Binomial Complexities and Parikh-Collinear Morphisms*

Michel Rigo[0000−0001−7463−8507], Manon Stipulanti[0000−0002−2805−2465], and Markus A. Whiteland[0000−0002−6006−9902]

Department of Mathematics, University of Liège, Liège, Belgium
{m.rigo,m.stipulanti,mwhiteland}@uliege.be

Abstract. Two words are k-binomially equivalent, if each word of length at most k occurs as a subword, or scattered factor, the same number of times in both words. The k-binomial complexity of an infinite word maps the natural n to the number of k-binomial equivalence classes represented by its factors of length n. Inspired by questions raised by Lejeune, we study the relationships between the k and $(k+1)$-binomial complexities; as well as the link with the usual factor complexity. We show that pure morphic words obtained by iterating a Parikh-collinear morphism, i.e. a morphism mapping all words to words with bounded abelian complexity, have bounded k-binomial complexity. In particular, we study the properties of the image of a Sturmian word by an iterate of the Thue–Morse morphism.

Keywords: Factor complexity · Abelian complexity · Binomial complexity · iterates of Thue–Morse morphism.

1 Introduction

When we are interested in the combinatorial structure of an infinite word $\mathbf{x}$ over a finite alphabet A, it is often useful to study its language $\mathcal{L}(\mathbf{x})$, i.e. the set of its factors, and in particular to look at factors of a given length n. We let $\mathcal{L}_n(\mathbf{x})$ denote $\mathcal{L}(\mathbf{x}) \cap A^n$. The usual *factor complexity* function $\mathsf{p}_{\mathbf{x}} \colon \mathbb{N} \to \mathbb{N}$ counts the number $\# \mathcal{L}_n(\mathbf{x})$ of words of length n occurring in $\mathbf{x}$. For instance, ultimately periodic words are characterized by a bounded factor complexity and Sturmian words are exactly those words satisfying $\mathsf{p}_{\mathbf{x}}(n) = n + 1$ for all n. For a general reference about word combinatorics, see, for instance, [2,13]. However, to highlight particular combinatorial properties of the infinite word of interest, other complexity measures such as abelian, k-abelian, cyclic, privileged, and k-binomial complexities have been introduced. See, for instance, [18,9,3,16,19]. In most cases, one considers the quotient of the language $\mathcal{L}(\mathbf{x})$ by a convenient equivalence relation $\sim$ and the corresponding complexity function therefore maps

* Manon Stipulanti is supported by the FNRS Research grant 1.B.397.20. Markus Whiteland is supported by the FNRS Research grant 1.B.466.21F. Markus Whiteland dedicates this paper to the memory of his father Alan Whiteland (1940–2021).

$n \in \mathbb{N}$ to $\#(\mathcal{L}_n(\mathbf{x})/\!\sim)$. For instance, a binary (non-periodic) word is balanced if and only if its abelian complexity is equal to the constant function 2. This paper focuses on the binomial complexity introduced in [19] and that is also the central theme of Lejeune's thesis [10].

A parallel can be drawn between the k-abelian complexity introduced by Karhumäki et al. [9] and the k-binomial complexity. In both cases, we have a series of refinements of the abelian equivalence already introduced by Erdős [5]. The fundamental difference is the following one. Let $k \geq 1$ be an integer. Two finite words u, v are *k-abelian equivalent* if, for each factor w of length at most k, we count the same number of occurrences of w in both words u and v. For *k-binomial equivalence*, we count the number of times each word w of length at most k occurs in u and v as a subword, i.e. scattered factor. Thus, in the first case, we are interested in sequences of k consecutive letters, whereas in the second case, we look at subsequences of length k extracted from a given word. We will thus make the important distinction between a *factor* of a word and a *subword*.

1.1 Binomial Coefficients and Complexity Functions

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 set of finite words (resp., non-empty finite words) over A. 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 $|w|_a$. 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

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

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

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

By convention, $\binom{u}{\varepsilon} = 1$. For more on these binomial coefficients, see, for instance, [13, Chapter 6]. Let $k \geq 1$ be an integer. Two words $u, v \in A^*$ are *k-binomially equivalent*, and we write $u \sim_k v$, if

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

Salomaa [20] introduces the *k-spectrum* of a word u which is a formal polynomial in non-commutative variables $\sum_{w \in A^{\leq k}} \binom{u}{w} w$. Thus two words are k-binomially equivalent if and only if they have the same k-spectrum. Observe that the word u is obtained as a permutation of the letters in v if and only if $u \sim_1 v$. In this case, we say that u and v are *abelian equivalent*.

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

It is immediate from the definition, that for all $k \geq 1$, $u \sim_{k+1} v$ implies $u \sim_k v$. Thus, for all n, we have the inequalities (illustrated by Fig. 1)

$$\mathsf{b}_{\mathbf{x}}^{(1)}(n) \leq \mathsf{b}_{\mathbf{x}}^{(2)}(n) \leq \cdots \leq \mathsf{b}_{\mathbf{x}}^{(k)}(n) \leq \mathsf{b}_{\mathbf{x}}^{(k+1)}(n) \leq \cdots \leq \mathsf{p}_{\mathbf{x}}(n). \tag{1}$$

1.2 Questions Addressed in This Paper

The k-binomial complexity function has been studied for particular infinite words: for $k \geq 2$, the k-binomial complexity of Sturmian words coincides with their factor complexity [19] and the same property holds true for the Tribonacci word [12]. Recently, the 2-binomial complexity of generalized Thue–Morse words was also computed [14]. The k-binomial complexity of the Thue–Morse word $\mathbf{t}$ is bounded by a constant (depending on k) [11], and more generally bounded k-binomial complexity holds for any fixed point of a prolongable Parikh-constant morphism ϕ [19], i.e. $\Psi(\phi(a)) = \Psi(\phi(b))$ for all letters a, b.

