Privileged Factors in the Thue-Morse Word – A Comparison of Privileged Words and Palindromes

Jarkko Peltomäki
jspelt@utu.fi

Turku Centre for Computer Science TUCS, 20520 Turku, Finland
University of Turku, Department of Mathematics and Statistics, 20014 Turku, Finland

Abstract

In this paper we continue the studies of [J. Peltomäki, Introducing Privileged Words: Privileged Complexity of Sturmian Words, Theor. Comp. Sci. (2013)] on so-called privileged words. In this earlier work the basic properties of privileged words were proved along with a result that relates privileged words with so-called rich words (see [A. Glen et al, Palindromic Richness, Eur. Jour. of Comb. 30 (2009) 510–531]) and a result characterizing Sturmian words using privileged words. The main result of this paper is a derivation of a recursive formula for the privileged complexity function of the Thue-Morse word. Using the formula it’s proven that this function is unbounded, but its values still have arbitrary large gaps of zeros. All in all the behavior of this complexity function differs radically from the Sturmian case investigated in the earlier work. In addition we further study the relation between rich words and privileged words and compare the behavior of palindromes and privileged words in infinite words.

Keywords: combinatorics on words, thue-morse word, palindromes, privileged words, return words, rich words

1 Introduction

In this paper we continue the study of so-called privileged words. The study of privileged words was initiated in articles [KLS11] and [Pel13], and they appear also in [FPZ12]. In [KLS11] privileged words were used as a technical tool in order to characterize aperiodic and minimal subshifts with bounded powers. This new class of words has also interest in its own right. In [Pel13] the author of this paper for example characterized Sturmian words using so-called privileged complexity function, which counts the number of privileged words of given length occurring in an infinite word.

The motivation for defining privileged words comes from the theory of so-called rich words [Gle+09]. Rich words are words containing maximal number of palindromes as factors, and they are characterized by the fact that in a rich word every palindrome is a complete first return to a shorter palindrome. By altering this condition slightly we have the definition of privileged words: a word is privileged if it is a complete first return to a shorter privileged word, the shortest
privileged words are the letters of the alphabet and the empty word. At first glance privileged words have nothing to do with palindromes. In general this is indeed true, but there are some connections to rich words. Namely in [KLS11] and [Pel13] it was observed that a word is rich if and only if its set of palindromes is exactly the set of its privileged words. This means in a sense that the number of privileged words which are not palindromes measures the so-called palindromic defect of a word, that is, this number measures the number of such positions in a word which don’t introduce new palindromes. Rich words are exactly the words with zero defect. Other than this connection to richness, privileged words have other similarities with palindromes. For instance a privileged prefix of a privileged word occurs also as a suffix, just as a palindromic prefix of a palindrome occurs also as a suffix.

In general privileged words behave differently. In this paper we show that the privileged complexity function of an infinite word may behave in a much more complex way than its palindromic complexity function. We investigate privileged factors of the Thue-Morse word, and prove that in the values of its privileged complexity function \( \mathcal{A}_t \) there are arbitrarily long but not infinite gaps of zeroes. For any palindromic complexity function such behavior is impossible. We also prove that the function \( \mathcal{A}_t \) is unbounded. In contrast it has been proven in [DZ00] and [All+03] that the palindromic complexity function of fixed points of primitive morphisms is bounded. To obtain the above results we derive recursive formulas for computing values of \( \mathcal{A}_t \). Various other kinds of complexity functions for the Thue-Morse word have been considered in the past. Recently in [GMS13] Goč, Mousavi and Shallit proved that there exists a system of recurrences for the number of unbordered factors of given length in the Thue-Morse word.

After defining the necessary notations and definitions, in Section 3 we compare palindromes and privileged words. First we sharpen further the connection to palindromic richness mentioned above. We also investigate when palindromic factors can be a proper subset of privileged factors in a word (remember in rich words palindromes are exactly the privileged words). We conclude our comparison by noting that in an infinite word with infinite defect privileged words and palindromes may behave radically differently.

In Section 4 we conclude study of the privileged complexity function of the Thue-Morse word initiated in [Pel13] and present the main results of this paper. We describe the relations between different kinds of privileged factors in the Thue-Morse word, and give a recursive formula for computing the values of its privileged complexity function. We study the asymptotic behavior of this complexity function, and find that it’s unbounded. We also verify a conjecture of [Pel13] which states that the values of the privileged complexity function of the Thue-Morse word contains arbitrarily long gaps of zeros.

In the last Section we briefly study the privileged palindrome complexity function of the Thue-Morse, that is, the number of privileged factors which are also palindromes. As in Section Section 4 we give a recursive formula for computing the values of this function, and further study its asymptotic behavior and gaps of zeros.

2 Preliminaries

In this text, we denote by \( A \) a finite alphabet, which is a finite non-empty set of symbols. The elements of \( A \) are called letters. A (finite) word over \( A \) is a sequence of letters. To the empty
sequence corresponds the \textit{empty word}, denoted by \( \epsilon \). The set of all finite words over \( A \) is denoted by \( A^* \). The set of non-empty words over \( A \) is the set \( A^+ := A^* \setminus \{ \epsilon \} \). A natural operation of words is concatenation. Under this operation \( A^* \) is a free monoid over \( A \). The letters occurring in the word \( w \) form the \textit{alphabet} of \( w \) denoted by \( \text{Alph}(w) \). From now on we assume that binary words are over the alphabet \( \{0, 1\} \). Given a finite word \( w = a_1 a_2 \cdots a_n \) of \( n \) letters, we say that the \textit{length} of \( w \), denoted by \( |w| \), is equal to \( n \). By convention the length of the empty word is 0. The set of all words of length \( n \) over the alphabet \( A \) is denoted \( A^n \).

An infinite word \( w \) over \( A \) is a function from the natural numbers with values in \( A \). We write consicely \( w = a_1 a_2 a_3 \cdots \) with \( a_i \in A \). The set of infinite words is denoted by \( A^\omega \). For infinite word \( w \) we denote \( |w| = \infty \). The infinite word \( w \) is said to be \textit{ultimately periodic} if it can be written in the form \( w = uv^n = uvv \cdots \) for some words \( u, v \in A^* \), \( v \neq \epsilon \). If \( u = \epsilon \), then \( w \) is said to be \textit{periodic}. An infinite word which is not ultimately periodic is said to be \textit{aperiodic}.

A finite word \( u \) is a \textit{factor} of the finite or infinite word \( w \) if it can be written that \( w = zuv \) for some \( z \in A^* \) and \( v \in A^* \cup A^\omega \). If \( z = \epsilon \), the factor \( u \) is called a \textit{prefix} of \( w \). If \( v = \epsilon \), then we say that \( u \) is a \textit{suffix} of \( w \). If word \( u \) is both a prefix and a suffix of \( w \), then \( u \) is a \textit{border} of \( w \). The set of factors of \( w \) is denoted by \( \mathcal{F}(w) \). The set \( \mathcal{F}_n(w) \) is defined to contain all factors of \( w \) of length \( n \). We call \( u \) a \textit{central factor} of \( w \) if there exists a factorization \( w = xuy \) with \( |x| = |y| \). If \( w = a_1 a_2 \cdots a_n \), then we denote \( w[i, j] = a_i \cdots a_j \) whenever the choices of positions \( i \) and \( j \) make sense. This notion is extended to infinite words in a natural way. An \textit{occurrence} of \( u \) in \( w \) is such a position \( i \), that \( w[i, i + |u| - 1] = u \). If such a position exists, we say that \( u \) \textit{occurs} in \( w \). If \( w \) has exactly one occurrence of \( u \), then we say that \( u \) is \textit{unioccurrent} in \( w \). We say that a position \( i \) introduces a factor \( u \) if \( w[i - |u| + 1, i] = u \), and \( u \) is unioccurrent in \( w[1, i] \). A \textit{complete first return} to the word \( w \) is a word starting and ending with \( u \), and containing exactly two occurrences of \( u \). A word which is a complete first return to some word is called a \textit{complete return word}. A \textit{complete return factor} is a factor of some word which is a complete return word.

An infinite word \( w \) is \textit{recurrent} if each of its factors occurs in it infinitely often. The word \( w \) is called \textit{uniformly recurrent} if each factor \( u \) occurs in it infinitely often, and the gap between two occurrences of \( u \) in \( w \) is bounded by a constant depending only on \( u \). Equivalently \( w \) is uniformly recurrent if for each factor \( u \) there exists an integer \( R \) such that every factor of \( w \) of length \( R \) contains an occurrence of \( u \).

