Characterizations of families of morphisms and words via binomial complexities*

Michel Rigo[†], Manon Stipulanti[‡], and Markus A. Whiteland[§¶]

Department of Mathematics, University of Liège,
Allée de la Découverte 12, 4000 Liège, Belgium

{m.rigo,m.stipulanti,mwhiteland}@uliege.be

Abstract

Two words are k-binomially equivalent if each subword of length at most k occurs the same number of times in both words. The k-binomial complexity of an infinite word is a counting function that maps n to the number of k-binomial equivalence classes represented by its factors of length n. Cassaigne et al. [Int. J. Found. Comput. S., 22(4) (2011)] characterized a family of morphisms, which we call Parikh-collinear, as those morphisms that map all words to words with bounded 1-binomial complexity. Firstly, we extend this characterization: they map words with bounded k-binomial complexity to words with bounded $(k+1)$-binomial complexity. As a consequence, fixed points of Parikh-collinear morphisms are shown to have bounded k-binomial complexity for all k. Secondly, we give a new characterization of Sturmian words with respect to their k-binomial complexity. Then we characterize recurrent words having, for some k, the same j-binomial complexity as the Thue–Morse word for all $j \leq k$. Finally, inspired by questions raised by Lejeune, we study the relationships between the k- and $(k+1)$-binomial complexities of infinite words; as well as the link with the usual factor complexity.

Keywords: Factor complexity, Abelian complexity, Binomial complexity, powers of the Thue–Morse morphism, Sturmian words.

2020 Mathematics Subject Classification: Primary: 68R15. Secondary: 05A05.

1 Introduction

The combinatorial structure of an infinite word $\mathbf{x} \in A^{\mathbb{N}}$ over a finite alphabet A may reveal important aspects of $\mathbf{x}$ itself. This structure is often studied through its language $\mathcal{L}(\mathbf{x})$, i.e., the set of its factors, and in particular to inspect the set $\mathcal{L}_n(\mathbf{x}) := \mathcal{L}(\mathbf{x}) \cap A^n$ of factors of length n. Even plain counting the cardinality of this set turns out to be a useful concept: with $p_{\mathbf{x}}$ the *factor complexity* function defined as $p_{\mathbf{x}} \colon \mathbb{N} \to \mathbb{N}$, $n \mapsto \#\mathcal{L}_n(\mathbf{x})$, the celebrated Morse–Hedlund theorem asserts that an infinite word $\mathbf{x}$ is aperiodic if and only if $p_{\mathbf{x}}(n) \geq n+1$ for all $n \geq 1$. For instance, the Thue–Morse word $\mathbf{t} = 01101001\cdots$ (also known as the Prouhet–Thue–Morse word), the fixed point of the morphism $\varphi \colon 0 \mapsto 01, 1 \mapsto 10$, is aperiodic because its factor complexity is given by

$$p_{\mathbf{t}}(2^m + r) = \begin{cases} 3 \cdot 2^m + 4(r-1), & \text{if } 1 \leq r \leq 2^{m-1}; \\ 4 \cdot 2^m + 2(r-1), & \text{if } 2^{m-1} < r \leq 2^m, \end{cases} \tag{1}$$

*Markus Whiteland dedicates this paper to the memory of his father Alan Whiteland (1940–2021).

[†]Supported by the FNRS Research grant T.196.23 (PDR)
[‡]Supported by the FNRS Research grant 1.C.104.24F.
[§]Supported by the FNRS Research grant 1.B.466.21F.
[¶]Corresponding author

for all $m \geq 0$ [9, §4]. The factor complexity has proved its importance in a number of areas of mathematics. For example, in number theory, Adamczewski and Bugeaud [2] proved that the base-b expansion of a real algebraic irrational number has a factor complexity function satisfying

$$\liminf_{n\to\infty} \frac{p(n)}{n} = \infty.$$

As a consequence, the number having $\mathbf{t}$ as base-2 expansion is transcendental.

One can conversely define families of words using the factor complexity function. For example, a word $\mathbf{x}$ is called *Sturmian* if $p_{\mathbf{x}}(n) = n+1$ for all n. Such words, studied also in this note, turn out to have many interesting properties. For general references about combinatorics on (Sturmian) words, see, for instance, [5, 9, 32].

Many variations of the factor complexity have been introduced. Some of these counting functions only take into account factors with specific properties such as palindromes or privileged words [16, 37]. Other functions count subwords extracted along subsequences of prescribed forms like maximal pattern or arithmetical complexities [25, 7]. Closely related to the subject discussed in this paper, abelian, k-abelian or cyclic complexities are functions of the form $n \mapsto \#(\mathcal{L}_n(\mathbf{x})/\sim)$ for the quotient by a relevant equivalence relation $\sim$ [42, 26, 10]. For more on abelian combinatorics on words (and related notions), we refer the reader to the recent excellent survey [18]. For each of the above complexity functions usual questions naturally arise:

- What is the complexity of well-known families of words such as Sturmian, Arnoux-Rauzy, automatic, (pure) morphic, or Toeplitz words?
- This leads to the more interesting problem of classifying or characterizing infinite words with respect to their complexity. As an example, Coven and Hedlund in [15] show that an infinite word is purely periodic if and only if its abelian complexity function attains the value 1. Further, a binary aperiodic word has its abelian complexity function equal to the constant function 2 if and only if it is Sturmian. Words with linear factor complexity are characterized in [11].
- What are the possible growth rates of the complexity function?
- Which non-periodic words may achieved the lowest complexity?

These questions have intrinsic theoretical interests but also provide particular insight about the combinatorial structure of the studied words. Depending on the properties of interest, one focuses on the appropriate complexity function. Infinite words with specific combinatorial properties are, for instance, sought to construct particular symbolic dynamical systems or tilings of the line. For instance, the Thue–Morse minimal subshift is completely characterized by its factor complexity together with its abelian complexity [40]. The Thue–Morse word, which is central in our paper, plays an important role in many areas of mathematics, e.g., see [4, 3].

In this paper, the complexity function of interest is built from binomial coefficients of words.

Definition 1.1. Let $u, w \in A^*$. The *binomial coefficient* of u and w is the number of times w occurs as a subword of u, i.e. writing $u = u_1 \cdots u_n$ with $u_i \in A$,

$$\binom{u}{w} = \#\left\{i_1 < i_2 < \cdots < i_{|w|} : u_{i_1} u_{i_2} \cdots u_{i_{|w|}} = w\right\}.$$

These binomial coefficients have proven to be useful in a variety of domains: generalizations of Pascal-like triangles [31], algebra and topology [38, 8], formal languages or relationship with the extensively studied Parikh matrices and Simon's congruence [6, 20, 22]. For more on these binomial coefficients, see, for instance, [32, §6].

We mention a well-known and actively researched problem related to binomial coefficients. A word u is k-*reconstructible* whenever the knowledge of the binomial coefficients $\binom{u}{v}$, for all subwords v of length k, uniquely determines u. Inspired by the problem of reconstructing graphs

from vertex-deleted subgraphs, the famous reconstruction problem is to determine the function $f(n) = k$ where k is the least integer for which all words of length n (over a given alphabet) are k-reconstructible. For articles on this problem, we mention [19, 24, 34] and references therein.

Let us now introduce our main object of study. Let $k \geq 1$ be an integer. The k-binomial complexity function introduced in [43] is the central theme of Lejeune's thesis [28]. It is built on the k-binomial equivalence where factors are distinguished with respect to the number of occurring subwords.

Definition 1.2. Two words $u, v \in A^*$ are k-*binomially equivalent*, and we write $u \sim_k v$, if

$$\binom{u}{x} = \binom{v}{x}, \quad \forall x \in A^{\leq k}.$$

As an example the words $u = 0110$ and $v = 1001$ are 2-binomially equivalent because for $z \in \{u, v\}$

$$\binom{z}{0} = 2,\ \binom{z}{1} = 2,\ \binom{z}{00} = 1,\ \binom{z}{01} = 2,\ \binom{z}{10} = 2,\ \binom{z}{11} = 1$$

and $u \not\sim_3 v$ because 011 is a subword of u and not of v.

In [34, Lem. 1], it is observed that one may replace the condition $\forall x \in A^{\leq k}$ with $\forall x \in A^k$ as soon as $|u|, |v| \geq k$. Observe that the word u is obtained as a permutation of the letters in v if and only if $u \sim_1 v$. The latter relation is the *abelian equivalence* already introduced by Erdős [17]. This leads to the following definition, introduced in [43].

Definition 1.3. Let $k \geq 1$ be an integer. The k-*binomial complexity function* of an infinite word $\mathbf{x}$ is defined as $b_{\mathbf{x}}^{(k)} \colon \mathbb{N} \to \mathbb{N}$, $n \mapsto \#(\mathcal{L}_n(\mathbf{x})/\sim_k)$.

As an example, the first few values for the Thue–Morse word $\mathbf{t}$ are given in Table 1.

	0	1	2	3	4	5	6	7	8	9	10
$b_{\mathbf{t}}^{(1)}$	1	2	3	2	3	2	3	2	3	2	3
$b_{\mathbf{t}}^{(2)}$	1	2	4	6	9	8	8	8	9	8	8
$p_{\mathbf{t}}$	1	2	4	6	10	12	16	20	22	24	28

Table 1: The first few values of $b_{\mathbf{t}}^{(1)}$, $b_{\mathbf{t}}^{(2)}$ and $p_{\mathbf{t}}$.

It is clear that we have a series of refinements of the abelian equivalence: for all $k \geq 1$, $u \sim_{k+1} v$ implies $u \sim_k v$. Thus, for all n, we have the inequalities

$$b_{\mathbf{x}}^{(1)}(n) \leq b_{\mathbf{x}}^{(2)}(n) \leq \cdots \leq b_{\mathbf{x}}^{(k)}(n) \leq b_{\mathbf{x}}^{(k+1)}(n) \leq \cdots \leq p_{\mathbf{x}}(n). \tag{2}$$

The study of the k-binomial complexity function has so far been studied for restricted families of words. For example, for $k \geq 2$, the k-binomial complexity of Sturmian words coincides with their factor complexity [43] (recalled here as Theorem 2.8) and the same property holds for the Tribonacci word [30]. For any $k \geq 2$, fixed points of Parikh-constant morphisms (see the next part for a definition) are known to have bounded k-binomial complexity [43]. Recently, the k-binomial complexities of the Thue–Morse word [29] (given in (3)) and the 2-binomial complexities of generalized Thue–Morse words was also computed [33]. That is the extent to which the notion has been studied.

We remark that a better understanding of the k-binomial complexity may give information about the language $\mathcal{L}(\mathbf{x})$ of an infinite word $\mathbf{x}$ for which the reconstruction problem could be solved. The aim is to restrict the reconstruction problem to the language of an infinite word having a k-binomial complexity of the same order as its factor complexity. Indeed, if $b_{\mathbf{x}}^{(k)} = p_{\mathbf{x}}$ for some k, then for any two distinct factors y, z of $\mathbf{x}$, there exists a subword v of length k such that $\binom{y}{v} \neq \binom{z}{v}$.

Finally a parallel can be drawn between the k-abelian complexity introduced by Karhumäki et al. [26] and the k-binomial complexity. In both cases, we have a series of refinements (2) of the abelian equivalence. The fundamental difference is the following one. Two finite words u, v are k-*abelian equivalent* if, for each word w of length at most k, we count the same number of occurrences of the factor w in both words u and v. We thus make the important distinction between a *factor* and a *subword* of a word. Many properties of the k-abelian complexity have been recently and extensively studied such as growth and fluctuations, k-abelian palindromes, variation of Morse–Hedlund theorem, etc. [13, 12, 27]. This is to be contrasted with the limited knowledge we have on the k-binomial complexity function. Indeed, part of our motivation for this work stems from this rather limited state of the art as described above.

1.1 Our Results

We present three kinds of results: a new characterization of Parikh-collinear morphisms and links with bounded binomial complexities; a characterization of recurrent words with the same j-binomial complexities as the Thue–Morse word for $j = 1, \ldots, k$; study of the relationships existing between $\mathsf{b}_{\mathbf{w}}^{(k)}$ and $\mathsf{b}_{\mathbf{w}}^{(k+1)}$. This paper improves upon the preliminary conference version [44]: not only do we provide proofs of results announced therein, but we also significantly extend them.

• Morphisms mapping all infinite words to words with bounded abelian complexity have been characterized in [14]. Such a morphism $f\colon A^* \to B^*$ is said to be *Parikh-collinear*: for all letters $a, b \in A$, there is $r_{a,b} \in \mathbb{Q}$ such that $\Psi(f(b)) = r_{a,b}\Psi(f(a))$, where $\Psi(u)$ denotes the *Parikh vector* of a word u (see Section 2 for definitions). In Section 3, we obtain several new characterizations of Parikh-collinear morphisms. Connecting this with the series of inequalities (2), we show with Theorem 3.5 that a morphism is Parikh-collinear if and only if it maps all words with bounded k-binomial complexity to words with bounded $(k+1)$-binomial complexity.

It is known that any fixed point of a prolongable *Parikh-constant morphism* $f : A^* \to A^*$, i.e., $\Psi(f(a)) = \Psi(f(b))$ for all letters $a, b \in A$, has a bounded k-binomial complexity [43]. Any Parikh-constant morphism is obviously Parikh-collinear. As a direct consequence of our characterization of Parikh-collinear morphisms, Corollary 3.6 extends the previous result: bounded k-binomial complexity holds for any fixed point of a prolongable Parikh-collinear morphism.

• We then turn to words sharing their binomial complexities with the Thue–Morse word $\mathbf{t}$. From the above discussion (the Thue–Morse morphism φ is Parikh-constant), for all $j \geq 1$, the j-binomial complexity of $\mathbf{t}$ is bounded by a constant depending on j. But more is known, the exact value of $\mathsf{b}_{\mathbf{t}}^{(j)}(n)$ computed in [29] is given by

$$\mathsf{b}_{\mathbf{t}}^{(j)}(n) = \begin{cases} \mathsf{p}_{\mathbf{t}}(n), & \text{if } n < 2^j; \\ 3 \cdot 2^j - 3, & \text{if } n \equiv 0 \pmod{2^j} \text{ and } n \geq 2^j; \\ 3 \cdot 2^j - 4, & \text{otherwise,} \end{cases} \tag{3}$$

where the factor complexity $\mathsf{p}_{\mathbf{t}}$ of $\mathbf{t}$ is given by (1). Considering $j = 1$ in (3), words having the same abelian complexity as the Thue–Morse words have been characterized in [40] as follows. The abelian complexity of an aperiodic word $\mathbf{x} \in \{0,1\}^{\mathbb{N}}$ is, for $n > 0$, $\mathsf{b}_{\mathbf{x}}^{(1)}(n) = 3$ if n is even, and $\mathsf{b}_{\mathbf{x}}^{(1)}(n) = 2$ if n is odd, if and only if there exists a word $\mathbf{y}$ such that $\mathbf{x} = u\varphi(\mathbf{y})$ with $u \in \{\varepsilon, 0, 1\}$. Sections 4 and 5 are about binomial properties of iterates of φ. We generalize the latter result and obtain a characterization of words having the same j-binomial complexity as the Thue–Morse word $\mathbf{t}$ for all $j \leq k$. Except for a remark in [18] (see Theorem 2.9), such a result together with Theorem 2.11 are the first where binomial complexity leads to the characterization of combinatorial families of words. In this paper, with Theorem 2.11, we observe that a word $\mathbf{x}$ is Sturmian if and only if $\mathsf{b}_{\mathbf{x}}^{(2)}(n) = n + 1$, for all n. We make the statements about words sharing the same j-binomial complexities as $\mathbf{t}$ more precise.

Let k be an integer and let $\mathbf{y}$ be an aperiodic binary word. With Theorem 4.2 we show that for $\mathbf{x} = u\varphi^k(\mathbf{y})$ we have, for all $j \leq k$, $\mathsf{b}_{\mathbf{x}}^{(j)} = \mathsf{b}_{\mathbf{t}}^{(j)}$ which is given by (3), where u is a (possibly

empty) proper suffix of $\varphi^k(0)$ or $\varphi^k(1)$. Conversely, with Theorem 5.2, if $b_x^{(j)} = b_t^{(j)}$ for all $j \leq k$ for a *recurrent* word $\mathbf{x}$, i.e., each factor of $\mathbf{x}$ appears infinitely often, then $\mathbf{x} = u\varphi^k(\mathbf{y})$ where u is a proper suffix of $\varphi^k(0)$ or $\varphi^k(1)$ and $\mathbf{y}$ is some aperiodic binary word.

• In general, not much is known about the general behavior or fluctuations that can be expected for the k-binomial complexity of an infinite word. In particular, computing the k-binomial complexity of a particular infinite word remains quite challenging. It would also be desirable to compare in some ways the k- and $(k+1)$-binomial complexities of a word.

Definition 1.4. For two functions $f, g\colon \mathbb{N} \to \mathbb{N}$, we write $f \prec g$ when the relation $f(n) < g(n)$ holds for infinitely many $n \in \mathbb{N}$.

We define $\prec$ this way because for some words, the 2-binomial complexity attains the factor complexity infinitely often while it is less than the factor complexity infinitely often. See end of Section 7.1 for a discussion.

As an example, a consequence of Proposition 4.17 is that $b_x^{(k)} \prec b_x^{(k+1)}$ for $\mathbf{x} = \varphi^k(\mathbf{y})$ with $\mathbf{y}$ aperiodic. Our reflection is here driven by the following questions inspired by Lejeune's questions [28, pp. 115–117] that are natural to consider in view of (2).

Question A. Does there exist an infinite word $\mathbf{w}$ such that, for all $k \geq 1$, $b_w^{(k)}$ is unbounded and $b_w^{(k)} \prec b_w^{(k+1)}$? If the answer is positive, can we find a (pure) morphic such word $\mathbf{w}$?

From (2), notice that $b_w^{(k)}$ is unbounded, for all $k \geq 1$, if and only if the abelian complexity $b_w^{(1)}$ is unbounded. Even though the Thue–Morse word $\mathbf{t}$ is such that, for all $k \geq 1$, $b_t^{(k)} \prec b_t^{(k+1)}$, $b_t^{(k)}$ remains bounded (3). So $\mathbf{t}$ is not a satisfying answer to Question A. However, in Section 6, we provide several positive answers to this question.

Question B. For each $\ell \geq 1$, does there exist a word $\mathbf{w}$ (depending on ℓ) such that $b_w^{(1)} \prec b_w^{(2)} \prec \cdots \prec b_w^{(\ell-1)} \prec b_w^{(\ell)} = p_w$? If the answer is positive, is there a (pure) morphic such word $\mathbf{w}$?

Putting together results from Sections 4 and 7 we fully answer Question B: Theorem 4.2 and Proposition 4.17 provide a word $\mathbf{x} = \varphi^k(\mathbf{y})$ for which $b_x^{(1)} \prec b_x^{(2)} \prec \cdots \prec b_x^{(k-1)} \prec b_x^{(k)} \prec b_x^{(k+1)}$, while assuming that $\mathbf{y}$ above is Sturmian, we show that $b_x^{(k+2)} = p_x$. We remark that iterates of φ applied to Sturmian words have been studied (among other words) in [21]. We observe that our construction leads to words with bounded abelian complexity. Question B is then strengthened in Section 7 where we ask for words with unbounded abelian complexity. We give a pure morphic answer when $\ell = 3$.

2 Preliminaries

Let us now give precise definitions and notation. For any integer k, we let A^k (resp., $A^{\leq k}$; resp., $A^{<k}$) denote the set of words of length exactly (resp., at most; resp., less than) k over A. We let A^* (resp., A^+) denote the semigroup of finite words (resp., non-empty finite words) over A equipped with concatenation. We let ε denote the empty word. The length of the word w is denoted by $|w|$ and the number of occurrences of a letter a in w is denoted by $|w|_a$. For binary words u, v (always over $\{0,1\}$ in this note, unless otherwise stated), we refer to $|u|_1$ as the *weight* of u and we say that u is *lighter* (resp., *heavier*) than v whenever $|u|_1 < |v|_1$ (resp., $|u|_1 > |v|_1$). For instance, if $b_y^{(1)}(n) = 2$, then there are only two kinds of factors in $\mathbf{y}$: the light ones and the heavy ones. A language L is said to be *balanced* if, for all words $u, v \in L$ of the same length and all letters a, we have $\big||u|_a - |v|_a\big| \leq 1$. In particular, an infinite word $\mathbf{z}$ is *balanced* if $\mathcal{L}(\mathbf{z})$ is balanced.

We let $\overline{\cdot}$ denote the (binary) complementation morphism defined by $\overline{a} = 1 - a$, for $a \in \{0,1\}$. Writing $A = \{a_1, \ldots, a_k\}$ and fixing the order $a_1 < a_2 < \cdots < a_k$ on the letters, the *Parikh vector* of a word $w \in A^*$ is defined as the column vector

$$\Psi(u) = (|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_k})^\intercal.$$

Using a classical "length-n sliding window" argument or extending factors of length n to factors of length $n+1$, one has the following.

Lemma 2.1 (Folklore). *For any binary word* $\mathbf{y}$ *over* $\{0,1\}$*, for all* $n \geq 0$*, we have*

$$\mathsf{b}_{\mathbf{y}}^{(1)}(n) = 1 + \max_{u,v \in \mathcal{L}_n(\mathbf{y})} \big||u|_1 - |v|_1\big| \quad \textit{and} \quad \big|\mathsf{b}_{\mathbf{y}}^{(1)}(n+1) - \mathsf{b}_{\mathbf{y}}^{(1)}(n)\big| \leq 1.$$

2.1 Binomial Equivalence

We first collect some useful results on k-binomial equivalence. Note that $\sim_k$ is a congruence, i.e., for $u, v, x\, y \in A^*$, $u \sim_k v$ and $x \sim_k y$ implies $ux \sim_k vy$. In particular, $A^*/\sim_k$ is a monoid. In fact, it is a cancellative monoid (see [29, Lemma 10]; cancellativity also follows from $A^*/\sim_k$ being isomorphic to a subsemigroup of the special linear group $\mathrm{SL}((k+1)n^k, \mathbb{Z})$ where n is the cardinality of the alphabet A [43]):

Lemma 2.2 (Cancellation property). *Let* u, v, w *be words over* A*. We have*

$$v \sim_k w \Leftrightarrow uv \sim_k uw \textit{ and } v \sim_k w \Leftrightarrow vu \sim_k wu.$$

We will also need the following result characterizing k-binomial commutation among words of equal length.

