String attractors for factors of the Thue-Morse word

Francesco Dolce

FIT, Czech Technical University in Prague, Czech Republic
dolcefra@fit.cvut.cz

Abstract. In 2020 Kutsukake et al. showed that every for every $n \geq 4$ the prefix of length 2^n of the Thue-Morse word has a string attractor of size 4. In this paper we extend their result by constructing a smallest string attractor for any given factor of the Thue-Morse word. In particular, we show that these string attractors have size at most 5 and that this upper bound is sharp.

Keywords: string attractors · Thue-Morse word · factorial languages

1 Introduction

String attractors were introduced by Kempa and Prezza in [6] in the context of dictionary-based data compression. A string attractor for a word w is a set of positions of the word such that all factors of w have an occurrence containing at least one of the elements of the set. Intuitively, the more repetitive is w the lower is the size of a smallest string attractor for w. Actually, the smallest size of a string attractor for a word is a lower bound for several other repetitiveness measures associated with the most common compression schemes, including the number of phrases in the LZ77 parsing and the number of equal-letter runs produced by the Burrows-Wheeler Transform (see [6,12,10]).

While it is trivial to construct a string attractor for a given word (e.g., by taking all possible positions), finding a smallest one is a NP-complete problem.

Mantaci et al. studied in [10] the size of a smallest string attractor of several infinite families of words. In particular they showed that every standard Sturmian word different than a letter has a smallest string attractor of size 2 (see also [5] for a generalization of this results to episturmian words), while the de Brujin word of length n has a smallest string attractor of size $\frac{n}{\log n}$. In the same paper they also studied the well-known Thue-Morse word $\mathbf{t}$, also known as Prouhet-Thue-Morse word, since first studied by Prouhet before being rediscovered by Thue and Morse, between others (see [13,16,11]). In a preliminary version of their paper ([9]) Mantaci et al. conjectured that prefixes of size 2^n of $\mathbf{t}$ have a smallest string attractor of size n. This conjecture has been proven to be wrong by Katsukake et al. in [7], who showed that for any such prefix it is possible to find a string attractor of size at most 4.

Schaeffer and Shallit introduced in [15] the notion of string attractor profile function for infinite words by evaluating the size of a smallest attractor for each prefix (see also [14]). If instead of prefixes we consider a generic factor of a (finite or infinite) word the situation get more complicated. Indeed, the measure of a smallest string attractor is not monotone, meaning that a factor w of a word u can have a smallest string attractor bigger than a string attractor of u (see Proposition 2).

In this article we prove and explicitly construct a smallest string attractor for any given factor of the Thue-Morse word. In particular, our main result is the following.

Theorem 1. *Let w be a non-empty finite factor of $\mathtt{t}$. Then there exists a string attractor for w of size at most 5.*

2 Preliminaries

For all undefined notation we refer to [8]. Let $\mathcal{A}$ be an *alphabet*, that is is a finite set of symbols called *letters*. A (finite) *word* over $\mathcal{A}$ of *length* n is a concatenation $u = u_1 \cdots u_n$, where $u_i \in \mathcal{A}$ for all $i \in \{1, \ldots, n\}$. The length of u is denoted by $|u|$. The set of all finite words over $\mathcal{A}$ together with the operation of concatenation form a monoid, denoted by $\mathcal{A}^*$, whose neutral element is the *empty word* ε. We also denote $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$. Similarly, given a set of words $S \subset \mathcal{A}^*$, we denote by S^* (resp., S^+) the set of all possible concatenations (resp., non-empty concatenations) of elements of S. When $\mathcal{A} = \{\mathtt{a}, \mathtt{b}\}$ is a binary alphabet we denote by $\overline{w}$ the word obtained from w by changing every $\mathtt{a}$ in $\mathtt{b}$ and vice versa. Formally $\overline{w}$ is obtained from w by applying the involution $\bar{\cdot} : a \mapsto b;\, b \mapsto a$.

Let $u = pfs$ for some $p, f, s \in \mathcal{A}^*$. We call p a *prefix* of w, s a *suffix* of w and f a *factor* of w. The prefix p (resp. suffix s) is called *proper* if it is different than u. If both p and s are non-empty we call f an *internal factor* of u. The set $\mathrm{Pref}(u)$ (resp., $\mathrm{Suf}(u)$) is the set of all non-empty prefixes (resp., suffixes) of u. The *language* of u, denoted by $\mathcal{L}(u)$, is the set of all finite factors of u.

An *infinite word* over $\mathcal{A}$ is a sequence $\mathbf{u} = u_1 u_2 \cdots$, where $u_i \in \mathcal{A}$ for every positive integer i. The notions above (prefix, suffix, etc.) naturally extend to infinite words.

Example 1. The Thue-Morse word is the infinite binary word

$$\mathtt{t} = \lim_{n \to \infty} t_n = \mathtt{abbabaabbaababbabaababbaabbabaabbaababbaabbabaabab} \cdots,$$

where $t_0 = \mathtt{a}$ and $t_{n+1} = t_n \, \overline{t_n}$ for any $n > 0$. Note that for any $n \in \mathbb{N}$ we have $|t_n| = |\overline{t_n}| = 2^n$.

Given a set $M \subset \mathbb{Z}$ and an integer $q \in \mathbb{Z}$, we denote $M + q = \{m + q \mid m \in M\}$.

3 String attractors

Let $w, u \in \mathcal{A}^+$, with $w \in \mathcal{L}(u)$, we say that w has an *occurrence starting at position* i in u, if it is possible to write $w = u_i u_{i+1} \cdots u_{i+|w|-1}$, with the convention that the empty word has an occurrence at every position. Clearly a word

w could have multiple occurrences in u. Given a position j with $1 \leq j \leq |u|$, we also say that an occurrence of w in u *contains* the position j if such occurrence starts at position i with $i \leq j < i + |w|$.

Example 2. Let us consider the words t_n as in Example 1. The word $w = \texttt{bba}$ has three occurrences in $t_4 = \texttt{abbabaabbaababba}$ starting respectively at positions 2, 8 and 14. The second occurrence is the only one containing the position 10.

Given a word $u \in \mathcal{A}^+$ a set Γ of positions is a *string attractor* for u if for every factor $w \in \mathcal{L}(u)$ there exists a $\gamma \in \Gamma$ such that at least one occurrence of w is of the form $w = u_i u_{i+1} \cdots u_{i+|w|-1}$ with $i \leq \gamma < i + |w|$.

The set $\{1, 2, \ldots, |u|\}$ is trivially a string attractor for a word u. On the other hand, a trivial lower bound for the size of a string attractor is given by the number of different letters appearing in u. Moreover, if Γ is a string attractor for u, so is Γ' for every superset $\Gamma' \supset \Gamma$. Note that a word can have different string attractors of the same size and, more generally, different string attractors that are not included into each other.