Let \( A \) and \( B \) be two alphabets. A \textit{morphism} from \( A^* \) to \( B^* \) is a mapping \( \varphi : A^* \to B^* \) such that \( \varphi(uxv) = \varphi(u)\varphi(v) \) for all words \( u, v \in A^* \). Morphism \( \varphi \) is said to be \textit{prolongable} on letter \( a \) if \( \varphi(a) \) begins with letter \( a \) and \( |\varphi(a)| > 1 \). Then clearly \( \varphi^n(a) \) is a prefix of \( \varphi^{n+1}(a) \), so we obtain a unique fixed point \( \varphi^\omega(a) := \lim_{n \to \infty} \varphi^n(a) \).

The word \( \partial(u, i, j) \), where \( i + j < |u| \), is obtained from the word \( u \) by deleting \( i \) letters from the beginning, and \( j \) letters from the end. Let \( \varphi \) be a morphism with fixed point \( w = \varphi^\omega(a) \). We say that a factor \( u \) of \( w \) admits an \textit{interpretation} \( s = (x_0 x_1 \ldots x_{n+1}, i, j) \) by \( \varphi \) if \( u = \partial(\varphi(x_0 x_1 \ldots x_{n+1}), i, j) \) where \( x_i \) are letters, \( 0 \leq i < |\varphi(x_0)| \) and \( 0 \leq j < |\varphi(x_{n+1})| \). The word \( x_0 x_1 \ldots x_{n+1} \) is called the \textit{ancestor} of the interpretation \( s \). In this paper we are focused on the Thue-Morse morphism, and in this particular case all long enough factors of the fixed point have a unique interpretation. In this case it’s convenient to just talk about the interpretation and the ancestor of \( u \) by \( \varphi \). In a factor \( u \) of \( w \) we often separate images of letters by bar lines. For
example if $\varphi : 0 \mapsto 01, 1 \mapsto 10$ is the Thue-Morse morphism, then the word 01100 has ancestor 010, and we place bars as follows: 01|10|0. If a factor has a unique interpretation, then there is only one way to place the bar lines in that factor.

The reversal $\bar{w}$ of $w = a_1 a_2 \cdots a_n$ is the word $\bar{w} = a_n \cdots a_2 a_1$. If $\bar{w} = w$, then we say that $w$ is a palindrome. By convention the empty word is a palindrome. The set of palindromes of $w$ is denoted by $\text{Pal}(w)$. Moreover we define $\text{Pal}_n(w) = \text{Pal}(w) \cap \mathcal{F}_n(w)$. We say that a word $w$ is closed under reversal if for each $u \in \mathcal{F}(w)$ it holds that $\bar{u} \in \mathcal{F}(w)$. It is well-known that a finite word $w$ contains at most $|w| + 1$ distinct palindromic factors. The equality is attained if each position of $w$ introduces exactly one new palindrome. The (palindromic) defect $\mathcal{D}(w)$ of $w$ is the number $|w| + 1 - |\text{Pal}(w)|$. The defect measures the abundance of palindromes in $w$. If $\mathcal{D}(w) = 0$ (i.e. it contains the maximum possible number of distinct palindromes), then $w$ is called rich. The defect for infinite word $w$ is defined as $\mathcal{D}(w) = \sup \{ \mathcal{D}(u) : u \text{ is a prefix of } w \}$. The defect of a finite word $w$ is equal to the number of prefixes $u$ of $w$ such that $u$ doesn’t have a unioccurrent longest palindromic suffix. For a good reference on defect and rich words see [Gle+09].

Privileged words were first defined in [KLS11] and further developed in [Pel13]. The set of privileged words over alphabet $A$, denoted $\text{Pri}(A)$, is defined as follows:

1. $u \in \text{Pri}(A)$ if $u$ is a complete first return to a shorter privileged word $v \in \text{Pri}(A)$,
2. the shortest privileged words are the letters $A$ and the empty word $\epsilon$.

The set of privileged factors of a word $w$ is denoted $\text{Pri}(w)$. We define that $\text{Pri}_n(w) = \text{Pri}(w) \cap \mathcal{F}_n(w)$. Privileged words have the following elementary properties proved in [Pel13]:

**Lemma 2.1.** Let $w$ be a privileged word, and $u$ its any privileged prefix (respectively suffix). Then $u$ is a suffix (respectively prefix) of $w$. \hfill $\square$

**Lemma 2.2.** Let $w$ be a privileged word, and $u$ its longest proper privileged prefix (or suffix). Then $w$ is a complete first return to $u$. \hfill $\square$

**Lemma 2.3.** Every position in a word introduces exactly one new privileged factor \hfill $\square$

### 3 A Comparison of Palindromes and Privileged Words

In this section we compare the behavior of palindromes and privileged words and the behavior of the respective complexity functions in infinite words.

Let $w$ be an infinite word. In [Pel13] the following relation between the sets $\text{Pal}(w)$ and $\text{Pri}(w)$ was proved:

**Proposition 3.1.** [Pel13] A finite or infinite word $w$ is rich if and only if $\text{Pri}(w) = \text{Pal}(w)$. \hfill $\square$

First we strengthen this result a bit. Now if a word is rich, then by the previous Proposition it obviously follows that the palindromic and privileged complexity functions of $w$ coincide. Next we prove the surprising fact that the converse is also true. We start with a Lemma which is interesting in its own right.

**Lemma 3.2.** Let $w$ be a finite or infinite word. If $w$ is not rich, then there exists a shortest privileged factor $u$ which is not a palindrome. Moreover $\text{Pal}_w(n) = \text{Pri}_w(n)$ for $0 \leq n < |w|$ and $\text{Pal}_w(|u|) \subset \text{Pri}_w(|u|)$.
Proof. If \( w \) is not rich, then there exists a position \( n \) such that no new palindrome in position \( n \) is introduced. However position \( n \) introduces a new privileged factor, which thus can’t be a palindrome. Hence there exists a shortest privileged factor \( u \) which is not a palindrome. By the minimality of \( |u| \) it follows that \( \text{Pri}_w(n) \subseteq \text{Pal}_w(n) \) for all \( 0 \leq n < |u| \). Let then \( p, |p| > 1 \), be a minimal length palindrome which is not privileged. Let \( q \) be the longest proper palindromic suffix of \( p \). By minimality \( q \) is privileged. As \( p \) is not privileged, it has as a suffix a complete first return to \( q \), say \( v \). As \( q \) is the longest palindromic suffix of \( p, v \) is not a palindrome. By minimality of \( |u| \) we have that \( |p| > |v| \geq |u| \), so \( \text{Pal}_w(n) \subseteq \text{Pri}_w(n) \) for all \( 0 \leq n \leq |u| \). \( \square \)

**Proposition 3.3.** A finite or infinite word \( w \) is rich if and only if \( \text{Pal}_w(n) = \text{A}_w(n) \) for all \( 0 \leq n \leq |w| \).

Proof. The fact that the condition is necessary follows from Proposition 3.1. Assume that \( \text{Pal}_w(n) = \text{A}_w(n) \) for all \( 0 \leq n \leq |w| \). If \( w \) is not rich, then by Lemma 3.2 there exists such \( n \) that \( \text{Pal}_w(n) \subset \text{Pri}_w(n) \), so \( \text{Pal}_w(n) < \text{A}_w(n) \), which is not possible. Therefore \( w \) is rich. \( \square \)

Now for instance the Thue-Morse word (see Section 4) has as a factor the word 00101100 which is privileged and not palindromic, and the palindrome 00101100110100 which is privileged. Thus for a word \( w \) its possible that neither of the sets \( \text{Pal}(w) \) and \( \text{Pri}(w) \) is included in the other. By Lemma 3.2 it follows that if \( \text{Pri}(w) \subseteq \text{Pal}(w) \), then \( \text{Pri}(w) = \text{Pal}(w) \), i.e. \( w \) is rich. Next its natural to ask if there are examples of infinite words \( w \) such that \( \text{Pal}(w) \) is properly contained in \( \text{Pri}(w) \). It turns out that this is possible, but not in the case of uniformly recurrent words containing infinitely many palindromes. We begin with a simple observation.

**Lemma 3.4.** Let \( w \) be a recurrent infinite word. If \( \text{Pal}(w) \subset \text{Pri}(w) \), then \( w \) has infinite defect.