In this work, we generalize the above property of the fixed points of Parikh-constant morphisms to what we call *Parikh-collinear* morphisms ϕ: for all letters a, b, there is a rational number $r_{a,b}$ such that $\Psi(\phi(a)) = r_{a,b}\Psi(\phi(b))$. Such morphisms were characterized in [4]; see Theorem 16. In Section 3.1, we provide a new characterization of these morphisms in terms of the binomial complexity: they map all words with bounded k-binomial complexity to words with bounded $(k+1)$-binomial complexity. Finally, Corollary 18 shows that fixed points of Parikh-collinear morphisms have bounded k-binomial complexity. (See Fig. 1 for an illustration.)

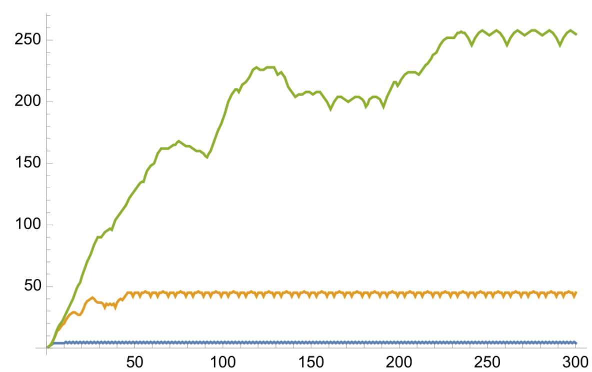

Fig. 1. The functions $\mathsf{b}_{\mathbf{x}}^{(k)}$, $k \in \{1, 2, 3\}$, where $\mathbf{x}$ is the fixed point of the morphism $0 \mapsto 000111$, $1 \mapsto 0110$. This morphism has the property of being Parikh-collinear.

For all $j \geq 1$, the exact value of $\mathsf{b}_{\mathbf{t}}^{(j)}(n)$ computed in [11] is given by

$$\mathsf{b}_{\mathbf{t}}^{(j)}(n) = \begin{cases} \mathsf{p}_{\mathbf{t}}(n) & \text{if } n \leq 2^j - 1; \\ 3 \cdot 2^j - 3, & \text{if } n \equiv 0 \pmod{2^j} \text{ and } n \geq 2^j; \\ 3 \cdot 2^j - 4, & \text{otherwise.} \end{cases} \tag{2}$$

We show in Theorem 23 that such a behavior is not specific to $\mathbf{t}$, but appears for a large class of words. Let φ be the Thue–Morse morphism. For any aperiodic binary word $\mathbf{y}$, the word $\mathbf{x} = \varphi^k(\mathbf{y})$ is such that $\mathsf{b}_{\mathbf{x}}^{(j)}(n) = \mathsf{b}_{\mathbf{t}}^{(j)}(n)$ for all $j \leq k$ and $n \geq 2^j$.

In general, not much is known about the general behavior or the properties that can be expected for the k-binomial complexity. In particular, computing the k-binomial complexity of a particular infinite word remains challenging, see, for instance, Fig. 1 to grasp the difficulty. It would also be desirable to compare in some ways k and $(k+1)$-binomial complexities of a word. For two functions $\mathsf{f}, \mathsf{g} \colon \mathbb{N} \to \mathbb{N}$, we write $\mathsf{f} \prec \mathsf{g}$ when the relation $\mathsf{f}(n) < \mathsf{g}(n)$ holds for infinitely many $n \in \mathbb{N}$. Our reflexion is driven by the following questions inspired by Lejeune's questions [10, pp. 115–117] that are natural to consider in view of (1). [1]

Question A. *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* $\mathbf{w}$*?*

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

Section 4 is about binomial properties of iterates of φ. Going further than (2), we also study the $(k+1)$- and $(k+2)$-binomial complexity of words of the form $\mathbf{x} = \varphi^k(\mathbf{y})$ with $\mathbf{y}$ aperiodic. In Section 4.1 we prove Theorem 23 mentioned above. In Section 4.2, we characterize $(k+1)$-binomial equivalence in $\mathbf{x}$ with Proposition 34. As a consequence of it, we get that $\mathsf{b}_{\mathbf{x}}^{(k)} \prec \mathsf{b}_{\mathbf{x}}^{(k+1)}$. We made these considerations because one can wonder if the factor complexity can be achieved (dismissing the trivial cases of periodic words or fixed points of Parikh-constant morphisms).

Question B. *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* $\mathbf{w}$*?*

---

[1] We choose $\prec$ because, e.g., for the period-doubling word $\mathbf{pd}$ there exist two subsequences such that $\mathsf{b}_{\mathbf{pd}}^{(2)}(n_i) = \mathsf{p}_{\mathbf{pd}}(n_i)$ and $\mathsf{b}_{\mathbf{pd}}^{(2)}(m_i) < \mathsf{p}_{\mathbf{pd}}(m_i)$ [10, Prop. 4.5.1].

Putting together results from Sections 4 and 5 we fully answer Question B: Theorem 23 and Proposition 34 give $\mathsf{b}_{\mathbf{x}}^{(1)} \prec \mathsf{b}_{\mathbf{x}}^{(2)} \prec \cdots \prec \mathsf{b}_{\mathbf{x}}^{(k-1)} \prec \mathsf{b}_{\mathbf{x}}^{(k)} \prec \mathsf{b}_{\mathbf{x}}^{(k+1)}$, while assuming that $\mathbf{y}$ above is Sturmian, we show that $\mathsf{b}_{\mathbf{x}}^{(k+2)} = \mathsf{p}_{\mathbf{x}}$. Iterates of φ applied to Sturmian words are studied (among other words) in [7]. Our construction leads to words with bounded abelian complexity. Question B is then strengthened in Section 5 where we ask for words with unbounded abelian complexity. We give a pure morphic answer when $\ell = 3$.

