On Occurrence-Preserving Morphisms

Kaisei Kishi
Department of Information Science and Technology, Kyushu University, Fukuoka, Japan

Peaker Guo
M&D Data Science Center, Institute of Integrated Research, Institute of Science Tokyo, Japan

Cristian Urbina
Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland
Center for Biotechnology and Bioengineering (CeBiB), Santiago, Chile

Hideo Bannai
M&D Data Science Center, Institute of Integrated Research, Institute of Science Tokyo, Japan

Abstract

A *morphism* is a mapping that transforms words through letter-wise substitution, where each symbol is consistently replaced by a fixed word. In the field of combinatorics on words, one topic that has attracted considerable attention is the characterization of morphisms that preserve specific properties, such as overlap-freeness, square-freeness, lexicographic order, and primitivity. Continuing this direction, we initiate the study on *occurrence-preserving morphisms*, which address the following fundamental question: given a morphism ϕ, two words u and v, and $k \geq 1$, under what conditions does the number of occurrences of u in v equal the number of occurrences of $\phi^k(u)$ in $\phi^k(v)$? To answer this question, we introduce the notion of *interference-free morphisms*, examine their properties, and uncover a connection to *recognizable morphisms*. We then present a precise characterization of occurrence-preserving morphisms in terms of interference-freeness. As applications of our characterization, we first show that there exists a bijection between the starting positions of the occurrences of u in v and those of $\phi^k(u)$ in $\phi^k(v)$. We then apply the characterization to the Fibonacci and Thue-Morse words to identify their *minimal unique substrings (MUSs)*. Finally, we exploit the connection between MUSs and *net occurrences* to simplify existing proofs on net occurrences in these words.

2012 ACM Subject Classification Mathematics of computing → Combinatorics on words

Keywords and phrases Property-preserving morphisms, interference-free morphisms, recognizable morphisms, injective morphisms, Fibonacci words, Thue-Morse words, minimal unique substrings (MUSs), net occurrences

Digital Object Identifier 10.4230/LIPIcs.CPM.2026.24

Related Version *Full Version*: `https://arxiv.org/abs/2605.18034`

Funding *Cristian Urbina*: Polish National Science Center, grant no. 2022/46/E/ST6/00463; Basal Funds FB0001 and AFB240001, ANID, Chile; and FONDECYT Project 1-230755, ANID, Chile. *Hideo Bannai*: JSPS KAKENHI Grant Number JP24K02899.

1 Introduction

A morphism is a structure-preserving mapping, where the preserved structure is concatenation in the context of combinatorics on words. More precisely, it is a mapping that transforms words through letter-wise substitution, with each symbol consistently replaced by a fixed image word. Morphisms are fundamental objects that have been used to define infinite sequences and generate repetitive patterns [27, 33]. Notable classes of morphisms include *injective morphisms*, which play an important role in coding theory, and *recognizable morphisms*, which have been extensively studied in dynamical systems and formal language theory [1]. More recently, the study of morphisms in relation to repetitiveness measures has gained considerable attention [6, 9–11].

A recurring theme in the literature on morphisms is identifying properties that remain preserved under repeated applications of a morphism. Such studies are particularly useful for generating infinite words that guarantee desired combinatorial properties. For example, Berstel and Séébold [3] showed that a morphism h maps overlap-free words to overlap-free words if and only if $h(\mathtt{abbabaab})$ is overlap-free. Subsequent work on overlap-free morphisms includes [31, 32]. Other morphism-preserved properties have also been studied, such as lexicographic order [30], palindromic richness [8], Abelian power-freeness [7, 20], square-freeness [16], primitivity [15, 17], BWT runs [10], and Sturmian words [4]. Continuing this line of research, we initiate the study of *occurrence-preserving morphisms*, which address the following fundamental question: given a morphism ϕ, two words u and v, and $k \geq 1$,

under what conditions does the number of occurrences of u in v equal the number of occurrences of $\phi^k(u)$ in $\phi^k(v)$?

Understanding this question has direct applications in proving properties that are constrained by the number of occurrences of factors (substrings), such as *minimal unique substrings (MUSs)* and *net occurrences*. A MUS is a unique substring whose every proper substring is repeated, whereas a net occurrence is an occurrence of a repeated substring whose every proper super-string is unique. Both notions have been extensively studied for their combinatorial properties and efficient algorithms: see, for example, [18, 22, 24, 25] for MUSs, and [12, 14, 19, 21, 23, 29] for net occurrences.

Our results. In this work, we make three main contributions. First, to answer the above question, we introduce the notion of *interference-free morphisms*, analyze their properties, and uncover a connection to recognizable morphisms (Theorem 19). Second, we provide a precise characterization of occurrence-preserving morphisms in terms of interference-freeness (Theorem 22). Third, as an application, we apply this characterization to the Fibonacci and Thue-Morse words to identify their *minimal unique substrings (MUSs)* (Theorem 32 and Theorem 34). We further exploit the connection between MUSs and *net occurrences* [23] to simplify existing proofs on net occurrences in these words [13]. In the process, we establish new properties of these morphisms and words that may have independent interest.

2 Preliminaries

Basics. Let Σ be an ordered alphabet. We assume $\Sigma = \{\mathtt{a}, \mathtt{b}\}$ when $|\Sigma| = 2$. A *word (or string)* is an element of Σ^*. The length of a word w is denoted as $|w|$. Let ε denote the *empty word* of length 0. We use $w[i]$ to denote the i^{th} character of a word w. Let $u \cdot v = u\,v$ denote the *concatenation* of two words, u and v. A *factor* (or *substring*) of a word w of length n, starting at position i and ending at position j, is written as $w[i \dots j]$. A factor $w[1 \dots j]$ is called a *prefix* of w, while $w[i \dots n]$ is called a *suffix* of w. A factor u of w is a *proper* factor if $u \neq w$. For two words u and w, let $\mathrm{occ}_w(u) = \{i \mid w[i \dots i + |u| - 1] = u\}$ be the set of (starting positions of) *occurrences* of u in w, and let $\#\mathrm{occ}_w(u) = |\,\mathrm{occ}_w(u)|$. When convenient, we identify an occurrence $i \in \mathrm{occ}_w(u)$ with its corresponding interval $[i, i + |u| - 1]$ in w. For convenience, we treat the empty word ε as occurring $|w| + 1$ times in w: before position 1, between positions i and $i + 1$ for $1 \leq i \leq |w|$, and after position $|w|$. We represent the i^{th} such occurrence by the interval $[i, i - 1]$ for $1 \leq i \leq |w| + 1$. For a word w, let $w^R = w[|w|] \cdots w[1]$ denote the reverse of w. For a non-empty word w, a sequence of non-empty words $(x_k)_{k=1}^m$ is referred to as a *factorization* of w if $w = x_1 \cdots x_m$. For a word w, a word of the form $w[i \dots |w|] \cdot w[1 \dots i - 1]$, for some $1 \leq i \leq |w|$, is called a *rotation* of w. Let $\mathcal{R}(w)$ denote the multiset of all $|w|$ rotations of w. For a word w with $|w| \geq 2$, let $w^{\lhd} = w[1 \dots |w| - 1]$ denote its longest proper prefix.

Morphisms. Let Σ and Γ be two alphabets. A *morphism* is a map $\phi : \Sigma^* \to \Gamma^*$ such that $\phi(uv) = \phi(u)\ \phi(v)$ for all $u, v \in \Sigma^*$. For each $c \in \Sigma$, $\phi(c)$ is called an *image*[1] of ϕ, and let $\mathsf{Im}(\phi) = \{\phi(c) \mid c \in \Sigma\}$ be the set of images of ϕ. Letting $\Sigma = \{\alpha_1, \ldots, \alpha_\sigma\}$, ϕ can be equivalently specified by $\alpha_1 \mapsto \phi(\alpha_1), \ldots, \alpha_\sigma \mapsto \phi(\alpha_\sigma)$. Let $\phi^0(u) = u$ and let $\phi^i(u) = \phi(\phi^{i-1}(u))$ for each $i \geq 1$. Further, ϕ is *non-erasing* if $\phi(a) \neq \varepsilon$ for all $a \in \Sigma$; ϕ is *injective*[2] if $\phi(u) \neq \phi(v)$ for all $u \neq v \in \Sigma^*$; ϕ is *ℓ-uniform* if $|\phi(a)| = \ell$ for all $a \in \Sigma$. For a binary word w, let $\overline{w}$ denote the word obtained by applying the morphism $\mathtt{a} \mapsto \mathtt{b}, \mathtt{b} \mapsto \mathtt{a}$.

Fibonacci and Thue-Morse morphisms and words. Let φ denote the *Fibonacci morphism* defined by $\varphi(\mathtt{a}) = \mathtt{ab}$ and $\varphi(\mathtt{b}) = \mathtt{a}$. Let μ denote the *Thue-Morse morphism* defined by $\mu(\mathtt{a}) = \mathtt{ab}$ and $\mu(\mathtt{b}) = \mathtt{ba}$. For each $i \geq 1$, let $F_i = \varphi^{i-1}(\mathtt{b})$ be the (finite) *Fibonacci word of order i*. For each $i \geq 1$, let $\mathcal{T}_i = \mu^{i-1}(\mathtt{a})$ be the (finite) *Thue-Morse word of order i*. The Fibonacci and Thue-Morse words can also be obtained as follows: $F_1 = \mathtt{b}, F_2 = \mathtt{a}$, and $F_i = F_{i-1}\ F_{i-2}$ for each $i \geq 3$; $\mathcal{T}_1 = \mathtt{a}$ and $\mathcal{T}_i = \mathcal{T}_{i-1}\ \overline{\mathcal{T}_{i-1}}$ for each $i \geq 2$. Further, for each $i \geq 1$, let $f_i = |F_i|$, which equals the i^{th} Fibonacci number, and let $\tau_i = |\mathcal{T}_i| = 2^{i-1}$. We next review the following known properties of Fibonacci words.