Example 3. Let t_n and $\overline{t_n}$ be defined as in Example 1. The set $\Gamma_0 = \{1\}$ is a string attractor for both words $t_0 = \underline{\texttt{a}}$ and $\overline{t_0} = \underline{\texttt{b}}$ (the positions of the string attractor are underlined). Similarly, the set $\Gamma_1 = \{1, 2\}$ is a string attractor for $t_1 = \underline{\texttt{a}}\underline{\texttt{b}}$ and for $\overline{t_1} = \underline{\texttt{b}}\underline{\texttt{a}}$. Such string attractor is the smallest one, since both letters $\texttt{a}$ and $\texttt{b}$ must be covered.

The set $\Gamma_2 = \{1, 2, 4\}$ is a string attractor for the word $t_2 = \underline{\texttt{a}}\underline{\texttt{b}}\texttt{b}\underline{\texttt{a}}$ (resp., for $\overline{t_2}$). Notice that $\Gamma_2' = \{2, 4\}$ is also a string attractor for $t_2 = \texttt{a}\underline{\texttt{b}}\texttt{b}\underline{\texttt{a}}$ (resp., for $\overline{t_2}$). Since both letters appear in t_2, the minimal size for a string attractor is 2. It is easy to check that $\{2, 5, 7\}$ is a string attractor for the word $t_3 = \texttt{a}\underline{\texttt{b}}\texttt{ba}\underline{\texttt{b}}\texttt{a}\underline{\texttt{a}}\texttt{b}$ (resp., for $\overline{t_3}$), while the same word does not have any string attractor of size 2. A larger string attractor for t_3 is given by $\Gamma_3 = \{2, 3, 4, 6\}$.

It is possible to check that the sets $\Gamma_4 = \{4, 6, 8, 12\}$, $\Gamma_5 = \{8, 12, 16, 24\}$, $\Gamma_6 = \{16, 24, 32, 48\}$ and $\Gamma_7 = \{32, 48, 64, 96\}$ are smallest string attractors respectively for the words $t_4 = \texttt{abb}\underline{\texttt{a}}\texttt{b}\underline{\texttt{a}}\texttt{a}\underline{\texttt{b}}\texttt{baa}\underline{\texttt{b}}\texttt{abba}$ (resp., for $\overline{t_4}$), t_5 (resp., for $\overline{t_5}$), t_6 (resp. $\overline{t_6}$) and t_7 (resp., $\overline{t_7}$).

The following two interesting combinatorial results are proved in [10, Propositions 12 and 14].

Proposition 1 ([10]). *Let $u, v \in \mathcal{A}^+$, Γ_u a string attractor for u and Γ_v a string attractor for v. Then $\Gamma_u \cup \{|u|\} \cup (\Gamma_v + |u|)$ is a string attractor for uv.*

Example 4. Let t_2, $\overline{t_2}$ and Γ_2' as in Example 3. A string attractor for $t_3 = t_2\,\overline{t_2}$ is given by $\Gamma_2' \cup \{4\} \cup (\Gamma_2' + 4) = \{2, 4, 6, 8\}$. Note that such a string attractor is not a smallest one.

Proposition 2 ([10]). *The size of a smallest string attractor for a word is not a monotone measure.*

The previous proposition says that if w is a factor of u, then it could be possible for u to have a string attractor of size smaller than the size of a smallest string attractor for w.

Example 5. Let t_n be as in Example 1. As seen in Example 3, the word t_7 has a smallest string attractor of size 4. However, it is possible to check that the word $w = \texttt{abaababb}\underline{\texttt{a}}\texttt{abbabaa}\underline{\texttt{b}}\texttt{baababbaabbabaa}\underline{\texttt{b}}\texttt{abbabaabbaababb}\underline{\texttt{a}}\texttt{b} = \texttt{a}\,\overline{t_4}\,\overline{t_5}\,\texttt{b} \in \mathcal{L}(t_7)$ has no string attractor of size 4. Note that $\Gamma = \{9, 13, 25, 33, 41\}$ is a string attractor of size 5 of w.

4 Proof of the main result

In [2] Brlek shows several combinatorial results concerning the factors in $\mathcal{L}(\mathbf{t})$. In particular he provides an explicit formula of the factor complexity of $\mathtt{t}$. Part of it is stated in the following.

Proposition 3 ([2]). *Let $n \in \mathbb{N}$ and $w \in \mathcal{L}(t_n)$ with $|w| \geq 2^{n-2} + 1$. The word t_n has exactly one occurrence of w.*

An important ingredient of our proof is [7, Theorem 2].

Theorem 2 ([7]). *Let $n \geq 4$. The set*

$$\Gamma_n = \{2^{n-2},\ 3 \cdot 2^{n-3},\ 2^{n-1},\ 3 \cdot 2^{n-2}\}$$

is a string attractor both for t_n and $\overline{t_n}$.

Note that in their papers Kutsukake et al. only state the result for t_n, but the same argument actually works also for $\overline{t_n}$.

We are now ready to prove Theorem 1.

Proof of Theorem 1. It can easily checked that every factor of t_n (resp., of $\overline{t_n}$) with $n \leq 5$ has a string attractor of size at most 4. Let us suppose the property true for all factors in $\mathcal{L}(t_n) \cup \mathcal{L}(\overline{t_n})$ and let us consider the case of $w \in \mathcal{L}(t_{n+1}) = \mathcal{L}(t_n\, \overline{t_n})$, with $n \geq 6$ (the case $w \in \mathcal{L}(\overline{t_{n+1}})$ being symmetrical).

If $w \in \mathcal{L}(t_n) \cup \mathcal{L}(\overline{t_n})$, then the result follows by induction. Thus, we can suppose that w has an occurrence in t_{n+1} containing the center of $t_n\, \overline{t_n}$, i.e., the last letter of the prefix t_n. In the following remaining part of the proof we consider all possible such factors of t_{n+1} by increasing their size. The idea is to write each factor as $w = \lambda\, t\, \rho$, with the central factor t of the form t_k or $\overline{t_k}$ for a certain $k \in \mathbb{N}$, and $\lambda = st'$ and $\rho = t''p$, where $t', t'' \in \{t_i, \overline{t_i} \mid i \in \mathbb{N}\}^*$ (more precisely we have $t' = t'_{h_\ell} \cdots t'_{h_1}$ and $t'' = t''_{j_1} \cdots t''_{j_r}$ with $h_\ell < \ldots < h_1$ and $j_1 > \ldots > j_r$), and s (resp., p) is a suffix (resp., a prefix) of a some element in $\{t_i, \overline{t_i} \mid i \in \mathbb{N}\}$.

