 1)}$$

for some  $k \in \mathbb{N}$ . Then there exists  $u \in \mathbb{A}^+$  with  $|u| \leq (k - 1)\binom{k}{2}$  such that  $u^l$  is a factor of  $x$  for every  $l \geq 1$ .

**Proof.** Fix  $k$  such that  $\underline{d}(AP(x, k)) < (1 + \binom{k}{2})^{-1}$ . Since  $AP(x, 1) = \mathbb{N}$ , and the lower density of  $\mathbb{N}$  is 1, we have  $k \geq 2$ . We set  $M = (k - 1)\binom{k}{2}$ . We have to show there exists  $u \in \mathbb{A}^+$  with  $|u| \leq M$  such that  $u^l$  is a factor of  $x$  for every  $l \geq 1$ . By the pigeonhole principle, it suffices to show that for every  $l \in \mathbb{N}$  there exists  $u \in \mathbb{A}^+$  with  $|u| \leq M$  such that  $u^l$  is a factor of  $x$ .

So, let us fix  $l \in \mathbb{N}$ , and set  $N = (l + 1)M$ . Since  $\underline{d}(AP(x, k)) < (1 + \binom{k}{2})^{-1}$ , there exists an integer  $m > N$  such that  $\{m, m + 1, \dots, m + \binom{k}{2}\} \subset \mathbb{N} \setminus AP(x, k)$ .

For every  $j$  and  $r$  such that  $0 \leq j \leq k - 1$  and  $m \leq r \leq m + \binom{k}{2}$ , set

$$U_{j,r} = x_{jr+1}x_{jr+2} \cdots x_{(j+1)r},$$

so that  $|U_{j,r}| = r$  and  $U_{0,r}, U_{1,r}, \dots, U_{k-1,r}$  are the first  $k$  consecutive blocks of  $x$  of length  $r$ . Thus for each  $m \leq r \leq m + \binom{k}{2}$  there exist  $i$  and  $j$ , with  $0 \leq i < j \leq k - 1$ , such that  $U_{i,r} = U_{j,r}$ . By the pigeonhole principle, there exist  $r$  and  $s$ , with  $m \leq r < s \leq m + \binom{k}{2}$ , and  $i$  and  $j$ , with  $0 \leq i < j \leq k - 1$ , such that  $U_{i,r} = U_{j,r}$  and  $U_{i,s} = U_{j,s}$ .

Notice that  $(i + 1)r > is + 1$  and  $(j + 1)r > js + 1$ .

Let us now set  $w = x_{is+1}x_{is+2} \cdots x_{(i+1)r}$  and  $v = x_{js+1}x_{js+2} \cdots x_{(j+1)r}$  (see Figure 1). We have

$$|v| = (j + 1)r - js < (i + 1)r - is = |w|,$$

whence  $v$  is a border of  $w$ . Writing  $w = uv$ , we have

$$1 \leq |u| = |w| - |v| = (j - i)(s - r) \leq (k - 1)\binom{k}{2} = M,$$

and

$$|w| > |v| = r - j(s - r) \geq m - (k - 1) \binom{k}{2} = m - M > N - M = lM.$$

Thus,  $|w| > l|u|$  and, by Lemma 3,  $u^l$  is a prefix of  $w$ , and therefore  $u^l$  is a factor of  $x$ . ◀

A more precise consequence of Theorem 4 is the following.

► **Corollary 5.** *Let  $x$  be a uniformly recurrent word. Then either*

$$\underline{d}(AP(x, k)) \geq \frac{2}{2 + k(k - 1)} \tag{1}$$

for every  $k \in \mathbb{N}$  or there exists  $r > 0$ , such that

$$\underline{d}(P(x, k)) \geq r \tag{2}$$

for every  $k \in \mathbb{N}$ .

**Proof.** According to Theorem 4, if (1) does not hold for some  $k' \in \mathbb{N}$ , then  $x = u^\omega$  for some  $u$  with  $1 \leq |u| \leq (k' - 1) \binom{k'}{2}$ . Whence  $n|u| \in P(x, k)$  for each  $n, k \in \mathbb{N}$ . The result now follows by setting  $r = 1/|u|$ . ◀

Note that the  $\underline{d}(P(x, k)) \geq r$  for every  $k$  is stronger than just  $\underline{d}(P(x, k)) > 0$ . Conversely, if  $\underline{d}(P(x, k)) > 0$  for some  $k \geq 2$ , then  $\underline{d}(P(x, 2)) > 0$ , and from this it is immediate to see that  $x$  is periodic. In fact, something stronger is true: following the notation in the proof of Theorem 4, if there exists  $j \geq 1$  such that  $\underline{d}\{r \mid U_{0,r} = U_{j,r}\} > 0$ , then  $x$  is periodic. And  $\underline{d}(P(x, 2)) > 0$  is a special case of this assumption (when  $j = 1$ ).

