Journal of Integer Sequences, Vol. 22 (2019),
Article 19.7.8

Prefix Palindromic Length of the Thue-Morse Word

Anna E. Frid
Aix Marseille Univ., CNRS
Centrale Marseille, I2M
Marseille
France
anna.frid@univ-amu.fr

Abstract

The prefix palindromic length $\mathrm{PPL}_u(n)$ of an infinite word u is the minimal number of concatenated palindromes needed to express the prefix of length n of u. In a 2013 paper with Puzynina and Zamboni we stated the conjecture that $\mathrm{PPL}_u(n)$ is unbounded for every infinite word u that is not ultimately periodic. Up to now, the conjecture has been proven for almost all words, including all words avoiding some power p. However, even in that simple case the existing upper bound for the minimal number n such that $\mathrm{PPL}_u(n) > K$ is greater than any constant to the power K. Precise values of $\mathrm{PPL}_u(n)$ are not known even for simplest examples like the Fibonacci word.

In this paper, we give the first example of such a precise computation and compute the function of the prefix palindromic length of the Thue-Morse word, a famous test object for all functions on infinite words. It happens that this sequence is 2-regular, which raises the question if this fact can be generalized to all automatic sequences.

1 Introduction

By the usual definition, a palindrome is a finite word $p = p[1] \cdots p[n]$ over a finite alphabet such that $p[i] = p[n-i+1]$ for every i, like *level* or *abba*. We consider decompositions, or factorizations, of a finite word into a concatenation of palindromes: for example, $acbcadd = (a)(cbc)(a)(dd)$ and, at the same time, $acbcadd = (acbca)(dd)$. In particular, we are interested in the minimal number of palindromes needed for such a decomposition, which we call a *palindromic length* of a word. For example, the palindromic length of *abbaba* is equal to

3, since this word is not a concatenation of two palindromes, but $abbaba = (abba)(b)(a) = (a)(bb)(aba)$.

A decomposition into a minimal possible number of palindromes is called *optimal*.

In this paper, we are interested in the palindromic length of prefixes of an infinite word $u = u[1] \cdots u[n] \cdots$, denoted by $\mathrm{PPL}_u(n)$. The length of the shortest prefix of u of palindromic length k is denoted by $\mathrm{SP}_u(k)$ and can be considered as a kind of an inverse function to $\mathrm{PPL}_u(n)$. Clearly, $\mathrm{SP}_u(k)$ can be infinite: for example, if $u = abababab \cdots$, $\mathrm{SP}_u(k) = \infty$ for every $k \geq 3$.

The following conjecture was first formulated, in slightly different terms, in our 2013 paper with Puzynina and Zamboni [13].

Conjecture 1. For every non ultimately periodic word u, the function $\mathrm{PPL}_u(n)$ is unbounded, or equivalently, $\mathrm{SP}_u(k) < \infty$ for every $k \in \mathbb{N}$.

In fact, there were two versions of the conjecture considered in our paper [13], one concerning the prefix palindromic length and the other concerning the palindromic length of any factor of u. However, Saarela [15] later proved the equivalence of these two statements.

In the same initial paper [13], the conjecture was proved for the case when u is p-power-free for some p, as well as for the more general case when a so-called (p, l)-condition holds for some p and l. Due to the above-mentioned result by Saarela, this means that the conjecture is proven for almost all words, since almost all words contain as long p-power-free factors as needed. However, for some cases, the conjecture remains unsolved, and, for example, its proof for all Sturmian words [11] required a special technique.

Most published papers on palindromic length concern algorithmic aspects; in particular, there are several fast effective algorithms for computing $\mathrm{PPL}_u(n)$ [9, 5, 14].

The original proof of Conjecture 1 for the p-power-free words is not constructive. The upper bound for a length N such that $\mathrm{PPL}(N) \geq K$ for a given K is given as a solution of a transcendental equation and grows with K faster than any exponential function. However, this does not look to be the best possible bound. So, it is reasonable to state the following conjecture:

Conjecture 2. If a word u is p-power free for some p, then

$$\limsup \frac{\mathrm{PPL}_u(n)}{\ln n} > 0,$$

or, equivalently, $\mathrm{SP}_u(k) \leq C^k$ for some C. The constant C can be chosen independently of u as a function of p.

