The Smallest String Attractors of Fibonacci and Period-Doubling Words

Mutsunori Banbara
Graduate School of Informatics, Nagoya University, Japan

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

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

Dominik Köppl
Department of Computer Science and Engineering, University of Yamanashi, Kofu, Japan

Takuya Mieno
Graduate School of Informatics and Engineering, University of Electro-Communications, Chofu, Japan

Yoshio Okamoto
Graduate School of Informatics and Engineering, University of Electro-Communications, Chofu, Japan

Abstract

A string attractor of a string $T[1..|T|]$ is a set of positions Γ of T such that any substring w of T has an occurrence that crosses a position in Γ, i.e., there is a position i such that $w = T[i..i+|w|-1]$ and the intersection $[i, i+|w|-1] \cap \Gamma$ is nonempty. The size of the smallest string attractor of Fibonacci words is known to be 2. We completely characterize the set of all smallest string attractors of Fibonacci words, and show a recursive formula describing the $2^{n-4} + 2^{\lceil n/2 \rceil - 2}$ distinct position pairs that are the smallest string attractors of the nth Fibonacci word for $n \geq 7$. Similarly, the size of the smallest string attractor of period-doubling words is known to be 2. We also completely characterize the set of all smallest string attractors of period-doubling words, and show a formula describing the two distinct position pairs that are the smallest string attractors of the nth period-doubling word for $n \geq 2$. Our results show that strings with the same smallest attractor size can have a drastically different number of distinct smallest attractors.

2012 ACM Subject Classification Mathematics of computing → Combinatorics on words

Keywords and phrases String attractors, Fibonacci words, Period-doubling words, Combinatorics on words

Digital Object Identifier 10.4230/LIPIcs.CPM.2026.33

Related Version *Extended Version*: `https://arxiv.org/abs/2602.16152`

Funding *Hideo Bannai*: JSPS KAKENHI Grant Number JP24K02899.
Dominik Köppl: JSPS KAKENHI Grant Number JP25K21150.
Takuya Mieno: JSPS KAKENHI Grant Number JP24K20734.
Yoshio Okamoto: JSPS KAKENHI Grant Number JP23K10982.

1 Introduction

For any string T, a set Γ of positions is a *string attractor* (or simply attractor) of T, if any substring of T has an occurrence that *crosses* (i.e., contains) a position in Γ. String attractors [5] are an important combinatorial notion that captures the repetitiveness (compressiveness) of the string in the sense that dictionary compression algorithms can be viewed as algorithms that compute string attractors. It is known that the size of the smallest string attractor of T, denoted by $\gamma(T)$, bounds all dictionary compression measures from below, and is NP-hard to compute [5].

The study of string attractors on well-known families of strings has attracted much attention. Mantaci et al. [10, 11] initiated the study of string attractors on standard Sturmian words (which include the Fibonacci words), and Thue–Morse words. For any standard Sturmian word T, they showed that at least one of the sets $\Gamma_1 = \{\eta+1, \eta+2\}$ or $\Gamma_2 = \{|T|-\eta-3, |T|-\eta-2\}$ is a smallest string attractor of T, where η is the length of the longest proper palindromic prefix of $T[1..|T|-2]$. For de Bruijn words of length n, they showed that the smallest attractor size asymptotically approaches $n/\log n$. They also showed an attractor of size $\Theta(\log|\mathcal{T}_n|)$ for the nth Thue–Morse word $\mathcal{T}_n$. Later, Kutsukake et al. [7] showed that $\gamma(\mathcal{T}_n) = 4$ for all $n \geq 4$. In a more general setting, Schaeffer and Shallit [14] discussed string attractors of all prefixes of automatic sequences, including the period-doubling words, Thue–Morse words, and Tribonacci words. In particular, they showed that the size of the smallest attractor of any prefix of the period-doubling word longer than one is 2, using the Walnut theorem prover [13].

For the Fibonacci words, defined as $F_0 = \mathtt{b}$, $F_1 = \mathtt{a}$, $F_n = F_{n-1}F_{n-2}$, it holds that $\eta = f_{n-1} - 2$ (where $f_{n-1} = |F_{n-1}|$), and the above result of Mantaci et al. translates to the sets $\Gamma_1 = \{f_{n-1}-1, f_{n-1}\}$ or $\Gamma_2 = \{f_{n-2}-1, f_{n-2}\}$. Since Γ_2 cannot be an attractor (it can be shown that the length f_{n-1} suffix of F_n occurs uniquely and does not have an occurrence crossing a position in Γ_2), it follows that Γ_1 is an attractor of F_n. Hence, this gives an explicit pair of positions to be a smallest attractor of Fibonacci words. However, in general, a given string can have multiple smallest string attractors. For example, any single position of a unary string is an attractor. For a more interesting example, the set of all smallest string attractors of F_8 is shown in Figure 1. See also Figure 13 for more examples.

In this paper, we take a deeper look into the combinatorial aspects of string attractors of Fibonacci words and period-doubling words, and give a complete characterization of their smallest string attractors. Our characterization for Fibonacci words is based on occurrences of *singular words* [15] in the Fibonacci word. We show a recursive formula describing the set of all $2^{n-4} + 2^{\lceil n/2 \rceil - 2}$ smallest string attractors of the nth Fibonacci word for all $n \geq 7$. For the nth period-doubling word D_n defined as $D_0 = \mathtt{a}$ and $D_n = \phi(D_{n-1})$ for the morphism $\phi(\mathtt{a}) = \mathtt{ab}$ and $\phi(\mathtt{b}) = \mathtt{aa}$, we show that the position pairs $\{3 \cdot 2^{n-3}, 6 \cdot 2^{n-3}\}$ and $\{4 \cdot 2^{n-3}, 6 \cdot 2^{n-3}\}$ characterize the smallest attractors for $n \geq 3$. To the best of our knowledge, this is the first work which shows a complete characterization of the set of smallest string attractors for non-trivial families of strings.

It is also worth noting that, although the smallest attractor size for both the Fibonacci and period-doubling words is two, our results show that the *number* of smallest attractors can differ drastically, namely, from exponential in the length to constant. We do not yet have a clear interpretation of what this quantity captures, and its understanding remains an open and interesting question, as it may allow us to consider more fine-grained types of repetitiveness. Other related open problems are listed in Section 5.

Related Work

Mieno et al. [12] gave a complete characterization of smallest straight-line programs (SLPs) of Fibonacci words, showing that an SLP of a Fibonacci word is a smallest SLP if and only if it can be obtained by some implementation of the RePair algorithm [8], i.e., the recursive pairing of a most frequent adjacent symbol pair. The number of distinct smallest SLPs for F_n was shown to be $2\lfloor (n+1)/2 \rfloor - 2$ for $n \geq 3$.

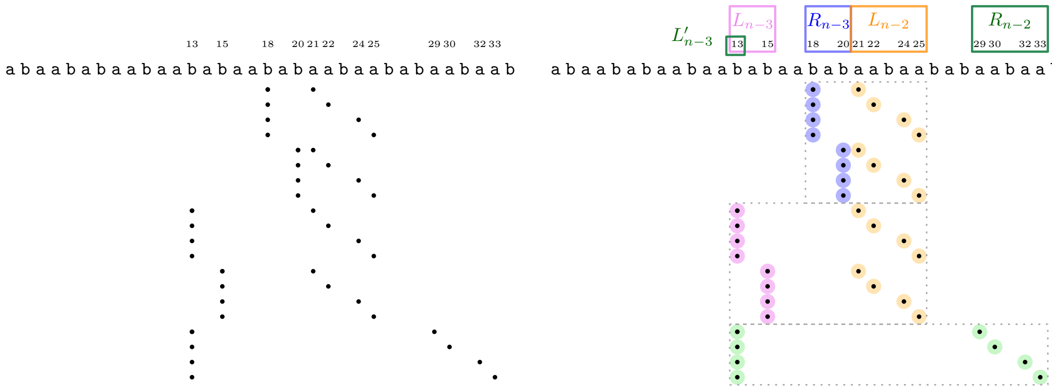

Figure 1 For $n = 8$. Left: all smallest attractors of F_n, where each pair of horizontally aligned dots represents an attractor. Right: characterization of these attractors according to Theorem 23; the three dotted boxes (top to bottom) correspond to Lemmas 25–27. (See Figure 18 in Appendix B for the case $n = 9$.)

We note that for the strings we have considered, the approach of Schaeffer and Shallit [14] and the software Walnut [13] can be applied, and it is possible to compute a finite state automaton for symbols in $(\{0,1\} \times \{0,1\} \times \{0,1\})^*$, such that accepting paths spell out a representation of the two attractor positions and the length of a prefix of the string. While the correctness of our claims can, in principle, be verified by analyzing and characterizing these paths, this would not seem to give us much insight into why the statements hold. This approach is discussed briefly in Appendix A.

Schaeffer and Shallit [14] and Cassaigne et al. [1] have studied another related attractor-based parameter *span*(w), defined as the smallest distance between the leftmost and rightmost positions in an attractor of w.

2 Preliminaries

Let $\Sigma = \{\mathtt{a}, \mathtt{b}\}$ be a binary alphabet. An element of Σ is called a symbol. An element of Σ^* is called a string. For a string x, let $|x|$ denote its length, and let ε denote the empty string, i.e., $|\varepsilon| = 0$. For any integer $1 \le i \le |x|$, let $x[i]$ denote the ith symbol of x, and for any $1 \le i \le j \le |x|$, let $x[i..j] = x[i] \cdots x[j]$, and $x[i..j) = x[i..j-1]$. Let x^R denote the *reverse* of x, i.e., if $x = x[1]x[2] \cdots x[|x|]$, then $x^R = x[|x|]x[|x|-1] \cdots x[1]$. The *complement* of a symbol in Σ is denoted as $\overline{\mathtt{a}} = \mathtt{b}$ and $\overline{\mathtt{b}} = \mathtt{a}$, and for any string $x \in \Sigma^*$, its complement is $\overline{x} = \overline{x[1]} \cdots \overline{x[|x|]}$.

For a set X of integers and another integer i, let $i \oplus X = X \oplus i := \{i + x : x \in X\}$ and $i \ominus X := \{i - x : x \in X\}$. For two sets X and Y, let $X \otimes Y := \{\{x, y\} : (x, y) \in X \times Y\}$. For a string T and an occurrence $w = T[i..j]$ of a substring w of T, we say that the occurrence of w *crosses* position k, if $i \le k \le j$. A *boundary* is a pair of consecutive positions sometimes identified with two strings ending and starting at the respective positions. We say that an occurrence of a string crosses the boundary if it crosses both positions of the boundary.

▶ **Definition 1** (String Attractors [5]). *A* string attractor*, or an* attractor *for short, of a string T is a set Γ of positions of T such that for any substring u of T, there exists an occurrence of u in T that crosses some position in Γ. We denote by $\mathcal{A}(T)$ the set of all smallest attractors of T.*

It is straightforward that string attractors are invariant under string reversal, meaning that the "mirrored" string attractor is a string attractor of the mirrored string.

▶ **Observation 2.** *If Γ is an attractor of a string w, then $(|w|+1) \ominus \Gamma$ is an attractor of w^R.*

A string w is a *minimal unique substring* (MUS) [4] of T if w is a substring of T that has a unique occurrence in T, and any proper substring of w has multiple occurrences in T. The following observation is straightforward.

