(k, λ)-anti-powers and other patterns in words

Amanda Burcroff
Department of Mathematics
University of Michigan
Ann Arbor, MI, USA
burcroff@umich.edu

Submitted: Jul 27, 2018; Accepted: Oct 23, 2018; Published: Nov 30, 2018

Abstract

Given a word, we are interested in the structure of its contiguous subwords split into k blocks of equal length, especially in the homogeneous and anti-homogeneous cases. We introduce the notion of $(\mu_1, \ldots, \mu_k)$-block-patterns, words of the form $w = w_1 \cdots w_k$ where, when $\{w_1, \ldots, w_k\}$ is partitioned via equality, there are μ_s sets of size s for each $s \in \{1, \ldots, k\}$. This is a generalization of the well-studied k-powers and the k-anti-powers recently introduced by Fici, Restivo, Silva, and Zamboni, as well as a refinement of the (k, λ)-anti-powers introduced by Defant. We generalize the anti-Ramsey-type results of Fici et al. to $(\mu_1, \ldots, \mu_k)$-block-patterns and improve their bounds on $N_\alpha(k, k)$, the minimum length such that every word of length $N_\alpha(k, k)$ on an alphabet of size α contains a k-power or k-anti-power. We also generalize their results on infinite words avoiding k-anti-powers to the case of (k, λ)-anti-powers. We provide a few results on the relation between α and $N_\alpha(k, k)$ and find the expected number of $(\mu_1, \ldots, \mu_k)$-block-patterns in a word of length n.

Mathematics Subject Classifications: 05A05, 68R15

1 Introduction

In 1975, Erdős, Simonivits, and Sós [4] introduced anti-Ramsey theory, the idea that sufficiently large partitioned structures cannot avoid anti-homogeneous substructures. Their investigation was initially graph-theoretic, but with time anti-Ramsey-type results have permeated many areas of combinatorics, including the studies of Sidon sets, canonical Ramsey theory, and the spectra of colorings [1, 10, 12]. The study of homogeneous and anti-homogeneous substructures can also be extended to *words*, finite or infinite (to the right) sequences of letters from a fixed alphabet. The substructures of interest are contiguous subwords, known as *factors.* A well-studied type of regularity in words concerns *k-powers*, that is, words of the form $u^k = uu \cdots u$ (concatenated k times) for some nonempty word u (see, for example, [7]). Recently Fici et al. [5] introduced a notion of anti-regularity in words through their definition of *k-anti-powers*.

Definition 1. Let $|u|$ denote the length of a word u. A *k-anti-power* is a word w of the form

$$w = w_1 w_2 \cdots w_k$$

such that $|w_1| = \cdots = |w_k|$ and $w_1, \ldots, w_k$ are distinct.

Fici et al. [5] were able to show several properties of anti-powers in words, including anti-Ramsey results concerning the existence of ℓ-powers or k-anti-powers. Gaetz [6] showed that the minimum $m_{k,j} > 0$ for which the factor of length $km_{k,j}$ beginning at the j^{th} index of the famous Thue-Morse word t is a k-anti-power, grows linearly in k; Defant [3] and Narayanan [9] studied the case $j = 0$. Defant also introduced the notion of (k, λ)-anti-powers, which are a generalization of k-anti-powers.

Definition 2. A (k, λ)-*anti-power* is a word w of the form

$$w = w_1 w_2 \cdots w_k$$

such that $|w_1| = \cdots = |w_k|$ and $|\{i : w_i = w_j\}| \leqslant \lambda$ for each fixed $j \in \{1, \ldots, k\}$.

Note that when $\lambda = 1$, this is precisely the definition of a k-anti-power. Whenever such a generalization is nontrivial, we prove that the results of Fici et al. in [5] concerning k-anti-powers generalize to the case of (k, λ)-anti-powers. In fact, many of these results can be strengthened by enforcing a particular structure on the partition of the blocks by equality. We generalize the notions of k-powers and k-anti-powers while refining the (k, λ)-anti-powers through the introduction of $(\mu_1, \ldots, \mu_k)$-*block-patterns*.

Definition 3. Let $\mu_1, \ldots, \mu_k$ be nonnegative integers satisfying $\sum_{s=1}^{k} s\mu_s = k$. A $(\mu_1, \cdots, \mu_k)$-*block-pattern* is a word of the form $w = w_1 \cdots w_k$ where, if the set $\{1, \ldots, k\}$ is partitioned via the rule $i \sim j \iff w_i = w_j$, there are μ_s parts of size s for all $1 \leqslant s \leqslant k$.

For example, 10 01 00 01 10 is a $(1, 2, 0, 0, 0)$-block-pattern. Let $P_{k,\leqslant\lambda}$ denote the set of k-tuples of natural numbers $(\mu_1, \ldots, \mu_k)$ such that $\sum_{s=1}^{k} s\mu_s = k$ and $\mu_s = 0$ for $s > \lambda$. These correspond to the partitions of k such that each part has size at most λ. We can relate $(\mu_1, \ldots, \mu_k)$-block-patterns to (k, λ)-anti-powers via the following observation.

Remark 4. Let $\mathcal{AP}_{\mathbb{A}}(k, \lambda)$ be the set of (k, λ)-anti-powers on an alphabet $\mathbb{A}$. Let $\mathcal{BP}_{\mathbb{A}}(\mu_1, \ldots, \mu_k)$ be the set of $(\mu_1, \ldots, \mu_k)$-block-patterns on $\mathbb{A}$. Then

$$\mathcal{AP}_{\mathbb{A}}(k, \lambda) = \bigcup_{(\mu_1, \ldots, \mu_k) \in P_{k,\leqslant\lambda}} \mathcal{BP}_{\mathbb{A}}(\mu_1, \ldots, \mu_k).$$

In particular, the k-anti-powers are precisely the $(k, 0, \ldots, 0)$-block-patterns, and moreover the k-powers are precisely the $(0, \ldots, 0, 1)$-block-patterns.

The generalizations of the anti-Ramsey results of Fici et al. in [5] to the case of $(\mu_1, \dots, \mu_k)$-block-patterns are the focus of Section 3. In particular, we obtain bounds on the sizes of words avoiding powers or block-patterns with at most σ pairs of equal blocks. In Section 4, we generalize the results of [5] on avoiding k-anti-powers in infinite words to (k, λ)-anti-powers. We also observe that Sturmian words have anti-powers of every order starting at each index.

A slight strengthening of the arguments of Fici et al. in [5] also provide better bounds for $N_\alpha(k,k)$, the smallest positive integer such that every word of length $N_\alpha(k,k)$ over an alphabet of size α contains a k-power or k-anti-power. Namely, it is shown in [5] that for $k > 2$, $k^2 - 1 \leqslant N_\alpha(k,k) \leqslant k^3\binom{k}{2}$. In Section 5, we improve both the lower and upper bounds according to the following theorem.

Theorem 27. For any $k > 3$,

$$2k^2 - 2k \leqslant N_\alpha(k,k) \leqslant (k^3 - k^2 + k)\binom{k}{2}.$$

We also investigate how the size of the alphabet affects $N_\alpha(k,k)$ in Section 5. In Section 6, we return to the more general setting and compute the expected number of $(\mu_1, \dots, \mu_k)$-block-patterns in a word of length n.