In this paper, we consider in detail the case of the Thue-Morse word A010060, a classical example of a word avoiding powers greater than 2 [3]. We give precise formulas for its prefix palindromic length and discuss its properties. This is a simple but necessary step before considering all p-power-free words, or all fixed points of uniform morphisms, or any other family of words containing the Thue-Morse word.

The results of this paper, in less detail, have been announced in the proceedings of DLT 2019 [10], together with some other results on the prefix palindromic length.

Throughout this paper, we use the notation $w(i..j] = w[i+1]..w[j]$ for a factor of a finite or infinite word w starting at position $i+1$ and ending at j.

The following lemma is a particular case of a statement by Saarela [15, L. 6]. We give its proof for the sake of completeness.

Lemma 3. *For every word u and for every $n \geq 0$, we have*

$$\mathrm{PPL}_u(n) - 1 \leq \mathrm{PPL}_u(n+1) \leq \mathrm{PPL}_u(n) + 1.$$

Proof. Consider the prefixes v and va of u of length n and $n+1$ respectively. Clearly, for any decomposition $u = p_1 \cdots p_k$ into k palindromes $ua = p_1 \cdots p_k a$ is a decomposition of ua into $k+1$ palindrome. On the other hand, for any palindromic decomposition $ua = q_1 \cdots q_k$, we have either $q_k = a$, and then $u = q_1 \cdots q_{k-1}$, or $q_k = ap_k a$, for a (possibly empty) palindrome p_k, and then $u = q_1 \cdots q_{k-1} a p_k$ is a decomposition of u into $k+1$ palindromes. If initial decompositions were optimal, this gives $\mathrm{PPL}_u(n+1) \leq \mathrm{PPL}_u(n) + 1$ and $\mathrm{PPL}_u(n) \leq \mathrm{PPL}_u(n+1) + 1$. $\square$

So, the first differences of the prefix palindromic length can be equal only to -1, 0, or 1, and the graph never jumps.

In this paper, it is convenient to consider the famous Thue-Morse word A010060

$$t = abbabaabbaababba \cdots$$

as the fixed point starting with a of the morphism

$$\tau : \begin{cases} a \to abba, \\ b \to baab. \end{cases}$$

Both images of letters under this morphism, which is the square of the usual Thue-Morse morphism $a \to ab, b \to ba$, are palindromes.

It is thus easy to see that every prefix of the Thue-Morse word of length 4^k is a palindrome, so that $\mathrm{PPL}_t(4^k) = 1$ for all $k \geq 0$. The first values of $\mathrm{PPL}_t(n)$ and of $\mathrm{SP}_t(k)$ are given below; also see the A307319 entry of the *On-Line Encyclopedia of Integer Sequences* (OEIS) [16].

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\mathrm{PPL}_t(n)$	1	2	2	1	2	3	3	2	3	4	3	2	3	3	2	1

As for the shortest prefix of a given palindromic length, we give its length in decimal and quaternary notation; also see A320429 in the OEIS [16].

k	1	2	3	4	5	6	7	8
$\mathrm{SP}_t(k)$	1	2	6	10	26	90	154	410
4-ary	1	2	12	22	122	1122	2122	12122

Now we are going to prove the self-similarity properties which we observe.

2 Recurrence relations

Theorem 4. *The following identities hold for all $n \geq 0$:*

$$\begin{aligned} \mathrm{PPL}_t(4n) &= \mathrm{PPL}_t(n), &(1)\\ \mathrm{PPL}_t(4n+1) &= \mathrm{PPL}_t(n) + 1, &(2)\\ \mathrm{PPL}_t(4n+2) &= \min(\mathrm{PPL}_t(n), \mathrm{PPL}_t(n+1)) + 2, &(3)\\ \mathrm{PPL}_t(4n+3) &= \mathrm{PPL}_t(n+1) + 1. &(4) \end{aligned}$$

To prove Theorem 4, we need several observations. First of all, the shortest non-empty palindrome factors in the Thue-Morse word are a, b, aa, bb, aba, bab, $abba$, $baab$. All palindromes of length more than 3 are of even length and have aa or bb in the center: if $t(i..i+2k]$ is a palindrome, then $t(i+k-1, i+k+1] = aa$ or bb.

Let us say that an occurrence of a palindrome $t(i..j]$ is of type (i', j') if i' is the residue of i and j' is the residue of j modulo 4. For example, the palindrome $t(5..7] = aa$ is of type $(1, 3)$, the palindrome $t(4, 8] = baab$ is of type $(0, 0)$, and the palindrome $t(7..9] = bb$ is of type $(3, 1)$.