1.3 Preliminaries

We collect some useful results on k-binomial equivalence. First 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 $ux \sim_k vy$.

Using a classical "length-n sliding window" argument, one has the following.

Lemma 2 (Folklore). *For any binary word* $\mathbf{y}$ *over* $\{0,1\}$*, we have*

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

Lemma 3 (Cancellation property). *Let* u, v, w *be words over* A*. We have*

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

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

Theorem 4 ([21, 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 [11, Lem. 30]. This could also be proved by induction using Theorem 4 with $x = \varphi^k(0)$, $y = \varphi^k(1)$.

Theorem 5 (Ochsenschläger [15]). *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 [11, Lem. 31] will be convenient. This can alternatively be proved using Theorem 4 combined with Ochsenschläger's result.

Lemma 6 (Transfer lemma). *Let* $k \geq 1$*. Let* $u, v,'$ *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 $u \in A^*$, $e \in B^*$,

$$\binom{f(u)}{e} = \sum_{\substack{a_1,\ldots,a_\ell \in A \\ \ell \leq |e|}} \binom{u}{a_1 \cdots a_\ell} \sum_{\substack{e = e_1 \cdots e_\ell \\ e_i \in B^+}} \prod_{i=1}^{\ell} \binom{f(a_i)}{e_i}. \qquad (3)$$

We recall the following lemma that appears in [21]; it is a straightforward generalization of an observation in [20]. We give a proof for the sake of completeness.

Lemma 7. *Let $\mathcal{C}$ 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. □

Theorem 8 ([19, 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 $\binom{u}{01} \neq \binom{v}{01}$.

2 Several Answers to Question A

One can give a rather direct answer to Question A. Indeed, let $\mathbf{c}$ be the binary Champernowne word, that is, the concatenation of the binary representations of the non-negative integers: $0\,1\,10\,11\,100\,101\,110\,111\cdots$. 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 5). 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. Therefore in the rest of the section we provide more "structured" words answering Question A.

2.1 A Non-Binary Pure Morphic Answer

Let $\varphi\colon 0 \mapsto 01, 1 \mapsto 10$ be the Thue–Morse morphism over $\{0,1\}$. Consider the morphism $g\colon \{a,0,1,\alpha\}^* \to \{a,0,1,\alpha\}^*$ defined by

$$a \mapsto a0\alpha,\ 0 \mapsto \varphi(0),\ 1 \mapsto \varphi(1),\ \alpha \mapsto \alpha^2.$$

We have $\mathbf{g} = g^\omega(a) = a\prod_{j=0}^{\infty} \varphi^j(0)\alpha^{2^j}$. We show that the word $\mathbf{g}$ answers Question A:

Proposition 9. *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

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

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 5, take $u_n = \varphi^k(0)\alpha^n$ and $v_n = \varphi^k(1)\alpha^n$. Consequently $\mathsf{b}_{\mathbf{g}}^{(k)} \prec \mathsf{b}_{\mathbf{g}}^{(k+1)}$ for all $k \geq 1$. □

2.2 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 $\alpha \mapsto 1$. We have the following:

Proposition 10. *The abelian complexity of $\tau(\mathbf{g})$ is unbounded and $\mathsf{b}^{(k)}_{\tau(\mathbf{g})} \prec \mathsf{b}^{(k+1)}_{\tau(\mathbf{g})}$ 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).

To show $\mathsf{b}^{(k)}_{\tau(\mathbf{g})} \prec \mathsf{b}^{(k+1)}_{\tau(\mathbf{g})}$ 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 $\tau(v_n)$. ◻

2.3 A Binary Uniformly Recurrent Answer

We note that none of the above words are *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 $\mathcal{L}_N(\mathbf{x})$). We recall a particular construction from Grillenberger [8] for uniformly recurrent words having arbitrary entropy. The word of interest is constructed as follows. Define $D_0 = \{0,1\}$. Assuming D_k is constructed, let u_k be the product of words of D_k in lexicographic order, assuming $0 < 1$. Define then $D_{k+1} := u_k D_k^2$. Now the sequence $(u_k)_{k\in\mathbb{N}}$ converges to a uniformly recurrent word $\mathbf{u} = 0100010101100111\cdots$.

Lemma 11. *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 3. For any word e of length $k+1$, we have

$$\begin{aligned}\binom{uu}{e} - \binom{vv}{e} &= 2\binom{u}{e} - 2\binom{v}{e} + \sum_{\substack{e=e_1e_2\\ e_i\in A^+}} \left[\binom{u}{e_1}\binom{u}{e_2} - \binom{v}{e_1}\binom{v}{e_2}\right] \\ &= 2\binom{u}{e} - 2\binom{v}{e}\end{aligned}$$

because $u \sim_k v$. Since $u \not\sim_{k+1} v$, there exists a word e of length $k+1$ such that $\binom{u}{e} \neq \binom{v}{e}$ which implies $\binom{uu}{e} \neq \binom{vv}{e}$, a contradiction.

Next we have $uv \sim_{k+1} vu$ by Theorem 4, and thus $z = u_j uv \sim_{k+1} u_j vu = w$ by Lemma 3. Similarly $z \sim_{k+2} w$ would imply $uv \sim_{k+2} vu$ and thus $u \sim_{k+1} v$ by Theorem 4, a contradiction. The claim follows. ◻

Theorem 12. *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.

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. □

3 An Interlude: Parikh-Collinear Morphisms

Definition 13 (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* $\Psi(f(b)) = r_{a,b}\Psi(f(a))$*.*

In this section, we show that, given an infinite fixed point of a prolongable Parikh-collinear morphism, its k-binomial complexity is bounded for each k.

Remark 14. 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 15. The morphism f defined by $0 \mapsto 000111$; $1 \mapsto 0110$ is Parikh-collinear since $\Psi(f(1)) = \frac{2}{3}\Psi(f(0))$. The first three binomial complexities are graphed in Fig. 1.

