

Various Properties of Sturmian Words

P. Baláži

This overview paper is devoted to Sturmian words. The first part summarizes different characterizations of Sturmian words. Besides the well known theorem of Hedlund and Morse it also includes recent results on the characterization of Sturmian words using return words or palindromes. The second part deals with substitution invariant Sturmian words, where we present our recent results. We generalize one-sided Sturmian words using the cut-and-project scheme and give a full characterization of substitution invariant Sturmian words.

Keywords: Sturmian words, mechanical words, 2-interval exchange map, palindromes, return words, substitutions.

1 Introduction

In recent years, the combinatorial properties of finite and infinite words have become significantly important in fields of physics, biology, mathematics and computer science. One of the first impulses for extensive research in this field was the discovery of quasi-crystals.

Normal crystal structures show rotational and translational symmetry. In 1982, however, Dan Shechtman discovered an aperiodic structure (which was formed by rapidly-quenched aluminum alloys) that has a perfect long-range order, but no three-dimensional translational periodicity (see e.g. [1] or [2]). Since then many stable and unstable aperiodic structures have been discovered. They are now known as quasi-crystals.

The problematics of aperiodic structures has been studied from various points of view and there are numerous relations with other applications (besides solid-state physics), such as pseudo-random number generators [3], pattern recognition and symbolic dynamical systems. Early results are reported in [4] or [5].

This paper is devoted to Sturmian words. Sturmian words are infinite words over a binary alphabet with exactly $n+1$ factors of length n for each $n \geq 0$. They represent the simplest family of quasi-crystals. The history of Sturmian words dates back to the astronomer J. Bernoulli III [6]. He considered the sequence $[(n+1)\alpha + 1/2] - [n\alpha + 1/2]$, where $n \geq 1$ and α is a positive irrational. Based on the continued fraction expansion of α , he gave (without proof) an explicit description of the terms of the sequence. There also exist some early works by Christoffel and Markoff.

A. A. Markoff was the first to prove the validity of Bernoulli's description. He did that in his work [7], where he described the terms of the sequence $[(n+1)\alpha] - [n\alpha] - [\alpha]$, $n \geq 1$ (later known as a mechanical sequence).

The first detailed investigation of Sturmian words is due to Hedlund and Morse [4], who studied such words from the point of view of symbolic dynamics and, in fact, introduced the term „Sturmian“; named after the mathematician Charles Francois Sturm.

It appears that there are several equivalent ways of constructing Sturmian words. We will describe some of them and show the relationship with other notions from the combinatorics on words such as palindromes and return words. Then we make some notes on an extension of Sturmian words using the cut-and-project scheme. The next section is devoted to the important question of the invariance of Sturmian words

on a substitution. In the last section we present some open problems related to generalizations of Sturmian words.

2 Sturmian words

An infinite one-sided word

$$w = (w_n)_{n \in N_0} = w_0 w_1 w_2 \dots,$$

is a sequence of letters from a finite set A which is called an alphabet. We use the notation N for the set of integers and $N_0 = N \cup \{0\}$. A finite word $v = v_0 v_1 \dots v_{n-1}$ is a finite string of letters from A and $|v| = n$ is the length of v . The set of all finite words over the alphabet A is denoted by A^* . A finite word $u \notin A^*$ is a factor of w if there exist $0 \leq k \leq l$ such that $u = w_k \dots w_l$. The empty word is denoted by ε . The set of factors of w of length n is written $L_n(w)$ and the set of all factors of w is denoted by $L(w)$. The set $L(w)$ is often called the language.

An infinite word w is ultimately periodic if there exist a word u and a word v such that $w = uv^\omega$ where v^ω is the infinite concatenation of the word v . It is periodic if u is the empty word. If the infinite word w does not have any of the previous forms we say that it is aperiodic. We say that u is a prefix (resp. suffix) of $v = v_0 v_1 \dots v_{n-1}$, $v \in L_n(w)$, if there exists $0 \leq l \leq n-1$ such that $u = v_0 \dots v_l$ (resp. $u = v_l v_{l+1} \dots v_{n-1}$). The infinite word w , $l \geq 0$, is a suffix of w .

An infinite word w is uniformly recurrent if for every integer k there exists an integer l such that each word of $L_k(w)$ occurs in every word of length l .