▶ **Observation 3.** *Any string attractor must contain at least one position that is crossed by the unique occurrence of a MUS.*

3 The Smallest Attractors of Fibonacci Words

3.1 Properties of Fibonacci Words and Singular Words

▶ **Definition 4** (F_n and f_n). *For each $n \geq 0$, the nth* Fibonacci word *is defined as $F_0 = \mathtt{b}, F_1 = \mathtt{a}$, and $F_n = F_{n-1}\ F_{n-2}$ for each $n \geq 2$. The nth* Fibonacci number *is defined as $f_0 = 1, f_1 = 1$, and $f_n = f_{n-1} + f_{n-2}$ for each $n \geq 2$. Note that $f_n = |F_n|$.*

▶ **Definition 5** (G_n, Δ_n). *For $n \geq 2$, let $G_n = F_n[1..f_n - 2]$ and $F_n = G_n\Delta_n$. For convenience, let $\Delta_0 = \mathtt{ab}$ and $\Delta_1 = \mathtt{ba}$.*

Below are known or simple observations concerning Fibonacci words.

▶ **Observation 6** (e.g. [2, 6, 9]). *The following propositions hold.*
1. *For $i \geq 2$, $\Delta_i = \Delta_{i \bmod 2}$.*
2. *For $i \geq 2$, G_i is a palindrome (by $G_i = G_{i-2}\Delta_{i-2}G_{i-3}\Delta_{i-3}G_{i-2}$ and induction).*
3. *For $i \geq 3$, $F_i = F_{i-1}F_{i-2} = F_{i-2}G_{i-1}\Delta_i$.*
4. *For $i \geq 5$,*

$$\begin{aligned} F_i = G_i\Delta_i = F_{i-1}G_{i-2}\Delta_i = G_{i-2}\Delta_{i-2}G_{i-3}\Delta_{i-3}G_{i-2}\Delta_i = G_{i-2}(F_{i-3})^R(F_{i-2})^R\Delta_i \\ = G_{i-2}(F_{i-1})^R\Delta_i. \end{aligned}$$

▶ **Definition 7** (Singular Words). *Let $S_{-1} = \varepsilon$, $S_0 = \mathtt{a}$, $S_1 = \mathtt{b}$, and $S_n = S_{n-2}S_{n-3}S_{n-2}$ for $n \geq 2$.*

Notice that $|F_n| = |S_n|$, and all singular words are palindromes. In the literature [3, 15], these words have been defined to be $S_n = F_n[f_n - 1]F_n[1..f_n - 1] = F_n[f_n - 1]G_nF_n[f_n - 1]$. However, we start from the above definition (Property 2 (4) in [15]) to simplify the presentation. The next lemma summarizes results shown in [15] that we utilize.

▶ **Lemma 8** ([15]). *For any $n \geq 0$, the following propositions hold.*

(1) *$S_n = S_{n \bmod 2}\left(\prod_{i=0}^{n-2} S_i\right)$.* *(Property 2 (12) in [15])*

(2) *S_n is not a substring of S_{n+1} nor of $\prod_{i=0}^{n-1} S_i$.* *(Properties 2 (2) and (13) in [15])*

(3) *Let $S_nS_{n+1} = u_1u_2u_3$ where u_1 is a non-empty proper prefix of S_n and u_3 is a non-empty proper suffix of S_{n+1}. Then, $u_2 \neq S_i$ for any $i \geq 0$.* *(Lemma 3 in [15])*

The following lemma is essentially the same as what has been shown by Wen and Wen (Theorem 1 in [15]) for the infinite Fibonacci string.

▶ **Lemma 9.** *For any $n \geq 1$, $F_n = \left(\prod_{i=0}^{n-2} S_i\right) S_{(n-1) \bmod 2}$.*

Proof. Induction on n. The equation holds for $n = 1$. Suppose it holds for all $n \le k$. Then,

$$F_{k+1} = F_k F_{k-1} = \left(\prod_{i=0}^{k-2} S_i\right) S_{(k-1) \bmod 2} \left(\prod_{i=0}^{k-3} S_i\right) S_{(k-2) \bmod 2} = \left(\prod_{i=0}^{k-1} S_i\right) S_{k \bmod 2}$$

where the last equation uses $S_{(k-1) \bmod 2}\left(\prod_{i=0}^{k-3} S_i\right) = S_{k-1}$ by Lemma 8 (1). Therefore, the equation holds for $n = k + 1$. ◀

▶ **Corollary 10.** *For any $k \ge 0$, S_k occurs uniquely in $S_0 \cdots S_k$ and the first (leftmost) occurrence of S_k in F_∞ spans the range $[f_{k+1}..f_{k+2} - 1]$.*

Proof. S_k cannot occur in $\prod_{i=0}^{k-1} S_i$ nor contain S_{k-1} (Lemma 8 (2)), nor cross the boundary of $S_{k-1}S_k$ (Lemma 8 (3)). The range follows from $|S_k| = f_k$ and $\sum_{i=0}^{k-1} f_i = f_{k+1} - 1$. ◀

▶ **Lemma 11.** *For any $n \ge 5$, S_{n-3} and S_{n-2} are the (only) MUSs of F_n.*

Proof. S_{n-3} occurs uniquely in $F_n = \left(\prod_{i=0}^{n-2} S_i\right) S_{(n-1) \bmod 2}$; It cannot occur after its first occurrence (characterized by Corollary 10) since it cannot occur in S_{n-2} (Lemma 8 (2)) nor cross the boundary of $S_{n-3}S_{n-2}$ as well as the boundary of $S_{n-2}S_{n-1}$ (Lemma 8 (3)). The latter also implies the uniqueness of S_{n-2} in F_n. We have $S_{n-3} = F_n[f_{n-2}..f_{n-1} - 1]$, $S_{n-2} = F_n[f_{n-1}..f_n - 1]$ from Corollary 10, and any substring containing them are unique. Since $F_n = G_{n-1}\Delta_{n-1}F_{n-2} = F_{n-2}G_{n-1}\Delta_n$ by Observation 6, it holds that $F_n[1..f_{n-1}-2] = F_n[f_{n-2} + 1..f_n - 2] = G_{n-1}$ and $F_n[f_{n-1} + 1..f_n] = F_{n-2}$ are repeating substrings. These are the maximal substrings of F_n not containing S_{n-3} nor S_{n-2}. All substrings of these maximal substrings are repeating as well, and include all proper substrings of S_{n-3} and S_{n-2}. Therefore S_{n-3} and S_{n-2} are MUSs. Also, since a substring containing a MUS must be unique, there can be no other MUSs in F_n. ◀

▶ **Definition 12** (U_n and V_n)**.** *For each $n \ge 5$, define $U_n = [f_{n-2}..f_{n-1} - 1]$ and $V_n = [f_{n-1}..f_n - 1]$, i.e., the ranges of positions crossed by the MUSs S_{n-3} and S_{n-2}, respectively.*

▶ **Proposition 13** ([10, Thm. 5])**.** *The size of the smallest string attractor of F_n is 2 for $n \ge 2$.*

For any $n \ge 5$, a Fibonacci word has exactly two MUSs and the ranges of positions crossed by their occurrence are disjoint (Corollary 10, Lemma 11, Definition 12). Since an occurrence of a unique substring must contain an attractor position (Observation 3) and from Proposition 13, the smallest attractor must consist of one position from each of the two MUS ranges.

▶ **Observation 14.** *For $n \ge 5$, $\mathcal{A}(F_n) \subseteq U_n \otimes V_n$.*

In Section 3.2, we first characterize a subset of position pairs in $U_n \otimes V_n$ that cannot be an attractor of F_n by arguments based on the occurrence of singular words in F_n. We then prove in Section 3.3 that all other position pairs are attractors of F_n.

3.2 Invalid Attractor Position Pairs

Here, we characterize position pairs in $U_n \otimes V_n$ that cannot be attractors of F_n by using occurrences of singular words in F_n. We first show that all occurrences of singular words S_i in F_n can be captured by their occurrences in our recursive definition (Definition 7). We interpret the recursion as a grammar rule, and show that each occurrence of singular words

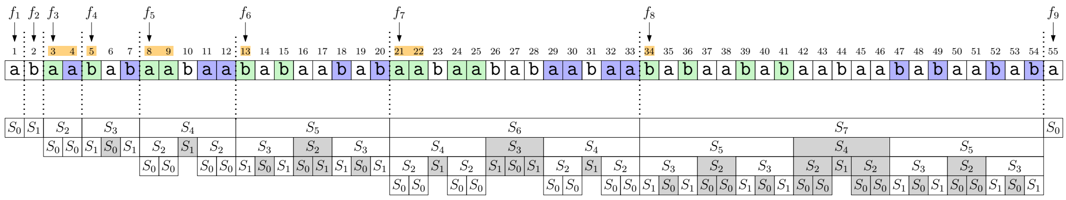

Figure 2 Illustration of Lemma 15 and Definition 18. Top: F_9. Bottom: the parse tree of F_9 based on singular words. For $2 \le k \le 7$: the gray blocks are descendants of the center child S_{k-3} in the parse tree for S_k, i.e. $S_k = S_{k-2}S_{k-3}S_{k-2}$. The green and blue blocks correspond to positions in L_k and R_k, respectively. (Notice that these positions are not crossed by any center child in the parse tree for S_k.) The positions corresponding to L'_k are highlighted directly in orange.

in F_n corresponds to a node in the parse tree of this grammar, where a node in the parse tree is considered to span exactly the interval of positions corresponding to the substring it derives. See Figure 2 for a concrete example.

▶ **Lemma 15.** *Consider the equations of Lemma 9 and Definition 7 as grammar rules that derive F_n for $n \ge 1$, i.e., $\{F_n \to \left(\prod_{i=0}^{n-2} S_i\right) S_{(n-1) \bmod 2}, S_{-1} \to \varepsilon, S_0 \to \mathtt{a}, S_1 \to \mathtt{b}\} \cup \{S_i \to S_{i-2}S_{i-3}S_{i-2} \mid i = 2, \ldots, n-2\}$. Then, for any $i \ge 0$, every occurrence of S_i in F_n corresponds to a node in the parse tree of this grammar.*

Proof. Due to Corollary 10, there is no occurrence of S_i in F_n before its first (leftmost) occurrence in a node of the parse tree. Suppose there is a later occurrence of S_i that does not correspond to a node in the parse tree of the grammar, and consider the lowest common ancestor (LCA) in the parse tree of the leftmost position and rightmost position of the occurrence. From the above observation, if the LCA is F_n, this implies that an occurrence of S_i crosses a boundary of S_kS_{k+1} for some $k \ge i$ which violates Lemma 8 (3). Therefore, the LCA must be S_k for some $k > i$. Furthermore, since S_i does not occur in S_{i+1} (Lemma 8 (2)), the LCA must be S_k for some $k > i+1$, thus, $i \le k-2$. By definition, $S_k = S_{k-2} \cdot S_{k-3} \cdot S_{k-2}$. Therefore, S_i must have an occurrence in S_k (i) crossing both boundaries between S_{k-2} and S_{k-3}, and between S_{k-3} and S_{k-2} (only when $i = k-2$), (ii) crossing only the boundary of S_{k-2} and S_{k-3}, or (iii) crossing only the boundary of S_{k-3} and S_{k-2}. Case (i) cannot happen since S_{k-3} is not a substring of S_{k-2} (Lemma 8 (2)). Case (ii) directly violates Lemma 8 (3) and Case (iii) also violates Lemma 8 (3) due to the fact that S_i is a palindrome and the statement of Lemma 8 (3) holds for S_kS_{k-1} as well. ◀

In the parse tree of F_n based on the grammar defined in the statement of Lemma 15, we say that the occurrence of S_i is a *center child* of S_{i+3} if it is derived by the rule $S_{i+3} \to S_{i+1}S_iS_{i+1}$. Note that $S_2 \to S_0S_0$ has no (or an empty) center child. We say a position p is crossed by a node in the parse tree if the occurrence of the string derived by the node crosses p.

