

# Factor versus palindromic complexity of uniformly recurrent infinite words

Peter Baláži\*, Zuzana Masáková, Edita Pelantová

*Department of Mathematics, FNSPE, Czech Technical University, Trojanova 13, 120 00 Praha 2, Czech Republic*

---

## Abstract

We study the relation between the palindromic and factor complexity of infinite words. We show that for uniformly recurrent words one has  $\mathcal{P}(n) + \mathcal{P}(n + 1) \leq \Delta\mathcal{C}(n) + 2$ , for all  $n \in \mathbb{N}$ . For a large class of words this is a better estimate of the palindromic complexity in terms of the factor complexity than the one presented in [J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik, Palindrome complexity, Theoret. Comput. Sci. 292 (2003) 9–31]. We provide several examples of infinite words for which our estimate reaches its upper bound. In particular, we derive an explicit prescription for the palindromic complexity of infinite words coding  $r$ -interval exchange transformations. If the permutation  $\pi$  connected with the transformation is given by  $\pi(k) = r + 1 - k$  for all  $k$ , then there is exactly one palindrome of every even length, and exactly  $r$  palindromes of every odd length.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Palindrome; Infinite words; Interval exchange transformations; Palindromic complexity; Factor complexity; Uniformly recurrent words

---

## 1. Introduction

Recently, palindromes have become a popular subject of study in the field of combinatorics on infinite words. Recall that a palindrome is a word which remains unchanged if read backwards. In natural language it is for example the word “madam” in English, or “krk” (neck) in Czech. We shall study infinite words  $u$  over a finite alphabet  $\mathcal{A}$ , i.e. sequences  $u = (u_n)_{n \in \mathbb{N}}$  where  $u_i \in \mathcal{A}$  for all  $i \in \mathbb{N} = \{0, 1, 2, \dots\}$ . A palindrome of length  $n$  in the infinite word  $u$  is a factor  $p = u_i u_{i+1} \cdots u_{i+n-1}$  such that  $u_i u_{i+1} \cdots u_{i+n-1} = u_{i+n-1} u_{i+n-2} \cdots u_i$ .

The attractiveness of palindromes increased when Droubay and Pirillo provided yet another equivalent definition of Sturmian words using palindromes. They have shown in [12] that an infinite word  $u$  is Sturmian if and only if  $u$  contains exactly one palindrome of every even length and exactly two palindromes of every odd length.

A strong motivation for the study of palindromes in infinite words appeared already before, in their application in the modelling of solid materials with long-range order, the so-called quasicrystals. In 1982, Shechtman et al. [20] discovered an aperiodic structure (which was formed by rapidly-quenched aluminium alloys) that has icosahedral rotational symmetry, but no three-dimensional translational invariance (see e.g. [6]). The existence of

---

\* Corresponding author.

*E-mail address:* [peter\\_balazi@centrum.cz](mailto:peter_balazi@centrum.cz) (P. Baláži).

such structures had been absolutely unexpected. Since then, many other stable and unstable aperiodic structures with crystallographically forbidden rotational symmetry have been discovered; they were named quasicrystals.

Since the discovery of quasicrystals there has been an increasing interest in the study of the spectral properties of non-periodic Schrödinger operators. One can assign to an infinite word  $u$  over an alphabet  $\mathcal{A}$ , which models a one-dimensional quasicrystal, a Schrödinger operator  $H$  acting on the Hilbert space  $\ell^2(\mathbb{Z})$  as follows:

$$(H\phi)(n) = \phi(n + 1) + \phi(n - 1) + V(u_n)\phi(n),$$

where  $V : \mathcal{A} \mapsto \mathbb{R}$  is an injection and represents a potential of the operator.

Many nice properties of these operators have been shown, and they are well understood at least in the one-dimensional case. The survey papers [9,19] map the history of this effort. One of the main tasks is to derive the spectral properties of the Schrödinger operator  $H$  from the properties of the sequence  $V(u_n)$ . The physical motivation behind this study is that the spectral properties of operators determine the conductivity of the given structure. Very roughly speaking, if the spectrum is pure point then the structure behaves as an insulator. In case of absolutely continuous spectrum the material becomes a conductor.

Generally, the task of describing the spectral properties of the operator with potential given by an arbitrary infinite word  $u$  is not a simple one. The relevance of the study of palindromes in infinite words has been proven by Hof et al. [15], who showed that the operators given by words having arbitrary large palindromes have purely singular continuous spectrum.

The aim of this article is to find a relation between factor and palindromic complexities of uniformly recurrent words. Let us first introduce the basic notions which will be used in what follows.

The set of all factors of length  $n$  of an infinite word  $u = u_0u_1u_2 \dots$  is denoted by

$$\mathcal{L}_n(u) = \{w_1 \dots w_n \mid \exists i \in \mathbb{N}, w_1 \dots w_n = u_i \dots u_{i+n-1}\}.$$

The set of all factors of  $u$ , including the empty word  $\varepsilon$ , is called the language of  $u$  and is denoted by

$$\mathcal{L}(u) = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n(u).$$

The variability of local configurations in the word  $u$  is characterized by the factor complexity, the function  $\mathcal{C} : \mathbb{N} \rightarrow \mathbb{N}$ , given by the prescription

$$\mathcal{C}(n) := \#\mathcal{L}_n(u).$$

It is known that if there is an  $n_0$  such that  $\mathcal{C}(n_0) \leq n_0$ , then the word  $u$  is eventually periodic, i.e. there exists  $k \in \mathbb{N}$  such that  $u_{k+n} = u_n$  for every  $n \geq n_0$ . Any aperiodic (i.e. not eventually periodic) word therefore satisfies  $\mathcal{C}(n) \geq n + 1$  for every  $n \in \mathbb{N}$ . Aperiodic words of minimal complexity  $\mathcal{C}(n) = n + 1$  are called Sturmian words. For a survey of different characteristics and properties of Sturmian words see [5].