► **Corollary 6.** *Let  $x$  be a uniformly recurrent word. If*

$$\underline{d}(AP(x, k)) < \frac{2}{2 + k(k - 1)}$$

for some  $k \in \mathbb{N}$ , then  $x$  is purely periodic.

Another direct consequence of Theorem 4 is the following.

► **Theorem 7.** *Let  $x$  be an infinite word. If  $x$  is  $\omega$ -power-free, then at every position of  $x$  start anti-powers of any order.*

**Proof.** Suppose that there exists a positive integer  $k$  and a suffix  $x'$  of  $x$  such that no prefix of  $x'$  is a  $k$ -anti-power. Then  $AP(x', k) = \emptyset$ , whence  $\underline{d}(AP(x', k)) = 0$ . By Theorem 4, there exists a factor  $u$  of  $x'$  such that  $u^l$  is a factor of  $x'$  for every  $l \geq 1$ , hence  $x$  is not  $\omega$ -power-free. ◀

From Theorems 2 and 7, we derive the following corollary.

► **Corollary 8.** *Let  $x$  be a uniformly recurrent aperiodic word. Then at every position of  $x$  start anti-powers of any order.*

## 4 Avoiding anti-powers

In this section we deal with avoidability of anti-powers.

► **Definition 9.** Given  $k > 1$ , we say that an infinite word  $x$  avoids  $k$ -anti-powers if no factor of  $x$  is a  $k$ -anti-power. That is, among any  $k$  consecutive blocks of the same length in  $x$ , at least two of them are equal. We say that an infinite word  $x$  avoids anti-powers if  $x$  avoids  $k$ -anti-powers for some  $k$ .

Periodic words avoid anti-powers, the period length being an upper bound for the maximal number of distinct consecutive blocks of the same length. In the following, we discuss the avoidability of anti-powers for aperiodic words. By Corollary 8, if an aperiodic word avoids anti-powers, then it cannot be uniformly recurrent.

Of course, any word containing at least two different letters cannot avoid 2-anti-powers. For 3-anti-powers, we have the following result.

► **Lemma 10.** *Let  $x$  be an infinite word. If  $x$  avoids 3-anti-powers, then  $x$  is a binary word.*

**Proof.** Suppose  $x$  avoids 3-anti-powers and contains three different letters. Then there is a factor of  $x$  of the form  $u = ab^n c$  with  $n \geq 1$  and  $a, b, c$  distinct letters. We will extend this factor to the right and force a 3-anti-power for every  $n$ . For  $n = 1$ , the word  $abc$  is already an anti-power. Take now  $n = 2$ . To avoid 3-anti-powers,  $abbc$  can only be extended to  $abccb$ . In the next step, the only option is  $abcbcb$ , and after that  $abcbcbcb$ . But now, the word  $abcbcbcbbyy'$  contains a 3-anti-power for every letters  $y, y'$ . Suppose now  $u = ab^n c$  with  $n \geq 3$ . If  $n$  is odd, let  $m = (n - 1)/2$  and note that  $u$  can be factored as  $ab^m \cdot b^{m+1} \cdot c$ , so that  $u$  will be extended to the right to a 3-anti-power of length  $3(m + 1)$ . If  $n$  is even,  $u$  can be factored as  $u = ab^m \cdot b^{m+1} \cdot bc$ , so that again  $u$  will be extended to the right to a 3-anti-power of length  $3(m + 1)$ . ◀

Hence, in what follows we will suppose that  $x$  is an infinite word over the binary alphabet  $\mathbb{A} = \{0, 1\}$ .

► **Proposition 11.** *Let  $x$  be an infinite word. If  $x$  avoids 3-anti-powers, then it cannot contain a factor of the form  $10^n 1$  or  $01^n 0$  with  $n > 1$ .*

**Proof.** Suppose that  $x$  contains a factor of the form  $u = 10^n 1$  with  $n > 1$  (the other situation is symmetric). The cases  $n = 2, 3, 4, 5$  can be checked by computer, so let us suppose  $n \geq 6$ .