Theorem 16 ([4, Thm. 11]). *A morphism* $f\colon A^* \to B^*$ *maps all infinite words to words with bounded abelian complexity if and only if it is Parikh-collinear.*

We extend the above theorem to the following one, where 0-*binomial complexity* has to be understood as the "equal length" equivalence relation.

Theorem 17. *A morphism* $f\colon A^* \to B^*$ *maps, for all* $k \geq 0$*, all words with bounded* k*-binomial complexity to words with bounded* $(k+1)$*-binomial complexity if and only if it is Parikh-collinear.*

Before proving this result in Section 3.2, let us mention a straightforward consequence, which generalizes [19, 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 15 is Parikh-collinear but not Parikh-constant.

Corollary 18. *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* $\mathsf{b}_{\mathbf{z}}^{(k)}(n) \leq C_{\mathbf{z},k}$ *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 16 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 19. We cannot relax the assumption on the rank of the adjacency matrix M_f. 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 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 0 and $5 \pm \sqrt{13}$, the latter two of which are strictly greater than 1. This means that the word has unbounded abelian complexity. Indeed, this follows from a deep result of Adamczewski [1, Thm. 1(ii)] combined with an observation in [17, Lem. 2.2]. Hence the word has unbounded $\mathsf{b}^{(k)}$ for all $k \geq 1$.

3.1 A Characterization of Parikh-Collinear Morphisms

To prove Theorem 17, we give further characterizations of Parikh-collinear morphisms. To this end, we require the following lemma where is defined a map g_e which is constant on any abelian equivalence class. Such a map is natural to consider in view of (3).

Lemma 20. *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 $i \in \{1, \dots, n\}$. For all $\alpha \in A$ and $\beta \in B$, define $I(\alpha, \beta) := \{i \in \{1, \dots, n\} \mid a_i = \alpha \text{ and } e_i = \beta\}$. 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

$$g_e(w) = \prod_{\substack{\alpha\in A\\ \beta\in B}} \prod_{i\in I(\alpha,\beta)} \binom{f(\alpha)}{\beta} = \prod_{\substack{\alpha\in A\\ \beta\in B}} \prod_{i\in I(\alpha,\beta)} r_\alpha \binom{f(0)}{\beta}$$

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

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. □

Proposition 21. *Let $f\colon A^* \to B^*$ be a morphism. The following are equivalent.*

(i) For all $k \geq 2$ and $u, v \in A^*$, $u \sim_{k-1} v$ implies $f(u) \sim_k f(v)$.
(ii) There exists an integer $k \geq 2$ such that for all $u, v \in A^*$, $u \sim_{k-1} v$ implies $f(u) \sim_k f(v)$.
(iii) For all $u, v \in A^*$, $u \sim_1 v$ implies $f(u) \sim_2 f(v)$.
(iv) f is Parikh-collinear.

Proof. Clearly (i) implies (ii). We show that (ii) implies (iii). There is nothing to prove if (ii) holds for $k = 2$, so assume that $k \geq 3$. We show that f also satisfies (ii) with $k - 1$ instead of k, and hence, by repeating the argument, f satisfies (ii) 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 4, 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 4, contrary to what was assumed.

Assuming (iii), we show that (iv) 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 (3),

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

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 (iv) implies (i). So let f be a Parikh-collinear morphism and $u \sim_{k-1} v$ with $k \ge 2$. We apply (3): for any word $e \in B^*$, we have

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

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

$$\sum_{a_1,\dots,a_k \in A} \binom{u}{a_1\cdots a_k} \prod_{i=1}^{k} \binom{f(a_i)}{e_i} = \sum_{a_1,\dots,a_k \in A} \binom{v}{a_1\cdots a_k} \prod_{i=1}^{k} \binom{f(a_i)}{e_i}. \quad (4)$$

Observe here that $\prod_{i=1}^{k} \binom{f(a_i)}{e_i} = g_e(a_1\cdots a_k)$ as defined in Lemma 20. Let $\mathcal{C}$ be an abelian equivalence class of a word in A^k. As the Parikh vector is constant on $\mathcal{C}$, let us write $\Psi(w) = \Psi_{\mathcal{C}}$ for all words $w \in \mathcal{C}$. We now have

$$\sum_{w \in A^k} \binom{u}{w} g_e(w) = \sum_{\mathcal{C}} \sum_{w \in \mathcal{C}} \binom{u}{w} g_e(w) \quad (5)$$

where $\mathcal{C}$ in the outer sum ranges over the abelian equivalence classes of words in A^k. By Lemma 20, $g_e(\cdot)$ is constant on $\mathcal{C}$, so write $g_e(w) = g_{\mathcal{C},e}$ for all words $w \in \mathcal{C}$. Then we obtain

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

by Lemma 7, where $\Psi_{\mathcal{C}} = (m_{\mathcal{C},a})_{a\in A}$. One obtains the same formula by replacing u with v, and equality indeed holds in (4) as $|u|_a = |v|_a$ for each letter $a \in A$. This concludes the proof. ☐

3.2 Proof of Theorem 17

The next result essentially appears in the proof of [4, Thm. 12]. We give a proof here for the sake of completeness.

Lemma 22. *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

$$\begin{aligned} c &\geq \big||f(u')| - |f(v)| + |f(v')|\big| \\ &\geq |f(u')| - \big||f(v')| - |f(v)|\big| = |f(u')| - |\langle M_f(\Psi(v') - \Psi(v)), \mathbf{1}\rangle|, \end{aligned}$$

where $\langle\cdot\,,\,\cdot\rangle$ denotes the inner product of vectors, and $\mathbf{1}$ is the all-ones-vector. Recall that $\mathbf{x}$ has bounded abelian complexity if and only if it is C-balanced for some C [17]. 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: A morphism $f\colon A^* \to B^*$ maps, for all $k \geq 0$, all words with bounded k-binomial complexity to words with bounded $(k+1)$-binomial complexity if and only if it is Parikh-collinear.