The usual way of defining Sturmian words is via the complexity function. Let w be an infinite word and C_w be a mapping $N_0 \rightarrow N$, such that $C_w(n)$ is the number of different factors of length n of w ; i.e. $C_w(n) = \#L_n(w)$. C_w is called the complexity and we say that the word w is Sturmian if $C_w(n) = n + 1$ for all n . Since $C_w(1) = 2$, Sturmian words are defined over a binary alphabet, say $A = \{0, 1\}$. Clearly $C_w(0) = 1$ for any w .

There have been some attempts to extend Sturmian words to words over alphabets with more than two letters, for instance [8] or [9], but none of these constructions show such nice properties as Sturmian words. However, the approach of Arnoux and Rauzy presented in [8] resulted in another interesting family of words called Arnoux-Rauzy sequences.

Another combinatorial definition of Sturmian words is based on the distribution of letters in the word. Let w be an infinite word, $u, v \in L(w)$ be two factors of the same length $|u| = |v|$ and function δ be defined

$$\delta(u, v) = \left| |u|_0 - |v|_0 \right|,$$

where $|u|_0$ denotes the number of occurrences of 0 in u . We say that w is balanced if and only if $\delta(u, v) \leq 1$ for all $u, v \in L(w)$ with $|u| = |v|$. Note that the structure of a balanced word over the alphabet $A = \{0, 1\}$ is formed, either by a block of 0's between two consecutive 1's, or by a block of 1's between two consecutive 0's. It is easy to see that the length of the blocks of 0's between two consecutive 1's (resp. blocks of 1's between two consecutive 0's) differs at most by 1 in a balanced word. As we will see balanced words are equivalent to Sturmian words.

In the literature one can find several constructions of binary words. We start with the construction presented by Hedlund and Morse in [4]. Let α, β be real numbers, where $\alpha \in (0, 1)$ and $\beta \in [0, 1)$. Then $\underline{s}_{\alpha, \beta} = \underline{s}_{\alpha, \beta}(n)$, $\bar{s}_{\alpha, \beta} = \bar{s}_{\alpha, \beta}(n)$ are sequences defined by,

$$\underline{s}_{\alpha, \beta}(n) := \lfloor (n+1)\alpha + \beta \rfloor - \lfloor n\alpha + \beta \rfloor, \text{ for } n \geq 0,$$

$$\bar{s}_{\alpha, \beta}(n) := \lceil (n+1)\alpha + \beta \rceil - \lceil n\alpha + \beta \rceil, \text{ for } n \geq 0.$$

These words are usually referred to as mechanical words and α is called the slope and β the intercept. Mechanical words of $\underline{s}_{\alpha, \beta}$ resp. $\bar{s}_{\alpha, \beta}$ have a nice geometrical interpretation. Consider the integer lattice Z^2 , the straight line $y = \alpha x + \beta$, $x \geq 0$, and two sequences of integer points $X_n = (n, \lfloor n\alpha + \beta \rfloor)$ and $Y_n = (n, \lceil n\alpha + \beta \rceil)$. The elements of X_n cover integer points of the lattice just below the line $y = \alpha x + \beta$ and the elements of Y_n cover points just above the line. If $\underline{s}_{\alpha, \beta}(n) = 0$ then the points X_n and X_{n+1} lie on the horizontal line, and if $\underline{s}_{\alpha, \beta}(n) = 1$ then they lie on the diagonal line. The same holds for $\bar{s}_{\alpha, \beta}(n)$ and Y_n . In fact, the sequences $\underline{s}_{\alpha, \beta}$, $\bar{s}_{\alpha, \beta}$ are the coding of a discretization of a straight line.

Note that if α is irrational then $\underline{s}_{\alpha, \beta}$ and $\bar{s}_{\alpha, \beta}$ differ by at most for two values of n . Clearly, this can happen only if $n\alpha + \beta$ is an integer for some n . If we consider $\beta = 0$ and $n = 0$, then $\underline{s}_{\alpha, 0}(0) = 0$ and $\bar{s}_{\alpha, 0}(0) = 1$ and we obtain the important special case of mechanical words

$$\underline{s}_{\alpha, 0} = 0c_\alpha, \bar{s}_{\alpha, 0} = 1c_\alpha; n \geq 0.$$

The infinite word c_α is called the characteristic sequence of α .