▶ **Lemma 16.** *Let $n \ge 5$ and $\{u, v\} \in U_n \otimes V_n$. If u or v is crossed by a center child in the parse tree of some S_i, then $\{u, v\}$ is not an attractor of F_n.*

Proof. See Figure 3. U_n and V_n correspond to the ranges of positions crossed by the MUSs S_{n-3} and S_{n-2} of F_n, respectively (Definition 12). Therefore, an attractor of size 2 of F_n must contain exactly one position from each of the ranges. By definition, $S_{n-3} = S_{n-5}S_{n-6}S_{n-5}$ and $S_{n-2} = S_{n-4}S_{n-5}S_{n-4}$. Here, S_{n-4} only occurs twice as children of S_{n-2}, and does not occur elsewhere in the strings S_{n-3} and S_{n-2} (Lemma 15). Therefore, an attractor position must be crossed by one of the two occurrences of S_{n-4}, and any position crossed by the center child S_{n-5} of S_{n-2} cannot be chosen as an attractor position, since it would not be

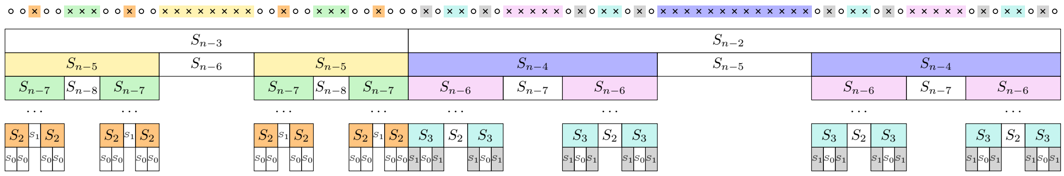

Figure 3 Illustration of the proof of Lemma 16 for $n = 11$. Circles and crosses at the top mark valid and invalid attractor positions, respectively. A cross highlighted in a given color indicates that the position is invalid because it fails to be crossed by the substring shown in the same color. (Note that some combinations of the positions marked by circles in S_{n-3} and S_{n-2} are still invalid; these are characterized in Lemma 17.)

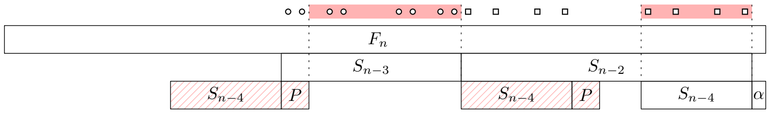

Figure 4 Illustration of the proof of Lemma 17 for $n = 9$, where $P = S_{n-3}[1..2]$ and $\alpha = S_{(n-1) \bmod 2}$. Circles (resp. squares) at the top mark positions in U_n (resp. V_n) that are not ruled out by Lemma 16. A circle-square pair in $U_n \otimes V_n$ is invalid if they are both in a red region since they fail to be crossed by $S_{n-4}\ P$. (All other circle-square pairs are indeed valid, which we prove in Section 3.3.)

crossed by S_{n-4}. Similarly, there are three occurrences of S_{n-5}; two as children of S_{n-3} and one as a center child of S_{n-2}. However, since we cannot choose an attractor position in the center child of S_{n-2}, an attractor position in S_{n-3} must be crossed by one of the two occurrences of S_{n-5}. Therefore, a position crossed by the center child S_{n-6} of S_{n-3} cannot be chosen as an attractor position. The argument can be repeated recursively; For $j = 1, \cdots, \lceil n/2 \rceil - 2$, given that the attractor positions must be crossed by occurrences of $S_{n-2j-1} = S_{n-2j-3}S_{n-2j-4}S_{n-2j-3}$ and $S_{n-2j} = S_{n-2j-2}S_{n-2j-3}S_{n-2j-2}$ that were not center children in the sub-tree of the parse tree respectively of S_{n-3} and S_{n-2}, attractor positions in the respective center children S_{n-2j-4} and S_{n-2j-3} cannot be chosen. ◀

▶ **Lemma 17.** *Let $n \geq 7$ and $\{u, v\} \in U_n \otimes V_n$. If $u \in U_n \setminus \{\min U_n, \min U_n + 1\}$ and $v \in V_n$ is in the right child of the first occurrence of S_{n-2}, then $\{u, v\}$ is not an attractor of F_n.*

Proof. See Figure 4. Consider the occurrences of $S_{n-4}(S_{n-3}[1..2])$ in F_n. Due to Lemma 15, S_{n-4} has only three occurrences in $S_{n-4}S_{n-3}S_{n-2}S_{(n-1) \bmod 2}$, and thus in F_n, two of which are the left and right child of $S_{n-2} = S_{n-4}S_{n-5}S_{n-4}$. Since the last occurrence is only followed by a single symbol $S_{(n-1) \bmod 2}$, it follows that there can only be two occurrences of $S_{n-4}(S_{n-3}[1..2])$ in F_n. Therefore, for an attractor position to be crossed by $S_{n-4}(S_{n-3}[1..2])$, if one of the first two positions in U_n (corresponding to the first two positions of S_{n-3}) is not chosen as an attractor position, then, we cannot choose a position in the right child of S_{n-2} for the other attractor position. Here, notice that we require $n \geq 7$, since otherwise, the second occurrence of $S_{n-4}(S_{n-3}[1..2])$ crosses the boundary of the center and right children of S_{n-2}. ◀

3.3 Valid Attractor Position Pairs

In this subsection, we show that the position pairs in $U_n \otimes V_n$ that are not invalidated by Lemma 16 nor Lemma 17 are indeed attractors of F_n.

▶ **Definition 18** (L_k, R_k, and L'_k). *For $k \geq 2$, we define three subsets of positions within the first occurrence of S_k; let u_k be the node in the parse tree corresponding to this occurrence. Let L_k be the set of positions in the left child of u_k that are <u>not</u> crossed by any center child. Similarly, let R_k be the set of positions in the right child of u_k that are <u>not</u> crossed by any center child. Further, let L'_k be the set of positions corresponding to the occurrence of $S_{2-(k \bmod 2)}$ in the leftmost path in the parse tree rooted at u_k (u_k itself when $u_k = S_2$).*

For example, $L_6 = \{21, 22, 24, 25\}$, $R_6 = \{29, 30, 32, 33\}$, $L'_6 = \{21, 22\}$. See Figure 2.

▶ **Observation 19.** *$L_2 = \{3\}$, $R_2 = \{4\}$, $L'_2 = \{3, 4\}$, $L_3 = \{5\}$, $R_3 = \{7\}$, $L'_3 = \{5\}$, and*

$$L_k = (L_{k-2} \cup R_{k-2}) \oplus f_k, \tag{1}$$

$$R_k = (L_{k-2} \cup R_{k-2}) \oplus f_{k+1} = L_k \oplus f_{k-1}, \tag{2}$$

$$L'_k = L'_{k-2} \oplus f_k. \tag{3}$$

for $k \geq 4$. For $k \geq 2$,

$$L'_k = \begin{cases} \{f_{k+1}, f_{k+1} + 1\} & \textit{if } k \textit{ is even,} \\ \{f_{k+1}\} & \textit{if } k \textit{ is odd.} \end{cases} \tag{4}$$

Proof. Equations (1) and (3) hold because, by Lemma 9, in the factorization $F_n = S_0 S_1 \cdots \underline{S_{k-2}} S_{k-1} \underline{S_k} \cdots S_{n-2} S_{(n-1) \bmod 2}$, the offset between the starting positions of the first occurrences of S_{k-2} and S_k is $|S_{k-2}S_{k-1}| = f_k$. Equation (4) follows directly from Equation (3). Equation (2) holds because, in $S_k = S_{k-2}S_{k-3}S_{k-2}$, the offset between the starting positions of the two occurrences of S_{k-2} is $|S_{k-2}S_{k-3}| = f_{k-1}$. ◀

The following is straightforward from the definitions and the above observation.

▶ **Observation 20.** *For $k \geq 2$, $\min L_k = f_{k+1}$, $\max L_k = 2f_k - 1$, $\min R_k = f_{k+1} + f_{k-1}$, and $\max R_k = f_{k+2} - 1$.*

The following holds since all singular words are palindromes and thus the parse tree rooted at each singular word is symmetric, implying that for any $k \geq 2$, L_k and R_k are a "mirrored" image of the other with respect to some mid-point.

▶ **Observation 21.** *For any $k \geq 2$, $R_k = (f_{k+3} - 1) \ominus L_k$, $L_{k+2} = (f_{k+4} - 1) \ominus (L_k \cup R_k)$.*

Using the sets of Definition 18, the results of Section 3.2 can be summarized as follows.

▶ **Corollary 22.** *For $n \geq 7$, $\mathcal{A}(F_n) \subseteq ((L_{n-3} \cup R_{n-3}) \otimes L_{n-2}) \cup (L'_{n-3} \otimes R_{n-2})$.*

Proof. Lemma 16 implies that an attractor position pair must be in $(L_{n-3} \cup R_{n-3}) \otimes (L_{n-2} \cup R_{n-2})$, where $L_{n-3} \cup R_{n-3} \subseteq U_n$ and $L_{n-2} \cup R_{n-2} \subseteq V_n$. Furthermore, Lemma 17 implies that if one of the attractor positions is in R_{n-2}, then the other position must be in the set $\{\min U_n, \min U_n + 1\} = \{f_{n-2}, f_{n-2} + 1\}$. If $n - 3$ is even, this set is exactly L'_{n-3}. If $n - 3$ is odd, L'_{n-3} only includes f_{n-2}, but then $f_{n-2} + 1$ is a center child and cannot be an attractor position by Lemma 16. Therefore, the other position will be in L'_{n-3}. These imply that $\mathcal{A}(F_n) \subseteq ((L_{n-3} \cup R_{n-3}) \otimes L_{n-2}) \cup (L'_{n-3} \otimes R_{n-2})$. ◀

Below, we show that all position pairs in this set are actually attractors of F_n.

▶ **Theorem 23.** *For any $n \geq 7$, $\mathcal{A}(F_n) = ((L_{n-3} \cup R_{n-3}) \otimes L_{n-2}) \cup (L'_{n-3} \otimes R_{n-2})$.*

▶ Note 24. Since $L'_{n-3} \subseteq L_{n-3}$ for $n \geq 6$, Theorem 23 implies $L'_{n-3} \otimes (L_{n-2} \cup R_{n-2}) \subseteq \mathcal{A}(F_n)$.