The mirror image, or reversal, of a finite word  $w = w_1 \dots w_n$  is the word  $\bar{w} = w_n \dots w_1$ . If the language  $\mathcal{L}(u)$  contains with every factor  $w$  also its mirror image  $\bar{w}$ , we say that  $\mathcal{L}(u)$  is closed (or invariant) under reversal.

The palindromic complexity of the infinite word  $u = (u_n)_{n \in \mathbb{N}}$  is a function  $\mathcal{P} : \mathbb{N} \rightarrow \mathbb{N}$  which counts the number of palindromes of each length. Formally,

$$\mathcal{P}(n) := \#\{w \in \mathcal{L}_n(u) \mid w = \bar{w}\}.$$

Trivially, one has  $\mathcal{P}(n) \leq \mathcal{C}(n)$ . A non-trivial result is an estimate of  $\mathcal{P}(n)$  using  $\mathcal{C}(n)$  provided in [1].

**Theorem 1.1** ([1]). *For arbitrary infinite word one has*

$$\mathcal{P}(n) \leq \frac{16}{n} \mathcal{C}\left(n + \left\lfloor \frac{n}{4} \right\rfloor\right), \quad \text{for all } n \in \mathbb{N}. \tag{1}$$

Let us mention that **Theorem 1.1** implies the result of [11]: The palindromic complexity of a fixed point of a primitive morphism is bounded.

In this paper we provide an estimate of  $\mathcal{P}(n)$  for uniformly recurrent words using the first difference  $\Delta\mathcal{C}(n) := \mathcal{C}(n + 1) - \mathcal{C}(n)$ . For words whose factor complexity is a polynomial of degree  $\leq 16$ , this estimate is better than that of (1). Let us recall that an infinite word  $u$  is uniformly recurrent if the gaps between consecutive occurrences of any

factor  $w \in \mathcal{L}(u)$  in the word  $u$  are bounded. Equivalently,  $u$  is uniformly recurrent if for every  $n \in \mathbb{N}$  there exists  $R(n) \in \mathbb{N}$  such that in an arbitrary segment of length  $R(n)$  in the word  $u$  one finds all factors of  $\mathcal{L}_n(u)$ , i.e.

$$\mathcal{L}_n(u) = \{u_i u_{i+1} \cdots u_{i+n-1} \mid k \leq i \leq k + R(n)\}, \quad \text{for all } k, n \in \mathbb{N}.$$

Let us mention that Sturmian words are example of uniformly recurrent words with language closed under reversal [12].

In Section 2 we show the following theorem.

**Theorem 1.2.** *Let  $u = (u_n)_{n \in \mathbb{N}}$  be an uniformly recurrent word.*

- (i) *If  $\mathcal{L}(u)$  is not closed under reversal, then  $\mathcal{P}(n) = 0$  for sufficiently large  $n$ .*
- (ii) *If  $\mathcal{L}(u)$  is closed under reversal, then*

$$\mathcal{P}(n) + \mathcal{P}(n + 1) \leq \Delta\mathcal{C}(n) + 2, \quad \text{for all } n \in \mathbb{N}.$$

It is interesting that equality in the latter estimate of the palindromic complexity holds for some known classes of infinite words, such as Arnoux–Rauzy words or fixed points of canonical substitutions associated to numeration systems with base  $\beta$ , where  $\beta$  is a Parry number [13]. We list some of these examples in Section 3. In Section 4 we show that the equality in the estimate is valid also for infinite words coding  $r$ -interval exchange transformations.

## 2. Proof of Theorem 1.2

First we show that unboundedness of the length of palindromes in an infinite uniformly recurrent word  $u$  implies that the language of  $u$  is invariant under mirror image.

**Lemma 2.1.** *Let  $u$  be an infinite word which is uniformly recurrent and such that  $\limsup_{n \rightarrow \infty} \mathcal{P}(n) > 0$ . Then  $\overline{\mathcal{L}(u)} = \mathcal{L}(u)$ .*

**Proof.** Let  $n \in \mathbb{N}$ . Consider  $R(n)$  from the definition of uniformly recurrent words. Let  $p$  be a palindrome of length greater than  $R(n)$ . It contains all factors of  $u$  of length  $n$ . In the same time it contains with every factor  $w$  also its mirror image. Thus  $\overline{\mathcal{L}_n(u)} = \mathcal{L}_n(u)$  for all  $n \in \mathbb{N}$ .  $\square$

The above lemma in fact proves (i) of Theorem 1.2. A crucial tool for the proof of (ii) is the notion of a Rauzy graph of an infinite word.

Let  $u = (u_n)_{n \in \mathbb{N}}$  be an infinite word,  $n \in \mathbb{N}$ . The Rauzy graph  $\Gamma_n$  of  $u$  is an directed graph whose set of vertices is  $\mathcal{L}_n(u)$  and set of edges is  $\mathcal{L}_{n+1}(u)$ . An edge  $e \in \mathcal{L}_{n+1}(u)$  starts at the vertex  $x$  and ends at the vertex  $y$ , if  $x$  is a prefix and  $y$  is a suffix of  $e$ .