Proposition 5. *Every occurrence of a palindromic factor of length not equal to one or three in the Thue-Morse word is of a type $(m, 4-m)$ for some $m \in \{0, 1, 2, 3\}$.*

Proof. Every such palindrome in the Thue-Morse word is of even length which we denote by $2k$, and every occurrence of it is of the form $t(i..i+2k]$. Its center $t(i+k-1, i+k+1]$ is equal to aa or bb, and these two words always appear in t at positions of the form $t(2l-1, 2l+1]$ for some $l \geq 1$. So, $i+k-1 = 2l-1$, meaning that $i = 2l-k$ and $i+2k = 2l+k$. So, modulo 4, we have $i + (i+2k) = 4l \equiv 0$, that is, $i \equiv -(i+2k)$. □

Note that the palindromes of odd length in the Thue-Morse word are, first, a and b, which can be of type $(0, 1)$, $(1, 2)$, $(2, 3)$ or $(3, 0)$, and second, aba and bab, which can only be of type $(2, 1)$ or $(3, 2)$.

Proposition 6. *Let $t(i..i+k]$ for $i > 0$ be a palindrome of length $k > 0$ and of type $(m, 4-m)$ for some $m \neq 0$. Then $t(i-1..i+k+1]$ is also a palindrome, as well as $t(i+1..i+k-1]$.*

Proof. The type of the palindrome is not $(0, 0)$, meaning that its first and last letters $t[i+1]$ and $t[i+k]$ are not the first and the last letters of τ-images of letters. Since these first and last letters are equal and their positions in τ-images of letters are symmetric and determine their four-blocks $abba$ or $baab$, the letters $t[i]$ and $t[i+k+1]$ are also equal, and thus $t(i-1..i+k+1]$ is a palindrome. As for $t(i+1..i+k-1]$, it is a palindrome since is obtained from the palindrome $t(i..i+k]$ by erasing the first and the last letters. □

Let us say that a decomposition of $t(0..4n]$ into palindromes is a 0-*decomposition* if all palindromes in it are of type $(0, 0)$. The minimal number of palindromes in a 0-decomposition is denoted by $\mathrm{PPL}_t^0(4n)$.

Proposition 7. *For every $n \geq 1$, we have* $\mathrm{PPL}_t(n) = \mathrm{PPL}_t^0(4n) \geq \mathrm{PPL}_t(4n)$.

Proof. It is sufficient to note that τ is a bijection between all palindromic decompositions of $t(0..n]$ and 0-decompositions of $t(0..4n]$. ☐

Proposition 8. *If* (3) *holds for $n = N - 1$, then*

$$\mathrm{PPL}_t(4N-2) > \mathrm{PPL}_t(4N). \tag{5}$$

Proof. The equality (3) means that $\mathrm{PPL}_t(4N-2) = \min(\mathrm{PPL}_t(N-1), \mathrm{PPL}_t(N)) + 2$, but since due to Lemma 3 we have $\mathrm{PPL}_t(N) \leq \mathrm{PPL}_t(N-1) + 1$, we also have $\min(\mathrm{PPL}_t(N-1), \mathrm{PPL}_t(N)) + 2 \geq \mathrm{PPL}_t(4N) + 1$. ☐

Now we can start the main proof of Theorem 4.

The proof is done by induction on n. Clearly, $\mathrm{PPL}_t(0) = 0$, $\mathrm{PPL}_t(1) = \mathrm{PPL}_t(4) = 1$, and $\mathrm{PPL}_t(2) = \mathrm{PPL}_t(3) = 2$, the equalities (1)–(4) hold for $n = 0$, and moreover, (1) is true for $n = 1$. Now suppose that they all, and, by Proposition 8, the equality (5), hold for all $n < N$, and (1) holds also for $n = N$. We fix an $N > 0$ and prove for it the following sequence of propositions.