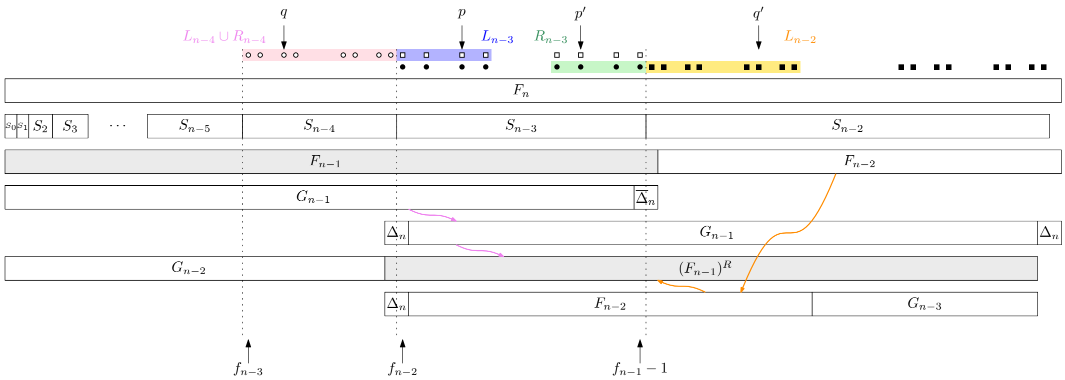

Figure 5 Illustration of Lemma 25 for $n = 10$. White and black markers at the top denote positions in $\mathcal{A}(F_{n-1})$ and $\mathcal{A}(F_n)$, respectively. Highlighted markers, together with the highlighted occurrences of F_{n-1} and $(F_{n-1})^R$, illustrate the mapping described in the proof. Pink and orange squiggly arrows illustrate Case 1 and Case 3, respectively. (Note that $\Delta_{n-1} = \overline{\Delta_n}$ and $(\Delta_{n-1})^R = \Delta_n$.)

Since Corollary 22 establishes $\mathcal{A}(F_n) \subseteq ((L_{n-3} \cup R_{n-3}) \otimes L_{n-2}) \cup (L'_{n-3} \otimes R_{n-2})$, we prove $\mathcal{A}(F_n) \supseteq ((L_{n-3} \cup R_{n-3}) \otimes L_{n-2}) \cup (L'_{n-3} \otimes R_{n-2})$ using the following three lemmas.

▶ **Lemma 25.** *For $n \geq 8$, if $(L_{n-4} \cup R_{n-4}) \otimes L_{n-3} \subseteq \mathcal{A}(F_{n-1})$, then $R_{n-3} \otimes L_{n-2} \subseteq \mathcal{A}(F_n)$.*

Proof. See Figure 5. For any $\{q, p\} \in (L_{n-4} \cup R_{n-4}) \otimes L_{n-3} \subseteq \mathcal{A}(F_{n-1})$ with $q < p$, let $p' = f_n - p - 1$ and $q' = f_n - q - 1$. Since $p \in L_{n-3}$ and $q \in L_{n-4} \cup R_{n-4}$, $p' = (f_n - 1) - p \in R_{n-3}$ and $q' = (f_n - 1) - q \in L_{n-2}$ by Observation 21. Thus, $\{p', q'\} \in R_{n-3} \otimes L_{n-2}$. Since the mapping $\{p, q\} \mapsto \{p', q'\}$ is bijective, it suffices to show that $\{p', q'\} \in \mathcal{A}(F_n)$.

From Observation 6, $G_n = F_{n-1}G_{n-2} = G_{n-2}(F_{n-1})^R$ is a prefix of F_n. Essentially, p', q' are positions in F_n that correspond to the positions p, q in F_{n-1}, mapped to the occurrence of $(F_{n-1})^R$, i.e., $p' = |G_n| - p + 1 = f_n - 2 - p + 1 = f_n - p - 1$ and $q' = |G_n| - q + 1 = f_n - 2 - q + 1 = f_n - q - 1$. Thus, by Observation 2, all substrings of $(F_{n-1})^R$ must have an occurrence that crosses p' or q'. In order to prove that $\{p', q'\}$ is an attractor of F_n, consider any substring $u = F_n[i..j]$ that does not cross p' nor q'. Notice that since $p \geq \min L_{n-3} = f_{n-2}$, we have $p' \leq f_n - f_{n-2} - 1 = f_{n-1} - 1$. Also, since $q \leq \max R_{n-4} = f_{n-2} - 1$, we have $q' \geq f_n - (f_{n-2} - 1) - 1 = f_{n-1}$.

1. If $j < p'$, then $j \leq f_{n-1} - 2$, so u is a substring of G_{n-1}, which is a substring of $(F_{n-1})^R$.
2. If $p' < i \leq j < q'$, then u is trivially a substring of $(F_{n-1})^R$.
3. If $q' < i$, then $i \geq f_{n-1} + 1$ so u is a substring of F_{n-2}, which is a substring of $(F_{n-1})^R$ because $(F_{n-1})^R = (\Delta_{n-1})^R G_{n-1} = (\Delta_{n-1})^R F_{n-2} G_{n-3}$.

In all these cases, u is a substring of $(F_{n-1})^R$, which means u has an occurrence that crosses p' or q'. Therefore, $\{p', q'\}$ is an attractor of F_n. ◀

▶ **Lemma 26.** *For $n \geq 7$, if $R_{n-3} \otimes L_{n-2} \subseteq \mathcal{A}(F_n)$, then $L_{n-3} \otimes L_{n-2} \subseteq \mathcal{A}(F_n)$.*

Proof. See Figure 6. For any $\{p, q\} \in R_{n-3} \otimes L_{n-2} \subseteq \mathcal{A}(F_n)$ with $p < q$, let $p' = p - f_{n-4}$. We have $\{p', q\} \in L_{n-3} \otimes L_{n-2}$ since $R_{n-3} = L_{n-3} \oplus f_{n-4}$ (Equation (2)). Since the mapping $\{p, q\} \mapsto \{p', q\}$ is bijective, it suffices to show that $\{p', q\} \in \mathcal{A}(F_n)$.

Since $\{p, q\}$ is an attractor of F_n, all substrings of F_n have an occurrence crossing p or q. Consider substrings of F_n that have occurrences crossing p but not q nor p'. Since $\min L_{n-3} = f_{n-2} \leq p' < p < q \leq \max L_{n-2} = 2f_{n-2} - 1$ and $F_n = \underbrace{\overbrace{F_{n-2}}^{f_{n-2}} G_{n-2}}_{2f_{n-2}-2} \Delta_n G_{n-3} \Delta_n$, all such occurrences are in a substring of G_{n-2}. Furthermore, they have an occurrence

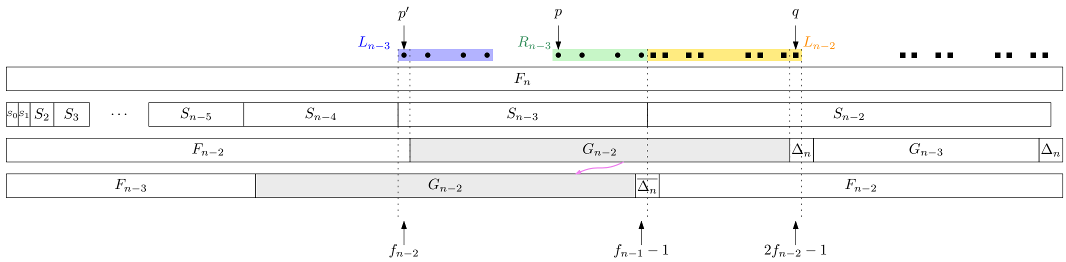

Figure 6 Illustration of Lemma 26 for $n = 10$. Black markers at the top denote positions in $\mathcal{A}(F_n)$. Relevant attractor positions and occurrences of G_{n-2} are highlighted.

crossing $p' = p - f_{n-4}$, since $F_n = F_{n-1}F_{n-2} = F_{n-3}G_{n-2}\Delta_{n-1}F_{n-2}$ and G_{n-2} also occurs $f_{n-4} = f_{n-2} - f_{n-3}$ positions to the left. Thus, all substrings of F_n have an occurrence crossing p' or q, and therefore, $\{p', q\}$ is an attractor of F_n. ◀

▶ **Lemma 27.** *For $n \geq 9$, if $L'_{n-5} \otimes (L_{n-4} \cup R_{n-4}) \subseteq \mathcal{A}(F_{n-2})$, then $L'_{n-3} \otimes R_{n-2} \subseteq \mathcal{A}(F_n)$.*

Proof. See Figure 7. For any $\{p, q\} \in L'_{n-5} \otimes (L_{n-4} \cup R_{n-4}) \subseteq \mathcal{A}(F_{n-2})$ with $p \in L'_{n-5}$ and $q \in L_{n-4} \cup R_{n-4}$, let $p' = p + f_{n-1}$, $q' = q + f_{n-1}$, and $p'' = p + f_{n-3} = p' - f_{n-2}$. Notice that $f_{n-2} \leq p'' \leq f_{n-2} + 1$, $p' \leq 2f_{n-2} + 1$, and $q' \leq f_n - 1$ can be seen from Equation (4) and Observation 20. It can be also seen that $p'' \in L'_{n-5} \oplus f_{n-3} = L'_{n-3}$ (Equation (3)), and $q' \in (L_{n-4} \cup R_{n-4}) \oplus f_{n-1} = R_{n-2}$ (Equation (2)). Thus, $\{p'', q'\} \in L'_{n-3} \otimes R_{n-2}$. Since the mapping $\{p, q\} \mapsto \{p'', q'\}$ is bijective, it suffices to show that $\{p'', q'\} \in \mathcal{A}(F_n)$.

Consider any substring $u = F_n[i..j]$ that does not cross p'' nor q' and the following cases.

1. If $j < p''$, then u is a substring of F_{n-2} since $p'' \leq f_{n-2} + 1$.
2. If $p'' < i \leq j < p'$, then u is a substring of F_{n-2} since $f_{n-2} \leq p'' < p' \leq 2f_{n-2} + 1$, and $(F_{n-2})^2$ is a prefix of $F_n = F_{n-1}F_{n-2} = F_{n-2}F_{n-3}F_{n-4}F_{n-5}F_{n-4} = (F_{n-2})^2F_{n-5}F_{n-4}$.
3. If $p'' < i < p' < j < q'$, then u is a substring of G_{n-1} which crosses p', since $f_{n-2} \leq p'' < q' \leq f_n - 1$ and $F_n = F_{n-2}G_{n-1}\Delta_n$.
4. If $p' < i$, then u is a substring of F_{n-2}, since $F_n = F_{n-1}F_{n-2}$ and $p' \geq 2f_{n-2} + 1 > f_{n-1}$.

