

Net Occurrences in Fibonacci and Thue-Morse Words

Peaker Guo ✉ 🏠 

School of Computing and Information Systems, The University of Melbourne, Parkville, Australia

Kaisei Kishi ✉

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

Abstract

A *net occurrence* of a repeated string in a text is an occurrence with unique left and right *extensions*, and the *net frequency* of the string is the number of its net occurrences in the text. Originally introduced for applications in Natural Language Processing, net frequency has recently gained attention for its algorithmic aspects. Guo et al. [CPM 2024] and Ohlebusch et al. [SPIRE 2024] focus on its computation in the offline setting, while Guo et al. [SPIRE 2024], Inenaga [arXiv 2024], and Mieno and Inenaga [CPM 2025] tackle the online counterpart. Mieno and Inenaga also characterize net occurrences in terms of the minimal unique substrings of the text. Additionally, Guo et al. [CPM 2024] initiate the study of net occurrences in Fibonacci words to establish a lower bound on the asymptotic running time of algorithms. Although there has been notable progress in algorithmic developments and some initial combinatorial insights, the combinatorial aspects of net occurrences have yet to be thoroughly examined. In this work, we make two key contributions. First, we confirm the conjecture that each Fibonacci word contains exactly three net occurrences. Second, we show that each Thue-Morse word contains exactly nine net occurrences. To achieve these results, we introduce the notion of *overlapping net occurrence cover*, which narrows down the candidate net occurrences in any text. Furthermore, we provide a precise characterization of occurrences of Fibonacci and Thue-Morse words of smaller order, offering structural insights that may have independent interest and potential applications in algorithm analysis and combinatorial properties of these words.

2012 ACM Subject Classification Mathematics of computing → Combinatorics on words

Keywords and phrases Fibonacci words, Thue-Morse words, net occurrence, net frequency, factorization

Digital Object Identifier 10.4230/LIPIcs.CPM.2025.16

Funding *Peaker Guo*: Supported by an Australian Government Research Training Program Scholarship.

Acknowledgements The authors thank Hideo Bannai and the organizers of the StringMasters workshop at CPM 2024 for initiating the collaboration between the authors during the workshop. The authors also thank Shunsuke Inenaga and William Umboh for their helpful advice.

1 Introduction

The work by Axel Thue at the beginning of the 20th century marked the beginning of the field of combinatorics on words [6]. Central to the field are two key objects that have attracted extensive research: Fibonacci words and Thue-Morse words [30]. These objects are remarkable for their rich combinatorial properties and applications in seemingly unrelated fields beyond combinatorics on words. Fibonacci words, for instance, have been used to establish lower bounds and analyze behaviors of string algorithms [24], while Thue-Morse words appear in diverse areas such as group theory, physics, and even chess [2]. They have also been used to prove properties related to repetitiveness measures [3, 5, 11, 27, 32].



© Peaker Guo and Kaisei Kishi;

licensed under Creative Commons License CC-BY 4.0

36th Annual Symposium on Combinatorial Pattern Matching (CPM 2025).

Editors: Paola Bonizzoni and Veli Mäkinen; Article No. 16; pp. 16:1–16:22

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Another key aspect of combinatorics on words involves identifying significant strings in a text. Different definitions of significance lead to different problem formulations. These significant strings could be repetitions [10], tandem repeats [20], or runs [4]. There is also a rich literature on the study of these significant strings in Fibonacci and Thue-Morse words [7, 12, 13, 22]. For many applications, frequency serves as a basis for significance measure. However, frequency alone can be misleading, as it may be inflated by occurrences of longer repeated strings. Consider the text `the_theoretical_the` as an example. The string `the` is the most frequent string of length three, but this is due to the fact that two of its occurrences are contained by the longer repeated string `_the`.

To address this issue, Lin and Yu [28, 29] introduced the notion of *net frequency* (NF), motivated by Natural Language Processing tasks. As reconceptualized by Guo et al. [18], a *net occurrence* of a repeated string in a text is an occurrence with unique left and right extensions, and the NF of the string is the number of its net occurrences in the text. In the earlier example, only the first occurrence of `the` is a net occurrence, reflecting the only occurrence that is not contained by a longer repeated string.

There has been a recent surge of interest in the computation of NF. Guo et al. [18] and Ohlebusch et al. [33] focus on the offline setting, while Guo et al. [19], Inenaga [23], and Mieno and Inenaga [31] extend the computation to the online setting. Mieno and Inenaga also characterize net occurrences in terms of the minimal unique substrings of the text. Additionally, Guo et al. [18] study net occurrences in Fibonacci words to establish a lower bound on the asymptotic running time of algorithms. Despite these advances, the combinatorial aspect of net occurrences has yet to be thoroughly investigated. It has been shown that there are *at least* three net occurrences in each Fibonacci word [18]. However, proving that these are the *only* three is more challenging and was only conjectured. Meanwhile, the net occurrences in each Thue-Morse word had not been investigated before – both of which we address in this work.

Our results. In this work, our main contribution is twofold. First, we confirm the conjecture by Guo et al. [18] that there are exactly three net occurrences in each Fibonacci word (Theorem 34). Second, we show that there are exactly nine net occurrences in each Thue-Morse word (Theorem 41). To achieve these results, we first introduce the concept of an *overlapping net occurrence cover*, which drastically reduces the number of occurrences that need to be examined when proving certain net occurrences are the only ones (Lemma 12). Additionally, we provide a precise characterization of occurrences of smaller-order Fibonacci and Thue-Morse words (Theorem 17 and Theorem 19). These findings could also be of independent interest, providing tools and insights for analyzing algorithms and exploring the combinatorial properties of these words. For example, they lead to methods to count the smaller-order occurrences (Corollary 18 and Corollary 21).

Other related work. Occurrences of Fibonacci and Thue-Morse words of smaller order have been previously studied. For Fibonacci words, these occurrences have been shown to be related to the Fibonacci representation of positive integers [22, 35]. For Thue-Morse words, these occurrences have been investigated using the binary representation of numbers and properties of the compact directed acyclic word graph (CDAWG) of each Thue-Morse word [34]. We emphasize that our work addresses occurrences of Fibonacci and Thue-Morse words of smaller order from a different angle than prior work: we provide a recurrence relation that precisely characterizes the occurrences, bypassing the need for other representations.

2 Preliminaries

Strings. Throughout, we consider the binary alphabet $\Sigma := \{\mathbf{a}, \mathbf{b}\}$. A *string* is an element of Σ^* . The length of a string S is denoted as $|S|$. Let ϵ denote the *empty string* of length 0. We use $S[i]$ to denote the i^{th} character of a string S . Let $[n]$ denote the set $\{1, 2, \dots, n\}$. Let ST be the *concatenation* of two strings, S and T . A substring of a string T of length n , starting at position $i \in [n]$ and ending at position $j \in [n]$, is written as $T[i \dots j]$. A substring $T[1 \dots j]$ is called a *prefix* of T , while $T[i \dots n]$ is called a *suffix* of T . A substring S of T is a *proper* substring if $S \neq T$. An *occurrence* in the text T of length n is a pair of starting and ending positions $(i, j) \in [n] \times [n]$. We say (i, j) is an *occurrence of string* S if $S = T[i \dots j]$, and i is an occurrence of S if $S = T[i \dots i + |S| - 1]$. An occurrence (i', j') is a *sub-occurrence* of (i, j) if $i \leq i' \leq j' \leq j$. An occurrence (i, j) is a *super-occurrence* of (i', j') if (i', j') is a sub-occurrence of (i, j) . Moreover, (i', j') is a *proper sub-occurrence* (or super-occurrence) of (i, j) if (i', j') is a sub-occurrence (or super-occurrence) of (i, j) and $i \neq i'$ or $j' \neq j$. Two occurrences (i, j) and (i', j') *overlap* if there exists a position k such that $i \leq k \leq j$ and $i' \leq k \leq j'$. For a non-empty string S , a sequence of non-empty strings $\mathcal{F} = (x_k)_{k=1}^m = (x_1, x_2, \dots, x_m)$ is referred to as a *factorization* of S if $S = x_1 x_2 \dots x_m$. Each string x_k is called a *factor* of \mathcal{F} . The *size* of \mathcal{F} , denoted by $|\mathcal{F}|$, is the number of factors in the factorization.

Net frequency and net occurrences. In a text T , the net frequency (NF) of a unique string in T is defined to be zero. The NF of a repeated string is the number of *net occurrences* in T .

► **Definition 1** (Net occurrence [18]). *In a text T , an occurrence (i, j) is a net occurrence if the corresponding string $T[i \dots j]$ is repeated, while both left extension $T[i - 1 \dots j]$ and right extension $T[i \dots j + 1]$ are unique. When $i = 1$, $T[i - 1 \dots j]$ is assumed to be unique; when $j = |T|$, $T[i \dots j + 1]$ is assumed to be unique.*

For an occurrence (i, j) in text T , we refer to $T[i - 1]$ and $T[j + 1]$ as the left and right extension *characters* of (i, j) , respectively. For a string S occurring in T , we say $x, y \in \Sigma$ are left and right extension *characters* of S if both strings xS and Sy also occur in T .

Fibonacci words. Let F_i denote the (finite) *Fibonacci word of order* i where $F_1 := \mathbf{b}$, $F_2 := \mathbf{a}$, and $F_i := F_{i-1} F_{i-2}$ for each $i \geq 3$. Let $f_i := |F_i|$ be the length of the Fibonacci word of order i , which is also the i^{th} Fibonacci number. We next review two useful results on F_i .

► **Lemma 2** ([12]). *F_i only occurs twice in $F_i F_i$.*

► **Lemma 3** ([32]). *The strings \mathbf{aaa} and \mathbf{bb} do not occur in F_i .*

The following result can be readily derived by repeatedly applying the definition of F_i .

► **Observation 4.** *For $1 \leq k \leq i$, there is a factorization of F_i where each factor is either F_k or F_{k+1} .*

For example, for $k = i - 2 \dots i - 5$, we have the following factorizations: $F_i = F_{i-1} F_{i-2} = F_{i-2} F_{i-3} F_{i-2} = F_{i-3} F_{i-4} F_{i-3} F_{i-3} F_{i-4} = F_{i-4} F_{i-5} F_{i-4} F_{i-4} F_{i-5} F_{i-4} F_{i-5} F_{i-4}$.

Thue-Morse words. For a binary string S , let \bar{S} denote the string obtained by simultaneously replacing each \mathbf{a} with \mathbf{b} and each \mathbf{b} with \mathbf{a} . Let \mathcal{T}_i be the (finite) *Thue-Morse word of order* i where $\mathcal{T}_1 := \mathbf{a}$ and $\mathcal{T}_i := \mathcal{T}_{i-1} \bar{\mathcal{T}}_{i-1}$ for each $i \geq 2$. Let $\tau_i := |\mathcal{T}_i| = 2^{i-1}$ be the length of the Thue-Morse word of order i . We next review two properties of each \mathcal{T}_i .

$\begin{array}{cccccccccccc} 1 & 2 & & 4 & & 6 & 7 & & 9 & & & 14 \\ \hline a & b & b & a & b & b & a & b & b & b & a & b & b & a \end{array}$

■ **Figure 1** An example for Definition 11. The set $\{(1, 6), (4, 9), (9, 14)\}$ is an ONOC, with each of its net occurrences underlined in blue; $\{(4, 6), (9, 9)\}$ is the corresponding set of BNSOs. Note that $(2, 7)$ is a net occurrence outside of this ONOC, underlined in orange.

- ▶ **Lemma 5** (Overlap-free [30]). \mathcal{T}_i has no overlapping occurrences of the same string.
- ▶ **Lemma 6** (Cube-free [30]). \mathcal{T}_i does not contain any string of the form xxx where x is a non-empty string.

The following result can be directly derived by repeatedly applying the definition of \mathcal{T}_i .

- ▶ **Observation 7.** For each $i \geq 2$ and $1 \leq j \leq i$, there is a factorization of \mathcal{T}_i where each factor is either $\mathcal{T}_{i-(j-1)}$ or $\overline{\mathcal{T}_{i-(j-1)}}$.

For example, for $2 \leq j \leq 3$, $\mathcal{T}_i = \mathcal{T}_{i-1} \overline{\mathcal{T}_{i-1}} = \mathcal{T}_{i-2} \overline{\mathcal{T}_{i-2}} \overline{\mathcal{T}_{i-2}} \mathcal{T}_{i-2}$. Figure 3 illustrates larger value of j . Also note that this result is analogous to Observation 4.

3 Overlapping Net Occurrence Cover

This section lays the foundation to prove the main results of this paper in the subsequent sections. Specifically, we aim to develop tools to show that certain net occurrences are the only ones in a text. To achieve this, we first provide two characteristics for non-net occurrences. The proofs in this section are presented in Appendix A.

- ▶ **Observation 8.** In a text T , if an occurrence (s, e) is a proper super-occurrence of a net occurrence, then (s, e) is not a net occurrence.

- ▶ **Observation 9.** In a text T , if an occurrence (s, e) is a proper sub-occurrence of a net occurrence, then (s, e) is not a net occurrence.

