On the 2-abelian Complexity of Thue–Morse Subwords

Florian Greinecker

greinecker@math.tugraz.at

October 2, 2014

Abstract

We show that the 2-abelian complexity of the infinite Thue–Morse word is 2-regular, and other properties of the 2-abelian complexity, most notably that it is a concatenation of palindromes. We also show sharp bounds for the length of unique extensions of subwords of size $n$, occurring in the Thue–Morse word.

1 Introduction

The main result of this paper is the following theorem.

Theorem 1.1. The 2-abelian complexity of the Thue–Morse word $t$ is 2-regular.

The infinite Thue–Morse word $t = 01101001101011010011001100101101001011\cdots$ is defined as $t := \lim_{n \to \infty} m^n(0)$ where $m$ is the morphism

$$m: 0 \mapsto 01, 1 \mapsto 10.$$ 

Theorem 1.1 combines two concepts: $\ell$-abelian complexity and $k$-regular sequences.

The $\ell$-abelian complexity is a complexity measure, which was first introduced in 1981 by Karhumäki [6]. The $\ell$-abelian complexity $P_{w,\ell}(n)$ of an infinite word $w$ lies between the factor complexity $P_{w,\ell}^{(\infty)}(n)$ and the abelian complexity $P_{w,1}(n)$. The abelian complexity (with $\ell = 1$) is $(P_{t,1}(n))_{n \geq 0} = (1, 2, 3)^\omega$. The factor complexity of the Thue–Morse word $t$ is well known [4], it is

$$P_{t,\ell}^{(\infty)}(0) = 1, \quad P_{t,\ell}^{(\infty)}(1) = 2, \quad P_{t,\ell}^{(\infty)}(2) = 4,$$

$$P_{t,\ell}^{(\infty)}(n) = \begin{cases} 4n - 2 \cdot 2^m - 4, & \text{if } 2 \cdot 2^m < n \leq 3 \cdot 2^m; \\ 2n + 4 \cdot 2^m - 2, & \text{if } 3 \cdot 2^m < n \leq 4 \cdot 2^m. \end{cases}$$

Before we define $\ell$-abelian complexity we need some vocabulary. For a word $w = w_0w_1\cdots w_n$ the prefix of length $\ell$ is defined as $\text{pref}_\ell(w) := w_0\cdots w_{\ell-1}$ while the suffix of length $\ell$ is $\text{suffix}_\ell(w) := w_{n-\ell+1}\cdots w_n$.

*The author is supported by the Austrian Science Fund (FWF): W1230, Doctoral Program “Discrete Mathematics”.

1

arXiv:1404.3906v2 [math.CO] 1 Oct 2014
We write $|w|$ to denote the length of a word $w$. If $v$ is a subword of $w$ the number of occurrences of $v$ in $w$ is denoted by $|w|_v$. We write $\mathbb{N}_0$ for the natural numbers, including 0.

**Definition 1.2.** For an integer $\ell \geq 1$, two words $u, v \in A^*$, for some alphabet $A$, are $\ell$-abelian equivalent if

- $\text{pref}_{\ell-1}(u) = \text{pref}_{\ell-1}(v)$ and $\text{suff}_{\ell-1}(u) = \text{suff}_{\ell-1}(v)$, and
- for all $w \in A^*$ with $|w| = \ell$ the number of occurrences of $w$ in $u$ and $v$ is equal, i.e. $|w|_w = |v|_w$.

We then write $u \equiv_\ell v$.

It is easy to check that $\ell$-abelian equivalence is indeed an equivalence relation. The first part of the definition, where we fix the prefix and suffix, guarantees that two $\ell$-abelian equivalent words are also $(\ell - 1)$-abelian equivalent.

**Example 1.3.** Let us take two words $w = 001011$ and $v = 001101$. We see that $w \equiv_2 v$ since $|w|_{00} = 1$, $|w|_{01} = 2$, $|w|_{10} = 1$, $|w|_{11} = 1$ and we get the same values for $v$. Furthermore both words have the same prefix and suffix. On the other hand $w \not\equiv_3 v$ since $|001011|_{010} = 1$ and $|001101|_{010} = 0$. Also the suffixes differ, $11 \neq 01$.

Since $\equiv_\ell$ is an equivalence relation it is natural to count equivalence classes.

We are interested in the number of $2$-abelian equivalence classes for words of a given length:

$$\mathcal{P}_t^{(2)}(n) := \#(T_n/\equiv_2).$$

Usually $\mathcal{P}_t^{(\ell)}(n)$ denotes the number of $\ell$-abelian equivalence classes of factors of $w$ of length $n$, where $w$ is an infinite word. In rest of this paper, we will only consider the $2$-abelian complexity of the infinite Thue–Morse word $t$. Therefore we will use the notation $\mathcal{P}_n := \mathcal{P}_t^{(2)}(n)$.

This sequence starts with the numbers

$$(\mathcal{P}_n)_{n \geq 0} = 1, 2, 4, 6, 8, 6, 8, 10, 8, 6, 8, 10, 10, 10, 10, 8, 8, 6, 8, 10, 8, 10, 10, 8, 10, 12, 12,$$

$10, 12, 12, 10, 10, 10, 8, 8, 6, 8, 10, 10, 12, 12, 10, 10, 12, 12, 14, 12, 12, 12, . . . .$

**Definition 1.4.** We assign to every word $w$ its equivalence class. To denote the $2$-abelian equivalence class of a word $w$ we use a $6$-tuple.

$$\text{con}: T \to \mathbb{N}_0^4 \times \{0, 1\}^2, \ w \mapsto (|w|_{00}, |w|_{01}, |w|_{10}, |w|_{11}, w_0, w_n).$$

**Example 1.5.** We have $\text{con}(w) = (1, 2, 3, 1, 1, 0)$ for $w = 10011010$.

Karhumäki, Saarela and Zamboni showed in [7] that for $n \geq 1$, $m \geq 0$ we have

$$\mathcal{P}_t^{(2)}(n) = O(\log n), \quad \mathcal{P}_t^{(2)}((2 \cdot 4^m + 4)/3) = \Theta(m) \quad \text{and} \quad \mathcal{P}_t^{(2)}(2^m + 1) \leq 8.$$
Let us now look at the other concept from Theorem 1.1, the $k$-regular sequences. Allouche and Shallit introduced $k$-regular sequences in 1990 [2] (and later wrote a sequel [3]). It is a well-known theorem by Eilenberg [5] that a sequence is $k$-automatic if and only if its $k$-kernel is finite.

**Definition 1.6.** Let $k \geq 2$ be an integer. The $k$-kernel of a sequence $(a(n))_{n \geq 0}$ is the set of subsequences

$$\{(a(k^e n + c))_{n \geq 0} \mid e \geq 0, \ 0 \leq c < k^e\}.$$

For example, the Thue-Morse sequence is 2-automatic. Allouche and Shallit took this characterization of $k$-automatic sequences via the kernel and extended it to $k$-regular sequences.

**Definition 1.7** (Allouche and Shallit). Let $k \geq 2$ be an integer. An integer sequence $(a(n))_{n \geq 0}$ is $k$-regular if the $\mathbb{Z}$-module generated by the $k$-kernel is finitely generated.

This definition turned out to be very useful and several papers have been written about $k$-regular sequences.

Recently, research has been conducted to investigate the regularity of the abelian complexity. Madill and Rampersad showed that the abelian complexity of the paperfolding word is 2-regular [8].

This article solves an open conjecture from Elise Vandomme, Aline Parreau and Michel Rigo [9]. Shortly after the discovery of our proof, they found one of their own [10], which uses the palindromic structure of the sequence.

This paper is organized as follows:

After some basic definitions in the rest of this section we will introduce **reading frames** in Section 2. Reading frames are a factorization of words into subsequences $v_i$ of the form $v_i = m^q(0)$ or $v_i = m^q(1)$ for some $q \in \mathbb{N}$, plus a prefix and suffix. Reading frames are a natural way to think about the Thue–Morse word since they preserve the morphism structure.

We use these reading frames in Section 3 to prove a theorem (Theorem 3.3) about unique extensions of Thue–Morse subwords. For a subword $w$ in $t$ there is sometimes only one possibility for the next (or previous) letters $x_1 \cdots x_n$ so that $wx_1 \cdots x_n$ (or $x_1 \cdots x_n w$) is again a subword of $t$. We give lower and upper bounds for the lengths of such unique extensions. This section can be skipped if one is only interested in the proof of the 2-regularity.

In Definition 1.4 we need 6 values to describe the 2-abelian equivalence class of a subword $w$, 4 binary values and 2 integer values. To simplify this 6-tuple of values we introduce the **off-beat frame** in Section 4. The off-beat frame is a shifted reading frame, which does not preserve the morphism structure but allows us to use only 3 values to denote the 2-abelian equivalence class of a subword $w$, 2 binary values and 1 integer value. Only the integer value $p(w)$, the number of pairs in a subword $w$, is nontrivial to determine.

Beside the off-beat frame we also introduce a **short coding**. The short coding is a way to encode words in the off-beat frame so that their important values can be seen on the first view.

In Section 5 we use the properties of pure off-beat words to prove a recursion (Theorem 5.2) on two types of sets, where $\text{pairs}(n)$ is the set of all possible values of $p(w)$ for subwords $w$ of length $n$. Once we have the recursion we can
use it to determine \( P_t(n) \) for all \( n \) (Theorem 5.4). This two theorems are used in all further proofs.