For Case 3, u has an occurrence that crosses p'', since $p'' = p' - f_{n-2}$, and there is an occurrence of G_{n-1}, f_{n-2} positions to the left (as a prefix of F_n). It remains to show that all substrings of F_{n-2} have an occurrence crossing p'' or q'. Since $\{p, q\}$ is an attractor of F_{n-2}, $F_n = F_{n-1}F_{n-2} = F_{n-2}F_{n-3}F_{n-2}$, and $\{p', q'\} = \{p, q\} \oplus f_{n-1}$, all substrings of F_{n-2} have an occurrence that crosses p' or q'. Furthermore, as was the case for Case 3, any substring of F_{n-2} crossing p' but not crossing q' is a substring of G_{n-1}, and has an occurrence crossing p''. Therefore, any substring will cross p'' or q', and $\{p'', q'\}$ is an attractor of F_n. ◀

Proof of Theorem 23. The proof is by induction. $\mathcal{A}(F_7)$ and $\mathcal{A}(F_8)$ can be confirmed by brute force computation:

$$\begin{aligned}
\mathcal{A}(F_7) &= ((L_4 \cup R_4) \otimes L_5) \cup (L'_4 \otimes R_5) \\
&= ((\{8,9\} \cup \{11,12\}) \otimes \{13,15\}) \cup (\{8,9\} \otimes \{18,20\}) \\
&= \{\{8,13\},\{8,15\},\{8,18\},\{8,20\},\{9,13\},\{9,15\},\{9,18\},\{9,20\}, \\
&\quad \{11,13\},\{11,15\},\{12,13\},\{12,15\}\}, \text{ and} \\
\mathcal{A}(F_8) &= ((L_5 \cup R_5) \otimes L_6) \cup (L'_5 \otimes R_6) \\
&= ((\{13,15\} \cup \{18,20\}) \otimes \{21,22,24,25\}) \cup (\{13\} \otimes \{29,30,32,33\}) \\
&= \{\{13,21\},\{13,22\},\{13,24\},\{13,25\},\{13,29\},\{13,30\},\{13,32\},\{13,33\}, \\
&\quad \{15,21\},\{15,22\},\{15,24\},\{15,25\},\{18,21\},\{18,22\},\{18,24\},\{18,25\}, \\
&\quad \{20,21\},\{20,22\},\{20,24\},\{20,25\}\}.
\end{aligned}$$

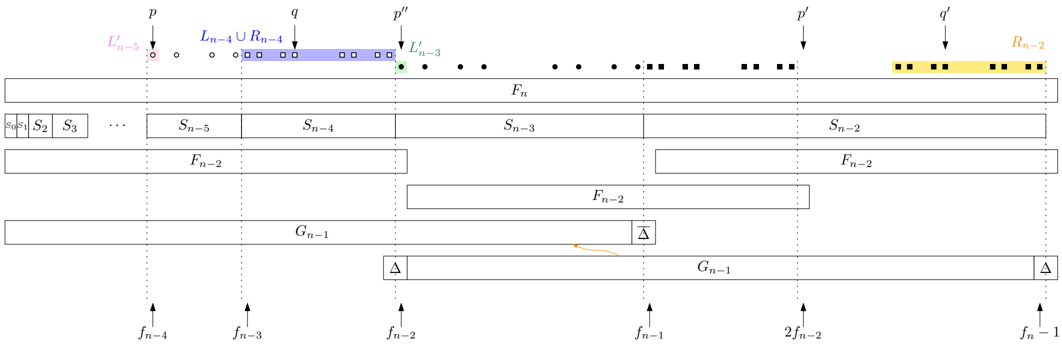

Figure 7 Illustration of Lemma 27 for $n = 10$. White and black markers at the top denote positions in $\mathcal{A}(F_{n-2})$ and $\mathcal{A}(F_n)$, respectively.

Under the induction hypothesis and Note 24, the conditions of Lemmas 25–27 hold for $n \geq 9$, and thus, together with Corollary 22 and the above base cases, the theorem holds. ◀

▶ **Corollary 28.** *For $k \geq 4$, $|\mathcal{A}(F_{2k-1})| = (2^{k-3}+1)2^{k-2}$ and $|\mathcal{A}(F_{2k})| = (2^{k-2}+1)2^{k-2}$.*

Proof. It is clear that $|L_2| = |R_2| = |L_3| = |R_3| = 1$, and $|L_i| = |R_i| = 2|L_{i-2}| = 2|R_{i-2}|$ for $i \geq 4$. Therefore, $|L_{2k-1}| = |R_{2k-1}| = 2^{k-2}$ and $|L_{2k}| = |R_{2k}| = 2^{k-1}$. Since $|L'_{2k-1}| = 1$ and $|L'_{2k}| = 2$, we have from Theorem 23 that $|\mathcal{A}(F_{2k-1})| = |((L_{2k-4} \cup R_{2k-4}) \otimes L_{2k-3}) \cup (L'_{2k-4} \otimes R_{2k-3})| = (2 \cdot 2^{k-3} \cdot 2^{k-3}) + 2 \cdot 2^{k-3} = (2^{k-3}+1)2^{k-2}$ and $|\mathcal{A}(F_{2k})| = |((L_{2k-3} \cup R_{2k-3}) \otimes L_{2k-2}) \cup (L'_{2k-3} \otimes R_{2k-2})| = (2 \cdot 2^{k-3} \cdot 2^{k-2}) + 2^{k-2} = (2^{k-2}+1)2^{k-2}$. ◀

4 The Smallest Attractors of Period-Doubling Words

▶ **Definition 29** (D_n)**.** *For each $n \geq 0$, the nth period-doubling word is defined as $D_0 = \mathtt{a}$, and $D_n = \phi(D_{n-1})$ where ϕ is a morphism defined by $\phi(\mathtt{a}) = \mathtt{ab}$ and $\phi(\mathtt{b}) = \mathtt{aa}$. Note that $|D_n| = 2^n$.*

▶ **Theorem 30.** *$\mathcal{A}(D_2) = \{\{2,3\},\{2,4\}\}$. For $n \geq 3$, $\mathcal{A}(D_n) = \{\{p_n, r_n\}, \{q_n, r_n\}\}$, where $p_n = 3 \cdot 2^{n-3}$, $q_n = 2^{n-1}$, and $r_n = 3 \cdot 2^{n-2}$.*

Schaeffer and Shallit [14] showed that for arbitrary prefixes of length at least 6 of the infinite period-doubling sequence, $\{p_n, r_n\}$ is a smallest attractor when the length is in $[2^n, 3 \cdot 2^{n-1})$, and $\{q_n, q_{n+1}\}$, when the length is in $[3 \cdot 2^{n-1}, 2^{n+1})$. We consider only the length 2^n prefixes, but characterize all smallest string attractors.

▶ **Observation 31.** *Since any occurrence of $\mathtt{b}$ in D_n must have come from an occurrence of $\mathtt{a}$ in D_{n-1} by $\phi(\mathtt{a}) = \mathtt{ab}$, an occurrence of $\mathtt{b}$ in D_n is always at an even position. Hence, $\mathtt{bb}$ cannot occur. This implies that an occurrence of $\mathtt{aa}$ starting at an odd position must be followed (provided it is not the end of the string) and preceded by $\mathtt{ab}$, since it could only have come from $\phi(\mathtt{aba})$.*

Furthermore, the above implies the following corollary.

▶ **Corollary 32.** *Any occurrence of a substring x in D_n with $|x| \geq 3$ has the same parity, i.e., $|\{i \bmod 2 \mid D_n[i..i+|x|) = x\}| = 1$.*

▶ **Observation 33.** *The following are known or easily verifiable properties of period-doubling words.*

(1) *For* $n \geq 1$, $(D_{n-1})^2 = D'_n\bar{c}$ *where* $D'_n = D_n[1..2^n)$ *and* $c = D_n[|D_n|]$.

(2) *For* $n \geq 2$, $D_n = D_{n-1}(D_{n-2})^2 = D'_{n-1}cD'_{n-1}\bar{c}$, *where* $c = D_{n-1}[|D_{n-1}|]$.

(3) *For any* i *such that* $0 \leq n - 2i < n$, $(D_{n-2i})^2$ *is a suffix of* D_n, *and for any* $0 \leq i < n$, D_i *is a prefix of* D_n.

We first show that the two position pairs are indeed attractors of D_n.

▶ **Lemma 34.** *For* $n \geq 3$, $\{\{p_n, r_n\}, \{q_n, r_n\}\} \subseteq \mathcal{A}(D_n)$

Proof. The statement can be verified by brute force for $n = 3$ and 4. Suppose the statement holds for all $3 \leq n < k$. Let $X = D_k[i..j]$ $(1 \leq i \leq j \leq 2^k)$ be an arbitrary substring of D_k, and let Y be the shortest even-length substring of D_k that contains the occurrence of X, starting at an odd position, and ending at an even position, i.e., $Y = D_k[2i' - 1..2j']$ where $i' = \lceil \frac{i}{2} \rceil$, $j' = \lceil \frac{j}{2} \rceil$. It holds that $Y = \phi(Y')$ where $Y' = D_{k-1}[i'..j']$. Since $\{p_{k-1}, r_{k-1}\}$ (resp. $\{q_{k-1}, r_{k-1}\}$) are attractors of D_{k-1}, Y' has an occurrence $D_{k-1}[i''..j''] = Y'$ crossing p_{k-1} (resp. q_{k-1}) or r_{k-1}, i.e., $i'' \leq s_{k-1} \leq j''$ for some $s_{k-1} \in \{p_{k-1}, q_{k-1}, r_{k-1}\}$. Then, since $2i''-1 < 2s_{k-1} = s_k \leq 2j''$, $Y[2..|Y|]$ in the corresponding occurrence of $Y = D_k[2i''-1..2j'']$ will cross s_k. Since $Y[2..|Y|-1]$ is a substring of X, X has an occurrence crossing s_k or ending just before s_k. See Figure 8. Thus, to prove that $\{p_k, r_k\}$ and $\{q_k, r_k\}$ are attractors of D_k, it remains to show that any substring X ending just before p_k (resp. q_k) or r_k will have an occurrence crossing p_k (resp. q_k) or r_k.

Since $p_k = |D_{k-2}D_{k-3}|$, $q_k = |D_{k-1}|$, and $D_k = D'_{k-1}cD'_{k-1}\bar{c}$ (Observation 33 (2)), any substring X ending just before p_k (resp. q_k) has an occurrence ending just before $p_k + 2^{k-1} = |D_{k-1}D_{k-2}D_{k-3}|$ (resp. $q_k + 2^{k-1} = |D_k|$). Since $D_k = D_{k-1}(D_{k-2})^2 = D_{k-1}D_{k-2}D_{k-3}(D_{k-4})^2$ and $r_k = |D_{k-1}D_{k-2}|$, any such X not crossing r_k must be a suffix of D'_{k-3} (resp. D'_{k-2}), having an occurrence ending just before r_k. See Figure 9.

Finally, we claim that any substring X ending just before r_k will have occurrences crossing both p_k and q_k, or r_k. We consider the following $O(k)$ disjoint cases depending on $|X|$.

1. $|X| > |D_{k-3}D'_{k-2}|$. See Figure 10 (a). Since the starting position of X is $r_k - |X| < r_k - |D_{k-3}D'_{k-2}| = 3 \cdot 2^{k-2} - 2^{k-3} - 2^{k-2} + 1 \leq p_k + 1$, X crosses both p_k and q_k.
2. $|D_{k-3}D'_{k-2}| \geq |X| \geq |(D_{k-4})^2|$. See Figure 10 (a) and Figure 11. Since $D_{k-3}D'_{k-2} = D_k[p_k + 1..r_k - 1] = D_k[p_k + 1 - |D_{k-3}|..r_k - 1 - |D_{k-3}|]$, it follows that $D_{k-3}D_{k-3}D'_{k-2}$ has a period of $|D_{k-3}|$. Since $|X| \geq |(D_{k-4})^2| > |D'_{k-3}|$, X starts with a suffix of D_{k-3} and therefore will have occurrences crossing p_k and q_k.
3. $|(D_{k-2i})^2| > |X| \geq |D_{k-2i}| = |(D_{k-2i-1})^2|$ for $i \geq 2, k - 2i - 1 \geq 0$. See Figure 10 (b) and Figure 12. Due to Observation 33 (3), there is an occurrence of $(D_{k-2i})^2$ centered at (i.e., the first copy ending at) both $r' = r_k - |(D_{k-2i})|$ and r_k. Since X is a suffix of $D_{k-2i}D'_{k-2i}$ and crosses r', it also has an occurrence crossing r_k.
4. $|(D_{k-2i-1})^2| > |D'_{k-2i}| \geq |X| \geq |D_{k-2i-1}| = |(D_{k-2i-2})^2|$ for $i \geq 2, k - 2i - 2 \geq 0$. See Figure 10 (b) and (c), and Figure 12. X is a suffix of D'_{k-2i}. There is an occurrence of $(D_{k-3})^2$ centered at both p_k and q_k. Due to Observation 33 (3), there must be an occurrence of $(D_{k-2i-1})^2$ also centered at p_k and q_k, and thus an occurrence of D'_{k-2i} (Observation 33 (1)) of which X is a suffix. Since $|X| \geq |D_{k-2i-1}| > |D'_{k-2i-1}|$, X has occurrences that cross p_k as well as q_k.
5. $|X| = 1$. $D_3[p_3] = D_3[q_3] = \mathtt{a}$ and $D_3[r_3] = \mathtt{b}$. It is easy to see that $D_k[p_k] = D_k[q_k] = c$ and $D_k[r_k] = \bar{c}$ from the definition of ϕ, so X will cross both p_k and q_k, or r_k.