▶ **Remark 10.** In a text, both a string and its substring can have positive NF, for example, in $abaababaabaab$, both $abaaba$ and $abaab$ have positive NF. However, this relationship does not hold for an occurrence and its sub-occurrence, as shown in the above two observations.

To show that a given set of net occurrences in T are the only ones in T , the above two observations allow us to ignore any occurrence that is either a sub-occurrence or a super-occurrence of a net occurrence. To fully use these two observations, we focus on the case when the given net occurrences “overlap” one another and collectively “cover” the text. Consequently, the only occurrences that need to be explicitly examined are the super-occurrences of those corresponding to the “overlapping regions” of these net occurrences. To formalize this, we introduce the following definition and lemma.

- ▶ **Definition 11** (ONOC and BNSO). Consider a text T and a set of c net occurrences in T : $\mathcal{C} = \{(i_1, j_1), (i_2, j_2), \dots, (i_c, j_c)\}$. We say \mathcal{C} is an overlapping net occurrence cover (ONOC) of T if $i_1 = 1$, $i_{k+1} \leq j_k$ for $1 \leq k \leq c - 1$, and $j_c = n$. Each occurrence in the set $\{(i_2, j_1), (i_3, j_2), \dots, (i_c, j_{c-1})\}$ is a bridging net sub-occurrence (BNSO) of \mathcal{C} .

An example of Definition 11 is shown in Figure 1.

► **Lemma 12.** *For a text T , if there exists an ONOC \mathcal{C} of T such that \mathcal{C} does not contain all the net occurrences in T , then each net occurrence in T outside of \mathcal{C} must be a super-occurrence of $(i - 1, j + 1)$, where (i, j) is a BNSO of \mathcal{C} .*

In the example in Figure 1, note that net occurrence $(2, 7)$ is indeed a super-occurrence of $(4 - 1, 6 + 1)$, where $(4, 6)$ is a BNSO.

In Section 6 and Section 7, we apply Lemma 12 in three steps. First, for a Fibonacci or Thue–Morse word, we show that an ONOC exists. Next, we examine the set of BNSOs of the ONOC. Finally, we prove that no super-occurrence of $(i - 1, j + 1)$ (where (i, j) is a BNSO) is a net occurrence, thus concluding that the ONOC already contains all the net occurrences in the text.

4 Occurrences of Fibonacci Words of Smaller Order

We study the occurrences of F_{i-j} in F_i for appropriate i and j . These results will help us prove the only net occurrences in F_i in Section 6 and may also be of independent interest.

When $j = 1$, with $F_i = F_{i-1} F_{i-2}$, we have one occurrence of F_{i-1} at position 1. The following result shows that this is the only one.

► **Lemma 13** ([32]). *F_{i-1} only occurs at position 1 in F_i for $i \geq 3$.*

The two factorizations in the following result reveal three occurrences of F_{i-2} in F_i .

► **Observation 14** ([18]). *For each $i \geq 6$,*

$$F_i = F_{i-2} F_{i-3} F_{i-2} \tag{1}$$

$$F_i = F_{i-2} F_{i-2} F_{i-5} F_{i-4}. \tag{2}$$

The following result confirms that these are the only three.

► **Lemma 15** ([32]). *F_{i-2} only occurs at positions 1, $f_{i-2} + 1$, and $f_{i-1} + 1$ in F_i for $i \geq 6$.*

We next provide the result when $j = 3$ and $i \geq 7$.

► **Lemma 16.** *F_{i-3} only occurs at positions 1, $f_{i-3} + 1$, $f_{i-2} + 1$, and $f_{i-1} + 1$ in F_i .*

Proof. From Lemma 15, notice that the second occurrence of F_{i-2} follows immediately after the first occurrence, and the second and the third occurrences of F_{i-2} overlap. Then, based on Equations (1)–(2), we consider the following three cases.

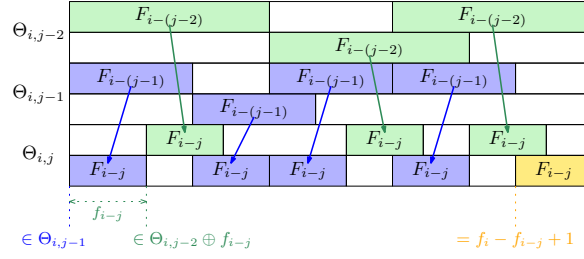
Case 1 F_{i-3} occurs within F_{i-2} . From Lemma 13, F_{i-3} only occurs at position 1 in F_{i-2} .

Thus, using Lemma 15, the only occurrences of F_{i-3} within F_{i-2} in F_i are at positions 1, $f_{i-2} + 1$, and $f_{i-1} + 1$.

Case 2 F_{i-3} occurs across the boundary of $F_{i-2} F_{i-3}$. Again from Lemma 15, the only occurrences of F_{i-3} within $F_{i-2} F_{i-3} = F_{i-1}$ are at positions 1, $f_{i-3} + 1$, and $f_{i-2} + 1$. Note that the occurrence at position $f_{i-3} + 1$ is the boundary-crossing one: we apply Equation (2) on F_{i-1} and obtain $F_{i-2} F_{i-3} = F_{i-3} F_{i-3} F_{i-6} F_{i-5}$.

Case 3 F_{i-3} occurs across the boundary of $F_{i-3} F_{i-2}$. Note that $F_{i-3} F_{i-2} = F_{i-3} F_{i-3} F_{i-4}$. Using Lemma 2, F_{i-3} does not occur in $F_{i-3} F_{i-3}$ and thus does not occur across the boundary of $F_{i-3} F_{i-2}$. ◀

We now present the main result of the section, illustrated in Figure 2. Before that, we define the following. For a set of integers A and another integer i , $A \oplus i$ denotes the set $\{a + i : a \in A\}$. We write $\max(A)$ for the maximum element of set A .



■ **Figure 2** An illustration of Theorem 17 when $j = 4$. Each row depicts a factorization of F_i with relevant factors highlighted in colors. The top two, middle two, and bottom two rows correspond to sets $\Theta_{i,j-2}$, $\Theta_{i,j-1}$ and $\Theta_{i,j}$, respectively. Each green and blue occurrence of F_{i-j} is introduced by an occurrence of $F_{i-(j-2)}$ and $F_{i-(j-1)}$, respectively. The yellow occurrence is the rightmost one.

► **Theorem 17.** Let $\Theta_{i,j}$ denote the set of the starting positions of the occurrences of F_{i-j} in F_i . Then, $\Theta_{i,0} = \Theta_{i,1} = \{1\}$, and for $2 \leq j \leq i-4$,

- when j is even, $\Theta_{i,j} = \Theta_{i,j-1} \cup (\Theta_{i,j-2} \oplus f_{i-j}) \cup \{f_i - f_{i-j} + 1\}$, where the three sets in the union are mutually disjoint, and $\max(\Theta_{i,j}) = f_i - f_{i-j} + 1$;
- when j is odd, $\Theta_{i,j} = \Theta_{i,j-1} \cup (\Theta_{i,j-2} \oplus f_{i-j})$ where the two sets in the union are disjoint, and $\max(\Theta_{i,j}) = f_i - f_{i-(j-1)} + 1$.

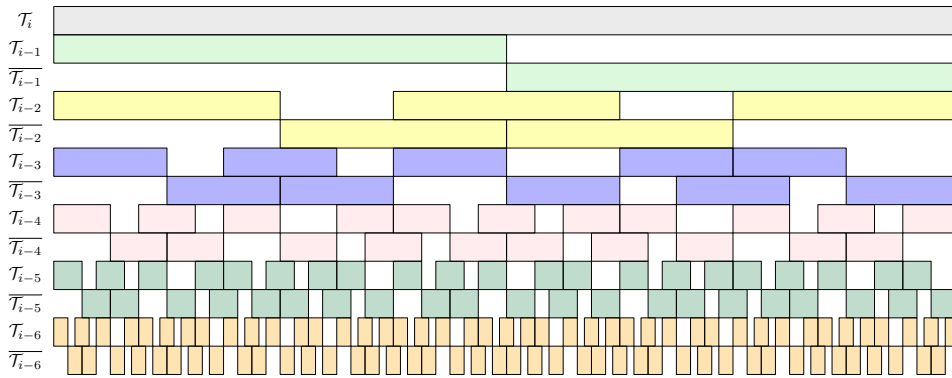
Proof. We proceed by induction on j .

Base cases. When $j = 0$, we have $\Theta_{i,0} = \{1\}$ trivially. When $j = 1$, it follows from Lemma 13 that $\Theta_{i,1} = \{1\}$. When $j = 2$, we obtain $\Theta_{i,2} = \{1, f_{i-2} + 1, f_{i-1} + 1\}$ by Lemma 15. Thus, $\Theta_{i,2} = \Theta_{i,1} \cup (\Theta_{i,0} \oplus f_{i-2}) \cup \{f_i - f_{i-2} + 1\}$, where the three sets are mutually disjoint, and $\max(\Theta_{i,2}) = f_{i-1} + 1 = f_i - f_{i-2} + 1$. Next, when $j = 3$, we have $\Theta_{i,3} = \{1, f_{i-3} + 1, f_{i-2} + 1, f_{i-1} + 1\}$ by Lemma 16. Hence, $\Theta_{i,3} = \Theta_{i,2} \cup (\Theta_{i,1} \oplus f_{i-3})$, $\Theta_{i,2} \cap (\Theta_{i,1} \oplus f_{i-3}) = \emptyset$, and $\max(\Theta_{i,3}) = f_{i-1} + 1 = f_i - f_{i-(3-1)} + 1$.

Inductive step. For each $4 \leq k \leq i-4$, assume the claim holds for $j = k-2$ and $k-1$, and we now prove the claim for $j = k$. We first prove the claim for even k . Define $\Lambda_{i,k} := \Theta_{i,k-1} \cup (\Theta_{i,k-2} \oplus f_{i-k}) \cup \{f_i - f_{i-k} + 1\}$. We aim to show that $\Theta_{i,k} = \Lambda_{i,k}$ by showing $\Lambda_{i,k} \subset \Theta_{i,k}$ and $\Theta_{i,k} \subset \Lambda_{i,k}$. Before we proceed, note that $F_{i-(k+2)}, F_{i-(k+1)}, F_{i-k}, F_{i-(k-1)}, F_{i-(k-2)}$ are consecutive Fibonacci words of *increasing* orders.

To prove that $\Lambda_{i,k} \subset \Theta_{i,k}$, we will show that each set in the union defining $\Lambda_{i,k}$ is contained in $\Theta_{i,k}$. Note that $\Theta_{i,k-1} \subset \Theta_{i,k}$ because F_{i-k} is a prefix of $F_{i-(k-1)} = F_{i-k} F_{i-(k+1)}$. Next, we have $(\Theta_{i,k-2} \oplus f_{i-k}) \subset \Theta_{i,k}$ because F_{i-k} occurs at position $f_{i-k} + 1$ in $F_{i-(k-2)} = F_{i-k} F_{i-k} F_{i-(k+3)} F_{i-(k+2)}$ where this factorization can be derived similarly to Equation (2). Lastly, by the induction hypothesis on $j = k-2$, we have $f_i - f_{i-(k-2)} + 1 \in \Theta_{i,k-2}$ and it is the rightmost occurrence of $F_{i-(k-2)}$ in F_i . Consider the factorization $F_{i-(k-2)} = F_{i-k} F_{i-(k+1)} F_{i-k}$, which can be derived similarly to Equation (1). Note that the rightmost occurrence of F_{i-k} in $F_{i-(k-2)}$ is at position $(f_i - f_{i-(k-2)} + 1) + (f_{i-k} + f_{i-(k+1)}) = (f_i - (f_{i-k} + f_{i-k} + f_{i-(k+1)}) + 1) + (f_{i-k} + f_{i-(k+1)}) = f_i - f_{i-k} + 1$ in F_i . Thus, $f_i - f_{i-k} + 1 \in \Theta_{i,k}$.

Next, we prove $\Theta_{i,k} \subset \Lambda_{i,k}$ by showing that each occurrence of F_{i-k} is in $\Lambda_{i,k}$. By Observation 4, there is a factorization of F_i where each factor is either $F_{i-(k-1)}$ or F_{i-k} . We now examine the occurrences of F_{i-k} based on this factorization. First, when there is an occurrence of F_{i-k} within $F_{i-(k-1)}$, this occurrence is in $\Theta_{i,k-1} \subset \Lambda_{i,k}$. Next, when there is an occurrence of F_{i-k} (underlined) across the boundary of $F_{i-(k-1)} F_{i-(k-1)} = F_{i-k} \underline{F_{i-k}} F_{i-(k+3)} F_{i-(k+2)} F_{i-(k+1)}$ or across the boundary of $F_{i-(k-1)} F_{i-k} = F_{i-(k-2)} = F_{i-k} \underline{F_{i-k}} F_{i-(k+3)} F_{i-(k+2)}$, then, by the fact that F_{i-k} only occurs at positions $1, f_{i-k} + 1$,



■ **Figure 3** An illustration of the occurrences of \mathcal{T}_{i-j} and $\overline{\mathcal{T}_{i-j}}$ in \mathcal{T}_i for $1 \leq j \leq 6$.

and $f_{i-(k-1)} + 1$ in $F_{i-(k-2)}$ (a direct generalization of Lemma 15), this occurrence of F_{i-k} must be in $(\Theta_{i,k-2} \oplus f_{i-k}) \subset \Lambda_{i,k}$. Finally, observe that there does not exist an occurrence of F_{i-k} across the boundary of $F_{i-k} F_{i-(k-1)} = F_{i-k} F_{i-k} F_{i-(k-1)}$ or across the boundary of $F_{i-k} F_{i-k}$, because otherwise, this would contradict Lemma 2.

