Balanced words

Laurent Vuillon

Abstract

This article presents a survey about balanced words. The balance property comes from combinatorics on words and is used as a characteristic property of the well-known Sturmian words. The main goal of this survey is to study various generalizations of this notion with applications and with open problems in number theory and in theoretical computer science. We also prove a new result about the generalized balance property of hypercubic billiard words.

1 Introduction

The balance property is a fine tool for studying words appearing in combinatorics on words [12, 26, 15], in billiard theory [70] and dynamical systems [20]. Two finite words of the same length on the alphabet $\{a, b\}$ have the balance property if the number of a's in the two words is almost the same. More precisely the difference between the number of a's in the two words is bounded by 1. By definition, an infinite word x is balanced if for any two finite words factors of x, the two words have the balance property.

This survey is motivated by the increasing number of results on balanced words. First of all, we have the following characterization. Infinite non-periodic balanced words on a binary alphabet are exactly Sturmian words [52, 12]. Furthermore, balanced words on more than two-letter alphabets appear in the statement of the famous Fraenkel conjecture with links to Beatty sequences and to number theory [36, 68]. Sometimes generalizations of Sturmian words are strongly non-balanced (the difference between the number of letter a in the two words is not bounded) as for a subclass of Arnoux-Rauzy words [26]. Conversely hypercubic billiards words

Key words and phrases : Balanced words, Sturmian words, complexity function, combinatorics on words, Beatty sequences, Arnoux-Rauzy words, hypercubic billiard words, heap of pieces, optimization problems.

(see [4, 70]) in dimension d are $(d-1)$-balanced; that is the difference between the number of a's in the two words is bounded by $d-1$. We also investigate the balance property for bidimensional words [16] and an alternative definition of the balance property [53]. In addition to that the periodic or non-periodic infinite words play an important role because of applications in optimization theory and discrete event systems [38, 51].

This survey is organized as follows. In section 2 we study the balance property of infinite words and the structure of balanced factors. In particular, we focus on the balance property of Sturmian words and we also construct balanced words in a geometrical way, namely the cutting words. We also state a theorem of Berstel and Séébold about Sturmian morphisms. We study the generalized balance property of Sturmian words and the balance property for words on alphabets of more than two letters. We present the Fraenkel conjecture. In section 3 we study results on the generalized balance property. We define the Arnoux-Rauzy words and the imbalance property result. We also recall the construction of billiard words in square and cubic tables. We show a new result of generalized balance property of hypercubic billiard words. We construct an example of hypercubic billiard words with maximal balance property. In the last subsection we show that the balance property in two dimensions is very restrictive and we define an alternative definition of the balance property. To close this survey, Section 4 is a presentation of heap of pieces and optimization problems in relation with balanced words.

2 Balance property of infinite words

2.1 Sturmian words

An *infinite word* x over a finite alphabet $\mathcal{A}$ is a mapping from the positive integers into $\mathcal{A}$. We write $x = x_1x_2\cdots$ where $x_i \in \mathcal{A}$ is the ith letter of x. A *factor* of x is word w such that $w = x_ix_{i+1}\cdots x_j$ for some i, j with $i \leq j$ and the length of w is $|w| = j-i+1$. The empty word is the word of length 0.The set of factors is denoted by $L(x)$ and the set of factors of length $n \geq 0$ is denoted by $L_n(x)$.

Hedlund and Morse (see [52]) define the *Sturmian words* by the notion of (subword) complexity :

Definition 1. *An infinite word x is a Sturmian word if the complexity of x is given by $p(n) = n+1$ for all $n \geq 0$.*

Here the *complexity function* $p : \mathbb{N} \to \mathbb{N}$ counts the number of distinct factors of length n of the infinite word x, i.e. $p(n) =$ Card $L_n(x)$ (see the survey [3]). That is, a Sturmian word contains exactly $n+1$ distinct factors of length n for all n. In particular, this definition implies that the Sturmian words are built on an alphabet with $p(1) = 2$ letters.