Proof. As \( \text{Pal}(w) \subset \text{Pri}(w) \) there exists a privileged factor \( u \) which is not a palindrome. Consider any factor \( v \) which is a complete first return to \( u \). Let \( p \) be the longest palindromic suffix of \( v \). By assumption \( p \) is also privileged. If \( |p| > |u| \), then \( p \) has \( \tilde{u} \) a prefix. Since \( p \) has a privileged suffix \( u \), it also has \( u \) as a prefix, so \( \tilde{u} = u \), which is impossible. Thus \( |p| < |u| \), so \( p \) is actually the longest palindromic suffix of \( u \). Hence any prefix of \( w \) having \( v \) as a suffix doesn’t mean that \( w \) has infinite defect. \( \square \)

Next we define an infinite binary word \( \kappa = \lim_{n \rightarrow \infty} u_n \) as the limit of the sequence \( u_0 = 00101100, u_{n+1} = u_n 0^m u_n \). It is clear that \( \kappa \) is recurrent and aperiodic, and contains infinitely many palindromes of the form \( 0^n \). The word \( \kappa \) is however not closed under reversal as \( (1011)^\sim = 1101 \) is not a factor. We claim that \( \text{Pal}_\kappa(n) = \{0^n, 10^{n-2}1\} \) for \( n \geq 7 \). Let \( p \in \text{Pal}_\kappa(n) \) for \( n \geq 7 \), and \( m \) be minimal such that \( p \) occurs in \( u_m \). As \( u_{m-1} \) starts and ends with 00101100, and 1101 is not a factor of \( \kappa \), we conclude that \( p \) must be a central factor of \( u_m \). There are thus only two possibilities, \( 0^m \) or \( 10^{m-2}1 \). By direct inspection the reader can verify that \( \text{Pal}_\kappa(6) = \{0^6\}, \text{Pal}_\kappa(5) = \{0^5\}, \text{Pal}_\kappa(4) = \{0000, 0110\} \) and \( \text{Pal}_\kappa(3) = \{000, 0101, 0110\} \). Thus we have proved the following:

**Lemma 3.5.** There exists an infinite recurrent aperiodic binary word \( w \) which is not closed under reversal containing infinitely many palindromes such that \( \text{Pal}(w) \subset \text{Pri}(w) \). \( \square \)

Let us then consider the Chacon word \( \lambda \), the fixed point of the (non-primitive) morphism \( 0 \mapsto 0010, 1 \mapsto 1 \) [Fer95]. The word \( \lambda \) is aperiodic, and also uniformly recurrent as the letter 0 occurs in
bounded gaps. By a direct verification one can show that the word \( \lambda \) doesn’t contain palindromes of length 13 or 14. Therefore \( \text{Pal}_\lambda(n) = \emptyset \) for all \( n \geq 13 \). There are total 23 palindromes in \( \lambda \). Using the same brute-force approach one can show that all palindromes in \( \lambda \) are privileged. Note that the Chacon word is not closed under reversal: for instance \((100100)^\sim = 001001 \not\in \mathcal{J}(\lambda)\) as 001001 can’t be properly factored over the set \{0010, 1\}. We have:

**Lemma 3.6.** There exists an infinite uniformly recurrent aperiodic binary word \( w \) which is not closed under reversal containing finitely many palindromes such that \( \text{Pal}(w) \subset \text{Pri}(w) \).

Next we recall a construction of [Ber+09]. Consider the infinite word \( \mu = \lim_{n \to \infty} u_n \), the limit of the sequence \( u_0 = 01, u_{n+1} = u_n 23 \bar{u}_n \). The word \( \mu \) is uniformly recurrent, aperiodic, closed under reversal and contains only finitely many palindromes, namely only the letters 0, 1, 2 and 3. By applying the morphism

\[
\begin{align*}
0 & \mapsto 101, \\
1 & \mapsto 1001, \\
2 & \mapsto 10001, \\
3 & \mapsto 100001
\end{align*}
\]

we obtain a uniformly recurrent aperiodic binary word which is closed under reversal and contains only finitely many palindromes (longest is of length 12). By direct inspection it can be verified that each palindrome in \( \mu \) is privileged. Hence we have:

**Lemma 3.7.** There exists an infinite uniformly recurrent aperiodic binary word \( w \) which is closed under reversal containing finitely many palindromes such that \( \text{Pal}(w) \subset \text{Pri}(w) \).

However it turns out that if a uniformly recurrent word contains infinitely many palindromes, then the inclusion can’t be proper. Note that such a word is necessarily closed under reversal.

**Proposition 3.8.** Let \( w \) be a uniformly recurrent word containing infinitely many palindromes. If \( \text{Pal}(w) \subseteq \text{Pri}(w) \), then \( \text{Pal}(w) = \text{Pri}(w) \), that is, \( w \) is rich.

**Proof.** Assume on the contrary that \( \text{Pal}(w) \subseteq \text{Pri}(w) \) and that \( w \) is not rich. Then there exists a privileged factor \( u \) which is not a palindrome. Since \( w \) is uniformly recurrent \( u \) is a factor of some palindrome \( p \). Clearly \( u \) can’t be a central factor of \( p \). Thus there exists a central factor \( q \) of \( p \) which begins with \( u \) and ends with \( \bar{u} \) (or \( q \) begins with \( \bar{u} \) and ends with \( u \), but this case is symmetric). It is immediate that \( q \) is a palindrome. Thus by the assumption \( q \) is privileged. As \( q \) has as a prefix the privileged word \( u \), the prefix \( u \) also occurs as a suffix of \( q \). Hence \( u \) is a palindrome, a contradiction.

Note that in the proof uniform recurrence was only needed to establish that \( u \) is a factor of some palindrome. Thus it is necessary to only suppose that every privileged factor occurs in some palindrome to obtain the result.

It can be proven that the palindromic complexity function of a fixed point of a primitive morphism is bounded (see [DZ00] and [All+03]). In this respect palindromes and privileged words behave radically differently. It is not true that the privileged complexity function of a fixed point of a primitive morphism is necessarily bounded. In Proposition 4.22 in the next section we prove that the privileged complexity function of the Thue-Morse word is not bounded. Note that the Thue-Morse word has infinite defect.
4 The Privileged Complexity of the Thue-Morse Word

In this section we prove a recursive formula for the privileged complexity function of the Thue-Morse word. Moreover we study the asymptotic behavior of the function, and the occurrences of zeros in its values.

Let $t = 0110100110010110 \ldots$ be the infinite Thue-Morse word (see [AS]). The word $t$ is a fixed point of the morphism $\varphi$ and its square $\theta = \varphi^2$.

$$\varphi: \begin{array}{c} 0 \mapsto 01 \\ 1 \mapsto 10 \end{array} \quad \theta: \begin{array}{c} 0 \mapsto 0110 \\ 1 \mapsto 1001 \end{array}$$

The word $t$ has the following well-known property (an overlap is a factor of the form $auaua$ where $a$ is a letter):

**Theorem 4.1.** The Thue-Morse word $t$ doesn’t contain overlaps, i.e. it is overlap-free.

By this Theorem the longest privileged proper border of a privileged factor $w$ of $t$ can’t overlap with itself in $w$. In what follows, we implicitly assume this fact. Using overlap-freeness by mere inspection we obtain the following:

**Lemma 4.2.** Every factor of $t$ of length at least four admits a unique interpretation by $\varphi$. Every factor of $t$ of length at least seven admits a unique interpretation by $\theta$.

Using this Lemma we are able to prove the following important Proposition:

**Proposition 4.3.** Let $w, u \in \mathcal{F}(t)$ be such that $|w| \geq |u| \geq 2$. Then $|\theta(w)|_{\theta(u)} = |w|_u$.

**Proof.** Clearly always $|\theta(w)|_{\theta(u)} \geq |w|_u$. Say $\theta(u)$ occurs in $\theta(w)$, so $\theta(w) = a\theta(u)\beta$ for some words $a$ and $\beta$. There must exist words $\lambda$ and $\mu$ such that $\theta(\lambda) = a$ and $\theta(\mu) = \beta$, since otherwise $\theta(u)$ would admit two interpretations by $\theta$, which is impossible by Lemma 4.2 as $|\theta(u)| \geq 8$. Hence $w = \lambda u \mu$. This proves that $|\theta(w)|_{\theta(u)} \leq |w|_u$.

We denote $X_\theta = \{0110, 1001\}$. A factorization of a factor $u$ over $X_\theta$ is a placement of bar lines with respect to the morphism $\theta$. We say that a word $w$ matches over $X_\theta$ if there exists a word $u$ such that $\theta(u) = w$.

The following interesting result was proved in [BPS].

**Proposition 4.4.** [BPS] Every factor in the Thue-Morse word has 3 or 4 complete returns.
for \( n > 1 \). Using overlap-freeness, we can easily see that \( \text{Pri}_0(1) = \{0\} \), \( \text{Pri}_0(2) = \{00\} \), \( \text{Pri}_0(3) = \{010\} \) and \( \text{Pri}_0(4) = \{0110\} \). Hence \( A(1) = A(2) = A(3) = A(4) = 2 \). Next we state the main result of this paper.