Equipped with Theorem 5.2 and Theorem 5.4 we prove the main Theorem 1.1 in Section 6 by showing 13 relations. For each of the 13 relations the calculations are similar, but since we have to look at three cases for each of them a bit lengthy.

Finally we show some additional properties of \( P_t(n) \) in Section 7, most notably that \( P_t(n) \) is a concatenation of longer and longer palindromes. Again we use Theorem 5.2 and Theorem 5.4 to do this. We also show that \( P_t(n) \) is unbounded.

We will use the fact that \( t \) is overlap-free (ref. [1], Theorem 1.6.1). An overlap is a word of the form \( xwxwx \), where \( w \) is a word, possibly empty, and \( x \) is a single letter. The word \( t \) is also cubefree, i.e. it contains no word of the form \( www \) where \( w \) is a nonempty subword.

The set of all finite subwords of the Thue–Morse word will be denoted by \( T \), while \( T_n \) stands for the set of Thue–Morse subwords of length \( n \). As usual, we use \( \overline{w} \) to denote the bitwise negation of a word \( w \) and call \( \overline{w} \) the complement of \( w \).

Example 1.8. The pattern \( x\overline{x}x \) can stand for \( 01011 \) or \( 10100 \) depending on the assignment of \( x \). So does \( \overline{x}_\overline{x} \).

The complement of a pattern encodes the same words. If we want a pattern out of a word, we do this via the morphism

\[
\text{pat}: 0 \mapsto x, \quad 1 \mapsto \overline{x}.
\]

We say a word \( w \) and a pattern \( p \) are equal \((p = w)\) if \( p = w \) for one assignment of \( x \). We can also concatenate patterns in the same way as words.

Example 1.9. With \( p = x \) and \( q = \overline{x}_\overline{x} \) we have \( pq = x\overline{x}xx \) and \( \overline{q}p = x\overline{x}x\overline{x} \).

2 Reading frames

From its definition via the substitution \( m \) it is clear that the Thue–Morse word is composed of copies of its first \( 2^q \) letters, \( q \in \mathbb{N} \), and their complements. To denote the special role of these words we define \( f_{2^q} := \text{pat}(t_{01}\cdots t_{2^q-1}) \).

Example 2.1. We take \( q = 2 \) and get \( f_{2^2} = x\overline{x}_x \) which gives us

\[
t = 011010010110010110101 \cdots = 011\overline{0}_100\overline{1}_1001_1011_101\overline{1}_1011_1011 \cdots
\]

A property that will be very useful is that the word \( f_{2^q} \) has the image \( m(f_{2^q}) = f_{2^q+1} \) and the preimage \( m^{-1}(f_{2^q}) = f_{2^q-1} \).

Definition 2.2. A \( 2^q \)-reading frame of a word \( w \) is a factorization of \( w \) into words \( w = v_1 \cdots v_m \), where \( v_1, \cdots, v_m \) are words of length \( 2^q \) plus a prefix \( p \) and a suffix \( s \) with \( |p|, |s| < 2^q \), so that \( v_i \) is either \( v_i = f_{2^q} \) or \( v_i = \overline{f}_{2^q} \), the prefix \( p \) is either \( p = \text{pref}_{|p|}(f_{2^q}) \) or \( p = \text{pref}_{|p|}(\overline{f}_{2^q}) \) and the suffix \( s \) is either \( s = \text{pref}_{|s|}(f_{2^q}) \) or \( s = \text{pref}_{|s|}(\overline{f}_{2^q}) \).
We call $p$, $s$ and the $v_i$ frame words, especially the $v_i$ are called complete frame words. The 1-frame is the trivial frame. The number of $2^q$-reading frames is at most $2^q$. If there is only one $2^q$-reading frame it is called the extensible reading frame. If there is an extensible $2^q$-frame, we call the $2^{q-1}$-frame that we get by splitting every complete $2^q$-frame word into two complete $2^{q-1}$-frame words, extensible too. Every word $w$ has at least one extensible reading frame since the trivial frame is extensible. If we write down extensions, we use a gray font for filled up letters.

**Example 2.3.** The word $0101$ has two $2$-frames: $0\underline{1}\underline{0}11$ and $01\underline{1}0\underline{1}1$, but only one extensible $4$-frame: $0\underline{1}\underline{1}0\underline{1}1$. We can not extend the $0\underline{1}\underline{1}0\underline{1}1$ frame since we would get $0\underline{1}10\underline{1}11$ but it is cubefree. Therefore the extensible $2$-frame is $0\underline{1}\underline{1}0\underline{1}1$. And since every complete $4$-frame word has the form $xxx$ it can not be $01010$, so the extensible $4$-frame is $0\underline{1}\underline{1}0\underline{1}1$.

In some sense, which we will clarify in a moment, the extensible $2^q$-reading frame is the correct one.

**Lemma 2.4.** A word $w \in T$ has a $2^q$-reading frame for every $q \in \mathbb{N}_0$.

**Proof.** By definition $w$ occurs somewhere in $t$. But $t$ can be read in a $2^q$-reading frame for every $q$ therefore $w$ can be read in a $2^q$-reading frame too. ∎

**Corollary 2.5.** A word $w$ is not in $T$ if there exists an integer $q$, so that $w$ has no $2^q$-frame.

**Example 2.6.** The words $xxx$, $xx\underline{x}xx$ and $x\underline{x}xx$ are not in $T$ since $t$ is overlap-free. We can also prove it with frames. The words $xxx$ and $xx\underline{x}xx$ are not in $T$ since they have no $2$-frame. The word $x\underline{x}xx$ is not in $T$ since it has no $4$-frame.

An extension of a word $w \in T$ is a pair of words $(v_1, v_2)$ with $v_1, v_2 \in T$, $|v_1| + |v_2| > 0$ so that $v_1wv_2 \in T$. An extension $(v_1, v_2)$ is unique if for all pairs $(y_1, y_2) \in \{0, 1\}^* \times \{0, 1\}^*$ with $|v_1| = |y_1|, |v_2| = |y_2|$ and $(v_1, v_2) \neq (y_1, y_2)$ it follows $y_1w_{y_2} \notin T$.

Every word $w \in T$ has extensions of arbitrary length. One natural way to find such extensions is to fill up the prefix and suffix to complete frame words in the $2^q$-reading frame. The extensible $2^q$-reading frame can be extended to arbitrary length, while the other $2^q$-reading frames (and their $2^{q-1}$ subframes) give words which are not in $T$. In this sense the extensible reading frame is the “correct” one. Going back to the Example 2.3 we get the following.

**Example 2.7.** We have $101\underline{0}110 \in T$ while $1\underline{0}1\underline{0}110 \notin T$ since $101010$ has no $4$-frame.

### 3 Maximal extensible reading frames

There is a maximal extensible reading frame (abbreviated as MERF), since if $w$ is a subword of $f_{2^q}$ it can not determine a $2^{q+1}$-frame. For small cases it is easy to determine the maximal extensible reading frame by hand, while longer words can be reduced to the short cases as preimages under the morphism $m$. 

5
Table 1: MERFs for all nonempty words in $T$ up to length 4.

| Pattern | MERF   | Pattern | MERF   |
|---------|--------|---------|--------|
| $\mathbf{a}$ | 1-frame | $\mathbf{a} \mathbf{x} \mathbf{x}$ | 4-frame |
| $\mathbf{a} \mathbf{a}$ | 2-frame | $\mathbf{x} \mathbf{a} \mathbf{x} \mathbf{x}$ | 2-frame |
| $\mathbf{a} \mathbf{a} \mathbf{a}$ | 1-frame | $\mathbf{x} \mathbf{x} \mathbf{x} \mathbf{x}$ | 4-frame |
| $\mathbf{a} \mathbf{a} \mathbf{x}$ | 2-frame | $\mathbf{x} \mathbf{x} \mathbf{x} \mathbf{x}$ | 4-frame |
| $\mathbf{a} \mathbf{a} \mathbf{x} \mathbf{x}$ | 1-frame | $\mathbf{x} \mathbf{x} \mathbf{x} \mathbf{x} \mathbf{x}$ | 2-frame |
| $\mathbf{a} \mathbf{a} \mathbf{x} \mathbf{x} \mathbf{x}$ | 2-frame |

The extensible reading frame of a subword $v$ of $w$ also determines the extensible reading frame of $w$. Therefore every word of length at least 4 has an extensible 2-frame. We can now formulate an algorithm to determine the MERF of a word $w$.

The algorithm determines the extensible 2-frame of the word and fills up the prefix and the suffix of $w$ to complete frame words, then takes the preimage of the new word and repeats those steps till it reaches a word with no extensible 2-frame. In every step the frame size doubles and the algorithm will need $q$ steps if the MERF has size $2^q$.

In every step there will be at most two new letters before the word size is halved. So the words will get shorter in every step until they have a length of 4 or shorter. Since the algorithm terminates for all words in Table 1 it will terminate in general.