Theorem 2.3 ([48, Thm. 3.5]). *Let* $k \geq 2$ *and* $x, y \in A^*$ *such that* $|x| = |y|$*. Then* $xy \sim_k yx$ *if and only if* $x \sim_{k-1} y$*.*

A proof of the next result can be conveniently found in [29, Lem. 30].

Theorem 2.4 (Ochsenschläger [36]). *Let* $\varphi\colon 0 \mapsto 01, 1 \mapsto 10$ *be the Thue–Morse morphism. For all* $k \geq 1$*, we have* $\varphi^k(0) \sim_k \varphi^k(1)$ *and* $\varphi^k(0) \not\sim_{k+1} \varphi^k(1)$*.*

The following result from [29, Lem. 31] will be of use. It can alternatively be proved using Theorem 2.3 combined with Ochsenschläger's result.

Lemma 2.5 (Transfer lemma). *Let* $k \geq 1$*. Let* u, v, v' *be three non-empty words such that* $|v| = |v'|$*. We have* $\varphi^{k-1}(u)\varphi^k(v) \sim_k \varphi^k(v')\varphi^{k-1}(u)$*.*

It is an exercise to see that, for an arbitrary morphism $f\colon A^* \to B^*$, we have, for all $u \in A^*$, $e \in B^*$,

$$\binom{f(u)}{e} = \sum_{\substack{a_1,\ldots,a_\ell \in A \\ \ell \leq |e|}} \binom{u}{a_1 \cdots a_\ell} \sum_{\substack{e = e_1 \cdots e_\ell \\ e_i \in B^+}} \prod_{i=1}^{\ell} \binom{f(a_i)}{e_i}. \tag{4}$$

The next result will turn out to be useful in several places of the paper.

Lemma 2.6. *Let* $x, y \in A^*$ *be two* k*-binomially equivalent words. For any integer* $n \geq 0$ *and any word* $e \in A^*$ *of length* $k+1$*, we have*

$$\binom{x^n}{e} - \binom{y^n}{e} = n\left[\binom{x}{e} - \binom{y}{e}\right].$$

In particular, for all $n \geq 1$*,* $x \sim_{k+1} y$ *if and only if* $x^n \sim_{k+1} y^n$*.*

Proof. For any words $u, v, w \in A^*$, we have

$$\binom{uv}{w} = \binom{u}{w} + \binom{v}{w} + \sum_{\substack{w = w_1 w_2 \\ w_i \neq \varepsilon}} \binom{u}{w_1}\binom{v}{w_2}.$$

To show the statement, we proceed by induction and we make use of the previous formula. The statement is trivially true for $n \in \{0, 1\}$. By the previous formula (with $(u, v, w) = (x^n, x, e)$ and $(u, v, w) = (y^n, y, e)$ respectively) and the induction hypothesis, we obtain

$$\binom{x^{n+1}}{e} - \binom{y^{n+1}}{e} = (n+1)\left[\binom{x}{e} - \binom{y}{e}\right] + \sum_{\substack{e=e_1e_2 \\ e_i \neq \varepsilon}} \left[\binom{x^n}{e_1}\binom{x}{e_2} - \binom{y^n}{e_1}\binom{y}{e_2}\right].$$

Since $x \sim_k y$ and $|e| = k+1$, the sum in the right-hand term is zero and we obtain the desired result. □

We recall the following lemma that appears in [48]; it is a straightforward generalization of an observation in [46]. We give a proof for the sake of completeness.

Lemma 2.7. *Let $\mathcal{C} \in A^*/\sim_1$ be an abelian equivalence class of non-empty words with Parikh vector $(m_a)_{a\in A}$. Then, for any word $u \in A^*$, we have $\sum_{w\in\mathcal{C}} \binom{u}{w} = \prod_{a\in A} \binom{|u|_a}{m_a}$.*

Proof. The sum on the left counts the number of ways one can choose a subword w of u so that $\Psi(w) = (m_a)_{a\in A}$. On the other hand, for a vector $(m_a)_{a\in A}$, any choice of m_a many distinct a's in u for each $a \in A$ gives rise to a subword of u having Parikh vector $(m_a)_{a\in A}$. The number of distinct such choices is the product on the right. □

2.2 Binomial equivalence in Sturmian words

The following result links the factor complexity and the 2-binomial complexity of Sturmian words.

Theorem 2.8 ([43, Thm. 7])**.** *For any Sturmian word $\mathbf{s}$, we have $\mathsf{b}_{\mathbf{s}}^{(2)} = \mathsf{p}_{\mathbf{s}}$.*

In particular, the theorem implies that for two distinct equal-length factors u, v of a Sturmian word, we have either $u \not\sim_1 v$, or $\binom{u}{01} \neq \binom{v}{01}$. It further implies that $\mathsf{b}_{\mathbf{s}}^{(k)}(n) = n+1$ for all $k \geq 2$. In the survey paper [18] on abelian combinatorics on words, Fici and Puzynina derive a characterization of Sturmian words from Theorem 2.8:

Theorem 2.9 ([18, Rem. 80])**.** *Let $\mathbf{x}$ be an infinite word. The following are equivalent:*

1. *$\mathbf{x}$ is Sturmian;*
2. *for all $n \geq 1$ and some $k \geq 2$, $\mathsf{b}_{\mathbf{x}}^{(1)}(n) = 2$ and $\mathsf{b}_{\mathbf{x}}^{(k)}(n) = n+1$;*
3. *for all $n \geq 1$ and $k \geq 2$, $\mathsf{b}_{\mathbf{x}}^{(1)}(n) = 2$ and $\mathsf{b}_{\mathbf{x}}^{(k)}(n) = n+1$.*

In fact, the second property in the above can be weakened to "$\mathsf{b}_{\mathbf{x}}^{(1)}(n) = 2$ for all $n \geq 1$ and $\sup_{k,n\in\mathbb{N}} \mathsf{b}_{\mathbf{x}}^{(k)}(n) = \infty$" using the same arguments [1]. In the following we show that the assumption of balancedness can be removed from the second point; Sturmian words are characterized by their k-binomial complexity for any fixed integer $k \geq 2$. We first recall the following crucial observation, which can be found in part of [15, Lem. 4.02'[2]]

Lemma 2.10. *Let $\mathbf{z}$ be an infinite binary word. Let $N \geq 2$ be such that*

1. *$\mathcal{L}(\mathbf{z}) \cap A^{<N}$ is balanced;*
2. *$\mathcal{L}(\mathbf{z}) \cap A^N$ is unbalanced;*
3. *$\mathsf{p}_{\mathbf{z}}(N) = N+1$;*

Then $\mathbf{z}$ is ultimately periodic.

[1] Indeed, the former property implies $\mathbf{x}$ is balanced and binary, and the latter implies that $\mathbf{x}$ is aperiodic.
[2] The statement has a fourth condition, which does not affect the conclusion appearing here.

Theorem 2.11. *Let* $\mathbf{z}$ *be an infinite word such that for some* $k \geq 2$, $\mathbf{b}_{\mathbf{z}}^{(k)} = n + 1$ *for all* n. *Then* $\mathbf{z}$ *is Sturmian.*

Proof. Assume that $\mathbf{b}_{\mathbf{z}}^{(k)}(n) = n + 1$ for all $n \geq 0$. In particular, $\mathbf{z}$ is binary and also aperiodic because $\mathbf{p}_{\mathbf{z}}(n) \geq \mathbf{b}_{\mathbf{z}}^{(k)}(n) = n+1$. This also implies $\mathbf{b}_{\mathbf{z}}^{(1)}(n) \geq 2$ for all $n \geq 1$. To get a contradiction, assume that $\mathbf{z}$ is unbalanced. Hence there exists a minimal integer $N \geq 2$ such that $\mathbf{b}_{\mathbf{z}}^{(1)}(N) = 3$. There is a pair (u, v) of factors of $\mathbf{z}$ of length N such that $|u|_1 - |v|_1 = 2$. The minimality of N implies that this pair is unique and of the form $(1w1, 0w0)$. For details, see [15, Lem. 3.06].

Let $|w|_1 = r$. Let $x \in \mathcal{L}_{N-2}(\mathbf{z})$. If $|x|_1 = r-1$, then $0x$ and $x0$ do not belong to $\mathcal{L}_{N-1}(\mathbf{z})$. Indeed, $1w, w1$ belong to the latter set and $|1w|_1 = |w1|_1 = r+1$ but by minimality of N, the set $\mathcal{L}_{N-1}(\mathbf{z})$ is balanced. In that case, x is preceded and followed by 1. Similarly, if $y \in \mathcal{L}_{N-2}(\mathbf{z})$ and $|y|_1 = r+1$, then y is preceded and followed by 0. This means that $\mathcal{L}_N(\mathbf{z}) \setminus \{0w0, 1w1\}$ is a subset of

$$\{1x1 : |x|_1 = r-1, |x| = N-2\} \cup \{0y0 : |y|_1 = r+1, |y| = N-2\} \cup \bigcup_{a \in \{0,1\}} \{az\overline{a} : |z|_1 = r, |z| = N-2\}$$

where all words have weight $r+1$. Now, observe that $\mathcal{L}(\mathbf{z}) \cap A^{<N}$ is a balanced set (by minimality of N) and that is also the case of $\mathcal{L}_N(\mathbf{z}) \setminus \{1w1\}$. The union of these two sets is factorial. By [41, Thm. 3.1], there exists a Sturmian word $\mathbf{s}$ such that

$$(\mathcal{L}(\mathbf{z}) \cap A^{<N}) \cup (\mathcal{L}_N(\mathbf{z}) \setminus \{1w1\}) \subset \mathcal{L}(\mathbf{s}).$$

As a consequence of Theorem 2.8, any two distinct words in the left-hand side set are not k-binomially equivalent. Also, $1w1$ is not abelian (and thus not k-binomially) equivalent to any word in $\mathcal{L}_N(\mathbf{z}) \setminus \{1w1\}$. In particular, since $\mathbf{b}_{\mathbf{z}}^{(k)}(n) = n + 1$ for $n \leq N$,

$$\mathcal{L}(\mathbf{z}) \cap A^{<N} = \mathcal{L}(\mathbf{s}) \cap A^{<N}$$

and $\#(\mathcal{L}_N(\mathbf{z}) \setminus \{1w1\}) = N$. Therefore, $\#(\mathcal{L}_N(\mathbf{z})) = N+1$. The word $\mathbf{z}$ now fulfills all the conditions of Lemma 2.10 implying the contradiction that $\mathbf{z}$ is ultimately periodic. □

3 Parikh-Collinear Morphisms via Binomial Complexities

In this section, we obtain a new characterization of Parikh-collinear morphisms and show that, given an infinite fixed point of a prolongable Parikh-collinear morphism, its k-binomial complexity is bounded for each k. Note that the automaticity of such fixed points is discussed in [45].

Definition 3.1 (Parikh-collinear morphisms)**.** A morphism $f\colon A^* \to B^*$ is said to be *Parikh-collinear* if, for all letters $a, b \in A$, there is $r_{a,b} \in \mathbb{Q}$ such that $\Psi(f(b)) = r_{a,b}\Psi(f(a))$.

Remark 3.2. Given a morphism $f\colon A^* \to B^*$, its *adjacency matrix* M_f is the matrix of size $\#B \times \#A$ defined by $(M_f)_{b,a} = |f(a)|_b$ for all $a \in A$, $b \in B$. Observe that f is a Parikh-collinear morphism if and only if M_f has rank 1 (unless it is totally erasing). We observe that for any word $u \in A^*$, we have that $\Psi(f(u)) = M_f\Psi(u)$.

Example 3.3. The morphism f defined by $0 \mapsto 000111; 1 \mapsto 0110$ is Parikh-collinear since $\Psi(f(1)) = \frac{2}{3}\Psi(f(0))$.

Theorem 3.4 ([14, Thm. 11])**.** *A morphism* $f\colon A^* \to B^*$ *is Parikh-collinear if and only if it maps all infinite words to words with bounded abelian complexity.*

We extend the above theorem to the following one. We say that a morphism $f\colon A^* \to B^*$ *satisfies* P_k if f maps all words with bounded k-binomial complexity to words with bounded $(k+1)$-binomial complexity. Note that, for $k = 0$, *0-binomial complexity* has to be understood as the "equal length" equivalence relation. So the 0-binomial complexity of an infinite word is the constant function 1 and Theorem 3.4 can be restated as f *is Parikh-collinear if and only if* f *satisfies* P_0.

Theorem 3.5. *Let* $f\colon A^* \to B^*$ *be a morphism. The following are equivalent.*

(i) The morphism f *is Parikh-collinear.*

(ii) For all $k \geq 0$, f *satisfies* P_k.

(iii) There exists an integer $k \geq 0$ *such that* f *satisfies* P_k.

Before proving this result in Section 3.2, let us mention a straightforward consequence, which generalizes [43, Thm. 13] from Parikh-constant to Parikh-collinear morphisms. For example, the Thue–Morse morphism is Parikh-constant and thus Parikh-collinear but the morphism of Example 3.3 is Parikh-collinear but not Parikh-constant.

Corollary 3.6. *Let* $\mathbf{z}$ *be a fixed point of a Parikh-collinear morphism. For any* $k \geq 1$ *there exists a constant* $C_{\mathbf{z},k} \in \mathbb{N}$ *such that* $\mathfrak{b}_{\mathbf{z}}^{(k)}(n) \leq C_{\mathbf{z},k}$ *for all* $n \in \mathbb{N}$.

Proof. Let $f\colon A^* \to A^*$ be a Parikh-collinear morphism whose fixed point is $\mathbf{z}$. Since $f(\mathbf{z}) = \mathbf{z}$, Theorem 3.4 implies that $\mathbf{z}$ has bounded abelian complexity. For any $k \geq 1$, we have that $\mathbf{z} = f(f^{k-1}(\mathbf{z}))$ implying that $\mathbf{z}$ has bounded k-binomial complexity by induction and the previous theorem. □

Remark 3.7. We cannot relax the (implicit) assumption on the rank of the adjacency matrix M_f in Corollary 3.6. For example, the morphism $f : \{0,1,2\}^* \to \{0,1,2\}^*$ defined by $0 \mapsto 0^3 2^3$, $1 \mapsto 0^3 1^3 2$, $2 \mapsto 2^4 0^6 1^3$ has an adjacency matrix of rank 2. The fixed point $\mathbf{x}$ starting with 0 is aperiodic as $f^n(0)$ is readily seen to be right special for all $n \geq 0$. Yet, its adjacency matrix has eigenvalues $\theta_1 = 5 + \sqrt{13}$, $\theta_2 = 5 - \sqrt{13}$, and 0, and the former two are greater than 1. This means that the word has unbounded abelian complexity. Indeed, a deep result of Adamczewski on balances in primitive pure morphic words [1, Thm. 13(ii)] implies that the "lim sup"-growth of the function

$$n \mapsto \max_{a \in \Sigma, u, v \in \mathcal{L}_n(\mathbf{x})} \left\{ \left| |u|_a - |v|_a \right| \right\}$$

grows as $\Theta(n^{\log_{\theta_1} \theta_2})$, where $\log_{\theta_1} \theta_2 \approx 0{,}15448$. It follows (see, e.g., [40, Lem. 2.2]) that $\mathfrak{b}_{\mathbf{x}}^{(k)}$ is unbounded for each $k \geq 1$.

3.1 An Intermediate Characterization of Parikh-Collinearity

To prove Theorem 3.5, we give further characterizations of Parikh-collinear morphisms. To this end, we require the following lemma where we define a map g_e which is constant on any abelian equivalence class. Notice that such a map appears within (4).

Lemma 3.8. *Let* A, B *be finite alphabets with* $\#A \geq 2$. *Let* $f\colon A^* \to B^*$ *be a Parikh-collinear morphism. For a word* $e = e_1 \cdots e_n$ *of length* n *over* B, *define* $g_e\colon A^n \to \mathbb{N}$ *by*

$$g_e(a_1 \cdots a_n) := \prod_{i=1}^{n} \binom{f(a_i)}{e_i}.$$

Then, for all words $w, w' \in A^n$ *with* $w \sim_1 w'$, *we have* $g_e(w) = g_e(w')$.

Proof. Write $w = a_1 \cdots a_n$ with $a_i \in A$ for all $i \in \{1, \ldots, n\}$. For all $\alpha \in A$ and $\beta \in B$, define $I(\alpha, \beta) := \{i \in \{1, \ldots, n\} \mid a_i = \alpha \text{ and } e_i = \beta\}$. We get

$$g_e(w) = \prod_{\substack{\alpha \in A \\ \beta \in B}} \prod_{i \in I(\alpha,\beta)} \binom{f(\alpha)}{\beta}.$$

The claim is trivial if f maps all words to ε, so let $0 \in A$ be a letter for which $|f(0)| \neq 0$. Since the morphism f is Parikh-collinear, for all $\alpha \in A$ and all $\beta \in B$, there exists $r_\alpha \in \mathbb{Q}$ such that $\binom{f(\alpha)}{\beta} = r_\alpha \binom{f(0)}{\beta}$. We now get

$$\begin{aligned} g_e(w) &= \prod_{\substack{\alpha\in A\\ \beta\in B}} \prod_{i\in I(\alpha,\beta)} \binom{f(\alpha)}{\beta} = \prod_{\substack{\alpha\in A\\ \beta\in B}} \prod_{i\in I(\alpha,\beta)} r_\alpha \binom{f(0)}{\beta} \\ &= \left(\prod_{\substack{\alpha\in A\\ \beta\in B}} \prod_{i\in I(\alpha,\beta)} \binom{f(0)}{\beta}\right)\left(\prod_{\substack{\alpha\in A\\ \beta\in B}} \prod_{i\in I(\alpha,\beta)} r_\alpha\right). \end{aligned}$$

For any letter $\beta \in B$, the definition of $I(\alpha, \beta)$ gives

$$\prod_{\alpha\in A} \prod_{i\in I(\alpha,\beta)} \binom{f(0)}{\beta} = \binom{f(0)}{\beta}^{|e|_\beta}.$$

Similarly, for any letter $\alpha \in A$, the definition of $I(\alpha, \beta)$ yields

$$\prod_{\beta\in B} \prod_{i\in I(\alpha,\beta)} r_\alpha = r_\alpha^{|w|_\alpha}.$$

Thus

$$g_e(w) = \left(\prod_{\beta\in B} \binom{f(0)}{\beta}^{|e|_\beta}\right)\left(\prod_{\alpha\in A} r_\alpha^{|w|_\alpha}\right).$$

Observe that the first factor in this product only depends on (the Parikh vector of) e — in particular, not on w — as the morphism f is fixed. Similarly, the second factor in the product depends solely on the Parikh vector of w, not on the word w itself. The desired result follows. □

We now characterize Parikh-collinear morphisms by means of binomial complexities.

Proposition 3.9. *Let $f\colon A^* \to B^*$ be a morphism. The following are equivalent.*

(i) The morphism f is Parikh-collinear.

(ii) For all $k \geq 2$ and $u, v \in A^*$, $u \sim_{k-1} v$ implies $f(u) \sim_k f(v)$.

(iii) There exists an integer $k \geq 2$ such that for all $u, v \in A^*$, $u \sim_{k-1} v$ implies $f(u) \sim_k f(v)$.

(iv) For all $u, v \in A^*$, $u \sim_1 v$ implies $f(u) \sim_2 f(v)$.

Proof. Clearly (ii) implies (iii). We show that (iii) implies (iv). There is nothing to prove if (iii) holds for $k = 2$, so assume that $k \geq 3$. We show that f also satisfies (iii) with $k - 1$ instead of k, and hence, by repeating the argument, f satisfies (iii) with $k = 2$. Assume to the contrary that there exists a pair u, v such that $u \sim_{k-2} v$ but $f(u) \not\sim_{k-1} f(v)$. Since u and v are abelian equivalent ($k-2 \geq 1$) they have equal length, so by Theorem 2.3, we have that $uv \sim_{k-1} vu$. Then, since f has the property for k, we have $f(u)f(v) \sim_k f(v)f(u)$. Furthermore, $f(u)$ and $f(v)$ have the same length (due to $u \sim_1 v$). This implies that $f(u) \sim_{k-1} f(v)$ by the converse part of Theorem 2.3, contrary to what was assumed.

Assuming (iv), we show that (i) holds. Let x, y be distinct letters from A. Since $xy \sim_1 yx$, we have $f(xy) \sim_2 f(yx)$ by assumption. In other words, for all $s, t \in B$ we have, applying (4),

$$\begin{aligned} 0 = \binom{f(xy)}{st} - \binom{f(yx)}{st} &= \sum_{\substack{a_1,\dots,a_\ell\in A\\ \ell\leq 2}} \left[\binom{xy}{a_1\cdots a_\ell} - \binom{yx}{a_1\cdots a_\ell}\right] \sum_{\substack{st=b_1\cdots b_\ell\\ b_i\in B^+}} \prod_{i=1}^{\ell} \binom{f(a_i)}{b_i} \\ &= \sum_{a_1,a_2\in A} \left(\binom{xy}{a_1a_2} - \binom{yx}{a_1a_2}\right)\binom{f(a_1)}{s}\binom{f(a_2)}{t} = \binom{f(x)}{s}\binom{f(y)}{t} - \binom{f(y)}{s}\binom{f(x)}{t}, \end{aligned}$$

where in the third equality we use $\binom{xy}{a} = \binom{yx}{a}$ for all $a \in A$ (since $xy \sim_1 yx$). Summing over $s \in B$, we get $|f(x)|\binom{f(y)}{t} = |f(y)|\binom{f(x)}{t}$ for all $t \in B$. Now x and y were chosen arbitrarily from the alphabet A. If $|f(x)| = 0$ for all $x \in A$, then f is clearly Parikh-collinear. If there is a letter x for which $|f(x)| > 0$, we may write $\left(\binom{f(y)}{t}\right)_{t\in B} = \frac{|f(y)|}{|f(x)|}\left(\binom{f(x)}{t}\right)_{t\in B}$ for each $y \in A$. In other words, f is Parikh-collinear.