2 Preliminaries

Let $\mathbb{N} = \{1, 2, 3, \dots\}$. The i^{th} letter of a word x is denoted $x[i]$, and for $i < j$ the contiguous substring beginning at the i^{th} letter and ending with the j^{th} is denoted $x[i..j]$. A word v is a *factor* of x if $x = uvw$ for words u and w. In the case that u is empty, v is a *prefix* of x, and if w is empty, then v is a *suffix* of x. The suffix of x beginning at the j^{th} index of x is denoted $x_{(j)}$. If w is both a prefix and suffix of x, then w is a *border* of x.

A word is called *recurrent* if every finite factor appears infinitely many times in the word. A word x is called *uniformly recurrent* if for every finite factor u, there exists a positive integer n_u such that u appears in every factor of x of length n_u. A word x is called *eventually periodic* if there exists an index $j \geqslant 0$ and a finite word u such that $x_{(j)} = u^\omega$; otherwise x is called *aperiodic*. A word is called *ω-power-free* if for every finite factor u, there exists an $\ell \in \mathbb{N}$ such that u^ℓ is not a factor. Note that a word that avoids k-powers for some $k \in \mathbb{N}$ is ω-power-free, but the converse is not necessarily true.

Let $[\alpha] = \{1, \dots, \alpha\}$. The *lower density* and *upper density* of a subset S of $\mathbb{N}$ are given respectively by

$$\underline{d}(S) = \liminf_{n\to\infty} \frac{|S \cap [n]|}{n} \quad \text{and} \quad \overline{d}(S) = \limsup_{n\to\infty} \frac{|S \cap [n]|}{n}.$$

3 Generalization of an Anti-Ramsey result to $(\mu_1, \ldots, \mu_k)$-block-patterns

A main result of Fici et al. [5] is that every infinite word contains either powers of all orders or anti-powers of all orders. Since powers are homogeneous substructures whereas the anti-powers are anti-homogeneous, one may wonder if similar results can be demonstrated for substructures between these extremes. We will generalize their result to the case of $(\mu_1, \ldots, \mu_k)$-block-patterns in infinite words. The density bounds rely on the number of pairs of equal blocks that are forced in the prefixes of length $km, \ldots, k(m+\beta)$ for some m, β. The following definition is created to account for these pairs.

Definition 5. Let $D(x, k, \sigma)$ be the set of $m \in \mathbb{N}$ such that the prefix of the word x of length km is a $(\mu_1, \ldots, \mu_k)$-block-pattern satisfying $\sum_{s=1}^{k} \mu_s \binom{s}{2} \leqslant \sigma$.

Note that $D(x, k, \sigma)$ is closed downward with respect to the dominance order. That is, if $m \in D(x, k, \sigma)$ and $m' \in \mathbb{N}$ are such that $x[1..km]$ is a $(\mu_1, \ldots, \mu_k)$-block-pattern and $x[1..km']$ is a $(\mu'_1, \ldots, \mu'_k)$-block-pattern satisfying $\sum_{s=1}^{\ell} \mu_s \geqslant \sum_{s=1}^{\ell} \mu'_s$ for each $\ell \in \{1, \ldots, k\}$, then $m' \in D(x, k, \sigma)$.

For the proof of Theorem 7, we make use of the following lemma of Fici, Restivo, Silva, and Zamboni.

Lemma 6. ([5], Lemma 3) *Let v be a border of a word w and let u be the word such that $w = uv$. If ℓ is an integer such that $|w| \geqslant \ell|u|$, then u^ℓ is a prefix of w.*

Theorem 7. *Let x be an infinite word such that*

$$\overline{d}(D(x, k, \sigma)) \geqslant \left(1 + \left\lfloor \frac{1}{\sigma}\binom{k}{2} \right\rfloor\right)^{-1}$$

for some $k, \sigma \in \mathbb{N}$. For every ℓ, there is a word u with $|u| \leqslant (k-1)\left\lfloor \frac{1}{\sigma}\binom{k}{2}\right\rfloor$ such that u^ℓ is a factor of x.

Proof. Fix such a k and σ. Fix an arbitrary $\ell \geqslant 1$, and let $\beta = \left\lfloor \frac{1}{\sigma}\binom{k}{2}\right\rfloor$. By the condition on the upper density of $D(x, k, \sigma)$, there exists some integer $m > \ell(k-1)\beta$ such that $\{m, m+1, \ldots, m+\beta\} \subset D(x, k, \sigma)$. Following [5], for every $j \in \{0, \ldots, k-1\}$ and $r \in \{m, \ldots, m+\beta\}$, set

$$U_{j,r} = x[jr+1..(j+1)r].$$

That is, $U_{0,r} \cdots U_{k-1,r} = x[1..kr]$. Since $\{m, m+1, \ldots, m+\beta\} \subset D(x, k, \sigma)$, we are guaranteed at least $(\beta+1)\sigma > \binom{k}{2}$ triples (i, j, r) such that $i < j$ and $U_{i,r} = U_{j,r}$. By the Pigeonhole Principle, there exist i, j, r, s such that $0 \leqslant i < j \leqslant k-1$, $m \leqslant r < s \leqslant \beta+1$, $U_{i,r} = U_{j,r}$, and $U_{i,s} = U_{j,s}$.

Setting $w = x[is+1..(i+1)r]$ and $v = x[js+1..(j+1)r]$, we have

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

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

$$1 \leqslant |u| = |w| - |v| = (j-i)(s-r) \leqslant (k-1)\beta$$

while

$$|v| = r - j(s-r) \geqslant m - (k-1)\beta \geqslant \ell(k-1)\beta \geqslant (\ell-1)|u|.$$

Hence, $|w| = |u| + |v| \geqslant \ell|u|$. By Lemma 6, u^ℓ is a factor of x. □

Theorem 7 can be applied to the special case of (k,λ)-anti-powers. The definition of (k,λ)-anti-powers suggests the following generalization of $\mathrm{AP}(x,k)$, the set of integers m such that the prefix of x of length km is a k-anti-power.

Definition 8. Let $\mathrm{AP}(x,k,\lambda)$ be the set of $m \in \mathbb{N}$ such that the prefix of the word x of length km is a (k,λ)-anti-power.

Note that $\mathrm{AP}(x,k,1) = \mathrm{AP}(x,k)$.

Corollary 9. *Let x be an infinite word such that*

$$\underline{d}(\mathrm{AP}(x,k,\lambda)) < \left(1 + \left\lfloor \frac{k^2-k}{\lambda^2+\lambda} \right\rfloor\right)^{-1}$$

for some $k,\lambda \in \mathbb{N}$. For every ℓ, there is a word u with $|u| \leqslant (k-1)\left\lfloor \frac{k^2-k}{\lambda^2+\lambda} \right\rfloor$ such that u^ℓ is a factor of x.