**Algorithm 1** Determines the MERF of a word $w \in T$ and fills the MERF

```plaintext
procedure: MERF($w$)
  $q \leftarrow 0$
  $w' \leftarrow w$
  while $w'$ has a nontrivial reading frame do
    $q \leftarrow q + 1$
    $w' \leftarrow \text{FillFrame}(w')$  \{Determines and fills the extensible 2-frame\}
    $w' \leftarrow m^{-1}(w')$
  end while
  return $m^q(w'), \text{ 2}^q \text{-frame}.$
```

To decide whether $w$ has a nontrivial reading frame we can use Table 1 as lookup table, since there are only 6 words (3 patterns) with a trivial reading frame. For $\text{FillFrame}(w')$ we use the same lookup table at the first 4 letters of $w'$ to determine the 2-frame, than we find the frame prefix and suffix and fill them up. The original word $w$ will occur only once as subword in $m^q(w')$.

**Example 3.1.** What is $MERF(0110010)$?
We start with $w = \overline{0110010}$ and $q = 0$. Then we enter the while loop and get

$q = 1,$ $\text{FillFrame}(w') = \overline{0110010} \overline{0110010}$, $m^{-1}(w') = \overline{0110010}$,
$q = 2,$ $\text{FillFrame}(w') = \overline{10101010}$, $m^{-1}(w') = \overline{10101010}$,
$q = 3,$ $\text{FillFrame}(w') = \overline{10101010}$, $m^{-1}(w') = \overline{10101010}$.
which gives $\text{MERF}(0110010): \texttt{0110011100101110}$ 8-frame.

As a consequence of the algorithm we have the following lemma:

**Lemma 3.2.** A word $w$ of length $|w| = 2^q + r$, $r < 2^q$, $q > 0$, has an extensible $2^{q-1}$-frame.

**Proof.** The output word of the algorithm is at least as long as the input word.

The while loop of the algorithm will only end if it reaches $w = xx$ or $w = xxx$. Hence the output word will have length $2 \cdot 2^q$ or $3 \cdot 2^q$ for $i \in \mathbb{N}_0$. But $3 \cdot 2^q \geq 2^q$ implies $i \geq q - 1$.

Equipped with the algorithm, we are ready to prove the main theorem of this section. As usual we define $a \text{ Mod } b := a - \lfloor \frac{a}{b} \rfloor b$.

**Theorem 3.3.** A word $w$ of length $n := |w| = 2^q + r$, $r < 2^q$ uniquely determines at least $\mu(n)$ and at most $\pi(n)$ letters where

$$
\mu(n) := \begin{cases} 
0 & \text{for } n = 1 \\
0 & \text{for } n = 2 \\
0 & \text{for } n = 3 \\
-n \text{ Mod } 2^{q-1} & \text{for } n > 3
\end{cases} \quad \text{and } \pi(n) := \begin{cases} 
0 & \text{for } n = 1 \\
2 & \text{for } n = 2 \\
1 & \text{for } n = 3 \\
2^{\lfloor \log_2(n-2) \rfloor + 2} - n & \text{for } n > 3.
\end{cases}
$$

**Proof.** Table 1 allows us to check the cases with $n \leq 3$. Then we take a look at the function $\mu(n)$. According to Lemma 3.2 the word $w$ has an extensible $2^{q-1}$-frame and therefore determines at least $2^{q-1} \cdot i - n$ letters for some $i \in \mathbb{N}_0$. But the smallest positive value of $2^{q-1} \cdot i - n$ is exactly $-n \text{ Mod } 2^{q-1}$. To show that it is actually possible to obtain this value for $r \leq 2^{q-1}$ and $r > 2^{q-1}$, take the first $n$ letters of $f_{2q}, \overline{f_{2q}}, f_{2q-1}$ and $f_{2q}, f_{2q}$, respectively.

To analyze the function $\pi(n)$ we insert the word $w$ in a $2^{q-1}$-frame, which exists according to Lemma 3.2. A word of length $2^q$ or $2^q + 1$ can determine 3 frame words in the extensible $2^{q-1}$ frame. So after $q - 1$ iterations of the while loop we have a word of length 3. We enter the while loop again, extend the word (in the best case) to length 4 and then map it via $m^{-1}$ to $xx$. So after $q$ iterations we determined $2 \cdot 2^q$ letters.

If the word $w$ has length $2^q + 2 \leq |w| \leq 2^q + 2^{q-1} + 1$ it can determine 4 frame words. So after $q - 1$ iterations we have a word of length 4 which (in the best case) has a 4-frame and gives therefore 2 further iterations before we end up in $xx$. Here we determined $2 \cdot 2^{q+1}$ letters.

If $2^q + 2^{q-1} + 2 \leq |w| \leq 2^{q+1} - 1$ the word $w$ can determine 5 frame words, so we have a word of length 5 after $q - 1$ iterations, extend it to length 6, map it to length 3 and (in the best case) extend it to length 4, before it is mapped to $xx$. Again we determined $2 \cdot 2^{q+1}$ letters.

In each of these cases we determined $2^{\lfloor \log_2(n-2) \rfloor + 2} - n$ new letters, but we always assumed a best case. What is left is to show that there is always a word $w$, with $|w| = 2^q + r$, $r < 2^q$, so that the best case occurs. The first $n$ letters of $\text{suffix}(f_{2q+1})f_{2q+1}f_{2q+1}f_{2q+1}f_{2q+1}f_{2q+1}$ form such a word.

One way to measure the quality of an unique extension is to look at the relative length $\frac{|w'|}{|w|}$, where $w$ is the input and $w'$ is the output of the algorithm. The relative length satisfies the inequality $1 \leq \frac{|w'|}{|w|} < 4$. 

7
We have $|w'| = 1$ for the words $v = f_{2^q} f_{2^q}$ of length $2 \cdot 2^q$ and $w = f_{2^q} f_{2^q} f_{2^q}$ of length $3 \cdot 2^q$.

For the upper limit we look at the words $w_q = \text{suff}_1(f_{2^q-1}) f_{2^q-1} f_{2^q-1}$. These words of length $2^q + 2$ have an unique extension $w'_q$ of length $4 \cdot 2^q$. Therefore we have $\lim_{q \to \infty} \frac{|w'_q|}{|w_q|} = \lim_{q \to \infty} \frac{4 \cdot 2^q}{2^q + 2} = 4$.

Example 3.4. The word $w = 011010101$ is a word of length 8 with no extension $\left( \frac{|w'|}{|w|} = 1 \right)$, while $v = 0_01101$ is a word of length 6 which can be extended to the word $v' = 10010110_01101001$, of length 16 $\left( \frac{|w'|}{|v|} = 8/3 \right)$.

4 The off-beat frame

In this section we will simplify $\text{con}(w)$ from Definition 1.4. From now on we call the extensible 2-reading frame also beat frame. We can get an other reading frame if we shift the beat frame one letter. This new reading fame is called off-beat frame.

Example 4.1. A word 01011 can be read in the beat frame $01_02_1$ or in the off-beat frame $01_02_1$.

While the only two complete frame words in the beat frame are 01 and 10 we have the four complete frame words 00, 01, 10 and 11 in the off-beat frame. We call the off-beat frame words 00 and 11 pairs. The easiest way to find the off-beat frame of a word is to look for pairs.

We define a short coding for off-beat frame words: $01_02_1 \mapsto D$, $00_02_1 \mapsto E$ and finally $0_0 0_1 0_1 0_1 \mapsto S$.

An off-beat word with no prefix and suffix in the off-beat frame is called pure off-beat word. The study of pure off-beat words will turn out to be crucial for the rest of the paper. There is no $S$ in the short coding of a pure off-beat word and all pure off-beat words have even length.

Example 4.2. The word $v = 1100$ is a pure off-beat word since it has the off-beat frame $v = 11_02_1$ and the short coding EE. On the other hand the word $w = 01001$ is not a pure off-beat word since the off-beat frame $w = 01_02_1$ has a single letter suffix and therefore the short coding DES.

If an off-beat frame word ends with one letter, the next one starts with another letter since $xx$ in the off-beat frame would be $xx$ in the beat frame which can not occur. This fact allows us to recover a word $w$ from its short coding if we know a single letter of $w$, and to recover $\text{pat}(w)$ from the short coding too.

Example 4.3. Take the word $w = 1001011$. It contains two pairs $1 01_02_1$ and has therefore the off-beat frame $w = 10_02_01_02_1$. The short coding SEDE gives the pattern $10_02_01_02_1$ since $t$ is cubefree. If we know $w_0$ we can recover $w$ from its short coding.

Thus we can switch between patterns in the off-beat frame and the short coding. We will use this in the following proofs.

Lemma 4.4. The off-beat frame has following properties:
• The sequence DD can not occur.
• The sequence DEED can not occur.
• The sequence EEEE can not occur.

Proof. • We showed in Example 2.3 that the word $xTxxT$ is in the beat frame.
• The word $xTxxxxT$ has no 4-frame.
• The word $xxxxxxxT$ has no 8-frame.

So at least every second letter is an E and at most $3/4$ of the letters are E. This gives an upper bound for the growth of $P_n$.

Lemma 4.5. The pairs 00 and 11 alternate in the off-beat frame.

Proof. As consequence of Lemma 4.4 two consecutive E have either the form EE or EDE which gives the patterns $xxTxxT$ and $xTxxxxT$.