To complete the proof, we show that (i) implies (ii). So let f be a Parikh-collinear morphism and $u \sim_{k-1} v$ with $k \geq 2$. We again apply (4): for any word $e \in B^*$, we have

$$\binom{f(u)}{e} - \binom{f(v)}{e} = \sum_{\substack{a_1,\ldots,a_\ell \in A \\ \ell \leq |e|}} \left(\binom{u}{a_1 \cdots a_\ell} - \binom{v}{a_1 \cdots a_\ell}\right) \sum_{\substack{e = e_1 \cdots e_\ell \\ e_i \in B^+}} \prod_{i=1}^{\ell} \binom{f(a_i)}{e_i}.$$

Notice that for words $e \in B^{<k}$, we have $\binom{u}{a_1\cdots a_\ell} = \binom{v}{a_1\cdots a_\ell}$ since $u \sim_{k-1} v$, which in turn gives $\binom{f(u)}{e} = \binom{f(v)}{e}$. So to show that $f(u) \sim_k f(v)$, it suffices to consider words $e \in B^k$. By assumption, for $\ell < k$, we again have $\binom{u}{a_1\cdots a_\ell} = \binom{v}{a_1\cdots a_\ell}$. Therefore, we have $\binom{f(u)}{e} = \binom{f(v)}{e}$ if and only if

$$\sum_{a_1,\ldots,a_k \in A} \binom{u}{a_1 \cdots a_k} \prod_{i=1}^{k} \binom{f(a_i)}{e_i} = \sum_{a_1,\ldots,a_k \in A} \binom{v}{a_1 \cdots a_k} \prod_{i=1}^{k} \binom{f(a_i)}{e_i}. \tag{5}$$

Observe here that $\prod_{i=1}^{k} \binom{f(a_i)}{e_i} = g_e(a_1 \cdots a_k)$ as defined in Lemma 3.8. Let $\mathcal{C}$ be an abelian equivalence class in $A^k/\sim_1$. By Lemma 3.8, $g_e(\cdot)$ is constant on $\mathcal{C}$, so write $g_e(w) = g_{\mathcal{C},e}$ for all words $w \in \mathcal{C}$. For each $w \in \mathcal{C}$ we may write $\Psi(w) = (m_{\mathcal{C},a})_{a\in A}$. We now have

$$\sum_{w\in A^k} \binom{u}{w} g_e(w) = \sum_{\mathcal{C}\in A^k/\sim_1} \sum_{w\in\mathcal{C}} \binom{u}{w} g_e(w) = \sum_{\mathcal{C}\in A^k/\sim_1} g_{\mathcal{C},e} \sum_{w\in\mathcal{C}} \binom{u}{w} = \sum_{\mathcal{C}\in A^k/\sim_1} g_{\mathcal{C},e} \prod_{a\in A} \binom{|u|_a}{m_{\mathcal{C},a}},$$

where the last equality is from Lemma 2.7. One obtains the same formula by replacing u with v, and equality indeed holds in (5) as $|u|_a = |v|_a$ for each letter $a \in A$. This concludes the proof. □

Remark 3.10. In [34, Lem. 5], the authors show that, for a morphism f such that $f(a) \sim_h f(b)$ for all $a, b \in A$, for all words $u, v \in A^*$ with $u \sim_k v$ we have that $f(u) \sim_{k+h} f(v)$. Towards the converse, assume that f is a morphism for which the conclusion holds (for all $k \geq 1$ but fixed $h \geq 1$). Then we necessarily have $f(a)^m f(b)^n \sim_{h+1} f(b)^n f(a)^m$ for all $a, b \in A$, $m, n \geq 1$. From this we infer that, e.g., $f(a)^{|f(b)|} \sim_h f(b)^{|f(a)|}$ (as a corollary of Theorem 2.3). In particular, if f is uniform, we have $f(a) \sim_h f(b)$ for all $a, b \in A$. It would be interesting to characterize the non-uniform morphisms f with this property. For example, one can take any Parikh-collinear morphism g; then $f = g^h$ is such a morphism. We highly suspect that these are not the only such morphisms.

3.2 Proof of Theorem 3.5

We require the following technical result, which essentially appears in the proof of [14, Thm. 12]. We give a proof here for the sake of completeness.

Lemma 3.11. *Let* $\mathbf{x}$ *be a an infinite word over* A *with bounded abelian complexity. Let* $f : A^* \to B^*$ *be a morphism and assume* $\mathbf{y} = f(\mathbf{x})$ *is an infinite word. Then for all* $c \in \mathbb{N}$ *there exists* $D_{\mathbf{x},c} \in \mathbb{N}$ *such that if* $\big||f(u)| - |f(v)|\big| \leq c$, *for some* $u, v \in \mathcal{L}(\mathbf{x})$, *then* $\big||u| - |v|\big| \leq D_{\mathbf{x},c}$.

Proof. Assume without loss of generality that $|u| \geq |v|$ and write $u = u'v'$ with $|v'| = |v|$. Let M_f be the adjacency matrix of f. If $\big||f(u)| - |f(v)|\big| \leq c$, we have by the reverse triangle inequality

$$c \geq \big||f(u')| - |f(v)| + |f(v')|\big| \geq |f(u')| - \big||f(v')| - |f(v)|\big| = |f(u')| - |\langle M_f(\Psi(v') - \Psi(v)), \vec{1}\rangle|,$$

where $\langle \cdot\,,\,\cdot \rangle$ denotes the inner product of vectors, and $\vec{1}$ is the all-ones-vector. Recall that $\mathbf{x}$ has bounded abelian complexity if and only if it is C-balanced for some C [40]. Hence, as v and v' are factors of the same length, $\Psi(v') - \Psi(v)$ attains finitely many distinct integer points (in particular, belonging to $[-C, C]^{\#A}$). So does $M_f(\Psi(v') - \Psi(v))$. We therefore obtain $|f(u')| \leq D$ for some $D \in \mathbb{N}$. We deduce that u' is bounded in length as well: indeed, let $a \in A$ be a letter occurring infinitely often in $\mathbf{x}$ and for which $f(a) \neq \varepsilon$ (such a letter exists because $f(\mathbf{x})$ is infinite). Since $\mathbf{x}$ is balanced, we deduce that all long enough factors of $\mathbf{x}$ contain more than $|u'|$ occurrences of a. We let $D_{\mathbf{x},c}$ be this bound on $|u'|$ to conclude the proof. □

We are now ready to prove the main result of this section, characterizing Parikh-collinear morphisms in terms the property P_k defined at the beginning of Section 3.

Proof of Theorem 3.5. Let us first show that (i) implies (ii). Assume thus that f is Parikh-collinear. Theorem 3.4 implies that f maps all words (i.e., all words with bounded 0-binomial complexity) to words with bounded 1-binomial complexity. Let $k \geq 1$ and let $\mathbf{x}$ be a word with bounded k-binomial complexity. Let $n \in \mathbb{N}$. Any length-n factor of $f(\mathbf{x})$ can be written as $pf(u)s$, where the word u is a factor of $\mathbf{x}$, p is a suffix of $f(a)$ and s is a prefix of $f(b)$ for some letters $a, b \in A$. Here $n - 2m < |f(u)| \leq n$, where $m := \max_{a \in A} |f(a)|$. The $(k+1)$-binomial equivalence class of $pf(u)s$ is completely determined by the words p, s, and the k-binomial equivalence class of $f(u)$, which itself is determined by the abelian equivalence class of u by Proposition 3.9.

The former two words p and s are drawn from a finite set, as their lengths are bounded by the constant m (depending on f). The length of u can be chosen from an interval whose length is uniformly bounded in n. Indeed, assume we have equal length factors $w = pf(u)s$ and $w' = p'f(v)s'$. As observed above, $n \geq |f(u)|$ and $|f(v)| > n - 2m$, so that $\big||f(u)| - |f(v)|\big| < 2m$. Applying Lemma 3.11 (by assumption, $\mathbf{x}$ has bounded k-binomial complexity and thus, $\mathbf{x}$ has bounded abelian complexity by (2)) there exists a bound D such that $\big||u| - |v|\big| \leq D$ uniformly in n. Since the number of k-binomial equivalence classes in $\mathbf{x}$ of each length is uniformly bounded by assumption, and the number of admissible lengths for u above is bounded, we conclude that the number of choices for the k-binomial equivalence class of u is bounded. We have shown that the number of $(k+1)$-binomial equivalence classes among factors of length n in $f(\mathbf{x})$ is determined from a bounded amount of information (not depending on n), as was to be shown. Consequently, f satisfies P_k.

Notice that (ii) trivially implies (iii).

Let us turn to the last implication, namely (iii) implies (i). Assume (iii) holds, that is, for some integer $k \geq 0$, f satisfies P_k. If $k = 0$, then f maps all words to words with bounded 1-binomial complexity, so f is Parikh-collinear by Theorem 3.4. Assume that $k \geq 1$, and towards a contradiction, assume further that f is not Parikh-collinear. By Proposition 3.9, there exist words u, v with $u \sim_k v$ and $f(u) \not\sim_{k+1} f(v)$. Write $U = f(u)$ and $V = f(v)$. Now define the word $\mathbf{x} = uvu^2v^2u^3v^3 \cdots u^n v^n \cdots$ and consider

$$f(\mathbf{x}) = UVU^2V^2U^3V^3 \cdots U^n V^n \cdots .$$

Below we show that $\mathbf{x}$ has bounded k-binomial complexity, while $f(\mathbf{x})$ has unbounded $(k+1)$-binomial complexity, which is enough to contradict (iii).

Since u and v are k-binomially equivalent, $b_{\mathbf{x}}^{(k)}$ is bounded. (To see this, one may apply arguments similar to those developed in the first part of the proof.) Let us prove the second. For each integer n, $f(\mathbf{x})$ contains the factors $U^r V^{n-r}$ with $r \in \{0, \ldots, n\}$. These factors are actually all $(k+1)$-binomially inequivalent. Indeed, assume towards a contradiction that $U^r V^{n-r} \sim_{k+1} U^s V^{n-s}$ for some $r, s \in \{0, \ldots, n\}$ with $r > s$. By the Cancellation property (Lemma 2.2), we obtain $U^{r-s} \sim_{k+1} V^{r-s}$. Lemma 2.6 then implies that $U \sim_{k+1} V$, which is a contradiction. Consequently, $b_{f(\mathbf{x})}^{(k+1)}$ is unbounded, as desired. □

4 Binomial Properties of the Thue–Morse Morphism, Part I

In this section, we consider binomial complexities of iterates of the Thue–Morse morphism φ on aperiodic binary words. The section is split into three subsections. To state the main result, we define the following.

Definition 4.1. Let $\mathbf{x}$ be a binary word and $k \geq 1$ an integer. We say that $\mathbf{x}$ has property $\mathcal{TMB}(k)$ if, for all $1 \leq j \leq k$, we have $\mathsf{b}_{\mathbf{x}}^{(j)} = \mathsf{b}_{\mathbf{t}}^{(j)}$.

Recall that the exact values for $\mathsf{b}_{\mathbf{t}}^{(j)}$ were computed in [29, Thm. 6] (and are given by (3)). The main result of Section 4.2 is the following theorem, which can be seen as a generalization of the aforementioned result.

Theorem 4.2. *Let k be an integer and let $\mathbf{y}$ be an aperiodic binary word. Then the word $\mathbf{x} = \varphi^k(\mathbf{y})$ (and any of its suffixes) has property $\mathcal{TMB}(k)$.*

The application of φ^k to a word changes the j-binomial complexity, $j \leq k$, to that of the Thue–Morse word. Putting this bluntly, the binomial complexities of the original word play no role in the j-binomial complexities of the image word (for small j).

The topic of Section 4.3 is to characterize the k- and $(k+1)$-binomial equivalence among factors of words of the form $\varphi^k(\mathbf{y})$ (Theorem 4.12 and Proposition 4.17). In the latter, we see that structure of $\mathbf{y}$ already appears to affect the $(k+1)$-binomial complexity of $\varphi^k(\mathbf{y})$. This allows to conclude, for example, that $\mathsf{b}^{(k)} \prec \mathsf{b}^{(k+1)}$ (Corollary 4.18) for words of this form. Throughout the rest of this section we fix $\mathbf{x}$ and $\mathbf{y}$ to be as in Theorem 4.2.

We begin with a subsection introducing a convenient tool, called abelian Rauzy graphs, which we use throughout the current and the following section.

4.1 Abelian Rauzy graphs

For an infinite word $\mathbf{z} \in A^{\mathbb{N}}$, consider a sliding window of length n: as the window shifts, one goes from heavier factors to lighter factors and vice versa. One can consider a directed labeled graph $G = (V, E)$ capturing its progress: the vertices are the Parikh vectors of factors of length n, and there is an edge from $\vec{x}$ to $\vec{y}$ labeled with $(a, b) \in A \times A$, if there exists $aub \in \mathcal{L}_{n+1}(\mathbf{z})$ such that $\Psi(au) = \vec{x}$ and $\Psi(ub) = \vec{y}$. We call G the *abelian Rauzy graph (of order n)*. Such graphs were considered already in [39].

Remark 4.3. The abelian Rauzy graph $G = (V, E)$ defined here is a quotient of the usual Rauzy graph of $\mathbf{z}$ of order n. The latter one is defined as $R = (V', E')$ where $V' = \mathcal{L}_n(\mathbf{z})$ and there is an edge from au to ub of label (a, b) whenever $aub \in \mathcal{L}_{n+1}(\mathbf{z})$. The Parikh map $\Psi : V' \to V$ is a morphism of graphs: any labeled path in R is mapped to a path in G with same label.

We describe some properties of abelian Rauzy graphs.

Observation 4.4. Let $G = (V, E)$ be the abelian Rauzy graph of order n of an infinite word $\mathbf{z} \in A^{\mathbb{N}}$.

- The number of vertices is $\#V = \mathsf{b}_{\mathbf{z}}^{(1)}(n)$;
- An edge with label (a, b) corresponds to an increase in weight if and only if $ab = 01$;
- An edge with label (a, b) corresponds to a decrease in weight if and only if $ab = 10$;
- An edge is a loop if and only if $a = b$.
- A right special factor of length n gives rise to a *right special vertex*: a vertex with two outgoing edges for which the labels have the same first component. If $\mathbf{z}$ is binary, then one of the two edges is a loop.

- Each vertex has at least one outgoing edge (possibly a loop). Furthermore, if the word $\mathbf{z}$ is aperiodic, each vertex has at least one outgoing edge that is not a loop, and there is at least one vertex with two outgoing edges. In particular, the graph has at least $\#V + 1$ edges.

Example 4.5. Let us consider the abelian Rauzy graphs of the Thue–Morse word. For a fixed n, we identify the vertices of G, that is, the Parikh vectors of factors of length n, with their second components. Indeed, for a binary word u of a fixed length n, we have $\Psi(u) = (n - |u|_1, |u|_1)^\intercal$. We show that, for all $m \geq 1$, G_{2m} and G_{2m+1} take forms as depicted in Fig. 1.

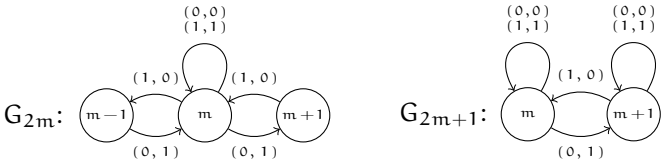

Figure 1: The abelian Rauzy graphs of the Thue–Morse word at even and odd lengths.

Let us write $\mathbf{t} = 01101001\cdots = a_0a_1a_2\cdots$. We first consider G_{2m}. It is well-known and plain to see that $\mathbf{t}$ is closed under complementation; for all $u \in \mathcal{L}(\mathbf{t})$, we have $\overline{u} \in \mathcal{L}(\mathbf{t})$. Let $au, a \in \{0,1\}$, $au \in \mathcal{L}_m(\mathbf{t})$, be right special in $\mathbf{t}$. Consequently, $\varphi(au)a \in \mathcal{L}(\mathbf{t})$; hence G_{2m} contains the loop $m \xrightarrow{(a,a)} m$. By complementing such a factor, we also find $m \xrightarrow{(\overline{a},\overline{a})} m$ in G_{2m}. To see that, e.g., there is no loop at $m-1$, we note that any $m-1$-factor is of the form $0\varphi(u')0$, and appears at an odd position in $\mathbf{t}$. Hence there is certainly no loop with label $(1,1)$ at $m-1$. Neither can there be a loop with label $(0,0)$, as 00 is not the image of a letter.

We then inspect G_{2m+1}. The following statements are easy to prove using, e.g., the automatic prover `Walnut` [35]. For a comprehensive take on the usage of `Walnut`, we recommend the book [47].

- For all $m \geq 1$ there exists a length-$(2m+2)$ factor u starting at an even index in $\mathbf{t}$, and which begins and ends with 1.
  The following `Walnut` formula returns 'TRUE':
  `eval OddL11 "Am (m>0) => Ej T[2*j]=@1 & T[2*j+2*m+1]=@1";`

- For all $m \geq 0$ there exists a length-$(2m+2)$ factor starting at an odd index in $\mathbf{t}$, and which begins and ends with 0.
  The following `Walnut` formula returns 'TRUE':
  `eval OddL00 "Am Ej j>0 & T[2*j-1]=@0 & T[2*j+2*m]=@0";`

Then the first item implies that G_{2m+1} contains the loop $m \xrightarrow{(1,1)} m$. Indeed, the length-$(2m+1)$ prefix of u in the first item has weight m (because $u = \varphi(u')01$ for some $|u'| = m$). Similarly the second item implies that m has a loop with label $(0,0)$. Recalling that $\mathbf{t}$ is closed under complementation, G_{2m+1} is seen to be of the claimed form.

We show the structure of the abelian Rauzy graphs of Sturmian words in Proposition 7.8.

We shall make use of the following general lemma, a part of which appears as [40, Lem. 3.2.]

Lemma 4.6. *Let $\mathbf{z}$ be an aperiodic binary word. Then, for every $n \geq 1$, the set of edge labels of the abelian Rauzy graph G_n contains $(0,1)$, $(1,0)$, and either $(0,0)$ or $(1,1)$. Furthermore, if (a,a) does not appear as a label of a loop in G_n, then G_{n+1} contains a loop with the label $(\overline{a},\overline{a})$.*

Proof. Since $\mathbf{z}$ is aperiodic, it must have a connected component containing a loop (a right special vertex) and at least two vertices (by a theorem of Coven and Hedlund [15]). An edge from a lighter vertex to a heavier one has label $(0,1)$, and $(1,0)$ appears as the label from the heavier to the lighter one.

Assume that G_n does not contain a loop with label $(0,0)$. Consider the walk W in G_n defined by $\mathbf{z}$: in particular, consider the strongly connected subgraph of G_n comprising the vertices and

edges that W traverses infinitely many times. There is an edge from the second lightest vertex to the lightest one 1, labeled with $(1,0)$. This means that the factor of length $n+1$ corresponding to this edge begins with 1 and ends with 0. Since $(0,0)$ does not appear as a label in G_n and from the lightest vertex there is no outgoing edge with label $(1,0)$, the edge that W takes from 1 has label $(\cdot,1)$. Thus the factor of length $n+2$ begins and ends with 1. This gives an edge in G_{n+1} with label $(1,1)$. □

4.2 The First k Binomial Complexities

We begin by defining the notion of φ^j-factorizations of factors of $\mathbf{x}$. This will be used throughout this and the next section.

Definition 4.7. For any factor u of $\varphi^j(\mathbf{y})$ of length at least 2^j-1 there exist $a,b\in\{0,1\}$ and $z\in\{0,1\}^*$ with $azb\in\mathcal{L}(\mathbf{y})$ such that $u=p\varphi^j(z)s$ for some proper suffix p of $\varphi^j(a)$ and some proper prefix s of $\varphi^j(b)$. (Note that z could be empty.) The triple $(p,\varphi^j(z),s)$ is called a φ^j-*factorization*[3] of u. The word azb (resp., zb; az; z) is said to be the corresponding φ^j-*ancestor* of u when p,s are non-empty (resp., $p=\varepsilon$ and $s\neq\varepsilon$; $p\neq\varepsilon$ and $s=\varepsilon$; $p=s=\varepsilon$).

Since the words $\varphi^j(0)$ and $\varphi^j(1)$ begin with different letters, we notice that if $s\neq\varepsilon$ in a φ^j-factorization of a word, then the letter b is uniquely determined. Similarly the jth images of the letters end with distinct letters (for j fixed), whence the letter a is uniquely determined once $p\neq\varepsilon$.

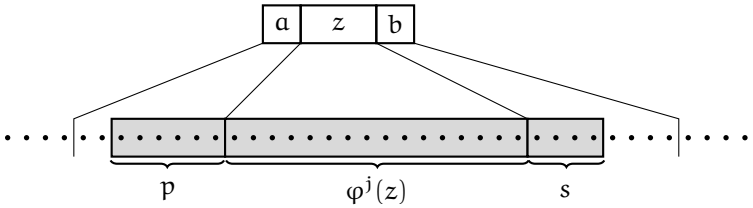

Figure 2: A φ^j-factorization and its φ^j-ancestor.

The following lemma says that an aperiodic word of the form $\varphi^k(\mathbf{y})$ has the same short factors as the Thue–Morse word.

Lemma 4.8. *For an aperiodic binary word* $\mathbf{y}$ *and integer* k*, we have* $\mathcal{L}_n(\varphi^k(\mathbf{y}))=\mathcal{L}_n(\mathbf{t})$ *for all* $n\leq 2^k$.

Proof. Let $\mathbf{x}=\varphi^k(\mathbf{y})$. The claim is trivial for $k=0$ so assume $k\geq 1$. Any factor of $\mathbf{x}$ of length at most 2^k appears as a factor of $\varphi^k(v)$, where $v\in\mathcal{L}_2(\mathbf{y})$. Since $\mathcal{L}_2(\mathbf{t})=\{0,1\}^2$, all such factors appear in $\varphi^k(\mathbf{t})=\mathbf{t}$. In particular we have shown $\mathcal{L}_n(\mathbf{x})\subseteq\mathcal{L}_n(\mathbf{t})$ for $n\leq 2^k$.