Since the center of $t_n\, \overline{t_n}$ is also the center of $t_{n-2}\, \overline{t_{n-2}}$, if $w \in \mathcal{L}(t_{n-2}\overline{t_{n-2}}) = \mathcal{L}(t_{n-1})$, we can conclude by induction. Let us thus suppose that $w \notin \mathcal{L}(t_{n-1})$. We can write w either as $w = \lambda\, t_{n-2}\, \rho$, with $\lambda \in \mathrm{Suf}(t_{n-1}\, \overline{t_{n-2}})$ and $\rho \in \mathrm{Pref}(\overline{t_n})$, or as $w = \lambda\, \overline{t_{n-2}}\, \rho$ with $\lambda \in \mathrm{Suf}(t_n)$ and $\rho \in \mathrm{Pref}(t_{n-2}\, t_{n-1})$. Let us focus on the former case. Since $|\lambda t_{n-2}| > 2^{n-2}$ then $w \notin \mathcal{L}(t_n)$ according to Proposition 3.

In the following we extend, step by step, ρ to the right, and, for every fixed ρ, we extend λ to the left. While the first step is fully developed, we let the reader check the details of Steps 2 to 7.

Step 1. Let us start considering $\rho = p$ with $p \in \mathrm{Pref}(\overline{t_n})$.

i) Let $w = s\, t_{n-2}\, p$, with $s \in \mathrm{Suf}(\overline{t_{n-4}})$ and $p \in \mathrm{Pref}(\overline{t_{n-4}})$. Then

$$\Gamma = \Gamma_{n-2} + |\lambda|,$$

where $\lambda = s$, is a string attractor for w (see Figure 1, where we represent only the central factor $\overline{t_{n-1}}\,\overline{t_{n-1}}$ of t_{n+1}). Indeed, let $v \in \mathcal{L}(w)$. If $v \in \mathcal{L}(t_{n-2})$, then, by Theorem 2, one of its occurrences contains at least one of the positions of the string attractor Γ_{n-2} shifted by $|\lambda| = |s|$ (see Figure 1).
If v has an occurrence appearing to the left of the left-most position in Γ, then $v \in \mathcal{L}(\overline{t_{n-4}}\, t_{n-4}) \subset \mathcal{L}(t_{n-2})$, and thus it also has another occurrence containing at least one of the positions of $\Gamma_{n-2}+|\lambda|$ (see Figure 1). Similarly, if v has an occurrence appearing to the right of the right-most position in Γ, then $v \in \mathcal{L}(t_{n-4}\, \overline{t_{n-4}}) \subset \mathcal{L}(t_{n-2})$ and thus v has another occurrence containing at least one of the positions of $\Gamma_{n-2} + |\lambda|$.

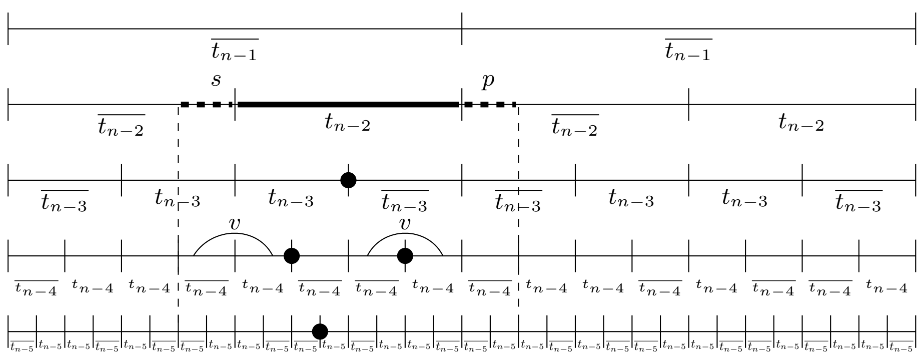

Fig. 1. A smallest string attractor Γ for a factor $w = s\, t_{n-2}\, p$ of t_{n+1}, with $s \in \mathrm{Suf}(\overline{t_{n-4}})$ and $p \in \mathrm{Pref}(\overline{t_{n-4}})$.

ii) Let $w = s\, \overline{t_{n-4}}\, t_{n-2}\, p$, with $s \in \mathrm{Suf}(t_{n-4})$ and $p \in \mathrm{Pref}(\overline{t_{n-4}})$. Then

$$\Gamma = \big(\big(\Gamma_{n-2} \setminus \{3 \cdot 2^{n-4}\}\big) \cup \{0\}\big) + |\lambda|,$$

where $\lambda = s\, \overline{t_{n-4}}$, is a string attractor for w (see Figure 2, where we represent only the central factor $\overline{t_{n-1}}\,\overline{t_{n-1}}$ of t_{n+1}). Indeed, let $v \in \mathcal{L}(w)$. If $v \in \mathcal{L}(t_{n-2})$, then, using Theorem 2, we have that v has an occurrence containing at least one of the positions of Γ_{n-2} shifted by $|\lambda|$; if the only position

contained by such occurrence is $3 \cdot 2^{n-4} + |\lambda|$, then $v \in \mathcal{L}(\overline{t_{n-4}}\, t_{n-4})$, hence v has another occurrence containing the position $|\lambda|$.
If v has an occurrence appearing to the left of the left-most position in Γ, then $v \in \mathcal{L}(t_{n-4}\, \overline{t_{n-4}}) \subset \mathcal{L}(t_{n-2})$ and we can conclude using again Theorem 2. If v has an occurrence appearing to the right of the right-most position of Γ, i.e., $2^{n-3} + |\lambda|$, then $v \in \mathcal{L}(\overline{t_{n-4}}\, t_{n-4}\, \overline{t_{n-4}})$: either v is fully contained in $\mathcal{L}(t_{n-4}\, \overline{t_{n-4}}) \subset \mathcal{L}(t_{n-2})$ and we can conclude, or v has another occurrence containing the position $|\lambda|$ (see Figure 2).

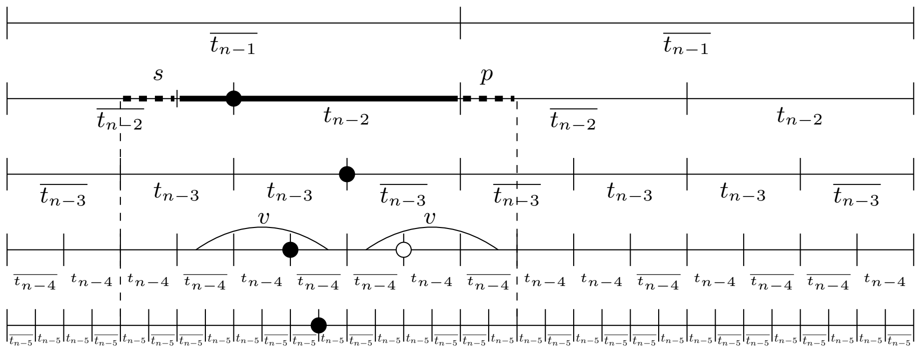

Fig. 2. A smallest string attractor Γ for a factor $w = s\, \overline{t_{n-4}}\, t_{n-2}\, p$ of t_{n+1}, with $s \in \mathrm{Suf}(t_{n-4})$ and $p \in \mathrm{Pref}(\overline{t_{n-4}})$. The position of $\Gamma_{n-2} + |\lambda|$ that is not in Γ is in white.