If the word  $u$  is uniformly recurrent, the graph  $\Gamma_n$  is strongly connected for every  $n \in \mathbb{N}$ , i.e. there exists a directed path from every vertex  $x$  to every vertex  $y$  of the graph.

The outdegree of a vertex  $x \in \mathcal{L}_n(u)$  is the number of edges which start in  $x$ . It is denoted by  $\text{deg}_+(x)$ ,

$$\text{deg}_+(x) := \#\{a \in \mathcal{A} \mid xa \in \mathcal{L}_{n+1}(u)\}.$$

Similarly, we define the indegree of  $x$  as

$$\text{deg}_-(x) := \#\{a \in \mathcal{A} \mid ax \in \mathcal{L}_{n+1}(u)\}.$$

The sum of outdegrees over all vertices is equal to the number of edges in every directed graph. Similarly, it holds for indegree. In particular, for the Rauzy graph we have

$$\sum_{x \in \mathcal{L}_n(u)} \text{deg}_+(x) = \#\mathcal{L}_{n+1}(u) = \sum_{x \in \mathcal{L}_n(u)} \text{deg}_-(x).$$

Since  $\Delta\mathcal{C}(n) = \#\mathcal{L}_{n+1}(u) - \#\mathcal{L}_n(u)$ , we obtain

$$\Delta\mathcal{C}(n) = \sum_{x \in \mathcal{L}_{n+1}(u)} (\text{deg}_+(x) - 1) = \sum_{x \in \mathcal{L}_n(u)} (\text{deg}_-(x) - 1). \tag{2}$$

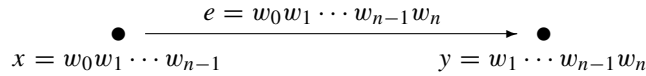


Fig. 1. Incidence relation between an edge and vertices in a Rauzy graph.

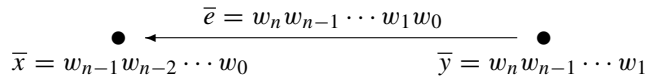


Fig. 2. The action of the mapping  $\rho$  on the edge and the vertices of Fig. 1.

A non-zero contribution to  $\Delta\mathcal{C}(n)$  is therefore given only by those factors  $x \in \mathcal{L}_n(u)$  for which  $\text{deg}_+(x) \geq 2$ , i.e. such that there exist distinct letters  $a, b \in \mathcal{A}$  satisfying  $xa, xb \in \mathcal{L}_{n+1}(u)$ . A factor of  $u$  which has at least two extensions to the right is called a right special factor of  $u$ . Similarly one can define a left special factor, and the relation (2) can be rewritten as

$$\Delta\mathcal{C}(n) = \sum_{x \in \mathcal{L}_n(u), x \text{ right special}} (\text{deg}_+(x) - 1) = \sum_{x \in \mathcal{L}_n(u), x \text{ left special}} (\text{deg}_-(x) - 1).$$

**Proof of (ii) of Theorem 1.2.** Suppose that the language of the infinite word  $u$  is closed under reversal. Consider the operation  $\rho$  which to every vertex of the graph associates  $\rho(x) = \bar{x}$  and to every edge associates  $\rho(e) = \bar{e}$  (see Fig. 2).

This operation maps the Rauzy graph  $\Gamma_n$  onto itself. Obviously,

$$\begin{aligned} \mathcal{P}(n) &= \#\{z \in \mathcal{L}_n(u) \mid \rho(z) = z\}, \\ \mathcal{P}(n + 1) &= \#\{e \in \mathcal{L}_{n+1}(u) \mid \rho(e) = e\}. \end{aligned}$$

We shall be interested in the paths leading between special factors. More precisely, we shall call a *simple path* a directed path  $w = v_0 v_1 \dots v_k$ , such that its initial vertex,  $v_0$ , and its final vertex,  $v_k$ , are left or right special factors, and the other vertices are not special factors, i.e.  $\text{deg}_+(v_i) = \text{deg}_-(v_i) = 1$  for  $i = 1, 2, \dots, k - 1$ . A special factor is considered as a simple path of length 0. Since the infinite word  $u$  is uniformly recurrent, the graph  $\Gamma_n$  is strongly connected, and therefore every vertex and every edge belongs to a simple path.

For an edge  $e$  satisfying  $\rho(e) = e$  we find the simple path  $w$  which contains  $e$ . Since  $\rho(e) = e$ , the operation  $\rho$  must map the path  $w$  onto itself. Similarly, if for a vertex  $z$  it holds that  $\rho(z) = z$ , then the simple path containing  $z$  is mapped by  $\rho$  onto itself.

To give an upper bound on  $\mathcal{P}(n) + \mathcal{P}(n + 1)$  therefore consists in finding the number of simple paths in the Rauzy graph  $\Gamma_n$  which are mapped by  $\rho$  onto itself. Indeed it suffices to study the so-called reduced Rauzy graph.

The set  $V$  of vertices of the reduced Rauzy graph is formed by all  $x \in \mathcal{L}_n(u)$  which are either left or right special factors of  $u$ . Each simple path which connects two vertices  $x, y \in V$  in the original Rauzy graph is replaced by the oriented edge from  $x$  to  $y$  in the reduced Rauzy graph. The operation  $\rho$  maps the reduced Rauzy graph onto itself.