Proof (of Theorem 17).
If $f\colon A \to B^*$ maps all words with bounded 0-binomial complexity (i.e., all words) to words with bounded 1-binomial complexity, then f is Parikh-collinear by Theorem 16.

Assume thus that f is Parikh-collinear. Theorem 16 implies that f maps all words (i.e., all words with bounded 0-binomial complexity) to words with bounded 1-binomial complexity. Let then $k \geq 1$ 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 k-binomial equivalence class of u by Proposition 21.

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 22 (by assumption, $\mathbf{x}$ has bounded k-binomial complexity and thus, $\mathbf{x}$ has bounded abelian complexity by (1)) 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. ☐

4 Binomial Properties of the Thue–Morse Morphism

In this section, we consider binomial complexities of iterates of the Thue-Morse morphism φ on aperiodic binary words. Repeated application of Theorem 17 shows that, for any $k \geq 1$ and any binary word $\mathbf{y}$, the k-binomial complexity function of the word $\varphi^k(\mathbf{y})$ is bounded. In Section 4.1 we make this result much more precise:

Theorem 23. *Let j, k be integers with $1 \leq j \leq k$ and let $\mathbf{y}$ be an aperiodic binary word. Let $\mathbf{x} = \varphi^k(\mathbf{y})$. For all $n \geq 2^j$, we have $\mathsf{b}_{\mathbf{x}}^{(j)}(n) = \mathsf{b}_{\mathbf{t}}^{(j)}(n)$ which is given by (2) and, for $n < 2^j$, $\mathsf{b}_{\mathbf{x}}^{(j)}(n) = \mathsf{p}_{\mathbf{x}}(n)$.*

This is a generalization of [11, Thm. 6], which says that, for all $j \geq 1$, the j-binomial complexity of the Thue–Morse word $\mathbf{t}$ is given by (2). It implies that $\mathsf{b}_{\mathbf{x}}^{(1)} \prec \mathsf{b}_{\mathbf{x}}^{(2)} \prec \cdots \prec \mathsf{b}_{\mathbf{x}}^{(k)}$. The aim of Section 4.2 is to go one step further and get $\mathsf{b}_{\mathbf{x}}^{(k)} \prec \mathsf{b}_{\mathbf{x}}^{(k+1)}$. To do so, we characterize k-binomial and $(k+1)$-binomial equivalence among factors of $\mathbf{x}$ (Theorem 29 and Proposition 34).

4.1 The First k Binomial Complexities

Before proving Theorem 23, we require the following general lemma about aperiodic binary words.

Lemma 24. *Let $\mathbf{z}$ be an aperiodic binary word. Then for all $n \geq 2$ we have $\mathcal{L}_n(\mathbf{z}) \cap L \neq \emptyset$ for each $L \in \{0A^*1, 1A^*0, 0A^*0 \cup 1A^*1\}$. Furthermore, for all $n \geq 2$ and $a \in \{0,1\}$, we have*

$$(\mathcal{L}_n(\mathbf{z}) \cap aA^*a) \cup (\mathcal{L}_{n+1}(\mathbf{z}) \cap \overline{a}A^*\overline{a}) \neq \emptyset.$$

Proof. If $\mathcal{L}_n(\mathbf{z}) \cap aA^*\overline{a} = \emptyset$ for some n, then $\mathbf{z}$ is ultimately periodic: for all $m \geq 0$, if $\mathbf{z}_m = a$, then $\mathbf{z}_{m+kn-1} = a$ for all $k \geq 1$. Consequently, for each $0 \leq m \leq n-1$, the word $(\mathbf{z}_{m+kn-1})_{k\geq 1}$ is either 0^ω or $0^\ell 1^\omega$ for some $\ell \geq 0$. It follows that $\mathbf{z}$ is eventually periodic. Also, since $\mathbf{z}$ is aperiodic, there is a right special factor of length $n-1 \geq 1$ of the form av or $\overline{a}v$, in which case $ava \in \mathcal{L}_n(\mathbf{z}) \cap aA^*a \neq \emptyset$ (resp., $\overline{a}v\overline{a} \in \mathcal{L}_n(\mathbf{z}) \cap \overline{a}A^*\overline{a} \neq \emptyset$).

Let us prove the second part of the statement. Assume for a contradiction that $\mathcal{L}_n(\mathbf{z}) \cap 0A^*0 = \emptyset = \mathcal{L}_{n+1}(\mathbf{z}) \cap 1A^*1$ for some $n \geq 2$. Consider a factor of the form $z = 1z_1 \cdots z_{n-1}z_n \cdots z_{2n-1}$ of length $2n$. Since $\mathcal{L}_{n+1}(\mathbf{z}) \cap 1A^*1 = \emptyset$, we have $z_n = 0$. Further, since $\mathcal{L}_n(\mathbf{z}) \cap 0A^*0 = \emptyset$, we have $z_1 = 1$. Repeating the argument we have $z_{n+i-1} = 0$ and $z_i = 1$ for all $i \geq 1$ which is a contradiction when $i = 1$ and $i = n$. □

Definition 25. *Let $j \geq 0$. 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[2] *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, whence the letter a is uniquely determined once $p \neq \varepsilon$.

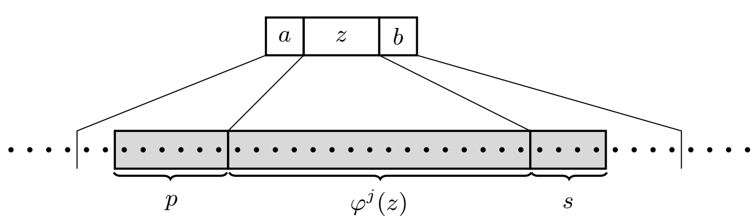

Fig. 2. A φ^j-factorization and its φ^j-ancestor.