**Theorem 4.5.** The privileged complexity function \( A(n) \) of the Thue-Morse word satisfies

\[
A(0) = 1, A(1) = A(2) = A(3) = A(4) = 2, \\
\frac{1}{2} A(4n) = 3A_{00}(n) + A_{010}(n) + A_{010}(n + 1) + A_{0110}(n + 1) \quad \text{for } n \geq 2, \\
\frac{1}{2} A(4n - 2) = A_{00}(4(n - 1)) + A_{010}(4n) + A_{0110}(4n) \quad \text{for } n \geq 2, \\
A(2n + 1) = 0 \quad \text{for } n \geq 2.
\]

**Proof.** The claim follows from Corollaries 4.14, 4.17 and 4.20 which are proved below. \( \square \)

In the next table we give some values for \( A(n) \) for even \( n \) computed using the above formulas.

| 2-16 | 18-32 | 34-48 | 50-64 | 66-80 | 82-96 | 98-112 | 114-128 |
|------|-------|-------|-------|-------|-------|--------|--------|
| 2    | 2     | 14    | 0     | 2     | 0     | 16     | 0      |
| 2    | 2     | 6     | 0     | 2     | 0     | 8      | 0      |
| 4    | 4     | 4     | 0     | 2     | 4     | 4      | 6      |
| 8    | 8     | 8     | 0     | 2     | 12    | 4      | 18     |
| 8    | 8     | 8     | 0     | 2     | 12    | 4      | 18     |
| 4    | 4     | 4     | 0     | 2     | 4     | 4      | 6      |
| 0    | 6     | 2     | 0     | 2     | 4     | 4      | 8      |
| 0    | 14    | 2     | 0     | 2     | 12    | 4      | 24     |

It’s interesting to compare the privileged complexity with the palindromic complexity:

**Theorem 4.6.** [All+03], [BBL08] The palindromic complexity function \( P(n) \) of the Thue-Morse word satisfies

\[
P(0) = 1, P(1) = P(2) = P(3) = P(4) = 2, \\
P(4n) = P(4n - 2) = P(n) + P(n + 1) \quad \text{for } n \geq 2, \\
P(2n + 1) = 0 \quad \text{for } n \geq 2.
\]

\( \square \)

Next we list all complete first returns to 00, 010 and 0110. These words are needed later on. We leave it to the reader to verify that these actually are factors of 

| \( a_1 = 00101100 \), \( a_2 = 00110100 \), \( a_3 = 001100 \) and \( a_4 = 00\) | \( \beta_1 = 01011010, \beta_2 = 010110011010, \beta_3 = 010010 \) and \( \beta_4 = 0100110010 \). Complete first returns to 010 \( \gamma_1 = 0110011101010, \gamma_2 = 01101001011010, \gamma_3 = 01100101101010 \) and \( \gamma_4 = 01101001011010 \). By Proposition 4.4 these are all complete first returns to 00, 010 and 0110 (this fact is also easily verified directly). We see that we have at least two privileged factors beginning with 0 of odd length, namely 0 and 010. It turns out that there are no more:
Proposition 4.7. [Pel13] $\mathcal{A}(2n + 1) = 0$ for $n \geq 2$.

Proof. We may focus on privileged factors beginning with $0$. Let $w, |w| > 4$, be a privileged factor of $t$ beginning with $0$. Now $w$ begins with some of the three privileged words $00$, $010$ or $0110$. With respect to the morphism $\varphi$ the bar lines must be placed as follows: $0|0, 01|0 = 0|10$ and $01|10$. Now if $w$ begins with $00$ (respectively $0110$), then it also ends with $00$ (respectively $0110$). By the placement of the bar lines we immediately see that if $w$ begins with $00$ or $0110$, then it has even length. Assume then that $w$ begins with $010$. As $|w| > 4$, $w$ has as a prefix a complete first return to $010$, i.e. some of the words $\beta_1, \beta_2, \beta_3$ or $\beta_4$. The bar line placements of these words (with respect to $\varphi$) are $\beta_1 = 01|01|10|10, \beta_2 = 01|01|10|01|10|10, \beta_3 = 0|01|01|01|0$ and $\beta_4 = 0|10|01|10|01|0$. If $w$ begins with some $\beta_i$, then it also ends with $\beta_i$. From the placement of the bar lines we see that $|w|$ is necessarily even. \hfill \box

Next we characterize the different classes of privileged factors in the Thue-Morse word.

Lemma 4.8. Let $w \in \mathcal{P}ri_{00}(n)$ for some $n \geq 1$. Then

(i) $4 \mid |w|$ $\iff 1w110$ or $011w1$ is a factor of $t$ which matches over $X_\theta$ $\iff w$ begins with $a_1$ or $a_2$,

(ii) $4 \mid |w|$ $\iff 1w1$ is a factor of $t$ which matches over $X_\theta$ $\iff w$ begins with $a_3$ or $a_4$.

Proof. By Lemma 4.2 all factors of $t$ of length at least seven admit a unique interpretation by $\theta$, so in $a_1, a_2$ and $a_4$ there is a unique way to place bar lines: $a_1 = 001|0110|0, a_2 = 0|0110|100$ and $a_4 = 001|0110|100$. For $a_3$ there are potentially two ways to place bar lines: $a_3 = 001|0100$ and $a_3 = 0|0110|00$. However the latter is not possible as $(0110)^3$ is not a factor of $t$.

(i) Assume that $4 \mid |w|$. If $w$ would begin with $a_4$, then it would also end in $a_4$. From the placement of the bar lines it can be seen that this isn’t possible: it would follow that $4 \mid |w|$. Similarly $w$ can’t begin with $a_3$. By the placement of the bar lines we have that $1w110$ or $011w1$ are factors of $t$ and match over $X_\theta$. On the other hand if $w$ begins with $a_1$ or $a_2$, then $1w110$ or $011w1$ are factors of $t$ and have to match over $t$. Then clearly $4 \mid |w|$.

(ii) Assume that $4 \mid |w|$. By (i) $w$ has to begin with $a_3$ or $a_4$. In either case $1w1 \in \mathcal{F}(t)$, and $1w1$ matches over $X_\theta$. The other direction is also clear: if $w$ begins with $a_3$ or $a_4$, then by (i) it must be that $4 \mid |w|$.

Lemma 4.9. Let $w \in \mathcal{P}ri_{010}(n)$ for some $n \geq 1$. Then

(i) $4 \mid |w|$ $\iff 10w01$ is a factor of $t$ which matches over $X_\theta$ $\iff w$ begins with $\beta_1$ or $\beta_2$,

(ii) $4 \mid |w|, 2 \mid |w|$ $\iff 011w110$ is a factor of $t$ which matches over $X_\theta$ $\iff w$ begins with $\beta_3$ or $\beta_4$,

(iii) $4 \mid |w|, 2 \mid |w|$ $\iff w = 010$.

Proof. From Proposition 4.7 it follows that (iii) holds.

As in the previous proof, we know the placements of the bar lines in $\beta_1, \beta_2, \beta_3$ and $\beta_4$: $\beta_1 = 01|0110|10, \beta_2 = 01|0110|01|10|10, \beta_3 = 0|1001|0$ and $\beta_4 = 0|1001|1001|0$.

(i) Assume that $4 \mid |w|$. As in the previous proof, from the placement of the bar lines we see that $w$ can’t begin with $\beta_3$ or $\beta_4$, and hence it must start with $\beta_1$ or $\beta_2$. Then $10w01$ is a factor if $t$ which matches over $X_\theta$. Again the unique placement of the bar lines implies that the converse is also true.
Lemma 4.10. Let \( w \in \mathcal{Pri}_{0110}(n) \) for some \( n \geq 1 \). Then

(i) \( 4 \mid |w| \iff w \text{ matches over } X_8 \iff w \text{ begins with } \gamma_1 \text{ or } \gamma_2. \)

(ii) \( 4 \mid |w| \iff 10w \text{ or } w01 \text{ is a factor of } t \text{ which matches over } X_8 \iff w \text{ begins with } \gamma_3 \text{ or } \gamma_4. \)

Proof. The placements of the bar lines is known: \( \gamma_1 = 0110|0110, \gamma_2 = 0110|1001|0110, \gamma_3 = 01|1001|0110 \) and \( \gamma_4 = 0110|1001|10. \) As in the two previous proofs, by looking at the placements of the bar lines, the claim straightforwardly follows.