Now, consider a position $x \in (\Theta_{i,k-2} \oplus f_{i-k})$. Assume, by contradiction, that $x \in \Theta_{i,k-1}$, then, we have $F_{i-(k-2)} = F_{i-k} F_{i-(k-1)}$, which contradicts $F_{i-(k-2)} = F_{i-(k-1)} F_{i-k} \neq F_{i-k} F_{i-(k-1)}$. This is analogous to $F_{i-1} F_{i-2} \neq F_{i-2} F_{i-1}$, which follows from the “near-commutative property” of Fibonacci words [26]. Thus, $\Theta_{i,k-1} \cap (\Theta_{i,k-2} \oplus f_{i-k}) = \emptyset$. Next, by the induction hypothesis, $\max(\Theta_{i,k-1} \cup \Theta_{i,k-2}) = f_i - f_{i-(k-2)} + 1$. Note that $(f_i - f_{i-(k-2)} + 1) + f_{i-k} < f_i - f_{i-k} + 1$. Therefore, $f_i - f_{i-k} + 1 \notin \Theta_{i,k-1} \cup (\Theta_{i,k-2} \oplus f_{i-k})$ and $\max(\Theta_{i,k}) = f_i - f_{i-k} + 1$.

The proof for odd k is very similar to even k with the difference being that we do not need to consider $f_i - f_{i-k} + 1$ for odd k . ◀

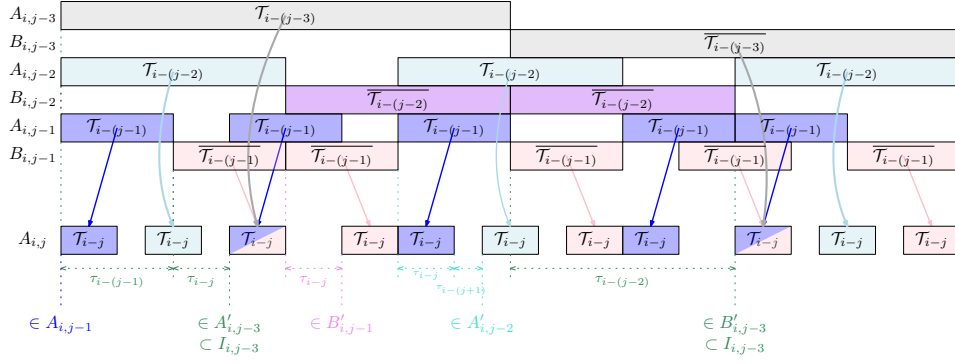
In the following result, the case where $0 \leq j \leq i - 4$ has been addressed in [35], while the case where $i - 3 \leq j \leq i - 1$ is straightforward. Our characterization in Theorem 17 can offer an alternative simpler proof for this result.

▶ **Corollary 18.** Consider F_i and $0 \leq j \leq i - 1$. Let $\theta_{i,j}$ denote the number of occurrences of F_{i-j} in F_i , and define $f_{-1} := 1$ and $f_0 := 0$ for convenience. Then,

$$\theta_{i,j} = \begin{cases} f_{j+2} - (j \bmod 2) & \text{if } 0 \leq j \leq i - 4; \\ f_{j+1} & \text{if } i - 3 \leq j \leq i - 2; \\ f_{j-1} & \text{if } j = i - 1. \end{cases}$$

5 Occurrences of Thue-Morse Words of Smaller Order

We study the occurrences of \mathcal{T}_{i-j} and $\overline{\mathcal{T}_{i-j}}$ in each \mathcal{T}_i for appropriate i and j (the occurrences are shown in Figure 3 for $1 \leq j \leq 6$). These results will help us identify the net occurrences in each \mathcal{T}_i in Section 6 and may also be of independent interest. We now present the main result of the section, illustrated in Figure 4.



■ **Figure 4** An illustration of Theorem 19. Each dark blue, pink, and light blue occurrence of \mathcal{T}_{i-j} is introduced by an occurrence of $\mathcal{T}_{i-(j-1)}$, $\overline{\mathcal{T}_{i-(j-1)}}$, and $\mathcal{T}_{i-(j-2)}$ respectively. Each occurrence of \mathcal{T}_{i-j} that is both dark blue and pink indicates that it is introduced by both an occurrence of $\mathcal{T}_{i-(j-1)}$ and an occurrence of $\overline{\mathcal{T}_{i-(j-1)}}$.

► **Theorem 19.** For each $i \geq 2$ and $0 \leq j \leq i-1$, let $A_{i,j}$ and $B_{i,j}$ denote the set of the starting positions of the occurrences of \mathcal{T}_{i-j} and $\overline{\mathcal{T}_{i-j}}$ in \mathcal{T}_i , respectively. Then, $A_{i,0} = A_{i,1} = \{1\}$. For each $j \geq 2$, we define

$$\begin{aligned}
 B'_{i,j-1} &:= B_{i,j-1} \oplus \tau_{i-j}, & A'_{i,j-2} &:= A_{i,j-2} \oplus (\tau_{i-j} + \tau_{i-(j+1)}), \\
 I_{i,j-3} &:= \begin{cases} \emptyset, & j = 2, \\ A'_{i,j-3} \cup B'_{i,j-3}, & j \geq 3, \text{ where} \end{cases} \\
 A'_{i,j-3} &:= A_{i,j-3} \oplus (\tau_{i-(j-1)} + \tau_{i-j}), & B'_{i,j-3} &:= B_{i,j-3} \oplus \tau_{i-(j-2)}.
 \end{aligned}$$

Then, $A_{i,j} = A_{i,j-1} \cup B'_{i,j-1} \cup A'_{i,j-2}$ with

$$A_{i,j-1} \cap B'_{i,j-1} = I_{i,j-3}, \quad A_{i,j-1} \cap A'_{i,j-2} = \emptyset, \quad \text{and} \quad B'_{i,j-1} \cap A'_{i,j-2} = \emptyset.$$

Proof. We proceed by induction on j .

Base cases. When $j = 0$, the claim holds trivially. When $j = 1$, note that \mathcal{T}_{i-1} only occurs at position 1 because it cannot occur at position $\tau_{i-1} + 1$ (where $\overline{\mathcal{T}_{i-1}}$ occurs), and any other occurrences of \mathcal{T}_{i-1} would overlap with its occurrence at position 1, contradicting Lemma 5. When $j = 2$, observe that $A_{i,2-1} = \{1\}$, $B'_{i,2-1} = \{\tau_{i-1} + \tau_{i-2} + 1\}$ and $A'_{i,2-2} = \{\tau_{i-2} + \tau_{i-3} + 1\}$ are mutually disjoint. Further, there are no occurrences of \mathcal{T}_{i-2} outside of $A_{i,2} = A_{i,1} \cup B'_{i,1} \cup A'_{i,0}$ because any such occurrences would contradict Lemma 5.

Inductive step. For each $3 \leq k \leq i-1$, assume the claim holds for $j = k-3$, $k-2$, and $k-1$ and we now prove the claim for $j = k$. Define $V_{i,k} := A_{i,k-1} \cup B'_{i,k-1} \cup A'_{i,k-2}$. We prove $A_{i,k} = V_{i,k}$ by showing $A_{i,k} \subset V_{i,k}$ and $V_{i,k} \subset A_{i,k}$.

To prove $V_{i,k} \subset A_{i,k}$, we will show that each set in the union defining $V_{i,k}$ is contained in $A_{i,k}$. Clearly, $A_{i,k-1} \subset A_{i,k}$ because \mathcal{T}_{i-j} is a prefix of $\overline{\mathcal{T}_{i-(j-1)}} = \overline{\mathcal{T}_{i-j} \mathcal{T}_{i-j}}$. Similarly, $B'_{i,k-1} \subset A_{i,k}$ because \mathcal{T}_{i-j} is a suffix of $\overline{\mathcal{T}_{i-(j-1)}} = \overline{\mathcal{T}_{i-j} \mathcal{T}_{i-j}}$. Lastly, $A'_{i,k-2} \subset A_{i,k}$ because \mathcal{T}_{i-j} occurs at position $\tau_{i-k} + \tau_{i-(k+1)}$ of

$$\mathcal{T}_{i-(k-2)} = \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-(k+1)}} \mathcal{T}_{i-k} \mathcal{T}_{i-(k+1)} \mathcal{T}_{i-k} \tag{3}$$

Next, we prove $A_{i,k} \subset V_{i,k}$ by showing that each occurrence of \mathcal{T}_{i-k} is in $V_{i,k}$. By Observation 7, there is a factorization of \mathcal{T}_i where each factor is either $\mathcal{T}_{i-(k-1)}$ or $\overline{\mathcal{T}_{i-(k-1)}}$. We thus consider the following four cases.

- Case 1** When \mathcal{T}_{i-k} occurs within $\mathcal{T}_{i-(k-1)}$ $\mathcal{T}_{i-(k-1)} = \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}}$, by the overlap-free property (Lemma 5), positions 1 and $\tau_{i-(k-1)} + 1$ are the only two occurrences of \mathcal{T}_{i-k} . They are both contained in $A_{i,k-1}$, while the latter is also in $B'_{i,k-1}$. (The overlap-free property will be used similarly in the remaining three cases.)
- Case 2** When \mathcal{T}_{i-k} occurs within $\overline{\mathcal{T}_{i-(k-1)}} \mathcal{T}_{i-(k-1)} = \overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k}$, positions $\tau_{i-k} + 1$ and $\tau_{i-(k-1)} + \tau_{i-k} + 1$ are the only two occurrences of \mathcal{T}_{i-k} . They are both contained in $B'_{i,k-1}$, while the former is also in $A_{i,k-1}$.
- Case 3** When \mathcal{T}_{i-k} occurs within $\mathcal{T}_{i-(k-1)} \overline{\mathcal{T}_{i-(k-1)}} = \mathcal{T}_{i-(k-2)}$, by Equation (3), positions 1, $\tau_{i-k} + \tau_{i-(k+1)} + 1$ and $\tau_{i-(k-1)} + \tau_{i-k} + 1$ are the only occurrences of \mathcal{T}_{i-k} : the first and third are both contained in $A_{i,k-1}$, the second is in $A'_{i,k-2}$, and the third is also in $B'_{i,k-1}$.
- Case 4** When \mathcal{T}_{i-k} occurs within $\overline{\mathcal{T}_{i-(k-1)}} \mathcal{T}_{i-(k-1)} = \overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k} \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}}$, position τ_{i-k} and $\tau_{i-(k-1)}$ are the only two occurrences of \mathcal{T}_{i-k} . The former is contained in $B'_{i,k-1}$, while the latter is in $A_{i,k-1}$.

After examining the above four cases, we conclude that $A_{i,k} \subset V_{i,k}$, and thus $A_{i,k} = V_{i,k}$. Next we will prove $A_{i,k-1} \cap B'_{i,k-1} = I_{i,k-3}$ by showing $A_{i,k-1} \cap B'_{i,k-1} \subset I_{i,k-3}$ and $I_{i,k-3} \subset A_{i,k-1} \cap B'_{i,k-1}$. Recall that $I_{i,k-3} := A'_{i,k-3} \cup B'_{i,k-3}$.

First, we prove $A_{i,k-1} \cap B'_{i,k-1} \subset I_{i,k-3}$ by establishing that if an occurrence of \mathcal{T}_{i-k} is in $A_{i,k-1} \cap B'_{i,k-1}$, then this occurrence is in $I_{i,k-3}$. First observe that in Cases 1–2, some occurrences of \mathcal{T}_{i-k} are contained in both $A_{i,k-1}$ and $B'_{i,k-1}$. By Lemma 5, $\overline{\mathcal{T}_{i-(k-3)}}$ and $\mathcal{T}_{i-(k-3)}$ do not overlap in \mathcal{T}_i , it follows that, for each occurrence of $\overline{\mathcal{T}_{i-(k-3)}}$ in \mathcal{T}_i , there is only one occurrence of \mathcal{T}_{i-k} contained in $B'_{i,k-3}$. Similarly, for each occurrence of $\mathcal{T}_{i-(k-3)}$ in \mathcal{T}_i , there is only one occurrence of \mathcal{T}_{i-k} contained in $A'_{i,k-3}$. Now, consider the factorizations:

$$\begin{aligned} \overline{\mathcal{T}_{i-(k-3)}} &= \overline{\mathcal{T}_{i-(k-1)}} \mathcal{T}_{i-(k-1)} \overline{\mathcal{T}_{i-(k-1)}} \mathcal{T}_{i-(k-1)} \overline{\mathcal{T}_{i-(k-1)}}, \text{ and} \\ \mathcal{T}_{i-(k-3)} &= \mathcal{T}_{i-(k-1)} \overline{\mathcal{T}_{i-(k-1)}} \mathcal{T}_{i-(k-1)} \overline{\mathcal{T}_{i-(k-1)}} \mathcal{T}_{i-(k-1)}. \end{aligned}$$

Observe that the set of occurrences of \mathcal{T}_{i-k} in Case 1 is a subset of $B'_{i,k-3}$ since $\overline{\mathcal{T}_{i-(k-1)}} \mathcal{T}_{i-(k-1)}$ occurs at position $\tau_{i-(k-1)} + 1$ in $\overline{\mathcal{T}_{i-(k-3)}}$. Similarly, the set of occurrences of \mathcal{T}_{i-k} in Case 2 is a subset of $A'_{i,k-3}$ since $\mathcal{T}_{i-(k-1)} \overline{\mathcal{T}_{i-(k-1)}}$ occurs at position $\tau_{i-(k-1)} + 1$ in $\mathcal{T}_{i-(k-3)}$.