Proposition 9. *An optimal decomposition into palindromes of the prefix $t(0..4N+1]$ cannot end with a palindrome of length 3.*

Proof. Suppose the opposite: some optimal decomposition of $t(0..4N+1]$ ends with the palindrome $t(4N-2..4N+1]$. This palindrome is preceded by an optimal decomposition of $t(0..4N-2]$. So, $\mathrm{PPL}_t(4N+1) = \mathrm{PPL}_t(4N-2) + 1$; but by (5) applied to $N-1$, which we can use by the induction hypothesis, $\mathrm{PPL}_t(4N-2) > \mathrm{PPL}_t(4N)$. So, $\mathrm{PPL}_t(4N+1) > \mathrm{PPL}_t(4N) + 1$, contradicting to Lemma 3. ☐

Proposition 10. *There exists an optimal decomposition into palindromes of the prefix $t(0..4N+2]$ that does not end with a palindrome of length 3.*

Proof. The opposite would mean that all optimal decompositions of $t(0..4N+2]$ end with the palindrome $t(4N-1..4N+2]$ preceded by an optimal decomposition of $t(0..4N-1]$. So, $\mathrm{PPL}_t(4N+2) = \mathrm{PPL}_t(4N-1) + 1$; by the induction hypothesis, $\mathrm{PPL}_t(4N-1) = \mathrm{PPL}_t(4N) + 1$. So, $\mathrm{PPL}_t(4N+2) = \mathrm{PPL}_t(4N) + 2$, and thus another optimal decomposition of $t(0..4N+2]$ can be obtained as an optimal decomposition of $t(0..4N]$ followed by two palindromes of length 1. A contradiction. ☐

Proposition 11. *For every $m \in \{1, 2, 3\}$, the equality holds*

$$\mathrm{PPL}_t(4N+m) = \min(\mathrm{PPL}_t(4N+m-1), \mathrm{PPL}_t(4N+m+1)) + 1.$$

Proof. Consider an optimal decomposition $t(0..4N+m] = p_1 \cdots p_k$, where $k = \mathrm{PPL}_t(4N+m)$. Denote the ends of palindromes as $0 = e_0 < e_1 < \cdots < e_k = 4N + m$, so that $p_i = t(e_{i-1}, e_i]$

for each i. Since $m \neq 0$ and due to Proposition 5, there exist some palindromes of length 1 or 3 in this decomposition. Let p_j be the last of them.

Suppose first that $j = k$. Then due to the two previous propositions, p_k can be taken of length 1 not 3, so that $t(0..4N + m - 1] = p_1 \cdots p_{k-1}$ is decomposable to $k - 1$ palindromes. Due to Lemma 3, we have $\mathrm{PPL}_t(4N+m-1) = k-1$, and thus $\mathrm{PPL}_t(4N+m) = \mathrm{PPL}_t(4N+m-1)+1$. Again due to Lemma 3, we have $\mathrm{PPL}_t(4N + m + 1) \geq \mathrm{PPL}_t(4N + m) - 1 = \mathrm{PPL}_t(4N + m - 1)$, and so the statement holds.

Now suppose that $j < k$, so that $e_j \equiv -e_{j+1} \equiv e_{j+2} \equiv \cdots \equiv (-1)^{k-j} e_k \pmod 4$. Here p_j is the last palindrome in an optimal decomposition of $p_1 \cdots p_j$ and it is of length 1 or 3. But if $e_j \equiv 1$ or $2 \pmod 4$, p_j can be taken of length 1 due to the two previous propositions applied to some smaller length; and if $e_j \equiv 3 \pmod 4$, it is of length 1 since the suffix of length 3 of $t(0..4n + 3])$ is equal to *abb* or to *baa*, so, it is not a palindrome. So, anyway, we can take p_j of length one: $p_j = t(e_j - 1, e_j]$.

Since $e_j \equiv \pm e_k$ and $e_k \equiv m \neq 0 \pmod 4$, we may apply Proposition 6 and see that $p'_j = t(e_j - 1..e_{j+1} + 1]$ is a palindrome, as well as $p'_{j+1} = t(e_{j+1} + 1..e_{j+2} - 1]$ and so on up to $p'_{k-1} = t(e_{k-1} + (-1)^{k-j}..e_k - (-1)^{k-j}]$. So, $p_1 \cdots p_{j-1} p'_j \cdots p'_{k-1}$ is a decomposition of $t(0..4N + m - (-1)^{k-j}]$ into $k - 1$ palindromes. So, as above, $\mathrm{PPL}_t(4N + m) = \mathrm{PPL}_t(4N + m - (-1)^{k-j}) + 1$, and since $\mathrm{PPL}_t(4N + m + (-1)^{k-j}) \geq \mathrm{PPL}_t(4N + m) - 1 = \mathrm{PPL}_t(4N + m - (-1)^{k-j})$, the proposition holds. □

Proposition 12. *Every optimal palindromic decomposition of* $t(0..4N+4]$ *is a* 0*-decomposition, and thus* $\mathrm{PPL}_t(4N + 4) = \mathrm{PPL}_t(N + 1)$.