Combining the results of the three previous Lemmas, we get the following:

Corollary 4.11. Let \( w \in \mathcal{Pri}(n) \) with \( n > 5 \), and \( u \) its longest privileged proper prefix. Then \( 4 \mid |w| \) if and only if \( 4 \mid |u|. \)

Let us then define the following functions \( f_1, g_1 : \{0,1\}^* \rightarrow \{0,1\}^* : f_1(w) = \partial(\theta(1w), 1, 3) \) and \( g_1(w) = \partial(\theta(w1), 3, 1). \)

Lemma 4.12. Let \( n \geq 2 \). The function \( f_1 \) is a bijection \( \mathcal{Pri}_{00}(n) \rightarrow \mathcal{Pri}_{a_1}(4n) \) and the function \( g_1 \) is a bijection \( \mathcal{Pri}_{00}(n) \rightarrow \mathcal{Pri}_{a_2}(4n). \)

Proof. We will first prove the claim for \( f_1 \). If \( n = 2 \), then \( \mathcal{Pri}_{00}(2) = \{00\} \) and \( \mathcal{Pri}_{a_1}(8) = \{a_1\}, \) so the claim indeed holds. The latter part of this proof shows that if \( \mathcal{Pri}_{a_1}(4n) \neq \emptyset \), then also \( \mathcal{Pri}_{00}(n) \neq \emptyset. \) Thus the claim holds also if \( n = 3, 4 \), as then \( \mathcal{Pri}_{00}(n) = \emptyset. \) Assume that \( n > 5. \)

Let \( v \in \mathcal{Pri}_{00}(n) \), and let \( v \) be its longest privileged proper prefix. Note that now \( |v| \geq 2. \) As \( v \) begins with 00 it follows by induction that \( f_1(v) \in \mathcal{Pri}_{a_1}(t). \) As \( v \) is always preceded by letter 1, \( w = vw'1v \), and thus

\[
\begin{align*}
\theta(w) &= \partial(\theta(1v)\theta(w')\theta(1v), 1, 3) = \partial(1v, 1, 3)110\theta(w')1\partial(1v, 1, 3) = f_1(v)110\theta(w')1f_1(v)
\end{align*}
\]

By Lemma 4.8 the factor \( f_1(v) \) is always preceded by 1 and followed by 110. Thus from the previous equality it follows that

\[
\theta(w) = \theta(v)\theta(w')\theta(1v).
\]

Now if \( f_1(v) \) would occur more than twice in \( f_1(w) \), then \( \theta(v) \) would occur more than twice in \( \theta(w) \), and by Proposition 4.3 it would follow that \( v \) would occur more than twice in \( w \) which is not possible. Hence \( f_1(w) \) is a complete first return to the privileged word \( f_1(v) \), and thus \( f_1(w) \in \mathcal{Pri}_{a_1}(4n). \)

Assume then that \( w \in \mathcal{Pri}_{a_1}(4n) \). By Lemma 4.8 there exists \( z \in f(n+1) \) such that \( \theta(z) = 1w110. \) Write \( u = \partial(z, 1, 0) \). Then \( f_1(w) = w \). Let \( v \) be the longest privileged proper prefix of \( w \). By Corollary 4.11 we have that \( 4 \mid |v|. \) Thus by induction there exists \( s \in \mathcal{Pri}_{00}(t) \) such that \( f_1(s) = v \). Thus \( f_1(u) = w = f_1(s) \cdots f_1(s) \), and so \( u \) begins and ends with \( s \). Now if \( s \) would occur more than twice in \( u \), then as \( s \) is always preceded by 1, \( f_1(s) = v \) would occur more than twice in \( w \) which is impossible. Thus \( u \) is a complete first return to \( s \), so \( u \in \mathcal{Pri}_{00}(n). \)

Now the fact that the claim for the function \( g_1 \) holds follows from the fact that \( f_1(w)\sim = g_1(\tilde{w}) \), and that \( \tilde{a}_1 = a_2. \)
Lemma 4.13. Let \( n \geq 1 \). The function \( f_2 : \text{Pri}_0(4n - 2) \to \text{Pri}_{1001}(4n) \), \( f_2(w) = 1w1 \) is a bijection.

Proof. If \( n = 1 \), then \( \text{Pri}_0(2) = \{00\} \) and \( \text{Pri}_{1001}(4) = \{1001\} \), so the claim holds. Assume that \( n \geq 2 \). Let \( w \in \text{Pri}_0(4n - 2) \). As the factor 00 is always preceded and followed by the letter 1, \( 1w1 \in f(t) \), and \( 1w1 \) begins and ends with 0101. Let \( v \) be the longest privileged proper prefix of \( w \). By Corollary 4.11 it holds that \( 4 \mid |v| \). Thus by induction the word \( 1v1 \) is privileged. The word \( 1w1 \) is a complete first return to \( 1v1 \), as otherwise \( w \) would contain more than two occurrences of \( v \). Thus \( 1w1 \in \text{Pri}_{1001}(4n) \).

Let then \( 1w1 \in \text{Pri}_{1001}(4n) \). Again by applying Corollary 4.11 to the longest privileged proper prefix of \( 1w1 \) we get that \( w \in \text{Pri}_0(4n - 2) \).

\[ \text{Corollary 4.14.} \quad \mathcal{A}_{00}(4n) = 2\mathcal{A}_{00}(n) \text{ and } \mathcal{A}_{00}(4n - 2) = \mathcal{A}_{0110}(4n) \text{ for all } n \geq 2. \]

Proof. As the ranges of the functions \( f_1 : \text{Pri}_0(n) \to \text{Pri}_{10}(n) \) and \( g_n : \text{Pri}_{00}(n) \to \text{Pri}_{10}(n) \) are disjoint, the claim follows since by Lemma 4.8 \( \text{Pri}_{10}(4n) = \text{Pri}_{10}(4n) \cup \text{Pri}_{10}(4n) \). The other equality follows directly from Lemma 4.13 as \( \mathcal{A}_{00}(m) = \mathcal{A}_{0110}(m) \) for all \( m \geq 4 \).

\[ \text{Lemma 4.15.} \quad \text{Let } n \geq 2. \text{ The function } f_3 : \text{Pri}_{101}(n + 1) \cup \text{Pri}_{1001}(n + 1) \to \text{Pri}_{1010}(4n), f_3(w) = \partial(\theta(w), 2, 2) \text{ is a bijection.} \]

Proof. If \( n = 2 \), then \( \text{Pri}_{101}(3) = \{101\}, \text{Pri}_{1001}(3) = \emptyset \) and \( \text{Pri}_{010}(8) = \{\beta_1\} \). If \( n = 3 \), then \( \text{Pri}_{101}(4) = \emptyset, \text{Pri}_{1001}(4) = \{1001\} \) and \( \text{Pri}_{0111}(12) = \{\beta_2\} \). We may assume that \( n \geq 4 \).

Let \( w \in \text{Pri}_{101}(n + 1) \cup \text{Pri}_{1001}(n + 1) \), and \( v \) its longest privileged proper prefix. By Corollary 4.11 and induction \( f_3(v) \in \text{Pri}_{1010}(t) \) (note that Lemma 4.9 has a symmetric version where roles of 0 and 1 are exchanged). As \( v \) is a prefix and a suffix of \( w, f_3(w) \) starts and ends with \( f_3(v) \). By Lemma 4.9 the word \( f_3(v) \) is always preceded by 01 and followed by 01. Thus if \( f_3(w) \) contained more than two occurrences of \( f_3(v) \), then Proposition 4.3 would imply that \( w \) contains more than two occurrences of \( v \) which would be a contradiction. We conclude that \( f_3(w) \in \text{Pri}_{1010}(4n) \).

Let then \( w \in \text{Pri}_{1010}(4n) \). By Lemma 4.9 there’s such a word \( u \) that \( f_3(u) = w \). Let \( v \) be the longest privileged proper prefix of \( w \). By Corollary 4.11 we may apply induction to obtain a word \( s \in \text{Pri}_{101}(t) \cup \text{Pri}_{1001}(t) \) such that \( f_3(s) = v \). We may write \( f_3(u) = w = f_3(s) \cdot f_3(s) \). By Lemma 4.9 \( f_3(s) \) is always preceded by 01 and followed by 01. Therefore \( u \) begins and ends with \( s \). Now if \( u \) would contain a third occurrence of \( s \), then \( w \) would contain a third occurrence of \( v \), which is not possible. Hence \( u \in \text{Pri}_{101}(n + 1) \cup \text{Pri}_{1001}(n + 1) \).

\[ \text{Lemma 4.16.} \quad \text{Let } n \geq 2. \text{ The function } f_4 : \text{Pri}_{101}(4n - 2) \to \text{Pri}_{1010}(4n), f_4(w) = 0w0 \text{ is a bijection.} \]

Proof. If \( n = 2 \), then \( \text{Pri}_{101}(6) = \{E(\beta_3)\} \) and \( \text{Pri}_{1010}(8) = \{\beta_1\} \). If \( n = 3 \), then \( \text{Pri}_{101}(10) = \{E(\beta_4)\} \), and \( \text{Pri}_{1010}(12) = \{\beta_2\} \). Thus it can be assumed that \( n \geq 4 \).

Let \( w \in \text{Pri}_{101}(4n - 2) \), and \( v \) its longest privileged proper prefix. By Corollary 4.11 and induction \( f_4(v) \in \text{Pri}_{1010}(t) \). By Lemma 4.9 the factor \( f_4(v) \) is always preceded and followed by letter 0. Thus it can be written that \( f_4(w) = f_4(v) \cdot f_4(v) \). If there was a third occurrence of \( f_4(v) \) in \( f_4(w) \), then in \( w \) there would be at least three occurrences of \( v \), which is false. Therefore \( f_4(w) \in \text{Pri}_{1010}(4n) \).