Next, we prove $I_{i,k-3} \subset A_{i,k-1} \cap B'_{i,k-1}$ by contraposition. Specifically, instead of directly showing that “if an occurrence of \mathcal{T}_{i-k} is in $I_{i,k-3}$, then this occurrence is in $A_{i,k-1} \cap B'_{i,k-1}$ ”, we prove the equivalent contrapositive: “if an occurrence of \mathcal{T}_{i-k} is not in $A_{i,k-1} \cap B'_{i,k-1}$, then this occurrence is not in $I_{i,k-3}$ ”. First observe that \mathcal{T}_{i-k} occurs at position 1 in $\mathcal{T}_{i-(k-1)} = \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}}$ and occurs at position $\tau_{i-k} + 1$ in $\overline{\mathcal{T}_{i-(k-1)}} = \overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k}$. Next, occurrences of $\mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}}$ and $\overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k}$ overlap in \mathcal{T}_i to form occurrences of $\overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}}$ (see Cases 1–2). Hence, if an occurrence of \mathcal{T}_{i-k} is not in $A_{i,k-1} \cap B'_{i,k-1}$, then this occurrence is not at position $\tau_{i-k} + 1$ in $\overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}}$. Next, consider the factorizations:

$$\begin{aligned} \overline{\mathcal{T}_{i-(k-3)}} &= \overline{\mathcal{T}_{i-(k-1)}} \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}} \overline{\mathcal{T}_{i-(k-1)}}, \text{ and} \\ \mathcal{T}_{i-(k-3)} &= \mathcal{T}_{i-(k-1)} \overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k} \mathcal{T}_{i-(k-1)}. \end{aligned}$$

Since $\overline{\mathcal{T}_{i-(k-3)}}$ and $\mathcal{T}_{i-(k-3)}$ do not overlap in \mathcal{T}_i , we know that $\overline{\mathcal{T}_{i-k}} \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}}$ only occurs at position $\tau_{i-(k-1)} + \tau_{i-k} + 1$ in $\overline{\mathcal{T}_{i-(k-3)}}$ and at position $\tau_{i-(k-1)} + 1$ in $\mathcal{T}_{i-(k-3)}$. Thus, if an occurrence of \mathcal{T}_{i-k} is not in $A_{i,k-1} \cap B'_{i,k-1}$, then this occurrence is not in $I_{i,k-3}$. Therefore, we conclude that $I_{i,k-3} \subset A_{i,k-1} \cap B'_{i,k-1}$. ◀

The analogous characterization of $B_{i,j}$ is presented as follows, which can be proven in a way similar to the proof of Theorem 19.

16:10 Net Occurrences in Fibonacci and Thue-Morse Words

► **Corollary 20.** For each $i \geq 2$ and $0 \leq j \leq i-1$, we have $B_{i,0} = \emptyset$, $B_{i,1} = \{\tau_{i-1} + 1\}$. For each $j \geq 2$, we define

$$\begin{aligned} A''_{i,j-1} &:= A_{i,j-1} \oplus \tau_{i-j}, & B''_{i,j-2} &:= B_{i,j-2} \oplus (\tau_{i-j} + \tau_{i-(j+1)}), \\ I'_{i,j-3} &:= \begin{cases} \emptyset, & j = 2, \\ A''_{i,j-3} \cup B''_{i,j-3}, & j \geq 3, \end{cases} \text{ where} \\ B''_{i,j-3} &:= B_{i,j-3} \oplus (\tau_{i-(j-1)} + \tau_{i-j}), & A''_{i,j-3} &:= A_{i,j-3} \oplus \tau_{i-(j-2)}. \end{aligned}$$

Then, $B_{i,j} = B_{i,j-1} \cup A''_{i,j-1} \cup B''_{i,j-2}$ with

$$B_{i,j-1} \cap A''_{i,j-1} = I'_{i,j-3}, \quad B_{i,j-1} \cap B''_{i,j-2} = \emptyset, \quad \text{and} \quad A''_{i,j-1} \cap B''_{i,j-2} = \emptyset. \quad \lrcorner$$

We next use Theorem 19 and Corollary 20 to count the number of occurrences of \mathcal{T}_{i-j} and $\overline{\mathcal{T}_{i-j}}$ in \mathcal{T}_i .

► **Corollary 21.** For $i \geq 2$, consider \mathcal{T}_i and $0 \leq j \leq i-1$. Let a_j and b_j denote the number of occurrences of \mathcal{T}_{i-j} and $\overline{\mathcal{T}_{i-j}}$ in \mathcal{T}_i , respectively. Then,

- $a_0 = a_1 = 1$ and $a_j = a_{j-1} + 2 \cdot a_{j-2}$ for each $j \geq 2$;
- $b_0 = 0$ and $b_j = b_{j-1} + a_{j-1}$ for each $j \geq 1$.

Proof. We proceed by induction on j . For $1 \leq j \leq 2$, the claim holds trivially. For $j \geq 3$, we have $|I_{i,j-3}| = |A_{i,j-3}| + |B_{i,j-3}| = a_{j-3} + b_{j-3} = b_{j-2}$ and $|A_{i,j}| = |A_{i,j-1}| + |B'_{i,j-1}| + |A'_{i,j-2}| - |I_{i,j-3}| = a_{j-1} + b_{j-1} + a_{j-2} - b_{j-2} = a_{j-1} + (b_{j-1} - b_{j-2}) + a_{j-2} = a_{j-1} + 2 \cdot a_{j-2}$. by induction hypothesis. We can prove b_j similarly with Corollary 20. ◀

► **Remark 22.** For each $i \geq 2$, a_j is the $(j+1)^{\text{th}}$ Jacobsthal Number: Sequence A001045 of the On-Line Encyclopedia of Integer Sequences (<https://oeis.org/A001045>).

With the characterization of the occurrences of \mathcal{T}_{i-j} and $\overline{\mathcal{T}_{i-j}}$ in \mathcal{T}_i , a natural next step is to investigate the structure of the strings that surround \mathcal{T}_{i-j} and $\overline{\mathcal{T}_{i-j}}$, which correspond to the blank areas in each row in Figure 3. In Appendix D, we explore two smallest factorizations of \mathcal{T}_i , each containing all occurrences of \mathcal{T}_{i-j} and $\overline{\mathcal{T}_{i-j}}$, respectively. The remaining factors in these factorizations represent the surrounding strings.

6 Net Occurrences in Fibonacci Words

In this section, we prove that there are only three net occurrences in each F_i , using the results on the occurrences of Fibonacci words of smaller order from Section 4, the notion of ONOC from Section 3, and new properties that we will develop in this section. We begin by reviewing the following results. Some of the proofs in this section are presented in Appendix B.

► **Lemma 23** ([18]). For $i \geq 7$, let $Q_i := F_{i-5}F_{i-6} \cdots F_3F_2$, $\Delta(0) := \mathbf{ba}$, $\Delta(1) := \mathbf{ab}$. Then,

$$F_{i-4} F_{i-5} = Q_i \Delta(1 - (i \bmod 2)) \text{ and} \quad (4)$$

$$F_{i-5} F_{i-4} = Q_i \Delta(i \bmod 2). \quad (5)$$

► **Lemma 24** ([18]). For each $i \geq 7$, the following are net occurrences in F_i :

- one occurrence of F_{i-2} at position $f_{i-1} + 1$;
- two occurrences of $F_{i-2} Q_i$ at positions 1 and $f_{i-2} + 1$.

Meanwhile, the two occurrences of F_{i-2} at positions 1 and $f_{i-2} + 1$ are not net occurrences.

F_i					
F_{i-2}	F_{i-2}		F_{i-5}	F_{i-4}	
F_{i-2}	F_{i-3}	F_{i-2}			
F_{i-2}	F_{i-2}			Q_i	Δ'
F_{i-2}	Q_i	Δ	F_{i-4}	Q_i	Δ'

■ **Figure 5** An illustration of several factorizations of F_i from Observation 14 and Lemma 23 where $\Delta := \Delta(1 - (i \bmod 2))$ and $\Delta' := \Delta(i \bmod 2)$. Net occurrences of F_{i-2} and $F_{i-2} Q_i$ are in yellow and green, respectively. Super-occurrences of the two BNSOs are shown as arrows.

With Lemma 15, we know that these three are the only occurrences of F_{i-2} in F_i . Similarly, we now strengthen Lemma 24 by showing the following result.

► **Lemma 25.** *For each $i \geq 7$, $F_{i-2} Q_i$ only occurs at positions 1 and $f_{i-2} + 1$ in F_i .*

By combining Lemma 15, Lemma 24, and Lemma 25, we conclude that F_{i-2} has only one net occurrence and $F_{i-2} Q_i$ only has two net occurrences.

► **Lemma 26.** *For each $i \geq 7$, the net occurrences identified in Lemma 24 are the only net occurrences of F_{i-2} and $F_{i-2} Q_i$ in F_i .*

It remains to show that there are no additional net occurrences in each F_i . To achieve this, we use the results from Section 3. First, observe that the three net occurrences in Lemma 24 form an ONOC of F_i . The two BNSOs of this ONOC correspond to an occurrence of Q_i and an occurrence of $F_{i-4} Q_i$, respectively. See Figure 5 for an illustration. Next, we aim to show that no super-occurrences of these two occurrences can be a net occurrence. To establish this, we analyze the super-occurrences of the occurrences of F_{i-3} in Lemma 32. This result covers the examination of the super-occurrences of the occurrence of $F_{i-4} Q_i$, since F_{i-3} is a prefix of $F_{i-4} Q_i = F_{i-3} Q_{i-1}$. Furthermore, Lemma 32 helps examining super-occurrences of the occurrence of Q_i in Lemma 33.

To prove these two lemmas, we introduce some properties of F_{i-3} and Q_i , which are proved in Appendix B.

► **Lemma 27.** $|Q_i| = f_{i-3} - 2$.

► **Lemma 28.** *For a substring S of F_i , if $F_{i-2} Q_i$ is a proper substring of S , then S is unique.*

► **Lemma 29.** F_{i-3} is always followed by Q_{i-1} in F_i .

► **Lemma 30.** $F_{i-3} F_{i-6} F_{i-5}$ and its length- $(f_{i-2} - 1)$ prefix are both unique in F_i .

► **Lemma 31.** *The length- $(f_{i-3} - 1)$ prefix of F_{i-3} is always followed by $F_{i-3}[f_{i-3}]$ in F_i .*

Now, we introduce the two crucial lemmas motivated earlier.

► **Lemma 32.** *Consider an occurrence (s, e) in F_i and let $S := F_i[s \dots e]$. If (s, e) is a super-occurrence of an occurrence of F_{i-3} , and S is neither F_{i-2} nor $F_{i-2} Q_i$, then (s, e) is not a net occurrence.*

16:12 Net Occurrences in Fibonacci and Thue-Morse Words

F_{i-3}	Q_{i-1}	Δ'	F_{i-3}	F_{i-2}
F_{i-3}	F_{i-3}	Q_{i-1}	Δ	F_{i-2}
F_{i-2}	F_{i-3}	Q_{i-1}	Δ'	F_{i-5} F_{i-4}
F_{i-1}	F_{i-3}	Q_{i-1}	Δ'	

(a) Case (1).

F_{i-3}	Q_{i-1}	Δ'	F_{i-3}	F_{i-2}
F_{i-3}	F_{i-3}	Q_{i-1}	Δ	F_{i-2}
F_{i-2}	F_{i-3}	Q_{i-1}	Δ'	F_{i-5} F_{i-4}
F_{i-2}	F_{i-3}	F_{i-3}	Q_{i-1}	Δ'

(b) Case (2).

F_{i-3}	F_{i-4}	F_{i-3}	F_{i-2}
F_{i-3}	F_{i-3}	F_{i-6} F_{i-5}	F_{i-2}
F_{i-2}	F_{i-3}	F_{i-4}	F_{i-5} F_{i-4}
F_{i-1}	F_{i-3}	F_{i-4}	

(c) Case (3).

F_{i-3}	F_{i-4}	F_{i-2}	F_{i-5} F_{i-4}
F_{i-3}	F_{i-3}	F_{i-6} F_{i-5}	F_{i-2}
F_{i-4}	Q_i	Δ'	F_{i-3} F_{i-4} F_{i-5} F_{i-4}
F_{i-2}	Q_i	Δ	F_{i-3} F_{i-4}

(d) Case (4).