For example, here you find the beginning of a Sturmian word:

$$x = abaaabaabaaabaaabaabaaabaabaaabaaabaab\cdots$$

It is easy to check the first values of the complexity function of x:

$$p(1) = \text{ Card } \{a, b\} = 2,$$

$$p(2) = \text{ Card } \{ab, ba, aa\} = 3,$$

$$p(3) = \text{ Card } \{aba, baa, aaa, aab\} = 4,$$

$$p(4) = \text{ Card } \{abaa, baaa, aaab, aaba, baab\} = 5 \cdots .$$

This combinatorial definition of Sturmian words leads to many characterizations [50, 10]. There are two synthesis references on Sturmian words, a chapter in the Lothaire 2 [12] and the book of the Marseille group [20].

2.2 Balance property

The following characterization of Sturmian words is important because it leads to many generalizations used in computer science.

Theorem 1 (Hedlund, Morse). *A non-periodic infinite word x is a Sturmian word if and only if for all factors w, w' of x of the same length, the difference between the number of a in w and the number of a in w' is bounded by* 1*.*

More formally, $\forall n \in \mathbb{N}, \forall w, w' \in L_n(x)$ $||w|_a - |w'|_a| \leq 1$ where $|w|_a$ denotes the number of distinct occurrences of the letter a in the word w. This characterization is based on the balance property of the infinite word x. By definition a finite or infinite word u is *balanced* if for all factors w and w' of same length of u we have $||w|_a - |w'|_a| \leq 1$.

Recent results of Jenkinson and Zamboni present three characterizations of finite balanced words w that can be extended on an infinite periodic balanced word w^ω in terms of ordering of an orbit either lexicographically or with respect to a norm [44]. In an article of O'Bryant there is a fine study of Sturmian words using algebraic properties of permutation that orders fractional parts [23]. This paper extends the results in [66, 2] on the three-distance theorem.

A geometrical method of producing balanced words is via cutting words. Cutting words on regular tilings are codings of natural geometrical objects appearing in billiard theory [4, 70], combinatorics on words [11, 28, 43] and dynamical systems [20]. The simplest case is given by a half-line with slope α in a unit square grid. This geometrical object is coded in order to build an infinite word u called a *cutting word*: we code the intersection of a given half-line with the unit square grid by the letter a (resp. b) if the intersection follows in vertical (resp. horizontal) line (see Figure 1).

Note that this coding is not well-defined on points with both integer coordinates. On these points we could choose either the coding ab or the coding ba. But in order to have balanced words we must code each intersection of the half-line and integer points by the same coding ([38, 51, 12]). For example, if we take the half-line D $y = x, x \geq 0$ then D intersects the unit grid only at integer points. We have to make the same choice for each intersection of D with integer points. If we choose ab then

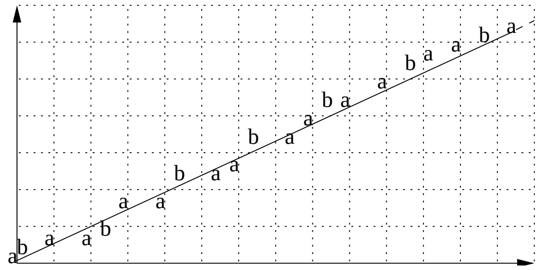

Figure 1: Cutting word.

the cutting word is $(ab)^\omega$ and if we choose ba then the cutting word is $(ba)^\omega$. And we can check that both infinite words are balanced words.

Of course, if α is rational then the preceding construction gives periodic balanced words. In order to construct non-periodic balanced words, we take α irrational and then the coding gives exactly the Sturmian words.

2.3 Iterated morphisms and balance property

A classical method of generating infinite words in computer science and in mathematics is the use of iterated morphisms (or substitutions in mathematician vocabulary) (see [29, 64, 55]). A *morphism* letter to word associates to each letter a word on a given alphabet. For example, the *Fibonacci morphism* on the two letter alphabet $\mathcal{A} = \{a, b\}$ is defined by $\sigma(a) = ab$ and $\sigma(b) = a$ (see for example [8]). The following rule is used to apply the morphism: if w is a finite word on $L(x)$ and $w = w_1 w_2 \cdots w_n$ where the w_i's are letters then $\sigma(w) = \sigma(w_1 w_2 \cdots w_n) = \sigma(w_1)\sigma(w_2)\cdots\sigma(w_n))$. In addition to that, one denotes by $\sigma^n(a)$, n iterations of the morphism applied to the letter a. To illustrate this definition, $\sigma^2(a)$ is equal to $\sigma(\sigma(a)) = \sigma(ab) = aba$.