Proof. Suppose the opposite; then the last palindrome in the optimal decomposition that is not of type $(0, 0)$ is of type $(m, 0)$ and thus is of length 1 or 3. Since the proof of Theorem 4 proceeds by induction on N, this proposition is true for all $n < N$, and thus the palindrome of type $(m, 0)$ is the very last palindrome of the optimal decomposition. Since the suffix of length 3 of $t(0..4N + 4]$ is equal to *bba* or *aab* and thus is not a palindrome, the last palindrome of the optimal decomposition is of length 1, meaning that $\mathrm{PPL}_t(4N + 4) = \mathrm{PPL}_t(4N + 3) + 1$. Now let us use Proposition 11 applied to $m = 3, 2, 1$; every time we get $\mathrm{PPL}_t(4N + m) = \mathrm{PPL}_t(4N + m - 1) + 1$. Using all these equalities together, we get $\mathrm{PPL}_t(4N + 4) = \mathrm{PPL}_t(4N) + 4$, which is impossible since $\mathrm{PPL}_t(4N) = \mathrm{PPL}_t(N)$ and $\mathrm{PPL}_t(4N + 4) \leq \mathrm{PPL}_t(N + 1) \leq \mathrm{PPL}_t(N) + 1$. A contradiction. □

We have proven (1) for $n = N + 1$. It remains to prove (2)–(4) for $n = N$. Indeed, we know that

$$-1 \leq \mathrm{PPL}_t(4N + 4) - \mathrm{PPL}_t(4N) = \mathrm{PPL}_t(N + 1) - \mathrm{PPL}_t(N) \leq 1. \tag{6}$$

Now to prove (2) suppose by contrary that $\mathrm{PPL}_t(4N + 1) \leq \mathrm{PPL}_t(4N) = \mathrm{PPL}_t(N)$. Due to Proposition 11, this means that $\mathrm{PPL}_t(4N + 1) = \mathrm{PPL}_t(4N + 2) + 1$, that is, $\mathrm{PPL}_t(4N +$

2) $< \mathrm{PPL}_t(4N)$, and, again by Proposition 11, $\mathrm{PPL}_t(N+1) = \mathrm{PPL}_t(4N+2) - 2$. Thus, $\mathrm{PPL}_t(N) - \mathrm{PPL}_t(N+1) \geq 3$, a contradiction to (6). So, (2) is proven.

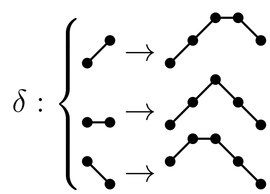
The equality (4) is proven symmetrically. Now (3) follows from both these equalities in combination with Proposition 11, completing the proof of Theorem 4.

3 Corollaries

The first differences $(d_t(n))_{n=0}^{\infty}$ of the prefix palindromic length are defined by $d_t(n) = \mathrm{PPL}_t(n+1) - \mathrm{PPL}_t(n)$; here we set $\mathrm{PPL}_t(0) = 0$. Due to Lemma 3, $d_t(n) \in \{-1, 0, +1\}$ for every n; so, it is a sequence over a finite alphabet, which we prefer to denote by $\{\searrow, \relbar, \nearrow\}$. We write these symbols joining the ends of intervals from left to right, so that the sequence $(d_t(n))$ becomes the plot of $\mathrm{PPL}_t(n)$.

The following corollary of Theorem 4 is more or less straightforward.

Corollary 13. *The sequence* $(d_t(n))$ *is the fixed point of the morphism*



Proof. Theorem 4 immediately means that $\mathrm{PPL}_t(4n), \ldots, \mathrm{PPL}_t(4n+4)$ are determined by $\mathrm{PPL}_t(n)$ and $\mathrm{PPL}_t(n+1)$, and moreover, $d_t(4n)$, ..., $d_t(4n+3)$ are determined by $d_t(n)$. This means exactly that the sequence $d_t(n)$ is a fixed point of a morphism of length 4. The equality (2) means that the first symbol of any morphic image of δ is $\nearrow$; the equality (4) means that the last symbol of any morphic image of δ is $\searrow$; the two symbols in the middle can be found from (3) and depend on $d_t(n)$. $\square$

With the previous corollary, we can draw the plot of $\mathrm{PPL}_t(n)$ as the fixed point of δ.