The set  $V$  of vertices of the reduced Rauzy graph can be divided into disjoint cycles of the mapping  $\rho$ . Since  $\rho^2 = \text{Id}$ , the cycles are either of length 1 or 2. The cycles of length 1 are given by special factors invariant under  $\rho$ , i.e. special factors, which are themselves palindromes. Let us denote their number by  $\alpha$ , and denote the number of cycles of length 2 by  $\beta$ . Note that the number of vertices in the reduced Rauzy graph (i.e. left or right special factors in  $\Gamma_n$ ) is  $\alpha + 2\beta$ .

If there is an edge from  $x$  to  $y$ , where  $x$  and  $y$  belong to different cycles, then there is another edge leading from  $\rho(y)$  to  $\rho(x)$ . Thus there are at least two edges between vertices of two different cycles. The mapping  $\rho$  disjointly divides vertices of the reduced graph into  $\alpha + \beta$  cycles. Since the reduced Rauzy graph is strongly connected, at least  $\alpha + \beta - 1$  pairs of cycles have to be connected by an edge; therefore the number of edges, which lead between vertices of different cycles, is at least  $2(\alpha + \beta - 1)$ . These edges correspond in the original Rauzy graph  $\Gamma_n$  to the simple paths of non-zero length which are not mapped by  $\rho$  onto themselves.

As we have said, the number of palindromes of length  $n$  and  $n + 1$  is bounded by the number of simple paths in  $\Gamma_n$ , which are mapped by  $\rho$  onto themselves. We thus have

$$\mathcal{P}(n) + \mathcal{P}(n + 1) \leq \sum_{x \text{ is left or right special}} \deg_+(x) - 2(\alpha + \beta - 1) + \alpha,$$

where the first summand is the number of all simple paths of non-zero length in  $\Gamma_n$ , the second summand estimates the number of simple paths of non-zero length which are not mapped onto itself, and the third one is the number of simple paths of zero length invariant under  $\rho$  (i.e. palindromic special factors). We obtain

$$\begin{aligned} \mathcal{P}(n) + \mathcal{P}(n + 1) &\leq \sum_{x \text{ is left or right special}} \deg_+(x) - (\alpha + 2\beta) + 2 \\ &= \sum_{x \text{ is left or right special}} (\deg_+(x) - 1) + 2 = \Delta\mathcal{C}(n) + 2, \end{aligned}$$

where we have used that  $\alpha + 2\beta$  is the number of left or right special factors in  $\Gamma_n$ . This completes the proof.  $\square$

### 3. Examples of infinite words with maximal $\mathcal{P}(n) + \mathcal{P}(n + 1)$

In this section we present several examples of infinite words which satisfy  $\limsup_{n \rightarrow \infty} \mathcal{P}(n) > 0$  and

$$\mathcal{P}(n) + \mathcal{P}(n + 1) = \Delta\mathcal{C}(n) + 2. \tag{3}$$

These are in a sense words with maximal number of palindromes.

**1. Arnoux–Rauzy sequences.** Arnoux–Rauzy sequences are generalizations of Sturmian words for an alphabet with more than two letters. An infinite word  $u$  over an  $r$ -letter alphabet is called Arnoux–Rauzy of order  $r$ , if for every  $n \in \mathbb{N}$  there exists exactly one left special factor, say  $w_L$ , and exactly one right special factor, say  $w_R$ , of length  $n$ , and they satisfy  $\deg_+(w_R) = \deg_-(w_L) = r$ . Note that Arnoux–Rauzy sequences of order 2 are precisely the Sturmian words. Directly from the definition one can deduce that the factor complexity is  $\mathcal{C}(n) = (r - 1)n + 1$  for all  $n \in \mathbb{N}$ . In [10] it was shown that

$$\mathcal{P}(n) = \begin{cases} r, & \text{if } n \text{ is odd,} \\ 1, & \text{if } n \text{ is even.} \end{cases}$$

Since  $\Delta\mathcal{C}(n) = r - 1$ , we obtain

$$\mathcal{P}(n) + \mathcal{P}(n + 1) = r + 1 = \Delta\mathcal{C}(n) + 2,$$

and thus the Arnoux–Rauzy words satisfy (3).

**2. Complementation–symmetric sequences.** In [1] it was shown that complementation–symmetric sequences, with factor complexity  $\mathcal{C}(n) = 2n$  for all  $n \in \mathbb{N}$ , satisfy  $\mathcal{P}(n) = 2$  for all  $n \geq 1$ . Recall that a complementation–symmetric sequence on a two-letter alphabet, say  $\mathcal{A} = \{a, b\}$ , is a sequence such that for any factor occurring in it, the word obtained by changing  $a$ 's into  $b$ 's and vice versa, is also a factor. Since  $\Delta\mathcal{C}(n) = 2$ , we have again

$$\mathcal{P}(n) + \mathcal{P}(n + 1) = 4 = \Delta\mathcal{C}(n) + 2.$$

**3. Words associated with  $\beta$ -integers.** In [2] one studies palindromes in words associated with  $\beta$ -integers, i.e. positive real numbers which have vanishing fractional part in the numeration system with base  $\beta$ . For the description of the words  $u_\beta$  we introduce the Rényi expansion of 1.

Let  $\beta$  be a fixed real number,  $\beta > 1$ . Denote by  $T_\beta$  the mapping  $T_\beta : [0, 1] \rightarrow [0, 1)$ , given by the prescription

$$T_\beta(x) := \beta x - \lfloor \beta x \rfloor.$$

The sequence

$$d_\beta(1) = t_1 t_2 t_3 \cdots, \quad \text{where } t_i := \lfloor \beta T_\beta^{i-1}(1) \rfloor, \quad i = 1, 2, 3, \dots$$

is called the Rényi expansion of 1. If  $d_\beta(1)$  is eventually periodic, then  $\beta$  is called a Parry number.