iii) Let $w = s\, t_{n-3}\, t_{n-2}\, p$, with $s \in \mathrm{Suf}(\overline{t_{n-5}})$ and $p \in \mathrm{Pref}(\overline{t_{n-4}})$. Then

$$\Gamma = \left(\left(\Gamma_{n-2} \setminus \{2^{n-4},\, 3 \cdot 2^{n-4}\}\right) \cup \{-2^{n-4},\, 0\}\right) + |\lambda|,$$

where $\lambda = s\, t_{n-3}$, is a string attractor for w. Indeed, let $v \in \mathcal{L}(w)$. Similarly to the previous case, if $v \in \mathcal{L}(t_{n-2})$, then, by Theorem 2, v has an occurrence containing at least one of the positions of Γ_{n-2} shifted by $|\lambda|$; if the only position contained in the occurrence of v is $2^{n-4}+|\lambda|$, then $v \in \mathcal{L}(\overline{t_{n-4}}\, t_{n-4})$ and thus there is another occurrence of v containing the position $-2^{n-4}+|\lambda|$; if the only position contained in the occurrence of v is $3 \cdot 2^{n-4} + |\lambda|$, then $v \in \mathcal{L}(\overline{t_{n-4}}\, t_{n-4})$ and thus there is another occurrence of v containing the position $|\lambda|$.
If v appears to the left of the left-most position in Γ, then we can conclude since $v \in \mathcal{L}(\overline{t_{n-5}}\, \overline{t_{n-5}}\, t_{n-5}) \subset \mathcal{L}(t_{n-2})$. If v appears between the positions $-2^{n-4} + |\lambda|$ and $|\lambda|$, then $v \in \mathcal{L}(\overline{t_{n-4}}) \subset \mathcal{L}(t_{n-2})$. If v appears to the right of the right-most position in Γ, i.e., $2^{n-3} + |\lambda|$, then $v \in \mathcal{L}(\overline{t_{n-4}}\, t_{n-4}\, \overline{t_{n-4}})$: either v is fully contained in $\mathcal{L}(t_{n-4}\, \overline{t_{n-4}}) \subset \mathcal{L}(t_{n-2})$, or v has another occurrence containing the position $|\lambda|$.

iv) Let $w = s\, \overline{t_{n-5}}\, t_{n-3}\, t_{n-2}\, p$, with $s \in \mathrm{Suf}(t_{n-3}\, \overline{t_{n-4}}\, t_{n-5})$ and $p \in \mathrm{Pref}(\overline{t_{n-4}})$. Then

$$\Gamma = \left(\left(\Gamma_{n-2} \setminus \{3 \cdot 2^{n-5},\, 3 \cdot 2^{n-4}\}\right) \cup \{-2^{n-3}, 0\}\right) + |\lambda|,$$

where $\lambda = s\,\overline{t_{n-5}}\,t_{n-3}$, is a string attractor for w. Indeed, let $v \in \mathcal{L}(w)$. As above, if $v \in \mathcal{L}(t_{n-2})$, then, by Theorem 2, v has an occurrence containing at least one of the positions of Γ_{n-2} shifted by $|\lambda|$; if the only position contained in the occurrence is $3 \cdot 2^{n-5} + |\lambda|$ (resp., $3 \cdot 2^{n-4} + |\lambda|$) then v has another occurrence containing the position $-2^{n-3} + |\lambda|$ (resp., $|\lambda|$).
If v appears to the left of the left-most position of Γ, then $v \in \mathcal{L}(t_{n-3}\,\overline{t_{n-3}}) = \mathcal{L}(t_{n-2})$ and we can conclude. If v appears between the positions $-2^{n-3}+|\lambda|$ and $|\lambda|$, then $v \in \mathcal{L}(t_{n-3}) \subset \mathcal{L}(t_{n-2})$ and we can conclude. If v appears to the right of the right-most position of Γ (i.e., $2^{n-3}+|\lambda|$) then it is contained in$\mathcal{L}(\overline{t_{n-4}}\,t_{n-4}\,\overline{t_{n-4}})$: either v is fully contained in $\mathcal{L}(t_{n-4}\,\overline{t_{n-4}}) \subset \mathcal{L}(t_{n-2})$ and we can conclude, or v has another occurrence containing the position $|\lambda|$.

v) Let $w = s\,t_{n-3}\,\overline{t_{n-1}}\,p$, with $s \in \mathrm{Suf}(\overline{t_{n-3}})$ and $p \in \mathrm{Pref}(\overline{t_{n-4}})$. Then

$$\Gamma = \left(\left(\Gamma_{n-1} \setminus \{3 \cdot 2^{n-4}\}\right) \cup \{-2^{n-4}\}\right) + |\lambda|.$$

where $\lambda = s\,t_{n-3}$, is a string attractor for w. Indeed, let $v \in \mathcal{L}(w)$. If $v \in \mathcal{L}(\overline{t_{n-1}})$ then, by Theorem 2, v has an occurrence containing at least one of the positions of Γ_{n-1} shifted by $|\lambda|$; if the only position contained in the occurrence is $3 \cdot 2^{n-4} + |\lambda|$, then $v \in \mathcal{L}(t_{n-4}\,\overline{t_{n-4}})$ and thus there is another occurrence of v containing the position $-2^{n-4} + |\lambda|$.
If v appears to the left of the left-most position of Γ (resp., between $-2^{n-4}+|\lambda|$ and $2^{n-3} + |\lambda|$; resp., to the right of the right-most position of Γ), then it is contained in $\mathcal{L}(\overline{t_{n-4}}\,t_{n-4}\,t_{n-4})$ (resp., $v \in \mathcal{L}(\overline{t_{n-4}}\,\overline{t_{n-4}}\,t_{n-4})$; resp., $v \in \mathcal{L}(\overline{t_{n-4}}\,t_{n-4}\,\overline{t_{n-4}})$) thus it is also contained in $\mathcal{L}(\overline{t_{n-1}})$ and we can conclude.