Since $\mathbf{y}$ is aperiodic it contains both 01 and 10 and either of 00 and 11. If $\mathbf{y}$ contains all factors of length 2, then clearly the considered languages are equal. So assume without loss of generality that 11 does not appear in $\mathbf{y}$. Consider the factors of length at most 2^k of $\varphi^k(11)=\varphi^{k-1}(1)\varphi^k(0)\varphi^{k-1}(0)$. Such a factor is either a factor of $\varphi^{k-1}(1)\varphi^k(0)$ or $\varphi^k(0)\varphi^{k-1}(0)$, both of which are factors of $\varphi^k(00)$. Hence all factors of length at most 2^k appearing in $\mathbf{t}$ appear in $\mathbf{x}$ as well; this concludes the proof. □

We are in the position to prove the main result of this subsection.

Proof of Theorem 4.2. Let $j\in\{1,\ldots,k\}$. We first prove the claim for $\mathbf{x}$ (and afterwards the claim for any of its suffixes). As the factors of length at most 2^j of $\mathbf{x}=\varphi^j(\varphi^{k-j}(\mathbf{y}))$ coincide with those

[3]We warn the reader that the term φ-factorization has a different meaning in [29]. Our φ^j-factorization corresponds to their "factorization of order j".

of $\mathbf{t}$ by the above lemma, the j-binomial complexity of $\mathbf{x}$ coincides with that of the Thue–Morse word's for $n < 2^j$.

In the remaining of the proof we let $n \geq 2^j$. We show that $\mathcal{L}_n(\mathbf{t})/\sim_j = \mathcal{L}_n(\mathbf{x})/\sim_j$ by double inclusion, which suffices for the claim since Theorem 4.2 holds true for $\mathbf{x} = \mathbf{t}$.

Let $u \in \mathcal{L}(\mathbf{x})$; we show that there exists $v \in \mathcal{L}(\mathbf{t})$ such that $u \sim_j v$. To this end, let $\mathbf{z} = \varphi^{k-j}(\mathbf{y})$ so that $\mathbf{x} = \varphi^j(\mathbf{z})$. Let u have φ^j-factorization $p\varphi^j(u')s$ with φ^j-ancestor $au'b \in \mathcal{L}(\mathbf{z})$. The Thue–Morse word contains a factor $av'b$, where $|v'| = |u'|$ (see, e.g., [29, Prop. 33] or the abelian Rauzy graphs in Example 4.5). It follows that $\mathbf{t}$ contains the factor $v := p\varphi^j(v')s$. Now $u \sim_j v$ because $\varphi^j(u') \sim_j \varphi^j(v')$ by Theorem 2.4.

Let then $u \in \mathcal{L}(\mathbf{t})$ have φ^j-factorization $p\varphi^j(u')s$ with φ^j-ancestor $au'b \in \mathcal{L}(\mathbf{t})$. As before we show that there exists $v \in \mathcal{L}(\mathbf{x})$ such that $u \sim_j v$. As a consequence of Lemma 4.6, $\mathbf{z}$ contains, at each length, factors from both the languages $0A^*1$ and $1A^*0$. Hence, if a and b above are distinct, we may argue as in the previous paragraph to obtain the desired conclusion. Assume thus that $a = b$. Again Lemma 4.6 implies that $\mathbf{z}$ contains a factor of length $|u'| + 2$ in the language $1A^*1 \cup 0A^*0$. Assume without loss of generality that it contains a factor from $0A^*0$. Then, if $a = b = 0$, we may again argue as in the previous paragraph. So assume now that $a = b = 1$ and $\mathcal{L}_{|u|'+2}(\mathbf{z}) \cap 1A^*1 = \emptyset$. Notice that Lemma 4.6 implies that $\mathcal{L}_{|u|'+2}(\mathbf{z}) \cap 0A^*0 \neq \emptyset$ and, further, $\mathcal{L}_{|u|'+2\pm1}(\mathbf{z}) \cap 0A^*0 \neq \emptyset$. To conclude with the claim for $\mathbf{x}$, we have four cases to consider depending on the length of p and s which can be less or equal, or greater than 2^{j-1}.

Case 1: Assume that p is a suffix of $\varphi^{j-1}(0)$ and s is a prefix of $\varphi^{j-1}(1)$. For all v' such that $|v'| = |u'| - 1$, $\varphi^j(u') \sim_j \varphi^j(v'1)$ by Theorem 2.4. By the Transfer Lemma (Lemma 2.5), $\varphi^j(v'1) \sim_j \varphi^{j-1}(1)\varphi^j(v')\varphi^{j-1}(0)$. Consequently

$$u \sim_j p\varphi^{j-1}(1)\varphi^j(v')\varphi^{j-1}(0)s =: v,$$

where $p\varphi^{j-1}(1)$ is a suffix of $\varphi^j(0)$ and $\varphi^{j-1}(0)s$ is a prefix of $\varphi^j(0)$. Hence v is a factor of $\varphi^j(0v'0)$. Recall that a factor of the form $0v'0$ appears in $\mathbf{z}$ by assumption, and thus $\varphi^j(0v'0)$ appears in $\mathbf{x}$. To recap, we have shown a factor v of $\mathbf{x}$ j-binomially equivalent to u.

Case 2: Assume that $p = p'\varphi^{j-1}(0)$ where p' is a suffix of $\varphi^{j-1}(1)$ and s is a prefix of $\varphi^{j-1}(1)$. For all v' such that $|u'| = |v'|$, applying Theorem 2.4 and Lemma 2.5,

$$u \sim_j p'\varphi^j(v')\varphi^{j-1}(0)s =: v.$$

Hence v is a factor of $\varphi^j(0v'0)$, and such a factor appears in $\mathbf{z}$ by assumption. We conclude as above.

Case 3: Assume that p is a suffix of $\varphi^{j-1}(0)$ and $s = \varphi^{j-1}(1)s'$ where s' is a prefix of $\varphi^{j-1}(0)$. For all v' such that $|u'| = |v'|$, applying Theorem 2.4 and Lemma 2.5, we have that $u \sim_j p\varphi^{j-1}(1)\varphi^j(v')s' =: v$ and the conclusion is the same as in the previous case.

Case 4: Assume that $p = p'\varphi^{j-1}(0)$ and $s = \varphi^{j-1}(1)s'$ where p' is a suffix of $\varphi^{j-1}(1)$ and s' is a prefix of $\varphi^{j-1}(0)$. For all v' such that $|v'| = |u'| + 1$, applying Theorem 2.4 and Lemma 2.5,

$$u \sim_j p'\varphi^{j-1}(0)\varphi^{j-1}(1)\varphi^j(u')s' \sim_j p'\varphi^j(w')s' =: v,$$

Hence v is a factor of $\varphi^j(0w'0)$ and the conclusion is similar to Case 1.

To conclude the proof, we consider the case of a suffix $\mathbf{w}$ of $\mathbf{x}$. Now $\mathbf{w}$ has a suffix of the form $\varphi^k(\mathbf{y}')$, where $\mathbf{y}'$ is a suffix of $\mathbf{y}$. Notice now that $\mathcal{L}(\mathbf{x}) \supseteq \mathcal{L}(\mathbf{w}) \supseteq \mathcal{L}(\varphi^k(\mathbf{y}'))$. The theorem applies to both $\mathbf{x}$ and $\varphi^k(\mathbf{y}')$ by the previous part. Hence $\mathbf{w}$ has property $\mathcal{TMB}(k)$ also. □

Remark 4.9. If $\mathbf{y}$ is an aperiodic infinite word, then for all $a, b \in \{0, 1\}$ and $n \geq 2$ we have $\mathcal{L}_n(\varphi(\mathbf{y})) \cap aA^*b \neq \emptyset$. Indeed, for $a \neq b$ the claim follows from Lemma 4.6. For $a = b$, we observe the following: for even length factors $n = 2\ell$, $\ell \geq 1$, a factor $\overline{a}ya$ of $\mathbf{y}$ of length $\ell+1$ (which exists by Lemma 4.6) gives a factor $\overline{a}a\varphi(y)a\overline{a}$ in $\mathbf{z}$, hence we have the factor aza with $|z| = 2\ell - 2$.

For odd length factors $n = 2\ell + 1$, $\ell \geq 1$, we have that a factor of the form cyc, $|y| = \ell - 1$, of $\mathbf{y}$ (such a factor exists for some $c \in \{0, 1\}$ by Lemma 4.6) gives $c\bar{c}\varphi(y)c\bar{c}$. Consequently $\mathbf{z}$ contains a factor in aA^*a of length n as well.

Applying this observation in the above proof to $\mathbf{z}$ when $j < k$, we have $\mathcal{L}_n(\mathbf{z}) \cap 1A^*1 \neq \emptyset$ for all $n \geq 2$, and thus case analysis at the end of the proof is only necessary for the case $j = k$.

4.3 On the k- and $(k+1)$-Binomial Equivalence

The previous subsection was dealing with the j-binomial equivalence in $\mathbf{x} = \varphi^k(\mathbf{y})$, where $\mathbf{y}$ is an aperiodic binary word and $j \leq k$. Here, we are concerned with the $(k+1)$-binomial equivalence in such words. To this end, we need to have more control on the k-binomial equivalence in $\mathbf{x}$. First, we have a closer look at the φ^j-factorizations of a word and in particular at the associated prefixes and suffixes.

Definition 4.10 ([29, Def. 43])**.** Let $j \geq 1$. Let us define the equivalence relation $\equiv_j$ on $A^{<2^j} \times A^{<2^j}$ by $(p_1, s_1) \equiv_j (p_2, s_2)$ whenever there exists $a \in A$ such that one of the following situations occurs:

1. $|p_1| + |s_1| = |p_2| + |s_2|$ and
   (a) $(p_1, s_1) = (p_2, s_2)$;
   (b) $(p_1, \varphi^{j-1}(a)s_1) = (p_2\varphi^{j-1}(a), s_2)$;
   (c) $(p_2, \varphi^{j-1}(a)s_2) = (p_1\varphi^{j-1}(a), s_1)$;
   (d) $(p_1, s_1) = (s_2, p_2) = (\varphi^{j-1}(a), \varphi^{j-1}(\overline{a}))$;
2. $\big||p_1| + |s_1| - (|p_2| + |s_2|)\big| = 2^j$ and
   (a) $(p_1, s_1) = (p_2\varphi^{j-1}(a), \varphi^{j-1}(\bar{a})s_2)$;
   (b) $(p_2, s_2) = (p_1\varphi^{j-1}(a), \varphi^{j-1}(\bar{a})s_1)$.

The next lemma is essentially [29, Lem. 40 and 41] (except that with an arbitrary word $\mathbf{y}$ instead of the Thue–Morse word $\mathbf{t}$, we cannot use the fact that $\mathbf{t}$ is overlap-free, so factors such as 10101 may appear in $\mathbf{y}$). To each φ^j-factorization there is a natural corresponding φ^{j-1}-factorization, though two φ^j-factorizations may correspond to the same φ^{j-1}-factorization. The next lemma also describes how such factorizations are related.

Lemma 4.11. *Let $j \geq 1$. Let u be a factor of $\varphi^j(\mathbf{y})$ such that $|u| \geq 2^j - 1$. Then u has at most two φ^j-factorizations. Let further u have a φ^j-factorization of the form $(p, \varphi^j(z), s)$ and z_0zz_{n+1} being the corresponding φ^j-ancestor (where according to Definition 4.7 z_0, z_{n+1} or z could be empty). The factor u has a unique φ^j-factorization if and only if the word z_0zz_{n+1} contains both letters 0 and 1. Moreover, if there is another φ^j-factorization $(p', \varphi^j(z'), s')$ with φ^j-ancestor $z'_0z'z'_{m+1}$, then $(p, s) \equiv_j (p', s')$ with $\big||p| - |p'|\big| = \big||s| - |s'|\big| = 2^{j-1}$, $z_0zz_{n+1} = a^{n+2}$, and $z'_0z'z'_{m+1} = \overline{a}^{m+2}$ for some $a \in \{0, 1\}$.*

Otherwise stated, the φ^j-factorization is not unique if and only if u is a factor of $\varphi^{j-1}(x)$ with $x \in (01)^* \cup (10)^* \cup 1(01)^* \cup 0(10)^*$.

Proof. Since $|u| \geq 2^j - 1$, u has a factor of the form $\varphi^{j-1}(a)$ and thus at least one φ^j-factorization of the prescribed form exists with $z = z_1 \cdots z_n$ and $n \geq 0$ ($n = 0$ if $z = \varepsilon$).

We first prove the claim for uniqueness by induction on j. For $j = 1$, assume that $u = z_0\varphi(z_1) \cdots \varphi(z_n)z_{n+1}$ with $z_0, z_{n+1} \in \{0, 1, \varepsilon\}$. Suppose, as in the statement, that both letters 0 and 1 occur in $z_0 \cdots z_{n+1}$. Then we have $z_iz_{i+1} = 01$ (or similarly 10) for some i. This means that u contains the factor 11 forcing uniqueness of this kind of a factorization: $11 \notin \{\varphi(0), \varphi(1)\}$. Assume that the property holds true up to $j-1$ and prove it for $j \geq 2$. Let $u = p\varphi^j(z_1) \cdots \varphi^j(z_n)s$ be a φ^j-factorization and assume that $z_iz_{i+1} = 01$ for some i. To this factorization, we have a corresponding factorization of the form

$$u = p\varphi^{j-1}(z_1)\varphi^{j-1}(\overline{z_1}) \cdots \varphi^{j-1}(z_n)\varphi^{j-1}(\overline{z_n})s.$$

Notice that p is a suffix of $\varphi^{j-1}(\overline{z_0})$ if $|p| < 2^{j-1}$ and otherwise, $p = p'\varphi^{j-1}(\overline{z_0})$ with p' a suffix of $\varphi^{j-1}(z_0)$. Similarly, s is a prefix of $\varphi^{j-1}(z_{n+1})$ if $|s| < 2^{j-1}$ and otherwise, $s = \varphi^{j-1}(z_{n+1})s'$ with s' a prefix of $\varphi^{j-1}(\overline{z_{n+1}})$. Observe that $z_i\overline{z_i}z_{i+1}\overline{z_{i+1}} = 0110$. So by the induction hypothesis, the φ^{j-1}-factorization of u is unique. There are at most two φ^j-factorizations corresponding to a φ^{j-1}-factorization. But since $\varphi^{j-1}(1)\varphi^{j-1}(1) \notin \{\varphi^j(0), \varphi^j(1)\}$, the claimed uniqueness follows.

We then prove the claim for non-unique factorizations. Assume that $z_0 = z_1 = \cdots = z_{n+1} = 0$. Then

$$u = p\varphi^j(0)\cdots\varphi^j(0)s = p\varphi^{j-1}(0)\varphi^{j-1}(1)\cdots\varphi^{j-1}(0)\varphi^{j-1}(1)s$$

with p (resp., s) a suffix (resp., prefix) of $\varphi^j(0)$. If $|p| \geq 2^{j-1}$, then $p = p'\varphi^{j-1}(1)$ with p' a suffix of $\varphi^{j-1}(0)$ (and thus, a suffix of $\varphi^j(1)$), otherwise set $p' = p\varphi^{j-1}(0)$. Similarly, if $|s| \geq 2^{j-1}$, then $s = \varphi^{j-1}(0)s'$ with s' a prefix of $\varphi^{j-1}(1)$, otherwise $s' = \varphi^{j-1}(1)s$. Notice that the corresponding φ^{j-1}-factorization of u is unique since the φ^{j-1}-ancestor is not a power of a letter: if $n \neq 0$ then the claim is clear. Otherwise $|u| = |ps| \geq 2^j - 1$; this implies that either $|p| \geq 2^{j-1}$ or $|s| \geq 2^{j-1}$. Assuming the latter (the other case being symmetric), we have that $s = \varphi^{j-1}(0)s'$ with s' a prefix of $\varphi^{j-1}(1)$. If $s' \neq \varepsilon$, then the φ^{j-1}-ancestor contains both letters. If $s' = \varepsilon$, then $p \neq \varepsilon$, and then again the φ^{j-1}-factorization contains both letters.

Now u can also be written as

$$p'\varphi^{j-1}(1)\varphi^{j-1}(0)\cdots\varphi^{j-1}(1)\varphi^{j-1}(0)s' = p'\varphi^j(1)\cdots\varphi^j(1)s'.$$

There are no other φ^j-factorizations due to the uniqueness of the φ^{j-1} factorization of u. To conclude the claim in this case, a straightforward case analysis shows that $(p, s) \equiv_j (p', s')$ with $\big||p| - |p'|\big| = \big||s| - |s'|\big| = 2^{j-1}$:
If $|p| \geq 2^{j-1}$ and if $|s| \geq 2^{j-1}$, then $(p, s) = (p'\varphi^{j-1}(1), \varphi^{j-1}(0)s')$.
If $|p| \geq 2^{j-1}$ and if $|s| < 2^{j-1}$, then $(p, \varphi^{j-1}(1)s) = (p'\varphi^{j-1}(1), s')$.
If $|p| < 2^{j-1}$ and if $|s| \geq 2^{j-1}$, then $(p\varphi^{j-1}(0), s) = (p', \varphi^{j-1}(0)s')$.
If $|p| < 2^{j-1}$ and if $|s| < 2^{j-1}$, then $(p\varphi^{j-1}(0), \varphi^{j-1}(1)s) = (p', s')$. □

We have the following theorem, the proof of which is essentially the proof of [29, Thm. 48]. Indeed, the lemmas in [29] leading to its proof do not require that the factors u and v are from the Thue–Morse word, only that they have φ^j-factorizations. We note that [29, Thm. 48] is stated for $j \geq 3$. However, the statement holds also for $j = 1$ (trivially) and for $j = 2$ as it is essentially a restatement of [29, Thm. 34] obtained by closely inspecting its proof.

Theorem 4.12. *Let* $\mathbf{y}$ *be an aperiodic binary word. Let* $k \geq j \geq 1$. *Let* u *and* v *be equal-length factors of* $\mathbf{x} = \varphi^k(\mathbf{y})$ *with* φ^j*-factorizations* $u = p_1\varphi^j(z)s_1$ *and* $v = p_2\varphi^j(z')s_2$. *Then* $u \sim_j v$ *if and only if* $(p_1, s_1) \equiv_j (p_2, s_2)$.

We then turn to the $(k+1)$-binomial equivalence in $\mathbf{x}$. A straightforward consequence of (4) together with the identities $\sum_{x\in A^\ell}\binom{u}{x} = \binom{|u|}{\ell}$, $\ell \geq 1$, is the following observation.

Lemma 4.13. *Let* $u \in \{0,1\}^*$. *Then*

$$\binom{\varphi(u)}{0} = |u|; \quad \binom{\varphi(u)}{01} = |u|_0 + \binom{|u|}{2}; \quad \binom{\varphi(u)}{011} = \binom{u}{01} + \binom{|u|_0}{2} + \binom{|u|}{3}.$$

Proof. Observe that $\binom{\varphi(a)}{011} = 0 = \binom{\varphi(a)}{11}$ for both $a \in \{0,1\}$. Similarly $\binom{\varphi(a)}{b} = 1$ for letters $a, b \in \{0,1\}$. Therefore

$$\begin{aligned}\binom{\varphi(u)}{011} &= \sum_{x_1,x_2\in A}\binom{u}{x_1x_2}\sum_{\substack{011=e_1e_2\\ e_i\in A^+}}\binom{\varphi(x_1)}{e_1}\binom{\varphi(x_2)}{e_2} + \sum_{|x|=3}\binom{u}{x}\\ &= \binom{u}{00} + \binom{u}{01} + \binom{|u|}{3}.\end{aligned}$$

and the claim follows. □

The next technical lemma has an important role in studying the $(k+1)$-binomial equivalence.

Lemma 4.14. *Let u, v be two binary words of equal length. For $k \geq 1$, we have*

$$\binom{\varphi^k(u)}{01^k} - \binom{\varphi^k(v)}{01^k} = 2^{(k-1)(k-2)/2}(|u|_0 - |v|_0).$$

In particular, $u \not\sim_1 v$ implies $\varphi^k(u) \not\sim_{k+1} \varphi^k(v)$. Moreover, if $u \sim_1 v$, for $k \geq 1$, we have

$$\binom{\varphi^k(u)}{01^{k+1}} - \binom{\varphi^k(v)}{01^{k+1}} = 2^{(k-1)(k-2)/2}\left(\binom{u}{01} - \binom{v}{01}\right).$$

In particular, $u \not\sim_2 v$ implies $\varphi^k(u) \not\sim_{k+2} \varphi^k(v)$.

Proof. The case $k = 1$ is deduced from Lemma 4.13. Then assume $k \geq 2$. We encourage the reader to refer to [29] for details that would be too long to reproduce here. From [29, Rem. 23], we have the following expression

$$\binom{\varphi^k(u)}{01^k} - \binom{\varphi^k(v)}{01^k} = \sum_{x \in f^k(01^k)} m_{f^k(01^k)}(x)\left[\binom{u}{x} - \binom{v}{x}\right],$$

where the map f is defined to take into account the multiple ways factors 01 or 10 may occur in a word: $f(u)$ is a multiset of words of length shorter than u; see [29, Def. 15 and 17]. We let the coefficient $m_{f^k(01^k)}(x)$ denote the multiplicity of x as an element of the multiset $f^k(01^k)$. It can be shown that the multiset $f^k(01^k)$ only contains the elements 0 and 1. Therefore we obtain

$$\binom{\varphi^k(u)}{01^k} - \binom{\varphi^k(v)}{01^k} = m_{f^k(01^k)}(0)\ (|u|_0 - |v|_0) + m_{f^k(01^k)}(1)\ (|u|_1 - |v|_1).$$

To conclude the proof, we use two facts. The first is that $|u|_1 - |v|_1 = -(|u|_0 - |v|_0)$ since u, v have equal length. The second is that

$$m_{f^k(01^k)}(0) - m_{f^k(01^k)}(1) = m_{f^{k-1}(01^k)}(01) - m_{f^{k-1}(01^k)}(10) = 2^{(k-1)(k-2)/2},$$

which follows from [29, Prop. 28]. For the second part, the same reasoning may be applied to obtain

$$\binom{\varphi^k(u)}{01^{k+1}} - \binom{\varphi^k(v)}{01^{k+1}} = \sum_{x \in f^k(01^{k+1})} m_{f^k(01^{k+1})}(x)\left[\binom{u}{x} - \binom{v}{x}\right].$$

The multiset $f^k(01^{k+1})$ only contains $0, 1, 00, 01, 10, 11$. But since it is assumed that $u \sim_1 v$, the only (potentially) non-zero terms in the sum correspond to $x \in \{01, 10\}$. Then the observation $\binom{u}{01} - \binom{v}{01} = \binom{v}{10} - \binom{u}{10}$ following from Lemma 2.7 suffices to conclude. □

Next we consider the structure of factors of the image of an arbitrary binary word $\mathbf{y}$.