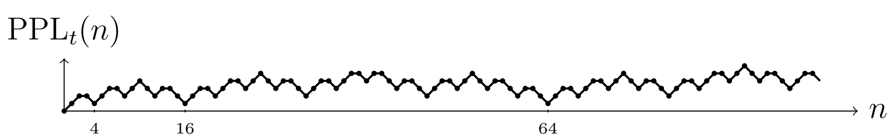


Figure 1: $\mathrm{PPL}_t(n)$

The next proposition can be obtained from Theorem 4 by elementary computations. Recall that $\mathrm{SP}(k) = \mathrm{SP}_t(k)$ is the length n of the shortest prefix of t such that its palindromic length $\mathrm{PPL}_t(n)$ is equal to k.

Proposition 14. *We have* $\mathrm{SP}(1) = 1$, $\mathrm{SP}(2) = 2$, $\mathrm{SP}(3) = 6$ *and for all* $k > 0$,

$$\mathrm{SP}(k+3) = 16\,\mathrm{SP}(k) - 6.$$

Proof. Let us introduce $\mathrm{SP}_2(k)$ as the minimal number n such that $\mathrm{PPL}_t(n) = \mathrm{PPL}_t(n+1) = k$. By definition, $\mathrm{SP}_2(k) \geq \mathrm{SP}(k)$. The first values of $\mathrm{SP}(k)$ and $\mathrm{SP}_2(k)$ are given below.

k	1	2	3	4	5
$\mathrm{SP}(k)$	1	2	6	10	26
$\mathrm{SP}_2(k)$	∞	2	6	22	38

From the definition of the morphism δ we immediately see that a new value $n = \mathrm{SP}(k)$ can appear either in the middle of the δ-image of $d_t(n') = d_t(\mathrm{SP}_2(k-2))$, or in the middle of the δ-image of $d_t(n'')$, where $n'' = \mathrm{SP}(k-1) - 1$. The latter case is also the only possible way to get a new value $n = \mathrm{SP}_2(k)$. So,

$$\mathrm{SP}(k) = \min(4\,\mathrm{SP}_2(k-2) + 2, 4\,\mathrm{SP}(k-1) - 2), \tag{7}$$

$$\mathrm{SP}_2(k) = 4\,\mathrm{SP}(k-1) - 2. \tag{8}$$

As we see from the table, for $3 \leq k \leq 5$, we have $\mathrm{SP}(k-1) \leq SP_2(k-1) < \mathrm{SP}(k)$. The first inequality is obvious, but let us prove the second one by induction. Its base is observed for $3 \leq k \leq 5$, so, consider $k \geq 6$ such that for all $k' = k-3, k-2, k-1$ we have $\mathrm{SP}_2(k'-1) < \mathrm{SP}(k')$.

In particular, $\mathrm{SP}_2(k-4) < \mathrm{SP}(k-3)$, so due to (7), we have $\mathrm{SP}(k-2) = 4\,\mathrm{SP}_2(k-4)+2$, and so due to (8),

$$\mathrm{SP}_2(k-1) = 16\,\mathrm{SP}_2(k-4) + 6. \tag{9}$$

On the other hand, we have $\mathrm{SP}_2(k-2) < \mathrm{SP}(k-1)$, so, (7) becomes $\mathrm{SP}(k) = 4\,\mathrm{SP}_2(k-2)+2$, and together with (8) this gives

$$\mathrm{SP}(k) = 16\,\mathrm{SP}(k-3) - 6. \tag{10}$$

Combining (9), (10), the induction base $\mathrm{SP}_2(k-4) < \mathrm{SP}(k-3)$ and the fact that all the values are integers, we obtain that $\mathrm{SP}_2(k-1) < \mathrm{SP}(k)$ for all $k \geq 3$. We also see that (10) is true for all $k \geq 4$, proving this proposition. □

The following corollary of the previous proposition can be proved by straightforward induction.

Corollary 15. *In the 4-ary numeration system, we have* $\mathrm{SP}(3k+2) = ((12)^k 2)_4$ *for all* $k \geq 0$*;* $\mathrm{SP}(3k) = (1(12)^{k-1}2)_4$ *for all* $k \geq 1$*;* $\mathrm{SP}(3k+1) = (2(12)^{k-1}2)_4$ *for all* $k \geq 1$.