▶ **Observation 1** ([13, 26]). *The following hold for F_i.*

- *Neither $\mathtt{aaa}$ nor $\mathtt{bb}$ occurs in any F_i.*
- *Let $\Delta_0 = \mathtt{ba}$, $\Delta_1 = \mathtt{ab}$, and $\Delta_i = \Delta_{(i \bmod 2)}$ for $i \geq 2$. Define $G_i = F_i[1 \ldots f_i - 2]$ for each $i \geq 3$. Then, $F_i = G_i\ \Delta_i$ and $\mathrm{occ}_{F_i}(G_{i-1}) = \{1, f_{i-2} + 1\}$ for each $i \geq 7$.*

MUSs and net occurrences. Consider a string w and a unique substring u of w. Let $[i, j]$ be the only occurrence of u in w. We say u is a *minimal unique substring (MUS)* of w if both strings $w[i+1 \ldots j]$ and $w[i \ldots j-1]$ are repeated in w. Let $\mathsf{MUS}(w)$ denote the set of MUSs of w. An occurrence $[i, j]$ in w is a *net occurrence* if the corresponding string $w[i \ldots j]$ is repeated, while both left extension $w[i-1 \ldots j]$ and right extension $w[i \ldots j+1]$ are unique. When $i = 1$, $w[i-1 \ldots j]$ is assumed to be unique; when $j = |w|$, $w[i \ldots j+1]$ is assumed to be unique. Let $\mathsf{NO}(w)$ denote the set of net occurrences in w.

3 Interference-Free Morphisms

To fully characterize occurrence-preserving morphisms, we first introduce the notion of *interference-free* morphisms and establish some of their key properties in this section. To motivate the definition, we begin with an example illustrating that, after applying a morphism ϕ to words u and v, the occurrences of u in v may fail to be preserved when certain forms of "interference" take place between $\phi(u)$ and images of ϕ.

▶ **Example 2.** In (1) and (2) of Figure 1, we have $\#\mathrm{occ}_v(u) = \#\mathrm{occ}_{\phi(v)}(\phi(u)) = 2$, and $\#\mathrm{occ}_v(u) = 2 < 3 = \#\mathrm{occ}_{\phi(v)}(\phi(u))$. A third occurrence of $\phi(u)$ (underlined in red in (2)) emerges in $\phi(v)$ because $\phi(u) = \mathtt{ab} \cdot \mathtt{a}$ and $\mathtt{a}$ is a proper prefix of $\phi(\mathtt{a}) = \mathtt{ab}$. In (3) and (4) of Figure 1, we have $\#\mathrm{occ}_v(u) = 1 < 2 = \#\mathrm{occ}_{\phi(v)}(\phi(u))$, and $\#\mathrm{occ}_v(u) = \#\mathrm{occ}_{\phi(v)}(\phi(u)) = 1$. A second occurrence of $\phi(u)$ (underlined in red in (3)) emerges in $\phi(v)$ because $\phi(u) = \mathtt{a} \cdot \mathtt{ba} \cdot \mathtt{b}$, $\mathtt{a}$ is a proper suffix of $\phi(\mathtt{b}) = \mathtt{ba}$, and $\mathtt{b}$ is a proper prefix of $\phi(\mathtt{b}) = \mathtt{ba}$.

[1] In this paper, we use the term *image* only for $\phi(c)$ with $c \in \Sigma$, and not for $\phi(u)$ with $u \in \Sigma^*$.
[2] Note that this defines a stronger notion of injectivity than simply requiring that for all $c \neq c' \in \Sigma$, $\phi(c) \neq \phi(c')$.

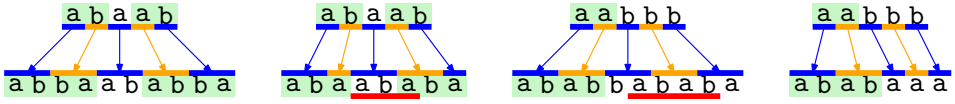

Figure 1 Illustration of Example 2. In each of (1)–(4), we show v at the top, $\phi(v)$ at the bottom, and in the middle, arrows indicate the mapping from each $v[i]$ to $\phi(v[i])$; the substring u in v, and the corresponding $\phi(u)$ in $\phi(v)$, are highlighted in green. Specifically, in (1) and (2), $u = \mathtt{ab}$ and $v = \mathtt{abaab}$, whereas in (3) and (4), $u = \mathtt{aa}$ and $v = \mathtt{aabbb}$. Moreover, $\phi : \mathtt{a} \mapsto \mathtt{ab}, \mathtt{b} \mapsto \mathtt{ba}$ in (1) and (3); $\phi : \mathtt{a} \mapsto \mathtt{ab}, \mathtt{b} \mapsto \mathtt{a}$ in (2) and (4).

To start formalizing the idea from the motivating example, we first define the following and review a key property of injective morphisms.

▶ **Definition 3** (Image Factorizations). *Let $\phi : \Sigma^* \to \Gamma^*$ be a morphism and let $w \in \Gamma^*$ be a word. We say w admits a* ϕ-image factorization *if $w = X_1 \cdots X_n$, where each $X_i \in \mathsf{Im}(\phi)$. When the morphism is clear from context, we simply say that w admits an* image factorization.

▶ **Lemma 4** ([2]). *Let $\phi : \Sigma^* \to \Gamma^*$ be a morphism and let $u \in \Sigma^*$ be a word. If ϕ is injective, then $\phi(u)$ admits a unique image factorization.*

A set X of words is a *code* if each non-empty word $w \in X^*$ admits a unique factorization into words of X. The above lemma implies that if ϕ is injective, then $\mathsf{Im}(\phi)$ forms a code.

We now give precise definitions of what was previously referred to as "interference".

▶ **Definition 5** (Interfered Image Factorizations). *Let $\phi : \Sigma^* \to \Gamma^*$ be a morphism and let $w \in \Gamma^*$ be a word. We say w admits an* interfered ϕ-image factorization *if $w = x \cdot y \cdot z$, where x is a* proper *suffix of some image in $\mathsf{Im}(\phi)$, y admits an image factorization, z is a* proper *prefix of some image in $\mathsf{Im}(\phi)$, and $x \cdot z \neq \varepsilon$. When the morphism is clear from context, we simply say that w admits an* interfered image factorization.

In Example 2, $\mathtt{ab} \cdot \mathtt{a}$ is an interfered image factorization of $\varphi(\mathtt{ab})$ and $\mathtt{a} \cdot \mathtt{ba} \cdot \mathtt{b}$ is an interfered image factorization of $\mu(\mathtt{aa})$. Beyond the cases illustrated in Example 2 and formalized in Definition 8, another type of "interference" can occur. We give an example and then formalize this notion below.

▶ **Example 6.** Let ϕ be a variant of the Thue-Morse morphism [28, A036577] defined as $\phi(\mathtt{a}) = \mathtt{abc}$, $\phi(\mathtt{b}) = \mathtt{ac}$, $\phi(\mathtt{c}) = \mathtt{b}$; let $u = \mathtt{c}$, and $v = \mathtt{ca}$. Then, $\phi(u) = \mathtt{b}$, $\phi(v) = \mathtt{babc}$, and $\#\mathrm{occ}_v(u) = 1 < 2 = \#\mathrm{occ}_{\phi(v)}(\phi(u))$.

▶ **Definition 7** (Inner Image Factor). *Let $\phi : \Sigma^* \to \Gamma^*$ be a morphism and let $w \in \Gamma^*$ be a word. We say w is an* inner ϕ-image factor *if w is a proper factor of an image $\phi(c)$ for some $c \in \Sigma$, but is neither a prefix nor a suffix of $\phi(c)$. When the morphism is clear from context, we simply say that w is an* inner image factor.

In Example 6, $\mathtt{b}$ is a proper factor of $\phi(\mathtt{a}) = \mathtt{abc}$, but is neither a prefix nor a suffix of $\mathtt{abc}$. Hence $\mathtt{b}$ is an inner image factor. Having formalized the two types of "interference", we are now ready to define interference-free morphisms.

▶ **Definition 8** (Interference-Free Morphisms). *Let $\phi : \Sigma^* \to \Gamma^*$ be an injective morphism and let $\mathcal{L} \subseteq \Sigma^*$. We say ϕ is* interference-free on $\mathcal{L}$ *if for every non-empty word $u \in \mathcal{L}$,*

- *$\phi(u)$ does* not *admit an interfered image factorization, and*
- *$\phi(u)$ is* not *an inner image factor.*

If $\mathcal{L} = \Sigma^*$, we say that ϕ is strongly *interference-free.*

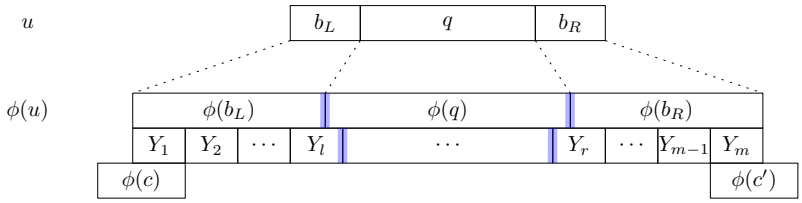

Figure 2 Illustration of Case (1) in the proof of Lemma 11.