Crisp et al. study, in their work [10], another way of constructing characteristic sequences. Consider again the integer lattice Z^2 and the straight line $y = \rho x$, where ρ is a positive irrational and $x \geq 0$. We label the intersections of the line $y = \rho x$ with verticals of the grid using 0, and we label by 1 the intersections of $y = \rho x$ with horizontals. The sequence of labels forms the so called cutting sequence and is denoted by S_ρ . It can be shown that $c_\alpha = S_\rho$ if and only if $\rho = \alpha/(1 - \alpha)$ (see e.g. [10]).

One can often encounter the term r -interval exchange map in connection with infinite words. Properties of words generated by the r -interval exchange map (called the coding of the r -interval exchange map) are studied from different aspects in [11], [12] or [13]. Let us closely mention the 2-interval exchange map and its reference to Sturmian words.

Let $\alpha \in (0, 1)$ be an irrational number and $x \in [0, 1)$. Let $I_1 = [0, 1 - \alpha)$, $I_2 = [1 - \alpha, 1)$ be the decompositions of the interval $[0, 1)$. The map T_α , given by the rule

$$T_\alpha(x) = \begin{cases} x + \alpha, & \text{for } x \in [0, 1 - \alpha), \\ x + \alpha - 1, & \text{for } x \in [1 - \alpha, 1), \end{cases}$$

is called the 2-interval exchange map. It can be written in a more complex way

$$T_\alpha(x) = x + \alpha - \lfloor x + \alpha \rfloor = \{x + \alpha\}, \forall x \in [0, 1).$$

For the n -th iteration of the 2-interval exchange map T_α we obtain

$$T_\alpha^{(n)}(x) = T_{n\alpha}(x) = \{n\alpha + x\}.$$

Since α is irrational, T_α does not have any fixed point. It is not difficult to check that

$$\lfloor (n+1)\alpha + x \rfloor = \lfloor n\alpha + x \rfloor \Leftrightarrow T_\alpha^{(n)}(x) \in [0, 1 - \alpha)$$

which implies that

$$\underline{s}_{\alpha, \beta}(n) = \begin{cases} 0 & \text{if } T_\alpha^{(n)}(\beta) \in [0, 1 - \alpha), \\ 1 & \text{if } T_\alpha^{(n)}(\beta) \in [1 - \alpha, 1). \end{cases}$$

Let us show the most famous example of Sturmian words, the Fibonacci word.

Example 2.1: Let φ be the map given by $\varphi: 0 \mapsto 01, 1 \mapsto 0$ with the property $\varphi(uv) = \varphi(u)\varphi(v)$ for any finite words u, v over the alphabet $\{0, 1\}$. Define the n -th finite Fibonacci word f_n in the following way: $f_0 = 0$ and for all $n \geq 0$,

$$f_{n+1} = \varphi(f_n).$$

Since $f_0 = 0$ and $\varphi(0)$ starts with 0, then f_n is the prefix of f_{n+1} for each $n \geq 0$. The Fibonacci word is defined as the limit of the sequence of words f_n and thus

$$\lim_{n \rightarrow \infty} f_n = 010010100100101001010010010\dots$$

Note that the further applications of the map f_n do not change the Fibonacci word. Observe that the length $|f_n|$ is the n -th element of the Fibonacci sequence F_n , ($F_n = F_{n-1} + F_{n-2}$; $F_0 = 1, F_1 = 2$), for all $n \geq 0$. It can be shown that the complexity of the Fibonacci word equals to $n+1$, thus the word is Sturmian. The Fibonacci word also coincides with the mechanical word with slope $1/\tau^2$ and zero intercept, where $\tau = 1/2(1 + \sqrt{5})$ is the golden mean.

We finalize this part with a theorem by Hedlund and Morse [4] which states that Sturmian, balanced and mechanical words are indeed equivalent.

Theorem 2.1: (Hedlund, Morse). Let w be an infinite word over the alphabet $A = \{0, 1\}$. The following conditions are equivalent:

1. w is Sturmian;
2. w is balanced and aperiodic;
3. there exist an irrational α , $\alpha \in (0, 1)$ and a real $\beta \in [0, 1)$ such that $w = \underline{s}_{\alpha, \beta}$ or $w = \bar{s}_{\alpha, \beta}$, for all $n \geq 0$.