F_{i-3}	Q_{i+1}	Δ	F_{i-2}
F_{i-3}	F_{i-3}	F_{i-6} F_{i-5}	F_{i-2}
F_{i-2}	F_{i-3}	Q_{i+1}	Δ'
F_{i-1}	F_{i-3}	F_{i-4}	

(e) Case (5).

F_{i-3}	Q_{i+1}	Δ	F_{i-2}
F_{i-3}	F_{i-3}	F_{i-6} F_{i-5}	F_{i-2}
F_{i-2}	F_{i-3}	Q_{i+1}	Δ'
F_{i-1}	F_{i-3}	F_{i-4}	

(f) Case (6).

F_{i-3}	Q_{i+1}	Δ	F_{i-2}
F_{i-3}	F_{i-3}	F_{i-6} F_{i-5}	F_{i-2}
F_{i-2}	F_{i-3}	Q_{i+1}	Δ'
F_{i-1}	F_{i-3}	F_{i-4}	

(g) Case (7).

■ **Figure 6** Illustration of the proof of Lemma 32. In each case, four factorizations of F_i are shown, each focusing on one occurrence of F_{i-3} , highlighted in green. For $S = X F_{i-3} Y$, each discussed X and Y is shown in pink and blue, respectively. Each discussed left or right extension character of S is shown in yellow. Recall that $\Delta := \Delta(1 - (i \bmod 2))$ and $\Delta' := \Delta(i \bmod 2)$.

Proof. The proof is illustrated in Figure 6. Consider strings X and Y such that $S = X F_{i-3} Y$ and $X Y \neq \epsilon$. We examine the following cases depending on $|Y|$. Note that $|Q_{i-1}| = f_{i-4} - 2$ and $|Q_{i+1}| = f_{i-2} - 2$ from Lemma 27.

- (1) $|Y| < |Q_{i-1}|$. Using Lemma 29, note that Y is a prefix of Q_{i-1} in this case. This means the right extension character of S is always $Q_{i-1}[|Y| + 1]$. Thus, no occurrence of S is a net occurrence.
- (2) $|Y| = |Q_{i-1}|$. Using Lemma 29, $F_{i-3} Y = F_{i-3} Q_{i-1}$ always holds in this case. Next, if $F_{i-3} Y$ occurs at position 1, $f_{i-2} + 1$, or $f_{i-1} + 1$, then the right extension character is always $\Delta(i \bmod 2)[1]$. On the other hand, if $F_{i-3} Y$ occurs at position $f_{i-3} + 1$, we examine the left extension character of $S = X F_{i-3} Y$. Notice that $|X| \leq f_{i-3}$ always holds in this case (and S becomes a prefix of F_i when $|X| = f_{i-3}$). Now, observe that if $F_{i-3} Y$ occurs at position $f_{i-3} + 1$ or $f_{i-1} + 1$, the left extension character of S is always $F_{i-3}[f_{i-3} - |X| - 1]$. Thus, this occurrence of S is also not a net occurrence.
- (3) $|Q_{i-1}| < |Y| < f_{i-4}$. If $F_{i-3} Y$ occurs at positions 1, $f_{i-2} + 1$, or $f_{i-1} + 1$, observe that occurrences of F_{i-3} at these three positions are always followed by F_{i-4} . Thus, Y is a prefix of F_{i-4} and the right extension character of S is always $F_{i-4}[|Y| + 1]$. So these three occurrences of S are not net occurrences. If $F_{i-3} Y$ occurs at position $f_{i-3} + 1$, since $|Y| = |Q_{i-1}| + 1 = f_{i-4} - 1$, we have $|F_{i-3} Y| = f_{i-3} + f_{i-4} - 1 = f_{i-2} - 1$. By Lemma 30, $F_{i-3} Y$ is unique, which means S is unique.
- (4) $|Y| = f_{i-4}$. If $F_{i-3} Y$ occurs at position 1, then X is empty and $S = F_{i-3} F_{i-4} = F_{i-2}$. If $F_{i-3} Y$ occurs at positions $f_{i-3} + 1$, then $F_{i-3} Y = F_{i-3} F_{i-6} F_{i-5}$ is unique (Lemma 30) so S is also unique. If $F_{i-3} Y$ occurs at position $f_{i-2} + 1$, then X ends with $\Delta(i \bmod 2)$.

If $F_{i-3}Y$ occurs at position $f_{i-1} + 1$, then X ends with $\Delta(1 - (i \bmod 2))$. Now, since $\Delta(i \bmod 2) F_{i-2}$ and $\Delta(1 - (i \bmod 2)) F_{i-2}$ are both unique by Lemma 15, S is also unique if $F_{i-3}Y$ occurs at these two positions.

- (5) $f_{i-4} < |Y| < |Q_{i+1}|$. First observe that the occurrences of F_{i-3} at positions 1 and $f_{i-2} + 1$ are both followed by $F_{i-4} Q_i = Q_{i+1}$. Thus, if $F_{i-3}Y$ occurs at these two positions, then Y is a prefix of Q_{i+1} and the right extension character of S is always $Q_{i+1}[|Y| + 1]$. So these two occurrences of S are not net occurrences. Next, if $F_{i-3} Y$ occurs at position $f_{i-3} + 1$, then $F_{i-3} Y$ is unique because $F_{i-3} F_{i-6} F_{i-5}$ is a prefix of $F_{i-3}Y$ and the former is unique by Lemma 30. Thus, S is unique. Finally note that, $F_{i-3}Y$ cannot occur at position $f_{i-1} + 1$ because $|Y| > |F_{i-4}|$ and $F_{i-3} F_{i-4}$ is a suffix of F_i .
- (6) $|Y| = |Q_{i+1}|$. Similar to the previous case, if $F_{i-3}Y$ occurs at position $f_{i-3} + 1$, then $F_{i-3}Y$ is unique, and $F_{i-3}Y$ cannot occur at position $f_{i-1} + 1$. If $F_{i-3}Y$ occurs at position 1, then X is empty and $S = F_{i-3}Y = F_{i-3} Q_{i+1} = F_{i-2} Q_i$. If $F_{i-3}Y$ occurs at position $f_{i-2} + 1$, then $S = X F_{i-2} Q_i$, which is unique by Lemma 28.
- (7) $|Y| > |Q_{i+1}|$. Similar to the previous two cases, if $F_{i-3}Y$ occurs at position $f_{i-3} + 1$, then $F_{i-3}Y$ is unique, and $F_{i-3}Y$ cannot occur at position $f_{i-1} + 1$. If $F_{i-3}Y$ occurs at position $f_{i-2} + 1$, then $F_{i-3}Y$ is a prefix of $F_{i-3} F_{i-2} = F_{i-2} F_{i-5} F_{i-4} = F_{i-2} Q_i \Delta(i \bmod 2)$, which is unique by Lemma 28. Thus, S is unique. Finally, since $|Y| > |Q_{i+1}| > f_{i-2} - 1$, if $F_{i-3}Y$ occurs at positions 1, then Y begins with the length- $(f_{i-2} - 1)$ prefix of $F_{i-3} F_{i-6} F_{i-5}$, which is unique by Lemma 30. Thus, S is also unique. ◀

► **Lemma 33.** Consider an occurrence (s, e) in F_i . If (s, e) is a proper super-occurrence of the occurrence of Q_i at position $f_{i-2} + 1$, then (s, e) is not a net occurrence.

Proof. Consider strings X and Y such that $S = XQ_iY$ and $X Y \neq \epsilon$. When $|Y| \geq 2$, notice that F_{i-3} occurs in S . By Lemma 32, (s, e) is not a net occurrence. Now, we consider the case when $|Y| = 1$. Note that Q_iY is precisely the length- $(f_{i-3} - 1)$ prefix of F_{i-3} . Thus, by Lemma 31, Q_iY is always followed by the same right extension character, $F_{i-3}[f_{i-3}]$, which means (s, e) is not a net occurrence. ◀

Finally, the main result follows from Lemma 24, Lemma 12, Lemma 32 and Lemma 33.

► **Theorem 34.** The three net occurrences identified in Lemma 24 are the only ones in F_i .

7 Net Occurrences in Thue-Morse Words

In this section, we prove the only nine net occurrences in each \mathcal{T}_i using the results on the occurrences of Thue-Morse words of smaller order from Section 5, the notion of ONOC from Section 3, and new results that we will introduce in this section. We will first show that each occurrence of each string in \mathcal{P}_i (defined below) is a net occurrence in \mathcal{T}_i , then show that they are the only ones.

► **Definition 35.** For each $i \geq 5$, define $\mathcal{P}_i := \{\mathcal{T}_{i-2}, \overline{\mathcal{T}_{i-2}}, \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}}, \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-3}\}$.

We next show several factorizations of \mathcal{T}_i , proved in Appendix C.

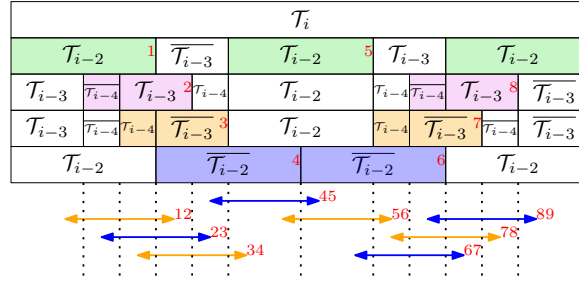
► **Lemma 36.** For each $i \geq 5$:

$$\mathcal{T}_i = \mathcal{T}_{i-2} \overline{\mathcal{T}_{i-2}} \overline{\mathcal{T}_{i-2}} \mathcal{T}_{i-2} \tag{6}$$

$$\mathcal{T}_i = \mathcal{T}_{i-2} \overline{\mathcal{T}_{i-3}} \mathcal{T}_{i-2} \mathcal{T}_{i-3} \mathcal{T}_{i-2} \tag{7}$$

$$\mathcal{T}_i = \mathcal{T}_{i-3} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}} \mathcal{T}_{i-2} \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}} \overline{\mathcal{T}_{i-4}} \overline{\mathcal{T}_{i-3}} \tag{8}$$

$$\mathcal{T}_i = \mathcal{T}_{i-3} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-3} \mathcal{T}_{i-4} \mathcal{T}_{i-2} \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-3} \overline{\mathcal{T}_{i-3}} \tag{9}$$



■ **Figure 7** An illustration of several factorizations of \mathcal{T}_i from Lemma 36. Net occurrences of each string in Definition 35 are highlighted in a separate color. Super-occurrences of the eight BNSOs are shown as colored arrows, blue for \mathcal{T}_{i-3} and orange for $\overline{\mathcal{T}_{i-3}}$ (see Lemma 40). Each net occurrence is numbered in red at the top-right corner, and each arrow with label ij corresponds to an overlap between the i^{th} and j^{th} net occurrences.

The following two results immediately follow from Theorem 19. Note that Corollary 37 also appears in [34].

► **Corollary 37.** For each $i \geq 5$:

- \mathcal{T}_{i-2} only occurs at positions 1, $\tau_{i-2} + \tau_{i-3} + 1$ and $\tau_{i-1} + \tau_{i-2} + 1$ in \mathcal{T}_i .
- $\overline{\mathcal{T}_{i-2}}$ only occurs at positions $\tau_{i-2} + 1$ and $\tau_{i-1} + 1$ in \mathcal{T}_i .
- $\mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}}$ only occurs at positions $\tau_{i-3} + \tau_{i-4} + 1$ and $\tau_{i-1} + \tau_{i-3} + 1$ in \mathcal{T}_i .
- $\overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-3}$ only occurs at positions $\tau_{i-3} + 1$ and $\tau_{i-1} + \tau_{i-3} + \tau_{i-4} + 1$ in \mathcal{T}_i .

► **Corollary 38.** \mathcal{T}_{i-3} only occurs at positions 1, $\tau_{i-3} + \tau_{i-4} + 1$, $\tau_{i-2} + \tau_{i-3} + 1$, $\tau_{i-1} + \tau_{i-3} + 1$, and $\tau_{i-1} + \tau_{i-2} + 1$ in \mathcal{T}_i .

We now identify the nine net occurrences in \mathcal{T}_i .

► **Lemma 39.** Each occurrence of each string in \mathcal{P}_i is a net occurrence in \mathcal{T}_i .

Proof. We proceed by examining the left and right extension characters of each occurrence of each string in \mathcal{P}_i .

Since \mathcal{T}_{i-2} is a prefix and a suffix of \mathcal{T}_i , by the definition of occurrences, the occurrence of \mathcal{T}_{i-2} at positions 1 has a unique left extension character, and the occurrence of \mathcal{T}_{i-2} at positions $\tau_{i-1} + \tau_{i-2} + 1$ has a unique right extension character. Next, note that the right extension character of the occurrence of \mathcal{T}_{i-2} at position 1 differs from that of the occurrence at position $\tau_{i-2} + \tau_{i-3} + 1$ because $\overline{\mathcal{T}_{i-3}}[1] \neq \mathcal{T}_{i-3}[1]$. Similarly, the left extension character of the occurrence at position $\tau_{i-2} + \tau_{i-3} + 1$ differs from that of the occurrence at position $\tau_{i-1} + \tau_{i-2} + 1$, because $\overline{\mathcal{T}_{i-3}}[\tau_{i-3}] \neq \mathcal{T}_{i-3}[\tau_{i-3}]$. Hence, all three occurrences of \mathcal{T}_{i-2} are net occurrences.