Let \( 0w0 \in \text{Pri}_{1010}(4n) \). Again by applying Corollary 4.11 to the longest privileged proper prefix of \( 0w0 \), we get that \( w \in \text{Pri}_{101}(4n - 2) \).
Corollary 4.17. \( A_{010}(4n) = A_{010}(n + 1) + A_{0110}(n + 1) \) and \( A_{010}(4n - 2) = A_{010}(4n) \) for all \( n \geq 2 \).

Proof. This follows directly from Lemmas 4.15 and 4.16 as \( A_{101}(n) = A_{010}(n) \) and \( A_{0011}(n) = A_{0110}(n) \) for all \( n \geq 0 \).

Lemma 4.18. Let \( n \geq 2 \). The function \( \theta : \text{Pri}_{00}(n) \cup \text{Pri}_{010}(n) \to \text{Pri}_{0110}(4n) \) is a bijection.

Proof. If \( n = 2 \), then \( \text{Pri}_{00}(2) = \{00\}, \text{Pri}_{010}(2) = \varnothing \) and \( \text{Pri}_{0110}(8) = \{\gamma_1\} \). If \( n = 3 \), then \( \text{Pri}_{00}(3) = \varnothing, \text{Pri}_{010}(3) = \{010\} \) and \( \text{Pri}_{0110}(12) = \{\gamma_2\} \). By the argument at the end of the proof if \( \text{Pri}_{0110}(4n) \neq \varnothing \), then \( \text{Pri}_{00}(n) \cup \text{Pri}_{010}(n) \neq \varnothing \). As \( \text{Pri}_{00}(n) \cup \text{Pri}_{010}(n) = \varnothing \) when \( n = 4 \), it can be further assumed that \( n > 5 \).

Let \( w \in \text{Pri}_{00}(n) \cup \text{Pri}_{010}(n) \), and \( v \) its longest privileged proper prefix. Now \( |v| \geq 2 \). Once again by Corollary 4.11 and induction \( \theta(v) \in \text{Pri}_{0110}(4n) \). By Proposition 4.3 the word \( \theta(w) \) must be a complete first return to \( \theta(v) \), that is, \( \theta(w) \in \text{Pri}_{0110}(4n) \).

Lemma 4.19. Let \( n \geq 2 \). The function \( f_4 : \text{Pri}_{11}(4n) \to \text{Pri}_{0110}(4n + 2), f_4(w) = 0w0 \) is a bijection.

Proof. If \( n = 2 \), then \( \text{Pri}_{11}(8) = \{E(\alpha_1), E(\alpha_2)\} \) and \( \text{Pri}_{0110}(10) = \{\gamma_3, \gamma_4\} \). The rest of the proof is by induction along the lines of the proof of Lemma 4.13.

Corollary 4.20. \( A_{0110}(4n) = A_{00}(n) + A_{010}(n) \) and \( A_{0110}(4n - 2) = A_{00}(4(n - 1)) \) for all \( n \geq 2 \).

Proof. The result follows from Lemmas 4.18 and 4.19 as \( A_{00}(n) = A_{11}(n) \) for all \( n \geq 0 \).

Corollaries 4.14, 4.17 and 4.20 together prove Theorem 4.5. In the following we characterize the primitive privileged factors:

Proposition 4.21. The only non-primitive privileged factors of \( t \) beginning with 0 are \( 00, \beta_3, \gamma_1 \) and \( \gamma_2^2 \).

Proof. Let \( w \) be a non-primitive privileged factor of \( t \) beginning with 0. Then clearly \( w = u^2 \) for some privileged factor \( u \). If \(|u| = 1\), then \( w = 00 \). If \(|u| > 1\), then \( u \) can’t begin with 0 as otherwise \( w \) would have \( 0^4 \) as a central factor. Hence \( u \) begins with 010 or 0110. If \(|u| = 3, 4\), then \( w = \beta_3, \gamma_1 \). We may assume that \(|u| > 5\). Then as \(|u| \) is even, we have that \( 4 \mid |w| \), so \( 4 \mid |u| \) by Corollary 4.11. By Lemma 4.9 if \( u \) begins with 010, then \( u \) begins with \( \beta_1 \) or \( \beta_2 \), and so \( w \) has \( \beta_1^2 \) or \( \beta_2^2 \) as a central factor. This is however impossible as neither \( \beta_1^2 \) nor \( \beta_2^2 \) is a factor of \( t \). Thus \( u \) must begin with 0110, and begin with \( \gamma_1 \) or \( \gamma_2 \). The word \( \gamma_1 \) is non-primitive, and thus \( \gamma_1^2 \notin F(t) \).

As \( \gamma_2 \) is primitive, and \( \gamma_2^2 \) is not, we may assume that \( u \) begins with \( \gamma_2 \) and that \( u \neq \gamma_2 \). Thus \( w \) has \( \gamma_2^2 = \theta(010)^2 \) as a central factor. By Lemma 4.18 \( w = \theta(v) \) where \( v \) is a privileged word beginning with 010. As 0102 must be preceded by 011 and followed by 110, we have that \( w \) has \( \theta(011010 \cdot 010110) = \theta^2(010) \) as a central factor. As 010 can’t be preceded and followed by the same letter, we have that \( v \) ends with \( \beta_1 \) and begins with \( \beta_2 \), or symmetrically \( v \) ends with \( \beta_2 \) and begins with \( \beta_1 \). Either case is impossible as \( v \) is privileged.

Next we study the asymptotic behavior of the function \( A \).

Proposition 4.22. \( \limsup_{n \to \infty} A(n) = \infty \) and \( \liminf_{n \to \infty} A(n) = 0 \).
Proof. The fact that the inferior limit is 0 already follows from Proposition 4.7. We will prove that when \( n \geq 6 \),

\[
\mathcal{A}(2^n) = \begin{cases} 
3 \cdot 2^{\frac{n}{2}(n-1)}, & \text{if } n \text{ is odd,} \\
0, & \text{if } n \text{ is even,}
\end{cases}
\]

which proves the claim.

As \( 2^{n-2} + 1 \) is odd, by Corollary 4.17 it holds for \( n > 3 \) that

\[
\mathcal{A}_{010}(2^n) = \mathcal{A}_{010}(2^{n-2} + 1) + \mathcal{A}_{0110}(2^{n-2} + 1) = 0.
\]

Thus

\[
\frac{1}{2} \mathcal{A}(2^n) = 3\mathcal{A}_{00}(2^{n-2}).
\]

Now \( \mathcal{A}_{00}(2) = 1 \) and \( \mathcal{A}_{00}(4) = 0 \). By Corollary 4.14 \( \mathcal{A}_{00}(2^n) = 2\mathcal{A}_{00}(2^{n-2}) \), so for all \( n \geq 1 \),

\[
\mathcal{A}_{00}(2^n) = \begin{cases} 
2^{\frac{n}{2}(n-1)}, & \text{if } n \text{ is odd,} \\
0, & \text{if } n \text{ is even,}
\end{cases}
\]

proving the desired equality. \( \square \)

Our next aim is to show that there exists arbitrarily long gaps of zeros in the values of \( \mathcal{A} \). For this we need several Lemmas.

Let us define an integer sequence \( \langle a_n \rangle \) as follows: \( a_1 = 14 \) and \( a_n = 4(a_{n-2}) + 2(-1)^n \) for \( n > 1 \). The first few terms of the sequence are 14, 50, 190, 754, 3006, . . . Note that \( a_n \) is always even, and not divisible by four.

**Lemma 4.23.** If \( n \) is even, then \( \mathcal{A}_{00}(a_n - 2) = \mathcal{A}_{010}(a_n - 2) = 0 \) and \( \mathcal{A}_{0110}(a_n - 2) = 1 \). If \( n > 1 \) is odd, then \( \mathcal{A}_{00}(a_n - 2) = \mathcal{A}_{0110}(a_n - 2) = 0 \) and \( \mathcal{A}_{010} = 1 \). Moreover if \( n > 1 \), then \( \mathcal{A}(a_n - 2) = 2 \).

**Proof.** Using the formulas of Corollaries 4.14, 4.17 and 4.20 it’s readily verified that the claim holds for \( n = 2, 3 \).