Proof. Fix k and λ as above. Note that $\mathbb{N} \setminus \mathrm{AP}(x,k,\lambda) \subseteq D\left(x,k,\binom{\lambda+1}{2}\right)$. Hence,

$$\underline{d}(\mathrm{AP}(x,k,\lambda)) < \left(1 + \left\lfloor \frac{k^2-k}{\lambda^2+\lambda} \right\rfloor\right)^{-1}$$

implies

$$\overline{d}\left(D\left(x,k,\binom{\lambda+1}{2}\right)\right) \geqslant \frac{\left\lfloor \frac{k^2-k}{\lambda^2+\lambda} \right\rfloor}{1+\left\lfloor \frac{k^2-k}{\lambda^2+\lambda} \right\rfloor} \geqslant \frac{1}{1+\left\lfloor \binom{\lambda+1}{2}^{-1}\binom{k}{2} \right\rfloor}.$$

This shows that x satisfies the conditions of Theorem 7 for the same k and $\sigma = \binom{\lambda+1}{2}$. □

In the case that our alphabet is finite, there are finitely many factors of length at most $(k-1)\left\lfloor \frac{k^2-k}{\lambda^2+\lambda} \right\rfloor$. Thus, the Pigeonhole Principle allows us to choose a word u that works for every ℓ in Theorem 9.

Corollary 10. *Let x be an infinite word on a finite alphabet such that*

$$\underline{d}(\mathrm{AP}(x,k,\lambda)) < \left(1 + \left\lfloor \frac{k^2-k}{\lambda^2+\lambda} \right\rfloor\right)^{-1}$$

for some $k,\lambda \in \mathbb{N}$. There is a word u with $|u| \leqslant (k-1)\left\lfloor \frac{k^2-k}{\lambda^2+\lambda} \right\rfloor$ such that u^ℓ is a factor of x for every $\ell > 0$. In particular, x is not ω-power-free.

There is a $\lambda = 1$ analogue to Corollary 9 in [5] (their Theorem 4), which claims under the same density condition that x is not ω-power-free. Though the condition that the alphabet is finite is not explicitly stated, their result is false for infinite alphabets. In fact, there exist ω-power-free words which avoid k-anti-power prefixes for some fixed $k \in \mathbb{N}$. These words also show that Theorem 6 of [5], which states that ω-power-free words have anti-powers of every order beginning at each index, is false when infinite alphabets are allowed. Theorem 11 provides a counterexample to Theorems 4 and 6 in [5] when infinite alphabets are permitted.

Theorem 11. *There exists an ω-power-free word x on an infinite alphabet such that $\mathrm{AP}(x, k)$ is empty for some $k \in \mathbb{N}$.*

Proof. Let $y = \prod_{i=1}^{\infty} (a_i)^{2^i}$. Since there are finitely many appearances of each letter a_i, y is clearly ω-power-free. Note that if $2^{i+1} - 2^i = 2^i \geqslant 4m$ for some block length m and some i satisfying $2^{i+1} < km$, then two blocks of the prefix of length km must equal a_i^m. Hence, $m \notin \mathrm{AP}(x, k)$. For $k \geqslant 17$, such an i always exists. $\mathrm{AP}(x, k)$ is empty for $k \geqslant 17$, despite x being ω-power-free. $\square$

We return to a modified version of the proof of Theorem 7 in order to find bounds on the length of words avoiding k-powers and k-anti-powers.

Theorem 12. *For all integers $\ell > 1, k > 1, \sigma \geqslant 1$ there exists $N'_\alpha(\ell, k, \sigma)$ such that every word of length $N'_\alpha(\ell, k, \sigma)$ on $[\alpha]$ contains an ℓ-power or $(\mu_1, \dots, \mu_k)$-block-pattern satisfying $\sum_{s=1}^{k} \mu_s \binom{s}{2} \leqslant \sigma$. Moreover,*

$$k\left(k - \left\lfloor \frac{1}{2}(\sqrt{8\sigma + 1} + 1)\right\rfloor\right) \leqslant N'_\alpha(k, k, \sigma) \leqslant \left\lfloor \frac{1}{\sigma}\binom{k}{2}\right\rfloor (k^3 - k^2 + k).$$

Proof. As in [5], the upper bound follows from the proof of the infinite case in Theorem 7. Let $\beta = \left\lfloor \frac{1}{\sigma}\binom{k}{2}\right\rfloor$. Let x be any word of length $\beta(k^3 - k^2 + k)$. For each $r \in \{(k^2 - k)\beta, \dots, (k^2 - k + 1)\beta\}$, consider the first k consecutive blocks of length r in x, denoted by $U_{0,r}, U_{1,r}, \dots, U_{k-1,r}$. If x does not contain any element of $D(x, k, \sigma)$, then there exist i, j, r, s such that $0 \leqslant i < j \leqslant k - 1$, $m \leqslant r < s \leqslant \beta + 1$, $U_{i,r} = U_{j,r}$ and $U_{i,s} = U_{j,s}$. Setting $w = x[is + 1..(i + 1)r]$ and $v = x[js + 1..(j + 1)r]$, we have that v is a border of w. Writing $w = uv$, we have $|u| \leqslant (k - 1)\beta$ and

$$|w| = |u| + |v| \geqslant |u| + r - j(s - r) \geqslant |u| + (k - 1)^2\beta \geqslant k|u|.$$

By Lemma 6, we get that u^k is a factor of x, i.e., x contains a k-power. The length of x is chosen to accommodate k blocks of size at most $(k^2 - k + 1)\beta$.

The lower bound is proven via a construction; we will show that the word

$$x = 0^{k-1}(10^{k-1})^{k - \left\lfloor \frac{1}{2}(\sqrt{8\sigma+1}+1)\right\rfloor - 1}$$

avoids k-powers and $(\mu_1, \dots, \mu_k)$-block-patterns with $\sum_{s=1}^{k} \mu_s \binom{s}{2} \leqslant \sigma$. Since $\sigma \geqslant 1$, we have

$$k - \left\lfloor \frac{1}{2}(\sqrt{8\sigma + 1} + 1)\right\rfloor - 1 \leqslant k - 1.$$

If u^k were a factor of x, either u would contain the letter 1, contradicting the fact that that x has at most $k-1$ copies of the letter 1, or $u = 0^m$ for some $m \geqslant 1$, contradicting the fact that x has no factor equal to 0^k. Hence, x avoids k-powers. We can see that for every factor v of length km, at least $\lfloor \frac{1}{2}(\sqrt{8\sigma+1}+1) \rfloor + 1$ blocks of v are equal to 0^m. v is a $(\mu_1, \ldots, \mu_k)$-block-pattern with

$$\sum_{s=1}^{k} \mu_s \binom{s}{2} \geqslant \binom{\lfloor \frac{1}{2}(\sqrt{8\sigma+1}+1) \rfloor + 1}{2} > \frac{1}{8}(\sqrt{8\sigma+1}+1)(\sqrt{8\sigma+1}) \geqslant \sigma. \qquad \square$$

We can specialize Theorem 12 to the case of (k, λ)-anti-powers.