There exist several proofs of the theorem. The original proof [4] is of combinatorial nature, while the other by Lunnon and Pleasants [14] is based on geometrical considerations.

2.1 Other characteristics of Sturmian words

We have mentioned several equivalent definitions of Sturmian words as those with minimal complexity, balanced aperiodic sequences and mechanical words. In the past few years, there have been successful attempts to find a new characterization of Sturmian words. The first one, which we will describe uses return words, while the second uses palindromes.

Let w be a one-sided infinite word and u a factor of w . We say that a finite word v is the return word over u if vu is a factor of w , u is a prefix of vu and there are exactly 2 occurrences of u

in vu . In other words, the return word v over u starts with the occurrence of u and ends just before the next occurrence of u .

Example 2.2: Let 0100101001001010010100100101... be the Fibonacci word. The set of return words over 0101 contains words 01010010 and 01010. For clarity, here is the Fibonacci word with indicated return words over 0101 :
 01001010010010100100100101...

Vuillon [15] observed that the number of return words indicates whether a word is Sturmian or not. He showed the following theorem.

Theorem 2.2: *A binary infinite word w is Sturmian if and only if the set of return words over u has exactly two elements for every non empty word u .*

Note that the proof of the necessary condition includes a nice application of Rauzy graphs, which are often used for investigation of growth of the complexity in infinite words [8].

Let us focus on palindromes. A palindrome is a finite word that reads the same backwards as forwards. For instance, these are the first palindromes of the Fibonacci word: $\varepsilon, 0, 1, 00, 010, 101, 1001, 00100, 01010, \dots$ In [16], Droubay and Pirillo showed the characterization of Sturmian words by observing the number of palindromes of even and odd length.

Theorem 2.3: *An infinite word is Sturmian if and only if, for each nonnegative integer n , there is exactly one palindrome of length n , if n is even, and there are exactly two palindromes of length n , if n is odd.*

The mapping that assigns to an integer n the number of palindromes of length n in a word is called the palindromic complexity.

In the context of palindromes and Sturmian words, let us draw attention to paper [17]. The authors proved that the number of palindromes in a word n of length n is less or is equal to $n+1$. Note that this holds for any kind of words (not necessary Sturmian) over arbitrary alphabets. However in the case of Sturmian words it was shown in [17] that the number of all palindromes in a factor of length n is equal to $n+1$. The reader may like to try finding words of length n over a 2-letter alphabet with the number of palindromes less than $n+1$; this is not as trivial as it may appear to be.

2.2 Bidirectional Sturmian words

In the previous sections we have outlined several characteristics of Sturmian words. However we have limited ourselves, from the very first definitions, to one-sided infinite words. This restriction is typical for a large number of papers, in spite the fact that the definitions of notions like balanced words, mechanical words etc. can be extended, very naturally, to both sides. The question is, whether the above listed theorems still hold for bidirectional infinite words. Let us show a way of generating bidirectional infinite words from which it will be clear that such a generalization is possible.

Consider η, ε fixed positive irrational numbers, $\eta \neq -\varepsilon$, $\beta \in [0,1)$ and the set

$$\Sigma_{\varepsilon,\eta}(\beta - 1, \beta) = \{m + n\eta \mid \beta - 1 < m - n\varepsilon \leq \beta, m, n \in \mathbb{Z}\} \dagger$$

It follows that m has to satisfy

$$\beta + n\varepsilon - 1 < m \leq \beta + n\varepsilon,$$

hence $m = \lfloor \beta + n\varepsilon \rfloor$, i.e. any element of $\Sigma_{\varepsilon,\eta}(\beta - 1, \beta)$ has the form $\lfloor \beta + n\varepsilon \rfloor + n\eta$, for some $n \in \mathbb{Z}$. Using this fact we can write

$$\Sigma_{\varepsilon,\eta}(\beta - 1, \beta) = \{x_n := \lfloor \beta + n\varepsilon \rfloor + n\eta \mid n \in \mathbb{Z}\},$$

thus the elements of $\Sigma_{\varepsilon,\eta}(\beta - 1, \beta)$ form an increasing sequence $(x_n)_{n \in \mathbb{Z}}$. We can compute the lengths between two consecutive points of $(x_n)_{n \in \mathbb{Z}}$ as

$$x_{n+1} - x_n = \lfloor (n+1)\varepsilon + \beta \rfloor - \lfloor n\varepsilon + \beta \rfloor + \eta.$$