Definition 4.15. For $n \geq 1$ we let $\mathcal{S}(n) = \mathcal{L}_n(\mathbf{y})$. Further, for all $a, b \in \{\varepsilon, 0, 1\}$ such that $ab \neq \varepsilon$, we define $\mathcal{S}_{a,b}(n) = \mathcal{L}_{n+|ab|}(\mathbf{y}) \cap a\mathcal{A}^*b$. We call these sets *factorization classes of order* n.

Consider now a factor u of $\varphi(\mathbf{y})$. We associate with u some factorization classes as follows. Let $a\varphi(u')b$ be the φ-factorization of u with φ-ancestor $au'b \in \mathcal{L}(\mathbf{y})$. If $ab = \varepsilon$, we associate the factorization class $\mathcal{S}(|u'|)$. For $ab \neq \varepsilon$, we have that u is a factor of $\varphi(\overline{a}u'b)$. In this case we associate the factorization class $\mathcal{S}_{\overline{a},b}(|u'|)$. If u is associated with a factorization class $\mathcal{T}$, we write $u \models \mathcal{T}$, otherwise we write $u \not\models \mathcal{T}$.

Observe that $u \models \mathcal{S}(n)$ implies that $|u| = 2n$. Also, for $ab \neq \varepsilon$, $u \models \mathcal{S}_{a,b}(n)$ implies that $|u| = 2n + |ab|$. Notice also that a factor u of $\varphi(\mathbf{y})$ can be associated with several factorization classes: take, e.g., $(10)^\ell 1 = 1(01)^\ell$ which is associated with both $\mathcal{S}_{\varepsilon,1}(\ell)$ and $\mathcal{S}_{0,\varepsilon}(\ell)$, or $(01)^{\ell+1} = 0(10)^\ell 1$ which is associated with both $\mathcal{S}(\ell+1)$ and $\mathcal{S}_{1,1}(\ell)$.

Lemma 4.16. *For two 2-binomially equivalent factors $u, v \in \mathcal{L}(\varphi(\mathbf{y}))$, if $u \models \mathcal{T}$ for some factorization class $\mathcal{T}$, then $v \models \mathcal{T}$. Furthermore, a factor u of $\mathbf{y}$ is associated with distinct factorization classes if and only if $u \in L = (01)^* \cup (10)^* \cup 1(01)^* \cup 0(10)^*$.*

Proof. **Even-length factors.** Let $u \sim_2 v$ with $|u| = 2n$. If $u \models \mathcal{S}_{\overline{a},a}(n-1)$ with $a \in \{0,1\}$, then u is of the form $a\varphi(x)a$ with $|x| = n-1$, whence $|u|_a = n+1$. Factors $v' \not\models \mathcal{S}_{\overline{a},a}(n-1)$ of length $2n$ have $|v'|_a \leq n$ by inspection. Hence also $v \models \mathcal{S}_{\overline{a},a}(n-1)$. The above arguments also show that u is associated with exactly one factorization class. For the latter claim, we note that u has even length and begins and ends with the same letter, so it cannot appear in the language L.

Assume then that $u \not\models \mathcal{S}_{\overline{a},a}(n-1)$, $a \in \{0,1\}$. Then $v \not\models \mathcal{S}_{\overline{a},a}(n-1)$, $a \in \{0,1\}$ by the previous observation. Notice that we may assume $n \geq 2$ as otherwise we have $|u| = 2$ and the claim is trivial (2-binomial equivalence is equality in this case). We compare the values of $\binom{y}{01}$ for y associated with $\mathcal{S}_{1,1}(n-1)$, $\mathcal{S}_{0,0}(n-1)$, and $\mathcal{S}(n)$, respectively.

Case 1: $y \models \mathcal{S}_{1,1}(n-1)$. We have $\binom{y}{01} \geq \binom{n}{2} + n$, and equality holds for $y = (01)^n$. Indeed, say $y = 0\varphi(x)1$ for some $x \in \{0,1\}^{n-1}$. Then we have by Lemma 4.13

$$\binom{y}{01} = \binom{\varphi(x)}{01} + |\varphi(x)|_0 + |\varphi(x)1|_1 = |x|_0 + \binom{|x|}{2} + 2|x| + 1 = |x|_0 + \binom{n}{2} + n,$$

since $|x| = n-1$. Equality now holds when $|x|_0 = 0$, i.e., $x = 1^{n-1}$.

Case 2: $y \models \mathcal{S}_{0,0}(n-1)$. We have $\binom{y}{01} \leq \binom{n}{2}$, and equality holds when $y = (10)^n$. Indeed, say $y = 1\varphi(x)0$ for some $x \in \{0,1\}^{n-1}$. Then

$$\binom{y}{01} = \binom{\varphi(x)}{01} = |x|_0 + \binom{|x|}{2} = |x|_0 + \binom{n}{2} - (n-1).$$

Since $|x| = n-1$, we have $\binom{y}{01} \leq \binom{n}{2}$. Equality holds when $x = 0^{n-1}$.

Case 3: $y \models \mathcal{S}(n)$. We have $\binom{n}{2} \leq \binom{y}{01} \leq \binom{n}{2} + n$. The former equality is attained with $y = (10)^n$ and the latter with $y = (01)^n$. Indeed, say $y = \varphi(x')$ for some $x' \in \{0,1\}^n$. We have $\binom{y}{01} = \binom{n}{2} + |x'|_0$ from Lemma 4.13. Therefore, $\binom{n}{2} \leq \binom{y}{01} \leq \binom{n}{2} + n$. The former equality is attained with $x' = 1^n$ and the latter with $x' = 0^n$.

We conclude that u and v are associated with a common factorization class. In fact, the latter claim is also implied from the above: a word can be associated with two (and only two) factorization classes if and only if it appears in L. This concludes the proof in the case of even length factors.

Odd-length factors. Assume without loss of generality that $u \models \mathcal{S}_{a,\varepsilon}(n)$ with $u = a\varphi(u')$ of length $2n+1$. Recalling that $|\varphi(u')|_0 = |u'| = n$, if $u \sim_2 v$ with u and v associated with distinct factorization classes, then necessarily $v \in \mathcal{S}_{\varepsilon,a}$, say $v = \varphi(v')a$. We show that this is impossible, unless $u = v \in L$.

Indeed, assuming that we have 2-binomial equivalence, we have

$$\binom{a\varphi(u')}{01} = \binom{\varphi(u')}{01} + \delta_0(a)\binom{\varphi(u')}{1} = |u'|_0 + \binom{n}{2} + \delta_0(a)n \tag{6}$$

which is equal to

$$\binom{\varphi(v')a}{01} = \binom{\varphi(v')}{01} + \delta_1(a)\binom{\varphi(v')}{0} = |v'|_0 + \binom{n}{2} + \delta_1(a)n \tag{7}$$

where $\delta_a(b) = 1$ if $a = b$, otherwise $\delta_a(b) = 0$. Rearranging, we get $|u'|_0 - |v'|_0 = (\delta_1(a) - \delta_0(a))n \in \{\pm n\}$. This implies, without loss of generality, that $u' = 0^n$, $v' = 1^n$, and $a = 1$. But then $u = 1(01)^n = (10)^n 1 = v \in L$, as claimed. □

The next result characterizes $(k+1)$-binomial equivalence in $\mathbf{x} = \varphi^k(\mathbf{y})$ when $\mathbf{y}$ is an arbitrary binary word.

Proposition 4.17. *Let u and v be factors of length at least $2^k - 1$ of $\mathbf{x} = \varphi^k(\mathbf{y})$ with the φ^k-factorizations $u = p_1\varphi^k(z)s_1$ and $v = p_2\varphi^k(z')s_2$. Then $u \sim_{k+1} v$ and $u \neq v$ if and only if $z \sim_1 z'$, $z' \neq z$, and $(p_1, s_1) = (p_2, s_2)$.*

Notice that the proposition claims that those factors of $\mathbf{x}$ having at least two φ^k-factorizations are $(k+1)$-binomially equivalent only to themselves (in $\mathcal{L}(\mathbf{x})$).

Proof. The "if"-part of the statement follows by a repeated application of Proposition 3.9 on the Thue–Morse morphism together with the fact that the morphism is injective.

Let us assume that $u \sim_{k+1} v$ for some distinct factors. It follows that $u \sim_k v$, which implies that $(p_1, s_1) \equiv_k (p_2, s_2)$ by Theorem 4.12. Next we show that $(p_1, s_1) = (p_2, s_2)$ and $z \sim_1 z'$. We have the following case distinction from Definition 4.10:

(1)(a): We have that $(p_1, s_1) = (p_2, s_2)$. By deleting the common prefix p_1 and suffix s_1, we are left with the equivalent statement $\varphi^k(z) \sim_{k+1} \varphi^k(z')$. If $z \not\sim_1 z'$, then we have a contradiction with Lemma 4.14. The desired result follows in this case.

In the remaining cases, we assume towards a contradiction that $(p_1, s_1) \neq (p_2, s_2)$.

(1)(b): Suppose that $(p_1, s_2) = (p_2\varphi^{k-1}(a), \varphi^{k-1}(a)s_1)$. Deleting the common prefixes p_2 and suffixes s_1, we are left with $\varphi^{k-1}(a\varphi(z)) \sim_{k+1} \varphi^{k-1}(\varphi(z')a)$. Now $a\varphi(z) \sim_1 \varphi(z')a$, but $a\varphi(z) \not\sim_2 \varphi(z')a$ by Lemma 4.16 (otherwise $a\varphi(z) = \varphi(z')a$ and thus $u = v$ contrary to the assumption). Lemma 4.14 then implies that $\varphi^{k-1}(a\varphi(z)) \not\sim_{k+1} \varphi^{k-1}(\varphi(z')a)$, which is a contradiction.

(1)(c): Suppose that $(p_2, \varphi^{k-1}(a)s_2) = (p_1\varphi^{k-1}(a), s_1)$. This is symmetric to the previous case.

(1)(d): Suppose that $(p_1, s_1) = (s_2, p_2) = (\varphi^{k-1}(a), \varphi^{k-1}(\overline{a}))$. We thus have directly

$$\varphi^{k-1}(a\varphi(z)\overline{a}) \sim_{k+1} \varphi^{k-1}(\overline{a}\varphi(z)a).$$

The claim follows by an argument similar to that of in Case (1)(b).

(2)(a): Suppose that $(p_1, s_1) = (p_2\varphi^{k-1}(a), \varphi^{k-1}(\bar{a})s_2)$. After removing common prefixes and suffixes, we are left with $\varphi^{k-1}(a\varphi(z)\overline{a}) \sim_{k+1} \varphi^{k-1}(\varphi(z'))$. We have that $a\varphi(z)\overline{a} \sim_1 \varphi(z')$, but by Lemma 4.16 $a\varphi(z)\overline{a} \not\sim_2 \varphi(z')$ (otherwise $z = \overline{a}^\ell$ and $z' = a^{\ell+1}$, implying that $u = v$, a contradiction). This is again a contradiction by Lemma 4.14.

(2)(b): Suppose that $(p_2, s_2) = (p_1\varphi^{j-1}(a), \varphi^{j-1}(\bar{a})s_1)$. This is symmetric to the previous case. □

Notice that Theorem 4.2 and Proposition 4.17 have the following corollary:

Corollary 4.18. *Let $\mathbf{x} = \varphi^k(\mathbf{y})$, where $\mathbf{y}$ is an arbitrary aperiodic binary word. We have*

$$\mathsf{b}_{\mathbf{x}}^{(1)} \prec \mathsf{b}_{\mathbf{x}}^{(2)} \prec \cdots \prec \mathsf{b}_{\mathbf{x}}^{(k)} \prec \mathsf{b}_{\mathbf{x}}^{(k+1)}.$$

Proof. Recall that $\mathbf{y}$ contains arbitrarily long factors of the form $\overline{a}za$, $a \in \{0, 1\}$, by Lemma 4.6. Therefore $\mathbf{x}$ contains the k-binomially equivalent (by Lemma 2.5) factors $\varphi^{k-1}(a)\varphi^k(z)$ and $\varphi^k(z)\varphi^{k-1}(a)$. However, by Proposition 4.17 these factors are either not $(k+1)$-binomially equivalent, or $\varphi^{k-1}(a)\varphi^k(z) = \varphi^k(z)\varphi^{k-1}(a)$. The latter happens when $\varphi^k(z) = \varphi^{k-1}(a)^\ell$ for some $\ell \geq 0$, and thus only when $\ell = 0$ and $z = \varepsilon$. (Indeed, it is not hard to prove that if w is primitive so is $\varphi(w)$.) This observation suffices for showing $\mathsf{b}_{\mathbf{x}}^{(k)} \prec \mathsf{b}_{\mathbf{x}}^{(k+1)}$. The rest of the claim follows by Theorem 4.2. □

5 Binomial Properties of the Thue–Morse Morphism, Part II

In this section we consider a complementary result to Theorem 4.2, which partially extends the following theorem of Richomme, Saari, and Zamboni [40].

Theorem 5.1 ([40, Thm 3.3])**.** *Let $\mathbf{x}$ be an aperiodic word. Then $\mathsf{b}_{\mathbf{x}}^{(1)} = \mathsf{b}_{\mathbf{t}}^{(1)}$ if and only if there exists a binary word $\mathbf{y}$ and $a \in \{\varepsilon, 0, 1\}$ such that $\mathbf{x} = a\varphi(\mathbf{y})$.*

Notice that Theorem 4.2 is a generalization of the "if"-direction. The following Theorem 5.2 is a partial generalization in the other direction. It is proved in Section 5.2. Recall Definition 4.1: For an integer $k \geq 1$, a binary word $\mathbf{x}$ has property $\mathcal{TMB}(k)$ if, for all $1 \leq j \leq k$, we have $b_{\mathbf{x}}^{(j)} = b_{\mathbf{t}}^{(j)}$.

Theorem 5.2. *Let* $\mathbf{x}$ *be a recurrent binary word having property* $\mathcal{TMB}(k)$ *for some* $k \geq 1$. *Then there exists a binary word* $\mathbf{y}$ *such that* $\mathbf{x} = u\varphi^k(\mathbf{y})$, *where* u *is a proper suffix of* $\varphi^k(0)$ *or* $\varphi^k(1)$.

To prove the theorem, we first derive a formula for counting $(k+1)$-binomial equivalence classes of words that are of the form $\varphi^k(\mathbf{y})$ for $\mathbf{y}$ aperiodic in Section 5.1.

5.1 A formula for counting $(k+1)$-binomial complexities

For a binary word $\mathbf{y}$ we define

$$\begin{aligned} X_{\mathbf{y}}(n) &:= \{(a, \Psi(u), b) \colon a, b \in \{0,1\}, aub \in \mathcal{L}_{n+1}(\mathbf{y})\}, \\ Y_{\mathbf{y},L}(n) &:= \{(a, \Psi(u)) \colon a \in \{0,1\}, au \in \mathcal{L}_{n+1}(\mathbf{y})\}, \\ Y_{\mathbf{y},R}(n) &:= \{(\Psi(u), a) \colon a \in \{0,1\}, ua \in \mathcal{L}_{n+1}(\mathbf{y})\}, \\ Y_{\mathbf{y}}(n) &:= Y_{\mathbf{y},L} \cup Y_{\mathbf{y},R}. \end{aligned}$$

Observation 5.3. Let $G_n = (V, E)$ be the abelian Rauzy graph of $\mathbf{y}$ (of order n).

- E is in one-to-one correspondence with $X_{\mathbf{y}}(n)$, namely, $\vec{x} \xrightarrow{(a,b)} \vec{y}$ is identified with $(a, \vec{x} - \Psi(a), b)$. In particular, $\#E = \#X_{\mathbf{y}}(n)$.
- The set $Y_{\mathbf{y},R}(n)$ is in one-to-one correspondence with $E/\equiv_R$, where $\equiv_R$ is the equivalence relation defined by the (surjective) mapping $E \to Y_{\mathbf{y},R}(n)$ (meaning, the equivalence classes are the full preimages of elements of $Y_{\mathbf{y},R}(n)$),

  $$\left(\vec{x} \xrightarrow{(a,b)} \vec{y}\right) \mapsto (\vec{x}, b).$$

  In particular, $\#Y_{\mathbf{y},R}(n) \leq \#E$. Similarly $Y_{\mathbf{y},L}(n)$ is in one-to-one correspondence with $E/\equiv_L$, where $\equiv_L$ is the equivalence relation defined by the (surjective) mapping

  $$\left(\vec{x} \xrightarrow{(a,b)} \vec{y}\right) \mapsto (a, \vec{y}).$$

  In particular, $\#Y_{\mathbf{y},L}(n) \leq \#E$.
- Each equivalence class in $E/\equiv_L$ contains at most two elements: two edges are equivalent if their target vertices and the first components of the labels are equal. Hence the equivalence relation can only identify a non-loop edge with a loop.
- Note that any loop at a vertex v with label $(0,0)$ can only be equivalent under either $\equiv_L$ or $\equiv_R$ to an edge between v and a lighter vertex. Similarly a loop with label $(1,1)$ can only be equivalent to an edge between v and a heavier vertex. In particular, a loop with $(0,0)$ (resp., $(1,1)$) on the lightest (resp., heaviest) vertex is not equivalent to any other edge under either $\equiv_L$ or $\equiv_R$.

Example 5.4. Recall the abelian Rauzy graphs of the Thue–Morse word from Example 4.5. The edges correspond exactly to $X_{\mathbf{t}}(n)$. The equivalence classes of $E/\equiv_R$ (resp., $E/\equiv_L$) corresponding to $Y_{\mathbf{t},R}(n)$ (resp., $Y_{\mathbf{t},L}(n)$) containing at least two elements are listed below:

$$n = 2m\colon\ Y_{\mathbf{t},R} : \left\{m \xrightarrow{(0,0)} m, m \xrightarrow{(1,0)} m-1\right\},\quad \left\{m \xrightarrow{(1,1)} m, m \xrightarrow{(0,1)} m+1\right\};$$

$$Y_{\mathbf{t},L} : \left\{m \xrightarrow{(0,0)} m, m-1 \xrightarrow{(0,1)} m\right\},\quad \left\{m \xrightarrow{(1,1)} m, m+1 \xrightarrow{(1,0)} m\right\}.$$

$n = 2m-1$: $Y_{t,R} : \left\{ m \xrightarrow{(1,1)} m, m \xrightarrow{(0,1)} m+1 \right\}$, $\left\{ m+1 \xrightarrow{(0,0)} m+1, m+1 \xrightarrow{(1,0)} m \right\}$;

$Y_{t,L} : \left\{ m \xrightarrow{(1,1)} m, m+1 \xrightarrow{(1,0)} m \right\}$, $\left\{ m+1 \xrightarrow{(0,0)} m+1, m \xrightarrow{(0,1)} m+1 \right\}$.

We may now establish a formula for counting the $(k+1)$-binomial complexity of the kth image of a word $\mathbf{y}$ under the Thue–Morse morphism. This will turn out to be key in proving a converse to Theorem 4.2.

Proposition 5.5. *Let* $\mathbf{y}$ *be an infinite binary word and let* $\mathbf{x} = \varphi^k(\mathbf{y})$ *with* $k \geq 1$. *Let* $m = \max\{n \in \mathbb{N} \colon 0^n \text{ and } 1^n \in \mathcal{L}(\mathbf{y})\}$ *and* $m' = \max\{n \in \mathbb{N} \colon 0^n \text{ or } 1^n \in \mathcal{L}(\mathbf{y})\}$, *where we allow* m *and* m' *to equal* ∞. *We have* $\mathsf{b}_{\mathbf{x}}^{(k+1)}(r) = \mathsf{p}_{\mathbf{t}}(r)$ *for all* $0 \leq r < 2^k$. *Setting* $Z(n,0) := (2^k-1)\#X_{\mathbf{y}}(n) + \mathsf{b}_{\mathbf{y}}^{(1)}(n)$, *for all* $n \geq 1$ *we have*

$$\mathsf{b}_{\mathbf{x}}^{(k+1)}(2^k n) = Z(n,0) - \begin{cases} 2^k, & \text{if } n < m; \\ 1, & \text{if } n = m < m'; \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

For all $n \geq 1$ *and* $0 < r < 2^k$, *setting* $Z(n,r) := (r-1)\#X_{\mathbf{y}}(n+1) + (2^k - r - 1)\#X_{\mathbf{y}}(n) + \#Y_{\mathbf{y}}(n)$, *we have*

$$\mathsf{b}_{\mathbf{x}}^{(k+1)}(2^k n + r) = Z(n,r) - \begin{cases} 2^k, & \text{if } n+1 < m; \\ (2^k - r + 1), & \text{if } n+1 = m < m'; \\ (2^k - 2(r-1)), & \text{if } n+1 = m = m' \text{ and } r \leq 2^{k-1}; \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Proof. Lemma 4.8 implies the formula for lengths less than 2^k. The proof strategy to establish formulas (8) and (9) is as follows. For $n \geq 1$ and $0 \leq r < 2^k$, we first obtain an upper bound $Z(n,r)$ on $\mathsf{b}_{\mathbf{x}}^{(k+1)}(2^k n + r)$ by counting the different φ^k-factorizations up to the equivalence implied by Proposition 4.17. Then we establish the exact formula by subtracting the number of $(k+1)$-binomial classes that admit several φ^k-factorizations as counted above. To do so, we use the following argument. By Proposition 4.17 and the observation made right after its statement, those factors of $\mathbf{x}$ that admit several φ^k-factorizations are $(k+1)$-binomially equivalent only to themselves. In fact, such factors are well-understood by Lemma 4.11; they only admit two distinct φ^k-factorizations. Hence, counting the number of factors that have two φ^k-factorizations and subtracting that number from the term $Z(n,r)$ gives the number of $(k+1)$-binomial equivalence classes.