Let \( n \) be even, so \( a_n = 4(a_{n-1} - 2) + 2 \). Using induction, Proposition 4.7 and the formulas mentioned above we get that

\[
\begin{align*}
\mathcal{A}_{00}(a_n - 2) &= 2\mathcal{A}_{00}(a_{n-1} - 2) = 0, \\
\mathcal{A}_{010}(a_n - 2) &= \mathcal{A}_{010}(a_{n-1} - 1) + \mathcal{A}_{0110}(a_{n-1} - 1) = 0 \text{ and} \\
\mathcal{A}_{0110}(a_n - 2) &= \mathcal{A}_{00}(a_{n-1} - 2) + \mathcal{A}_{010}(a_{n-1} - 2) = \mathcal{A}_{010}(a_{n-1} - 2) = 1.
\end{align*}
\]

Let \( n > 1 \) be odd. Then \( a_n = 4(a_{n-1} - 2) - 2 \). Similarly as above

\[
\begin{align*}
\mathcal{A}_{00}(a_n - 2) &= 2\mathcal{A}_{00}(a_{n-1} - 3) = 0, \\
\mathcal{A}_{010}(a_n - 2) &= \mathcal{A}_{010}(a_{n-1} - 2) + \mathcal{A}_{0110}(a_{n-1} - 2) = \mathcal{A}_{0110}(a_{n-1} - 2) = 1 \text{ and} \\
\mathcal{A}_{0110}(a_n - 2) &= \mathcal{A}_{00}(a_{n-1} - 3) + \mathcal{A}_{010}(a_{n-1} - 3) = 0.
\end{align*}
\]

It clearly follows that \( \mathcal{A}(a_n - 2) = 2 \) for \( n > 1 \). \( \square \)
In the case that \( n = 1 \), using it can be verified that \( A_{00}(12) = 0 \) and \( A_{010}(12) = A_{0110}(12) = 1 \), so that \( A(12) = 4 \). Hence in particular \( A(a_n - 2) \neq 0 \) for all \( n > 1 \).

**Lemma 4.24.** \( A(2^n + 2) = 2 \) for all even \( n \).

*Proof.* We will prove that \( A_{00}(2^n + 2) = A_{0110}(2^n + 2) = 0 \) and \( A_{010}(2^n + 2) = 1 \). The claim follows from this. This claim is readily verified in the case than \( n = 0 \). Let then \( n \geq 2 \) be even. By the formulas of Corollaries 4.14, 4.17 and 4.20 we obtain by applying induction that

\[
\begin{align*}
A_{00}(2^n + 2) &= A_{0110}(2^n + 4) = A_{00}(2^{n-2} + 1) + A_{010}(2^{n-2} + 1) = 0, \\
A_{010}(2^n + 2) &= A_{010}(2^n + 4) = A_{010}(2^{n-2} + 2) + A_{0110}(2^{n-2} + 2) = 1 \text{ and} \\
A_{0110}(2^n + 2) &= A_{00}(2^n) = 0,
\end{align*}
\]

where the last equality was proven correct in the proof of Proposition 4.22. \( \square \)

Finally we have proven enough Lemmas in order to prove the following result:

**Proposition 4.25.** For all \( n \geq 1 \) if \( a_n - 1 \leq k \leq 2^{2(n+1)} + 1 \), then \( A(k) = 0 \). Also for all \( n \geq 1 \) \( A(a_n - 2) \neq 0 \) and \( A(2^{2n+1} + 2) \neq 0 \).

*Proof.* By inspection it can be verified that indeed if \( a_1 - 1 = 13 \leq k \leq 17 = 2^{2(2+1)} + 1 \), then \( A(k) = 0 \). Let \( n > 1 \). Assume that \( a_1 \leq k \leq 2^{2(n+1)} \). If \( k \) is odd, then \( A(k) = 0 \). Suppose that \( k \) is even. Assume that four divides \( k \). Then it follows that \( a_{n-1} - 1 \leq k/4 \leq 2^{2n} \), so by the induction hypothesis \( A(k/4) = 0 \). By Theorem 4.5

\[
\frac{1}{2}A(k) = 3A_{00}(k/4) + A_{010}(k/4) + A_{010}(k/4 + 1) + A_{0110}(k/4 + 1) = 0.
\]

Assume that four does not divide \( k \). Now \( a_{n-1} - 1 \leq \frac{k+2}{4} \leq 2^{2n} \) and \( a_{n-1} - 2 \leq \frac{k-2}{4} \leq 2^{2n} \). Now using the already familiar formulas we get

\[
\frac{1}{2}A(k) = A_{00}(k - 2) + A_{010}(k + 2) + A_{0110}(k + 2)
= 2A_{00}\left(\frac{k-2}{4}\right) + A_{010}\left(\frac{k+2}{4} + 1\right) + A_{0110}\left(\frac{k+2}{4} + 1\right) + A_{010}\left(\frac{k+2}{4}\right) + A_{010}\left(\frac{k+2}{4}\right)
= 2A_{010}\left(\frac{k-2}{4}\right),
\]

where the last equality follows from the induction hypothesis. Now if \( k \neq a_n \), then \( a_{n-1} - 1 \leq \frac{k-2}{4} \), so by the induction hypothesis \( A(k) = 0 \). If \( k = a_n \), then \( \frac{k-2}{4} = a_{n-1} - 2 \). Then however by Lemma 4.23 \( A_{00}(a_{n-1} - 2) = 0 \), so also \( A(k) = 0 \).

The claim now follows as \( a_n - 1 \) and \( 2^{2(n+1)} + 1 \) are odd. Earlier it was proved that \( A(a_n - 2), A(2^{2n+1} + 2) \neq 0 \). \( \square \)

Straightforwardly using induction it can be proved that for \( n \geq 3 \) it holds that \( a_n < 2^{2n+1} + 2^{2n} < 2^{2(n+1)} \), so in particular if for \( n \geq 3 \) it holds that \( 2^{2n+1} + 2^{2n} \leq k < 2^{2(n+1)} \), then \( A(k) = 0 \). This verifies the following result which was conjectured in [Pel13]:

**Corollary 4.26.** There exists arbitrarily long (but not infinite) gaps of zeros in the values of the privileged complexity function of the Thue-Morse word. \( \square \)
This Corollary raises a natural question: If the privileged complexity function of a word \( w \) contains arbitrarily large gaps of zeros, does it follow that \( \limsup_{n \to \infty} A(n) = \infty \)? It’s conceivable that the large gaps force large values of \( A \) between the gaps. On the other hand the gaps could occur so sparsely that \( A \) is still bounded. The author wasn’t able to answer this question. All examples so far of words with large gaps seem to have unbounded privileged complexity.

The privileged complexity function of the Thue-Morse word is complicated. Even though the Thue-Morse morphism has really nice properties, finding the recursive formula for the function is a long task. On the other hand without the nice properties of the morphism, the work may not have been possible at all. Indeed if the morphism wasn’t uniform, then it would have been harder to calculate the length of the privileged factors. Other crucial property of the morphism is its circularity: every image of a letter is uniquely determined by its first or last letter. The author thinks that it could be possible to obtain results on the privileged complexity of fixed points of primitive uniform circular morphisms other than Thue-Morse.

5 Privileged Palindrome Complexity

The privileged palindrome complexity function \( B(n) \) counts the number of factors of length \( n \) which are privileged and palindromic. In this section we will focus only on the privileged palindrome complexity of the Thue-Morse word. As in the previous section we denote \( M_u(n) = \text{Pri}_i(n) \cap \text{Pal}_t(n) \cap u \cdot \{0, 1\}^* \), and \( B_u(n) = |M_u(n)| \). Again it suffices to consider factors beginning with letter 0.

**Lemma 5.1.** Let \( w \in \text{Pal}(t) \). Then \( w \in \text{Pal}(4n) \) for some \( n \) if and only if \( w \) matches over \( X_\varphi = \{01, 10\} \).

**Proof.** The claim holds for all palindromes of length less or equal to four: 0, 1, 00, 11, 010, 101, 0110 and 1001.

Let \( w \in \text{Pal}(4n) \) be shortest such palindrome that it doesn’t match over \( X_\varphi \). Suppose first that \( w \) begins with 00. Then \( w = 001w100 \). Now \( 1w1 \in \text{Pal}(4(n - 1)) \), so by the minimality of \( |w| \) we have that \( 1w1 \) matches over \( X_\varphi \). As \( |1w1| \geq 4 \), by Lemma 4.2 it has a unique interpretation by \( \varphi \). Hence it would follow that 00 matches over \( X_\varphi \) which is absurd. Say \( w \) begins with 01 (the case that it begins with 10 is symmetric). It can be written that \( w = 01w10 \). As \( w \) doesn’t match over \( X_\varphi \), it follows that neither does \( w' \in \text{Pal}(4(n - 1)) \), which is a contradiction with the minimality of \( |w| \).

Suppose then that \( w \) matches over \( X_\varphi \). Then clearly \( |w| = n \) is even. Suppose moreover that \( 4 \nmid |w| \), and that \( |w| \) is minimal. It can be written that \( w = 01w10 \) (the symmetric case being \( w = 10w0'1 \)), so \( w' \in \text{Pal}(4(n - 1) - 2) \), and \( w' \) matches over \( X_\varphi \). This is a contradiction with the minimality of \( |w| \).