Now we define a function $p: T \to \mathbb{N}_0$, $w \mapsto |w|_{00} + |w|_{11}$, which counts the pairs in a word $w$, and a function

$$r(w) := \begin{cases} 0, & \text{if } w_0w_1 \text{ is in the off-beat frame;} \\ 1, & \text{if } w_0w_1 \text{ is in in the beat frame;} \end{cases}$$

which determines the reading frame of $w$. A word $w$ is a pure off-beat word if $r(w) = 0$ and $|w|$ is even. We could also define $r(w)$ via the short coding: $r(w) = 1$ if the short coding of $w$ starts with S and $r(w) = 0$ otherwise.

With the two functions $p(w)$ and $r(w)$ we can collect all information necessary to determine the 2-abelian equivalence class of a word $w$ in a 3-tuple:

$$\text{tup}: T \to \mathbb{N}_0 \times \{0, 1\}^2, \ w \mapsto (|w|_{00}, p(w), r(w)).$$

Example 4.6. Let us look at Example 1.5 again. For the word $w = 10011010$ we have now $\text{tup}(w) = (1, 2, 1)$.

In the next theorem we show that we can recover $\text{con}(w)$ from $\text{tup}(w)$. We will use the XOR operator $\oplus$ and the Iverson bracket $[\square]$ which is 1 if the statement $\square$ is true and 0 otherwise.

Theorem 4.7. There is a function $h$ so that

$$h(\text{tup}(w)) = \text{con}(w).$$

For two words $v, w \in T$ with $v \neq w$, we have $h(\text{tup}(v)) = h(\text{tup}(w))$ if and only if $v_0 = w_0$, $p(v) = p(w)$ and $p(v)$ is even.

Proof. The basic idea is to use parity arguments. If we take $w$ and erase one letter from every pair, we get a sequence of length $|w| - p(w)$ which alternates between 0 and 1 and starts with $w_0$. Since the sequence has the same number of 01 and 10 as $w$, we can use it to determine $|w|_{01}$, $|w|_{10}$ and $w_n$.

The pairs form already an alternating sequence (Lemma 4.5), so we only need to identify the first pair. In the off-beat frame a word can start either with E = xx or DE = xTxx. In both cases the first pair is $w_0w_0$. 

9
In the beat frame a word starts with $\mathbf{SE} = x\overline{FF}$ or $\mathbf{SDE} = x\overline{F}F\overline{F}$. In both cases the first pair is $\overline{m}m\overline{m}$. This allows us to determine $|w|_{00}$, $|w|_{11}$. We can also give these values in an explicit form as

$$h: (w_0, p(w), r(w)) \mapsto (a, b, c, d, w_0, e)$$

with

$$a = \lfloor \frac{p(w)}{2} \rfloor + \lfloor p(w) \text{ odd} \rfloor[w_0 = 0]$$

$$b = \lfloor \frac{|w| - p(w)}{2} \rfloor + \lfloor |w| - p(w) \text{ even}\rfloor[w_0 = 0]$$

$$c = \lfloor \frac{|w| - p(w)}{2} \rfloor + \lfloor |w| - p(w) \text{ even}\rfloor[w_0 = 1]$$

$$d = \lceil \frac{p(w)}{2} \rceil + \lfloor p(w) \text{ odd}\rfloor[w_0 = 1]$$

$$e = \lfloor |w| - p(w) \text{ even} \rfloor \oplus w_0.$$ 

Let $v, w \in T$ with $v \neq w$ be two words that belong to the same equivalence class. Then $v_0 = w_0$ and $p(v) = p(w)$, so they can only differ in the reading frame with $r(v) \neq r(w)$. The reading frame determines the first pair in the alternating pair sequence. If $p(w)$ is even the numbers $|w|_{00}$ and $|w|_{11}$ do not depend on $r(w)$. So $h(tup(v)) = h(tup(w))$ for $v \neq w$ if and only if $v_0 = w_0$, $p(v) = p(w)$ and $p(w)$ is even. \hfill \Box

The idea of Theorem 4.7 is to gather more information in less memory. We need two boolean and four integer variables for $\text{con}(w)$ while $\text{tup}(w)$ uses only one integer and two boolean variables.

**Example 4.8.** We have $\text{con}(w) = \text{con}(v)$ for the two words $w = 011001$ and $v = 001011$ but $\text{tup}(w) \neq \text{tup}(v)$ since $r(w) = 1$ and $r(v) = 0$. So tup can distinguish more words than con.

With Theorem 4.7 we can determine $T_n$ if we know the possible values of $\text{tup}(w)$, $w \in T_n$. The boolean variables can assume all possible values, a word $w \in T_n$ can start either with 0 or 1 and can be in the beat frame or in the off-beat frame. The difficult part is to find the possible values of $p(w)$. We will obtain them in the next section.

## 5 On pairs

We are interested in $\text{pairs}(n) := \{p(w) \mid w \in T_n\}$. It will emerge that we need a second set $\text{PAIRS}(2n) := \{p(w) \mid r(w) = 0, w \in T_{2n}\}$. The value $\text{PAIRS}(n)$ is undefined for odd $n$. The elements of $\text{PAIRS}(2n)$ are the possible numbers of $E$ in pure off-beat words of length $2n$.

**Example 5.1.** Let us determine $\text{pairs}(6)$ and $\text{PAIRS}(6)$.

| Pattern | Coding | $p(w)$ |
|---------|--------|--------|
| $\overline{FF}FF\overline{F}$ | EDE | 2 |
| $\overline{FF}FF\overline{F}$ | EEE | 3 |
| $\overline{FF}FF\overline{F}$ | EED | 2 |
| $\overline{FF}FF\overline{F}$ | DED | 1 |

| Pattern | Coding | $p(w)$ |
|---------|--------|--------|
| $\overline{FF}FF\overline{F}$ | DEE | 2 |
| $\overline{FF}FF\overline{F}$ | SDES | 1 |
| $\overline{FF}FF\overline{F}$ | SEES | 2 |
| $\overline{FF}FF\overline{F}$ | SDES | 1 |
Theorem 5.2. For $n \geq 4$ the sets $\text{pairs}(n)$ and $\text{PAIRS}(n)$ fulfill the recursions

\[
\begin{align*}
\text{PAIRS}(2n) &= n - \text{pairs}(n + 1) \\
\text{pairs}(2n + 1) &= \text{PAIRS}(2n) \\
\text{pairs}(2n) &= \text{PAIRS}(2n) \cup \text{PAIRS}(2n - 2)
\end{align*}
\]

with $n - \text{pairs}(n + 1) := \{ n - x \mid x \in \text{pairs}(n + 1) \}$.

Proof. The proof works only for $n \geq 4$ since we distinguish between beat and off-beat frame. But words of length $n < 4$ may have no defined 2-frame.

Let $w \in T_{n+1}$. We have $|w|_{01} + |w|_{10} = n - p(w)$ since $w$ has $n$ subwords of length 2. The image $m(w)$ has length $2n + 2$ and is in the beat frame. We remove the prefix and the suffix to get a pure off-beat word $w'$ of length $2n$. A pair 00 in $w'$ corresponds to a 10 in $w$ and a pair 11 in $w'$ corresponds to a 01 in $w$. Thus $p(w') = n - p(w)$. Since the steps to get from $w$ to $w'$ are bijective, the two sets $\text{PAIRS}(2n)$ and $n - \text{pairs}(n + 1)$ are equal.

All words $w \in T_{2n+1}$ are of the form $w'S$ or $Sw'$ where $w'$ is a pure off-beat word of length $2n$. Since $p(w') = p(w'S) = p(Sw')$ the bijection $w' \mapsto w'S$ proves $\text{pairs}(2n + 1) = \text{PAIRS}(2n)$.

Every word in $w \in T_{2n}$ is either of the form $w'S$ or $Sw'$, where $w'$ is a pure off-beat word of length $2n$ and $w''$ is a pure off-beat word of length $2n - 2$. Again, adding and removing $S$ is a bijection which does not change the number of pairs. Therefore $\text{pairs}(2n) = \text{PAIRS}(2n) \cup \text{PAIRS}(2n - 2)$. \(\square\)

If we assume that the sets $\text{pairs}(n)$ and $\text{PAIRS}(n)$ are nonempty and their elements are nonnegative integers, we can use Theorem 5.2 as definition of $\text{pairs}(n)$ and $\text{PAIRS}(n)$ for $n < 4$. For example $\text{pairs}(1) = \text{PAIRS}(0) = 0 - \text{PAIRS}(1)$ so every element of $\text{pairs}(1)$ has an additive inverse in $\text{pairs}(1)$ but since the elements are nonnegative $\text{pairs}(1) = \{ 0 \}$. All small values can be found in Table 3.

We write $\{ a; b \}$ to denote the interval of all integers between $a$ and $b$, including both. The integer interval $\{ a; b \}$ has cardinality $\# \{ a; b \} = b - a + 1$. For a set $S$ we define $\#_2 S$ as $\#_2 S := \# \{ s \in S \mid s = 2k, k \in \mathbb{Z} \}$, the number of even elements in $S$.

Lemma 5.3. The sets $\text{pairs}(n)$ and $\text{PAIRS}(2n)$ are integer intervals.