The term $\lfloor (n+1)\varepsilon + \beta \rfloor - \lfloor n\varepsilon + \beta \rfloor$ takes only two values 0 and 1 for each $n \in \mathbb{Z}$. Thus there are only two distances between neighbors in $\Sigma_{\varepsilon,\eta}(\beta - 1, \beta)$, namely $1 + \eta$ and η . Let us define a sequence $(w_n)_{n \in \mathbb{Z}}$ of 0's and 1's by

$$w_n = \begin{cases} 0 & \text{if } x_{n+1} - x_n = \eta, \\ 1 & \text{if } x_{n+1} - x_n = 1 + \eta. \end{cases}$$

The sequence $(w_n)_{n \in \mathbb{Z}}$ is a pointed bidirectional infinite word

$$(w_n)_{n \in \mathbb{Z}} = \dots w_{-2} w_{-1} | w_0 w_1 w_2 \dots,$$

where $|$ denotes a delimiter. Since the distances between the consecutive points in $\Sigma_{\varepsilon,\eta}(\beta - 1, \beta)$ are ordered as the mechanical word $\underline{s}_{\varepsilon,\beta}$ then w (i.e. the infinite word to the right from the delimiter) is the Sturmian word. From the construction it is clear that the language of $w_0 w_1 w_2 \dots$ is the same as the language of $\dots w_{-2} w_{-1}$. The complexity function C_w has been defined for the right sided infinite words, however the definition can be naturally extended to left sided infinite words, say $w' = \dots w_{-2} w_{-1}$ as $C_{w'}(n) = \# L_n(w')$. Since $C_w(n) = \# L_n(w) = \# L_n(w') = C_{w'}(n)$ the word $\dots w_{-2} w_{-1}$ is also Sturmian. Since an arbitrary shift of the whole interval $(\beta - 1, \beta)$ (i.e. we are keeping the unit length of the interval) does not change the set $\Sigma_{\varepsilon,\eta}(\beta - 1, \beta)$, it just shifts the position of the delimiter, we can conclude that the word $(w_n)_{n \in \mathbb{Z}}$ is the pointed bidirectional infinite Sturmian word. From now on we will consider only bidirectional infinite words.

Note that the set $\Sigma_{\varepsilon,\eta}(\beta - 1, \beta)$ is called the cut-and-project set and the interval $(\beta - 1, \beta)$ is an acceptance window.

There are a couple of papers dealing with cut-and-project sets. In [18] it is shown that a cut-and-project set has either two or three distances between adjacent points; two distances correspond to the case of the unit length of the acceptance window and the distances form a Sturmian word. On the other hand, the words corresponding to the cut-and-project sets with three distances are exactly those which arise from the coding of the 3-interval exchange map, and vice versa. In [19], [20] the authors study substitution properties and the substitutivity of cut-and-project sets.

3 Sturmian words and substitutions

Let us take a look at Sturmian words from a different point of view. One can see in Example 2.1 that there exist Sturmian words which are generated by certain maps (here we will call them substitutions). In fact, the mentioned Fibonacci word is a fixed point of the map φ . The question is whether this is a general property of all Sturmian words, or whether there exists a class of Sturmian words invariant under substitutions. Let us now state basic definitions and then we give an overview of the most interesting results.

A morphism φ is a map of A^* into itself satisfying $\varphi(uv) = \varphi(u)\varphi(v)$ for each $u, v \in A^*$. The morphism is called non-erasing if $\varphi(a_i)$ is not an empty word for any $a_i \in A$. A non-erasing morphism φ is called a substitution.