Let’s have a closer look at the following functions

\[
\begin{align*}
f_2 & : w \mapsto 1w1 \\
f_3 & : w \mapsto \theta(\theta(w), 2, 2) \\
f_4 & : w \mapsto 0w0
\end{align*}
\]

defined in Lemmas 4.13, 4.15 and 4.16. Obviously if \( \mu \) is any of the functions \( f_2, f_3 \) or \( \theta \), then \( \mu(w) \) is a palindrome if and only if \( w \) is a palindrome. Similar property holds also for \( f_3 \) when
restricted onto $M_{101}(n + 1) \cup M_{001}(n + 1)$ because by Lemma 4.9 for each word $w$ in the range of this function it holds that $w$ is always preceded by 10 and followed by 01. Thus we have that the following functions are bijections:

$$
\begin{align*}
&f_2 : M_{00}(4n - 2) \to M_{1001}(4n), w \mapsto 1w1, \\
&f_3 : M_{101}(n + 1) \cup M_{001}(n + 1) \to M_{010}(4n), w \mapsto \partial(\theta(w), 2, 2), \\
&f_4 : M_{101}(4n - 2) \to M_{010}(4n), w \mapsto 0w0, \\
&f_4 : M_{11}(4n) \to M_{0110}(4n + 2), w \mapsto 0w0, \\
&\theta : M_{00}(n) \cup M_{010}(n) \to M_{0110}(4n).
\end{align*}
$$

We have thus proved the following formulas:

$$
\begin{align*}
B_{00}(4n - 2) &= B_{0110}(4n), \\
B_{1010}(4n) &= B_{00}(n + 1) + B_{0110}(n + 1), \\
B_{010}(4n - 2) &= B_{00}(4n), \\
B_{0110}(4n - 2) &= B_{00}(4(n - 1)), \\
B_{0110}(4n) &= B_{00}(n) + B_{010}(n),
\end{align*}
$$

for $n \geq 2$. We are still missing a formula for $B_{00}(4n)$. However by Lemma 5.1 we have that $M_{00}(4n) = \emptyset$, and thus $B_{00}(4n) = 0$. By putting together these formulas we get the following result:

**Theorem 5.2.** The privileged palindrome complexity function $B(n)$ of the Thue-Morse word satisfies

\[
B(0) = 1, B(1) = B(2) = B(3) = B(4) = 2,
\]

\[
\frac{1}{2}B(4n) = B_{00}(n) + B_{010}(n) + B_{010}(n + 1) + B_{0110}(n + 1) \quad \text{for } n \geq 2,
\]

\[
B(4n - 2) = B(4n) \quad \text{for } n \geq 2,
\]

\[
B(2n + 1) = 0 \quad \text{for } n \geq 2.
\]

As in the previous section for the function $A$, we study next the asymptotic behavior of the function $B$ and the occurrences of zeros in its values.

Let us define an integer sequence $(b_n)$ as follows: $b_1 = 6$ and $b_n = 4b_{n-1} - 2$ for $n > 1$. The first few terms of the sequence are 6, 22, 86, 342, 1366, . . . Note that $b_n$ is always even, and not divisible by four.

**Lemma 5.3.** $B(b_n) = 4$ for all $n \geq 1$.

**Proof.** By inspection $B_0(6) = 1, B_{1010}(6) = 1$ and $B_{0110}(6) = 0$, so $B(6) = 4$. We will prove that $B_0(b_n) = 2$ and $B_{010}(b_n) = B_{0110}(b_n) = 0$ for all $n > 1$. The claim follows from this. Now

$$
\begin{align*}
B_{00}(b_n) &= B_{0110}(b_n + 2) = B_{00}(b_{n-1}) + B_{0110}(b_{n-1}) \\
B_{010}(b_n) &= B_{010}(b_n + 2) = B_{010}(b_{n-1} + 1) + B_{0110}(b_{n-1} + 1), \\
B_{0110}(b_n) &= B_{00}(b_{n-2}) = B_{00}(b_{n-1} - 1).
\end{align*}
$$

So the claim is indeed true.
Proposition 5.4. The function $B(n)$ takes values in $\{0, 1, 2, 4\}$, and the values 0, 2 and 4 are attained infinitely often.

Proof. As $B(0) = 1, B(1) = B(2) = B(3) = B(4) = 2$, we need only to consider the values $B(4n)$. If $n = 2$, then $B(8) = 4$. Let $n > 4$ be even. Then by Theorem 5.2

$$\frac{1}{2} B(4n) = B_{00}(n) + B_{10}(n).$$

It suffices to prove that if $B_{00}(n) \neq 0$ then $B_{010}(n) = 0$. We are only interested in the case that $n$ is not divisible by four, as otherwise $B_{00}(n) = 0$. Now

$$B_{00}(n) = B_{0110}(n + 2) = B_{10} \left( \frac{n + 2}{4} \right) + B_{010} \left( \frac{n + 2}{4} \right) \text{ and}$$

$$B_{010}(n) = B_{010}(n + 2) = B_{010} \left( \frac{n + 2}{4} + 1 \right) + B_{010} \left( \frac{n + 2}{4} + 1 \right).$$

Clearly if $B_{00}(n) \neq 0$, then $(n + 2)/4$ is even, and thus $B_{010}(n) = 0$.

Let $n$ be odd. Then

$$\frac{1}{2} B(4n) = B_{010}(n + 1) + B_{0110}(n + 1).$$

Again it suffices to prove that if $B_{010}(n) \neq 0$, then $B_{0110}(n) = 0$. As $B_{00}(n) = 0$, when $n$ is not divisible by four, we need to consider only the case where $n$ is divisible by four. Now

$$B_{010}(n) = B_{10} \left( \frac{n}{4} + 1 \right) + B_{010} \left( \frac{n}{4} + 1 \right) \text{ and}$$

$$B_{0110}(n) = B_{00} \left( \frac{n}{4} + 1 \right) + B_{010} \left( \frac{n}{4} \right).$$

Obviously if $B_{010}(n) \neq 0$, then $n/4$ is odd, and thus $B_{0110}(n) = 0$.

By Lemma 5.3 the function $B$ takes value 4 infinitely often. Moreover arguments of Lemmas 4.23 and 4.24 work if the function $A$ is replaced with the function $B$. Thus the value 2 is also attained infinitely often.

The function $B(n)$ has the same gaps of zeros as described in Proposition 4.25 because the arguments of Lemmas 4.23 and 4.24 work if the function $A$ is replaced with the function $B$.

References

[AS] J.-P. Allouche and J. Shallit. “The ubiquitous Prouhet-Thue-Morse sequence”. In: Sequences and their applications (Singapore, 1998). Springer Ser. Discrete Math. Theor. Comput. Sci. Springer, pp. 1–16.

[All+03] J.-P. Allouche et al. “Palindrome Complexity”. In: Theoretical Computer Science 292 (2003), pp. 9–31.

[BPS] L. Balková, E. Pelantová, and W. Steiner. “Return Words in the Thue-Morse and Other Sequences”. Preprint arXiv:math/0608603v1.

[Ber+09] J. Berstel et al. “Infinite Words without Palindrome”. Preprint arXiv:0903:2382. 2009.
[BBL08] A. Blondin-Massé, S. Brlek, and S. Labbé. “Palindromic Lacunas of the Thue-Morse Word”. In: Proceeedings of GASCom 2008, 6th International Conference on Random Generation of Combinatorial Structures. (Arezzo, Italia, 16–20 June 2008). 2008, pp. 53–67.

[DZ00] D. Damanik and D. Zare. “Palindrome Complexity Bounds for Primitive Substitution Sequences”. In: Discrete Mathematics 222 (2000), pp. 259–261.

[Fer95] S. Ferenczi. “Les transformations de Chacon: combinatoire, structure géométrique, lien avec les systèmes de complexité 2n+1”. In: Bull. Soc. Math. France 123 (1995), pp. 271–292.

[FPZ12] A. Frid, S. Puzynina, and L. Zamboni. “On Minimal Factorizations of Words as Products of Palindromes”. Preprint arXiv:1210.6179. 2012.

[Gle+09] A. Glen et al. “Palindromic Richness”. In: European Journal of Combinatorics 30 (2009), pp. 510–531.

[GMS13] D. Goč, H. Mousavi, and J. Shallit. “On the Number of Unbordered Factors”. In: Language and Automata Theory and Applications - 7th International Conference, LATA 2013, Bilbao, Spain, April 2-5, 2013. Proceedings. Ed. by A.-H. Dediu, C. Martin-Vide, and B. Truthe. Vol. 7810. Springer, 2013, pp. 299–310.

[KLS11] J. Kellendonk, D. Lenz, and J. Savinien. “A Characterization of Subshifts with Bounded Powers”. Preprint arXiv:1111:1609. 2011.

[Pel13] J. Peltonäki. “Introducing Privileged Words: Privileged Complexity of Sturmian Words”. In: Theoretical Computer Science (2013).