Proof. This is true for $n < 10$ (cf. Table 3). All other values can be calculated using Theorem 5.2. In most cases it is obvious that integer intervals are mapped to integer intervals, we just have to show that $\text{PAIRS}(2n) \cup \text{PAIRS}(2n - 2)$ is an integer interval. But this is true, since we know from the definition of $\text{PAIRS}(2n)$ that the upper and lower limit of two consecutive sets can differ only by 1. \(\square\)
Lemma 5.4. For $n \geq 4$ the number of 2-abelian equivalence classes is given by

$$
\begin{align*}
\mathcal{P}_{2n+1} &= 2(2\#\text{PAIRS}(2n) - \#_2\text{PAIRS}(2n)) \\
\mathcal{P}_{2n} &= 2(\#\text{PAIRS}(2n) + \#\text{PAIRS}(2n - 2) - \\
&\quad \#_2(\text{PAIRS}(2n) \cap \text{PAIRS}(2n - 2))).
\end{align*}
$$

Proof. This is an immediate consequence of Theorem 4.7 and Theorem 5.2. First we look at $\mathcal{P}_{2n}$. We want to find all possible values of $\text{tup}(w)$ for $w \in T_{2n}$. We have $\#\text{PAIRS}(2n)$ possibilities to choose $p(w)$ in the off-beat frame and we have $\#\text{PAIRS}(2n - 2)$ possibilities to choose $p(w)$ in the beat frame. If there is an even value $p(w)$ in both frames, the off-beat frame and the beat frame give the same equivalence class so we subtract $\#_2(\text{PAIRS}(2n) \cap \text{PAIRS}(2n - 2))$. Finally we multiply by 2 since $\varepsilon_0$ can be 0 or 1.

We use the same argument for $\mathcal{P}_{2n+1}$ but $\text{PAIRS}(2n)$ is the set of numbers of possible pairs in both frames and therefore also the intersection. \hfill \Box

Now we look at consecutive sets, because for even numbers the recursion needs two sets. If we know $\text{pairs}(n)$ and $\text{pairs}(n + 1)$, we can use Theorem 5.2 to determine $\text{PAIRS}(2n - 2)$ and $\text{PAIRS}(2n)$ and thus $\text{pairs}(2n - 1)$, $\text{pairs}(2n)$ and $\text{pairs}(2n + 1)$. With Lemma 5.4 we can also determine $\mathcal{P}_{2n-1}$, $\mathcal{P}_{2n}$ and $\mathcal{P}_{2n+1}$.

It is clear from the definition of $\text{pairs}(n)$ that the sequence $\min \text{pairs}(n)$ is monotonically increasing in steps of 0 or 1:

$$\min \text{pairs}(n + 1) - \min \text{pairs}(n) \in \{0, 1\}.$$

This also holds for $\max \text{pairs}(n)$, $\min \text{PAIRS}(n)$ and $\max \text{PAIRS}(n)$. So if $\text{pairs}(n) = \{a; b\}$ we have four possibilities for $\text{pairs}(n + 1)$:

$$\{a; b\}, \; \{a + 1; b\}, \; \{a; b + 1\} \; \text{and} \; \{a + 1; b + 1\}.$$

Now we use Theorem 5.2 to make Table 2.

Example 5.5. Let us take a look at the first case in Table 2. We start with the two sets $\text{pairs}(n) = \{a; b\}$ and $\text{pairs}(n + 1) = \{a; b\}$. Now we use Equation (1) of Theorem 5.2 and get

$$\text{PAIRS}(2n) = n - \text{pairs}(n + 1) = n - \{a; b\} = \{n - b; n - a\}$$

and

$$\text{PAIRS}(2n - 2) = n - 1 - \text{pairs}(n) = n - 1 - \{a; b\} = \{n - b - 1; n - a - 1\}.$$

Then we use Theorem 5.2 Equation (2) to get

$$\text{pairs}(2n - 1) = \text{PAIRS}(2n - 2) + \text{pairs}(2n - 1) = \{n - b - 1; n - a - 1\}$$

and

$$\text{pairs}(2n + 1) = \text{PAIRS}(2n) = \{n - b; n - a\}.$$

Finally we use use Theorem 5.2 Equation (3) and get

$$\text{pairs}(2n) = \text{PAIRS}(2n - 2) \cup \text{PAIRS}(2n) = \{n - b - 1; n - a - 1\} \cup \{n - b; n - a\} = \{n - b - 1; n - a\}.$$

We continue in the same manner to get the complete Table 2.
Table 2: Behavior of integer intervals under the maps \( \text{pairs}(n+i) \) and \( \text{PAIRS}(n) \).

| Case | \( \text{pairs}(n+i) \) | \( \text{PAIRS}(2n+i) \) | \( \text{PAIRS}(4n+i) \) | \( \text{PAIRS}(4n+i) \) |
|------|-----------------|-----------------|-----------------|-----------------|
| I.   | \( \{n-b-1; n-a-1\} \) | \( \{n-a-1; n+b\} \) | \( \{n+a-1; n+b\} \) | \( \{n+a-1; n+b\} \) |
|     | \{a;b\}          | \{n-b; n-a\}    | \{n-b; n-a\}    | \{n-b; n-a\}    |
|     | \{a;b\}          | \{n-b-1; n-a-1\} | \{n-a-1; n+b\}  | \{n+a-1; n+b\}  |
|     | \{a+1;b\}       | \{n-b-1; n-a-1\} | \{n+a-1; n+b\}  | \{n+a-1; n+b\}  |
|     | \{a+1; b+1\}    | \{n-b-1; n-a-1\} | \{n+a-1; n+b\}  | \{n+a-1; n+b\}  |

Table 3: Small values of the functions

If we observe small values of our sequence \( \text{pairs}(n) \) and see that the IV. case does not occur. The column \( \text{PAIRS}(2n+i) \) of Table 2 shows that case IV. can also not occur as image of smaller values. Therefore we have just proved:

\( \text{PAIRS}(2n-2) \neq \text{PAIRS}(2n) \) and \( \text{pairs}(n+1) \neq 1 + \text{pairs}(n) \). \hspace{1cm} (4)

From Table 2 we also get:

**Lemma 5.6.** If we know \( \text{pairs}(n) \) for all \( n \in \{a; b\} \) we can determine \( \text{pairs}(n') \) and \( \mathcal{P}_{n'} \) for \( n' \in \{2a - 1; 2b - 1\} \) and \( \text{PAIRS}(n'') \) for \( n'' \in \{2a - 2; 2b - 2\} \).

**Remark.** An especially interesting case of Lemma 5.6 is to go from \( \text{pairs}(n) \), \( n \in \{2^{a-1} + 1; 2^a + 1\} \) to \( \text{pairs}(n') \), \( n' \in \{2^a + 1; 2^{a+1} + 1\} \). This is the idea behind thee proof of Theorem 7.3.

# The sequence \( \mathcal{P}_n \) is 2-regular

We will prove Theorem 1.1 by proving the following 13 relations which generate all sequences for the \( Z \)-kernel. So any sequence \( \mathcal{P}_{2n+r} \) is a linear combination of \( \mathcal{P}_{2n+1}, \mathcal{P}_{4n+2}, \mathcal{P}_{4n+3}, \mathcal{P}_{8n}, \mathcal{P}_{8n+2}, \mathcal{P}_{8n+3}, \mathcal{P}_{8n+6}, \mathcal{P}_{8n+7}, \mathcal{P}_{16n+3} \).
1. \( P_{4n+1} = P_{2n+1} \)
2. \( P_{8n+4} = P_{8n+3} + P_{4n+3} - P_{4n+2} \)
3. \( P_{16n} = P_{8n} \)
4. \( P_{16n+2} = P_{8n+2} \)
5. \( P_{16n+6} = -P_{16n+3} + P_{8n+3} + 3P_{8n+2} + P_{4n+3} - 2P_{4n+2} - P_{2n+1} \)
6. \( P_{16n+7} = -P_{16n+3} + P_{8n+3} + 3P_{8n+2} + 2P_{4n+3} - 3P_{4n+2} - P_{2n+1} \)
7. \( P_{16n+8} = P_{8n+2} + P_{4n+3} - P_{2n+1} \)
8. \( P_{16n+10} = P_{8n+2} + P_{4n+3} - P_{2n+1} \)
9. \( P_{16n+11} = -P_{16n+3} + 3P_{8n+2} + P_{4n+3} - 2P_{2n+1} \)
10. \( P_{16n+14} = P_{16n+3} + P_{8n+7} - P_{8n+3} - P_{8n+2} - P_{4n+3} + 3P_{4n+2} - P_{2n+1} \)
11. \( P_{16n+15} = P_{16n+3} + 2P_{8n+7} - 3P_{8n+6} - 2P_{8n+3} + 6P_{4n+2} - 3P_{2n+1} \)
12. \( P_{32n+3} = P_{8n+3} \)
13. \( P_{32n+19} = -P_{16n+3} + P_{8n+3} + 3P_{8n+2} + 2P_{4n+3} - 3P_{4n+2} - P_{2n+1} \)

These relations have been found by computer experiments by Parreau, Rigo and Vandomme. We will prove them one by one by a four step approach.

Table 4 will be used in the second and third step. It has been generated in the same manner as Table 2 starting with all three possibilities for the relative sizes of \( \text{pairs}(n+1) \) and \( \text{pairs}(n+2) \). (For Relation 3 we need a different table.)

**First** We move all terms of the relation to the right hand side and replace them using Lemma 5.4. Since all \( P_n \) are even (for \( n > 0 \)), as consequence of Lemma 5.4, we can divide the whole equation by 2.