The infinite word  $u_\beta$ , which codes the distances between neighboring  $\beta$ -integers, is the fixed point of a morphism over a finite alphabet. The morphisms are of two types, according to the type of the Parry number  $\beta$ .

- If  $d_\beta(1) = t_1 \cdots t_m 0^\omega$ , with  $t_m \neq 0$ , then  $\beta$  is called a simple Parry number. In this case  $u_\beta$  is the fixed point of the substitution  $\varphi = \varphi_\beta$  over the alphabet  $\mathcal{A} = \{0, 1, \dots, m - 1\}$ , given by

$$\begin{aligned} \varphi(0) &= 0^{t_1} 1, \\ \varphi(1) &= 0^{t_2} 2, \\ &\vdots \\ \varphi(m-2) &= 0^{t_{m-1}} (m-1), \\ \varphi(m-1) &= 0^{t_m}. \end{aligned}$$

- If  $d_\beta(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega$ , where  $m, p$  are minimal indices which allow such notation, then  $u_\beta$  is the fixed point of the substitution  $\varphi = \varphi_\beta$  over the alphabet  $\mathcal{A} = \{0, 1, \dots, m + p - 1\}$ , given by

$$\begin{aligned} \varphi(0) &= 0^{t_1} 1, \\ \varphi(1) &= 0^{t_2} 2, \\ &\vdots \\ \varphi(m-1) &= 0^{t_m} m, \\ &\vdots \\ \varphi(m+p-2) &= 0^{t_{m+p-1}} (m+p-1), \\ \varphi(m+p-1) &= 0^{t_{m+p}} m. \end{aligned}$$

One can easily show that the infinite word  $u_\beta$  is uniformly recurrent. The condition of invariance of the language of  $u_\beta$  under reversal is described in [13] for the case of simple Parry number, i.e.  $d_\beta(1) = t_1 \cdots t_m 0^\omega$ . It is shown that  $\mathcal{L}(u_\beta)$  is closed under reversal if and only if  $t_1 = \cdots = t_{m-1} \geq t_m$ . For the case  $d_\beta(1) = t_1 \cdots t_m (t_{m+1} \cdots t_{m+p})^\omega$  it is shown in [4] that the language of  $u_\beta$  is closed under reversal if and only if  $m = p = 1$ . Papers [13,3] show that, if  $u_\beta$  has its language invariant under reversal, then

$$\mathcal{P}(n+2) - \mathcal{P}(n) = \Delta^2 \mathcal{C}(n) = \Delta \mathcal{C}(n+1) - \Delta \mathcal{C}(n),$$

which allows one to derive the validity of (3).

While in examples 1 and 2 the second difference  $\Delta^2 \mathcal{C}(n) \equiv 0$ , for the words  $u_\beta$  it holds that  $\Delta^2 \mathcal{C}(n) \in \{-1, 0, 1\}$  and all three values are reached infinitely many times.

**4. Words coding  $r$ -interval exchange transformation.** Another possible generalization of Sturmian words are words coding a bijective transformation of the interval  $[0, 1)$  onto itself, known under the name  $r$ -interval exchange. Let us recall the definition of an interval exchange map. It can be found together with some properties in [17,18].

Given  $r$  positive numbers  $\alpha_1, \alpha_2, \dots, \alpha_r$  such that  $\sum_{i=1}^r \alpha_i = 1$ , they define a partition of the interval  $I = [0, 1)$  into  $r$  intervals

$$I_k = \left[ \sum_{i=1}^{k-1} \alpha_i, \sum_{i=1}^k \alpha_i \right), \quad k = 1, 2, \dots, r.$$

Let  $\pi$  denote a permutation of the set  $\{1, 2, \dots, r\}$ . The interval exchange transformation associated with  $\alpha_1, \dots, \alpha_r$  and  $\pi$  is defined as the map  $T : I \rightarrow I$  which exchanges the intervals  $I_k$  according to the permutation  $\pi$ ,

$$T(x) = x + \sum_{j < \pi(k)} \alpha_{\pi^{-1}(j)} - \sum_{j < k} \alpha_j, \quad \text{for } x \in I_k.$$

For  $x_0 \in I$ , the sequence  $(T^n(x_0))_{n \in \mathbb{Z}}$  is called the orbit of  $x_0$  under  $T$ . The infinite bidirectional word  $(u_n)_{n \in \mathbb{Z}}$  over the alphabet  $\mathcal{A} = \{1, \dots, r\}$  associated to the orbit  $(T^n(x_0))_{n \in \mathbb{Z}}$  is defined as

$$u_n = k \in \mathcal{A} \iff T^n(x_0) \in I_k.$$

The complexity of the word corresponding to any  $r$ -interval exchange transformation satisfies  $\mathcal{C}(n) \leq n(r - 1) + 1$ , for all  $n \in \mathbb{N}$ . Here we focus on the non-degenerated case, i.e. on mappings  $T$  for which the complexity of the word

associated to the orbit of arbitrary  $x_0 \in I$  satisfies  $\mathcal{C}(n) = (r - 1)n + 1$ , for all  $n \in \mathbb{N}$ . This property is ensured by additional conditions (denoted by  $\mathfrak{P}$ ) on the parameters of the map  $T$ .