The definition is illustrated on the left of Figure 3. (The case where $\phi(u)$ is an inner image factor can be regarded as an edge case that does not affect the intuition, and thus omitted from the figure.) We now observe the following on the Fibonacci morphism.

▶ **Observation 9.** *φ is* not *interference-free on $\{F_i : i \geq 5$ and i is odd$\}$*

Proof. Let $n = |F_i|$ and consider the factorization of $\varphi(F_i) = F_{i+1} = X_1 \cdots X_n$ where each $X_j = \varphi(F_i[j]) \in \mathsf{Im}(\varphi)$ for $1 \leq j \leq n$. Note that $i+1$ is even. By Observation 1, $\mathtt{aba}$ is a suffix of even Fibonacci words. Thus, $X_{n-1} = \varphi(\mathtt{a}) = \mathtt{ab}$ and $X_n = \varphi(\mathtt{b}) = \mathtt{a}$. It follows that $\varphi(F_i)$ admits an interfered image factorization $\varphi(F_i) = x \cdot y \cdot z$ with $x = \varepsilon$, $y = X_1 \dots X_{n-1}$, and $z = X_n = \mathtt{a}$ being a proper prefix of $\varphi(\mathtt{a}) = \mathtt{ab}$. Therefore, φ is not interference-free on the desired set. ◀

This, combined with μ not being interference-free on $\{\mathtt{aa}\}$ (Example 2), gives the following.

▶ **Observation 10.** *The Fibonacci morphism φ and Thue-Morse morphism μ are not strongly interference-free.*

3.1 Properties of Interference-Free Morphisms

In this subsection, we examine several key properties of interference-free morphisms. When proving interference-freeness, it is often more convenient to use the following lemma rather than verify the definition directly. In what follows, b_L and b_R serve as the left and right interference "barriers".

▶ **Lemma 11.** *Let $\phi : \Sigma^* \to \Gamma^*$ be an injective morphism and let $u \in \Sigma^*$. If there exist $b_L, q, b_R \in \Sigma^*$ such that $u = b_L \cdot q \cdot b_R$ and ϕ is interference-free on $\{b_L, b_R\}$, then ϕ is interference-free on $\{u\}$.*

Proof. We proceed by contrapositive and assume that ϕ is not interference-free on $\{u\}$. Then one of the following two cases occurs.

(1) $\phi(u) = \phi(b_L) \cdot \phi(q) \cdot \phi(b_R)$ admits an interfered image factorization $\phi(u) = Y_1 \cdots Y_m$, where Y_1 is a proper suffix of $\phi(c)$ for some $c \in \Sigma$, $Y_i \in \mathsf{Im}(\phi)$ for each $2 \leq i \leq m-1$, Y_m is a proper prefix of $\phi(c')$ for some $c' \in \Sigma$, and $Y_1 \cdot Y_m \neq \varepsilon$; or

(2) $\phi(u)$ is an inner image factor.

In Case (1), if $\phi(u)$ admits an interfered image factorization, see Figure 2 for an illustration of this case. Let l be the smallest index such that $\phi(b_L)$ is a prefix of $Y_1 \cdots Y_l$, and let r be the largest index such that $\phi(b_R)$ is a suffix of $Y_r \cdots Y_m$. (Note that these do not have to be proper prefix/suffix relations; the non-proper case corresponds to the situation where the blue vertical lines are aligned in Figure 2.) Since at least one of Y_1 and Y_m is non-empty, when $Y_1 \neq \varepsilon$, ϕ is not interference-free on $\{b_L\}$; when $Y_m \neq \varepsilon$, ϕ is not interference-free on $\{b_R\}$. In Case (2), if $\phi(u)$ is an inner image factor, then $\phi(b_L)$ and $\phi(b_R)$ are also inner image factors, since both are factors of $\phi(u)$. Therefore, we have shown by contrapositive that ϕ is interference-free on $\{u\}$. ◀

With Lemma 11, one can often verify that ϕ is interference-free on $\{u\}$ simply by examining suitable choices of b_L and b_R. We next demonstrate the usefulness of Lemma 11 through the following corollaries.

▶ **Corollary 12.** *φ is interference-free on $\mathcal{L}_F = \{F_i : i \geq 4$ and i is even$\}$.*

Proof. First note that, for any $i \geq 2$, ab is a prefix of $\varphi(F_i)$. Next, for each $F_i \in \mathcal{L}_F$, $\varphi(F_i) = F_{i+1}$ is an odd Fibonacci word. Thus, by Observation 1, ab is also a suffix of $\varphi(F_i)$. Since φ is interference-free on {ab} by definition, it follows from Lemma 11 that φ is also interference-free on $\mathcal{L}_F$. ◀

▶ **Corollary 13.** *μ is interference-free on $\mathcal{L}_\mathcal{T} = \{\mathcal{T}_i : i \geq 4\}$.*

Proof. For each $\mathcal{T}_i \in \mathcal{L}_\mathcal{T}$, $\mu(\mathcal{T}_i) = \mathcal{T}_{i+1}$. Observe that abba is a prefix of $\mu(\mathcal{T}_i)$, baab is a suffix of $\mu(\mathcal{T}_i)$ if i is odd, and abba is a suffix of $\mu(\mathcal{T}_i)$ if i is even. Since μ is interference-free on {abba, baab} by definition, it follows from Lemma 11 that μ is interference-free on $\mathcal{L}_\mathcal{T}$. ◀

Lemma 11 also leads to the following simple yet powerful characterization.

▶ **Lemma 14.** *An injective morphism $\phi : \Sigma^* \to \Gamma^*$ is strongly interference-free if and only if ϕ is interference-free on Σ.*

Proof. ($\Rightarrow$) If ϕ is strongly interference-free, then ϕ is interference-free on Σ^*, in particular, ϕ is interference-free on Σ. ($\Leftarrow$) If ϕ is interference-free on Σ, then, for each $u \in \Sigma^*$, we know that ϕ is interference-free on $\{u[1], u[|u|]\}$. It follows that, by Lemma 11, ϕ is interference-free on $\{u\}$. Thus, ϕ is strongly interference-free. ◀

With Lemma 14, we can alternatively prove Observation 10 by showing that the Fibonacci morphism φ is not interference-free on {b} and the Thue-Morse morphism μ is not interference-free on {a}. Further, the following result also follows directly from Lemma 14.

▶ **Corollary 15.** *The following injective morphisms are strongly interference-free:*
- *Mephisto-Waltz [28, A064990]:* a ↦ aab, b ↦ bba*;*
- *Thue-Morse-Morse [28, A189718]:* a ↦ abb, b ↦ baa*;*
- *Last nonzero digit [28, A080846]:* a ↦ aba, b ↦ abb*.*

3.2 Interference-Free and Recognizable Morphisms

In this subsection, we uncover a connection between interference-free and recognizable morphisms. We adapt the definition from [10], while extracting the core idea as follows.

▶ **Definition 16** (Circular Image Factorization)**.** *Let $\phi : \Sigma^* \to \Gamma^*$ be a morphism and let $w \in \Gamma^*$ be a word. We say w admits a* circular ϕ-image factorization *if $w = q \cdot r \cdot p$, where r admits an image factorization, $p \cdot q = \phi(c)$ for some $c \in \Sigma$,* and[3] *$p \neq \varepsilon$. When the morphism is clear from the context, we simply say that w admits a* circular image factorization*.*

▶ **Definition 17** (Recognizable Morphisms)**.** *Let $\phi : \Sigma^* \to \Gamma^*$ be an injective morphism and let $\mathcal{L} \subseteq \Sigma^*$. We say ϕ is* recognizable on $\mathcal{L}$ *if for every non-empty word $u \in \mathcal{L}$ and every rotation $w' \in \mathcal{R}(w)$, where $w = \phi(u)$, w' admits a unique circular image factorization. If $\mathcal{L} = \Sigma^*$, we simply say that ϕ is recognizable.*

[3] The condition $p \neq \varepsilon$ is necessary for uniqueness of circular image factorizations. Otherwise, any image factorization $w = X_1 \cdots X_n$ would always yield at least two distinct circular image factorizations, by setting $q = \varepsilon$, $r = X_1 \cdots X_{n-1}$, and $p = X_n$; or $q = X_1$, $r = X_2 \cdots X_n$, and $p = \varepsilon$.

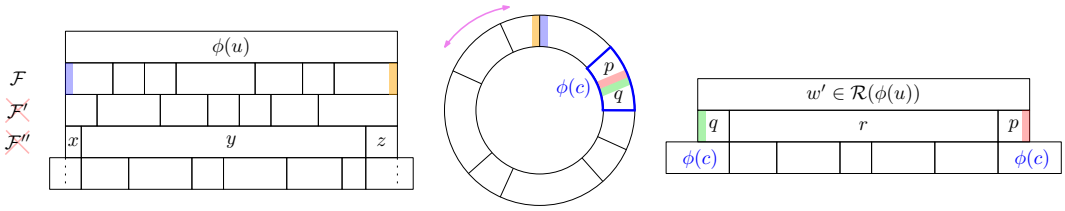

Figure 3 Various illustrations of concepts and results from Section 3. Each unlabeled solid rectangle or annular sector represents an image of ϕ. Left: for Lemma 4 and Definition 8. If ϕ is injective, then $\phi(u)$ admits the unique image factorization $\mathcal{F} = \phi(u[1]) \cdots \phi(u[|u|])$, and no alternative image factorization $\mathcal{F}'$ is possible. If, moreover, ϕ is interference-free on $\{u\}$, then no interfered image factorization $\mathcal{F}''$ is possible either. Right: for Definition 17. Given a rotation w' and a circular image factorization of w', by folding the (linear) word and connecting its two ends (marked in green and red), we obtain the circular word shown in the middle. Middle: for Remark 18.