Thus, $\{\{p_k, r_k\}, \{q_k, r_k\}\} \subseteq \mathcal{A}(D_k)$, proving the lemma. ◀

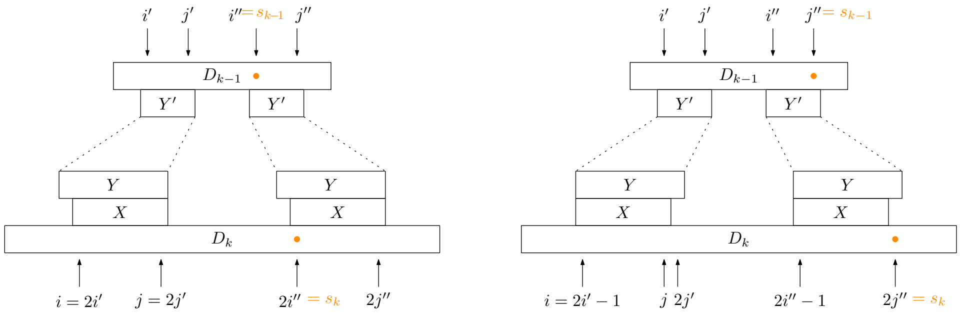

Figure 8 Illustration of the first paragraph of the proof of Lemma 34. Left: $X = Y[2..|Y|]$ and $i'' = s_{k-1}$. Right: $X = [1..|Y| - 1]$ and $j'' = s_{k-1}$. (Note that if $X = Y$, $X = Y[2..|Y|]$, or $X = [1..|Y| - 1]$ and $j'' < s_{k-1}$, then X has an occurrence crossing s_k.)

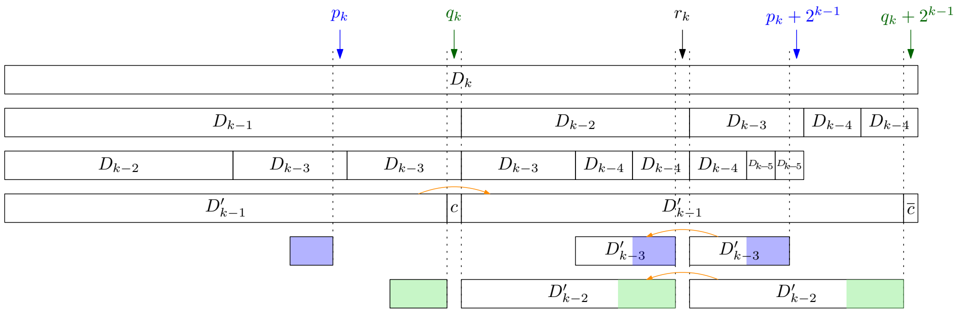

Figure 9 Illustration of the second paragraph of the proof of Lemma 34. Blue and green blocks represent substrings that have an occurrence ending just before p_k and q_k, respectively, but do not have an occurrence crossing r_k when considering its occurrence in the second half of the string via D'_{k-1}. Such substrings have occurrences ending just before r_k.

Next, we will show that no other position pairs can be attractors of D_n with the help of the following lemma.

▶ **Lemma 35.** *For $n \geq 4$, let $\{p, q\} \in \mathcal{A}(D_n)$ and $D_n[p] = \mathtt{a}$ and $D_n[q] = \mathtt{b}$. Then, p, q are both even and $\{p', q'\} = \{p/2, q/2\} \in \mathcal{A}(D_{n-1})$.*

Proof. The strings aaa, aab, baa, bab, abab, aabaa, aabab, babab, babaa, ababab are all substrings of D_n for $n \geq 4$. It is clear that q is even from Observation 31.

We first claim that $D_n[p+1] \neq \mathtt{b}$. Suppose to the contrary that $D_n[p+1] = \mathtt{b}$. Then, for aaa to have an occurrence crossing p, $D_n[p-2..p+1] = \mathtt{aa}\underline{\mathtt{a}}\mathtt{b}$, where the underlined a corresponds to position p. From Observation 31, we have $D_n[q-1..q] = \mathtt{a}\underline{\mathtt{b}}$, where the underlined b corresponds to position q. In order for the substrings baa and bab to have an occurrence crossing one of these positions, $D_n[q-3..q+4] = \mathtt{aba}\underline{\mathtt{b}}\mathtt{aaab}$, where the bold a and ab are due to Observation 31. Next, in order for the substrings aabaa and aabab to have an occurrence crossing p or q, $D_n[p-2..p+3] = \mathtt{aa}\underline{\mathtt{a}}\mathtt{baa}$ and $D_n[q-5..q+4] = \mathtt{aababa}\underline{\mathtt{b}}\mathtt{aaab}$. However, then, there cannot be an occurrence of ababab that crosses p or q contradicting that they are an attractor.

Next, we claim that p is even. If it is odd, then, from the above arguments, $D_n[p..p+3] = \underline{\mathtt{a}}\mathtt{aab}$. We also have $D_n[q-1..q] = \mathtt{a}\underline{\mathtt{b}}$. For aab and abab to cross p or q, $D_n[q-3..q+2] = \mathtt{aaa}\underline{\mathtt{b}}\mathtt{ab}$. Furthermore, for aabaa to cross p or q, $D_n[p-3..p+3] = \mathtt{aab}\underline{\mathtt{a}}\mathtt{aab}$. However, then at least one of babab and babaa cannot cross p or q.

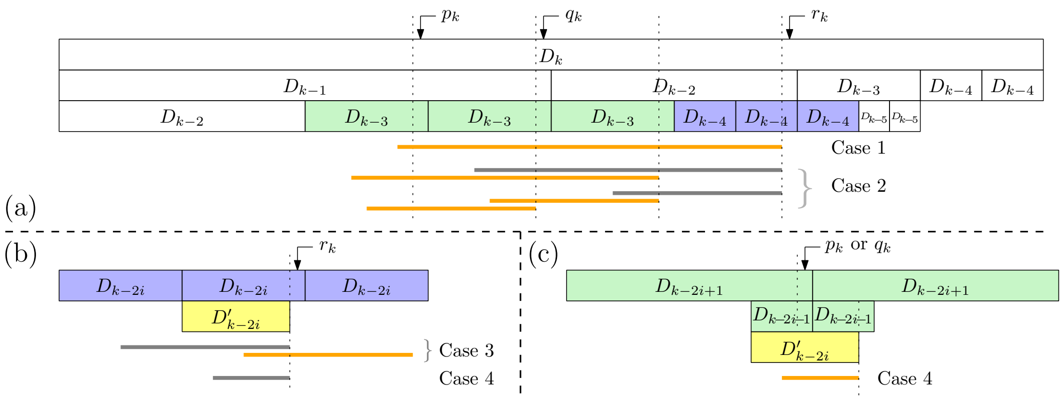

Figure 10 Illustration of the proof of Lemma 34. There are 4 main cases depending on the length of X which is a suffix of $D_k[1..r_k)$ and shown in gray (except for Case 1, which already crosses both p_k and q_k), and the occurrence crossing both p_k and q_k, or r_k is shown in orange.

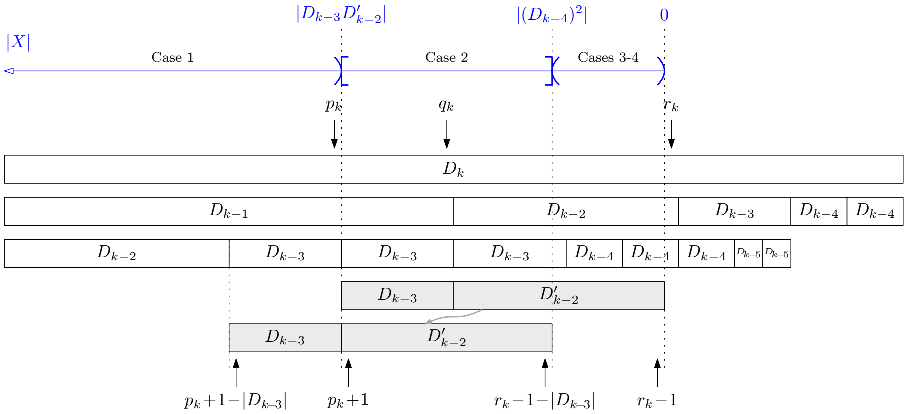

Figure 11 Illustration of Case 2 in the proof of Lemma 34.

Finally, consider any substring X' of D_{n-1}. If $|X'| = 1$, then, from the above arguments, $D_{n-1}[p'] = \mathtt{b}$ since $D_n[p] = \mathtt{a}$ could not have been derived by $\phi(\mathtt{a})$, and $D_{n-1}[q'] = \mathtt{a}$ since $D_n[q] = \mathtt{b}$ must have been derived by $\phi(\mathtt{a})$. Hence, X' crosses p' or q'. If $|X'| \geq 2$, consider the corresponding substring $X = \phi(X')$ in D_n. Since $\{p, q\}$ is an attractor of D_n, X has an occurrence crossing p or q. Furthermore, since $|X| \geq 4$ and the parity of its occurrences is determined uniquely (Corollary 32), an occurrence of X at position k in D_n crossing p or q implies an occurrence of X' at position $\lceil k/2 \rceil$ in D_{n-1} crossing p' or q'. ◀

Proof of Theorem 30. Assume the statement of Theorem 30 holds for all $3 \leq n < k$. Lemma 35 implies that any element other than $\{p_k, r_k\}$ or $\{q_k, r_k\}$ in $\mathcal{A}(D_k)$ would imply an element in $\mathcal{A}(D_{k-1})$ not in $\{\{p_{k-1}, r_{k-1}\}, \{q_{k-1}, r_{k-1}\}\}$ contradicting the induction hypothesis. Together with Lemma 34, this completes the proof. ◀

While we studied in earlier sections the smallest attractors of Fibonacci words and period-doubling words of higher orders (F_n for $n \geq 7$ and D_n for $n \geq 3$), for completeness, we list the smallest attractors of lower orders in Figure 13.

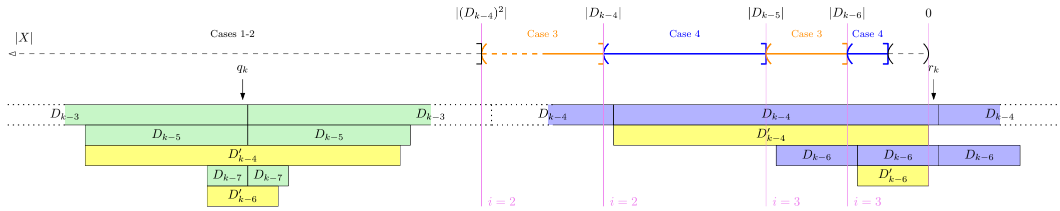

Figure 12 Illustration of Cases 3–4 in the proof of Lemma 34 for $i = 2$ and $i = 3$. The highlighted substrings follow the same color pattern in (b) and (c) of Figure 10.