**Example 6.1** (Relation 1). To make the steps clearer we will prove Relation 1 as example. To prove \( P_{4n+1} = P_{2n+1} \) we move all terms to the right side.

\[
0 = -P_{4n+1} + P_{2n+1}
\]

Now we Lemma 5.4 and get

\[
0 = -2(2\#\text{PAIRS}(4n) - \#\text{PAIRS}(4n)) + 2(2\#\text{PAIRS}(2n) - \#\text{PAIRS}(2n)).
\]

We can divide this equation by 2 to get

\[
0 = -(2\#\text{PAIRS}(4n) - \#\text{PAIRS}(4n)) + (2\#\text{PAIRS}(2n) - \#\text{PAIRS}(2n)).
\]

**Second** We use Table 4 to substitute the sets \#PAIRS\((n)\) whose cardinality we want to calculate (and not the sets \#PAIRS\((n)\)).

Since there are three cases we will now have three equations. We calculate the cardinalities on the right hand side in all three cases, simplify until we have a single integer and put the result into a triple \((\text{rhs}_1, \text{rhs}_2, \text{rhs}_3)\), where \( \text{rhs}_i \) is the integer on the right hand side in the \( i \)-th case.
Example 6.2 (Relation 1, continued). We start with

\[ 0 = -(2\# \text{PAIRS}(4n) - \# \text{PAIRS}(2n) + 2\# \text{PAIRS}(2n) - \# \text{PAIRS}(2n)). \]

We look at case I. of Table 4 and substitute the sets \( \# \text{PAIRS}(2n) = \{n-b; n-a\} \) and \( \# \text{PAIRS}(4n) = \{n+a; n+b\} \) to get

\[ 0 = -2\# \{n+a; n+b\} + 2\# \{n-b; n-a\} - \# \text{PAIRS}(2n). \]

Now we calculate the cardinalities. Since \( \# \{a; b\} = b - a + 1 \) we get

\[ 0 = -(b-a +1) + \# \text{PAIRS}(4n) + 2(b-a +1) - \# \text{PAIRS}(2n). \]

This can be simplified to

\[ 0 = \# \text{PAIRS}(4n) - \# \text{PAIRS}(2n) \]

so the integer in case I. is 0. The other two cases are identical and we get

\[ 0 = \# \text{PAIRS}(4n) - \# \text{PAIRS}(2n) + (0, 0, 0). \]

This is a shorthand for three (identical) equations. Most times two or even three cases will be identical.

Third We will use Table 4 again to substitute the sets \( \# \text{PAIRS}(n) \) in all three cases and intersect them if necessary.

Example 6.3 (Relation 1, continued). In all three cases we get

\[ 0 = \# \{n+a; n+b\} - \# \{n-b; n-a\}. \]

Fourth Now we deal with the number of even elements in integer intervals. If two intervals of the same size with different signs contain the same number of even elements they cancel. If the do not cancel (because their borders have the wrong parity) or have different sizes we split off Iverson brackets from the beginning or end of the larger set until both sets have the same parity of limits and the same cardinality

\[ \# \{a; b\} = \lceil a \text{ even} \rceil + \# \{a+1; b\}; \quad \# \{a; b\} = \# \{a; b-1\} + \lceil b \text{ even} \rceil. \]

To check two intervals of the same size with different signs contain the same number of even elements we use the following procedure, which we call normalization:

- Replace all even numbers by 0 and all odd numbers by 1.
- Change all “−” signs to “+” signs.
- If an \( a \) occurs in the upper limit of a set we change the upper and lower limit.

It is easy to see that parity of the interval borders does not change during the procedure, basically we calculate modulo 2. It does however change the size of the intervals so its important to use normalization only on intervals of the same size.
Example 6.4. We want to show that
\[ 0 = -\#_2\{5n + a + 1; 5n + b + 2\} + \#_2\{3n - b; 3n - a + 1\}. \]

Both sides have the same cardinality \( b - a + 2 \), therefore we can normalize them:
\[
-\#_2\{5n+a+1;5n+b+2\} + \#_2\{3n-b;3n-a+1\} = \\
= -\#_2\{n+a+1;n+b\} + \#_2\{n-b;n-a+1\} = \\
= -\#_2\{n+a+1;n+b\} + \#_2\{n+b;n+a+1\} = \\
= -\#_2\{n+a+1;n+b\} + \#_2\{n+a+1;n+b\} = 0.
\]

Example 6.5 (Relation 1, continued). In all three cases we have
\[ 0 = \#_2\{n+a;n+b\} - \#_2\{n-b;n-a\}. \]

Since the sets have the same cardinality so we normalize an get
\[ 0 = \#_2\{n+a;n+b\} - \#_2\{n-b;n-a\} = \\
= \#_2\{n+a;n+b\} - \#_2\{n+b;n+a\} = \\
= \#_2\{n+a;n+b\} - \#_2\{n+a;n+b\} 
\]

which is true.

The only non mechanical step in the whole procedure is step four. Sometimes we have to split intervals before we can normalize. This will be mentioned in the proof. After this four steps we will have an equation which is trivially true.

Proof of Theorem 1.1.

1. This is shown in the example.

2. We will show the first relation in more detail we want to prove
\[ 0 = -P_{8n+4} + P_{8n+3} + P_{4n+3} - P_{4n+2}. \]

Using Lemma 5.4 (and dividing by 2) this is equivalent to
\[
0 = -\#\text{PAIRS}(8n+4) - \#\text{PAIRS}(8n+2) + \#\text{PAIRS}(8n+4) \cap \text{PAIRS}(8n+2) + 2\#\text{PAIRS}(8n+2) - \#_2\text{PAIRS}(8n+2) + 2\#\text{PAIRS}(4n+2) - \#_2\text{PAIRS}(4n+2) - \#_2\text{PAIRS}(4n+2) - \#_2\text{PAIRS}(4n+2) \cap \text{PAIRS}(4n). \]

For the second step we substitute the intervals from Table 4. With the shorthand \( C := b - a + 1 \) we get in the first case (where \( \text{PAIRS}(n+1) = \{a; b\} \) and \( \text{PAIRS}(n + 2) = \{a; b\} \))
\[
0 = -2C - 2 + \#_2\text{PAIRS}(8n+4) \cap \text{PAIRS}(8n+2) + 2C + 2 - \#_2\text{PAIRS}(8n+4) + 2C + 2 - \#_2\text{PAIRS}(4n+2) - 2C - 1 + \#_2\text{PAIRS}(4n+2) \cap \text{PAIRS}(4n). \]

Most terms cancel and we have 1 remaining on the right hand side. It turns out this is also true in the two other cases so our triple is \((1, 1, 1)\) and we get
\[
0 = + \#_2\text{PAIRS}(8n+4) \cap \text{PAIRS}(8n+2) - \#_2\text{PAIRS}(8n+2) \\
- \#_2\text{PAIRS}(4n+2) + \#_2\text{PAIRS}(4n+2) \cap \text{PAIRS}(4n) + (1, 1, 1). \]
4. Similar to relation 3. but with values from Table 4. The triple is \(0, 3n-a+1\). We can get the values needed if we extend Table 2.

Now we apply the third step, build the intersection and get in all three cases

\[
0 = + \#_2\{3n-b+1, 3n-a+1\} - \#_2\{3n-b, 3n-a+1\} - \#_2\{n+a, n+b+1\} + \#_2\{n+a, n+b\} + 1.
\]

In the fourth step we can not normalize, so we split off Iverson brackets and get

\[
0 = + \#_2\{3n-b+1, 3n-a+1\} - [3n-b \text{ even}] - \#_2\{3n-b+1, 3n-a+1\} - \#_2\{n+a, n+b\} + 1.
\]

The intervals cancel and we are left with the true equation

\[
0 = -[3n-b \text{ even}] - [n+b+1 \text{ even}] + 1,
\]

in all three cases.

3. We can get the values needed if we extend Table 2.

| \(i\) | \(\text{PAIRS}(8n+i)\) | \(\text{PAIRS}(16n+i)\) |
|------|-----------------|-----------------|
| -2   | \(\{3n-b-1; 3n-a\}\) | \(\{5n+a-1; 5n+b\}\) |
| 0    | \(\{3n-b; 3n-a\}\)    | \(\{5n+a; 5n+b\}\)    |

The triple is \((0, 0, 0)\). The rest is a straightforward. In the fourth step we just have to normalize.

4. Similar to relation 3. but with values from Table 4. The triple is \((0, 0, 0)\) and we just have to normalize.

5. After the first two steps we get the triple \((0, 1, 0)\). Please note that the sets in cases I. and III. are identical, so we can ignore case III. In the first case we have

\[
0 = -3 \#_2\{3n-b-3n-a\} + 2 \#_2\{n+a; n+b\} + \#_2\{n-b; n-a\} + \#_2\{n+a; n+b+1\} + \#_2\{5n+a; 5n+b+1\} - \#_2\{3n-b; 3n-a+1\}
\]

where all terms in a single line normalize to zero.