▶ Remark 18. The intuition behind Theorem 19 is illustrated in Figure 3. If ϕ is interference-free on $\{u\}$, then by folding the (linear) word $\phi(u)$ and connecting its two ends (marked in blue and yellow on the left), we obtain the circular word shown in the middle. Crucially, regardless of how the circular word is rotated, producing different rotations in $\mathcal{R}(\phi(u))$, the condition for recognizability remains satisfied. Hence, with interference-freeness, each unique (linear) image factorization induces a unique circular image factorization. ⌟

▶ **Theorem 19.** *Let $\phi : \Sigma^* \to \Gamma^*$ be an injective morphism and let $\mathcal{L} \subseteq \Sigma^*$. If ϕ is interference-free on $\mathcal{L}$, then ϕ is recognizable on $\mathcal{L}$.*

Proof. We proceed by contrapositive and prove that, for each non-empty word $u \in \mathcal{L}$, if ϕ is *not* recognizable on $\{u\}$, then ϕ is *not* interference-free on $\{u\}$.

Let $w = \phi(u)$. We have assumed that ϕ is not recognizable on $\{u\}$, which means there exist some $w' \in \mathcal{R}(w)$ such that either w' does not admit a circular image factorization, or its factorization is not unique. We first show that each $w' \in \mathcal{R}(w)$ does admit a circular image factorization. Let $w = X_1 \cdots X_n$ be an image factorization. Since w' is a rotation of w, there exists an index $1 \leq c \leq n$ such that $w' = X_c^{suf} \cdot X_{c+1} \cdots X_n \cdot X_1 \cdots X_{c-1} \cdot X_c^{pref}$, where $X_c^{pref} \neq \varepsilon$ and $X_c^{pref} \cdot X_c^{suf} = X_c$. Setting $q = X_c^{suf}$, $r = X_{c+1} \cdots X_n \cdot X_1 \cdots X_{c-1}$, and $p = X_c^{pref}$, we have shown that w' admits a circular image factorization.

Since such a factorization exists for each $w' \in \mathcal{R}(w)$, our assumption implies that there must exist w' whose circular image factorization is not unique. Hence, consider one such factorization of w', given by $w' = Y_1^{suf} \cdot Y_2 \cdots Y_m \cdot Y_1^{pref}$, where $Y_i \in \mathsf{Im}(\phi)$ for $1 \leq i \leq m$ and $Y_1^{pref} \cdot Y_1^{suf} = Y_1$. Further, define index $1 \leq k \leq m$, words Y_k^{pref} and Y_k^{suf} such that $Y_k^{pref} \cdot Y_k^{suf} = Y_k$, and $Y_1^{suf} \cdot Y_2 \cdots Y_k^{pref} = X_c^{suf} \cdot X_{c+1} \cdots X_n$. Now, rotating w' back to w gives $w = Y_k^{suf} \cdot Y_{k+1} \cdots Y_m \cdot Y_1 \cdots Y_{k-1} \cdot Y_k^{pref}$. We now prove that ϕ is not interference-free on $\{u\}$ by considering this factorization of w in three cases.

(1) $Y_k^{pref} = Y_k$ and $Y_k^{suf} = \varepsilon$. In this case, $w = X_1 \cdots X_n = Y_{k+1} \cdots Y_m \cdot Y_1 \cdots Y_{k-1} \cdot Y_k$. The injectivity of ϕ implies that the factorizations for w must be equivalent, contradicting that the two factorizations for w' were distinct.

(2) $Y_k^{pref} = \varepsilon$ and $Y_k^{suf} = Y_k$. By a symmetric argument, ϕ is again not injective.

(3) $Y_k^{pref} \neq \varepsilon$ and $Y_k^{suf} \neq \varepsilon$. In this case, let $x = Y_k^{suf}$, $y = Y_{k+1} \cdots Y_m \cdot Y_1 \cdots Y_{k-1}$, and $z = Y_k^{pref}$. Then, ϕ is not interference-free on $\{u\}$ since x is a proper suffix of $Y_k \in \mathsf{Im}(\phi)$, $y \in \mathsf{Im}(\phi)^*$, z is a proper prefix of $Y_k \in \mathsf{Im}(\phi)$, and $x \cdot z = Y_k \neq \varepsilon$. Thus, $w = \phi(u)$ admits an interfered image factorization.

Therefore, we have shown that if ϕ is not recognizable on $\{u\}$, then ϕ is not interference-free on $\{u\}$. This completes the proof. ◀

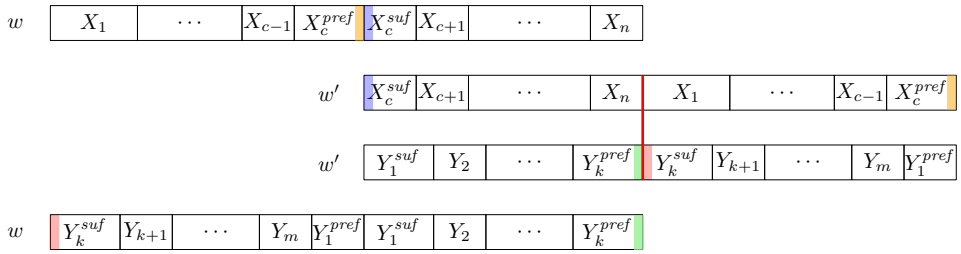

Figure 4 Illustration of the proof of Theorem 19.

Theorem 19 naturally extends to the following, more general result.

▶ **Corollary 20.** *Every strongly interference-free morphism is recognizable.*

Having established that interference-freeness implies recognizability, it is natural to ask whether the converse also holds. In [1], the Fibonacci morphism was shown to be recognizable. Together with Observation 9, this implies a morphism φ and an infinite family $\mathcal{L}$ of words such that φ is recognizable on $\mathcal{L}$ but not interference-free on $\mathcal{L}$. Thus, recognizability does not imply interference-freeness. We summarize the resulting strict hierarchy below.

▶ Remark 21. The following inclusion relation among morphisms holds:

$$\texttt{INTERFERENCE\_FREE} \subsetneq \texttt{RECOGNIZABLE} \subsetneq \texttt{INJECTIVE}.$$

Moreover, the Thue-Morse morphism is injective but not recognizable, whereas the Fibonacci morphism is recognizable but not strongly interference-free. ⌟

4 Occurrence-Preserving Morphisms

After introducing and discussing interference-free morphisms, we are ready to present the main result, which establishes sufficient conditions under which factor occurrences are preserved.

▶ **Theorem 22.** *Let $\phi : \Sigma^* \to \Gamma^*$ be an injective morphism, let $u, v \in \Sigma^*$ be two words, and let $k \geq 1$ be an integer. If ϕ is interference-free on $\bigcup_{i=0}^{k-1}\{\phi^i(u)\}$, then,*

$$\#\mathrm{occ}_v(u) = \#\mathrm{occ}_{\phi^k(v)}(\phi^k(u)).$$

To prove this theorem, we first observe that it suffices to establish the case $k = 1$. The general case then follows by induction on k, using an argument analogous to that for the base case $k = 1$. Accordingly, we aim to show the following.

▶ **Proposition 23.** *Let $\phi : \Sigma^* \to \Gamma^*$ be an injective morphism and let $u, v \in \Sigma^*$ be two words. If ϕ is interference-free on $\{u\}$, then $\#\mathrm{occ}_v(u) = \#\mathrm{occ}_{\phi(v)}(\phi(u))$.*

We begin with the following lemma, which intuitively states that applying a non-erasing morphism cannot "eliminate" existing factor occurrences.

▶ **Lemma 24.** *Let $\phi : \Sigma^* \to \Gamma^*$ be a non-erasing morphism and let $u, v \in \Sigma^*$ be two words. Then $\#\mathrm{occ}_v(u) \leq \#\mathrm{occ}_{\phi(v)}(\phi(u))$.*

Proof. For each occurrence of u in v, there exist words $v^{pref}, v^{suf} \in \Sigma^*$ such that $v = v^{pref} \cdot u \cdot v^{suf}$. Applying ϕ to both sides gives $\phi(v) = \phi(v^{pref}) \cdot \phi(u) \cdot \phi(v^{suf})$. Hence, each occurrence of u in v yields an occurrence of $\phi(u)$ in $\phi(v)$. Moreover, distinct occurrences of u in v yield distinct occurrences of $\phi(u)$ in $\phi(v)$, since they correspond to different prefixes v^{pref}. Therefore, $\#\mathrm{occ}_v(u) \leq \#\mathrm{occ}_{\phi(v)}(\phi(u))$. ◀

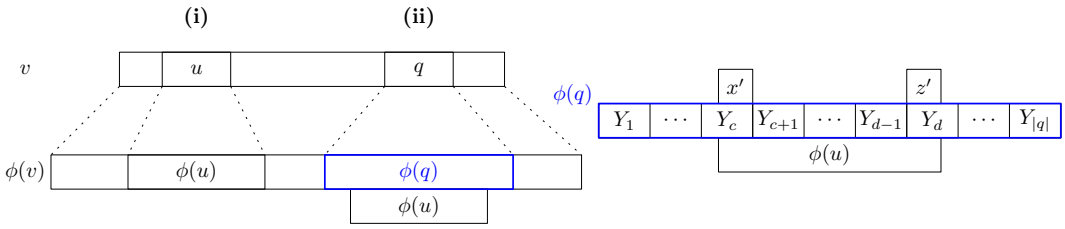

Figure 5 Left: illustration of Lemma 26. Right: illustration of the proof of Proposition 23.

Note that the non-erasing condition is not explicitly stated in Theorem 22; the following observation shows that it is already implied by injectivity.