- ( $\mathfrak{P}$ ) 1.  $\alpha_1, \dots, \alpha_r$  are linearly independent over  $\mathbb{Q}$ ,
- 2.  $\pi\{1, \dots, k\} \neq \{1, \dots, k\}$  for each  $k = 1, 2, \dots, r - 1$ .

If the conditions ( $\mathfrak{P}$ ) are fulfilled, then the set  $\{T^n(x_0)\}_{n \in \mathbb{Z}}$  is dense in  $I$  for each  $x_0 \in I$  and the dynamical system associated to the transformation  $T$  is minimal. This implies that the infinite word corresponding to the sequence  $(T^n(x_0))_{n \in \mathbb{Z}}$  is uniformly recurrent.

Another important consequence of ( $\mathfrak{P}$ ) is that the language of the word  $(u_n)_{n \in \mathbb{Z}}$  corresponding to  $(T^n(x_0))_{n \in \mathbb{Z}}$  does not depend on the position of the starting point  $x_0$ , but only on the transformation  $T$ . Therefore the notation  $\mathcal{L}(T)$ , which we adopt here, is justified. We know that  $\mathcal{L}(T_1) = \mathcal{L}(T_2)$  only if  $T_1$  and  $T_2$  coincide.

If  $r = 2$ , the permutation satisfying  $\mathfrak{P}$  is  $\pi(1) = 2, \pi(2) = 1$  and the corresponding word is Sturmian. On the other hand, every Sturmian word can be obtained as a coding of a 2-interval exchange transformation.

If  $r = 3$ , then condition 2 of  $\mathfrak{P}$  is satisfied by three permutations. One can easily see that only the permutation  $\pi(1) = 3, \pi(2) = 2, \pi(3) = 1$  gives an infinite word with language invariant under reversal. Such words can be geometrically represented by cut-and-project sequences [14].

For general  $r$ , the language of the infinite word  $u$  is closed under reversal if and only if

$$\pi(1) = r, \quad \pi(2) = r - 1, \quad \dots, \quad \pi(r) = 1. \tag{4}$$

Only for such a permutation does the infinite word  $u$  coding the corresponding interval exchange transformation satisfy  $\limsup_{n \rightarrow \infty} \mathcal{P}(n) > 0$ .

The palindromic complexity in words coding a 3-interval exchange map was described in [10]. In Section 4 we generalize this result for any  $r$ . We show that for words coding an  $r$ -interval exchange transformation with permutation (4) the equality (3) holds.

#### 4. Words coding interval exchange transformation

In this section we will be dealing only with such transformations  $T$  of  $r$ -intervals for which the permutation  $\pi$  satisfies (4). In this case the transformation has the form

$$T(x) = x + \sum_{j>k} \alpha_j - \sum_{j<k} \alpha_j \quad \text{for } x \in I_k. \tag{5}$$

It is known that there exists an interval  $I_w \subset I_{w_0}$  for every word  $w = w_0 w_1 \dots w_{n-1} \in \mathcal{L}(T)$  such that the sequence of points  $x, T(x), \dots, T^{n-1}(x)$  is coded by the same word  $w$  for each  $x \in I_w$ . Note that the boundaries of the interval  $I_w$  belong to the set  $\mathbb{Z}[\alpha_1, \dots, \alpha_r] = \{\sum k_i \alpha_i \mid k_i \in \mathbb{Z}\}$ .

Let us denote the decomposition of the interval  $I = [0, 1)$  by the transformation  $T^{-1}$  by  $\tilde{I}_1, \tilde{I}_2, \dots, \tilde{I}_r$  and analogously  $\tilde{I}_w$  for an arbitrary  $w \in \mathcal{L}(T^{-1})$ .

Clearly,  $\tilde{I}_{\pi^{-1}(j)} = T(I_j)$  for each  $j \in \{1, \dots, r\}$ . Since  $\pi$  is of the form (4), it follows that  $I_j = [a, b)$  implies  $\tilde{I}_{\pi^{-1}(j)} = T(I_j) = [1 - b, 1 - a)$ . The same relation is therefore valid for any factor  $w \in \mathcal{L}(T)$ ,

$$I_w = [a, b) \implies \tilde{I}_{\pi^{-1}(w)} = [1 - b, 1 - a), \tag{6}$$

where  $\pi^{-1}$  is viewed as a morphism of  $\mathcal{A}^*$ .

Now we have everything prepared for determination of the palindromic complexity.

**Theorem 4.1.** *Let  $\alpha_1, \dots, \alpha_r$  be positive real numbers, linearly independent over  $\mathbb{Q}$ , and  $\pi$  the permutation given by (4). Then*

$$\mathcal{P}(n) = \begin{cases} 1 & \text{for each } n \text{ even,} \\ r & \text{for each } n \text{ odd.} \end{cases}$$



**Proof.** Consider a palindrome of even length in the form of

$$w_{n-1}w_{n-2} \dots w_0w_0 \dots w_{n-2}w_{n-1} \in \mathcal{L}(T).$$

This means that there exists  $x \in [0, 1)$  such that

$$\begin{aligned} x \in I_{w_0}, & \quad T(x) \in I_{w_1}, & \dots, & \quad T^{n-1}(x) \in I_{w_{n-1}}, \\ T^{-1}(x) \in I_{w_0}, & \quad T^{-2}(x) \in I_{w_1}, & \dots, & \quad T^{-n}(x) \in I_{w_{n-1}}. \end{aligned}$$