vi) Let $w = s\,\overline{t_{n-2}}\,\overline{t_{n-1}}\,p$, with $s \in \mathrm{Suf}(t_{n-4})$ and $p \in \mathrm{Pref}(\overline{t_{n-4}})$. This is the first case when it is not enough to "move" some of the positions of a string attractor of the form Γ_k, with $k \in \mathbb{N}$. Indeed, as shown in Example 5, in this case it is not possible to have a string attractor of size 4. On the other hand the set

$$\Gamma = \left(\left(\Gamma_{n-1} \setminus \{3 \cdot 2^{n-4}\}\right) \cup \{-2^{n-3},\, -2^{n-4}\}\right) + |\lambda|,$$

where $\lambda = s\,\overline{t_{n-2}}$, is a string attractor for w (see Figure 3). Indeed, let $v \in \mathcal{L}(w)$. If $v \in \mathcal{L}(\overline{t_{n-1}})$ then, by Theorem 2, v has an occurrence containing at least one of the positions of Γ_{n-1}; if the only position contained in the occurrence is $3 \cdot 2^{n-4} + |\lambda|$, then $v \in \mathcal{L}(t_{n-4}\,\overline{t_{n-4}})$ and thus there exists another occurrence of v containing $-2^{n-4} + |\lambda|$.
All the other cases are proved as in the previous cases: namely if v appears to the left of $-2^{n-3}+|\lambda|$, (resp., between $-2^{n-3}+|\lambda|$ and $-2^{n-4}+|\lambda|$; resp., between $-2^{n-4} + |\lambda|$ and $2^{n-3} + |\lambda|$; resp., to the right of $3 \cdot 2^{n-3} + |\lambda|$) then $v \in \mathcal{L}(t_{n-4}\,\overline{t_{n-4}}\,t_{n-4})$ (resp., $v \in \mathcal{L}(t_{n-4})$; resp., $v \in \mathcal{L}(\overline{t_{n-4}}\,\overline{t_{n-4}}\,t_{n-4})$; resp., $v \in \mathcal{L}(\overline{t_{n-4}}\,t_{n-4}\,\overline{t_{n-4}})$), thus it appears in $\mathcal{L}(\overline{t_{n-1}})$ and we can conclude using Theorem 2.

vii) Let $w = s\,t_{n-4}\,\overline{t_{n-2}}\,\overline{t_{n-1}}\,p$, with $s \in \mathrm{Suf}(t_{n-3}\,\overline{t_{n-4}})$ and $p \in \mathrm{Pref}(\overline{t_{n-4}})$. As in the previous case, it is possible to check that there exist no string

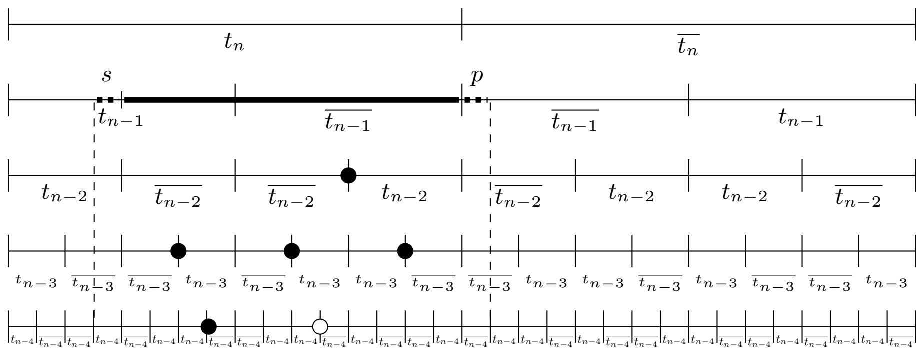

Fig. 3. A smallest string attractor Γ for a factor $w = s\,\overline{t_{n-2}}\,\overline{t_{n-1}}\,p$ of t_{n+1}, with $s \in \mathrm{Suf}(t_{n-4})$, and $p \in \mathrm{Pref}(\overline{t_{n-4}})$. The position of $\Gamma_{n-1} + |\lambda|$ that is not in Γ is in white.

attractor of size 4. However, the set

$$\Gamma = \big(\big(\Gamma_{n-1} \setminus \{3 \cdot 2^{n-4}\}\big) \cup \{-2^{n-2},\, -2^{n-4}\}\big) + |\lambda|,$$

where $\lambda = s\,t_{n-4}\,\overline{t_{n-2}}$, is a string attractor for w. Indeed, let $v \in \mathcal{L}(w)$. If $v \in \mathcal{L}(\overline{t_{n-1}})$, then, by Theorem 2, v has an occurrence containing at least one of the positions of Γ_{n-1}; if the only position contained in the occurrence is $3 \cdot 2^{n-4} + |\lambda|$, then $v \in \mathcal{L}(t_{n-4}\,\overline{t_{n-4}})$ and thus there exists another occurrence of v containing $-2^{n-4} + |\lambda|$.
All the other cases are proved as in the previous cases: namely if v appears to the left of $-2^{n-2} + |\lambda|$, (resp., between $-2^{n-2} + |\lambda|$ and $-2^{n-4} + |\lambda|$; resp., between $-2^{n-4} + |\lambda|$ and $2^{n-3} + |\lambda|$; resp., to the right of $3 \cdot 2^{n-3} + |\lambda|$) then $v \in \mathcal{L}(t_{n-2})$ (resp., $v \in \mathcal{L}(\overline{t_{n-4}}\,t_{n-4}\,t_{n-4})$; resp., $v \in \mathcal{L}(\overline{t_{n-4}}\,\overline{t_{n-4}}\,t_{n-4})$; resp., $v \in \mathcal{L}(\overline{t_{n-4}}\,t_{n-4}\,\overline{t_{n-4}})$), thus it appears in $\mathcal{L}(\overline{t_{n-1}})$ and we can conclude.

The seven cases above are summarized in Table 1. Note that in all previous cases $\overline{t_{n-4}}$ is a prefix of p, and hence a prefix of ρ.

Step 2. We now consider the case of ρ containing $\overline{t_{n-4}}$ as a proper prefix. The seven possible cases of factors are summarized in Table 2. Note that in all these cases t_{n-4} is a prefix of p, and hence $\overline{t_{n-3}} = \overline{t_{n-4}}\,t_{n-4}$ is a prefix of ρ.

Step 3. We now consider the case of ρ containing as a proper prefix $\overline{t_{n-3}}$. The six possible cases of factors are summarized in Table 3. Note that in all these cases t_{n-5} is a prefix of p, and hence $\overline{t_{n-3}}\,t_{n-5}$ is a prefix of ρ.

Step 4. We now consider the case of ρ containing as a proper prefix $\overline{t_{n-3}}\,t_{n-5}$. The six possible cases of factors are summarized in Table 4. Note that in all these cases $\overline{t_{n-5}}\,\overline{t_{n-4}}\,t_{n-4}$ is a prefix of p, and hence $\overline{t_{n-2}}\,t_{n-4}$ is a prefix of ρ.

Step 5. We now consider the case of ρ containing as a proper prefix $\overline{t_{n-2}}\,t_{n-4}$. Since $t_{n-2}\overline{t_{n-2}} = t_{n-1}$ is a factor of w, in the first two cases weconsider as starting point for constructing a string attractor Γ_{n-1} instead of Γ_{n-2} (for the remaining two cases we have as central factor $t = \overline{t_{n-1}}$, so we also use Γ_{n-1}).