In the second case we have

\[
0 = + \#_2\{5n+a+2; 5n+b+2\} + \#_2\{n-b; n-a\} - 2 \#_2\{3n-b; 3n-a\} - \#_2\{3n-b; 3n-a\} + 2 \#_2\{n+a+1; n+b\} - \#_2\{3n-b; 3n-a+1\} - \#_2\{n+a+1; n+b+1\} + \#_2\{5n+a; 5n+b+1\} + 1
\]
**Table 4: Values for the proof of Theorem 1.1.**

| Case | \(i\) | \(\text{pairs}(n+i)\) | \(\text{PAIRS}(2n+i)\) | \(\text{PAIRS}(4n+i)\) | \(\text{PAIRS}(8n+i)\) | \(\text{PAIRS}(16n+i)\) | \(\text{PAIRS}(32n+i)\) |
|------|--------|----------------|----------------|----------------|----------------|----------------|----------------|
| I.   | 0      | \{n-b; n-a\} | \{n+a; n+b\} | \{3n-b; 3n-a\} | \{5n+a; 5n+b\} | \{11n-b; 11n-a\} |
|      | 1 \{a; b\} | \{n-b+1; n-a+1\} | \{n+a; n+b+1\} | \{3n-b+1; 3n-a+1\} | \{5n+a+1; 5n+b+2\} | \{11n-b+1; 11n-a+2\} |
|      | 2 \{a; b\} | \{n-a+1\} | \{n+a+1; n+b+1\} | \{3n-b+1; 3n-a+2\} | \{5n+a+1; 5n+b+3\} | \{11n-b+1; 11n-a+3\} |
|      | 6 \{a; b\} | \{n-b+1; 3n-a\} | \{5n+a+2; 5n+b+3\} | \{11n-b+2; 11n-a+3\} |
| II.  | 8 \{a; b\} | \{n-b+3; 3n-a\} | \{5n+a+2; 5n+b+4\} | \{11n-b+2; 11n-a+4\} |
|      | 10   | \{5n+a+2; 5n+b+4\} | \{11n-b+3; 11n-a+4\} |
|      | 12   | \{5n+a+3; 5n+b+4\} | \{11n-b+4; 11n-a+5\} |
|      | 14   | \{5n+a+4; 5n+b+5\} | \{11n-b+4; 11n-a+6\} |
|      | 16   | \{5n+a+5; 5n+b+5\} | \{11n-b+5; 11n-a+6\} |
|      | 18   | \{11n-b+5; 11n-a+7\} |
| III. | 8 \{a; b\} | \{n-b+3; 3n-a\} | \{5n+a+3; 5n+b+3\} | \{11n-b+2; 11n-a+3\} |
|      | 10   | \{5n+a+3; 5n+b+4\} | \{11n-b+2; 11n-a+4\} |
|      | 12   | \{5n+a+4; 5n+b+4\} | \{11n-b+3; 11n-a+4\} |
|      | 14   | \{5n+a+5; 5n+b+5\} | \{11n-b+4; 11n-a+5\} |
|      | 16   | \{5n+a+5; 5n+b+5\} | \{11n-b+5; 11n-a+6\} |
|      | 18   | \{11n-b+5; 11n-a+6\} |
where all terms in the first line normalize to zero. We then split off Iverson brackets to get

\[
0 = - \#_2(3n-b;3n-a) + \#_2(5n+a;5n+b) + [5n+b+1 \text{ even}]
- \#_2(n+a+1; n+b) - [n+b+1 \text{ even}] + \#_2(n+a+1; n+b)
- \#_2(3n-b; 3n-a+1) + \#_2(n+a+1; n+b) + 1.
\]

The first two lines cancel. We split the first interval two times and get

\[
- \#_2(3n-b;3n-a-1) - [3n-a \text{ even}] - [3n-a+1 \text{ even}] + \#_2(n+a+1; n+b) + 1,
\]

and finally

\[
0 = -[3n-a \text{ even}] - [3n-a+1 \text{ even}] + 1.
\]

6. We already know that the Relation 5. is true so we subtract it from Relation 6. to get:

\[
\mathcal{P}_{16n+7} - \mathcal{P}_{16n+6} = \mathcal{P}_{4n+3} - \mathcal{P}_{4n+2}.
\]

Now we follow the usual procedure. The triple is \((0,0,0)\) and we have to split intervals before we can normalize to zero.

7. Straightforward. The triple is \((0,0,0)\) and we can normalize to zero.

8. We subtract Relation 7. from Relation 8. we get \(\mathcal{P}_{16n+10} - \mathcal{P}_{16n+8} = 0\). Again the triple is \((0,0,0)\) and we can normalize to zero.

9. If we subtract Relation 8. from Relation 9. we get:

\[
\mathcal{P}_{16n+11} - \mathcal{P}_{16n+10} = -\mathcal{P}_{16n+3} + 2\mathcal{P}_{8n+2} - \mathcal{P}_{2n+1}.
\]

Our triple is \((-1,-1,-1)\), we have to split and in all three cases we get

\[
0 = [5n+b+4 \text{ even}] + [5n+b+1 \text{ even}] - 1
\]
as final result.

10. The triple is \((0,-1,0)\). The rest of the calculation is lengthy, since for the first time all three cases are different, but not too hard. We have to split in the second and third case.

11. Another long calculation. Our triple is \((0,-2,-1)\). We have to split in the second and third case. We show as an example the second case:

\[
0 = -2 \#_2(3n-b;3n-a+2) + 2 \#_2(n-b; n-a) + 3 \#_2(3n-b;3n-a+1) - 3 \#_2(n+a+1; n+b)
+ 2 \#_2(3n-b;3n-a+1) - 3 \#_2(n+a+1; n+b) + \#_2(5n+a+5; 5n+b+5)
+ \#_2(n-b; n-a) - \#_2(5n+a; 5n+b+1) - 2
\]
The first two lines cancels.

\[
0 = \#_2\{5n+a+5;5n+b+5\} - \#_2\{5n+a+1;5n+b+1\} - [5n+a \text{ even}] \\
+ \#_2\{n-b;n-a-1\} + [n-a \text{ even}] - \#_2\{n+a+1;n+b\} \\
+ 2\#_2\{3n-b;3n-a+1\} - 2\#_2\{n+a+1;n+b\} - 2
\]

The first two lines cancel and after splitting the first interval in the last line two times we get the true equation

\[
0 = +2[3n-a+1 \text{ even}] + 2[3n-a \text{ even}] - 2.
\]

12. Straightforward. The triple is \((0,0,0)\) and we can normalize to zero.

13. We subtract Relation 6. from Relation 13. to get

\[
\mathcal{P}_{32n+19} - \mathcal{P}_{16n+7} = 0.
\]

The triple is \((0,0,0)\). We normalize to zero.

\[\square\]

### 7 Properties of \(\mathcal{P}_n\)

In this section we show three additional properties of \(\mathcal{P}_n\).

**Lemma 7.1.** For \(n \geq 4\) we have

\[
|\mathcal{P}_{n+1} - \mathcal{P}_n| \in \{-2,0,2\}.
\]

**Proof.** Due to Equation (4) we know that there are three possibilities for the relative sizes of \(\text{PAIRS}(2n-2)\) and \(\text{PAIRS}(2n)\):

1. \(\{a;b\}\)  
2. \(\{a;b\}\)  
3. \(\{a;b\}\)

\(\{a+1;b\}\) \(\{a;b+1\}\) \(\{a+1;b+1\}\)

Then we use Lemma 5.4 and with \(C := \#(\text{PAIRS}(2n) \cap \text{PAIRS}(2n-2))\) and \(E := \#_2(\text{PAIRS}(2n) \cap \text{PAIRS}(2n-2))\), we get

| Case | \(\mathcal{P}_{2n-1}\) | \(\mathcal{P}_{2n}\) | \(\mathcal{P}_{2n+1}\) |
|------|-----------------|-----------------|-----------------|
| 1.   | \(2(2C+1)-E-[a \text{ even}]\) | \(2((C+1)+C-E)\) | \(2(2C-E)\) |
| 2.   | \(2(2C-E)\) | \(2((C+1)+C-E)\) | \(2(2(C+1)-E-[b+1 \text{ even}])\) |
| 3.   | \(2(2(C+1)-E-[a \text{ even}])\) | \(2((C+1)+(C+1)-E)\) | \(2(2(C+1)-E-[b+1 \text{ even}])\) |

In all three cases, regardless if \(a\) and \(b+1\) are even or odd, Lemma 7.1 is true. \(\square\)

Now we want to prove that the sequence \(\mathcal{P}_n\) is unbounded.

**Lemma 7.2.** If \(\text{pairs}(n) = \{a;b\}\), \(\text{pairs}(n + 1) = \{a;b+1\}\) with \(n\) and \(b\) odd and \(a\) even and then the sequence \(a_0 = n\), \(a_{i+1} = 16a_i - 5\) satisfies

\[
\mathcal{P}_{a_n} = \mathcal{P}_{a_0} + 6n.
\]
Proof. If we start with two sets $\text{pairs}(n) = \{a; b\}$ and $\text{pairs}(n + 1) = \{a; b + 1\}$ then we have the two sets $\text{pairs}(16n - 5) = \{5n + a - 3; 5n + b - 1\}$ and $\text{pairs}(16n - 6) = \{5n + a - 3; 5n + b\}$ as consequence of Theorem 5.2.

So if we start in case III. with $\text{pairs}(n)$ and $\text{pairs}(n + 1)$ we will be again in case III. with $\text{pairs}(16n - 5)$ and $\text{pairs}(16n - 6)$. Therefore we can concatenate the whole process.