Corollary 13. *For all integers $\ell > 1, k > 1, \lambda \geqslant 1$, there exists $N_\alpha(\ell, k, \lambda)$ such that every word of length $N_\alpha(\ell, k, \lambda)$ on $[\alpha]$ contains an ℓ-power or (k, λ)-anti-power. Moreover,*

$$k(k-\lambda) \leqslant N_\alpha(k, k, \lambda) \leqslant \left\lfloor \frac{k^2-k}{\lambda^2+\lambda} \right\rfloor (k^3-k^2+k).$$

Proof. If a word avoids (k, λ)-anti-powers, then it avoids $(\mu_1, \ldots, \mu_k)$-block-patterns with $\sum_{s=1}^{k} \mu_s \binom{s}{2} \leqslant \binom{\lambda+1}{2}$. Applying Theorem 12 with $\sigma = \binom{\lambda+1}{2}$ yields the corresponding bounds. $\square$

In particular, this improves upon the upper bound for $N_\alpha(k, k)$ (in their notation, $N(k, k)$) in [5].

Corollary 14. *For all $k > 1$,*

$$N_\alpha(k, k) \leqslant (k^3-k^2+k)\binom{k}{2}.$$

4 Avoiding Anti-Powers

This section is devoted to generalizing the results of Fici et al. [5] on infinite words avoiding k-anti-powers to the case of (k, λ)-anti-powers. Many of these generalizations can be achieved using proofs similar to those in [5]. We also provide a condensed proof of the fact that the Sturmian words contain anti-powers of every order beginning at every index.

We begin with a straightforward lemma.

Lemma 15. *Suppose $k > \lambda > j > 1$. If a word avoids (k, λ)-anti-powers, then it avoids $(k-j, \lambda-j)$-anti-powers.*

Proof. It is enough to show that if a word avoids (k, λ)-anti-powers, then it avoids $(k-1, \lambda-1)$-anti-powers. Suppose that a word x contains a $(k-1, \lambda-1)$-anti-power w of length km. If we extend to the right by m letters, we obtain a (k, λ)-anti-power, since we increase the number of equal blocks, $|\{i : w_i = w_j\}|$ for any j, by at most 1. $\square$

Definition 16. We call an infinite word *constant* if it is of the form a^ω for some $a \in \mathbb{A}$.

In order to classify the words avoiding $(k, k-2)$-anti-powers, we will use two results of Fici, Restivo, Silva, and Zamboni. The case $k = 3$, that is, the avoidance of 3-anti-powers, is handled in their paper [5].

Lemma 17. ([5], Lemma 9) *Let x be an infinite word. If x avoids 3-anti-powers, then x is a binary word.*

Proposition 18. ([5], Proposition 10) *Let x be an infinite word. If x avoids 3-anti-powers, then it cannot contain a factor of the form 10^n1 or 01^n0 with $n > 1$.*

Theorem 19.

1. *For $k > 1$, the infinite words avoiding $(k, k-1)$-anti-powers are precisely the constant words.*
2. *For $k > 3$, infinite words avoiding $(k, k-2)$-anti-powers are the words that differ from a constant word in at most one position.*
3. *For $k > 3$, there exist infinite aperiodic words avoiding $(k, k-3)$-anti-powers.*

Proof. The first claim is trivial; merely note that the avoidance of $(k, k-1)$-anti-powers implies that every factor whose length is a multiple of k is a k-power.

For the second claim, let x be a word avoiding $(k, k-2)$-anti-powers. By Lemma 15, x avoids 3-anti-powers and $(4, 2)$-anti-powers. By Lemma 17, x is a binary word. Suppose, seeking a contradiction, that x has at least 2 instances of 1 and at least 2 instances of 0, i.e., x differs from a constant word in more than one position. Note x contains no factor of the form 1100 or 0011, as these are $(4, 2)$-anti-powers. Thus x has a factor of the form 10^a1^b0 or 01^a0^b1 for some $a \geqslant 1, b \geqslant 1$; without loss of generality assume it is the first. By Proposition 18, $a = b = 1$. However, under these conditions, x has a factor of the form 1010, which is itself a $(4, 2)$-anti-power.

For the third claim, we exhibit a family of infinite aperiodic words avoiding $(k, k-3)$-anti-powers. Let $\{\gamma_i\}_{i=1}^{n}$ be an increasing sequence such that $\gamma_{i+1} \geqslant (k+1)\gamma_i$ for all $i \in \mathbb{N}$. Define a word x as follows:

$$x[j] = \begin{cases} 1 & \text{if } j = \gamma_i \text{ for some } i; \\ 0 & \text{otherwise.} \end{cases}$$

We will show that x avoids $(k, k-3)$-anti-powers. Note that if $x[\ell+1..\ell+n]$ has at least two nonzero entries, then for some i we have

$$\ell + 1 \leqslant \gamma_i < (k+1)\gamma_i \leqslant \gamma_{i+1} \leqslant \ell + n.$$

This implies that $n > k\gamma_i \geqslant k(\ell+1)$, so $\ell + 1 \leqslant \frac{n}{k}$. Suppose, seeking a contradiction, that the k consecutive blocks $x[j+1..j+m], \ldots, x[j+(k-1)m+1..j+km]$ form a $(k, k-3)$-anti-power. At most $k-3$ of these blocks can be 0^m, so the word $x[j+m+1..j+km]$ has at least two nonzero entries. Thus, $j + m + 1 \leqslant \frac{(k-1)m}{k}$. It follows that $j + 1 < 0$, a contradiction. □

Theorem 20. *For all $k \geqslant 6$, there exist aperiodic recurrent words avoiding $(k, k-5)$-anti-powers.*

Proof. Let w be the limit of the sequence $w_0 = 0$, $w_{n+1} = w_n 1^{(k-3)|w_n|} w_n$. Note that each occurrence of w_n except the first is preceded and followed by $1^{(k-3)|w_n|}$. Let $v = v_1 v_2 \cdots v_k$ be a factor of w, where $|v_i| = \ell > 0$ for all $i \in \{1, \dots, k\}$. Let n be the largest integer such that

$$|w_n| = (k-1)^n < 2\ell < (k-1)^{n+1} = |w_{n+1}|.$$

Since w is recurrent, we can assume v appears after the first appearance of w_n.

We claim that at most four blocks of v can intersect an occurrence of w_n. Each occurrence of w_n intersects at most two blocks of v by the condition $2\ell > |w_n|$. Moreover, any three occurrences of w_n are separated by factors of $1^{(k-3)|w_n|}$ and $1^{(k-3)|w_{n+1}|}$,. As

$$|v| = k\ell < \frac{k}{2}|w_{n+1}| \leqslant (k-3)|w_{n+1}|,$$

v can intersect at most 2 occurrences of w_n. We can conclude that at most four blocks of v are not equal to 1^ℓ. □

We now restrict ourselves to the setting of k-anti-powers. In [5], Fici et al. question under what conditions aperiodic recurrent words can avoid k-anti-powers. It is known this is possible for $k \geqslant 6$ and impossible for $k \leqslant 3$, but nothing has been shown for $k = 4$ or 5. One class of words that we can exclude from this search are the aperiodic recurrent words, including the well-studied class of Sturmian words.

Definition 21. A *Sturmian word* is an infinite word x such that for all $n \in \mathbb{N}$, x has exactly $n + 1$ distinct factors of length n.

Note that Sturmian words are necessarily binary. An alternate characterization of the Sturmian words in terms of *irrationally mechanical words* was given by Morse and Hedlund [8] in 1938.