▶ **Observation 25.** *Let $\phi : \Sigma^* \to \Gamma^*$ be a morphism. If ϕ is injective then ϕ is non-erasing.*

Proof. We proceed by contrapositive. If ϕ is erasing, then there exists $c \in \Sigma$ such that $\phi(c) = \varepsilon$. Then, $\phi(cc) = \phi(c) = \varepsilon$ but $cc \neq c$, which means ϕ is not injective. ◀

The next lemma traces occurrences of $\phi(u)$ in $\phi(v)$ back to the corresponding occurrences in v, identifying the only two possible cases, illustrated on the left of Figure 5.

▶ **Lemma 26.** *Let $\phi : \Sigma^* \to \Gamma^*$ be an injective morphism and let $u, v \in \Sigma^*$ be two words. If $\phi(u)$ is a factor of $\phi(v)$, then at least one of the following two cases holds:*

(i) *u is a factor of v, or*

(ii) *there exists a factor q of v such that $q \neq u$ and $\phi(u)$ is a proper factor of $\phi(q)$.*

Proof. Assume, for contradiction, that $\phi(u)$ is a factor of $\phi(v)$ while neither Case (i) nor Case (ii) holds. Since u is not a factor of v, the only possible way for $\phi(u)$ to be a factor of $\phi(v)$ is that there exists a factor q' of v such that $q' \neq u$ and $\phi(q') = \phi(u)$. However, this contradicts the injectivity of ϕ. ◀

▶ Remark 27. In the context of Proposition 23, Case (i) corresponds to the situation in which the occurrence count is preserved. In contrast, Case (ii) corresponds to the situation in which "interference" causes an occurrence-count violation. In Lemma 26, only the injectivity of ϕ is assumed. In the proof of Proposition 23, we show that imposing the interference-free condition rules out the undesirable Case (ii). ⌟

Proof of Proposition 23. We aim to show that $\#\mathrm{occ}_v(u) \geq \#\mathrm{occ}_{\phi(v)}(\phi(u))$. We follow the two cases described in Lemma 26. In particular, we prove that Case (ii) is impossible and only Case (i) is possible under the interference-free condition in Theorem 22. Recall that q was introduced in Lemma 26. First observe that if $|q| = 1$, i.e., q is a symbol, then Case (ii) is impossible, as it would force one of the following two contradictions. For any $1 \leq i \leq |v|$,

- if $\phi(u)$ is a proper prefix or a proper suffix of $\phi(v[i])$, then $\phi(u)$ admits an interfered image factorization where $x = y = \varepsilon$ and z is a proper prefix of $\phi(v[i])$, or x is a proper suffix of $\phi(v[i])$ and $y = z = \varepsilon$;
- if $\phi(u)$ is a proper factor of $\phi(v[i])$ but is neither a prefix nor a suffix of $\phi(v[i])$, then $\phi(u)$ is an inner image factor.

Thus, we know that $|q| > 1$. We now show that if ϕ is interference-free on $\{u\}$ then Case (ii) is impossible. We proceed by contrapositive: assume Case (ii) does occur, and we will show that ϕ is not interference-free on $\{u\}$. Let $\phi(q) = Y_1 \cdots Y_{|q|}$ where $Y_i \in \mathsf{Im}(\phi)$ for $1 \leq i \leq |q|$. If $\phi(u)$ is a proper factor of $\phi(q)$, then there exists a factor of $\phi(q)$ of the form $Y_c \cdots Y_d$ (for some $1 \leq c < d \leq |q|$) such that $\phi(u)$ is a proper factor of $Y_c \cdots Y_d$ and $Y_{c+1} \cdots Y_{d-1}$ is a

proper factor of $\phi(u)$. (We define $Y_{c+1} \cdots Y_{d-1}$ to be ε when $d = c + 1$.) Hence, we have $\phi(u) = x' \cdot Y_{c+1} \cdots Y_{d-1} \cdot z'$, where x' is a suffix of Y_c, z' is a prefix of Y_d, and $x' \cdot z' \neq Y_c \cdot Y_d$ (illustrated on the right of Figure 5). Now we are ready to show that ϕ is not interference-free on $\{u\}$ by constructing an interfered image factorization of $\phi(u) = x \cdot y \cdot z$, where

$$x = \begin{cases} \varepsilon \text{ if } x' = Y_c, \\ x' \text{ otherwise;} \end{cases} \quad z = \begin{cases} \varepsilon \text{ if } z' = Y_d, \\ z' \text{ otherwise;} \end{cases} \quad y = \begin{cases} Y_{c+1} \cdots Y_{d-1} \text{ if } x = x' \text{ and } z = z', \\ Y_c \cdots Y_{d-1} \text{ if } x = \varepsilon \text{ and } z = z', \\ Y_{c+1} \cdots Y_d \text{ if } x = x' \text{ and } z = \varepsilon. \end{cases}$$

Thus, we have proved by contrapositive that Case (ii) is impossible when ϕ is interference-free on $\{u\}$. Therefore, only Case (i) is possible under the conditions in Theorem 22. In other words, we have shown that if $\phi(u)$ is a factor of $\phi(v)$, then u is a factor of v. This implies that $\#\mathrm{occ}_v(u) \geq \#\mathrm{occ}_{\phi(v)}(\phi(u))$. Combining with Lemma 24, we conclude that $\#\mathrm{occ}_v(u) = \#\mathrm{occ}_{\phi(v)}(\phi(u))$. ◀

Finally, Theorem 22 follows from Proposition 23 by induction.

5 Applications of Occurrence-Preserving Morphisms

In this section, we present several applications of the characterization of occurrence-preserving morphisms established in the previous section. As an immediate consequence of Theorem 22, occurrence preservation holds at a more granular level: not only is the number of occurrences preserved, but there also exists a bijection between the starting positions of the occurrences of u in v and those of $\phi^k(u)$ in $\phi^k(v)$.

▶ **Lemma 28.** *Let $\phi : \Sigma^* \to \Gamma^*$ be an injective morphism, let $u, v \in \Sigma^*$ be two words, and let $k \geq 1$ be an integer. If ϕ is interference-free on $\bigcup_{i=0}^{k-1} \{\phi^i(u)\}$, then, $p \in \mathrm{occ}_v(u)$ if and only if $|\phi^k(v[1 \ldots p-1])| + 1 \in \mathrm{occ}_{\phi^k(v)}(\phi^k(u))$.*

Proof. ($\Rightarrow$) If u occurs at position p in v, then v can be factorized as $v = v[1 \ldots p-1] \cdot u \cdot y$ for $y \in \Sigma^*$. It follows that $\phi^k(v) = \phi^k(v[1 \ldots p-1]) \cdot \phi^k(u) \cdot \phi^k(y)$, and thus $\phi^k(u)$ occurs at position $|\phi^k(v[1 \ldots p-1])| + 1$ in $\phi^k(v)$.

($\Leftarrow$) From the "$\Rightarrow$" direction we can infer that $\#\mathrm{occ}_v(u) \leq \#\mathrm{occ}_{\phi^k(v)}(\phi^k(u))$. By Theorem 22, we know that $\#\mathrm{occ}_v(u) = \#\mathrm{occ}_{\phi^k(v)}(\phi^k(u))$. Thus, there does not exist an occurrence p' of $\phi^k(u)$ in $\phi^k(v)$ such that $p' \neq |\phi^k(v[1 \ldots p-1])| + 1$ for any $p \in \mathrm{occ}_v(u)$. Therefore, the "$\Leftarrow$" direction also holds. ◀

Note that if ϕ is ℓ-uniform, then $|\phi^k(v[1 \ldots p-1])| + 1 = \ell^k(p-1) + 1$.

MUSs of Fibonacci and Thue-Morse Words

In this subsection, as another application of Theorem 22, we identify the minimal unique substrings (MUSs) of Fibonacci and Thue-Morse words.

MUSs of Fibonacci words. We first state two observations on $F_i^{\triangleleft} = F_i[1 \ldots f_i - 1]$.

▶ **Observation 29.** *For $i \geq k \geq 4$, $\#\mathrm{occ}_{F_i}(F_k^{\triangleleft}) = \#\mathrm{occ}_{F_i}(F_k)$.*

Proof. First observe that each occurrence of $F_k^{\triangleleft}$ is followed by either `a` or `b`. Hence $\#\mathrm{occ}_{F_i}(F_k^{\triangleleft}) = \#\mathrm{occ}_{F_i}(F_k^{\triangleleft} \cdot \mathtt{a}) + \#\mathrm{occ}_{F_i}(F_k^{\triangleleft} \cdot \mathtt{b})$. By Observation 1, when k is even, F_k ends with `aba`, so $F_k^{\triangleleft}$ ends with `ab` and $F_k^{\triangleleft} \cdot \mathtt{a} = F_k$. Further, by Observation 1, `bb` does not occur

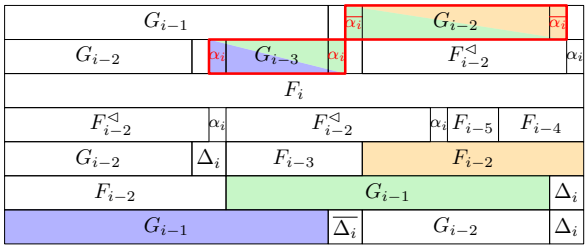

Figure 6 Illustration of Theorem 32. In the top two factorizations of F_i, each MUS is highlighted in a red rectangle. In the bottom three factorizations of F_i, each net occurrence is colored. Moreover, the colors of the two consecutive net occurrences together are used as the colors to highlight the occurrence of the MUS they correspond to. The uncolored factorization is used in the proof.