$\mathrm{Suf}(\cdot)$	t'	t	t''	$\mathrm{Pref}(\cdot)$	Γ'	Γ''	$\lvert\Gamma\rvert$
$\overline{t_{n-4}}$	ε	t_{n-2}	ε	$\overline{t_{n-4}}$	$\emptyset$	$\emptyset$	4
t_{n-4}	$\overline{t_{n-4}}$	t_{n-2}	ε	$\overline{t_{n-4}}$	$\{3\cdot 2^{n-4}\}$	$\{0\}$	4
$\overline{t_{n-5}}$	t_{n-3}	t_{n-2}	ε	$\overline{t_{n-4}}$	$\{2^{n-4}, 3\cdot 2^{n-4}\}$	$\{-2^{n-4}, 0\}$	4
$t_{n-3}\,\overline{t_{n-4}}\,t_{n-5}$	$\overline{t_{n-5}}\,t_{n-3}$	t_{n-2}	ε	$\overline{t_{n-4}}$	$\{3\cdot 2^{n-5}, 3\cdot 2^{n-4}\}$	$\{-2^{n-3}, 0\}$	4
$\overline{t_{n-3}}$	t_{n-3}	$\overline{t_{n-1}}$	ε	$\overline{t_{n-4}}$	$\{3\cdot 2^{n-4}\}$	$\{-2^{n-4}\}$	4
t_{n-4}	$\overline{t_{n-2}}$	$\overline{t_{n-1}}$	ε	$\overline{t_{n-4}}$	$\{3\cdot 2^{n-4}\}$	$\{-2^{n-3}, -2^{n-4}\}$	5
$t_{n-3}\,\overline{t_{n-4}}$	$t_{n-4}\,\overline{t_{n-2}}$	$\overline{t_{n-1}}$	ε	$\overline{t_{n-4}}$	$\{3\cdot 2^{n-4}\}$	$\{-2^{n-4}, -2^{n-2}\}$	5

Table 1. Summary of **Step 1** of the proof of Theorem 1. For a factor of the form $w = s\,t'\,t\,t\,p$, with $s \in \mathrm{Suf}(\cdot)$, $p \in \mathrm{Pref}(\cdot)$, a smallest string attractor is $\Gamma = ((\Gamma_k \setminus \Gamma') \cup \Gamma'') + |s\,t'|$, with k the integer such that $t = t_k$ or $\overline{t_k}$.

$\mathrm{Suf}(\cdot)$	t'	t	t''	$\mathrm{Pref}(\cdot)$	Γ'	Γ''	$\lvert\Gamma\rvert$
$\overline{t_{n-4}}$	ε	t_{n-2}	$\overline{t_{n-4}}$	t_{n-4}	$\{3\cdot 2^{n-5}\}$	$\{9\cdot 2^{n-5}\}$	4
t_{n-4}	$\overline{t_{n-4}}$	t_{n-2}	$\overline{t_{n-4}}$	t_{n-4}	$\{3\cdot 2^{n-5}, 3\cdot 2^{n-4}\}$	$\{0, 9\cdot 2^{n-5}\}$	4
$\overline{t_{n-5}}$	t_{n-3}	t_{n-2}	$\overline{t_{n-4}}$	t_{n-4}	$\{2^{n-4}\}$	$\{-2^{n-4}\}$	4
$t_{n-3}\,\overline{t_{n-4}}\,t_{n-5}$	$\overline{t_{n-5}}\,t_{n-3}$	t_{n-2}	$\overline{t_{n-4}}$	t_{n-4}	$\{2^{n-4}, 3\cdot 2^{n-5}\}$	$\{-2^{n-3}, -2^{n-4}\}$	4
$\overline{t_{n-3}}$	t_{n-3}	$\overline{t_{n-1}}$	$\overline{t_{n-4}}$	$t_{n-4}\,t_{n-3}\,t_{n-2}$	$\{3\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	$\{0, 2^{n-1}\}$	4
t_{n-4}	$\overline{t_{n-2}}$	$\overline{t_{n-1}}$	$\overline{t_{n-4}}$	$t_{n-4}\,t_{n-3}\,t_{n-2}$	$\{2^{n-3}, 3\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	$\{-2^{n-3}, 0, 2^{n-1}\}$	4
$t_{n-3}\,\overline{t_{n-4}}$	$t_{n-4}\,\overline{t_{n-2}}$	$\overline{t_{n-1}}$	$\overline{t_{n-4}}$	$t_{n-4}\,t_{n-3}\,t_{n-2}$	$\{3\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	$\{-2^{n-2}, 0, 2^{n-1}\}$	5

Table 2. Summary of **Step 2** of the proof of Theorem 1. For a factor of the form $w = s\,t'\,t\,t\,p$, with $s \in \mathrm{Suf}(\cdot)$, $p \in \mathrm{Pref}(\cdot)$, a smallest string attractor is $\Gamma = ((\Gamma_k \setminus \Gamma') \cup \Gamma'') + |s\,t'|$, with k the integer such that $t = t_k$ or $\overline{t_k}$.

The four possible cases of factors are summarized in Table 5. Note that in all these cases $\overline{t_{n-4}}\,\overline{t_{n-3}}$ is a prefix of p, and hence $\overline{t_{n-1}}$ is a prefix of ρ.

Step 6. We now consider the case of ρ containing as a proper prefix $\overline{t_{n-1}}$. As in the previous step, we have $t_{n-2}\,\overline{t_{n-2}} = t_{n-1}$ is a factor of w. For this reason, in the first of the three cases considered we construct the string attractor starting by a shift of Γ_{n-1} (for the remaining cases we also use Γ_{n-1} following the same construction of steps above). The three possible cases of factors are summarized in Table 6. Note that in all these cases t_{n-3} is a prefix of p.

Suf(·)	t'	t	t''	Pref(·)	Γ'	Γ''	$\lvert\Gamma\rvert$
$\overline{t_{n-4}}$	ε	t_{n-2}	$\overline{t_{n-3}}$	t_{n-5}	$\{2^{n-4}, 3\cdot 2^{n-4}\}$	$\{2^{n-2}, 5\cdot 2^{n-5}\}$	4
t_{n-4}	$\overline{t_{n-4}}$	t_{n-2}	$\overline{t_{n-3}}$	t_{n-5}	$\{2^{n-4}, 3\cdot 2^{n-5}, 3\cdot 2^{n-4}\}$	$\{-2^{n-5}, 2^{n-2}, 5\cdot 2^{n-4}\}$	4
$\overline{t_{n-5}}$	t_{n-3}	t_{n-2}	$\overline{t_{n-3}}$	t_{n-5}	$\{2^{n-4}, 3\cdot 2^{n-4}\}$	$\{-2^{n-4}, 2^{n-2}, 5\cdot 2^{n-4}\}$	5
$t_{n-3}\,\overline{t_{n-4}}\,t_{n-5}$	$\overline{t_{n-5}}\,t_{n-3}$	t_{n-2}	$\overline{t_{n-3}}$	t_{n-5}	$\{3\cdot 2^{n-5}, 3\cdot 2^{n-4}\}$	$\{-2^{n-3}, 5\cdot 2^{n-4}\}$	4
$\overline{t_{n-3}}$	t_{n-3}	$\overline{t_{n-1}}$	$\overline{t_{n-3}}$	$t_{n-3}\,t_{n-2}$	$\{3\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	$\{0, 2^{n-1}\}$	4
t_{n-2}	$\overline{t_{n-2}}$	$\overline{t_{n-1}}$	$\overline{t_{n-3}}$	$t_{n-3}\,t_{n-2}$	$\{2^{n-3}, 3\cdot 2^{n-4}\}$	$\{-2^{n-3}, 2^{n-1}\}$	4