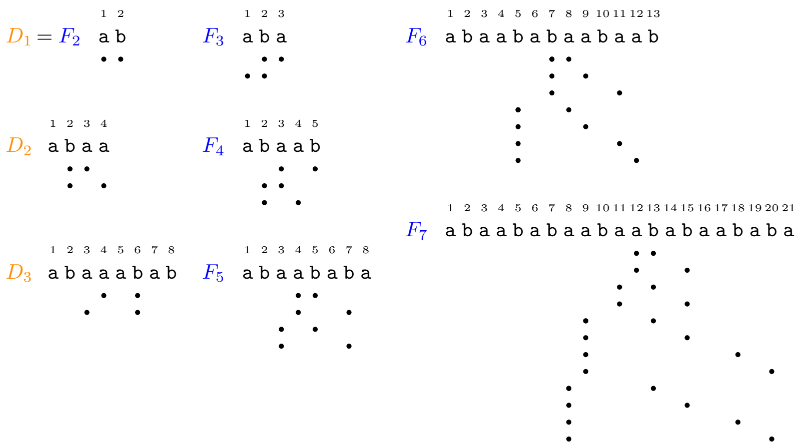

Figure 13 Smallest attractors of period-doubling words D_n for $n = 1, 2, 3$ and Fibonacci words F_n for $n = 3, 4, 5, 6, 7$. The visualization pattern follows that of Figure 1.

5 Discussion

Some natural related open problems are:

- Characterization of all smallest bidirectional macro schemes (BMS) for the same family of strings considered.
- Characterization of all smallest string attractors and BMS for arbitrary prefixes of the considered infinite sequence.

References

1. Julien Cassaigne, France Gheeraert, Antonio Restivo, Giuseppe Romana, Marinella Sciortino, and Manon Stipulanti. New string attractor-based complexities for infinite words. *J. Comb. Theory, Ser. A*, 208:105936, 2024. `doi:10.1016/J.JCTA.2024.105936`.
2. Aldo de Luca. A combinatorial property of the Fibonacci words. *Inf. Process. Lett.*, 12(4):193–195, 1981. `doi:10.1016/0020-0190(81)90099-5`.
3. Gabriele Fici. Factorizations of the Fibonacci infinite word. *J. Integer Seq.*, 18(9):15.9.3, 2015. URL: `https://cs.uwaterloo.ca/journals/JIS/VOL18/Fici/fici5.html`.
4. 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`.
5. Dominik Kempa and Nicola Prezza. At the roots of dictionary compression: string attractors. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 827–840. ACM, 2018. `doi:10.1145/3188745.3188814`.

6 Donald E. Knuth, James H. Morris Jr., and Vaughan R. Pratt. Fast pattern matching in strings. *SIAM J. Comput.*, 6(2):323–350, 1977. doi:10.1137/0206024.

7 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.

8 N. Jesper Larsson and Alistair Moffat. Offline dictionary-based compression. In *Data Compression Conference, DCC 1999, Snowbird, Utah, USA, March 29-31, 1999*, pages 296–305. IEEE Computer Society, 1999. doi:10.1109/DCC.1999.755679.

9 M. Lothaire. *Sturmian Words*, volume 90 of *Encyclopedia of Mathematics and its Applications*, pages 45–110. Cambridge University Press, 2002.

10 Sabrina Mantaci, Antonio Restivo, Giuseppe Romana, Giovanna Rosone, and Marinella Sciortino. String attractors and combinatorics on words. In Alessandra Cherubini, Nicoletta Sabadini, and Simone Tini, editors, *Proceedings of the 20th Italian Conference on Theoretical Computer Science, ICTCS 2019, Como, Italy, September 9-11, 2019*, volume 2504 of *CEUR Workshop Proceedings*, pages 57–71. CEUR-WS.org, 2019. URL: https://ceur-ws.org/Vol-2504/paper8.pdf.

11 Sabrina Mantaci, Antonio Restivo, Giuseppe Romana, Giovanna Rosone, and Marinella Sciortino. A combinatorial view on string attractors. *Theor. Comput. Sci.*, 850:236–248, 2021. doi:10.1016/J.TCS.2020.11.006.

12 Takuya Mieno, Shunsuke Inenaga, and Takashi Horiyama. RePair grammars are the smallest grammars for Fibonacci words. In Hideo Bannai and Jan Holub, editors, *33rd Annual Symposium on Combinatorial Pattern Matching, CPM 2022, Prague, Czech Republic, June 27-29, 2022*, volume 223 of *LIPIcs*, pages 26:1–26:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.CPM.2022.26.

13 Hamoon Mousavi. Automatic theorem proving in Walnut, 2021. arXiv:1603.06017.

14 Luke Schaeffer and Jeffrey Shallit. String attractors for automatic sequences, 2024. arXiv:2012.06840.

15 Zhi-Xiong Wen and Zhi-Ying Wen. Some properties of the singular words of the Fibonacci word. *Eur. J. Comb.*, 15(6):587–598, 1994. doi:10.1006/EUJC.1994.1060.

A Using Walnut to Characterize the Smallest Attractors

Using the approach by Schaeffer and Shallit [14], we can characterize the smallest attractors of period-doubling and Fibonacci words using the following Walnut program. Below, we only introduce the connection and do not give rigorous arguments.

```
def pdfaceq "k+i>=j & A u,v (u>=i & u<=j & v+j=u+k) => PD[u]=PD[v]":
def pdsa2 "(i1<n) & (i2<n) & Ak,l (k<=l & l<n) => (Er,s r<=s & s<n &
 (s+k=r+l) & $pdfaceq(k,l,s) & ((r<=i1 & i1<=s) | (r<=i2 & i2<=s)))":
def pdfa1 "An (n>=2) => Ei1,i2 $pdsa2(i1,i2,n)":
def pdfa2 "$pdsa2(i1,i2,n) & i1 < i2":

def fibfaceq "?msd_fib k+i>=j & A u,v (u>=i & u<=j & v+j=u+k) => F[u]=F[v]":
def fibsa2 "?msd_fib (i1<n) & (i2<n) & Ak,l (k<=l & l<n) =>
 (Er,s r<=s & s<n & (s+k=r+l) & $fibfaceq(k,l,s)
 & ((r<=i1 & i1<=s) | (r<=i2 & i2<=s)))":
def fibfa1 "?msd_fib An (n>=2) => Ei1,i2 $fibsa2(i1,i2,n)":
def fibfa2 "?msd_fib $fibsa2(i1,i2,n) & i1 < i2":
```

The code for the automaton `pdfa2` for period-doubling words is taken from [14]. The code for `fibfa2` only changes `PD` (period-doubling words) to `F` (Fibonacci words), and specifies the numeration system the corresponding automaton is defined for.

See Figures 14 and 15. While these automata characterize the set of all smallest string attractors of arbitrary prefixes of the period-doubling words or Fibonacci words, their interpretation is not straightforward. To limit the lengths of prefixes to the words we consider, we can modify the above to:

```
reg pow2 msd_2 "10*":
def pdsapow2 "(i1<n) & (i2<n) & (i1<i2) & $pow2(n) & Ak,l (k<=l & l<n) =>
 (Er,s r<=s & s<n & (s+k=r+l) & $pdfaceq(k,l,s)
 & ((r<=i1 & i1<=s) | (r<=i2 & i2<=s)))":

reg fibpow msd_fib "10*":
def fibsaf "?msd_fib (i1<n) & (i2<n) & (i1<i2) & $fibpow(n) &
 Ak,l (k<=l & l<n) => (Er,s r<=s & s<n & (s+k=r+l) & $fibfaceq(k,l,s)
 & ((r<=i1 & i1<=s) | (r<=i2 & i2<=s)))":
```

Figure 16 shows the sub-automaton `pdsapow2` of `pdfa2` for paths corresponding to prefixes of D_∞ of lengths 2^k, and can be used to verify the results of Theorem 30. For Fibonacci words, Figure 17 shows a sub-automaton `fibsaf` of `fibfa2` containing only the accepting paths for prefixes of length f_k, which can be used to verify the results of Theorem 23.

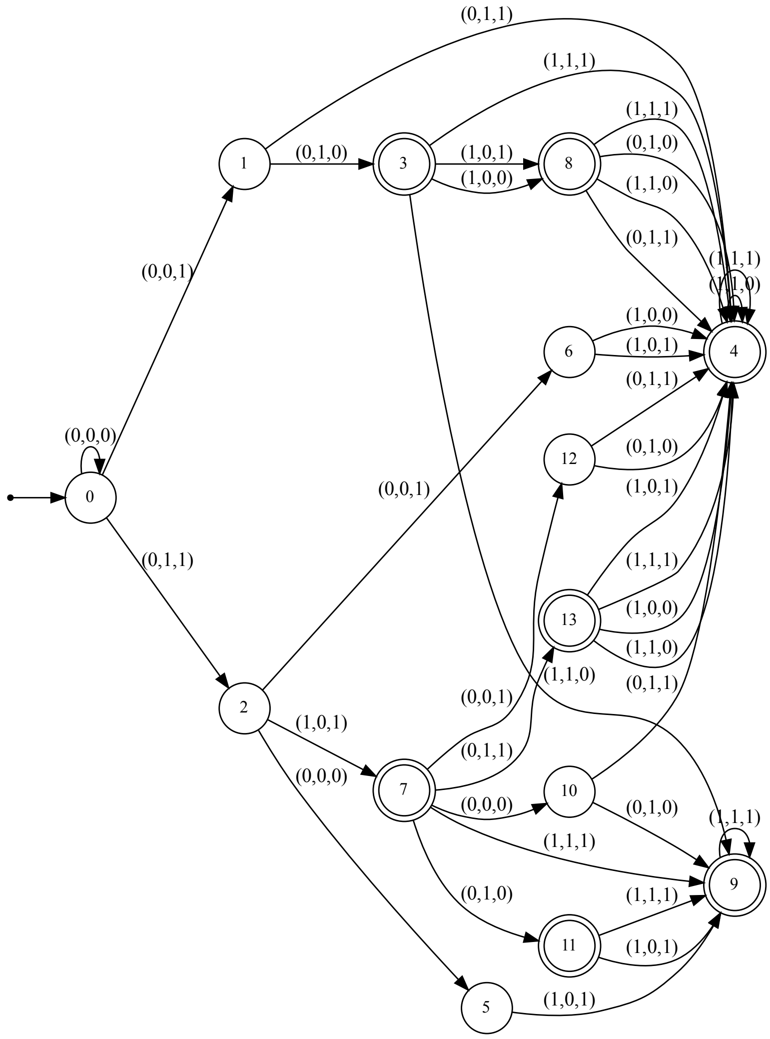

Figure 14 The automaton pdfa2 accepting a sequence $s \in (\{0,1\} \times \{0,1\} \times \{0,1\})^*$ if and only if $s[i] = (\mathsf{p}[i], \mathsf{q}[i], \mathsf{n}[i])$ where $\mathsf{p}, \mathsf{q}, \mathsf{n}$ are respectively the binary representation of integers $p-1$, $q-1$, n, satisfying $\{p, q\} \in \mathcal{A}(D_\infty[1..n])$ and $p < q$. Here, the -1 is due to the difference in the start index used for strings. By analyzing accepting paths for which $\mathsf{n} = 10^k$ implying $n = 2^k$ we can verify the statement of Theorem 30. See also Figure 16.