Hence  $x \in I_w$ , where  $w = w_0, \dots, w_{n-1}$ , and on the other side

$$\begin{aligned} x & \in T(I_{w_0}) = \tilde{I}_{\pi^{-1}(w_0)}, \\ T^{-1}(x) & \in T(I_{w_1}) = \tilde{I}_{\pi^{-1}(w_1)}, \\ & \vdots \\ T^{-n+1}(x) & \in T(I_{w_{n-1}}) = \tilde{I}_{\pi^{-1}(w_{n-1})}. \end{aligned}$$

It follows that  $x \in \tilde{I}_{\pi^{-1}(w_0w_1\dots w_{n-1})}$ . Thus  $x$  has to belong to the intersection of both intervals, i.e.  $x \in I_w \cap \tilde{I}_{\pi^{-1}(w)}$ . If  $I_w = [a, b)$ , then according to (6)

$$x \in [a, b) \cap [1 - b, 1 - a).$$

Now we use the simple fact that for every interval  $[a, b)$  it holds that

$$[a, b) \cap [s - b, s - a) \neq \emptyset \iff \frac{s}{2} \in [a, b). \tag{7}$$

Therefore

$$\frac{1}{2} \in I_w \cap \tilde{I}_{\pi^{-1}(w)}. \tag{8}$$

We have shown that every palindrome of even length arises from the coding of

$$T^{-n}\left(\frac{1}{2}\right), \dots, T^{-1}\left(\frac{1}{2}\right), \frac{1}{2}, T\left(\frac{1}{2}\right), \dots, T^{n-1}\left(\frac{1}{2}\right).$$

The fact that  $\{T^n(x)\}_{n \in \mathbb{Z}}$  is dense in  $[0, 1)$  implies that the previous sequence occurs in  $(T^n(x))_{n \in \mathbb{Z}}$ , for each  $x$ . Thus the coding of  $(T^n(x))_{n \in \mathbb{Z}}$  includes exactly one palindrome of even length for each  $n$ .

Consider now a palindrome of odd length in the form of

$$w_{n-1}w_{n-2} \dots w_1w_0w_1 \dots w_{n-2}w_{n-1} \in \mathcal{L}(T).$$

Again, this means that there exist  $x, y \in [0, 1)$  such that

$$\begin{aligned} x \in I_{w_0}, & \quad T(x) \in I_{w_1}, & \dots, & \quad T^{n-1}(x) \in I_{w_{n-1}}, \\ y \in I_{w_1}, & \quad T^{-1}(y) \in I_{w_0}, & \dots, & \quad T^{-n}(y) \in I_{w_{n-1}}. \end{aligned}$$

The first sequence is the coding of the word  $w = w_0w_1 \dots w_{n-1}$ , i.e.  $x \in I_w$ , and the following is true for the second one

$$\begin{aligned} y & \in T(I_{w_0}) = \tilde{I}_{\pi^{-1}(w_0)}, \\ T^{-1}(y) & \in T(I_{w_1}) = \tilde{I}_{\pi^{-1}(w_1)}, \\ & \vdots \\ T^{-n+1}(y) & \in T(I_{w_{n-1}}) = \tilde{I}_{\pi^{-1}(w_{n-1})}. \end{aligned}$$

Thus  $y \in \tilde{I}_{\pi^{-1}(w)}$ . If there exists a palindrome of odd length with the central letter  $w_0$  then it has to be  $y = T(x) = x + s_{w_0}$ , where  $s_{w_0} \neq 0$  is a shift of  $x \in I_{w_0}$  by the mapping  $T$ . Using (5) we have  $s_{w_0} = \sum_{j>w_0} \alpha_j - \sum_{j<w_0} \alpha_j$ . In other words we have  $x \in I_w$  and  $x \in \tilde{I}_{\pi^{-1}(w)} - s_{w_0}$ . If  $I_w = [c, d)$  then  $\tilde{I}_{\pi^{-1}(w)} = [1 - d, 1 - c)$ , and therefore  $x \in [c, d) \cap [1 - d - s_{w_0}, 1 - c - s_{w_0})$ . According to (7)

$$x_{w_0} := \frac{1 - s_{w_0}}{2} \in I_w \cap (\tilde{I}_{\pi^{-1}(w)} - s_{w_0}).$$



We have shown that the palindrome  $w_{n-1}w_{n-2} \dots w_1w_0w_1 \dots w_{n-2}w_{n-1}$  can be obtained by coding of following sequences:

$$T^{-n+1}(x_{w_0}), \dots, T^{-1}(x_{w_0}), x_{w_0}, T(x_{w_0}), \dots, T^{n-1}(x_{w_0}). \quad (9)$$

One may rewrite

$$x_{w_0} = \frac{1 - s_{w_0}}{2} = \frac{\sum_{j=1}^r \alpha_j - \sum_{j=w_0+1}^r \alpha_j + \sum_{j=1}^{w_0-1} \alpha_j}{2} = \sum_{j=1}^{w_0-1} \alpha_j + \frac{\alpha_{w_0}}{2}.$$

This means that the point  $x_{w_0}$ , which corresponds to central letter  $w_0$  in the palindrome of odd length, is lying in the middle of interval  $I_{w_0}$  associated to the central letter.

On the other hand, if  $x_{w_0}$  is the centre of one of the intervals  $I_1, \dots, I_r$ , the sequence (9) corresponds to a palindrome. Therefore  $\mathcal{P}(2n+1) = r$ .  $\square$

Note that according to the previous theorem the interval exchange transformation with a permutation satisfying (4) has the same palindromic complexity and also factor complexity as Arnoux–Rauzy words over  $r$  letters [10,16].