For $\overline{\mathcal{T}_{i-2}}$, a similar argument holds. the right extension characters satisfy $\overline{\mathcal{T}_{i-2}}[1] \neq \mathcal{T}_{i-2}[1]$ and the left extension characters satisfy $\mathcal{T}_{i-2}[\tau_{i-2}] \neq \overline{\mathcal{T}_{i-2}}[\tau_{i-2}]$. Thus, both occurrences of $\overline{\mathcal{T}_{i-2}}$ are net occurrences. For $\mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}}$, similarly, the right extension characters satisfy $\mathcal{T}_{i-2}[1] \neq \overline{\mathcal{T}_{i-4}}[1]$ and the left extension characters satisfy $\overline{\mathcal{T}_{i-4}}[\tau_{i-4}] = \overline{\mathcal{T}_{i-2}}[\tau_{i-2}] \neq \mathcal{T}_{i-2}[\tau_{i-2}]$. Thus, both occurrences of $\mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}}$ are net occurrences. Finally, for $\overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-3}$, once again, the right extension characters satisfy $\mathcal{T}_{i-4}[1] \neq \overline{\mathcal{T}_{i-3}}[1]$ and the left extension characters satisfy $\mathcal{T}_{i-3}[\tau_{i-3}] \neq \overline{\mathcal{T}_{i-4}}[\tau_{i-4}]$. Thus, both occurrences of $\overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-3}$ are net occurrences. ◀

To show that all other occurrences are not net occurrences, we follow Lemma 12. First note that the nine net occurrences we identified in Lemma 39 form an ONOC of \mathcal{T}_i . The eight BNSOs of this ONOC correspond to the occurrences of \mathcal{T}_{i-3} and $\overline{\mathcal{T}_{i-3}}$ shown in Figure 7. We next show that no super-occurrences of these occurrences are net occurrences to conclude that this ONOC already contains all the net occurrences in \mathcal{T}_i .

► **Lemma 40.** *Consider an occurrence (s, e) in \mathcal{T}_i and let $S := \mathcal{T}_i[s \dots e]$. If (s, e) is a proper super-occurrence of \mathcal{T}_{i-3} or $\overline{\mathcal{T}_{i-3}}$, and $S \notin \mathcal{P}_i$, then (s, e) is not a net occurrence.*

Proof. We first consider when (s, e) contains an occurrence of \mathcal{T}_{i-3} . Consider strings X and Y such that $S = X \mathcal{T}_{i-3} Y$ and $X Y \neq \epsilon$. Let position $u := s + |X|$ be the starting position of this occurrence of \mathcal{T}_{i-3} . Let $C := \{1, \tau_{i-2} + \tau_{i-3} + 1, \tau_{i-1} + \tau_{i-2} + 1\}$ and $D := \{\tau_{i-3} + \tau_{i-4} + 1, \tau_{i-1} + \tau_{i-3} + 1\}$. By Corollary 38, we have $u \in C \cup D$. We next examine the following cases depending on which set u belongs to and how large $|Y|$ is.

We first consider when $u \in C$.

- (a) $|Y| < \tau_{i-3}$. By Corollary 37, note that Y is always a prefix of $\overline{\mathcal{T}_{i-3}}$. This means the right extension character of S is always $\overline{\mathcal{T}_{i-3}}[|Y| + 1]$. Thus, (s, e) is not a net occurrence.
- (b) $|Y| = \tau_{i-3}$. By Corollary 37, $S \in \mathcal{P}_i$ in this case.
- (c) $|Y| > \tau_{i-3}$. By Corollary 37, (s, e) contains a net occurrence of $\mathcal{T}_{i-3} \overline{\mathcal{T}_{i-3}} = \mathcal{T}_{i-2}$ as a proper sub-occurrence. By Observation 8 and Lemma 39, (s, e) is not a net occurrence.

We next consider when $u \in D$.

- (a) $|Y| < \tau_{i-4}$. Recall that $\mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}} = \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-4} = \mathcal{T}_{i-3} \mathcal{T}_{i-4}$. By Corollary 37, note that Y is always a prefix of \mathcal{T}_{i-4} in this case. This means the right extension character of S is always $\mathcal{T}_{i-4}[|Y| + 1]$. Thus, (s, e) is not a net occurrence.
- (b) $|Y| = \tau_{i-4}$. Using Corollary 37, $S \in \mathcal{P}_i$ in this case.
- (c) $|Y| > \tau_{i-4}$. By Corollary 37, in this case (s, e) contains a net occurrence of $\mathcal{T}_{i-3} \mathcal{T}_{i-4} = \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-4} = \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}}$ as a proper sub-occurrence. Thus, by Observation 8 and Lemma 39, (s, e) is not a net occurrence.

We can prove the case when (s, e) contains an occurrence of $\overline{\mathcal{T}_{i-3}}$ similarly. ◀

Finally, the main result follows from Lemma 12, Lemma 39 and Lemma 40.

► **Theorem 41.** *The net occurrences in Lemma 39 are the only net occurrences in each \mathcal{T}_i .*

8 Conclusion and Future Work

In this work, we investigate net occurrences in Fibonacci and Thue-Morse words, making two main contributions. First, we confirm the conjecture that each Fibonacci word contains exactly three net occurrences. Second, we establish that each Thue-Morse word contains exactly nine net occurrences. To achieve these results, we first introduce the notion of an overlapping net occurrence cover and show how it can be used to prove that certain net occurrences in a text are the only ones. We then develop recurrence relations that precisely characterize the occurrences of Fibonacci and Thue-Morse words of smaller order, which could be of independent interest. As an application, we illustrate how these results facilitate the counting of small-order occurrences.

An avenue of future work is to extend our findings to study the net occurrences in k -bonacci words [15–17, 25] and Thue-Morse-like words [1, 9]. Furthermore, since both Fibonacci and Thue-Morse words can be defined via morphisms, one could also explore net occurrences in other morphic words [8, 14, 21]. Finally, the net occurrences have been characterized in terms of minimal unique substrings [31]; this viewpoint may offer alternative and potentially simpler proofs than those presented in Sections 6–7.

References

- 1 Ibai Aedo, Uwe Grimm, Yasushi Nagai, and Petra Staynova. Monochromatic arithmetic progressions in binary Thue-Morse-like words. *Theoretical Computer Science*, 934:65–80, 2022. doi:10.1016/J.TCS.2022.08.013.
- 2 Jean-Paul Allouche and Jeffrey O. Shallit. The ubiquitous Prouhet-Thue-Morse sequence. In Cunsheng Ding, Tor Helleseth, and Harald Niederreiter, editors, *Sequences and their Applications - Proceedings of SETA 1998, Singapore, December 14-17, 1998*, Discrete Mathematics and Theoretical Computer Science, pages 1–16. Springer, 1998. doi:10.1007/978-1-4471-0551-0_1.
- 3 Hideo Bannai, Mitsuru Funakoshi, Tomohiro I, Dominik Köppl, Takuya Mieno, and Takaaki Nishimoto. A separation of γ and b via Thue-Morse words. In Thierry Lecroq and Hélène Touzet, editors, *String Processing and Information Retrieval - 28th International Symposium, SPIRE 2021, Lille, France, October 4-6, 2021, Proceedings*, volume 12944 of *Lecture Notes in Computer Science*, pages 167–178. Springer, 2021. doi:10.1007/978-3-030-86692-1_14.
- 4 Hideo Bannai, Tomohiro I, Shunsuke Inenaga, Yuto Nakashima, Masayuki Takeda, and Kazuya Tsuruta. The “runs” theorem. *SIAM J. Comput.*, 46(5):1501–1514, 2017. doi:10.1137/15M1011032.
- 5 Hideo Bannai, Tomohiro I, and Yuto Nakashima. On the compressiveness of the Burrows-Wheeler transform. In Paola Bonizzoni and Veli Mäkinen, editors, *36th Annual Symposium on Combinatorial Pattern Matching, CPM 2025, June 17-19, 2025, Milano, Italy*, LIPIcs. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.48550/arXiv.2411.11298.
- 6 Jean Berstel and Dominique Perrin. The origins of combinatorics on words. *European Journal of Combinatorics*, 28(3):996–1022, 2007. doi:10.1016/J.EJC.2005.07.019.
- 7 Srečko Brlek. Enumeration of factors in the Thue-Morse word. *Discrete Applied Mathematics*, 24(1-3):83–96, 1989. doi:10.1016/0166-218X(92)90274-E.
- 8 Srečko 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.
- 9 Jin Chen, Zhi-Xiong Wen, and Wen Wu. On the additive complexity of a Thue-Morse-like sequence. *Discrete Applied Mathematics*, 260:98–108, 2019. doi:10.1016/J.DAM.2019.01.008.
- 10 Maxime Crochemore, Lucian Ilie, and Wojciech Rytter. Repetitions in strings: Algorithms and combinatorics. *Theoretical Computer Science*, 410(50):5227–5235, 2009. doi:10.1016/J.TCS.2009.08.024.
- 11 Francesco Dolce. String attractors for factors of the Thue-Morse word. In Anna E. Frid and Robert Mercas, editors, *Combinatorics on Words - 14th International Conference, WORDS 2023, Umeå, Sweden, June 12-16, 2023, Proceedings*, volume 13899 of *Lecture Notes in Computer Science*, pages 117–129. Springer, 2023. doi:10.1007/978-3-031-33180-0_9.
- 12 Xavier Droubay. Palindromes in the Fibonacci word. *Information Processing Letters*, 55(4):217–221, 1995. doi:10.1016/0020-0190(95)00080-V.
- 13 Aviezri S. Fraenkel and Jamie Simpson. The exact number of squares in Fibonacci words. *Theoretical Computer Science*, 218(1):95–106, 1999. doi:10.1016/S0304-3975(98)00252-7.
- 14 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.
- 15 Narges Ghareghani, Morteza Mohammad Noori, and Pouyeh Sharifani. Some properties of the k -bonacci words on infinite alphabet. *The Electronic Journal of Combinatorics*, 27(3):3, 2020. doi:10.37236/9406.

- 16 Narges Ghareghani and Pouyeh Sharifani. On square factors and critical factors of k -bonacci words on infinite alphabet. *Theoretical Computer Science*, 865:34–43, 2021. doi:10.1016/j.tcs.2021.02.027.
- 17 France Gheeraert, Giuseppe Romana, and Manon Stipulanti. String attractors of fixed points of k -bonacci-like morphisms. In Anna E. Frid and Robert Mercas, editors, *Combinatorics on Words - 14th International Conference, WORDS 2023, Umeå, Sweden, June 12-16, 2023, Proceedings*, volume 13899 of *Lecture Notes in Computer Science*, pages 192–205. Springer, 2023. doi:10.1007/978-3-031-33180-0_15.
- 18 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.
- 19 Peaker Guo, Seeun William Umboh, Anthony Wirth, and Justin Zobel. Online computation of string net frequency. In Zsuzsanna Lipták and Edleno Moura, editors, *String Processing and Information Retrieval - 31th International Symposium, SPIRE 2024, Puerto Vallarta, Mexico, September 23-25, 2024, Proceedings*, *Lecture Notes in Computer Science*. Springer, 2024. doi:10.1007/978-3-031-72200-4_12.
- 20 Dan Gusfield and Jens Stoye. Linear time algorithms for finding and representing all the tandem repeats in a string. *Journal of Computer and System Sciences*, 69(4):525–546, 2004. doi:10.1016/J.JCSS.2004.03.004.
- 21 Vesa Halava, Tero Harju, Tomi Kärki, and Michel Rigo. On the periodicity of morphic words. In Yuan Gao, Hanlin Lu, Shinnosuke Seki, and Sheng Yu, editors, *Developments in Language Theory, 14th International Conference, DLT 2010, London, ON, Canada, August 17-20, 2010. Proceedings*, volume 6224 of *Lecture Notes in Computer Science*, pages 209–217. Springer, 2010. doi:10.1007/978-3-642-14455-4_20.
- 22 Costas S. Iliopoulos, Dennis W. G. Moore, and William F. Smyth. A characterization of the squares in a Fibonacci string. *Theoretical Computer Science*, 172(1-2):281–291, 1997. doi:10.1016/S0304-3975(96)00141-7.
- 23 Shunsuke Inenaga. Faster and simpler online computation of string net frequency. *CoRR*, 2024. doi:10.48550/arXiv.2410.06837.
- 24 Hiroe Inoue, Yoshiaki Matsuoka, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda. Factorizing strings into repetitions. *Theory of Computing Systems*, 66(2):484–501, 2022. doi:10.1007/S00224-022-10070-3.
- 25 Marieh Jahannia, Morteza Mohammad Noori, Narad Rampersad, and Manon Stipulanti. Closed Ziv-Lempel factorization of the m -bonacci words. *Theoretical Computer Science*, 918:32–47, 2022. doi:10.1016/j.tcs.2022.03.019.
- 26 Donald E. Knuth, James H. Morris Jr., and Vaughan R. Pratt. Fast pattern matching in strings. *SIAM Journal on Computing*, 6(2):323–350, 1977. doi:10.1137/0206024.
- 27 Kanaru Kutsukake, Takuya Matsumoto, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda. On repetitiveness measures of Thue-Morse words. In Christina Boucher and Sharma V. Thankachan, editors, *String Processing and Information Retrieval - 27th International Symposium, SPIRE 2020, Orlando, FL, USA, October 13-15, 2020, Proceedings*, volume 12303 of *Lecture Notes in Computer Science*, pages 213–220. Springer, 2020. doi:10.1007/978-3-030-59212-7_15.
- 28 Yih-Jeng Lin and Ming-Shing Yu. Extracting Chinese frequent strings without dictionary from a Chinese corpus and its applications. *Journal of Information Science and Engineering*, 17(5):805–824, 2001. URL: https://jise.iis.sinica.edu.tw/JISESearch/pages/View/PaperView.jsf?keyId=86_1308.
- 29 Yih-Jeng Lin and Ming-Shing Yu. The properties and further applications of Chinese frequent strings. *International Journal of Computational Linguistics and Chinese Language Processing*, 9(1), 2004. URL: <http://www.aclclp.org.tw/clclp/v9n1/v9n1a7.pdf>.