[2] We warn the reader that the term φ-factorization has a different meaning in [11]. Our φ^j-factorization corresponds to their "factorization of order j".

Proof (of Theorem 23). Let $j \in \{1, \dots, k\}$. Notice all factors of length at most $2^j - 1$ of $\mathbf{x} = \varphi^k(\mathbf{y})$ occur already in the Thue–Morse word $\mathbf{t}$: such factors appear in factors of the form $\varphi^j(ab)$, $ab \in \mathcal{L}(\mathbf{y})$. Since $\varphi^j(ab)$ appears in the Thue–Morse word for all $a, b \in \{0, 1\}$, it follows from (2) that all such words are pairwise j-binomially non-equivalent. Hence we have shown that $\mathsf{b}_{\mathbf{x}}^{(j)}(n) = \mathsf{p}_{\mathbf{x}}(n)$ for $n \leq 2^j - 1$.

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 23 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., [11, Prop. 33]). 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 5.

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 $u \sim_j v$. By the previous lemma, $\mathbf{z}$ 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 the previous lemma says 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 by the previous lemma, $\mathcal{L}_{|u|'+2}\,\mathbf{z} \cap 0A^*0 \neq \emptyset$ and, further, $\mathcal{L}_{|u|'+2\pm 1}\,\mathbf{z} \cap 0A^*0 \neq \emptyset$. To conclude with the proof, 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 5. By the Transfer Lemma (Lemma 6), $\varphi^j(v'1) \sim_j \varphi^{j-1}(1)\varphi^j(v')\varphi^{j-1}(0)$. Consequently

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

where $p\varphi^{j-1}(1)$ is a suffix of $\varphi^j(0)$ and $\varphi(0)^{j-1}s$ 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 5 and Lemma 6,

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

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 5 and

Lemma 6, $u \sim_j p\varphi^{j-1}(1)\varphi^j(v')s' =: v$ 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 5 and Lemma 6,

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

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

□

Remark 26. 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 24. 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 24) 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 24) gives $c\overline{c}\varphi(y)c\overline{c}$. Consequently $\mathbf{z}$ contains a factor in aA^*a of length n.

Applying this observation to $\mathbf{z}$ when $j < k$ in the above proof shows that $\mathcal{L}_n(\mathbf{z}) \cap 1A^*1 \neq \emptyset$ for all $n \geq 2$, and thus some of the arguments are unnecessary in the case $j < k$.

4.2 The $(k+1)$-Binomial Complexity

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 27 ([11, Def. 43]). *Let $j \geq 1$. As usual, we let $\overline{\cdot}$ denote the complementation morphism defined by $\overline{a} = 1 - a$, for $a \in \{0,1\}$. 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) $(p_1, s_1) = (s_2, p_2) = (\varphi^{j-1}(a), \varphi^{j-1}(\bar{a}))$;
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 [11, 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 says that in such a case the φ^j-factorizations are related.

Lemma 28. *Let $j \geq 1$. Let u be a factor of $\varphi^j(\mathbf{y})$ such that $|u| \geq 2^j - 1$ with a φ^j-factorization of the form $(p, \varphi^j(z), s)$ and $z_0 z z_{n+1}$ being the corresponding φ^j-ancestor (where according to Definition 25 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 $z_0 z z_{n+1}$ contains both letters 0 and 1. Otherwise stated, the φ^j-factorization is not unique if and only if u is a factor of $\varphi^{j-1}(m)$ with $m \in (01)^* \cup (10)^* \cup 1(01)^* \cup 0(10)^*$. Moreover, when the φ^j-factorization is not unique, i.e. if there is another φ^j-factorization $(p', \varphi^j(z'), s')$, then $(p, s) \equiv_j (p', s')$.*

Proof. If $|u| \geq 2^j - 1$, u contains at least a factor $\varphi^{j-1}(a)$ 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 $z_i z_{i+1} = 01$ (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 $z_i z_{i+1} = 01$ 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

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

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 by the previous part. 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 $(p, s) \equiv_j (p', s')$:
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 [11, Thm. 48]. Indeed, the lemmas 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 [11, Thm. 48] is stated for $j \geq 3$. The case $j = 1$ is trivial. The case $j = 2$ is obtained by looking closely at the proof of [11, Thm. 34].

Theorem 29. *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 $\mathbf{x}$. We require some lemmas. A straightforward consequence of (3) together with the identities $\sum_{x\in A^\ell}\binom{u}{x} = \binom{|u|}{\ell}$, $\ell \geq 1$, is the following observation.

Lemma 30. *Let* $\varphi\colon 0 \mapsto 01, 1 \mapsto 10$ *be the Thue–Morse morphism. 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. For example, $\binom{\varphi(a)}{011} = 0 = \binom{\varphi(a)}{11}$ 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. □

Lemma 31. *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 30. Then assume $k \geq 2$. We encourage the reader to refer to [11] for details that would be too long to reproduce here. From [11, 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 [11, 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 with 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 [11, 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 $\binom{u}{01} - \binom{v}{01} = \binom{v}{10} - \binom{u}{10}$ suffices to conclude. □

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

Definition 32. *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 $\mathcal{S}_{a,b}(n) = \mathcal{L}_{n+|ab|}(\mathbf{y}) \cap aA^*b$. 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 33. *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 30

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

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 30. 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 $u = 1(01)^n = (10)^n1 = v \in L$, 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 34. *Let u and v be factors of length at least $2^k - 1$ of $\mathbf{x}$ 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 more than one φ^k-factorization 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 21 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 29. 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 27:

(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 31. 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 33 (otherwise $a\varphi(z) =$

$\varphi(z')a$ and thus $u = v$ contrary to the assumption). Lemma 31 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 33 $a\varphi(z)\overline{a} \not\sim_2 \varphi(z')$ (otherwise $z = \overline{a}^\ell$ and $z' = a^{\ell+1}$, implying that $u = v$, a contradiction). This is again a contradiction by Lemma 31.