The action of φ can be extended to bidirectional infinite words

$$(w_n)_{n \in \mathbb{Z}} = \dots w_{-2} w_{-1} | w_0 w_1 w_2 \dots$$

by

$$\dots \varphi(w_{-2}) \varphi(w_{-1}) | \varphi(w_0) \varphi(w_1) \varphi(w_2) \dots$$

We say that the word $(w_n)_{n \in \mathbb{Z}}$ is invariant under the substitution φ (or is a fixed point of φ) if

$$\dots \varphi(w_{-2}) \varphi(w_{-1}) | \varphi(w_0) \varphi(w_1) \varphi(w_2) \dots = \dots w_{-2} w_{-1} | w_0 w_1 w_2 \dots$$

Suppose that we have a substitution φ and there exist letters $a_i, a_j \in A$, such that $\varphi(a_i) = ua_i$ and $\varphi(a_j) = a_jv$ for some non-empty words $u, v \in A^*$. Then by repeated application of φ on the pair $a_i | a_j$ of letters separated by the delimiter $|$ we obtain words $\varphi^{(n)}(a_i) | \varphi^{(n)}(a_j)$, $n \in \mathbb{N}$ of increasing length.

Clearly

$$\varphi^{(n)}(a_i) | \varphi^{(n)}(a_j) = u_n \varphi^{(n-1)}(a_i) | \varphi^{(n-1)}(a_j) v_n,$$

for certain words $u_n, v_n \in A^*$. The limit of $\varphi^{(n)}(a_i) | \varphi^{(n)}(a_j)$ for $n \rightarrow \infty$ is an infinite bidirectional word $(w_n)_{n \in \mathbb{Z}}$ and we say that φ generates the word $(w_n)_{n \in \mathbb{Z}}$.

Let us define a weaker notion of a substitutive word. We say that $(w_n)_{n \in \mathbb{Z}} = \dots w_{-2} w_{-1} | w_0 w_1 w_2 \dots$ is a substitutive word, if there exists a substitution φ on an alphabet B with a fixed point $(v_n)_{n \in \mathbb{Z}} = \dots v_{-2} v_{-1} | v_0 v_1 v_2 \dots$ and a map $\chi: B \mapsto A$ such that $w_n = \chi(v_n)$, for each $n \in \mathbb{Z}$. Note that all fixed points of a substitution are substitutive. The opposite is not true.

Let $A = \{a_1, \dots, a_k\}$ be an alphabet. To a substitution φ one may assign a substitution matrix $\mathbf{A} \in \mathbb{N}^{k \times k}$ in the following way:

$\mathbf{A}_{ij} :=$ number of letters a_j in the word $\varphi(a_i)$.

The problem of invariance under a substitution (or the weaker notion of substitutivity) has motivated many papers. There are some partial results, where authors consider only one sided Sturmian words or characterize substitution invariant bidirectional Sturmian words depending on the slope α if the intercept $\beta = 0$.

Crisp et al. [10] carried on the work of Brown [21] and studied substitution invariant cutting sequences. They proved that the cutting sequence \mathbf{S}_ρ (resp. the mechanical sequence c_α) is substitution invariant if and only if the continuous fraction expansion of ρ (resp. α) has a certain form. The author in [22] used some of their results and simplified their condition on the invariance of the characteristic sequence c_α under a substitution. He showed that c_α , $\alpha \in (0, 1)$, is invariant under a substitution φ if and only if α is a quadratic irrational with conjugate $\alpha' \notin (0, 1)$. Such α is called a Sturm number. This result was shown independently by Allauzen, [23].

Parvaix [24] proved that bidirectional non-pointed Sturmian words with $\beta \neq 0$ are invariant under α substitution if and only if α is a Sturm number and the intercept β belongs to the quadratic field $Q(a) = \{a + ba' | a, b \in Q\}$.

A complete characterization of infinite one-sided substitution invariant Sturmian words was done by Yasutomi [25]. Berthé et al. [26] studied infinite words which arise from the coding of the 2-interval exchange map and gave an alternative proof of Yasutomi's result using Rauzy fractals associated with invertible primitive substitutions. The authors also de-

finied for every fixed Sturm number α a matrix $\mathbf{M}_\alpha \in \mathbb{Q}^{2 \times 2}$ that is called the generating matrix of α and is closely related to the smallest solution of a Pell equation. They showed that a Sturmian sequence $s_{\alpha, \beta}$ (α Sturm number, $\beta \in [0, 1)$) is a fixed point of a substitution with substitution matrix \mathbf{A} if and only if \mathbf{A} has the form \mathbf{M}_α^l , for some $l \geq 1$.