Another direct consequence of Proposition 14 is

Corollary 16. *We have*

$$\limsup \frac{\mathrm{PPL}_t(n)}{\ln n} = \frac{3}{4\ln 2},$$

whereas $\liminf \dfrac{\mathrm{PPL}_t(n)}{\ln n} = 0$ *since* $\mathrm{PPL}_t(4^m) = 1$ *for all* m.

4 Regularity

The sequence $(\mathrm{PPL}_t(n))$ is closely related to the Thue-Morse word, the most classical example of a 2-automatic sequence. In general, a sequence $w = (w[n])$ is called k-automatic if there exists a finite automaton such that for the input equal to the k-ary representation of n, the output is equal to $w[n]$. Equivalently, due to a theorem by Cobham [7], a sequence is k-automatic if and only if it is an image under a coding $c : \Sigma \to \Delta$ of a fixed point of a k-uniform morphism φ: $w = c(w')$, where $w' = \varphi(w')$ [1, Ch. 6]. So, the Thue-Morse word is 2-automatic since it is a fixed point of the 2-uniform morphism $a \to ab, b \to ba$, and the sequence $(d_t(n))$ is 4-automatic since it is a fixed point of δ. In both cases, the coding can be taken to be trivial: $c(x) = x$ for every letter x. It is also well-known that a sequence is k-automatic if and only if it is k^m-automatic for any integer m, so, in particular, being 2- and 4-automatic is the same property. So, the Thue-Morse word is also 4-automatic and the sequence $(d_t(n))$ is 2-automatic.

A more general notion of a *k-regular sequence* was introduced by Allouche and Shallit [2]; also see [1, Ch. 16]. A sequence $(a(n))$ is called k-regular (on $\mathbb{Z}$) if there exists a finite number of sequences $\{(a_1(n)), \dots, (a_s(n))\}$ such that for every integer $i \geq 0$ and $0 \leq b < k^i$ there exist $c_1, \dots, c_s \in \mathbb{Z}$ such that for all $n \geq 0$ we have

$$a(k^i n + b) = \sum_{1 \leq j \leq s} c_j a_j(n).$$

It is also known that a sequence is k-automatic if and only if it is k-regular and takes on finitely many values [1, Thm. 16.1.4]. Moreover, a sequence $a = (a(n))$ is k-regular if and only if there exist r sequences $a_1 = a, \dots, a_r$ and a matrix-valued morphism μ such that if

$$V(n) = \begin{pmatrix} a_1(n) \\ a_2(n) \\ \cdots \\ a_r(n) \end{pmatrix},$$

then

$$V(kn + b) = \mu(b) V(n)$$

for $0 \leq b < k$ [1, Thm. 16.1.3]. Many sequences related to k-automatic words are k-regular, as it was shown by Charlier, Rampersad and Shallit [6]. However, it seems that the general approach from the mentioned paper does not directly work for the palindromic length, so, we have to prove the following corollary only for the Thue-Morse word.

Corollary 17. *The sequence* $\mathrm{PPL}_t(n)$ *is 4-regular.*

Proof. Consider the auxiliary sequences $(d(n))$, which is $(d_t(n))$ interpreted as a sequence over the alphabet $\{-1, 0, 1\}$; $(e(n))$ defined by

$$e(n) = \begin{cases} 0, & \text{if } d(n) = -1; \\ 1, & \text{otherwise,} \end{cases}$$

and $(h(n))$ defined by

$$h(n) = \begin{cases} 1, & \text{if } d(n) = 1; \\ 0, & \text{otherwise.} \end{cases}$$

Consider the vector

$$V(n) = \begin{pmatrix} PPL_t(n) \\ d(n) \\ e(n) \\ h(n) \\ 1 \end{pmatrix}.$$

Then Theorem 4 and Corollary 13 immediately imply that

$$V(4n+b) = \mu(b)V(n)$$

for every $b = 0, \ldots, 3$, where

$$\mu(0) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \mu(1) = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mu(2) = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \mu(3) = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

As it is discussed above, these equalities are sufficient to claim that the first sequence in the vector, $PPL_t(n)$, is 4-regular. □

Remark 18. In fact, every sequence with k-automatic first differences is k-regular, which can be proven with a similar construction, perhaps with more auxiliary sequences like $(e(n))$ and $(h(n))$.

5 Conclusion

To my knowledge, the results of this paper are thus far, the only precise formulas for the prefix palindromic length of a non-trivial infinite word not constructed especially for that purpose.