Table 3. Summary of **Step 3** of the proof of Theorem 1. For a factor of the form $w = s\,t'\,t\,t\,p$, with $s \in \mathrm{Suf}(\cdot)$, $p \in \mathrm{Pref}(\cdot)$, a smallest string attractor is $\Gamma = ((\Gamma_k \setminus \Gamma') \cup \Gamma'') + |s\,t'|$, with k the integer such that $t = t_k$ or $\overline{t_k}$.

Suf(·)	t'	t	t''	Pref(·)	Γ'	Γ''	$\lvert\Gamma\rvert$
$\overline{t_{n-4}}$	ε	t_{n-2}	$\overline{t_{n-3}}\,t_{n-5}$	$\overline{t_{n-5}}\,\overline{t_{n-4}}\,t_{n-4}$	$\{2^{n-4}, 3\cdot 2^{n-5}\}$	$\{2^{n-2}, 3\cdot 2^{n-3}\}$	4
t_{n-4}	$\overline{t_{n-4}}$	t_{n-2}	$\overline{t_{n-3}}\,t_{n-5}$	$\overline{t_{n-5}}\,\overline{t_{n-4}}\,t_{n-4}$	$\{2^{n-4}, 3\cdot 2^{n-5}, 3\cdot 2^{n-4}\}$	$\{0, 2^{n-2}, 3\cdot 2^{n-4}\}$	4
$\overline{t_{n-5}}$	t_{n-3}	t_{n-2}	$\overline{t_{n-3}}\,t_{n-5}$	$\overline{t_{n-5}}\,\overline{t_{n-4}}\,t_{n-2}$	$\{2^{n-4}, 3\cdot 2^{n-5}\}$	$\{-2^{n-4}, 3\cdot 2^{n-3}\}$	4
$t_{n-3}\,\overline{t_{n-4}}\,t_{n-5}$	$\overline{t_{n-5}}\,t_{n-3}$	t_{n-2}	$\overline{t_{n-3}}t_{n-5}$	$\overline{t_{n-5}}\,\overline{t_{n-4}}\,t_{n-2}$	$\{3\cdot 2^{n-5}, 3\cdot 2^{n-4}\}$	$\{-2^{n-3}, 5\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	5
$\overline{t_{n-3}}$	t_{n-3}	$\overline{t_{n-1}}$	$\overline{t_{n-3}}\,t_{n-5}$	$\overline{t_{n-5}}\,\overline{t_{n-4}}\,t_{n-2}$	$\{3\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	$\{0, 2^{n-1}\}$	4
t_{n-2}	$\overline{t_{n-2}}$	$\overline{t_{n-1}}$	$\overline{t_{n-3}}t_{n-5}$	$\overline{t_{n-5}}\,\overline{t_{n-4}}\,t_{n-2}$	$\{2^{n-3}, 3\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	$\{-2^{n-3}, 0, 2^{n-1}\}$	4

Table 4. Summary of **Step 4** of the proof of Theorem 1. For a factor of the form $w = s\,t'\,t\,t\,p$, with $s \in \mathrm{Suf}(\cdot)$, $p \in \mathrm{Pref}(\cdot)$, a smallest string attractor is $\Gamma = ((\Gamma_k \setminus \Gamma') \cup \Gamma'') + |s\,t'|$, with k the integer such that $t = t_k$ or $\overline{t_k}$.

Step 7. As last step we consider the case of ρ containing as a proper prefix $\overline{t_{n-1}}\,t_{n-3}$. Similarly to the previous step, since $t_{n-2}\,\overline{t_{n-2}} = t_{n-1}$ is a factor of w, we construct the string attractor starting from a shift of Γ_{n-1} also in the first of the three cases. The three possible cases of factors are summarized in Table 7.

Thus we proved the result for all factors $w \in \mathcal{L}(t_{n+1})$ containing $x\,t_{n-2}\,y$, with x the last letter of $\overline{t_{n-2}}$ and y the first letter of $\overline{t_n}$, as a proper factor. The case $w = \lambda\,\overline{t_{n-2}}\,\rho$ with $\lambda \in \mathrm{Suf}(t_n)$ and $\rho \in \mathrm{Pref}(t_{n-2}\,t_{n-1})$ is proved in a symmetrical way.

Suf$(\cdot)$	t'	t	t''	Pref$(\cdot)$	Γ'	Γ''	$\lvert\Gamma\rvert$
$\overline{t_{n-4}}$	ε	t_{n-1}	t_{n-4}	$\overline{t_{n-4}}\,\overline{t_{n-3}}$	$\{3\cdot 2^{n-4}\}$	$\{2^{n-1}\}$	4
$\overline{t_{n-3}}\,t_{n-4}$	$\overline{t_{n-4}}$	t_{n-1}	t_{n-4}	$\overline{t_{n-4}}\,\overline{t_{n-3}}$	$\{3\cdot 2^{n-4}\}$	$\{0, 2^{n-1}\}$	5
t_{n-3}	ε	$\overline{t_{n-1}}$	$\overline{t_{n-2}}\,t_{n-4}$	$\overline{t_{n-4}}\,\overline{t_{n-3}}$	$\{3\cdot 2^{n-4}\}$	$\{2^{n-1}\}$	4
$t_{n-2}\,\overline{t_{n-3}}$	t_{n-3}	$\overline{t_{n-1}}$	$\overline{t_{n-2}}t_{n-4}$	$\overline{t_{n-4}}\,\overline{t_{n-3}}$	$\{3\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	$\{0, 2^{n-1}\}$	4

Table 5. Summary of **Step 5** of the proof of Theorem 1. For a factor of the form $w = s\,t'\,t\,t\,p$, with $s \in$ Suf$(\cdot)$, $p \in$ Pref$(\cdot)$, a smallest string attractor is $\Gamma = ((\Gamma_{n-1} \setminus \Gamma') \cup \Gamma'') + |s\,t'|$.