in F_i, which means $\#\mathrm{occ}_{F_i}(F_k^{\triangleleft} \cdot \mathtt{b}) = 0$. Thus, $\#\mathrm{occ}_{F_i}(F_k^{\triangleleft}) = \#\mathrm{occ}_{F_i}(F_k^{\triangleleft} \cdot \mathtt{a}) = \#\mathrm{occ}_{F_i}(F_k)$. The case where k is odd can be proved analogously using the facts that $\mathtt{aaa}$ does not occur in F_i and F_k ends with $\mathtt{aab}$. ◀

▶ **Observation 30.** *φ is interference-free on $\{F_i^{\triangleleft} : i \geq 4$ and i is odd$\}$.*

Proof. Since i is odd, F_i ends with $\mathtt{ab}$ and F_{i+1} ends with $\mathtt{ba}$ by Observation 1. With $F_i = F_i^{\triangleleft} \cdot \mathtt{b}$, applying φ, we obtain $F_{i+1} = \varphi(F_i) = \varphi(F_i^{\triangleleft})\varphi(\mathtt{b}) = \varphi(F_i^{\triangleleft}) \cdot \mathtt{a}$. Since $\varphi(\mathtt{b}) = \mathtt{a}$, we have $\varphi(F_i^{\triangleleft}) = F_{i+1}^{\triangleleft}$. The claim then follows using a similar argument for Corollary 12 and the fact that $\mathtt{ab}$ is both a prefix and a suffix of $\varphi(F_i^{\triangleleft})$. ◀

Occurrences of smaller-order Fibonacci words (i.e., occurrences of F_k in F_i for $k < i$) have been studied previously [13]. Using our characterization of occurrence-preserving morphisms, we show that $\#\mathrm{occ}_{F_i}(F_{i-d})$ is independent of i once the offset d is fixed.

▶ **Lemma 31.** *For i, j, d with $i \geq j$ and $j - d \geq 4$, $\#\mathrm{occ}_{F_i}(F_{i-d}) = \#\mathrm{occ}_{F_j}(F_{j-d})$.*

Proof. It suffices to prove the case $j = i - 1$, i.e., $\#\mathrm{occ}_{F_i}(F_{i-d}) = \#\mathrm{occ}_{F_{i-1}}(F_{i-1-d})$, and the result then follows. When $i - d$ is even, the desired equality holds by Corollary 12 and Theorem 22. For odd $i - d \geq 4$, we have

$$\#\mathrm{occ}_{F_i}(F_{i-d}) = \#\mathrm{occ}_{F_i}(F_{i-d}^{\triangleleft}) = \#\mathrm{occ}_{F_{i-1}}(F_{i-1-d}^{\triangleleft}) = \#\mathrm{occ}_{F_{i-1}}(F_{i-1-d}),$$

where the first and last equalities hold by Observation 29, while the second equality holds by Observation 30 and Theorem 22. This completes the proof. ◀

▶ **Theorem 32.** *For each $i \geq 6$, let $\alpha_i = \mathtt{a}$ if i is even, and $\alpha_i = \mathtt{b}$ if i is odd. Then, $\mathit{MUS}(F_i) = \{\alpha_i\, G_{i-3}\, \alpha_i,\ \overline{\alpha_i}\, G_{i-2}\, \overline{\alpha_i}\}$.*

Proof. We proceed to show the set equality by proving subset relations in both directions.

We first show that $\overline{\alpha_i}\, G_{i-2}\, \overline{\alpha_i} \in \mathsf{MUS}(F_i)$. Its unique occurrence is outlined in red in Figure 6. By Lemma 31, we have $\#\mathrm{occ}_{F_i}(F_{i-2}) = \#\mathrm{occ}_{F_6}(F_4) = 3$. The factorizations $F_i = F_{i-2}F_{i-2}F_{i-5}F_{i-4}$ and $F_i = F_{i-2}F_{i-3}F_{i-2}$ reveal all three occurrences of F_{i-2}. (In Figure 6, these two factorizations correspond to $F_i = F_{i-2}^{\triangleleft}\alpha_i F_{i-2}^{\triangleleft}\alpha_i F_{i-5}F_{i-4}$ and $F_i = G_{i-2}\Delta_i F_{i-3}F_{i-2}$.) Moreover, by Observation 29, we have $\#\mathrm{occ}_{F_i}(F_{i-2}^{\triangleleft}) = \#\mathrm{occ}_{F_i}(F_{i-2}) = 3$. Among the three occurrences of $F_{i-2}^{\triangleleft} = G_{i-2}\,\overline{\alpha_i}$ in F_i, exactly one is preceded by $\overline{\alpha_i}$ (namely the one corresponding to the occurrence of F_{i-2} highlighted in yellow in Figure 6). Hence, $\overline{\alpha_i}\, G_{i-2}\, \overline{\alpha_i}$ is unique in F_i. Further, $\overline{\alpha_i}\, G_{i-2}$ is not unique in F_i, since it is a substring of

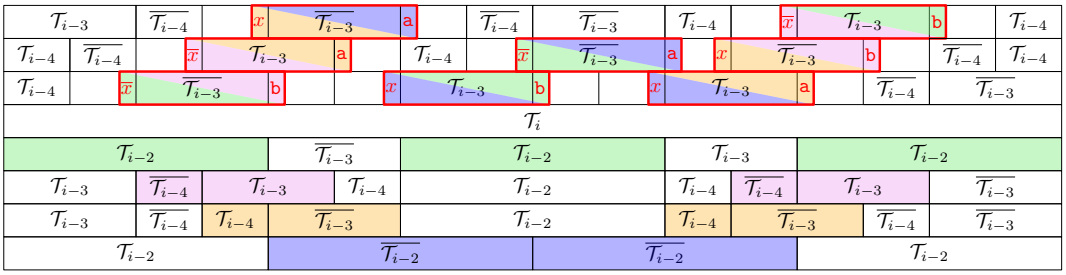

Figure 7 Illustration of Theorem 34, following the conventions of Figure 6. Let $x = \mathtt{b}$ for even i and $x = \mathtt{a}$ for odd i.

G_{i-1}, which occurs twice in F_i (by Observation 1). Now, since $\overline{\alpha_i}\, G_{i-2}\, \overline{\alpha_i}$ is unique, while both $G_{i-2}\, \overline{\alpha_i}$ and $\overline{\alpha_i}\, G_{i-2}$ are repeated, we have shown that $\overline{\alpha_i}\, G_{i-2}\, \overline{\alpha_i}$ is a MUS. By an analogous argument, $\alpha_i\, G_{i-3}\, \alpha_i \in \mathsf{MUS}(F_i)$ can be proved similarly.

We next show $\mathsf{MUS}(F_i) \subset \{\alpha_i\, G_{i-3}\, \alpha_i,\ \overline{\alpha_i}\, G_{i-2}\, \overline{\alpha_i}\}$. Let $[s, e]$ be an occurrence of a substring $P = F_i[s \dots e]$ of F_i such that $[s, e]$ is not a proper sub-occurrence of $[f_{i-2}, f_{i-1} - 1]$ or $[f_{i-1}, f_i - 1]$, which are the occurrences of $\alpha_i\, G_{i-3}\, \alpha_i$ and $\overline{\alpha_i}\, G_{i-2}\, \overline{\alpha_i}$, respectively. We claim that P is repeated and thus cannot be a MUS. We consider the following cases.

- When $e \le f_{i-1} - 2$, P is contained in the repeated substring $F_i[1 \dots f_{i-1} - 2] = G_{i-1}$, highlighted in blue in Figure 6.
- When $f_{i-2} + 1 \le s \le f_{i-1} - 1$ and $f_{i-1} \le e \le f_i - 2$, P is a contained in the repeated substring $F_i[f_{i-2} + 1 \dots f_i - 2] = G_{i-1}$, highlighted in green in Figure 6.
- When $f_{i-1}+1 \le s$ and $e = f_i$, P is contained in the repeated substring $F_i[f_{i-1}+1 \dots f_i] = F_{i-2}$, highlighted in yellow in Figure 6.

In all cases, P is repeated in F_i and thus cannot be a MUS. Therefore, we conclude that $\mathsf{MUS}(F_i) = \{\alpha_i\, G_{i-3}\, \alpha_i,\ \overline{\alpha_i}\, G_{i-2}\, \overline{\alpha_i}\}$. ◀

MUSs of Thue-Morse words. Before proving the MUSs of Thue-Morse words, we first state the following lemma, which serves as the Thue-Morse analogue of Lemma 31.