Even for famous and simple examples like Toeplitz words or the Fibonacci word A003849, lower bounds for the prefix palindromic length are difficult [12, 10]. The only more or less universal lower bounds for all p-power-free words are those from the first paper on the subject [13], with $\mathrm{SP}(k)$ growing faster than any constant to the power k. Later [12], some calculations allowed a reasonable exponential conjecture on the $\mathrm{SP}(k)$ of the Fibonacci word, but it is not clear how to prove it. So, the following more particular open questions can be added to general Conjectures 1 and 2.

Problem 19. Find a precise formula for the prefix palindromic length of the period-doubling word A096268, or a lower bound for its $\limsup$.

Problem 20. Find a precise formula for the prefix palindromic length of the Fibonacci word A003849, or a lower bound for its $\limsup$.

Problem 21. Is it true that the function $\mathrm{PPL}_u(n)$ is k-regular for any k-automatic word u? Fibonacci-regular for the Fibonacci word? For the definition and discussions of Fibonacci-regular sequences, see [8] and references therein.

References

[1] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, 2003.

[2] J.-P. Allouche and J. Shallit, The ring of k-regular sequences. *Theoret. Comput. Sci.* **98** (1992), 163–197.

[3] J.-P. Allouche and J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence. In: C. Ding, T. Helleseth and H. Niederreiter, eds., *Sequences and Their Applications: Proc. SETA 1998*, Springer, 1999, pp. 1–16.

[4] P. Ambrož, O. Kadlec, Z. Masáková, and E. Pelantová, Palindromic length of words and morphisms in class $\mathcal{P}$. *Theoret. Comp. Sci.* **780** (2019), 74–83.

[5] K. Borozdin, D. Kosolobov, M. Rubinchik, and A. M. Shur, Palindromic length in linear time. In: J. Kärkkäinen, J. Radoszewski, and W. Rytter, eds., *Proc. CPM 2017*, Dagstuhl Publishing, 2017, pp. 23:1–23:12.

[6] E. Charlier, N. Rampersad, and J. Shallit, Enumeration and decidable properties of automatic sequences, *Int. J. Found. Comput. Sci.* **23** (2012), 1035–1066.

[7] A. Cobham. Uniform tag sequences. *Math. Systems Theory* **6** (1972), 164–192.

[8] Chen Fei Du, H. Mousavi, L. Schaeffer, and J. Shallit, Decision Algorithms for Fibonacci-Automatic Words, III: Enumeration and Abelian Properties, *Int. J. Found. Comput. Sci.* **27** (2016), 943–963.

[9] G. Fici, T. Gagie, J. Kärkkäinen, and D. Kempa, A subquadratic algorithm for minimum palindromic factorization. *J. Discr. Alg.* **28** (2014), 41–48.

[10] A. E. Frid, First lower bounds on palindromic length, in *Developments in Language Theory: Proc. DLT 2019*, Lect. Notes in Comp. Sci., Vol. 11647, Springer, 2019, pp. 234–243.

[11] A. E. Frid, Sturmian numeration systems and decompositions to palindromes. *European J. Combin.* **71** (2018), 202–212.

[12] A. Frid, Representations of palindromes in the Fibonacci word, in *Proc. Numeration 2018*,https://numeration2018.sciencesconf.org/data/pages/num18_abstracts.pdf, pp. 9–12.

[13] A. Frid, S. Puzynina, and L. Zamboni. On palindromic factorization of words. *Advances in Appl. Math.* **50** (2013), 737–748.

[14] M. Rubinchik and A. M. Shur, EERTREE: An efficient data structure for processing palindromes in strings. *European J. Combin.* **68** (2018), 249–265.

[15] A. Saarela, Palindromic length in free monoids and free groups, in: *Proc. WORDS 2017*, Lect. Notes in Comp. Sci., Vol. 10432, Springer 2017, pp. 203–213.

[16] N. J. A. Sloane et al., The On-Line Encyclopedia of Integer Sequences, https://oeis.org.

2010 *Mathematics Subject Classification*: Primary 68R15.
Keywords: Thue-Morse word, palindrome, palindromic length, k-automatic word, k-regular sequence.

(Concerned with sequences A003849, A010060, A096268, A307319, and A320429.)

Received August 28 2019; revised version received November 28 2019. Published in *Journal of Integer Sequences*, November 30 2019.

Return to Journal of Integer Sequences home page.