Definition 22. The *upper mechanical word* $s_{\theta,x}$ and the *lower mechanical word* $s'_{\theta,x}$ with angle θ and initial position x are defined, respectively, by

$$s_{\theta,x}[n] = \begin{cases} 1 & \text{if } \theta(n-1) + x \in [1-\theta, 1) \mod 1 \\ 0 & \text{if } \theta(n-1) + x \in [0, 1-\theta) \mod 1 \end{cases}$$

$$s'_{\theta,x}[n] = \begin{cases} 1 & \text{if } \theta(n-1) + x \in (1-\theta, 1) \mod 1 \\ 0 & \text{if } \theta(n-1) + x \in [0, 1-\theta] \mod 1 \end{cases}$$

for some $\theta, x \in \mathbb{R}$. A word w is called *irrationally mechanical* if $w = s_{\theta,x}$ or $w = s'_{\theta,x}$ for some $x \in \mathbb{R}$ and irrational $\theta \in \mathbb{R}$.

Theorem 23. ([8]) *A word is Sturmian if and only if it is irrationally mechanical.*

Irrationally mechanical words can be interpreted through the lens of mathematical billiards. Consider the unit circle centered at the origin, parameterized by $g(t) = (\cos(2\pi t), \sin(2\pi t))$ for $t \in \mathbb{R}$. Place an (infinitesimal) ball at point $g(x)$ on the circle and shoot it in a straight trajectory toward $g(x+\theta)$. At each moment the ball "bounces off" the circle, it generates a 0 if it hits the point $g(x)$ for $x \in [0, 1-\theta)$ and a 1 otherwise. The sequence generated by the trajectory of such a ball is precisely the word $s_{\theta,x}$. For example, a trajectory generating the famous Fibonacci word is shown below.

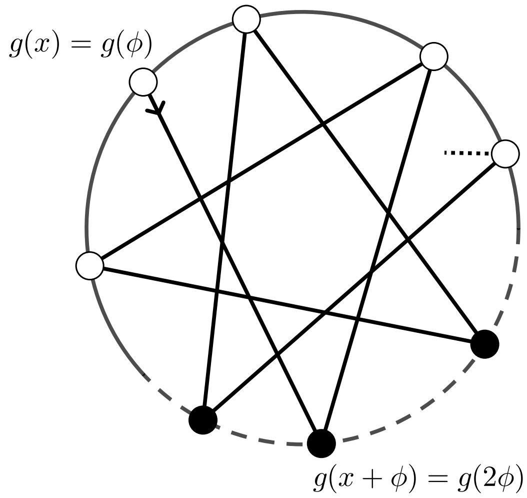

Figure 1: The trajectory associated with the Fibonacci word $s_{\phi,\phi} = 01001010\ldots$, where ϕ is the golden ratio $1.6180339\ldots$. A white point indicates that the letter 0 is generated, and a black point indicates that the letter 1 is generated. The Fibonacci word can also be generated as the limit of the sequence $\{S_n\}_{n=1}^{\infty}$, where $S_1 = 0$, $S_2 = 01$, and $S_n = S_{n-1}S_{n-2}$ for $n \geqslant 3$.

It is well-known that every Sturmian word is uniformly recurrent (See, for example, Proposition 2.2.30 of [7]). It is straightforward to see that any uniformly recurrent, aperiodic word is ω-power-free. We note that the following theorem of Fici et al. in [5] implies that any Sturmian word has a k-anti-power starting at every position.

Theorem 24. ([5], Theorem 6) *Let x be an ω-power-free word on a finite alphabet. For every $k > 1$ there is an occurrence of a k-anti-power starting at every position of x.*

We can also deduce a structural property of recurrent words avoiding k-anti-powers in terms of their representation as a *sesquipower.*

Definition 25. Given a sequence $\{v_n\}_{n=1}^{\infty}$ of finite words, define words w_n by $w_1 = v_1$ and $w_{n+1} = w_n v_n w_n$. The limit of the sequence of words $\{w_n\}_{n=1}^{\infty}$ is called the *sesquipower induced by the sequence* $\{v_n\}_{n=1}^{\infty}$.

It is well-known that an infinite word is recurrent if and only if it is a sesquipower (see, for example, [7]). We show that if an aperiodic recurrent word avoids k-anti-powers, then $\{v_n\}_{n=1}^{\infty}$ must contain words with powers of arbitrary order as factors.

Theorem 26. *Let x be the aperiodic sesquipower on a finite alphabet induced by $\{v_n\}_{n=1}^{\infty}$, and suppose x avoids k-anti-powers for some $k \geqslant 2$. There exists a word u of length at most $k-1$ such that for all $\ell > 0$, there is some $n > 0$ such that u^ℓ is a factor of v_n.*

Proof. Since x avoids k-anti-powers, Corollary 10 implies x is not ω-power-free. Thus, there is some factor u of x such that u^ℓ is a factor for every $\ell > 0$. We can assume u is not an m-power for any $m \geqslant 2$; otherwise, let $u = u_{(|u|/m)}$. Suppose, seeking a contradiction, that $|u| \geqslant k$. Note that the prefix of length $k(|u|+1)$ of u^{k+1} is a k-anti-power, contradicting the fact that u^{k+1} is a factor of x that avoids k-anti-powers. Thus, $|u| \leqslant k-1$.

We now know arbitrarily long powers of u occur in x, but, in fact, we can show that arbitrarily long powers of u occur in $\{v_n\}_{n=1}^{\infty}$. Since x is not periodic, there exists a k such that v_k is not a factor of u^ω. Let ℓ_0 be the largest power of u that is a factor of x_k. For sufficiently large ℓ, there is some $m \geqslant k$ such that u^ℓ is a factor of w_{m+1} but not of w_m. We can conclude $u^{\ell-2\ell_0-2}$ is a factor of v_{m+1}, as w_k is a border of w_m. As ℓ_0 is fixed, this implies that for all $\ell > 0$, u^ℓ is a factor of some v_m. □

To summarize, if x is an aperiodic recurrent word avoiding k-anti-powers for some $k \geqslant 2$, then x is not uniformly recurrent, hence not Sturmian, and is the sesquipower induced by a sequence $\{v_n\}_{n=1}^{\infty}$ where the words $\{v_n\}$ contain arbitrarily long powers of some word u of length at most $k-1$.

5 Avoiding Powers and Anti-Powers

In [5], Fici et al. show that for every $\ell, k > 1$, there exists $N_\alpha(\ell, k)$ such that every word of length $N_\alpha(\ell, k)$ on an alphabet of size α contains either an ℓ-power or a k-anti-power. They prove that for $k > 2$, one has $k^2 - 1 \leqslant N_\alpha(k,k) \leqslant k^3\binom{k}{2}$. We improve both these lower and upper bounds.

Theorem 27. *For any $k > 3$,*

$$2k^2 - 2k \leqslant N_\alpha(k,k) \leqslant (k^3 - k^2 + k)\binom{k}{2}.$$

Proof. The upper bound is precisely the statement of Corollary 14.