(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 23 and Proposition 34 have the following corollary:

Corollary 35. *Let $\mathbf{x} = \varphi^k(\mathbf{y})$, where $\mathbf{y}$ is an arbitrary aperiodic binary word. We have*

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

Proof. Recall that $\mathbf{y}$ contains arbitrarily long factors of the form $\overline{a}za$, $a \in \{0, 1\}$. Therefore $\mathbf{x}$ contains the k-binomially equivalent (by Lemma 6) factors $\varphi^{k-1}(a)\varphi^k(z)$ and $\varphi^k(z)\varphi^{k-1}(a)$. However, by Proposition 34 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 23. □

5 Answer to Question B and Beyond

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 8.

Remark 36. We notice that $\mathsf{b}_{\mathbf{x}}^{(1)} = \mathsf{p}_{\mathbf{x}}$ cannot be attained for an aperiodic word $\mathbf{x}$ (indeed, there must exist a factor ava, 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 $\{a_1, \ldots, a_m\}$ 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 φ and we prove the following result.

Theorem 37. *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*

$$\mathsf{b}_{\mathbf{s}_k}^{(1)} \prec \mathsf{b}_{\mathbf{s}_k}^{(2)} \prec \cdots \prec \mathsf{b}_{\mathbf{s}_k}^{(k+1)} \prec \mathsf{b}_{\mathbf{s}_k}^{(k+2)} = \mathsf{p}_{\mathbf{s}_k}.$$

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 17 (because $\mathbf{s}$ has bounded abelian complexity), and thus $\mathsf{b}_{\mathbf{s}_k}^{(k+1)} \prec \mathsf{p}_{\mathbf{s}_k}$. By Corollary 35, we need only to show that $\mathsf{b}_{\mathbf{s}_k}^{(k+2)} = \mathsf{p}_{\mathbf{s}_k}$.

Let u and v be distinct factors of $\mathbf{s}_k$. Assume they are $(k+2)$-binomially equivalent. By Proposition 34, 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 8. But then Lemma 31 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 38. In the above proof, since $\mathbf{s}$ is Sturmian, Theorem 8 says distinct factors are not 2-binomially equivalent. This means that Theorem 37 applies to and only to aperiodic words $\mathbf{s}$ such that $\mathsf{b}_{\mathbf{s}}^{(2)} = \mathsf{p}_{\mathbf{s}}$. The "only if"-part of the statement follows by a repeated application of Proposition 21 on the Thue–Morse morphism together with the fact that the morphism is injective.

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 $\mathsf{b}_{\mathbf{w}}^{(1)}$ is unbounded and*

$$\mathsf{b}_{\mathbf{w}}^{(1)} \prec \mathsf{b}_{\mathbf{w}}^{(2)} \prec \cdots \prec \mathsf{b}_{\mathbf{w}}^{(\ell-1)} \prec \mathsf{b}_{\mathbf{w}}^{(\ell)} = \mathsf{p}_{\mathbf{w}}?$$

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

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

Theorem 39. *The word $\mathbf{h} = 0112122122212222122222\cdots$ fixed point of the morphism $0 \mapsto 01$, $1 \mapsto 12$, and $2 \mapsto 2$ is such that its abelian complexity $\mathsf{b}_{\mathbf{h}}^{(1)}$ is unbounded and $\mathsf{b}_{\mathbf{h}}^{(1)} \prec \mathsf{b}_{\mathbf{h}}^{(2)} \prec \mathsf{b}_{\mathbf{h}}^{(3)} = \mathsf{p}_{\mathbf{h}}$.*

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

Proposition 40. *The abelian complexity $\mathsf{b}_{\mathbf{h}}^{(1)}$ of $\mathbf{h}$ is unbounded and $\mathsf{b}_{\mathbf{h}}^{(1)}(n) < \mathsf{b}_{\mathbf{h}}^{(2)}(n) < \mathsf{p}_{\mathbf{h}}(n)$ for all $n \geq 6$.*

Proof. We claim that $\mathsf{b}_{\mathbf{h}}^{(1)}$ 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, 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 [6] 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 41. *We have* $\mathsf{b}^{(3)}_{\mathbf{h}} = \mathsf{p}_{\mathbf{h}}$.

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 $u_i = 2^{x_i}12^{a_i}12^{a_i+1}1 \cdots 12^{a_i+t}12^{y_i}$ 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

$$\binom{u_i}{121} = \binom{v_i}{121}, \tag{8}$$

where $v_i = 12^{a_i}12^{a_i+1}1 \cdots 12^{a_i+t}1$ 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 (8) gives the claim. □

6 Concluding Remarks

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 $\mathsf{b}^{(m)}_{\mathbf{w}} = \mathsf{p}_{\mathbf{w}}$?

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. We can prove that the period doubling word

$\mathbf{pd} = 01000101010001\cdots$, fixed point of $\sigma : 0 \mapsto 01,\ 1 \mapsto 00$, has the following properties [10]. Its abelian complexity $\mathsf{b}^{(1)}_{\mathbf{pd}}$ is unbounded. For the 2-binomial complexity, we can show that $\mathsf{b}^{(2)}_{\mathbf{pd}}(2^n) = \mathsf{p}_{\mathbf{pd}}(2^n)$ for all n, but $\mathsf{b}^{(2)}_{\mathbf{pd}}(n) < \mathsf{p}_{\mathbf{pd}}(n)$ for all $n \neq 2^m$. Otherwise stated, $\mathsf{b}^{(1)}_{\mathbf{pd}} \prec \mathsf{b}^{(2)}_{\mathbf{pd}} \prec \mathsf{p}_{\mathbf{pd}}$. Computer experiments suggest that $\mathsf{b}^{(3)}_{\mathbf{pd}} \prec \mathsf{b}^{(4)}_{\mathbf{pd}} = \mathsf{p}_{\mathbf{pd}}$.