▶ **Lemma 33.** *For each $i \ge j$ and $j - d \ge 2$, we have $\#\mathrm{occ}_{\mathcal{T}_i}(\mathcal{T}_{i-d}) = \#\mathrm{occ}_{\mathcal{T}_j}(\mathcal{T}_{j-d})$ and $\#\mathrm{occ}_{\mathcal{T}_i}(\overline{\mathcal{T}_{i-d}}) = \#\mathrm{occ}_{\mathcal{T}_j}(\overline{\mathcal{T}_{j-d}})$.*

Proof. The first equality follows directly from Corollary 13 and Theorem 22. For the second equality, note that μ is interference-free on $\{\overline{\mathcal{T}_i} : i \ge 4\}$, as can be shown by an argument analogous to Corollary 13; combining this with Theorem 22 yields the desired result. ◀

▶ **Theorem 34.** *For a word w, let $\mathit{EXT}(w) = \{\mathtt{a}\, w\, \mathtt{a}, \mathtt{a}\, w\, \mathtt{b}, \mathtt{b}\, w\, \mathtt{a}, \mathtt{b}\, w\, \mathtt{b}\}$ be the set of extensions of w. Then, for each $i \ge 5$, $\mathit{MUS}(\mathcal{T}_i) = \mathit{EXT}(\mathcal{T}_{i-3}) \cup \mathit{EXT}(\overline{\mathcal{T}_{i-3}})$.*

Proof. As in the proof of Theorem 32, we establish the set equality by proving both inclusions. We first show that $\mathsf{EXT}(\mathcal{T}_{i-3}) \subset \mathsf{MUS}(\mathcal{T}_i)$. By Lemma 33, we have $\#\mathrm{occ}_{\mathcal{T}_i}(\mathcal{T}_{i-3}) = \#\mathrm{occ}_{\mathcal{T}_5}(\mathcal{T}_2) = 4$. Consider the four occurrences of $\mathcal{T}_{i-3}$ in $\mathcal{T}_i$. Each occurrence is preceded and followed by either $\mathcal{T}_{i-4}$ or $\overline{\mathcal{T}_{i-4}}$. More precisely, across the four occurrences, all four possible combinations occur: $\mathcal{T}_{i-4}\mathcal{T}_{i-3}\mathcal{T}_{i-4}$, $\overline{\mathcal{T}_{i-4}}\mathcal{T}_{i-3}\overline{\mathcal{T}_{i-4}}$, $\mathcal{T}_{i-4}\mathcal{T}_{i-3}\overline{\mathcal{T}_{i-4}}$, and $\overline{\mathcal{T}_{i-4}}\mathcal{T}_{i-3}\mathcal{T}_{i-4}$. Since $\mathcal{T}_{i-4}$ and $\overline{\mathcal{T}_{i-4}}$ have different first and last letters, It follows that each $u \in \mathsf{EXT}(\mathcal{T}_{i-3})$ is unique. Further, since $\mathcal{T}_{i-4}\mathcal{T}_{i-3}$, $\overline{\mathcal{T}_{i-4}}\mathcal{T}_{i-3}$, $\mathcal{T}_{i-3}\overline{\mathcal{T}_{i-4}}$, and $\mathcal{T}_{i-3}\mathcal{T}_{i-4}$ all occur twice in $\mathcal{T}_i$, it follows that $\mathtt{a}\,\mathcal{T}_{i-3}$, $\mathtt{b}\,\mathcal{T}_{i-3}$, $\mathcal{T}_{i-3}\,\mathtt{a}$, and $\mathcal{T}_{i-3}\,\mathtt{b}$ also all occur twice in $\mathcal{T}_i$. Therefore, $\mathsf{EXT}(\mathcal{T}_{i-3}) \subset \mathsf{MUS}(\mathcal{T}_i)$. With a similar argument, it can be shown that $\mathsf{EXT}(\overline{\mathcal{T}_{i-3}}) \subset \mathsf{MUS}(\mathcal{T}_i)$. All these eight MUSs are outlined in red in Figure 7.

We next prove $\mathsf{MUS}(\mathcal{T}_i) \subset \mathsf{EXT}(\mathcal{T}_{i-3}) \cup \mathsf{EXT}(\overline{\mathcal{T}_{i-3}})$. Let $[s, e]$ be an occurrence of a substring $P = \mathcal{T}_i[s \dots e]$ of $\mathcal{T}_i$ such that $[s, e]$ is not a proper sub-occurrence of any occurrences of substrings in $\mathsf{EXT}(\mathcal{T}_{i-3}) \cup \mathsf{EXT}(\overline{\mathcal{T}_{i-3}})$. We claim that P is repeated and thus cannot be a MUS. We consider the following cases.

- When $s \leq \tau_{i-3} - 1$ and $e \leq \tau_{i-2}$, P is contained in the repeated substring $\mathcal{T}_i[1 \dots \tau_{i-2}] = \mathcal{T}_{i-2}$, highlighted in green on the left of Figure 7.
- When $\tau_{i-3} + 1 \leq s \leq \tau_{i-3} + \tau_{i-4} - 1$ and $\tau_{i-2} + 1 \leq e \leq \tau_{i-2} + \tau_{i-4}$, P is contained in the repeated substring $\mathcal{T}_i[\tau_{i-3} + 1 \dots \tau_{i-2} + \tau_{i-4}] = \overline{\mathcal{T}_{i-4}}\,\mathcal{T}_{i-3}$, highlighted in pink on the left of Figure 7.
- When $\tau_{i-3} + \tau_{i-4} + 1 \leq s \leq \tau_{i-2} - 1$ and $\tau_{i-2} + \tau_{i-4} + 1 \leq e \leq \tau_{i-2} + \tau_{i-3}$, P is contained in the repeated substring $\mathcal{T}_i[\tau_{i-3} + \tau_{i-4} + 1 \dots \tau_{i-2} + \tau_{i-3}] = \mathcal{T}_{i-4}\,\overline{\mathcal{T}_{i-3}}$, highlighted in yellow on the left of Figure 7.
- When $\tau_{i-2} + 1 \leq s \leq \tau_{i-2} + \tau_{i-3} - 1$ and $\tau_{i-2} + \tau_{i-3} + 1 \leq e \leq \tau_{i-1}$, P is contained in the repeated substring $\mathcal{T}_i[\tau_{i-2} + 1 \dots \tau_{i-1}] = \overline{\mathcal{T}_{i-2}}$, highlighted in blue on the left of Figure 7.
- When $\tau_{i-2} + \tau_{i-3} + 1 \leq s \leq \tau_{i-1} - 1$ and $\tau_{i-1} + 1 \leq e \leq \tau_{i-1} + \tau_{i-3}$, P is contained in the repeated substring $\mathcal{T}_i[\tau_{i-2} + \tau_{i-3} + 1 \dots \tau_{i-1} + \tau_{i-3}] = \mathcal{T}_{i-2}$, highlighted in green in the middle of Figure 7.

The remaining four cases can be proved analogously, and are illustrated by the blue, yellow, pink, and green substrings on the right of Figure 7. In all cases, P is repeated in $\mathcal{T}_i$ and thus cannot be a MUS. Therefore, we conclude that $\mathsf{MUS}(\mathcal{T}_i) = \mathsf{EXT}(\mathcal{T}_{i-3}) \cup \mathsf{EXT}(\overline{\mathcal{T}_{i-3}})$. ◀

Finally, with the two theorems above, we can exploit the connection between MUSs and net occurrences (Lemma 35) to simplify existing proofs on net occurrences in these words [13] (Corollary 36). These results are also illustrated in Figure 6 and Figure 7.

▶ **Lemma 35** ([18, 23]). *For a string w, let $[i_1, j_1], [i_2, j_2], \dots, [i_m, j_m]$ be the sequence of MUS occurrences in w, ordered by increasing starting position. Then, $\mathsf{NO}(w) = \{[1, j_1 - 1], [i_1 + 1, j_2 - 1], \dots, [i_{m-1} + 1, j_m - 1], [i_m + 1, |w|]\}$.*

▶ **Corollary 36** ([13]). *For each $i \geq 7$, $\mathsf{NO}(F_i)$ consists of the two occurrences of G_{i-1} and the last occurrence of F_{i-2}. For each $i \geq 5$, $\mathsf{NO}(\mathcal{T}_i)$ consists of all occurrences of each of the following substrings: $\mathcal{T}_{i-2}$, $\overline{\mathcal{T}_{i-2}}$, $\mathcal{T}_{i-4}\,\overline{\mathcal{T}_{i-3}}$, and $\overline{\mathcal{T}_{i-4}}\,\mathcal{T}_{i-3}$.*

6 Conclusion and Future Work

In this work, we investigated morphisms that preserve factor occurrences. We introduced the notion of *interference-free morphisms* and studied their key properties. Building on this notion, we established sufficient conditions for *occurrence-preserving morphisms* and applied these conditions to identify the MUSs of the Fibonacci and Thue-Morse words. We now outline several directions for future work.

First, one could consider improving the conditions in Theorem 22 so that they are both sufficient and necessary. A key observation is that interference-freeness is defined with respect to a morphism ϕ and a single word u, whereas occurrence-preservation concerns ϕ and two words u and v. Thus, extending the definition of interference-freeness to involve the two words may lead to a tighter characterization. (A natural starting point would be to restrict attention to cases where $\#\mathrm{occ}_v(u) > 0$, since when $\#\mathrm{occ}_v(u) = 0$, occurrences can be trivially preserved even if interference-freeness does not hold.) Furthermore, it would be interesting to examine the class of morphisms that lie between recognizable and strongly interference-free morphisms in the inclusion hierarchy (Remark 21). Another natural direction is to investigate the class of infinite words generated as fixed points of strongly interference-free morphisms.

A second line of future work is algorithmic. For example, in Section 3.2 of [5], a particular class of morphisms is used in an algorithm for the inverse problem of overlap-graph construction. Since their morphisms are defined for a specific algorithmic setting, the assumptions on the morphisms are naturally more restrictive: for example, they require the morphisms to be uniform. Further, their definition explicitly excludes Case (ii) in our Lemma 26; in contrast, we do not impose such a restriction directly, but instead capture the essential behavior through the conditions in our Theorem 22. It would be worthwhile to explore other algorithmic settings in which interference-free morphisms may prove useful.

References

1 Marie-Pierre Béal, Dominique Perrin, and Antonio Restivo. Recognizability of morphisms. *Ergod. Th. & Dynam. Sys.*, 43(11):3578–3602, 2023. doi:10.1017/etds.2022.109.

2 Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*, volume 129 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, 2010.

3 Jean Berstel and Patrice Séébold. A characterization of overlap-free morphisms. *Discret. Appl. Math.*, 46(3):275–281, 1993. doi:10.1016/0166-218X(93)90107-Y.