## 5. Conclusions

The main result of this paper is the estimate of the palindromic complexity of infinite words in terms of their factor complexity. We have shown in [Theorem 1.2](#) that uniformly recurrent words with infinitely many palindromes satisfy the following relation:

$$\mathcal{P}(n) + \mathcal{P}(n+1) \leq \Delta\mathcal{C}(n) + 2, \quad \text{for all } n \in \mathbb{N}.$$

It is interesting to mention that the first difference of factor complexity was already useful for estimation of the frequencies of factors. In [8] it is shown that the frequencies of factors of length  $n$  in a recurrent word take at most  $3\Delta\mathcal{C}(n)$  values.

The second part of the paper is devoted to infinite words for which  $\mathcal{P}(n) + \mathcal{P}(n+1)$  in [Theorem 1.2](#) reaches the upper bound. We cite several examples of such infinite words among the words for which the palindromic and factor complexity is known. As a new result, we derive the palindromic complexity for infinite words coding  $r$ -interval exchange transformation and prove that for this class of infinite words the equality in the estimate holds, too.

According to our knowledge, all known examples of infinite words which satisfy the equality  $\mathcal{P}(n) + \mathcal{P}(n+1) = \Delta\mathcal{C}(n) + 2$  for  $n \in \mathbb{N}$  have sublinear factor complexity. A known example of an infinite word with higher factor complexity is the billiard sequences on three letters, for which  $\mathcal{C}(n) = n^2 + n + 1$ . As shown in [7], they satisfy  $\mathcal{P}(n) + \mathcal{P}(n+1) = 4$ , and thus billiard sequences do not reach the upper bound in [Theorem 1.2](#).

The proof of [Theorem 1.2](#) is based on the study of properties of the Rauzy graph and its behaviour with respect to the operation of mirror image on the language of the infinite word. It turns out that the Rauzy graphs of words reaching the upper bound in our estimate of palindromic complexity must have a very special form.

## Acknowledgments

The authors acknowledge the financial support of Czech Science Foundation GAČR 201/05/0169 and the Ministry of Education of the Czech Republic LC00602.

## References

- [1] J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik, Palindrome complexity, *Theoret. Comput. Sci.* 292 (2003) 9–31.
- [2] P. Ambrož, C. Frougny, Z. Masáková, E. Pelantová, Palindromic complexity of infinite words associated with simple Parry numbers, *Ann. Inst. Fourier* 56 (2007) 2131–2160.
- [3] L. Balková, Z. Masáková, Palindromic complexity of infinite words associated with non-simple Parry numbers, Preprint DI, arXiv:math/0611508v1 (2006).
- [4] J. Bernat, Computation of  $L_{\oplus}$  for several cubic Pisot numbers, *Discrete Math. Theor. Comput. Sci.* (2007) (in press).
- [5] J. Berstel, Recent results on extensions of Sturmian words, *Internat. J. Algebra Comput.* 12 (2002) 371–385.
- [6] E. Bombieri, J.E. Taylor, Which distributions of matter diffract? An initial investigation, *J. Phys.* 14 (Colloque C3) (1986) 19–28.

- [7] J.P. Borel, Complexity and palindromic complexity of billiards words, in: S. Brlek, C. Reutenauer (Eds.), *Proceedings of WORDS2005*, 2005, pp. 175–183.
- [8] M. Boshernitzan, A condition for unique ergodicity of minimal symbolic flows, *Ergodic Theory Dynam. Systems* 12 (1992) 425–428.
- [9] D. Damanik, Gordon-type arguments in the spectral theory of one-dimensional quasicrystals, in: M. Baake, R.V. Moody (Eds.), *Directions in Mathematical Quasicrystals*, in: CRM Monograph Series, vol. 13, AMS, Providence, RI, 2000, pp. 277–305.
- [10] D. Damanik, L.Q. Zamboni, Combinatorial properties of Arnoux–Rauzy subshifts and applications to Schrödinger operators, *Rev. Math. Phys.* 15 (2003) 745–763.
- [11] D. Damanik, D. Zare, Palindrome complexity bounds for primitive substitution sequences, *Discrete Math.* 222 (2000) 259–267.
- [12] X. Droubay, G. Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.* 223 (1999) 73–85.
- [13] Ch. Frougny, Z. Masáková, E. Pelantová, Complexity of infinite words associated with beta-expansions, *RAIRO Theor. Inform. Appl.* 38 (2004) 163–185.
- [14] L.S. Guimond, Z. Masáková, E. Pelantová, Combinatorial properties of infinite words associated with cut-and-project sequences, *J. Théor. Nombres Bordeaux* 15 (2003) 697–725.
- [15] A. Hof, O. Knill, B. Simon, Singular continuous spectrum for palindromic Schrödinger operators, *Comm. Math. Phys.* 174 (1995) 149–159.
- [16] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms, *Theoret. Comput. Sci.* 276 (2002) 281–313.
- [17] M.S. Keane, Interval exchange transformations, *Math. Z.* 141 (1975) 25–31.
- [18] G. Rauzy, Échanges d’intervalles et transformations induites, *Acta Arith.* 34 (1979) 315–328.
- [19] A. Sütő, Schrödinger difference equation with deterministic ergodic potential, in: F. Alex, D. Gratias (Eds.), *Beyond Quasicrystals* (Les Houches, 1994), Springer, Berlin, 1995, pp. 481–549.
- [20] D. Shechtman, L. Blech, D. Gratias, J.V. Cahn, Metallic phase with long-range orientational order and no translational symmetry, *Phys. Rev. Lett.* 53 (1984) 1951–1953.