For the lower bound, consider the word

$$x = 1(0^{k-1}1)^{k-2}0^{k-2}10^{k-2}(10^{k-1})^{k-2}1.$$

We begin by showing that the border $1(0^{k-1}1)^{k-2}0^{k-2}10^{k-2}$ of length k^2-2 avoids k-powers and k-anti-powers. In their proof that $k^2 - 1 \leqslant N_\alpha(k,k)$, Fici et al. [5] show that the word $(0^{k-1}1)^{k-2}0^{k-2}10^{k-1}$ avoids k-powers and k-anti-powers, so we need only check this border for k-power or k-anti-power prefixes. We can see immediately that there are no k-power prefixes: as $1 \leqslant m \leqslant k-1$, the first block of length m of the prefix of length km begins with 1 while the second begins with 0. Suppose, seeking a contradiction, that the prefix of length km is a k-anti-power for some m. Since the prefix of length km would need to contain at least $k-1$ instances of the letter 1 to distinguish the blocks, we require $km \geqslant 1+k(k-2)$. Hence, $m \geqslant k-1$. We know the block length m is at most $k-1$ since $1(0^{k-1}1)^{k-2}0^{k-2}10^{k-2}$ has length k^2-2. However, as $k(k-2)+1=(k-1)^2$, the last two blocks must be $0^{k-2}1$. The equality of these blocks contradicts the assumption that the prefix of length km is a k-anti-power. $1(0^{k-1}1)^{k-2}0^{k-2}10^{k-2}$ avoids both k-powers and k-anti-powers.

Thus, we need to consider only those factors of x intersecting nontrivially with the prefix and suffix of length k^2-2. Fix such a factor y of length km starting at position j.

Suppose, seeking a contradiction, that y is a k-power. Let $y_\ell = x[j+\ell m..j+(\ell+1)m-1]$ be the ℓ^{th} block of length m in y. Choose b such that the central letter 1 of x is contained in y_b. That is, $j+bm \leqslant k(k-1) \leqslant j+(b+1)m-1$. Since $k \geqslant 4$ and there are exactly two occurrences of the factor $10^{k-2}1$ in x, the block y_b cannot contain $10^{k-2}1$ as a factor. Note $y_\ell \neq 0^m$ for any nonnegative integers ℓ and m. As $k \geqslant 4$, one of $b-2, b+2 \in \{0,\ldots,k-1\}$; without loss of generality assume it is $b+2$. Thus, the two factors $y_{b-1}y_b$ and $y_{b+1}y_{b+2}$ of x each contain an occurrence of the factor $10^{k-2}1$. However, the only two occurrences of $10^{k-2}1$ in x intersect while $y_{b-1}y_b$ and $y_{b+1}y_{b+2}$ are disjoint, so we've reached a contradiction. Therefore, x avoid k-powers.

Now we show that such a factor y is not a k-anti-power. Suppose it were. Since y contains at least $k-1$ occurrences of the letter 1, it follows that $m \geqslant k-2$. In the case $m = k-2$, each block contains at most one occurrence of the letter 1, but there are only $k-1$ distinct such blocks. One can check $m \neq k-1, k$ by examining the period of the prefix/suffix of length $k(k-2)+1$ or the middle section of length $2k-1$. Thus, taking into consideration the length of x, we have $k+1 \leqslant m \leqslant 2k-3$. Consider all blocks except y_b (the block containing the central 1). Note that the letters of each block are determined by the number of leading 0's, which is at most $k-1$. If there are ℓ blocks preceding y_b, and y_{b-1} has z leading zeros, then the numbers of leading zeros for all blocks except y_b are given by the multiset

$$\{z+(\ell-1)m,\ldots,z+m,z,z-2m-2,z-3m-2,\ldots,z-(k-\ell)m-2\} \mod k \quad (*)$$

which has a repeated element if and only if the multiset

$$\{(l-1)m,\ldots,m,0,(k-2)m-2,(k-3)m-2,\ldots,\ell m-2\} \mod k$$

has a repeated element. Assuming this has no repeated element, we have

$$\{\ell m-2,\ldots,(k-2)m-2\} \subseteq \{\ell m,(\ell+1)m,\ldots,(k-1)m\}.$$

The left-hand side has size $k-\ell-1$ while the right-hand side has size at most $k-\ell$, and both are arithmetic progressions with difference m. Thus, either $\ell m-2 \equiv \ell m \mod k$ or $\ell m-2 \equiv (\ell+1)m \mod k$. In the former case, $2 \equiv 0 \mod k$, but this would imply $k=2$, contradicting the fact that $k \geqslant 4$. In the latter case, $m \equiv -2 \mod k$, but this also leads to a contradiction as $k+1 \leqslant m \leqslant 2k-3$. Therefore, x avoids k-powers and k-anti-powers. □

Note that the above bounds are independent of the alphabet size. This leads to two questions: does $N_\alpha(k,k)$ depend on the size of the alphabet, and if so, in what way? Note that $N_\alpha(k,k)$ is nondecreasing as α increases. The following values of $N_2(k,k)$ were computed by Shallit [11].

$$N_2(1,1) = 1 \qquad N_2(2,2) = 2 \qquad N_2(3,3) = 9 \qquad N_2(4,4) = 24 \qquad N_2(5,5) = 55$$

For small α and k, $N_\alpha(k,k)$ can be computed by testing all α-ary strings of small length with a computer. In particular, we were able to check that $N_4(3,3) = N_2(3,3) = 9$, which implies $N_\alpha(3,3) = 9$ for all $\alpha \geqslant 2$. This follows because a word of length 9 avoiding 3-anti-powers would use at most 4 letters. We also computed that $N_{11}(4,4) = N_2(4,4) = 24$. It is straightforward to check that a word of length 24 avoiding 4-anti-powers would use at most 11 letters. Thus, $N_\alpha(3,3)$ and $N_\alpha(4,4)$ are independent of α. It remains open if this is true for all k.

Another scenario to investigate is under what conditions a word can be extended (in a potentially larger alphabet) and still avoid k-powers and k-anti-powers. We aim to show that for large enough α, no word of length $N_\alpha(k,k) - 1$ can be extended (in a larger alphabet) and avoid k-powers and k-anti-powers. To do so, we require the following lemma.

Lemma 28. *If there exists $\alpha \geqslant 2$ such that $N_\alpha(k,k) < N_{\alpha+1}(k,k)$, then one of the following must hold:*

1. *Let $W_{k,\alpha}$ be the set of words on $[\alpha+1]$ of length $N_\alpha(k,k)$ that avoid k-powers and k-anti-powers. For every word $w \in W_{k,\alpha}$, the two factors of w of length $N_\alpha(k,k)-1$ each use exactly $\alpha+1$ letters.*

2. *There exists a word on $[\alpha]$ of the form*

$$w = u_1(1u_1)^{k-1}x_1 = u_2(2u_2)^{k-1}x_2 = \cdots = u_\alpha(\alpha u_\alpha)^{k-1}x_\alpha$$

that avoids k-powers and k-anti-powers, where $x_1, \ldots, x_\alpha, u_1, \ldots, u_\alpha$ are finite words, $|u_1| < \cdots < |u_\alpha|$, and for all $1 \leqslant i < j \leqslant \alpha$,

$$\gcd(|u_i|+1, |u_j|+1) \leqslant \frac{|u_j|+1}{k-1}.$$

Proof. Suppose that the first case does not hold. There is a word w of length $N_\alpha(k,k)-1$ on $[\alpha]$ such that $(\alpha+1)w$, the extension of w by the addition the letter $\alpha+1$ on the left, avoids k-powers and k-anti-powers. Since $|w|$ is maximal for words on $[\alpha]$ avoiding k-powers and k-anti-powers, the extension aw contains a k-power or k-anti-power for any $a \in [\alpha]$. As $(\alpha+1)w$ contains no k-anti-powers, neither does aw for any $a \in [\alpha]$. Thus, aw has a prefix that is a k-power for each $a \in [\alpha]$.