We complete this overview by giving a few notes on paper [27]. The authors have completely solved the question of the substitution invariance of pointed bidirectional Sturmian words. The main theorem shown in the paper is the following.

Theorem 3.1: *Let α be an irrational number, $\alpha \in (0, 1)$, $\beta \in [0, 1)$.*

The pointed bidirectional Sturmian word with slope α and intercept β is invariant under a non-trivial substitution if and only if the following three conditions are satisfied:

1. α is a Sturm number,
2. $\beta \in Q(\alpha)$,
3. $\alpha' \leq \beta' \leq 1 - \alpha'$ or $1 - \alpha' \leq \beta' \leq \alpha'$, where β' is the image of β under the Galois automorphism of the quadratic field $Q(\alpha)$.

Note that this result is analogous to those derived in [25] and [26] for the one-sided case.

The proof presented in [27] is constructive based on the cut-and-project scheme that was sketched in the paragraph devoted to bidirectional words. This approach has not been used yet in the study of substitution invariant Sturmian words. It turns out to be a good choice because the proof is simple. One of the advantages of the cut-and-project scheme is that the more difficult parts of the proof can be illustrated on vivid examples, which makes the whole paper more comprehensible.

We believe that methods similar to those in [27] can solve the question of the substitution invariance of words over a 3 letter alphabet, which arise from the coding of the 3-interval exchange map.

4 Open problems

We have summarized several interesting properties of Sturmian words and here we would like to highlight how the results about Sturmian words can help in further advances in the field of aperiodic words. As Sturmian words are the simplest aperiodic structures, the question of generalisation of results obtained for Sturmian words is particularly interesting. The direct generalization of Sturmian words are infinite words which arise from the coding of a 3-interval (resp. r -interval) exchange map. We believe that results and techniques used in the study of Sturmian words can be applied with success to open problems connected with words coding a 3-interval exchange map. Below you may find a list of the most interesting issues.

1. A necessary and sufficient condition on the substitution invariance of the words coding a 3-interval exchange map is still not known. The problem of finding a similar condition to that of Theorem 3.1 is still open, but the techniques used in [27] may bring some answers.
2. Giving a description of substitution matrices of substitutions which generate the words coding a 3-interval exchange map is a problem closely related to the previous one. The question of finding generating matrices of such

substitutions seems to be very challenging. This issue was solved in [26] for Sturmian words.

3. Although a description of palindromes in the words coding a 3-interval exchange map is already known [17], we believe that the use of the cut-and-project scheme can give an alternative proof of this result. Nevertheless the palindromic complexity is described only for a 2-interval exchange map (see theorem 2.3) and a 3-interval exchange map [28], but we believe that a general description can be obtained by considerations based on the cut-and-project scheme.

5 Acknowledgment

The author acknowledges partial funding from the Czech Science Foundation GA, ČR 201/05/0169. The author is also very obliged to Edita Pelantová and Zuzana Masáková for their help and discussions on the paper.