4 Jean Berstel and Patrice Séébold. A characterization of Sturmian morphisms. In Andrzej M. Borzyszkowski and Stefan Sokolowski, editors, *Mathematical Foundations of Computer Science 1993, 18th International Symposium, MFCS'93, Gdansk, Poland, August 30 - September 3, 1993, Proceedings*, volume 711 of *Lecture Notes in Computer Science*, pages 281–290. Springer, 1993. doi:10.1007/3-540-57182-5_20.

5 Marília D. V. Braga and Joao Meidanis. An algorithm that builds a set of strings given its overlap graph. In Sergio Rajsbaum, editor, *LATIN 2002: Theoretical Informatics, 5th Latin American Symposium, Cancun, Mexico, April 3-6, 2002, Proceedings*, volume 2286 of *Lecture Notes in Computer Science*, pages 52–63. Springer, 2002. doi:10.1007/3-540-45995-2_10.

6 Srecko Brlek, Andrea Frosini, Ilaria Mancini, Elisa Pergola, and Simone Rinaldi. Burrows-Wheeler Transform of words defined by morphisms. In Charles J. Colbourn, Roberto Grossi, and Nadia Pisanti, editors, *Combinatorial Algorithms - 30th International Workshop, IWOCA 2019, Pisa, Italy, July 23-25, 2019, Proceedings*, volume 11638 of *Lecture Notes in Computer Science*, pages 393–404. Springer, 2019. doi:10.1007/978-3-030-25005-8_32.

7 Arturo Carpi. On Abelian power-free morphisms. *Int. J. Algebra Comput.*, 3(2):151–168, 1993. doi:10.1142/S0218196793000123.

8 Francesco Dolce and Edita Pelantová. On morphisms preserving palindromic richness. *Fundam. Informaticae*, 185(1):1–25, 2022. doi:10.3233/FI-222102.

9 Gabriele Fici, Giuseppe Romana, Marinella Sciortino, and Cristian Urbina. On the impact of morphisms on BWT-runs. In Laurent Bulteau and Zsuzsanna Lipták, editors, *34th Annual Symposium on Combinatorial Pattern Matching, CPM 2023, June 26-28, 2023, Marne-la-Vallée, France*, volume 259 of *LIPIcs*, pages 10:1–10:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.CPM.2023.10.

10 Gabriele Fici, Giuseppe Romana, Marinella Sciortino, and Cristian Urbina. Morphisms and BWT-run sensitivity. In Pawel Gawrychowski, Filip Mazowiecki, and Michal Skrzypczak, editors, *50th International Symposium on Mathematical Foundations of Computer Science, MFCS 2025, August 25-29, 2025, Warsaw, Poland*, volume 345 of *LIPIcs*, pages 49:1–49:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/LIPIcs.MFCS.2025.49.

11 Andrea Frosini, Ilaria Mancini, Simone Rinaldi, Giuseppe Romana, and Marinella Sciortino. Logarithmic equal-letter runs for BWT of purely morphic words. 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 139–151. Springer, 2022. doi:10.1007/978-3-031-05578-2_11.

12 Peaker Guo, Patrick Eades, Anthony Wirth, and Justin Zobel. Exploiting new properties of string net frequency for efficient computation. In Shunsuke Inenaga and Simon J. Puglisi, editors, *35th Annual Symposium on Combinatorial Pattern Matching, CPM 2024, June 25-27, 2024, Fukuoka, Japan*, volume 296 of *LIPIcs*, pages 16:1–16:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.CPM.2024.16.

13 Peaker Guo and Kaisei Kishi. Net occurrences in Fibonacci and Thue-Morse words. In Paola Bonizzoni and Veli Mäkinen, editors, *36th Annual Symposium on Combinatorial Pattern Matching, CPM 2025, June 17-19, 2025, Milan, Italy*, volume 331 of *LIPIcs*, pages 16:1–16:22. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/LIPIcs.CPM.2025.16.

14 Peaker Guo, Seeun William Umboh, Anthony Wirth, and Justin Zobel. Online computation of string net frequency. In Zsuzsanna Lipták, Edleno Silva de Moura, Karina Figueroa, and Ricardo Baeza-Yates, editors, *String Processing and Information Retrieval - 31st International Symposium, SPIRE 2024, Puerto Vallarta, Mexico, September 23-25, 2024, Proceedings*, volume 14899 of *Lecture Notes in Computer Science*, pages 159–173. Springer, 2024. doi:10.1007/978-3-031-72200-4_12.

15 Stepan Holub and Martin Raska. Binary codes that do not preserve primitivity. *Arch. Formal Proofs*, 2023, 2023. doi:10.1007/s10817-023-09674-2.

16 H.K. Hsiao, Y.T. Yeh, and S.S. Yu. Square-free-preserving and primitive-preserving homomorphisms. *Acta Mathematica Hungarica*, 101(1):113–130, 2003. doi:10.1023/B:AMHU.0000003896.79824.c8.

17 C. C. Huang and S. S. Yu. Prefix-primitivity-preserving homomorphisms. *Discret. Math.*, 308(7):1025–1032, 2008. doi:10.1016/J.DISC.2007.03.057.

18 Lucian Ilie and William F. Smyth. Minimum unique substrings and maximum repeats. *Fundam. Informaticae*, 110(1-4):183–195, 2011. doi:10.3233/FI-2011-536.

19 Shunsuke Inenaga. Faster and simpler online computation of string net frequency. *CoRR*, 2024. doi:10.48550/arXiv.2410.06837.

20 Veikko Keränen. A powerful abelian square-free substitution over 4 letters. *Theor. Comput. Sci.*, 410(38-40):3893–3900, 2009. doi:10.1016/J.TCS.2009.05.027.

21 Kotaro Kimura and Tomohiro I. R-enum revisited: Speedup and extension for context-sensitive repeats and net frequencies. In Philip Bille and Nicola Prezza, editors, *37th Annual Symposium on Combinatorial Pattern Matching, CPM 2026, June 15-17, 2026, Copenhagen, Denmark*, LIPIcs. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026. doi:10.48550/arXiv.2511.11057.

22 Takuya Mieno, Yuta Fujishige, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda. Computing minimal unique substrings for a sliding window. *Algorithmica*, 84(3):670–693, 2022. doi:10.1007/S00453-021-00864-1.

23 Takuya Mieno and Shunsuke Inenaga. Space-efficient online computation of string net occurrences. In Paola Bonizzoni and Veli Mäkinen, editors, *36th Annual Symposium on Combinatorial Pattern Matching, CPM 2025, June 17-19, 2025, Milan, Italy*, volume 331 of *LIPIcs*, pages 23:1–23:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/LIPIcs.CPM.2025.23.

24 Takuya Mieno, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda. Shortest unique substring queries on run-length encoded strings. In Piotr Faliszewski, Anca Muscholl, and Rolf Niedermeier, editors, *41st International Symposium on Mathematical Foundations of Computer Science, MFCS 2016, August 22-26, 2016 - Kraków, Poland*, volume 58 of *LIPIcs*, pages 69:1–69:11. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPIcs.MFCS.2016.69.

25 Takuya Mieno, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda. Tight bounds on the maximum number of shortest unique substrings. In Juha Kärkkäinen, Jakub Radoszewski, and Wojciech Rytter, editors, *28th Annual Symposium on Combinatorial Pattern Matching, CPM 2017, July 4-6, 2017, Warsaw, Poland*, volume 78 of *LIPIcs*, pages 24:1–24:11. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017. doi:10.4230/LIPIcs.CPM.2017.24.

26 Gonzalo Navarro, Carlos Ochoa, and Nicola Prezza. On the approximation ratio of ordered parsings. *IEEE Trans. Inf. Theory*, 67(2):1008–1026, 2021. doi:10.1109/TIT.2020.3042746.

27 Gonzalo Navarro and Cristian Urbina. Repetitiveness measures based on string morphisms. *Theoretical Computer Science*, 1043:115259, 2025. doi:10.1016/j.tcs.2025.115259.

28 OEIS Foundation Inc. The On-Line Encyclopedia of Integer Sequences. Founded by Neil J. A. Sloane. URL: https://oeis.org.

29 Enno Ohlebusch, Thomas Büchler, and Jannik Olbrich. Faster computation of Chinese frequent strings and their net frequencies. In Zsuzsanna Lipták, Edleno Silva de Moura, Karina Figueroa, and Ricardo Baeza-Yates, editors, *String Processing and Information Retrieval - 31st International Symposium, SPIRE 2024, Puerto Vallarta, Mexico, September 23-25, 2024, Proceedings*, volume 14899 of *Lecture Notes in Computer Science*, pages 249–256. Springer, 2024. doi:10.1007/978-3-031-72200-4_19.

30 Gwénaël Richomme. On morphisms preserving infinite Lyndon words. *Discret. Math. Theor. Comput. Sci.*, 9(2), 2007. doi:10.46298/DMTCS.411.

31 Gwénaël Richomme and Patrice Séébold. Characterization of test-sets for overlap-free morphisms. *Discret. Appl. Math.*, 98(1-2):151–157, 1999. doi:10.1016/S0166-218X(99)00118-3.

32 Gwénaël Richomme and Francis Wlazinski. Overlap-free morphisms and finite test-sets. *Discret. Appl. Math.*, 143(1-3):92–109, 2004. doi:10.1016/J.DAM.2003.10.005.

33 G. Rozenberg and A. Salomaa. The mathematical theory of L systems. In Julius T. Tou, editor, *Advances in Information Systems Science: Volume 6*, pages 161–206. Springer US, Boston, MA, 1976. doi:10.1007/978-1-4615-8249-6_4.