Hence,

$$w = u_1(1u_1)^{k-1}x_1 = u_2(2u_2)^{k-1}x_2 = \cdots = u_\alpha(\alpha u_\alpha)^{k-1}x_\alpha,$$

where $x_1, \dots, x_\alpha, u_1, \dots, u_\alpha$ are finite words. Note that this implies $w[m(|u_\ell|+1)] = \ell$ for all $1 \leqslant m \leqslant k-1$ and $1 \leqslant \ell \leqslant k$. Without loss of generality (since the labels of the letters are arbitrary), we can assume $|u_1| < \cdots < |u_\alpha|$. Suppose, seeking a contradiction, that for some $1 \leqslant i < j \leqslant \alpha$, we have $\gcd(|u_i|+1, |u_j|+1) \geqslant \frac{|u_j|+1}{k-1}$. There is some $1 \leqslant m \leqslant k-1$ such that $m(|u_i|+1) \equiv 0 \mod (|u_j|+1)$. Hence, $m(|u_i|+1) = d(|u_j|+1)$ for some $1 \leqslant d \leqslant k-1$. However, this implies

$$i = w[m(|u_i|+1)] = w[d(|u_j|+1)] = j.$$

Since we assumed $i < j$, we've reached a contradiction. □

An investigation of the failure of the first case leads to the following corollary.

Corollary 29. *Suppose $\alpha > \frac{N_\alpha(k,k)}{k} - k + 3$. If a word w has a factor $u \neq w$ of length $N_\alpha(k,k) - 1$ that uses only α letters, w contains a k-power or k-anti-power.*

Proof. Suppose, seeking a contradiction, that w is as above but contains no k-power or k-anti-power. For all $1 \leqslant i < j \leqslant \alpha$, we have by Lemma 28 that

$$1 \leqslant \gcd(|u_i|+1, |u_j|+1) \leqslant \frac{|u_j|+1}{k-1}.$$

Thus, $|u_j| \geqslant k-2$ for all $j \geqslant 2$. Since the $|u_j|$'s are strictly increasing, this implies $|u_\alpha| \geqslant (\alpha-2) + (k-2) = \alpha + k - 4$. As $w = u_\alpha(\alpha u_\alpha)^{k-1}x_\alpha$, we have

$$|w| \geqslant k(\alpha + k - 4) + k - 1 = k\alpha + k^2 - 3k - 1.$$

Since w is a word on $[\alpha]$ avoiding k-powers and k-anti-powers, $k\alpha + k^2 - 3k - 1 \leqslant N_\alpha(k,k)$. If this inequality is not satisfied, then we can conclude w is as above but contains a k-power or k-anti-power. □

6 Block Patterns and Their Expectation

In this section, we return to the general setting of block-patterns to calculate the expected number of $(\mu_1, \dots, \mu_n)$-block-patterns in a word of length n on an alphabet of size α. The special case of this expectation for k-powers was calculated by Christodoulakis, Christou, Crochemore, and Iliopoulos in [2].

Theorem 30. ([2], Theorem 4.1) *On average, a word of length n has $\Theta(n)$ k-powers. More precisely, this number is*

$$(n+1)\frac{\alpha^{1-k}(1-\alpha^{(1-k)\lfloor \frac{n}{k} \rfloor})}{1-\alpha^{1-k}} - \frac{k}{\alpha^{k-1}}\left(\frac{1}{1-\alpha^{1-k}} - \frac{\lfloor \frac{n}{k} \rfloor \alpha^{(1-k)\lfloor \frac{n}{k} \rfloor}}{1-\alpha^{1-k}} + \frac{\alpha^{1-k}(1-\alpha^{(1-k)(\lfloor \frac{n}{k} \rfloor - 1)})}{(1-\alpha^{1-k})^2}\right).$$

Theorem 31. *On average, a word of length n has $O(n^2)$ and $\Omega(n)$ $(\mu_1, \dots, \mu_k)$-block-patterns. More precisely, the expected number of $(\mu_1, \dots, \mu_k)$-block-patterns is*

$$\sum_{m=1}^{\lfloor \frac{n}{k} \rfloor} (n+1-km)\frac{k!}{\mu_1! \cdots \mu_k!}\frac{1}{\alpha^{km}} \prod_{\ell=1}^{\mu_1+\cdots+\mu_k} (\alpha^m - (\ell-1)).$$

Proof. Let x be a word of length n, drawn uniformly at random. Let

$$X_{i,j} = \begin{cases} 1 & \text{if } x[i..j] \text{ is a } (\mu_1, \ldots, \mu_k)\text{-block-pattern;} \\ 0 & \text{otherwise.} \end{cases}$$

Let $N = \sum_{i \leqslant j} X_{i,j}$. That is, N is the number of $(\mu_1, \ldots, \mu_k)$-block-patterns in x. We have

$$\begin{aligned} \mathbb{E}[N] &= \mathbb{E}\left[\sum_{i=1}^{n-k+1} \sum_{j=i+1}^{n} X_{i,j}\right] \\ &= \sum_{i=1}^{n-k+1} \sum_{j=i+1}^{n} \mathbb{E}[X_{i,j}] \\ &= \sum_{i=1}^{n-k+1} \sum_{j=i+1}^{n} \mathbb{P}(x[i..j] \text{ is a } (\mu_1, \ldots, \mu_k)\text{-block-pattern}). \end{aligned}$$

Let us count the number of $(\mu_1, \ldots, \mu_k)$-block-patterns of length α^{j+1-i} on $[\alpha]$. Partition $[k]$ into unlabeled parts with μ_s parts of size s, and choose $\mu_1 + \cdots + \mu_k$ distinct ordered elements from $\alpha^{(j+1-i)/k}$. We can assign elements to parts by order of appearance of the parts, which will yield a $(\mu_1, \ldots, \mu_k)$-block-pattern. Moreover, the block-pattern is uniquely determined by the choice of an unlabeled partition and ordered m-tuple. Let $[A]$ denote the indicator function of the event A. We have

$$\begin{aligned} \mathbb{E}[N] &= \sum_{i=1}^{n-k+1} \sum_{j=i+1}^{n} \frac{k!}{\mu_1! \cdots \mu_k!} \frac{1}{\alpha^{j+1-i}} \prod_{\ell=1}^{\mu_1+\cdots+\mu_k} \left(\alpha^{(j+1-i)/k} - (\ell - 1)\right) [j+1-i \equiv 0 \bmod k] \\ &= \sum_{m=1}^{\lfloor \frac{n}{k} \rfloor} (n+1-km) \frac{k!}{\mu_1! \cdots \mu_k!} \frac{1}{\alpha^{km}} \prod_{\ell=1}^{\mu_1+\cdots+\mu_k} (\alpha^m - (\ell-1)). \end{aligned}$$