Suppose first  $n$  even, and write  $n = 2m$ . Since  $u = 10^{m-1} \cdot 0^m \cdot 01$ , any extension of  $u$  to the right will produce a 3-anti-power of length  $3m$ . If  $n$  is odd,  $n = 2m + 1$ , then we can write  $u = 10^{m-1} \cdot 0^m \cdot 001$ , so that any extension of  $u$  to the right will produce a 3-anti-power of length  $3m$ . ◀

► **Corollary 12.** *Let  $x$  be an infinite word avoiding 3-anti-powers. Then  $x$  is ultimately periodic.*

Actually, from Proposition 11, we have that an infinite word avoiding 3-anti-powers can only be of the form  $x = (01)^\omega$ ,  $x = 01^\omega$ , or  $x = 0^n 10^\omega$  for some  $n > 0$ , up to exchanging letters.

► **Proposition 13.** *There exist aperiodic words avoiding 4-anti-powers.*

**Proof.** We exhibit an example of an aperiodic word avoiding 4-anti-powers. Let  $(\alpha_i)_{i \geq 1}$  be an increasing sequence of positive integers with  $\alpha_{i+1} \geq 5\alpha_i$  for each  $i \geq 1$ . Let  $x \in \{0, 1\}^\mathbb{N}$

be defined by  $x_n = 1$  if  $n = \alpha_i$  for some  $i \geq 1$ , and  $x_n = 0$  otherwise. Clearly  $x$  is aperiodic. Moreover, given  $m \geq 0$  and  $n \in \mathbb{N}$ , if  $|x_{m+1}x_{m+2} \cdots x_{m+n}|_1 \geq 2$ , then for some  $i \geq 1$

$$m + 1 \leq \alpha_i < 5\alpha_i \leq \alpha_{i+1} \leq m + n$$

and hence  $n > 4\alpha_i \geq 4(m + 1)$  whence  $m + 1 < n/4$ . We claim that  $x$  avoids 4-anti-powers. In fact, suppose to the contrary that for some  $m \geq 0$  and  $n \in \mathbb{N}$  we have  $x_{m+1} \cdots x_{m+n}$ ,  $x_{m+n+1} \cdots x_{m+2n}$ ,  $x_{m+2n+1} \cdots x_{m+3n}$ , and  $x_{m+3n+1} \cdots x_{m+4n}$  are pairwise distinct. Then at least three of the four blocks must contain an occurrence of 1. Thus  $|x_{m+n+1} \cdots x_{m+4n}|_1 \geq 2$  from which it follows that  $m + n + 1 < 3n/4$  and hence  $m + 1 < 0$ , a contradiction. ◀

The word in the previous proposition is not recurrent. It is natural to ask whether there exist recurrent words avoiding 4-anti-powers. We do not know the answer. However, we can state the following result.

► **Proposition 14.** *There exist aperiodic recurrent words avoiding 6-anti-powers.*

**Proof.** We exhibit an example of an aperiodic recurrent word avoiding 6-anti-powers. Let  $w_0 = 0$  and  $w_n = w_{n-1}1^{3|w_{n-1}|}w_{n-1}$  for every  $n > 0$ . Let  $w$  be the infinite word obtained as the limit of the sequence of words  $(w_n)_{n \geq 1}$ . Then clearly  $w$  is recurrent. Without loss of generality, we can assume that each occurrence of  $w_n$  in  $w$  is preceded and followed by  $1^{3|w_n|}$ , since  $w$  avoids 6-anti-powers if  $1^\infty w$  does.

Let  $v = v_1v_2 \cdots v_6$  be a non-empty factor of  $w$  of length  $6k$ . Let  $n$  be the largest integer such that  $|w_n| = 5^n < 2k$ . By the hypothesis on  $n$ , no  $v_i$  can intersect two occurrences of  $w_n$ .

Suppose first that for some  $i$ ,  $v_i$  is contained as a factor in  $w_n$ . By the hypothesis on  $n$ , neither  $v_{i-1}v_i$  nor  $v_iv_{i+1}$  is contained in  $w_n$ . Since  $w_n$  is preceded and followed by  $1^{3|w_n|}$ , either  $v_{i-3}$  and  $v_{i-2}$  (if  $i \geq 4$ ) or  $v_{i+2}$  and  $v_{i+3}$  (if  $i < 4$ ) are both equal to  $1^k$ , so that  $v$  cannot be an anti-power.