- 30 M. Lothaire. *Combinatorics on words, Second Edition*. Cambridge mathematical library. Cambridge University Press, 1997.
- 31 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, Milano, Italy*, LIPIcs. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/LIPIcs.CPM.2025.23.
- 32 Gonzalo Navarro, Carlos Ochoa, and Nicola Prezza. On the approximation ratio of ordered parsings. *IEEE Transactions on Information Theory*, 67(2):1008–1026, 2021. doi:10.1109/TIT.2020.3042746.
- 33 Enno Ohlebusch, Thomas Büchler, and Jannik Olbrich. Faster computation of Chinese frequent strings and their net frequencies. In Zsuzsanna Lipták and Edleno Moura, editors, *String Processing and Information Retrieval - 31th International Symposium, SPIRE 2024, Puerto Vallarta, Mexico, September 23-25, 2024, Proceedings*, Lecture Notes in Computer Science. Springer, 2024. doi:10.1007/978-3-031-72200-4_19.
- 34 Jakub Radoszewski and Wojciech Rytter. On the structure of compacted subword graphs of Thue-Morse words and their applications. *Journal of Discrete Algorithms*, 11:15–24, 2012. doi:10.1016/J.JDA.2011.01.001.
- 35 Wojciech Rytter. The structure of subword graphs and suffix trees of Fibonacci words. *Theoretical Computer Science*, 363(2):211–223, 2006. doi:10.1016/j.tcs.2006.07.025.

A Proofs Omitted from Section 3

► **Observation 8.** *In a text T , if an occurrence (s, e) is a proper super-occurrence of a net occurrence, then (s, e) is not a net occurrence.*

Proof. Let (s', e') be the net occurrence, then $T[s' - 1 \dots e']$ and $T[s' \dots e' + 1]$ are both unique. Since $T[s \dots e]$ contains at least one of these two strings as a substring, $T[s \dots e]$ is also unique. Thus, (s, e) is not a net occurrence. ◀

► **Observation 9.** *In a text T , if an occurrence (s, e) is a proper sub-occurrence of a net occurrence, then (s, e) is not a net occurrence.*

Proof. Let (s', e') be the net occurrence, then $T[s' \dots e']$ is repeated. Since (s, e) is a proper sub-occurrence of (s', e') , $T[s - 1 \dots e]$ or $T[s \dots e + 1]$ is also repeated. Thus, (s, e) is not a net occurrence. ◀

► **Lemma 12.** *For a text T , if there exists an ONOC \mathcal{C} of T such that \mathcal{C} does not contain all the net occurrences in T , then each net occurrence in T outside of \mathcal{C} must be a super-occurrence of $(i - 1, j + 1)$, where (i, j) is a BNSO of \mathcal{C} .*

Proof. Let (s, e) be a net occurrence in T that is outside of \mathcal{C} . Assume, by contradiction, that (s, e) is not a super-occurrence of any occurrence $(i - 1, j + 1)$, where (i, j) is an occurrence in the set of BNSOs, $\{(i_2, j_1), (i_3, j_2), \dots, (i_c, j_{c-1})\}$. We consider the following two cases depending on the position of s .

First, when $i_{k+1} \leq s < i_{k+2}$ for some $0 \leq k \leq c - 2$. Given our assumption, (s, e) is not a super-occurrence of $(i_{k+2} - 1, j_{k+1} + 1)$, where (i_{k+2}, j_{k+1}) is a BNSO. It follows that e satisfies $e \leq j_{k+1}$, implying that (s, e) must be a sub-occurrence of (i_{k+1}, j_{k+1}) , which is a net occurrence in \mathcal{C} . Now, if (s, e) is a proper sub-occurrence of (i_{k+1}, j_{k+1}) , this contradicts Observation 9; if (s, e) is (i_{k+1}, j_{k+1}) , it contradicts the assumption that (s, e) is a net occurrence in T outside of \mathcal{C} . Second, when $s \geq i_c$. In this case, (s, e) is a sub-occurrence of (i_c, j_c) , the last net occurrence in \mathcal{C} .

In both cases, (s, e) must be a sub-occurrence of some net occurrence in \mathcal{C} , leading to a contradiction of either the assumption or Observation 9. Therefore, our initial assumption was false and we conclude that (s, e) is indeed a super-occurrence of $(i-1, j+1)$, where (i, j) is a BNSO of \mathcal{C} . ◀

B Proofs Omitted from Section 6

► **Lemma 25.** *For each $i \geq 7$, $F_{i-2} Q_i$ only occurs at positions 1 and $f_{i-2} + 1$ in F_i .*

Proof. By Lemma 15, there are only three positions where $F_{i-2} Q_i$ could occur. First observe that $F_{i-2} Q_i$ cannot occur at position $f_{i-1} + 1$. Next, by Observation 14 and Lemma 23, the occurrences of F_{i-2} at positions 1 and $f_{i-2} + 1$ are both followed by Q_i , thus $F_{i-2} Q_i$ only occurs at these two positions. ◀

► **Lemma 27.** $|Q_i| = f_{i-3} - 2$.

Proof. Note that $|Q_i| = \sum_{j=2}^{i-5} |F_j| = \left(\sum_{j=1}^{i-5} f_j \right) - f_1 = f_{i-3} - 1 - f_1 = f_{i-3} - 2$ where the third equality comes from the fact that $\sum_{j=1}^k f_j = f_{k+2} - 1$. ◀

► **Lemma 28.** *For a substring S of F_i , if $F_{i-2} Q_i$ is a proper substring of S , then S is unique.*

Proof. By Lemma 25 and Lemma 24, $F_{i-2} Q_i$ only occurs twice in F_i , and both are net occurrences, which means the extensions are unique. Thus, any string containing $F_{i-2} Q_i$ as a substring is also unique. ◀

► **Lemma 29.** F_{i-3} is always followed by Q_{i-1} in F_i .

Proof. Observe that the occurrence of F_{i-3} at position $f_{i-3} + 1$ is followed by $F_{i-6} F_{i-5} = Q_{i-1} \Delta(1 - (i \bmod 2))$ while the other three occurrences of F_{i-3} are all followed by $F_{i-4} = F_{i-5} F_{i-6} = Q_{i-1} \Delta(i \bmod 2)$. ◀

► **Lemma 30.** $F_{i-3} F_{i-6} F_{i-5}$ and its length- $(f_{i-2} - 1)$ prefix are both unique in F_i .

Proof. From the proof of Lemma 29, F_{i-3} is only followed by $F_{i-6} F_{i-5}$ once and by F_{i-4} three times, thus $F_{i-3} F_{i-6} F_{i-5}$ is unique. Next, by Lemma 27, $|F_{i-3} Q_{i-1}| = f_{i-3} + f_{i-4} - 2 = f_{i-2} - 2$. Also from the proof of Lemma 29, $F_{i-3} Q_{i-1}$ is only followed by $\Delta(1 - (i \bmod 2))$ once and by $\Delta(i \bmod 2)$ three times, thus, the length- $(f_{i-2} - 1)$ prefix of $F_{i-3} Q_{i-1} \Delta(1 - (i \bmod 2)) = F_{i-3} F_{i-6} F_{i-5}$ is also unique. ◀

► **Lemma 31.** *The length- $(f_{i-3} - 1)$ prefix of F_{i-3} is always followed by $F_{i-3}[f_{i-3}]$ in F_i .*

Proof. Let U be the length- $(f_{i-3} - 1)$ prefix of F_{i-3} . By Equation (1), $F_{i-3} = Q_i \Delta(1 - (i \bmod 2))$. Note that $U[|U| - 1]$ is always **a** because Q_i ends with $F_2 = \mathbf{a}$. When $U[|U|] = \mathbf{a}$, the right extension character of U is always **b**. This is because, if it were not, **aaa** would occur in F_i , contradicting Lemma 3. On the other hand, when $U[|U|] = \mathbf{b}$, the right extension character of U is always **a**. This is because, similarly, an occurrence of **bb** would contradict Lemma 3. Finally, observe that, whether $U[|U|]$ is **a** or **b**, $U[|U|]$ concatenated with the right extension character of U is exactly $\Delta(1 - (i \bmod 2))$. Therefore, the desired result follows. ◀

C Proof Omitted from Section 7

► **Lemma 36.** For each $i \geq 5$:

$$\mathcal{T}_i = \mathcal{T}_{i-2} \overline{\mathcal{T}_{i-2}} \overline{\mathcal{T}_{i-2}} \mathcal{T}_{i-2} \quad (6)$$

$$\mathcal{T}_i = \mathcal{T}_{i-2} \overline{\mathcal{T}_{i-3}} \mathcal{T}_{i-2} \mathcal{T}_{i-3} \mathcal{T}_{i-2} \quad (7)$$

$$\mathcal{T}_i = \mathcal{T}_{i-3} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}} \mathcal{T}_{i-2} \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}} \overline{\mathcal{T}_{i-4}} \overline{\mathcal{T}_{i-3}} \quad (8)$$

$$\mathcal{T}_i = \mathcal{T}_{i-3} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-3} \mathcal{T}_{i-4} \mathcal{T}_{i-2} \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-3} \overline{\mathcal{T}_{i-3}} \quad (9)$$

Proof. The first two follow from Equation (10) and Equation (11). We next proceed by repeatedly applying the definition of Thue-Morse words. Substituting $\mathcal{T}_{i-2} = \mathcal{T}_{i-3} \overline{\mathcal{T}_{i-3}} = \mathcal{T}_{i-3} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-4}$ and $\mathcal{T}_{i-3} \mathcal{T}_{i-2} = \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-3} \overline{\mathcal{T}_{i-3}} = \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-4}} \overline{\mathcal{T}_{i-3}} = \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-4}$ to Equation (10), we have Equation (8). Finally, observe that $\mathcal{T}_{i-4} \overline{\mathcal{T}_{i-3}} = \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-4} = \mathcal{T}_{i-3} \mathcal{T}_{i-4}$ and similarly, $\overline{\mathcal{T}_{i-3}} \mathcal{T}_{i-4} = \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-4} \overline{\mathcal{T}_{i-4}} = \overline{\mathcal{T}_{i-4}} \mathcal{T}_{i-3}$. Substituting them to Equation (8), we derive Equation (9). ◀

D A Factorization of Thue-Morse Word

First, we define a *smallest* factorization of a string as one that contains the fewest number of factors while satisfying certain conditions. In this section, we explore two smallest factorizations of \mathcal{T}_i , each containing all occurrences of \mathcal{T}_{i-j} and $\overline{\mathcal{T}_{i-j}}$, respectively. Observe that such factorizations exist due to the overlap-free property of each Thue-Morse word (Lemma 5).

► **Definition 42.** For each $i \geq 2$ and $0 \leq j \leq i-1$, we define the following.