Since there are only $\binom{n}{2} + n$ nonempty factors of x, we have $\mathbb{E}[N] = O(n^2)$. Note that the expectation is minimized for k-powers, where $\mu_k = 1$ and $\mu_s = 0$ for all $s < k$. Thus, from Theorem 30, we have $\mathbb{E}[N] = \Omega(n)$. ☐

Corollary 32. *On average, a word of length n has $\Theta(n^2)$ k-anti-powers. More precisely, the expected number of k-anti-powers is*

$$\sum_{m=1}^{\lfloor \frac{n}{k} \rfloor} (n+1-km) \prod_{\ell=0}^{k-1} \left(1 - \frac{\ell}{\alpha^m}\right).$$

Proof. Let x be a word of length n on an alphabet of size α. The formula follows Theorem 31 in the case $\mu_1 = k$ and $\mu_s = 0$ for $s > 1$. Restricting the sum (of nonnegative terms) to the range $\lfloor \frac{n}{4k} \rfloor \leqslant m \leqslant \lfloor \frac{3n}{4k} \rfloor$, we see

$$\mathbb{E}[\#(k\text{-anti-powers in } x)] \geqslant \left(\frac{n}{2k} - 1\right)\left(\frac{n}{4} + 1\right)\left(1 - \frac{k-1}{\alpha^{\lfloor \frac{n}{4k} \rfloor}}\right)^{k-1}.$$

For $n > 4k\left(1+\log_\alpha \frac{k-1}{1-2^{-k-1}}\right)$, we have $\left(1-\frac{k-1}{\alpha^{\lfloor \frac{n}{4k}\rfloor}}\right)^{k-1} > \frac{1}{2}$, hence

$$\mathbb{E}\left[\#(k\text{-anti-powers in } x)\right] > \frac{n^2}{16k} - \frac{n}{4} - \frac{1}{2} = \Omega(n^2).$$

There are $\binom{n}{2}+n$ nonempty factors of x, so we have $\mathbb{E}\left[\#(k\text{-anti-powers in } x)\right] = O(n^2)$. $\square$

7 Further Directions

Recall that Theorem 7 shows that having a small enough density of $(\mu_1, \ldots, \mu_k)$-block-pattern prefixes with few equal blocks implies the existence of arbitrarily long power prefixes. We believe that a strengthening of this argument could yield a lower bound on the density of $P(x, k)$, the set of $m \in \mathbb{N}$ such that the prefix of x of length km is a k-power. In Section 3, we also remark that Theorem 6 of [5], stating that if x is an ω-power-free word then $AP(x_{(j)}, k)$ is nonempty for every j and k, is false if we allow infinite alphabets. Perhaps there is a finer characterization of which ω-power-free words fail this condition.

As in the bounds found by Fici et al. [5], our upper and lower bounds for $N_\alpha(k, k)$ are polynomials in k whose degrees differ by 3. If it is the case that $N_\alpha(k, k)$ depends on α, such a dependence could be used to strengthen the bounds for $N_\alpha(k, k)$. Given the few known values of $N_\alpha(k, k)$, it seems plausible that k always divides $N_\alpha(k, k)$. On the other hand, if $N_\alpha(k, k)$ is independent of α, this alone would be an interesting structural property of the set of words avoiding k-powers and k-anti-powers achieving the length $N_\alpha(k, k)$ for arbitrary α. We believe the second case holds.

Conjecture 33. The quantity $N_\alpha(k, k)$ is independent of α.

Whether there exist aperiodic recurrent words avoiding 4 or 5 powers remains an open question. One may wish to investigate other large classes of words, such as the morphic words, and their potential to avoid k-anti-powers. A natural generalization is to find the structure of infinite words avoiding $(\mu_1, \ldots, \mu_k)$-block-patterns other than (k, λ)-anti-powers.

Lastly, fix a finite alphabet $\mathbb{A} = \{a_1, \ldots, a_\alpha\}$. The *Parikh vector* $\mathcal{P}(w) = (e_1, \ldots, e_\alpha)$ of a finite word w on $\mathbb{A}$ has entry e_i equal to the number of instances of a_i in w. Define an *abelian* $(\mu_1, \ldots, \mu_k)$*-block-pattern* to be a word of the form $w = w_1 \cdots w_k$ where, if the set $\{1, \ldots, k\}$ is partitioned via the rule $i \sim j \iff \mathcal{P}(w_i) = \mathcal{P}(w_j)$, there are μ_s parts of size s for all $1 \leqslant s \leqslant k$. One may ask questions similar to those addressed in this paper for abelian $(\mu_1, \ldots, \mu_k)$-block-patterns.

Acknowledgements

The author would like to thank Joe Gallian for his tireless efforts to foster a productive and engaging mathematical community. She extends her gratitude to Colin Defant and Samuel Judge for carefully reading through this article and providing helpful feedback, as well as to Jarkko Peltomäki for his comments. This research was conducted at the University of Minnesota, Duluth REU and was supported by NSF/DMS grant 1650947 and NSA grant H98230-18-1-0010.

References

[1] L. Babai. An anti-Ramsey theorem. *Graphs and Combinatorics*, 1(1):23-28, 1985.

[2] M. Christodoulakis, M. Christou, M. Crochemore, and C. Iliopoulos. On the average number of regularities in a word. *Theoretical Computer Science*, 525:3-9, 2014.

[3] C. Defant. Anti-power prefixes of the Thue-Morse word. *Electronic Journal of Combinatorics*, 24(1):#P32, 2017.

[4] P. Erdős, M. Simonovits, and V.T. Sós. Anti-Ramsey theorems, in: Finite and Infinite Sets. *Colloquium of the János Bolyai Mathematical Society*, pages 633-643, 1975.

[5] G. Fici, A. Restivo, M. Silva, and L. Q. Zamboni. Anti-powers in infinite words. *Journal of Combinatorial Theory, Series A*, 157:109-119, 2018.

[6] M. Gaetz. Anti-power j-fixes of the Thue-Morse word. arXiv:1808.01528 [math.CO], 2018.

[7] M. Lothaire. Algebraic combinatorics on words. *Encyclopedia of Mathematics and Its Applications*, 2011.

[8] M. Morse and G. A. Hedlund. Symbolic dynamics. *American Journal of Mathematics*, 11(1): 815-856, 1938.

[9] S. Narayanan. Functions on antipower prefix lengths of the Thue-Morse word. arXiv:1705.06310 [math.CO], 2018.

[10] R. Rado. Anti-Ramsey theorems, in: Finite and Infinite Sets. *Colloquium of the János Bolyai Mathematical Society*, pages 1159-1168, 1975.

[11] J. Shallit. *The On-Line Encyclopedia of Integer Sequences*. http://oeis.org, Sequence A274543.

[12] M. Simonovits and V.T. Sós. On restricted colorings of K_n. *Combinatorica*, 4(1): 101-110, 1984.