Furthermore, if $n$ and $b$ are odd and $a$ is even then $16n - 5$ and $5n + b - 1$ are odd and $5n + a - 3$ is even. Hence the sets $\#\text{pairs}(n) = \#\{a; b\}$ and $\#\text{pairs}(16n - 5) = \#\{a'; b'\}$ contain an even number of elements and we have:

$$\#\{a; b\} = 2s, \quad \#\{a'; b'\} = 2s + 2, \quad \#2\{a; b\} = s, \quad \#2\{a'; b'\} = s + 1.$$ Since the values $n$ and $16n - 5$ are both odd we know from Theorem 5.2 that $\text{PAIRS}(n - 1) = \text{PAIRS}(n)$ and $\text{PAIRS}(16n - 6) = \text{PAIRS}(16n - 5)$. Now we apply Lemma 5.4 and get $P_n = 6s$ and $P_{16n - 5} = 6s + 6$.

The sequence $P_3 = 6, P_{41} = 12, P_{683} = 18, \ldots$ is one example of such an unbounded sequence. In [7] they show $P_4(2)^{((2 \cdot 4^m + 4)/3)} = \Theta(m)$ which also proves that the sequence $P_n$ is unbounded.

The next theorem reveals a property of $P_n$.

| Case | $i$ | pairs$(n+i)$ | pairs$(2n+i)$ | $\#\text{pairs}(n+i)$ | $\#\text{pairs}(2n+i)$ |
|------|-----|--------------|---------------|----------------------|----------------------|
| I.   | $-1$ | $\{a; b\}$  | $\{a' + 1; b'\}$ | $\mathcal{C}$        | $\mathcal{C} + 1$   |
|      | $+1$ | $\{a; b\}$  | $\{a' + 1; b' + 1\}$ | $\mathcal{C}$        | $\mathcal{C}$       |
| II.  | $-1$ | $\{a + 1; b\}$ | $\{a' + 1; b'\}$ | $\mathcal{C} + 1$ | $\mathcal{C} + 1$   |
|      | $+1$ | $\{a; b\}$  | $\{a'; b\}$     | $\mathcal{C}$        | $\mathcal{C}$       |
| III. | $-1$ | $\{a; b\}$  | $\{a'; b'\}$    | $\mathcal{C}$        | $\mathcal{C}$       |
|      | $+1$ | $\{a; b + 1\}$ | $\{a'; b' + 1\}$ | $\mathcal{C} + 1$ | $\mathcal{C} + 1$   |

Table 5: The action of Theorem 5.2 on $\text{pairs}(n)$ and $\text{pairs}(n + 1)$

**Theorem 7.3.** The word

$$\mathcal{P}_{2^i+1} \mathcal{P}_{2^i+2} \mathcal{P}_{2^i+3} \cdots \mathcal{P}_{2^{i+1}+1}$$

is a palindrome, or equivalently

$$\mathcal{P}_{2^{i+1}+i} = \mathcal{P}_{2^{i+1}+1-i}$$

for $0 \leq i \leq 2^{q-1}$.

Proof. First we will show by induction that the sequence

$$\#\text{pairs}(2^q + 1), \#\text{pairs}(2^q + 2), \ldots, \#\text{pairs}(2^{q+1} + 1)$$

is a palindrome. The base case is

$$\#\text{pairs}(3) = 2, \#\text{pairs}(4) = 3, \#\text{pairs}(5) = 2.$$
Two sets \( \text{pairs} (2^q + 1 + i) \) and \( \text{pairs} (2^{q+1} + 1 - i) \) with \( 0 \leq i \leq 2^{q-1} \) are called corresponding sets. If a consecutive pair of sets is mapped to a consecutive triple of sets

\[
\begin{align*}
\text{pairs}(2^{q+1} + 2i) & \quad \text{pairs}(2^q + 2i) \\
\text{pairs}(2^q + i) & \quad \text{pairs}(2^{q+1} + 2i) \\
\text{pairs}(2^q + i + 1) & \quad \text{pairs}(2^{q+1} + 2i + 1)
\end{align*}
\]

with \( 1 \leq i \leq 3 \cdot 2^{q-1} \) then the corresponding pair of consecutive sets is mapped to a consecutive triple of sets

\[
\begin{align*}
\text{pairs}(2^{q+2} - 2i) & \quad \text{pairs}(2^{q+2} - 2i - 1) \\
\text{pairs}(2^{q+2} - 2i + 1) & \quad \text{pairs}(2^{q+2} - 2i + 2) \\
\text{pairs}(2^{q+2} - 2i + 2) & \quad \text{pairs}(2^{q+2} - 2i + 3)
\end{align*}
\]

Now we look at Table 5 and see that it is enough to know the relative sizes of consecutive sets do determine in which case we are. So if a consecutive pair is mapped to a consecutive triple via case II. the corresponding consecutive set pair is mapped to a corresponding consecutive set triple via case III. and vise versa. If we have a case I. map for the consecutive pair we also have a case I. map for the corresponding consecutive set pair. In all three cases the palindromic structure of the set cardinality is preserved.

Now we show that

\[
\#_2 \text{pairs}(2^q + 1), \#_2 \text{pairs}(2^q + 2), \ldots, \#_2 \text{pairs}(2^{q+1} + 1)
\]

is a palindrome too. We do this by showing that for two corresponding sets \( \text{pairs}(2^q + 1 + i) = \{a; b\} \) and \( \text{pairs}(2^{q+1} + 1 - i) = \{a'; b'\} \) we have

\[
a \equiv b' \pmod{2} \quad \text{and} \quad a' \equiv b \pmod{2}.
\]

Since two corresponding sets have the same cardinality we can conclude that

\[
\#_2 \{a; b\} = \#_2 \{a'; b'\}.
\]

Since \( P_3 = 6 \), Relation 1. tells us \( P_{2^q+1} = 6 \), so Equation (5) is true for our base case, the corresponding sets \( \text{pairs}(2^q + 1) \) and \( \text{pairs}(2^{q+1} + 1) \).

We use induction to go from \( \text{pairs}(2^q + 1 + i) \) and \( \text{pairs}(2^{q+1} + 1 - i) \) to \( \text{pairs}(2^{q+1} + 1 + (i + 1)) \) and \( \text{pairs}(2^{q+1} + 1 - (i + 1)) \). We have to check the three cases from Table 5 again:

In the first case nothing changes, we go from \( \{a; b\} \) and \( \{a'; b'\} \) to \( \{a; b\} \) and \( \{a'; b'\} \) and Equation 5 is trivially fulfilled.

In the second case we go from \( \{a; b\} \) and \( \{a'; b'\} \) to \( \{a + 1; b\} \) and \( \{a'; b' - 1\} \) and Equation 5 is fulfilled again.

In the third case the step is from \( \{a; b\} \) and \( \{a'; b'\} \) to \( \{a; b + 1\} \) and \( \{a' - 1; b'\} \) and Equation 5 holds.

Since

\[
\# \text{pairs}(2^q + 1 + i) = \# \text{pairs}(2^{q+1} + 1 - i)
\]

and also

\[
\#_2 \text{pairs}(2^q + 1 + i) = \#_2 \text{pairs}(2^{q+1} + 1 - i)
\]

with \( 0 \leq i \leq 2^{q-1} \) we can use Theorem 5.2 and to get

\[
\# \text{PAIRS}(2^q + 2i) = \# \text{PAIRS}(2^{q+1} - 2i)
\]
and also

\[ \#_2 \text{PAIRS}(2^q + 2i) = \#_2 \text{PAIRS}(2^{q+1} - 2i) \]

with \( 0 \leq i \leq 2^{q-2} \). Now we use Lemma 5.4 and get \( P_{2^q+1+i} = P_{2^{q+1}-i} \) with \( 0 \leq i \leq 2^{q-1} \).

References

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

[2] Jean-Paul Allouche and Jeffrey Shallit. “The ring of \( k \)-regular sequences”. In: 7. STACS 1990, Proceedings. Springer, 1990, pp. 12–23.

[3] Jean-Paul Allouche and Jeffrey Shallit. “The ring of \( k \)-regular sequences, II”. In: *Theoretical Computer Science* 307.1 (2003), pp. 3–29.

[4] Srečko Brlek. “Enumeration of factors in the Thue-Morse word”. In: *Discrete Applied Mathematics* 24.13 (1989), pp. 83 –96.

[5] S. Eilenberg. *Automata, languages, and machines*. Vol. A. Academic Press, New York and London, 1974.

[6] Juhani Karhumäki. “Generalized Parikh mappings and homomorphisms”. In: *Automata, Languages and Programming* 115 (1981), pp. 324–332.

[7] Juhani Karhumäki, Aleksi Saarela, and Luca Zamboni. *Variations of the Morse-Hedlund Theorem for \( k \)-Abelian Equivalence*. 2013. eprint: *arXiv*: 1302.3783.

[8] Blake Madill and Narad Rampersad. “The abelian complexity of the palindromic word”. In: *Discrete Mathematics* 313.7 (2013), pp. 831 –838.

[9] Aline Parreau, Michel Rigo, and Elise Vandomme. *A conjecture on the \( 2 \)-abelian complexity of the Thue-Morse word*. 2014. URL: *http://orbi.ulg.ac.be/handle/2268/162740*.

[10] Aline Parreau, Michel Rigo, Eric Rowland, and Elise Vandomme. *A new approach to the \( 2 \)-regularity of the \( \ell \)-abelian complexity of \( 2 \)-automatic sequences*. eprint: *arXiv*: 1405.3532.