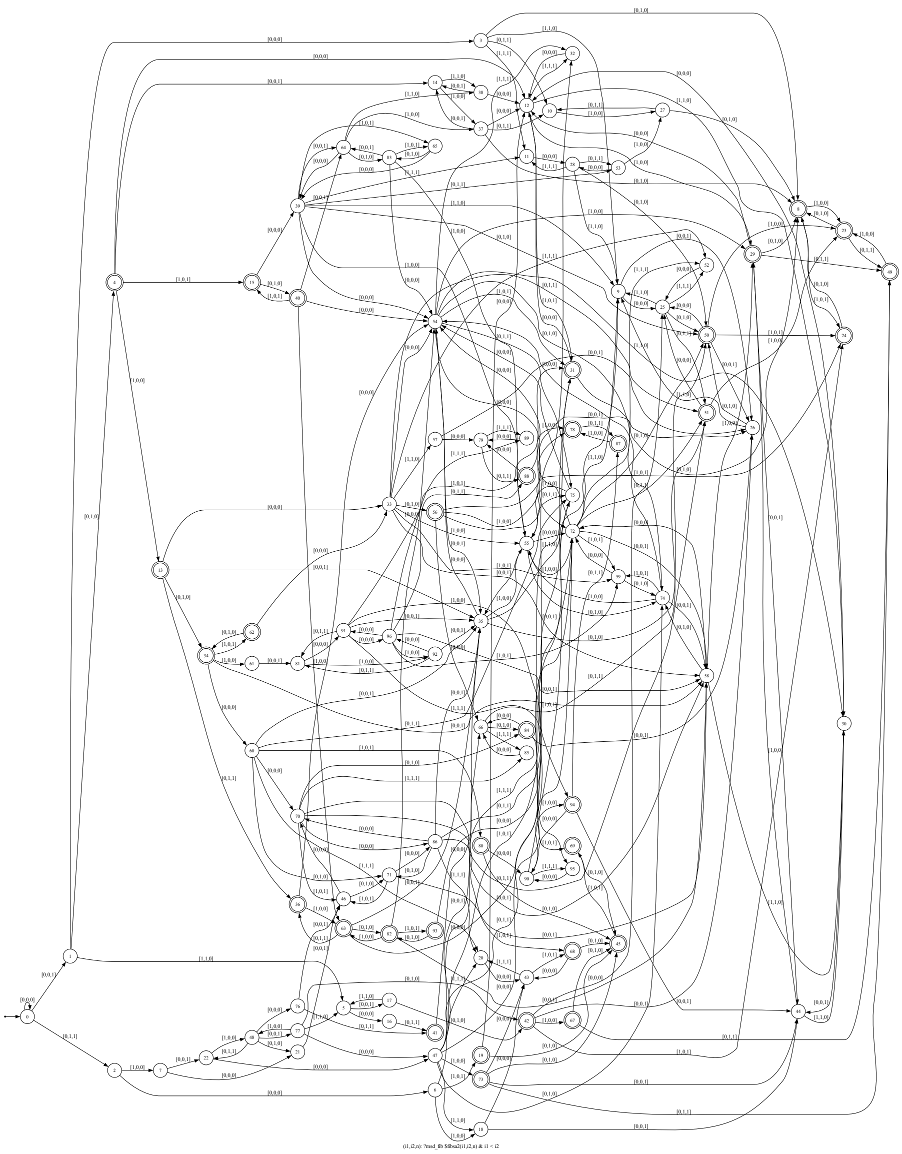

Figure 15 The automaton `fibsa2` accepting a sequence $s \in (\{0,1\} \times \{0,1\} \times \{0,1\})^*$ if and only if $s[i] = (\mathsf{p}[i], \mathsf{q}[i], \mathsf{n}[i])$ where $\mathsf{p}, \mathsf{q}, \mathsf{n}$ are respectively the bit representation, in Fibonacci numeration (Zeckendorf bit representation), of integers $p-1$, $q-1$, n satisfying $\{p,q\} \in \mathcal{A}(F_\infty[1..n])$ and p, q. Here, the -1 is due to the difference in the start index used for strings. The accepting paths for which $\mathsf{n} = 10^k$ implies $n = f_{k+1}$ and the values p and q implied from p and q should correspond to the statement of Theorem 23. See Figure 17 for a simplified automaton containing only such paths.

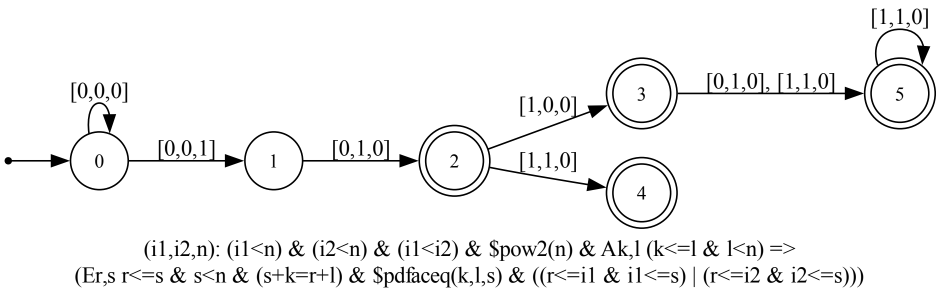

Figure 16 The automaton `pdsapow2` obtained by removing, from the automaton in Figure 14, all edges and nodes that are not part of accepting paths with $\mathtt{n} = 10^k$ implying $n = 2^k$. We can see that For $k = 2$, $\mathtt{p} = 001$ implying $p = 2$ and $\mathtt{q} = 010$ or 011 implying $q = 3$ or 4, i.e., $\mathcal{A}(D_k) = \{\{2,3\},\{2,4\}\}$. For $k \geq 3$, $\mathtt{p} = 00101^{k-3}$ or 00111^{k-3} implying $p = 3 \cdot 2^{k-3}$ or 2^{k-1}, and $\mathtt{q} = 01011^{k-3}$ implying $q = 3 \cdot 2^{k-2}$, i.e., $\mathcal{A}(D_k) = \{\{3 \cdot 2^{k-3}, 3 \cdot 2^{k-2}\}, \{2^{k-1}, 3 \cdot 2^{k-2}\}\}$, verifying the statement of Theorem 30.

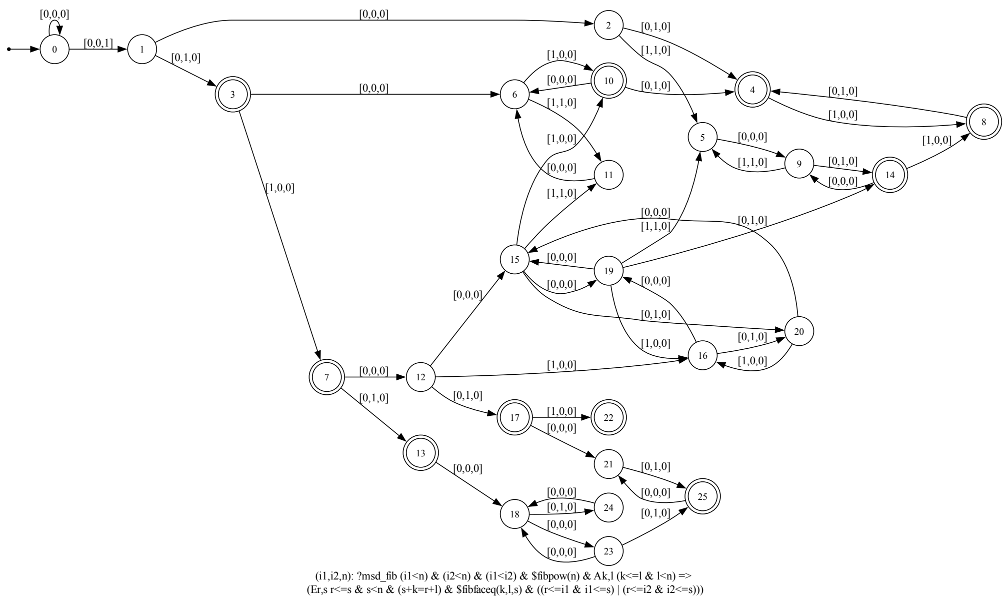

Figure 17 The automaton `fibsaf` obtained by removing, from the automaton in Figure 15, all edges and nodes that are not part of accepting paths with $\mathtt{n} = 10^k$ representing f_{k+1} in the Fibonacci numeration. Each accepting path of length n starting with $[0,0,1]$ represents a position pair in $\mathcal{A}(F_n)$, and the automaton can be used to verify Theorem 23 (in principle).

B Additional Figures

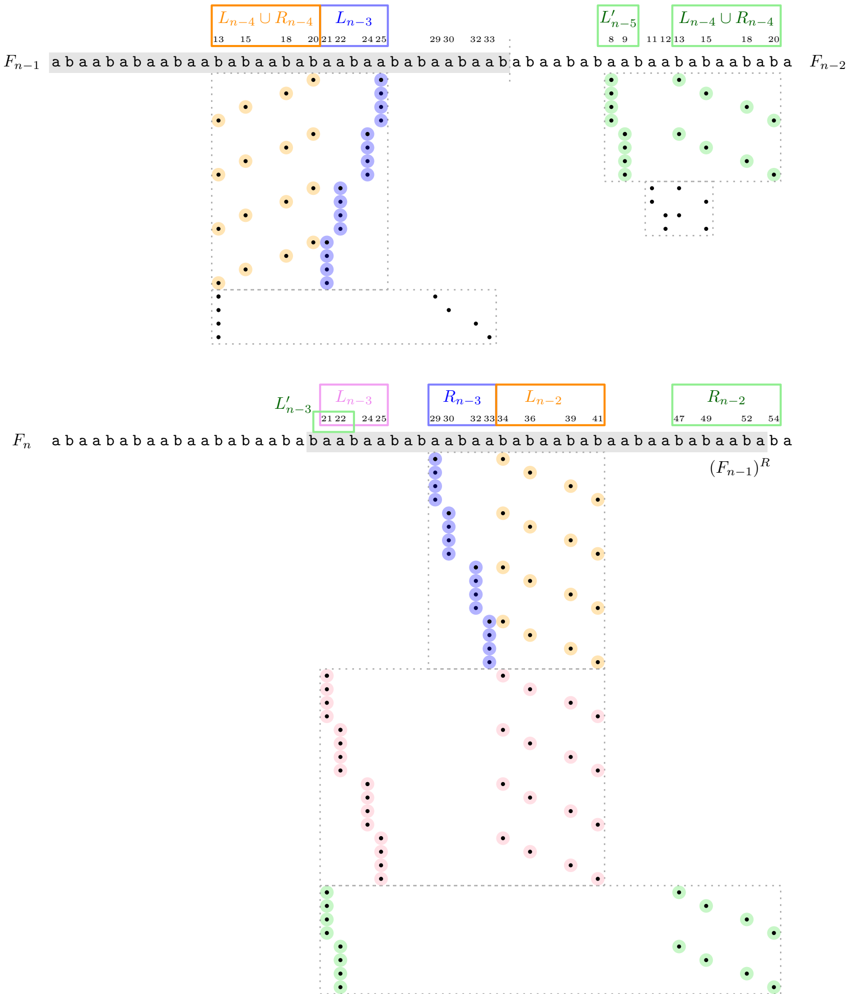

Figure 18 Extended version of Figure 1 for the case $n = 9$. For each of F_n (bottom), F_{n-1} (top left), and F_{n-2} (top right), all of their smallest attractors are indicated as pairs of horizontally aligned dots. Characterization of the attractors of F_n via Theorem 23 is illustrated by colors: the blue and yellow dots illustrate the mapping via the occurrences of F_{n-1} and $(F_{n-1})^R$ in Lemma 25; the pink dots illustrate the offset in Lemma 26; while the green dots illustrate Lemma 27.