If instead no  $v_i$  is contained as a factor in  $w_n$ , then one of the following cases must hold:

- (i) There is an occurrence of  $w_n$  intersecting  $v_i$  and the next occurrence of  $w_n$  intersects  $v_{i+1}$ . In this case, either  $v_{i-3}$  and  $v_{i-2}$  or  $v_{i+3}$  and  $v_{i+4}$  are both equal to  $1^k$ .
- (ii) There is an occurrence of  $w_n$  intersecting  $v_i$  and the next occurrence of  $w_n$  intersects  $v_{i+2}$ , so that  $v_{i+1} = 1^k$ . In this case, either  $v_{i-2}$  or  $v_{i+4}$  must be equal to  $1^k$ .
- (iii) There are two consecutive blocks  $v_i, v_{i+1}$  both equal to  $1^k$ .

In all cases,  $v$  cannot be an anti-power. ◀

## 5 Conclusions and open problems

We proved that every infinite word contains powers of any order or anti-powers of any order, that is, the existence of powers or anti-powers is an unavoidable regularity. This result can also be stated in the following finite version.

► **Theorem 15.** *For every integers  $k > 1$  and  $r > 1$  there exists  $N = N(k, r)$  such that every word of length  $N$  contains a  $k$ -power or an  $r$ -anti-power. Furthermore, for  $k > 2$ , one has  $k^2 - 2 \leq N(k, k) \leq k^5 + k^3$ .*

The upper bound follows from the proof of Theorem 4. For the lower bound, it is easy to prove that the word  $(0^{k-1}1)^{k-2}0^{k-2}10^{k-2}$  of length  $k^2 - 3$  avoids both  $k$ -powers and  $k$ -anti-powers, for any  $k > 2$ .

In the case of binary alphabet, it can be verified that  $N(2, 2) = 2$ ,  $N(3, 2) = 3$ ,  $N(2, 3) = 4$ ,  $N(3, 3) = 9$ ,  $N(4, 3) = 12$ ,  $N(3, 4) > 16$ ,  $N(4, 4) > 16$ . We do not know how these numbers



grow. Moreover, the bounds on  $N(k, r)$  given in Theorem 15 can probably be improved by a deeper analysis of the function  $N(k, r)$ .

Concerning the avoidability of anti-powers, we proved that there exist words avoiding 4 anti-powers and that there exist recurrent words avoiding 6-anti-powers. A natural problem is therefore that of determining what is the least  $k$  such that there exists a recurrent word avoiding  $k$ -anti-powers.

Another possible direction of investigation is related to the possible lengths of anti-powers appearing in a word. For an aperiodic uniformly recurrent word  $x$ , define  $ap(x, k) = \min(AP(x, k))$ , i.e., the minimum length  $m$  for which the prefix of  $x$  of length  $km$  is a  $k$ -anti-power. The first values of this function for the Thue-Morse word are displayed in Table 1 (where the value of  $ap(x, k)$  is the ratio between the length of the  $k$ -anti-power prefix and the order  $k$ ). We wonder whether it is possible to link the behavior of the function  $ap(x, k)$  to the combinatorics of the word  $x$ , at least for special classes of words. For example, the values reported in Table 1 suggest that for the Thue-Morse word the function  $ap(x, k)$  grows linearly in  $k$ .

**Acknowledgements.** We thank the anonymous referees for their careful reading of the paper and Filippo Mignosi for useful discussions.

---

#### References

- 1 J.-P. Allouche and J. Shallit. The ubiquitous prouhet-thue-morse sequence. In *Sequences and their Applications: Proceedings of SETA'98*, pages 1–16. Springer, 1999.
- 2 A. Condon, J. Mañuch, and C. Thachuk. The complexity of string partitioning. *J. Discrete Algorithms*, 32(C):24–43, May 2015. doi:10.1016/j.jda.2014.11.002.
- 3 A. de Luca and S. Varricchio. *Finiteness and Regularity in Semigroups and Formal Languages*. Monographs in Theoretical Computer Science (An EATCS Series). Springer, 1st edition, 1999.
- 4 H. Fernau, F. Manea, R. Mercas, and M. L. Schmid. Pattern Matching with Variables: Fast Algorithms and New Hardness Results. In *32nd International Symposium on Theoretical Aspects of Computer Science (STACS 2015)*, volume 30 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 302–315, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.STACS.2015.302.
- 5 M. Lothaire. *Combinatorics on Words, 2nd edition*. Cambridge Mathematical Library, Cambridge University Press, 1997.
- 6 M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.
- 7 M. L. Schmid. Computing equality-free and repetitive string factorisations. *Theoret. Comput. Sci.*, 618:42–51, March 2016. doi:10.1016/j.tcs.2016.01.006.
- 8 A. Thue. Uber unendliche zeichenreihen. *Kra. Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl.*, pages 1–22, 1906.