We first prove formula (8) by inspecting factors of length $2^k n$ for some $n \geq 1$. They are of the following two forms: either $\varphi^k(u)$, with $|u| = n$, or $p\varphi^k(v)s$, with $|v| = n-1$, p and s non-empty. Each abelian equivalence class in $\mathcal{L}_n(\mathbf{y})/\sim_1$ gives a $(k+1)$-equivalence class of factors of the first form by Proposition 3.9 (recall that φ is Parikh-collinear). Hence the term $\mathsf{b}_{\mathbf{y}}^{(1)}(n)$ in $Z(n,0)$. For the factors of the second form, we notice the following. Such a factor has the φ^k-ancestor avb, with $(a, \Psi(v), b) \in X_{\mathbf{y}}(n)$. On the other hand, any $(a, \Psi(v), b) \in X_{\mathbf{y}}(n)$ gives rise to $(2^k - 1)$ $(k+1)$-binomial equivalence classes, namely, those represented by the words

$$\operatorname{suff}_i(\varphi^k(a))\ \varphi^k(v)\ \operatorname{pref}_{2^k - i}(\varphi^k(b)),\ 1 \leq i < 2^k,$$

where, for a word w and $i \in \{1, \ldots, |w|\}$, we let $\operatorname{pref}_i(w)$ (resp., $\operatorname{suff}_i(w)$) denote the length-i prefix (resp., suffix) of w. Hence the term $(2^k - 1)\#X_{\mathbf{y}}(n)$ in the formula. Therefore we have established the upper bound $\mathsf{b}_{\mathbf{x}}^{(k+1)}(2^k n) \leq Z(n,0)$, with $Z(n,0) = (2^k-1)\#X_{\mathbf{y}}(n) + \mathsf{b}_{\mathbf{y}}^{(1)}(n)$.

As explained at the beginning of the proof, we now examine factors admitting several φ^k-factorizations and subtract their number from $Z(n,0)$ to establish formula (8). Let x be such a factor. Then it has, by Lemma 4.11, exactly two φ^k-factorizations, and we may write

$$p\varphi^k(u)s = x = p'\varphi^k(u')s'. \tag{10}$$

Here we note that $|ps| = 2^k$ if and only if p or s non-empty. Moreover, the corresponding φ^k-ancestors are powers of letters, and given in Table 2.

fact.	ancest.	conditions on p, s	2nd fact.	ancest.
$p\varphi^k(a^n)s$	a^n	$\|p\| = 0 = \|s\|$	$\varphi^{k-1}(a)\varphi^k(\overline{a}^{n-1})\varphi^{k-1}(\overline{a})$	$\overline{a}^{n+1}$
$p\varphi^k(a^{n-1})s$	a^{n+1}	$p = \varphi^{k-1}(\overline{a}), s = \varphi^{k-1}(a)$	$\varphi^k(\overline{a}^n)$	$\overline{a}^n$
$p\varphi^k(a^{n-1})s$	a^{n+1}	$\|p\| > 2^{k-1}, 0 < \|s\| < 2^{k-1}$	$p'\varphi^k(\overline{a}^{n-1})\varphi^{k-1}(\overline{a})s$	$\overline{a}^{n+1}$
$p\varphi^k(a^{n-1})s$	a^{n+1}	$0 < \|p\| < 2^{k-1}, \|s\| > 2^{k-1}$	$p\varphi^{k-1}(a)\varphi^k(\overline{a}^{n-1})s'$	$\overline{a}^{n+1}$

Table 2: Factors of length $2^k n$ of $\mathbf{x}$ admitting two φ^k-factorizations and their potential φ^k-ancestors. In the third row p' is defined by $p = p'\varphi^{k-1}(\overline{a})$, and in the fourth row, s' is defined by $s = \varphi^{k-1}(a)s'$.

In particular, for x to have two φ^k-factorizations (and so for a class to have been counted twice), a^n and $\overline{a}^n$ both must appear in $\mathbf{y}$, and at least one of a^{n+1} and $\overline{a}^{n+1}$ has to also appear in the word. We divide the proof into three cases.

Case 1. If $n > m$ or when $n = m = m'$, then as concluded above, there is no factor having several φ^k-factorizations, and the formula holds.

Case 2. Assume now that $n < m$, so both a^{n+1} and $\overline{a}^{n+1}$ appear in $\mathbf{y}$. Reusing Table 2 above, we see that any equivalence class corresponding to a factor having φ^k-ancestor a^{n+1} (or a^n) has been counted twice (and corresponds to a word with φ^k-ancestor $\overline{a}^{n+1}$ or $\overline{a}^n$). There are 2^k of those, whence the formula for $n < m$.

Case 3. Assume finally that $n = m$ and $m' > m$. Assume without loss of generality that a^{m+1} appears in $\mathbf{y}$. Therefore $\overline{a}^{m+1}$ does not appear in $\mathbf{y}$. Thus, if x has two φ^k-ancestors, one of them is $\overline{a}^m$. There is only one such factor, and this proves the remaining case in the formula.

We now turn to the proof of formula (9), and consider factors of the length $2^k n + r$, with $n \geq 1$ and $0 < r < 2^k$. Let us first establish the upper bound $Z(n, r)$ on $b_x^{(k+1)}(2^k n + r)$. Each element of $Y_y(n)$ gives rise to a unique $(k+1)$-factorization; for example, $(a, \Psi(u))$ gives the class represented by $\mathrm{suff}_r(\varphi^k(a))\varphi^k(u)$. Hence the term $\#Y_y(n)$ in $Z(n, r)$. Each element of $X_y(n+1)$ gives rise to $r-1$ many $(k+1)$-factorizations as follows: $(a, \Psi(u), b)$ gives

$$\mathrm{suff}_i(\varphi^k(a))\ \varphi^k(u)\ \mathrm{pref}_{r-i}(\varphi^k(b)), \quad 1 \leq i < r.$$

Similarly each element of $X_y(n)$ gives $2^k - r - 1$ elements, namely $(a, \Psi(u'), b)$ gives

$$\mathrm{suff}_i(\varphi^k(a))\ \varphi^k(u')\ \mathrm{pref}_{2^k+r-i}(\varphi^k(b)) \quad r < i < 2^k.$$

Hence $b_x^{(k+1)}(2^k n + r) \leq Z(n, r)$ with $Z(n, r) = (r-1)\#X_y(n+1) + (2^k - r - 1)\#X_y(n) + \#Y_y(n)$.

We again face the problem of over-counting. As previously, we count the number of factors that have two φ^k-factorizations. Let again x have two φ^k-factorizations as in (10), where now ps and $p's'$ are both non-empty. The φ^k-ancestor of each factorization is given in Table 3. We conclude that, for x to have two φ^k-factorizations (and so for a class to have been counted twice), we must have $n + 1 \leq m$. We divide the proof into three cases.

Case 1. Assume that $n + 1 > m$. As concluded above, there is no factor having several φ^k-factorizations, and the formula holds.

Case 2. Assume that $n+1 < m$. In this case we have 1^{n+2} and 0^{n+2} appearing in $\mathbf{y}$. We claim that any factor with a φ^k-ancestor a^{n+2} or a^{n+1} has also a φ^k-ancestor $\overline{a}^{n+2}$ or $\overline{a}^{n+1}$, and we show there are 2^k such factors. Hence the formula follows.

fact.	ancestor	conds. on p, s
$p\varphi^k(u)s$	a^{n+2}	if $\lvert ps\rvert = r$ and $p, s \neq \varepsilon$
	a^{n+1}	if $\lvert ps\rvert = r$ with $p = \varepsilon$ or $s = \varepsilon$, or if $\lvert ps\rvert > r$
$p'\varphi^k(u')s'$	$\overline{a}^{n+2}$	if $\lvert p's'\rvert = r$ and $p', s' \neq \varepsilon$
	$\overline{a}^{n+1}$	if $\lvert p's'\rvert = r$ with $p' = \varepsilon$ or $s' = \varepsilon$, or if $\lvert p's'\rvert > r$

Table 3: Factors of $\mathbf{x}$ of length $2^k n + r$, $0 < r < 2^k$, admitting two φ^k-factorizations together with their potential φ^k-ancestors.

conditions on p, s	φ^k-fact.	φ^k-ancestor
$\lvert p\rvert, \lvert s\rvert < 2^{k-1}$	$p'\varphi^k(\overline{a}^{n-1})s'$	$\overline{a}^{n+1}$
$\lvert p\rvert \geq 2^{k-1}$ and $\lvert s\rvert < 2^{k-1}$	$p'\varphi^k(\overline{a}^n)\varphi^{k-1}(\overline{a})s$	$\overline{a}^{n+2}$
$\lvert p\rvert < 2^{k-1}$ and $\lvert s\rvert \geq 2^{k-1}$	$p\varphi^{k-1}(a)\varphi^k(\overline{a}^n)s'$	$\overline{a}^{n+2}$

Table 4: The accompanying second φ^k-factorization of a factor $p\varphi^k(a^n)s$, $\lvert ps\rvert = r$, having two φ^k-factorization. Here p' and s' are suitably chosen; for example, in the first row we have $p' = p\varphi^{k-1}(a)$ and $s' = \varphi^{k-1}(\overline{a})s$.

First, if $x = p\varphi^k(a^n)s$, with $\lvert ps\rvert = r$, we have $x = p\varphi^{k-1}(a)\varphi^k(\overline{a}^{n-1})\varphi^{k-1}(\overline{a})s$. The other φ^k-factorization of x is given in Table 4. (Observe that $\lvert p\rvert \geq 2^{k-1}$ and $\lvert s\rvert \geq 2^{k-1}$ cannot simultaneously hold as $\lvert ps\rvert = r < 2^k$.)

Second, if $x = p\varphi^k(a^{n-1})s$ with $\lvert ps\rvert = 2^k + r$, $r < \lvert p\rvert < 2^k$, the other φ^k-factorization of x is given in Table 5, where p' and s' are again suitably chosen. This concludes the proof for this part,

conditions on p, s	φ^k-fact.	φ^k-ancestor
$\lvert p\rvert \geq 2^{k-1}, \lvert s\rvert < 2^{k-1}$	$p'\varphi^k(\overline{a}^{n-1})\varphi^{k-1}(\overline{a})s$	$\overline{a}^{n+1}$
$\lvert p\rvert = 2^{k-1}, \lvert s\rvert = 2^{k-1} + r$	$\varphi^k(\overline{a}^n)s'$	$\overline{a}^{n+1}$
$\lvert p\rvert, \lvert s\rvert > 2^{k-1}$	$p'\varphi^k(\overline{a}^n)s'$	$\overline{a}^{n+2}$
$\lvert p\rvert = 2^{k-1} + r, \lvert s\rvert = 2^{k-1}$	$p'\varphi^k(\overline{a}^n)$	$\overline{a}^{n+1}$
$\lvert p\rvert < 2^{k-1}, \lvert s\rvert \geq 2^{k-1}$	$p\varphi^{k-1}(a)\varphi^k(\overline{a}^{n-1})s'$	$\overline{a}^{n+1}$

Table 5: The accompanying second φ^k-factorization of a factor $p\varphi^k(a^{n-1})s$, $\lvert ps\rvert = 2^k + r$, $r < \lvert p\rvert < 2^k$, having two φ^k-factorizations. Here p' and s' are again suitably chosen and can be inferred from the other φ^k-ancestor and the length constraints.

as we have exhibited 2^k distinct factors, and there are no other possibilities. (Indeed, there are $r - 1$ factors having a^{n+2} as a φ^k-ancestor, and $2^k - r + 1$ factors having a^{n+1} as such.)

Case 3. Assume finally that $n + 1 = m$. We divide the proof into two subcases.

Case 3.1. Assume that $m < m'$. Let us assume that a^{n+1} appears in $\mathbf{y}$ but a^{n+2} does not. Then $\overline{a}^{n+2}$ does under the assumption. Notice that in the previous case there were exactly $r - 1$ factors having a^{n+2} as a φ^k-ancestor. Under our current assumption, these factors do not have this ancestor, but have instead the ancestor $\overline{a}^{n+2}$. They thus have only one φ^k-factorization. However, as before, the $2^{k-1} - r + 1$ factors with φ^k-ancestor a^{n+1} have a second φ^k-factorization. We conclude that the formula holds also in this case.

Case 3.2. Finally assume that $m' = m$. Then we have that neither a^{n+2} nor $\overline{a}^{n+2}$ appears in $\mathbf{y}$, while both a^{n+1} and $\overline{a}^{n+1}$ do. We thus need to count those factors that have both a^{n+1}

and $\overline{a}^{n+1}$ as φ^k-ancestors. Looking at the previous table, only the center row gives $\overline{a}^{n+2}$ as a φ^k-ancestor. Such factors appear when $|ps| = 2^k + r$ with $2^{k-1} < |p| < 2^{k-1} + r$, i.e., there are $r - 1$ of them whenever $r \leq 2^{k-1}$ (recall that $|p|, |s| < 2^k$). Symmetric arguments apply to factors having φ^k-ancestors $\overline{a}^{n+1}$ (i.e., exchanging the role of a and $\overline{a}$). We conclude that the number of factors having two φ^k-factorizations is $2^k - 2(r-1)$ when $r \leq 2^{k-1}$, as is claimed in the formula.

We are left with the case that $r > 2^{k-1}$. Here we show that no factor has two φ^k-factorizations with respect to $\mathbf{y}$. Since $r > 2^{k-1}$, we have $|ps| = 2^k + r > 2^k + 2^{k-1}$ with $|p|, |s| < 2^k$. It follows that for such φ^k-factorizations we must have $|p|, |s| > 2^{k-1}$, which only leaves the center row of the previous table. But, we already discarded these factors, so the proof is completed. □

5.2 A converse to Theorem 4.2

As announced at the beginning of the section, we now obtain a partial converse statement to Theorem 4.2. Before giving the proof, which is quite long and technical, we give a brief sketch of it. The proof is by induction on k. The induction hypothesis allows to conclude that $\mathbf{x}$ in the statement is essentially the kth image of a recurrent word $\mathbf{z}$. We then show that $\mathbf{z}$ has property $\mathcal{TMB}(1)$ using several times the formulas established in Proposition 5.5. The word $\mathbf{z}$ having property $\mathcal{TMB}(1)$ allows to show that $\mathbf{x}$ is essentially the $(k+1)$st image of another binary word $\mathbf{y}$ which then suffices for the claim by Theorem 5.2.

Proof of Theorem 5.2. Observe first that $\mathbf{x}$ is aperiodic; we shall implicitly use this fact throughout the proof. Indeed, if it was not aperiodic, it would be purely periodic by the recurrence assumption. However, purely periodic words have $\mathsf{b}^{(1)}(n) = 1$ for infinitely many n. This would contradict the assumption that $\mathbf{x}$ has property $\mathcal{TMB}(1)$.

We shall prove the claim by induction. So let first $k = 1$. Then Theorem 5.1 asserts that there exist $a \in \{\varepsilon, 0, 1\}$ and a binary word $\mathbf{y}$ such that $\mathbf{x} = a\varphi(\mathbf{y})$, which was to be proven. Assume then that the claim holds for some k and assume further that $\mathbf{x}$ has property $\mathcal{TMB}(k+1)$. It follows that $\mathbf{x} = u'\varphi^k(\mathbf{z})$, where u' is a proper (possibly empty) suffix of $\varphi^k(0)$ or $\varphi^k(1)$.

Claim 1. *If* $\mathbf{z}$ *has property* $\mathcal{TMB}(1)$*, then* $\mathbf{x}$ *is of the form* $\mathbf{x} = u\varphi^{k+1}(\mathbf{y})$ *with* u *a suffix of* $\varphi^{k+1}(0)$ *or* $\varphi^{k+1}(1)$.

Proof of claim: The assumption implies that $\mathbf{z} = b\varphi(\mathbf{y})$ for some $b \in \{\varepsilon, 0, 1\}$, whence $\mathbf{x} = u'\varphi^k(b)\varphi^{k+1}(\mathbf{y})$. Let y be a prefix of $\mathbf{y}$ that contains both letters 0 and 1. Then the factor $Y = \varphi^{k+1}(y)$ has a unique φ^{k+1}-factorization by Lemma 4.11. Now $u'\varphi^k(b)Y$ appears also in $\varphi^{k+1}(\mathbf{y})$ due to $\mathbf{x}$ being recurrent. In particular, it admits a φ^{k+1}-factorization, and since Y has a unique φ^{k+1}-factorization, we conclude that $u'\varphi^k(b)$ must be the suffix of $\varphi^{k+1}(0)$ or $\varphi^{k+1}(1)$. ■

To prove the theorem, it is thus enough to show that $\mathbf{z}$ has $\mathcal{TMB}(1)$. Indeed, then $\mathbf{x}$ is a suffix of the word of the form $\varphi^{k+1}(\mathbf{y})$ with $\mathbf{y}$ aperiodic, and Theorem 4.2 gives the claim. Notice that $\mathsf{b}_{\mathbf{x}}^{(k+1)} = \mathsf{b}_{\varphi^k(\mathbf{z})}^{(k+1)}$ again due to recurrence of $\mathbf{x}$. This fact is again used throughout the rest of the proof.

Notice now that $\mathbf{z}$ is also recurrent: if it is not recurrent, then it has a prefix w, containing both letters, which appears only once in $\mathbf{z}$. Let us write $\mathbf{z} = w\mathbf{z}'$. However, $\varphi^k(w)$ appears in $\varphi^k(\mathbf{z}')$ by the recurrence of $\mathbf{x}$. Since w contains both letters, the φ^k-factorization of $\varphi^k(w)$ is unique. But, since φ^k is injective, we must find w in $\mathbf{z}'$, a contradiction.

Let us assume towards a contradiction that $\mathbf{z}$ does not have $\mathcal{TMB}(1)$, and let n be the least integer for which

$$\mathsf{b}_{\mathbf{z}}^{(1)}(n) \neq \mathsf{b}_{\mathbf{t}}^{(1)}(n) = \begin{cases} 2, & \text{if } n \text{ is odd;} \\ 3, & \text{if } n \text{ is even.} \end{cases}$$

We now divide the proof into two cases, depending on the parity of n. As it appears, the case where n is even is easier to handle.

5.2.1 n is even

By definition of n, $b_{\mathbf{z}}^{(1)}(n-1) = b_{\mathbf{t}}^{(1)}(n-1) = 2$ and $b_{\mathbf{z}}^{(1)}(n) \neq 3$. Note however that $b_{\mathbf{z}}^{(1)}(n) \neq 1$ because $\mathbf{z}$ is aperiodic. Since $b_{\mathbf{z}}^{(1)}$ can increase or decrease by at most 1 between consecutive values, we conclude that $b_{\mathbf{z}}^{(1)}(n) = 2$.

Claim 2. *We have $n > m$, where m is as in Proposition 5.5.*

Proof of claim: If $n > 2$ (i.e., $n \geq 4$), then $m = 2$ (and $m' = 2$) because $b_{\mathbf{z}}^{(1)}(2) = 3$ implies that 00, $11 \in \mathcal{L}(\mathbf{z})$ while $b_{\mathbf{z}}^{(1)}(3) = 2$ implies that 000, $111 \notin \mathcal{L}(\mathbf{z})$. If $n = 2$ then $m = 1$. ■

We next show that $\#X_{\mathbf{z}}(n) \leq 5$. To this end, let $M_n = \max\{|u|_1 : u \in \mathcal{L}_n(\mathbf{z})\}$, i.e., the maximum weight among length-n factors of $\mathbf{y}$. Assume first that $M_{n-1} = M_n$. Then any factor $v \in \mathcal{L}_{n-1}(\mathbf{z})$ with $|v|_1 = M_{n-1}$ is followed and preceded by 0 (except possibly for the prefix, which is still followed by 0), as otherwise $M_n \neq M_{n-1}$. We conclude that

$$X_{\mathbf{z}}(n) \subseteq \left\{(0, \Psi(v), 0)\right\} \cup \left(\{0,1\} \times \left\{\Psi(v) + (1,-1)\right\} \times \{0,1\}\right),$$

so the claim follows.

Assume second that $M_{n-1} = M_n - 1$. Then each factor $v \in \mathcal{L}_{n-1}(\mathbf{z})$ with $|v|_1 = M_{n-1} - 1$ is followed and preceded by 1; this is because $b_{\mathbf{z}}^{(1)}(n-1) = 2 = b_{\mathbf{z}}^{(1)}(n)$. In this case

$$X_{\mathbf{z}}(n) \subseteq \left\{(1, \Psi(v), 1)\right\} \cup \left(\{0,1\} \times \left\{\Psi(v) + (-1,1)\right\} \times \{0,1\}\right).$$

We have shown $\#X_{\mathbf{z}}(n) \leq 5$. This however leads to a contradiction: applying formula (8), we find

$$3 \cdot 2^{k+1} - 3 = b_{\mathbf{t}}^{(k+1)}(2^k n) = b_{\mathbf{x}}^{(k+1)}(2^k n) = b_{\varphi^k(\mathbf{z})}^{(k+1)}(2^k n) \leq 5 \cdot (2^k - 1) + 2 = 3 \cdot 2^{k+1} - 2^k - 3,$$

where the leftmost equality follows from n being even and (3). We hence move to the case where n is odd.

5.2.2 n is odd

We have that $n \geq 3$ is odd (as $\mathbf{z}$ is binary). Since $b_{\mathbf{z}}^{(1)}(n-1) = 3$, we have $b_{\mathbf{z}}^{(1)}(n) \in \{3,4\}$ arguing as in the case when n was even.