References

- [1] Shechtman, D. et al.: "Metallic Phase with Long-Range Orientational Order and No Translational Symmetry." *Phys. Rev. Lett.* Vol. **53** (1984), p. 1951–1953.
- [2] Bombieri, E., Taylor, J. E.: "Which Distributions of Matter Diffract? An Initial Investigation." *J. Phys. Colloque C3*, Vol. **14** (1986), p. 19–28.
- [3] Guimond, L. S., Patera, J.: "Proving the Deterministic Period Breaking of Linear Congruential Generators Using Two Tile Quasicrystals." *Math. of Comput.*, Vol. **71** (2002), p. 319–332.
- [4] Hedlund, G., Morse, M.: "Symbolic Dynamics II: Sturmian Sequences." *Amer. J. Math.*, Vol. **61** (1940), p. 1–42.
- [5] Venkov, B. A.: *Elementary Number Theory*, Groningen: Wolters-Noordhoff, 1970.
- [6] Bernoulli, J.: "Sur une nouvelle espece de calcul." *Receuil pour les astronomes*, Vols. **1, 2**, Berlin, 1772.
- [7] Markoff, A. A.: "Sur une question de Jean Bernoulli." *Math. Ann.*, Vol. **19** (1882), p. 27–36.
- [8] Arnoux, P., Rauzy, G.: "Représentation géométrique de suites de complexité $2n+1$." *Bull. Soc. Mat. France*, Vol. **119** (1991), p. 199–215.
- [9] Coven, E. M.: "Sequences with Minimal Block Growth II." *Math. Sys. Theory*, Vol. **8** (1973), p. 376–382.
- [10] Crips, D. et al.: "Substitution Invariant Cutting Sequences." *J. de Th. des Nom. de Bordeaux*, Vol. **5** (1993), p. 123–137.
- [11] Adamczewski, B.: "Codages de rotations et phénomènes d'autosimilarité." *J. de Th. des Nom. de Bordeaux*, Vol. **14** (2002), p. 351–386.
- [12] Alessandri, B., Berthé, V.: "Three Distance Theorems and Combinatorics on Words." *L'Enseignement Mat.*, Vol. **44** (1998), p. 103–132.
- [13] Ferenczi, S. et al.: "Structure of Three Interval Exchange Transformations I: an Arithmetic Study." *Ann. Inst. Fourier*, Vol. **51** (2001), p. 861–901.
- [14] Lunnon, W. F., Pleasants, P. A. B.: "Characterization of Two-Distance Sequences." *J. Austral. Math. Soc.*, Vol. **53** (1992), p. 198–218.
- [15] Vuillon, L.: "A Characterization of Sturmian Words by Return Words." *European J. Combin.*, Vol. **22** (2001), p. 263–275.
- [16] Droubay, X., Pirillo, G.: "Palindromes and Sturmian Words." *Theoret. Comput. Sci.*, Vol. **223** (1999), p. 73–85.
- [17] Droubay, X. et al.: "Episturmian Words and some Constructions of de Luca and Rauzy." *Theoret. Comput. Sci.*, Vol. **225** (2001), p. 539–553.
- [18] Guimond, L. S. et al.: "Combinatorial Properties of Infinite Words Associated with Cut-and-Project Sequences." *J. de Th. des Nom. de Bordeaux*, Vol. **15** (2003), p. 697–725.
- [19] Baláži, P.: "Infinite Words Coding 3-Interval Exchange." Diploma work, 2003.
- [20] Masáková, Z. et al.: "Substitution Rules for Aperiodic Sequences of Cut and Project Type." *J. of Phys. A: Math. Gen.*, Vol. **33** (2000), p. 8867–8886.
- [21] Brown, T. C.: "A Characterization of the Quadratic Irrationals." *Canad. Math. Bull.*, Vol. **34** (1991), p. 36–41.
- [22] Baláži, P.: "Jednoduchá charakteristika Sturmových čísel." Proc. of 9th Math. Conf. of Tech. Uni., 2001, p. 5–15.
- [23] Allauzen, C.: "Une caractérisation simple des nombres de Sturm." *J. de Th. des Nom. de Bordeaux*, Vol. **10** (1998), p. 237–241.
- [24] Parvaix, B.: "Substitution Invariant Sturmian Bisequences." *J. de Th. des Nom. de Bordeaux*, Vol. **11** (1999), p. 201–210.
- [25] Yasutomi, S.: "On Sturmian Sequences which are Invariant under some Substitutions." *Number Theory and its Applications*, Vol. **2** (1999), p. 347–373.
- [26] Berthé, V. et al.: "Invertible Substitutions and Sturmian Words: An Application Of Rauzy Fractals, preprint 2003.
- [27] Baláži, P. et al.: "Complete Characterization of Substitution Invariant Sturmian Sequences." To appear in *Integers*, 2005.
- [28] Damanik D., Zamboni L. Q: *Arnoux-Rauzy Subshifts: Linear Recurrences, Powers, and Palindromes*, arXiv: math.CO/0208137v1.
- [29] Berstel, J.: "Recent Results on Extensions of Sturmian Words." *Internat. J. Algebra Comput.*, Vol. **12** (2002), p. 371–385.

Ing. Peter Baláži
e-mail: balazi@kml.fjfi.cvut.cz

Department of Mathematics

Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering
Trojanova 13
120 00 Prague 2, Czech Republic