Suf$(\cdot)$	t'	t	t''	Pref$(\cdot)$	Γ'	Γ''	$\lvert\Gamma\rvert$
$\overline{t_{n-2}}$	ε	t_{n-1}	t_{n-2}	t_{n-3}	$\{2^{n-3}, 3\cdot 2^{n-4}\}$	$\{2^{n-1}, 5\cdot 2^{n-3}\}$	4
t_{n-3}	ε	$\overline{t_{n-1}}$	$\overline{t_{n-1}}$	t_{n-3}	$\{3\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	$\{2^{n-1}, 7\cdot 2^{n-3}\}$	4
$t_{n-2}\,\overline{t_{n-3}}$	t_{n-3}	$\overline{t_{n-1}}$	$\overline{t_{n-1}}$	t_{n-3}	$\{3\cdot 2^{n-4}, 2^{n-2}, 3\cdot 2^{n-3}\}$	$\{0, 2^{n-1}, 3\cdot 2^{n-2}\}$	4

Table 6. Summary of **Step 6** of the proof of Theorem 1. For a factor of the form $w = s\,t'\,t\,t\,p$, with $s \in$ Suf$(\cdot)$, $p \in$ Pref$(\cdot)$, a smallest string attractor is $\Gamma = ((\Gamma_{n-1} \setminus \Gamma') \cup \Gamma'') + |s\,t'|$.

Suf$(\cdot)$	t'	t	t''	Pref$(\cdot)$	Γ'	Γ''	$\lvert\Gamma\rvert$
$\overline{t_{n-2}}$	ε	t_{n-1}	$t_{n-2}\,t_{n-3}$	$\overline{t_{n-3}}\,\overline{t_{n-2}}$	$\{3\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	$\{2^{n-1}, 3\cdot 2^{n-2}\}$	4
t_{n-3}	ε	$\overline{t_{n-1}}$	$\overline{t_{n-1}}\,t_{n-3}$	$\overline{t_{n-3}}\,\overline{t_{n-2}}$	$\{2^{n-3}, 3\cdot 2^{n-4}\}$	$\{2^{n-1}, 2^{n}\}$	4
$t_{n-2}\,\overline{t_{n-3}}$	t_{n-3}	$\overline{t_{n-1}}$	$\overline{t_{n-1}}\,t_{n-3}$	$\overline{t_{n-3}}\,\overline{t_{n-2}}$	$\{2^{n-3}, 3\cdot 2^{n-4}, 3\cdot 2^{n-3}\}$	$\{0, 2^{n-1}, 2^{n}\}$	4

Table 7. Summary of **Step 7** of the proof of Theorem 1. For a factor of the form $w = s\,t'\,t\,t\,p$, with $s \in$ Suf$(\cdot)$, $p \in$ Pref$(\cdot)$, a smallest string attractor is $\Gamma = ((\Gamma_{n-1} \setminus \Gamma') \cup \Gamma'') + |s\,t'|$.

5 Future works and different approaches

The Thue-Morse word has been generalized to larger alphabets in several different ways. One possible generalization is the one given in [2], where $\mathbf{t}_m$ is defined over an alphabet $\mathcal{A}_m = \{a_1, a_2, \ldots, a_m\}$ of cardinality m as the fixed point $\mathbf{t}_m = \lim_{n\to\infty} \varphi_m^n(a_1)$, where $\varphi_m(a_k) = a_k \cdots a_m a_1 \cdots a_{k-1}$ for every $1 \le k \le m$.

For instance, we have $\mathbf{t}_3 = \texttt{abcbcacabbcacababccababcabc}\cdots$ over the ternary alphabet $\{\texttt{a}, \texttt{b}, \texttt{c}\}$.

Conjecture 1. For every $m \in \mathbb{N}$ there exist an integer K_m such that every non-empty factor of $\mathbf{t}_m$ has a string attractor of size at most K_m.

Recently Dvořáková proved that every factor of an episturmian word has a sting attractor having size the number of distinct letters appearing in the factor (see [5]). In particular, every factor of a Sturmian word different from a letter has a string attractor of size 2. Such result is based on the construction of (standard) episturmian words by iterated palindromic closure (see [4]). We believe that a similar approach could be used also for the Thue-Morse word, using pseudo-palindromic closure instead (see [3,1]).

Acknowledgements. This research received funding from the Ministry of Education, Youth and Sports of the Czech Republic through the project CZ.02.1.01/0.0/0.0/16_019/0000765. To check some of the examples we used a code written by undergraduate student Veronika Hendrychová.

References

1. Alexandre Blondin Massé, Geneviève Paquin, Hugo Tremblay, and Laurent Vuillon. On generalized pseudostandard words over binary alphabets. *J. Integer Seq.*, 16(2):Article 13.2.11, 28, 2013.
2. Srečko Brlek. Enumeration of factors in the Thue-Morse word. *Discrete Applied Mathematics*, 24(1-3):83–96, 1989.
3. Aldo de Luca and Alessandro De Luca. Pseudopalindrome closure operators in free monoids. *Theoret. Comput. Sci.*, 362(1-3):282–300, 2006.
4. Xavier Droubay, Jacques Justin, and Giuseppe Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1-2):539–553, 2001.
5. Ľubomíra Dvořáková. String attractors of episturmian sequences. https://arxiv.org/pdf/2211.01660v2.pdf, 2022.
6. Dominik Kempa and Nicola Prezza. At the roots of dictionary compression: string attractors. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 827–840. ACM, New York, 2018.
7. Kanaru Kutsukake, Takuya Matsumoto, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda. On repetitiveness measures of Thue-Morse words. *Lecture Notes in Computer Science*, 12303 LNCS:213–220, 2020.
8. M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.
9. Sabrina Mantaci, Antonio Restivo, Giuseppe Romana, Giovanna Rosone, and Marinella Sciortino. String attractors and combinatorics on words. In *ICTS, CEUR Workshop Proceedins*, volume 2504, pages 57–71, 2019.
10. Sabrina Mantaci, Antonio Restivo, Giuseppe Romana, Giovanna Rosone, and Marinella Sciortino. A combinatorial view on string attractors. *Theoret. Comput. Sci.*, 850:236–248, 2021.
11. Harold Marston Morse. Recurrent geodesics on a surface of negative curvature. *Trans. Amer. Math. Soc.*, 22(1):84–100, 1921.

12. Gonzalo Navarro. Indexing highly repetitive collections. In *Combinatorial algorithms*, volume 7643 of *Lecture Notes in Comput. Sci.*, pages 274–279. Springer, Heidelberg, 2012.
13. Eugène Prouhet. Mémoire sur quelques relations entre les puissances des nombres. *C.R. Acad. Sci.*, 31:225, 1851.
14. Antonio Restivo, Giuseppe Romana, and Marinella Sciortino. String attractors and infinite words. In *LATIN 2022: Theoretical Informatics*, pages 426–442. Springer International Publishing, 2022.
15. Luke Schaeffer and Jeffrey Shallit. String attractors for automatic sequences. https://arxiv.org/pdf/2012.06840.pdf, 2021.
16. Axel Thue. Über unendliche zeichenreihenal. *orske vid. Selsk. Skr. Mat. Nat. Kl.*, 7:1–22, 1906.