Claim 3. *We have $b_{\mathbf{z}}^{(1)}(n) = 3$.*

Proof of claim: Assume for a contradiction that $b_{\mathbf{z}}^{(1)}(n) = 4$. As $\mathbf{z}$ recurrent, the abelian Rauzy graph G_n has a subgraph of the form depicted in Fig. 3. Further, since $\mathbf{z}$ is also aperiodic, it must

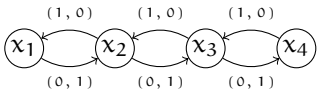

Figure 3: A subgraph of the order-n abelian Rauzy graph of $\mathbf{z}$ in Claim 3.

have at least one loop (at a right special vertex). We conclude that G_n has at least seven edges, that is, $\#X_{\mathbf{z}}(n) \geq 7$. Since n is odd we have, using (8) and recalling that $b_{\mathbf{x}}^{(k+1)} = b_{\varphi^k(\mathbf{z})}^{(k+1)}$

$$3 \cdot 2^{k+1} - 4 = b_{\mathbf{t}}^{(k+1)}(2^k n) = b_{\mathbf{x}}^{(k+1)}(2^k n) \geq \#X_{\mathbf{z}}(n)(2^k - 1) + b_{\mathbf{z}}^{(1)}(n) - 2^k \geq 6 \cdot 2^k - 3 = 3 \cdot 2^{k+1} - 3,$$

which is absurd. ■

Recall the entities m and m' from Proposition 5.5.

Claim 4. *We have $m = 2$. If $n \geq 5$, then $m' = 2$ also. Otherwise $n = 3$ and $m' > 2$.*

Proof of claim: For $n \geq 5$ one may proceed as in the proof of Claim 2. When $n = 3$, we still have that $\mathsf{b}_{\mathbf{z}}^{(1)}(2) = 3$ implies that both 00 and 11 appear in the word. However, $\mathsf{b}_{\mathbf{z}}^{(1)}(3) = 3$ implies that one of 000 or 111 appears in $\mathbf{z}$ while the other does not. ■

Claim 5. *We have that* $k = 1$ *and* $\#X_{\mathbf{z}}(n) = 5$.

Proof of claim: Consider $\mathsf{b}_{\mathbf{x}}^{(k+1)}(2^k n)$; applying (8) (using the previous claim) we have

$$(2^k - 1)\#X_{\mathbf{z}}(n) + \mathsf{b}_{\mathbf{z}}^{(1)}(n) = \mathsf{b}_{\mathbf{x}}^{(k+1)}(2^k n) = \mathsf{b}_{\mathbf{t}}^{(k+1)}(2^k n) = 3 \cdot 2^{k+1} - 4$$

because n is odd. By Claim 3, this is equivalent to

$$(6 - \#X_{\mathbf{z}}(n))2^k + \#X_{\mathbf{z}}(n) = 7.$$

Since $\mathbf{z}$ is aperiodic and recurrent and the abelian Rauzy graph G_n has three vertices, G_n must have at least five edges, i.e., $\#X_{\mathbf{z}}(n) \geq 5$. The only way to satisfy the above equality is when $k = 1$ and $\#X_{\mathbf{z}}(n) = 5$. Indeed, the function $x \mapsto (6 - x)2^k + x$ is strictly decreasing (as $k \geq 1$), and for $x = 6$, it yields 6 which is less than the right-hand side in the above equation. Therefore we must have $\#X_{\mathbf{z}}(n) \leq 5$. We conclude that $\#X_{\mathbf{z}}(n) = 5$. Plugging this into the above equation, we find that $k = 1$, as claimed. ■

The previous claim shows that G_n is a graph with three vertices and five edges. Since $\mathbf{z}$ is recurrent and aperiodic, G_n can be obtained, without loss of generality, by adding one loop to the graph depicted in Fig. 4.

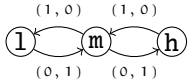

Figure 4: A subgraph of the order-n abelian Rauzy graph G_n of $\mathbf{z}$; G_n is assumed to have three vertices and five edges. Here the leftmost vertex $\mathtt{l}$ corresponds to the lightest abelian equivalence class, the rightmost vertex $\mathtt{h}$ to the heaviest, and the center vertex $\mathtt{m}$ to the remaining class.

For any vertex $v \in \{\mathtt{l}, \mathtt{m}, \mathtt{h}\}$ of G_n, we shall refer to factors of length n having their Parikh-vector corresponding to v as v-*factors*.

Recall that M_n is defined as the maximum weight among factors of length n.

Lemma 5.6. *The graph* G_n *contains either the loop* $\mathtt{h} \xrightarrow{(0,0)} \mathtt{h}$ *or the loop* $\mathtt{l} \xrightarrow{(1,1)} \mathtt{l}$.

Proof. Assume first that $M_n = M_{n-1}$. This implies that all the heaviest factors of length $n - 1$ are surrounded by 0s in $\mathbf{z}$ (meaning, preceded and followed by 0; notice that by assumption that the prefix is not a heavy factor). Hence, there is a factor $0u0$ in $\mathbf{z}$ which corresponds to the loop $\Psi(0u) \xrightarrow{(0,0)} \Psi(u0)$, where $0u$ is an $\mathtt{h}$-factor, because $|u0|_1 = |u|_1 = M_{n-1} = M_n$.

Assume then that $M_n = M_{n-1} + 1$. Since $\mathsf{b}_{\mathbf{z}}^{(1)}(n) = \mathsf{b}_{\mathbf{z}}^{(1)}(n-1) = 3$, we must have that the minimum weight of factors of length n is one greater than that of factors of length $n - 1$; thus all length-$(n-1)$ minimum-weight factors are surrounded by 1s in $\mathbf{z}$. (The only exception is the prefix, which is still followed by 1.) So any non-prefix occurrence of such a factor (recall $\mathbf{z}$ is recurrent) gives a loop on $\mathtt{l}$ with label $(1, 1)$ similar to the above. □

Applying (9) with $r = 1$ and $n + 1 \geq 4 > 2 = m$ (by Claim 4), we have

$$\#Y_{\mathbf{z}}(n) = \mathsf{b}_{\varphi(\mathbf{z})}^{(2)}(2n + 1) = \mathsf{b}_{\mathbf{x}}^{(2)}(2n + 1) = \mathsf{b}_{\mathbf{t}}^{(2)}(2n + 1) = 8.$$

We show that this is impossible. Recall that we add either of the loops $\mathtt{h} \xrightarrow{(0,0)} \mathtt{h}$ or $\mathtt{l} \xrightarrow{(1,1)} \mathtt{l}$ to the graph G_n. In either case we note the following: all $\mathtt{l}$-factors are followed by 1 and all $\mathtt{h}$-factors are followed by 0. In particular, inspecting factors of length $n + 1$, we only have two distinct

Parikh-vectors. Therefore, the graph G_{n+1} has only two vertices, i.e., $\mathsf{b}_{\mathbf{z}}^{(1)}(n+1) = 2$. The number of edges of such a graph is at most six: both vertices can have two loops and one outgoing edge to the other vertex. However, applying (8) (with $n+1 > m$), we find $\#X_{\mathbf{z}}(n+1) + 2 = 9$ because $n+1$ is even. But then $\#X_{\mathbf{z}}(n+1) = 7$, which is impossible. This final contradiction proves that n cannot be odd either. This concludes the proof of Theorem 5.2. □

6 Several Answers to Question A

In this section, we are interested in Question A. Namely, does there exist an infinite word $\mathbf{w}$ such that, for all $k \geq 1$, $\mathsf{b}_{\mathbf{w}}^{(k)}$ is unbounded and $\mathsf{b}_{\mathbf{w}}^{(k)} \prec \mathsf{b}_{\mathbf{w}}^{(k+1)}$? If the answer is positive, can we find a (pure) morphic such word $\mathbf{w}$?

One can give a rather direct answer to this question. Indeed, let $\mathbf{c}$ be the binary Champernowne word, that is, the concatenation of the binary representations of the non-negative integers: 0, 1, 10, 11, 100, 101, 110, 111, ... Notice that $\mathbf{c}$ contains all binary words. For each k, there exist two binary words u, v such that $u \sim_k v$ and $u \not\sim_{k+1} v$ (see, for instance, Theorem 2.4). Therefore, the same properties hold for ux and vx, for all $x \in \{0,1\}^*$, thus $\mathsf{b}_{\mathbf{c}}^{(k)} \prec \mathsf{b}_{\mathbf{c}}^{(k+1)}$ for all k. Clearly $\mathsf{b}_{\mathbf{c}}^{(1)}(n) = n+1$ is unbounded and so is $\mathsf{b}_{\mathbf{c}}^{(k)}$ for $k \geq 2$.

Observe that $\mathbf{c}$ is not morphic, nor *uniformly recurrent* (a word $\mathbf{x}$ is uniformly recurrent if for each $x \in \mathcal{L}(\mathbf{x})$ there exists $N \in \mathbb{N}$ such that x appears in all factors in $\mathcal{L}_N(\mathbf{x})$). Therefore in the rest of the section we provide more "structured" words answering Question A.

6.1 A Non-Binary Pure Morphic Answer

Consider the morphism $g\colon \{a,0,1,b\}^* \to \{a,0,1,b\}^*$ defined by

$$a \mapsto a0b,\ 0 \mapsto \varphi(0),\ 1 \mapsto \varphi(1),\ b \mapsto b^2$$

where φ is the Thue–Morse morphism. We have $\mathbf{g} = g^\omega(a) = a\prod_{j=0}^{\infty} \varphi^j(0)b^{2^j}$. We show that the word $\mathbf{g}$ answers Question A:

Proposition 6.1. *The abelian complexity of* $\mathbf{g}$ *is unbounded and* $\mathsf{b}_{\mathbf{g}}^{(k)} \prec \mathsf{b}_{\mathbf{g}}^{(k+1)}$ *for all* $k \geq 1$.

Proof. The abelian complexity of $\mathbf{g}$ is (at least) linear, since

$$\{|u|_b\colon u \in \mathcal{L}_n(\mathbf{g})\} = \{0,\dots,n\}.$$

Furthermore, for each $k \in \mathbb{N}$ there exist infinitely many words $u_n, v_n \in \mathcal{L}(\mathbf{g})$ such that $u_n \sim_k v_n$ but $u_n \not\sim_{k+1} v_n$: by Theorem 2.4, take $u_n = \varphi^k(0)b^n$ and $v_n = \varphi^k(1)b^n$. Consequently $\mathsf{b}_{\mathbf{g}}^{(k)} \prec \mathsf{b}_{\mathbf{g}}^{(k+1)}$ for all $k \geq 1$. □

6.2 A Binary Morphic Answer

Consider the word $\tau(\mathbf{g})$, where $\mathbf{g}$ is the word defined in the previous subsection, and τ is the coding $a \mapsto \varepsilon$, $0 \mapsto 0$, $1 \mapsto 1$, and $b \mapsto 1$. We have the following:

Proposition 6.2. *The abelian complexity of* $\tau(\mathbf{g})$ *is unbounded and* $\mathsf{b}_{\tau(\mathbf{g})}^{(k)} \prec \mathsf{b}_{\tau(\mathbf{g})}^{(k+1)}$ *for all* $k \geq 1$.

Proof. The word $\tau(\mathbf{g})$ has unbounded abelian complexity: it contains arbitrarily long words u for which $|u|_1 = \lfloor |u|/2 \rfloor$ (take factors of the Thue–Morse word for instance). Similarly it contains arbitrarily long powers of 1. Consequently, the word has unbounded abelian complexity (recall Lemma 2.1).

To show $\mathsf{b}_{\tau(\mathbf{g})}^{(k)} \prec \mathsf{b}_{\tau(\mathbf{g})}^{(k+1)}$ for all k, we notice that the same arguments as in the case of $\mathbf{g}$ can be applied verbatim with $\tau(u_n)$ and $\tau(v_n)$. □

6.3 A Binary Uniformly Recurrent Answer

We note that none of the above words are uniformly recurrent. A natural candidate for such a word is one that has relatively high factor complexity. Uniformly recurrent words having positive *topological entropy*[4] were studied by Grillenberger in [23]. A construction for uniformly recurrent positive entropy words appears in [9, §4.4.3]; this construction is simpler than that of Grillenberger's, though some properties are lost (see [9, §4.4.3] for a discussion). We recall this construction here. To attain a word with entropy between 0 and $\log d$, define $D_0 = \{0, 1, \ldots, d-1\}$ and let $(q_k)_{k\geq 0}$ be a sequence of positive integers. Assuming D_k is constructed, let u_k be the product of words of D_k in lexicographic order (assuming, e.g., $0 < 1 < \ldots < d-1$). Define then $D_{k+1} := u_k D_k^{q_k}$. The sequence $(u_k)_{k\in\mathbb{N}}$ converges to a uniformly recurrent word $\mathbf{u}$ having, with a suitable choice of (q_k), the prescribed entropy. We consider the word with $d = 2$ and $q_k = 2$ for all k (and are not interested in the entropy). Hence for us $\mathbf{u} = 0100010101100111\cdots$.

Lemma 6.3. *Let $k \geq 1$. If, for some $j \geq 0$, D_j contains two words u, v, such that $u \sim_k v$ and $u \not\sim_{k+1} v$, then D_{j+1} contains words x, y, z and w such that*

- *$x \sim_k y$ but $x \not\sim_{k+1} y$;*
- *$z \sim_{k+1} w$ but $z \not\sim_{k+2} w$.*

Proof. By definition, the set D_{j+1} contains the words $x = u_j uu$, $y = u_j vv$, $z = u_j uv$, and $w = u_j vu$.

We first consider the pair x, y. Since $\sim_k$ is a congruence, $x \sim_k y$. To see that $x \not\sim_{k+1} y$, assume the contrary, so that this equivalence reduces to $uu \sim_{k+1} vv$ by Lemma 2.2. Lemma 2.6 implies $u \sim_{k+1} v$, a contradiction.

Next we have $uv \sim_{k+1} vu$ by Theorem 2.3, and thus $z = u_j uv \sim_{k+1} u_j vu = w$ by Lemma 2.2. Similarly $z \sim_{k+2} w$ would imply $uv \sim_{k+2} vu$ and thus $u \sim_{k+1} v$ by Theorem 2.3, a contradiction. The claim follows. □

Theorem 6.4. *The abelian complexity of $\mathbf{u}$ is unbounded and $\mathsf{b}_{\mathbf{u}}^{(k)} \prec \mathsf{b}_{\mathbf{u}}^{(k+1)}$ for all $k \geq 1$.*

Proof. First we show that $\mathsf{b}_{\mathbf{u}}^{(1)}$ is unbounded. Assume, for some $j \geq 0$, that D_j contains words u, v with $|u|_0 - |v|_0 = 2^j$ (this holds for $j = 0$). Then by definition D_{j+1} contains the words $x = u_j uu$ and $y = u_j vv$, for which $|x|_0 - |y|_0 = 2(|u|_0 - |v|_0) = 2^{j+1}$. This observation suffices for the claim by Lemma 2.1.

We then prove the second part of the statement. Observe that D_1 contains the words 0101 and 0110, which are abelian equivalent, but not 2-binomially equivalent (as $\binom{0101}{01} = 3$ and $\binom{0110}{01} = 2$). The above lemma then implies that for all $k \geq 1$ and for all $j \geq k$, the set D_j contains words that are k-binomially equivalent, but not $(k+1)$-binomially equivalent. The claim follows. □

Remark 6.5. It can be shown that the word $\mathbf{u}$ above has topological entropy equal to 0. By modifying the arguments above suitably, the statement of the above theorem holds for any choice of d and (q_k) in the construction—as long as $q_k > 1$ for infinitely many k. Note that to attain a word with positive entropy, the sequence (q_k) must satisfy this property. Hence we have: *For any positive real number h there is a uniformly recurrent d-ary word (with $d = \max\{2, \lfloor h \rfloor + 1\}$) having entropy h, unbounded $\mathsf{b}^{(1)}$, and $\mathsf{b}^{(k)} \prec \mathsf{b}^{(k+1)}$ for all k.*

7 Answer to Question B and Beyond

In this section, we are interested in Question B. Namely, for each $\ell \geq 1$, does there exist a word $\mathbf{w}$ (depending on ℓ) such that $\mathsf{b}_{\mathbf{w}}^{(1)} \prec \mathsf{b}_{\mathbf{w}}^{(2)} \prec \cdots \prec \mathsf{b}_{\mathbf{w}}^{(\ell-1)} \prec \mathsf{b}_{\mathbf{w}}^{(\ell)} = \mathsf{p}_{\mathbf{w}}$? If the answer is positive, is there a (pure) morphic such word $\mathbf{w}$?

The word 0^ω gives $\mathsf{b}^{(1)} = \mathsf{p}$. The Fibonacci word $\mathbf{f} = 0100101001001010010\cdots$, the fixed point of the morphism $0 \mapsto 01, 1 \mapsto 0$, is a pure morphic word such that $2 = \mathsf{b}_{\mathbf{f}}^{(1)} \prec \mathsf{b}_{\mathbf{f}}^{(2)} = \mathsf{p}_{\mathbf{f}}$ by Theorem 2.8.

[4] The topological entropy of a word $\mathbf{x}$ is defined as the quantity $\lim_{n\to\infty} \frac{\log p_{\mathbf{x}}(n)}{n}$, which exists for any $\mathbf{x}$ (see [9, §4.3.2]).

Remark 7.1. We notice that $b_x^{(1)} = p_x$ cannot be attained for an aperiodic word $\mathbf{x}$ (indeed, there must exist a factor avb, with $a \in A$ and v containing a letter different to a, whence $av \sim_1 va$ with $av \neq va$). In fact, the only ultimately periodic words over an m-letter alphabet $\{a_1, \dots, a_m\}$ for which the equality holds are of the form $a_1^{n_1} a_2^{n_2} \cdots a_m^{\omega}$, $n_i \in \mathbb{N}$ (up to permutation of the letters).

To answer Question B for larger values of k, we take images of a Sturmian word $\mathbf{s}$ by a power of the Thue–Morse morphism φ and we prove the following result.

Theorem 7.2. *Let φ be the Thue–Morse morphism. Let $\mathbf{s}$ be a Sturmian word. For each $k \geq 0$, the word $\mathbf{s}_k := \varphi^k(\mathbf{s})$ has*

$$b_{\mathbf{s}_k}^{(1)} \prec b_{\mathbf{s}_k}^{(2)} \prec \cdots \prec b_{\mathbf{s}_k}^{(k+1)} \prec b_{\mathbf{s}_k}^{(k+2)} = p_{\mathbf{s}_k}.$$

In particular, putting the Fibonacci word for $\mathbf{s}$ gives a morphic positive answer to Question B.

Proof. Observe that $\mathbf{s}_k$ has bounded $(k+1)$-binomial complexity as a straightforward application of Theorem 3.5 (because $\mathbf{s}$ has bounded abelian complexity), and thus $b_{\mathbf{s}_k}^{(k+1)} \prec p_{\mathbf{s}_k}$. By Corollary 4.18, we need only to show that $b_{\mathbf{s}_k}^{(k+2)} = p_{\mathbf{s}_k}$.

Let u and v be distinct factors of $\mathbf{s}_k$. Assume they are $(k+2)$-binomially equivalent. By Proposition 4.17, we have that $u = p\varphi^k(z)s$, $v = p\varphi^k(z')s$ with $z \sim_1 z'$. If $z \neq z'$, then $z \not\sim_2 z'$ by Theorem 2.8. But then Lemma 4.14 implies that $\varphi^k(z) \not\sim_{k+2} \varphi^k(z')$, contradicting the assumption. Hence we deduce that $z = z'$, but then $u = v$ contrary to the assumption. □

Remark 7.3. In the above proof, since $\mathbf{s}$ is Sturmian, Theorem 2.8 says distinct factors are not 2-binomially equivalent. This means that Theorem 7.2 applies to and only to aperiodic words $\mathbf{s}$ such that $b_{\mathbf{s}}^{(2)} = p_{\mathbf{s}}$. The "only if"-part of the statement follows by a repeated application of Proposition 3.9 on the Thue–Morse morphism together with the fact that the morphism is injective.

7.1 Strengthening Question B

We answered Question B by providing a word with bounded abelian complexity. We can therefore strengthen the question with the following extra requirement.

Question C. For each $\ell \geq 1$, does there exist a word $\mathbf{w}$ (depending on ℓ) such that $b_{\mathbf{w}}^{(1)}$ is unbounded and

$$b_{\mathbf{w}}^{(1)} \prec b_{\mathbf{w}}^{(2)} \prec \cdots \prec b_{\mathbf{w}}^{(\ell-1)} \prec b_{\mathbf{w}}^{(\ell)} = p_{\mathbf{w}}?$$

If the answer is positive, can we find a (pure) morphic such word $\mathbf{w}$?

The following word answers the question for $\ell = 3$ in the positive.

Theorem 7.4. *The word $\mathbf{h} = 011212212221222122222\cdots$ fixed point of the morphism $0 \mapsto 01$, $1 \mapsto 12$, and $2 \mapsto 2$ is such that its abelian complexity $b_{\mathbf{h}}^{(1)}$ is unbounded and $b_{\mathbf{h}}^{(1)} \prec b_{\mathbf{h}}^{(2)} \prec b_{\mathbf{h}}^{(3)} = p_{\mathbf{h}}$.*

We obtain the previous theorem by combining the following two results.

Proposition 7.5. *The abelian complexity $b_{\mathbf{h}}^{(1)}$ of $\mathbf{h}$ is unbounded and $b_{\mathbf{h}}^{(1)}(n) < b_{\mathbf{h}}^{(2)}(n) < p_{\mathbf{h}}(n)$ for all $n \geq 6$.*

Proof. We claim that $b_{\mathbf{h}}^{(1)}$ is of the order $\Theta(\sqrt{n})$. Clearly it suffices to show the claim for the word $\mathbf{h}' = 0^{-1}\mathbf{h}$, as removing the first zero always removes exactly one abelian equivalence class: the only one that contains a zero. The resulting word $\mathbf{h}'$ is effectively a binary word; it is evident that the maximal number of 1's in a word of length n is attained by the prefix of $\mathbf{h}'$. This value equals the maximal m for which $\sum_{i=1}^{m} i = \binom{m+1}{2} \leq n$. Clearly $m = \Theta(\sqrt{n})$. By Lemma 2.1, we conclude that the abelian complexity of $\mathbf{h}$ is $\Theta(\sqrt{n})$.

Since the abelian complexity of $\mathbf{h}$ if unbounded, so is its 2-binomial complexity. However, the 2-binomial complexity does not equal the factor complexity at lengths $n \geq 6$: $\mathbf{h}$ contains both the factors $12^{n-2}1$ and $212^{n-4}12$ which are readily seen to be 2-binomially equivalent. (One may also invoke a result from [20] for binary alphabets.)