- Let $\mathcal{F}_{i,j}^A$ and $\mathcal{F}_{i,j}^B$ denote the smallest factorization of \mathcal{T}_i that contains all occurrences of \mathcal{T}_{i-j} and $\overline{\mathcal{T}_{i-j}}$ in \mathcal{T}_i , respectively.
- Define $(\mathcal{T}_{i-j})^- := \mathcal{T}_{i-(j+1)}$ and $(\overline{\mathcal{T}_{i-j}})^- := \overline{\mathcal{T}_{i-(j+1)}}$.
- Consider $\mathcal{F}_{i,j} \in \{\mathcal{F}_{i,j}^A, \mathcal{F}_{i,j}^B\}$ and suppose $\mathcal{F}_{i,j} = (x_t)_{t=1}^m$. We define two operators on $\mathcal{F}_{i,j}$:

$$(\mathcal{F}_{i,j})^- := \left((x_t)^- \right)_{t=1}^m \quad \text{and} \quad \overline{\mathcal{F}_{i,j}} := (\overline{x_t})_{t=1}^m.$$

- Consider two factorizations $\mathcal{X} = (x_k)_{k=1}^m$ and $\mathcal{Y} = (y_k)_{k=1}^\ell$. If $x_m = y_1 = \overline{\mathcal{T}_{i-j}}$, then

$$\mathcal{X} \boxplus \mathcal{Y} := (x_1, x_2, \dots, x_{m-1}, \overline{\mathcal{T}_{i-(j+1)}}, \mathcal{T}_{i-j}, \mathcal{T}_{i-(j+1)}, y_2, y_3, \dots, y_\ell). \quad \lrcorner$$

For the definition of operator \boxplus , note that $|\mathcal{X} \boxplus \mathcal{Y}| = |\mathcal{X}| + |\mathcal{Y}| + 1$ and

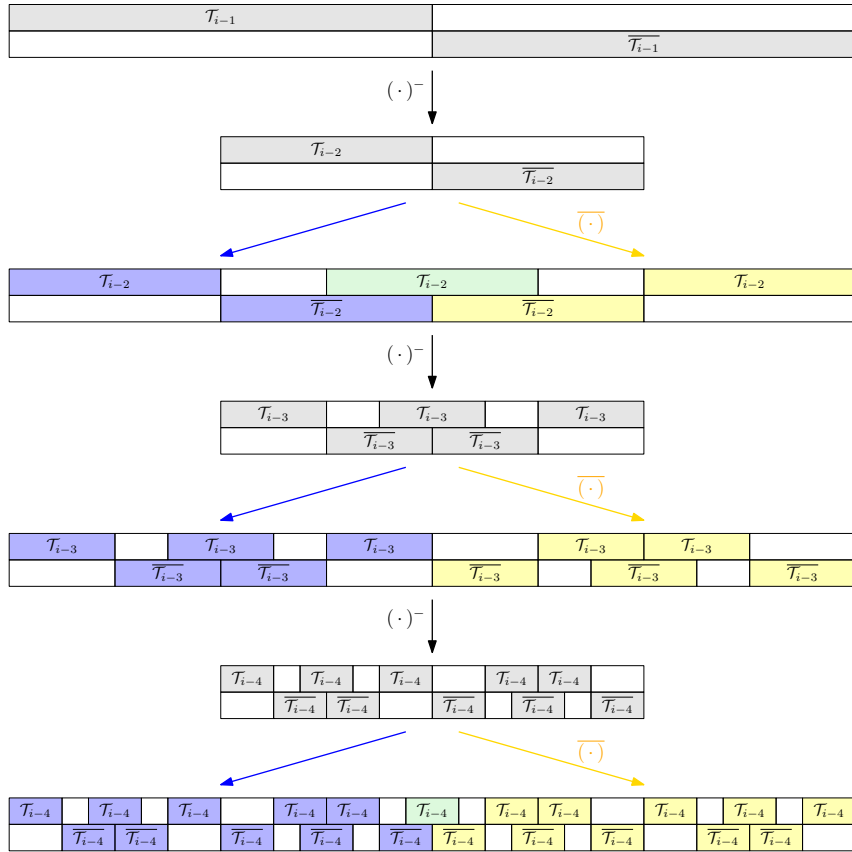
$$\overline{\mathcal{T}_{i-j}} \overline{\mathcal{T}_{i-j}} = \overline{\mathcal{T}_{i-(j+1)}} \mathcal{T}_{i-(j+1)} \overline{\mathcal{T}_{i-(j+1)}} \mathcal{T}_{i-(j+1)} = \overline{\mathcal{T}_{i-(j+1)}} \mathcal{T}_{i-j} \mathcal{T}_{i-(j+1)}.$$

We next introduce a simple characteristic of $\mathcal{F}_{i,j}^A$ and $\mathcal{F}_{i,j}^B$.

► **Observation 43.** For each $i \geq 2$ and $0 \leq j \leq i-1$, consider a factorization $\mathcal{X} = (x_k)_{k=1}^m$ of \mathcal{T}_i that contains all occurrences of \mathcal{T}_{i-j} (respectively, $\overline{\mathcal{T}_{i-j}}$). If no two consecutive factors are both different from \mathcal{T}_{i-j} (respectively, $\overline{\mathcal{T}_{i-j}}$), then \mathcal{X} is the smallest and thus $\mathcal{X} = \mathcal{F}_{i,j}^A$ (respectively, $\mathcal{X} = \mathcal{F}_{i,j}^B$).

The observation holds because, otherwise, we could merge the two consecutive factors and obtain a smaller factorization.

Now, we present the main result of the section, illustrated in Figure 8.



■ **Figure 8** An illustration of Theorem 44 for $1 \leq j \leq 4$. Operators $(\cdot)^-$ and $(\cdot)^{\bar{}}$ are defined in Definition 42. Notice the green occurrences of \mathcal{T}_{i-j} are introduced from \boxplus .

► **Theorem 44.** For $i \geq 2$ and $0 \leq j \leq i - 1$, the following statements hold.

- (1) For each $\mathcal{F}_{i,j} \in \{\mathcal{F}_{i,j}^A, \mathcal{F}_{i,j}^B\}$, let $\mathcal{F}_{i,j} = (x_t)_{t=1}^m$. Then each term x_t is in the set $\mathcal{B}_{i,j} := \{\mathcal{T}_{i-j}, \overline{\mathcal{T}_{i-j}}, \mathcal{T}_{i-(j+1)}, \overline{\mathcal{T}_{i-(j+1)}}\}$. Moreover, $x_1 = \mathcal{T}_{i-j}$. If j is even, then $x_m = \mathcal{T}_{i-j}$; otherwise, $x_m = \overline{\mathcal{T}_{i-j}}$.
- (2) $\mathcal{F}_{i,0}^A = (\mathcal{T}_i)$, $\mathcal{F}_{i,1}^A = (\mathcal{T}_{i-1}, \overline{\mathcal{T}_{i-1}})$, and for each $j \geq 2$,

$$\mathcal{F}_{i,j}^A = \begin{cases} (\mathcal{F}_{i,j-1}^A)^- \overline{(\mathcal{F}_{i,j-1}^B)^-}, & j \text{ is odd,} \\ (\mathcal{F}_{i,j-1}^A)^- \boxplus (\mathcal{F}_{i,j-1}^B)^-, & j \text{ is even.} \end{cases}$$

- (3) $\mathcal{F}_{i,0}^B = ()$, $\mathcal{F}_{i,1}^B = (\mathcal{T}_{i-1}, \overline{\mathcal{T}_{i-1}})$, and for each $j \geq 2$, $\mathcal{F}_{i,j}^B = (\mathcal{F}_{i,j-1}^B)^- \overline{(\mathcal{F}_{i,j-1}^A)^-}$.

Proof. We proceed by induction on j .

Base cases. The claim holds trivially for $j = 0$. When $j = 1$, $(\mathcal{T}_{i-1}, \overline{\mathcal{T}_{i-1}})$ is the smallest factorization following Observation 43 and Statement 1 holds. Next, note that $(\mathcal{F}_{i,1}^A)^- = (\mathcal{F}_{i,1}^B)^- = (\mathcal{T}_{i-2}, \overline{\mathcal{T}_{i-2}})$. When $j = 2$, it follows that

$$\mathcal{F}_{i,2}^A = (\mathcal{F}_{i,1}^A)^- \boxplus (\mathcal{F}_{i,1}^B)^- = (\mathcal{T}_{i-2}, \overline{\mathcal{T}_{i-2}}) \boxplus (\overline{\mathcal{T}_{i-2}}, \mathcal{T}_{i-2}) = (\mathcal{T}_{i-2}, \overline{\mathcal{T}_{i-2}}, \mathcal{T}_{i-2}, \overline{\mathcal{T}_{i-2}}) \quad (10)$$

and

$$\mathcal{F}_{i,2}^B = (\mathcal{F}_{i,1}^B)^- \overline{(\mathcal{F}_{i,1}^A)^-} = (\mathcal{T}_{i-2}, \overline{\mathcal{T}_{i-2}}) \overline{(\overline{\mathcal{T}_{i-2}}, \mathcal{T}_{i-2})} = (\mathcal{T}_{i-2}, \overline{\mathcal{T}_{i-2}}, \overline{\mathcal{T}_{i-2}}, \mathcal{T}_{i-2}) \quad (11)$$

Both factorizations are the smallest following Observation 43, and Statement 1 holds.

16:22 Net Occurrences in Fibonacci and Thue-Morse Words

Inductive step. Let k be an odd integer such that $3 \leq k \leq i-1$, and assume the claim holds for $j = k-1$. Specifically, assume $\mathcal{F}_{i,k-1}^A = (x_t)_{t=1}^m$, $\mathcal{F}_{i,k-1}^B = (y_t)_{t=1}^l$, $x_1 = y_1 = \mathcal{T}_{i-(k-1)}$, and $x_m = y_l = \mathcal{T}_{i-(k-1)}$. We now prove the result for $j = k$.

First, note that both $\mathcal{F}_{i,k-1}^A$ and $\mathcal{F}_{i,k-1}^B$ are factorizations of \mathcal{T}_i . Then, by the definition of operation $(\cdot)^-$, both $(\mathcal{F}_{i,k-1}^A)^-$ and $(\mathcal{F}_{i,k-1}^B)^-$ are factorizations of \mathcal{T}_{i-1} , and $(\mathcal{F}_{i,k-1}^B)^-$ is a factorization of $\overline{\mathcal{T}_{i-1}}$. Now, since $\mathcal{T}_i = \mathcal{T}_{i-1} \overline{\mathcal{T}_{i-1}}$, it follows that $\mathcal{Y}_{\text{odd}}^A := (\mathcal{F}_{i,k-1}^A)^- \overline{(\mathcal{F}_{i,k-1}^B)^-}$ is a factorization of \mathcal{T}_i . It remains to show that $\mathcal{Y}_{\text{odd}}^A = \mathcal{F}_{i,j}^A$. Observe that

$$\mathcal{Y}_{\text{odd}}^A = (\mathcal{F}_{i,k-1}^A)^- \overline{(\mathcal{F}_{i,k-1}^B)^-} = (x_1)^- \cdots (x_{m-1})^- \mathcal{T}_{i-k} \overline{\mathcal{T}_{i-k}} \overline{(y_2)^-} \cdots \overline{(y_l)^-}. \quad (12)$$

By the induction hypothesis on $\mathcal{F}_{i,k-1}^A$ and $\mathcal{F}_{i,k-1}^B$ and the definition of operation $(\cdot)^-$, we know that $(\mathcal{F}_{i,k-1}^A)^-$ contains all the occurrences of \mathcal{T}_{i-k} in \mathcal{T}_{i-1} , and $(\mathcal{F}_{i,k-1}^B)^-$ contains all the occurrences of \mathcal{T}_{i-k} in $\overline{\mathcal{T}_{i-1}}$. Since $\mathcal{T}_i = \mathcal{T}_{i-1} \overline{\mathcal{T}_{i-1}}$ and \mathcal{T}_i is overlap-free, it follows that $\mathcal{Y}_{\text{odd}}^A$ contains all the occurrences of \mathcal{T}_{i-k} . Moreover, no two consecutive factors of $\mathcal{Y}_{\text{odd}}^A$ are both different from \mathcal{T}_{i-k} , so by Observation 43, we conclude that $\mathcal{Y}_{\text{odd}}^A = \mathcal{F}_{i,j}^A$.

Next we show that all factors of $\mathcal{F}_{i,j}^A$ are elements of $\mathcal{B}_{i,k}$. Since $x_t \in \mathcal{B}_{i,k-1}$ for each $1 \leq t \leq m$ and $y_t \in \mathcal{B}_{i,k-1}$ for each $1 \leq t \leq l$, it follows from Equation (12) that all factors of $\mathcal{F}_{i,j}^A$ are elements of $\mathcal{B}_{i,k}$. Additionally, the first factor in $\mathcal{F}_{i,j}^A$ is $(x_1)^- = \mathcal{T}_{i-k}$, and the last factor in $\mathcal{F}_{i,j}^A$ is $\overline{(y_l)^-} = \overline{\mathcal{T}_{i-k}}$ since $k-1$ is even.

Similarly, we can show that $\mathcal{Y}_{\text{odd}}^B := (\mathcal{F}_{i,k-1}^B)^- \overline{(\mathcal{F}_{i,k-1}^A)^-}$ is a factorization of \mathcal{T}_i , that $\mathcal{Y}_{\text{odd}}^B = \mathcal{F}_{i,j}^B$, and that Statement 1 holds.

We can prove analogously when k is even. In this case, the operation \boxplus is used to ensure that $\mathcal{F}_{i,j}^A$ contains all occurrences of \mathcal{T}_{i-j} . Specifically, when k is even, we have $(x_m)^- = \overline{(y_1)^-} = \overline{\mathcal{T}_{i-k}}$, and there is a occurrence of \mathcal{T}_{i-k} within $\overline{\mathcal{T}_{i-k}} \overline{\mathcal{T}_{i-k}} = \overline{\mathcal{T}_{i-(k+1)}} \mathcal{T}_{i-k} \mathcal{T}_{i-(k+1)}$. \blacktriangleleft