Proposition 42. *Let* $\mathbf{w}$ *be the fixed point of an injective morphism* f *such that* M_f *is invertible. If there exist two distinct factors* u *and* v *of the same length such that* $u \sim_k v$*, then* $\mathsf{b}^{(k)}_{\mathbf{w}} \prec \mathsf{p}_{\mathbf{w}}$.

Proof. One can define an extended Parikh vector $\Psi_k(u)$ of size $|A| + |A|^2 + \cdots + |A|^k$ encoding the binomial coefficients for all subwords of length at most k. As in [12, Lemma 9], an extended adjacency matrix M'_f can be defined accordingly and it satisfies $M'_f \Psi_k(u) = \Psi_k(f(u))$. It can be shown that this matrix is block-triangular and the square blocks on the main diagonal are the Kronecker products of i copies of M_f: M_f, $M_f \otimes M_f$, $\ldots$, $M_f \otimes \cdots \otimes M_f$, for $i = 1, \ldots, k$. Since M_f is invertible, M'_f is also invertible (its determinant is a power of $\det(M_f)$). Using this fact, observe that $u \sim_k v$ if and only if $f(u) \sim_k f(v)$. So we have found infinitely many pairwise distinct factors $f^i(u)$ and $f^j(u)$ of the same length that are k-binomially equivalent. $\square$

References

1. Adamczewski, B.: Symbolic discrepancy and self-similar dynamics. Annales de l'Institut Fourier **54**(7), 2201–2234 (2004). https://doi.org/10.5802/aif.2079
2. Allouche, J.P., Shallit, J.: Automatic sequences: Theory, applications, generalizations. Cambridge University Press, Cambridge (2003). https://doi.org/10.1017/CBO9780511546563
3. Cassaigne, J., Fici, G., Sciortino, M., Zamboni, L.Q.: Cyclic complexity of words. J. Comb. Theory, Ser. A **145**, 36–56 (2017). https://doi.org/10.1016/j.jcta.2016.07.002
4. Cassaigne, J., Richomme, G., Saari, K., Zamboni, L.Q.: Avoiding Abelian powers in binary words with bounded Abelian complexity. International Journal of Foundations of Computer Science **22**(4), 905–920 (2011). https://doi.org/10.1142/S0129054111008489
5. Erdős, P.: Some unsolved problems. Mich. Math. J. **4**, 291–300 (1958)
6. Fossé, S., Richomme, G.: Some characterizations of Parikh matrix equivalent binary words. Inf. Process. Lett. **92**(2), 77–82 (2004)
7. Frid, A.: Applying a uniform marked morphism to a word. Discrete Mathematics & Theoretical Computer Science. DMTCS. An Electronic Journal **3**(3), 125–139 (1999). https://doi.org/10.46298/dmtcs.255
8. Grillenberger, C.: Constructions of strictly ergodic systems. I. Given entropy. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **25**, 323–334 (1973). https://doi.org/10.1007/BF00537161
9. Karhumäki, J., Saarela, A., Zamboni, L.Q.: On a generalization of abelian equivalence and complexity of infinite words. J. Comb. Theory, Ser. A **120**(8), 2189–2206 (2013). https://doi.org/10.1016/j.jcta.2013.08.008

10. Lejeune, M.: On the k-binomial equivalence of finite words and k-binomial complexity of infinite words. Ph.D. thesis, Univ. of Liège (2021), http://hdl.handle.net/2268/259266
11. Lejeune, M., Leroy, J., Rigo, M.: Computing the k-binomial complexity of the Thue-Morse word. J. Comb. Theory, Ser. A **176**, 44 (2020). https://doi.org/10.1016/j.jcta.2020.105284, id/No 105284
12. Lejeune, M., Rigo, M., Rosenfeld, M.: Templates for the k-binomial complexity of the Tribonacci word. Adv. Appl. Math. **112**, 26 (2020). https://doi.org/10.1016/j.aam.2019.101947, id/No 101947
13. Lothaire, M.: Combinatorics on Words. Cambridge Mathematical Library. Cambridge University Press (1997). https://doi.org/10.1017/CBO9780511566097
14. LÃij, X.T., Chen, J., Wen, Z.X., Wu, W.: On the 2-binomial complexity of the generalized Thue-Morse words (2021)
15. Ochsenschläger, P.: Binomialkoeffizienten und shuffle-zahlen. Technischer bericht, Fachbereicht Informatik, T.H. Darmstadt (1981)
16. PeltomÃďki, J.: Introducing privileged words: Privileged complexity of Sturmian words. Theoretical Computer Science **500**, 57–67 (2013). https://doi.org/https://doi.org/10.1016/j.tcs.2013.05.028
17. Richomme, G., Saari, K., Zamboni, L.Q.: Abelian complexity of minimal subshifts. Journal of the London Mathematical Society **83**(1), 79–95 (2011). https://doi.org/10.1112/jlms/jdq063
18. Rigo, M.: Relations on words. Indag. Math., New Ser. **28**(1), 183–204 (2017). https://doi.org/10.1016/j.indag.2016.11.018
19. Rigo, M., Salimov, P.: Another generalization of abelian equivalence: binomial complexity of infinite words. Theoretical Computer Science **601**, 47–57 (2015). https://doi.org/10.1016/j.tcs.2015.07.025
20. Salomaa, A.: Counting (scattered) subwords. Bull. Eur. Assoc. Theor. Comput. Sci. EATCS **81**, 165–179 (2003)
21. Whiteland, M.A.: Equations over the k-binomial monoids. In: WORDS 2021, Lecture Notes in Computer Science, vol. 12847, pp. 185–197. Springer, Cham. (2021). https://doi.org/10.1007/978-3-030-85088-3_16