Finally observe that the abelian complexity does not coincide with the 2-binomial complexity either: the factors 2^x12^y with $x+y = n-1$ are abelian equivalent but not 2-binomially equivalent. This ends the proof. □

Proposition 7.6. *We have* $\mathsf{b}_{\mathbf{h}}^{(3)} = \mathsf{p}_{\mathbf{h}}$.

Proof. We may again discard the first 0 of $\mathbf{h}$, as the prefix is the only factor containing a zero. Assume to the contrary that there exist 3-binomially equivalent distinct factors u_1 and u_2 in $\mathbf{h}' = 0^{-1}\mathbf{h}$. The two factors must contain the same number of 1's, and hence at least one under the assumption that they are distinct. If the factors are of the form $u_i = 2^{x_i}12^{y_i}$ with $x_1 \neq x_2$, then the factors are not even 2-binomially equivalent. So the words contain at least two 1's. By the structure of $\mathbf{h}$, we may write $u_i = 2^{x_i}12^{a_i}12^{a_i+1}1\cdots12^{a_i+t}12^{y_i}$ for some $t \geq 0$, $a_i \in \mathbb{N}$, $x_i < a_i$ and $y_i \leq a_i + t + 1$ for all $i \in \{1, 2\}$. If $a_1 = a_2$, then $x_1 \neq x_2$, and we again deduce that the factors are not even 2-binomially equivalent. So we must have $a_1 < a_2$ without loss of generality. We show that in this case the factors are not 3-binomially equivalent. Indeed, consider the coefficient $\binom{\cdot}{121}$. For $i = 1, 2$, we clearly have

$$\binom{u_i}{121} = \binom{v_i}{121}, \tag{11}$$

where $v_i = 12^{a_i}12^{a_i+1}1\cdots12^{a_i+t}1$ is obtained from u_i by deleting a prefix and a suffix. But, since $a_1 < a_2$, notice now that v_1 is a proper subword of v_2, meaning that each occurrence of 121 in v_1 has a corresponding occurrence in v_2. Clearly v_2 will have more occurrences of 121. This combined with (11) gives the claim. □

Remark 7.7. Consider the morphic words $\mathbf{h}_k = f_k^\omega(0)$, where $f_k : \{0, \ldots, k\} \to \{0, \ldots, k\}^*$ is defined by $i \mapsto i(i+1)$ for $i < k$, and $k \mapsto k$. It can be shown that each of the words $\mathbf{h}_k$, $k \geq 2$, has the property $\mathsf{b}^{(1)} \prec \mathsf{b}^{(2)} \prec \mathsf{b}^{(3)} = \mathsf{p}$ (this was shown for $\mathbf{h}_2$ in the above). Indeed, one may proceed by induction on k; we find all factors of f_k in f_{k+1} (up to renaming the letters $i \mapsto i+1$), so the first two relations hold immediately. To show the equality between $\mathsf{b}^{(3)}$ and p, one may proceed as in Proposition 7.6. If two factors were 3-binomially equivalent, they should contain at least one occurrence of 1 (otherwise they are factors of $1\prod_{i=0}^{m} f_k^i(2)$ for some large enough m, and are factors of $f_{k-1} = 0\prod_{i=0}^{\omega} f_{k-1}^i(1)$ after renaming letters). If they contain exactly one occurrence of 1, then inspecting the coefficients $\binom{\cdot}{1j}$, $j \geq 2$, reveals that the maximal suffixes not containing a 1 must be of the same length. Since they are both prefixes of the same word ($\prod_{i=0}^{m} f^i(2)$ for some large enough m), they must be equal. Hence, by cancellativity, we may remove the common suffix starting from 1 and keep 3-binomial equivalence, which is a contradiction. If the words contained at least two 1s, then inspecting the coefficients $\binom{\cdot}{121}$ would reveal that the first occurrence of 1 in both words must correspond to the same occurrence of 1 in $\mathbf{h}_k$ (much like in the proof of Proposition 7.6). Inspecting the coefficients $\binom{\cdot}{1j}$, one can proceed as in the case where only one 1 was assumed to conclude with a contradiction.

A complete answer to Question C is far from obvious; especially if one wishes to obtain a pure morphic word. Conversely, for a non-periodic morphic word $\mathbf{w}$ which is not the fixed point of a Parikh-collinear morphism, one can wonder about the existence of a minimal value m for which the binomial and factor complexities would coincide. Does there exists $m \in \mathbb{N}$ such that $\mathsf{b}_{\mathbf{w}}^{(m)} = \mathsf{p}_{\mathbf{w}}$?

Even with an apparently simple situation, it is far from obvious. As stated in the introduction, computing the k-binomial complexity of a particular infinite word remains challenging. The period-doubling word $\mathbf{pd} = 01000101010001\cdots$, the fixed point of $\sigma : 0 \mapsto 01,\ 1 \mapsto 00$, can be proved to have the following properties. Its abelian complexity $\mathsf{b}_{\mathbf{pd}}^{(1)}$ is unbounded [27, Lem. 4].

For the 2-binomial complexity, we have $\mathsf{b}^{(2)}_{\mathbf{pd}}(2^n) = \mathsf{p}_{\mathbf{pd}}(2^n)$ for all n, but $\mathsf{b}^{(2)}_{\mathbf{pd}}(n) < \mathsf{p}_{\mathbf{pd}}(n)$ for all $n \neq 2^m$ [28, Prop. 4.5.1]. Otherwise stated, $\mathsf{b}^{(1)}_{\mathbf{pd}} \prec \mathsf{b}^{(2)}_{\mathbf{pd}} \prec \mathsf{p}_{\mathbf{pd}}$. Computer experiments show that $\mathsf{b}^{(3)}_{\mathbf{pd}} \prec \mathsf{p}_{\mathbf{pd}}$ and suggest that $\mathsf{b}^{(4)}_{\mathbf{pd}} = \mathsf{p}_{\mathbf{pd}}$.

7.2 Completing the Binomial Complexities of φ^k Applied to a Sturmian word

For any $k \geq 1$, the results presented so far imply that we have the exact j-binomial complexity function of $\varphi^k(\mathbf{s})$, with $\mathbf{s}$ a Sturmian word, for each $j \neq k+1$. As a bonus, we compute the $(k+1)$-binomial complexity in Proposition 7.9. We first analyze the abelian Rauzy graphs of Sturmian words, after which we may apply Proposition 5.5 to obtain the exact $(k+1)$-binomial complexity as well.

A Sturmian word $\mathbf{s}$ has $\mathsf{b}^{(1)}_{\mathbf{s}}(n) = 2$ for all $n \geq 1$. Hence its abelian Rauzy graph has two vertices.

Proposition 7.8. *Let* $G_n = (V_n, E_n)$. *We have* $\#E_1 = 3$ *and* $\#E_n = 4$ *for all* $n \geq 2$. *For all* $n \geq 1$, *we have* $\#E_n/\equiv_L + \#E_n/\equiv_R = 6$.

Proof. Sturmian words are aperiodic, so the graph G_n is always strongly connected. It also always has a right special vertex.

The claim is plain to verify for G_1. Indeed, only one of the vertices can have a loop, and this loop can only be labeled with (a, a), where a is the letter corresponding to the vertex in question. The second claim is straightforward to check in this case.

We next consider G_n, $n \geq 2$. Since G_n is strongly connected and has a right special vertex, we conclude that G_n is obtained by adding (possibly zero) loops to one of the graphs in Fig. 5

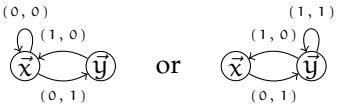

Figure 5: Possible subgraphs of a Sturmian word's abelian Rauzy graph.

By symmetry, we may assume it is the one on the left. We claim that we add exactly one loop to this graph. More precisely, the added loop is either $\vec{x} \xrightarrow{(1,1)} \vec{x}$ or $\vec{y} \xrightarrow{(0,0)} \vec{y}$.

First off, we cannot add the loop $\vec{y} \xrightarrow{(1,1)} \vec{y}$; otherwise we have the factors $x0$ and $y1$, where x is an $\vec{x}$-factor and y is a $\vec{y}$-factors, for which we have $|y1|_1 - |x0|_1 = 2$ contradicting balancedness at $n+1$. Towards a contradiction, we consider whether we add neither or both of the remaining admissible loops.

Assume first that we add both loops. Then we have that length $n+1$ factors $1x1$ and $0y0$, where $1x$, $x1$ are $\vec{x}$-factors and $0y$, $y0$ are $\vec{y}$-factors, having $|1x1|_1 = |0y0|_1$. However, we then have $|y|_1 - |x|_1 = 2$, which gives a contradiction with balancedness at $n-1$.

To complete the proof of the first claim, suppose we add neither of the loops. Inspecting G_n, we see that a factor of length n is an $\vec{x}$-factor if and only if it begins with 0. Consider the right special factor $0v$ of length n (i.e., $|v| = n-1 \geq 1$) (it begins with 0 by the form of the graph). Since $v1$ is a $\vec{y}$-factor, we deduce that v begins with a 1. But then $v0$ is a $\vec{x}$-factor beginning with 1, a contradiction.

The second part of the claim is now straightforward. The edges in the left-hand graph in Fig. 5 are all pairwise inequivalent under both $\equiv_L$ and $\equiv_R$. Both of the admissible loops to be added to the graph to obtain G_n are equivalent to the non-loop edges of the graph. Hence $\#E_n/\equiv_L + \#E_n/\equiv_R = 6$. □

For a Sturmian word $\mathbf{s}$, we have, by Proposition 7.8, $X_{\mathbf{s}}(1) = 3$, $X_{\mathbf{s}}(n) = 4$ for all $n \geq 2$, and $Y_{\mathbf{s}}(n) = 6$ for all $n \geq 1$. We also have that $1 = m < m'$ using the notation of Proposition 5.5. Hence, applying the proposition, we have:

Proposition 7.9. *For all* $n \geq 0$ *and* $0 \leq r < 2^k$

$$b^{(k+1)}_{\varphi^k(\mathbf{s})}(2^k n + r) = \begin{cases} p_{\mathbf{t}}(r), & \text{if } n = 0 \text{ and } 0 \leq r < 2^k; \\ 3 \cdot 2^k - 2, & \text{if } n = 1 \text{ and } r = 0; \\ 3 \cdot 2^k + r - 1, & \text{if } n = 1 \text{ and } r > 0; \\ 2^{k+2} - 2, & \text{otherwise.} \end{cases} \tag{12}$$

We thus conclude the following. For any $k \geq 1$ and a Sturmian word $\mathbf{s}$, we have $b^{(j)}_{\varphi^k(\mathbf{s})} = b^{(j)}_{\mathbf{t}}$ if $1 \leq j \leq k$ (Theorem 4.2); $b^{(k+1)}_{\varphi^k(\mathbf{s})}$ is as in (12); and $b^{(j)}_{\varphi^k(\mathbf{s})} = p_{\varphi^k(\mathbf{s})}$ if $j \geq k+2$ (Theorem 7.2). The exact value for $p_{\varphi^k(\mathbf{s})}(n)$ is given by $p_{\mathbf{t}}(n)$ when $n \leq 2^k$ (Lemma 4.8), and by $n + 2^{k+1} - 1$ for $n > 2^k$. The latter can be deduced by using the methods described in [21, §4.1].

Acknowledgments

We thank the reviewers for their comments improving the paper.

References

[1] Boris Adamczewski. Balances for fixed points of primitive substitutions. *Theoret. Comput. Sci.*, 307(1):47–75, 2003. doi:10.1016/S0304-3975(03)00092-6.

[2] Boris Adamczewski and Yann Bugeaud. On the complexity of algebraic numbers. I. Expansions in integer bases. *Ann. of Math. (2)*, 165(2):547–565, 2007. doi:10.4007/annals.2007.165.547.

[3] Jean-Paul Allouche. Thue, combinatorics on words, and conjectures inspired by the Thue-Morse sequence. *J. Théor. Nombres Bordeaux*, 27(2):375–388, 2015. URL: http://jtnb.cedram.org/item?id=JTNB_2015__27_2_375_0.

[4] Jean-Paul Allouche and Jeffrey Shallit. The ubiquitous Prouhet–Thue–Morse sequence. In C. Ding, T. Helleseth, and H. Niederreiter, editors, *Sequences and their Applications*, pages 1–16, London, 1999. Springer London. doi:10.1007/978-1-4471-0551-0_1.

[5] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences: Theory, applications, generalizations*. Cambridge University Press, Cambridge, 2003. doi:10.1017/CBO9780511546563.

[6] Adrian Atanasiu, Carlos Martín-Vide, and Alexandru Mateescu. On the injectivity of the Parikh matrix mapping. *Fund. Inform.*, 49(4):289–299, 2002.

[7] Sergei V. Avgustinovich, Dmitriĭ G. Fon-Der-Flaass, and Anna E. Frid. Arithmetical complexity of infinite words. In *Words, languages & combinatorics, III (Kyoto, 2000)*, pages 51–62. World Sci. Publ., River Edge, NJ, 2003. doi:10.1142/9789812704979_0004.

[8] Jean Berstel, Maxime Crochemore, and Jean-Éric Pin. Thue-Morse sequence and p-adic topology for the free monoid. *Discrete Math.*, 76(2):89–94, 1989. doi:10.1016/0012-365X(89)90302-6.

[9] Valérie Berthé and Michel Rigo, editors. *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010. doi:10.1017/CBO9780511777653.

[10] Julien Cassaigne, Gabriele Fici, Marinella Sciortino, and Luca Q. Zamboni. Cyclic complexity of words. *J. Comb. Theory, Ser. A*, 145:36–56, 2017. doi:10.1016/j.jcta.2016.07.002.

[11] Julien Cassaigne, Anna E. Frid, Svetlana Puzynina, and Luca Q. Zamboni. A characterization of words of linear complexity. *Proc. Amer. Math. Soc.*, 147(7):3103–3115, 2019. doi:10.1090/proc/14440.

[12] Julien Cassaigne, Juhani Karhumäki, and Svetlana Puzynina. On k-abelian palindromes. *Inf. Comput.*, 260:89–98, 2018. doi:10.1016/j.ic.2018.04.001.

[13] Julien Cassaigne, Juhani Karhumäki, and Aleksi Saarela. On growth and fluctuation of k-abelian complexity. *Eur. J. Comb.*, 65:92–105, 2017. doi:10.1016/j.ejc.2017.05.006.

[14] Julien Cassaigne, Gwéenaël Richomme, Kalle Saari, and Luca Q. Zamboni. Avoiding Abelian powers in binary words with bounded Abelian complexity. *Int. J. Found. Comput. S.*, 22(4):905–920, 2011. doi:10.1142/S0129054111008489.

[15] Ethan M. Coven and G. A. Hedlund. Sequences with minimal block growth. *Math. Syst. Theory*, 7(2):138–153, 1973. doi:10.1007/BF01762232.

[16] Xavier Droubay and Giuseppe Pirillo. Palindromes and Sturmian words. *Theoret. Comput. Sci.*, 223(1-2):73–85, 1999. doi:10.1016/S0304-3975(97)00188-6.

[17] Paul Erdős. Some unsolved problems. *Michigan Math. J.*, 4:291–300, 1958.

[18] Gabriele Fici and Svetlana Puzynina. Abelian combinatorics on words: a survey. *Comput. Sci. Rev.*, 47:Paper No. 100532, 21, 2023. doi:10.1016/j.cosrev.2022.100532.

[19] Pamela Fleischmann, Marie Lejeune, Florin Manea, Dirk Nowotka, and Michel Rigo. Reconstructing words from right-bounded-block words. *Int. J. Found. Comput. Sci.*, 32(6):619–640, 2021. doi:10.1142/S0129054121420016.

[20] Stéphane Fossé and Gwénaël Richomme. Some characterizations of Parikh matrix equivalent binary words. *Inform. Process. Lett.*, 92(2):77–82, 2004. doi:10.1016/j.ipl.2004.06.011.

[21] Anna Frid. Applying a uniform marked morphism to a word. *Discrete Math. Theor. Comput. Sci.*, 3(3):125–139, 1999. doi:10.46298/dmtcs.255.

[22] Paweł Gawrychowski, Maria Kosche, Tore Koß, Florin Manea, and Stefan Siemer. Efficiently testing Simon's congruence. In *38th International Symposium on Theoretical Aspects of Computer Science*, volume 187 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 34, 18. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021.

[23] Christian Grillenberger. Constructions of strictly ergodic systems. I. Given entropy. *Z. Wahrscheinlichkeit.*, 25:323–334, 1973. doi:10.1007/BF00537161.

[24] L. I. Kalašnik. The reconstruction of a word from fragments. In *Numerical mathematics and computer technology, No. IV (Russian)*, pages 56–57, 137. Akad. Nauk Ukrain. SSR Fiz.-Tehn-Inst. Nizkih Temperatur, Kharkov, 1973.

[25] Teturo Kamae and Luca Zamboni. Sequence entropy and the maximal pattern complexity of infinite words. *Ergodic Theory Dynam. Systems*, 22(4):1191–1199, 2002. doi:10.1017/S0143385702000585.

[26] Juhani Karhumäki, Aleksi Saarela, and Luca Q. Zamboni. On a generalization of abelian equivalence and complexity of infinite words. *J. Comb. Theory, Ser. A*, 120(8):2189–2206, 2013. doi:10.1016/j.jcta.2013.08.008.

[27] Juhani Karhumäki, Aleksi Saarela, and Luca Q. Zamboni. Variations of the Morse-Hedlund theorem for k-abelian equivalence. *Acta Cybernet.*, 23(1):175–189, 2017. doi:10.14232/actacyb.23.1.2017.11.

[28] Marie Lejeune. *On the* k*-binomial equivalence of finite words and* k*-binomial complexity of infinite words*. PhD thesis, Univ. of Liège, 2021. URL: http://hdl.handle.net/2268/259266.

[29] Marie Lejeune, Julien Leroy, and Michel Rigo. Computing the k-binomial complexity of the Thue–Morse word. *J. Comb. Theory, Ser. A*, 176:44, 2020. doi:10.1016/j.jcta.2020.105284.

[30] Marie Lejeune, Michel Rigo, and Matthieu Rosenfeld. Templates for the k-binomial complexity of the Tribonacci word. *Adv. Appl. Math.*, 112:26, 2020. doi:10.1016/j.aam.2019.101947.

[31] Julien Leroy, Michel Rigo, and Manon Stipulanti. Generalized Pascal triangle for binomial coefficients of words. *Adv. Appl. Math.*, 80:24–47, 2016. doi:10.1016/j.aam.2016.04.006.

[32] M. Lothaire. *Combinatorics on Words*. Cambridge Mathematical Library. Cambridge University Press, 1997. doi:10.1017/CBO9780511566097.

[33] Xiao-Tao Lü, Jin Chen, Zhi-Xiong Wen, and Wen Wu. On the 2-binomial complexity of the generalized Thue–Morse words, 2021. (preprint). doi:10.48550/ARXIV.2112.05347.

[34] Bennet Manvel, Aaron D. Meyerowitz, Allen J. Schwenk, Kenneth W. Smith, and Paul K. Stockmeyer. Reconstruction of sequences. *Discrete Math.*, 94(3):209–219, 1991. doi:10.1016/0012-365X(91)90026-X.

[35] Hamoon Mousavi. Automatic theorem proving in Walnut, 2016. doi:10.48550/ARXIV.1603.06017.

[36] Peter Ochsenschläger. Binomialkoeffizienten und Shuffle-Zahlen. Technischer Bericht, Fachbereich Informatik, T.H. Darmstadt, 1981.

[37] Jarkko Peltomäki. Introducing privileged words: Privileged complexity of Sturmian words. *Theor. Comput. Sci.*, 500:57–67, 2013. doi:10.1016/j.tcs.2013.05.028.

[38] Jean-Éric Pin and Pedro V. Silva. A noncommutative extension of Mahler's theorem on interpolation series. *European J. Combin.*, 36:564–578, 2014. doi:10.1016/j.ejc.2013.09.009.

[39] Gwénaël Richomme, Kalle Saari, and Luca Q. Zamboni. Balance and abelian complexity of the tribonacci word. *Adv. Appl. Math.*, 45(2):212–231, 2010. doi:10.1016/j.aam.2010.01.006.

[40] Gwénaël Richomme, Kalle Saari, and Luca Q. Zamboni. Abelian complexity of minimal subshifts. *J. Lond. Math. Soc.*, 83(1):79–95, 2011. doi:10.1112/jlms/jdq063.

[41] Gwénaël Richomme and Patrice Séébold. On factorially balanced sets of words. *Theoret. Comput. Sci.*, 412(39):5492–5497, 2011. doi:10.1016/j.tcs.2011.06.027.

[42] Michel Rigo. Relations on words. *Indag. Math., New Ser.*, 28(1):183–204, 2017. doi:10.1016/j.indag.2016.11.018.

[43] Michel Rigo and Pavel Salimov. Another generalization of abelian equivalence: binomial complexity of infinite words. *Theor. Comput. Sci.*, 601:47–57, 2015. doi:10.1016/j.tcs.2015.07.025.

[44] Michel Rigo, Manon Stipulanti, and Markus A. Whiteland. Binomial complexities and Parikh-collinear morphisms. In Volker Diekert and Mikhail V. Volkov, editors, *Developments in Language Theory - 26th International Conference, DLT 2022, Tampa, FL, USA, May 9-13, 2022, Proceedings*, volume 13257 of *Lecture Notes in Computer Science*, pages 251–262. Springer, 2022. doi:10.1007/978-3-031-05578-2_20.

[45] Michel Rigo, Manon Stipulanti, and Markus A. Whiteland. Automaticity and Parikh-collinear morphisms. In *Combinatorics on words*, volume 13899 of *Lecture Notes in Comput. Sci.*, pages 247–260. Springer, Cham, [2023] ©2023. doi:10.1007/978-3-031-33180-0_19.

[46] Arto Salomaa. Counting (scattered) subwords. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, 81:165–179, 2003.

[47] Jeffrey Shallit. *The Logical Approach to Automatic Sequences: Exploring Combinatorics on Words with Walnut*. London Mathematical Society Lecture Note Series. Cambridge University Press, 2022. doi:10.1017/9781108775267.

[48] Markus A. Whiteland. Equations over the k-binomial monoids. In *WORDS 2021*, volume 12847 of *LNCS*, pages 185–197. Springer, Cham., 2021. doi:10.1007/978-3-030-85088-3_16.