The *Fibonacci word* is defined as the fixed point of σ, that is the infinite word x such that $\sigma(x) = x$ [9, 64]. Here you find the beginning of the Fibonacci word:

$$x = abaababaabaababaababaabaabab\cdots$$

In fact, this morphism is called the Fibonacci morphism because the length of the finite word $\sigma^n(a)$ is equal to the Fibonacci number F_n. Note that the $\sigma^n(a)$ are prefixes of the Fibonacci word for all n. This morphism is then related to the linear recurrence $F_{n+2} = F_{n+1} + F_n$ with $F_{-1} = 1$ and $F_0 = 1$.

Sturmian morphisms are morphisms which leave invariant the class of Sturmian words [11]. That is a Sturmian morphism applied to any Sturmian word gives another Sturmian word. It is remarkable that we can check if a morphism is a Sturmian morphism by studying the balance property of the image of a finite word.

Theorem 2 (Berstel, Séébold). *A morphism ϕ is a Sturmian morphism if and only if the image of $W = abbabbababbaba$ by ϕ is a balanced word.*

This means that Berstel and Séébold check a property for an infinite class of morphisms on a word of length 14 (see [11]).

For example, the morphism of Fibonacci $\sigma(a) = ab$ and $\sigma(b) = a$ is a Sturmian morphism. Indeed $\sigma(W) = abaaabaaabaabaaabaab$ is a balanced word. To verify this property, it is sufficient to take all factors two by two with the same length (for instance $w = baaabaa$ and $w' = baabaab$) and to check that $||w|_a - |w'|_a| \leq 1$ (with the preceding choice of words, we compute $|baaabaa|_a = 5$ and $|baabaab|_a = 4$ in accordance with the formula).

3 Balance property with a word

With Fagnot [33], we study the balance property of Sturmian word not only with a single letter but with factors of x. We show that for a Sturmian word x the difference between the number of occurrences of a factor u of x in two factors of x of the same length is bounded by the length of u.

Theorem 3 (Fagnot, Vuillon). *If x is a Sturmian word then*

$$\forall u \in L(x) \textit{ with } |u| \geq 1, \forall n \in \mathbb{N}, \forall w, w' \in L_n(x),$$

$$||w|_u - |w'|_u| \leq |u|$$

where $|w|_u$ denotes the number of distinct occurrences of the word u in the word w.

We fulfil the balance property of Sturmian word by taking for u a word of length 1 (that is a letter of the alphabet). The article precises that if the continued fraction expansion of the slope α of the cutting word has bounded partial quotients then we find that the difference $||w|_u - |w'|_u|$ is uniformly bounded on the length of w.

The main interest of this approach is to understand well the structure of Sturmian words. This notion of balance property should be extended to more complicated phenomena like the balance property on infinite words in alphabets of more than a two letters.

In an article of Adamczewski, we find a deep link between generalized balance property of infinite words given by iterated morphisms and discrepancy function [1, 55]. More precisely, in the case of fixed points of primitive substitutions (primitive means that there exists an integer k such that all letters of the alphabet $\mathcal{A}$ appear in $\sigma^k(a), \forall a \in \mathcal{A}$) the maximal balance ($\max_{w,w' \in L_n(x)}(||w|_a - |w'|_a|)$ [1, 33]) is in part ruled by the spectrum of the incidence matrix associated with the substitution. The article presents also an interesting catalogue of substitutions and their spectra. In particular Adamczewski shows that the two notions of balance and discrepancy are strongly connected in case of linearly recurrent words (a word is said linearly recurrent if there exists an integer K such that for any of its factor w the difference between two successive occurrences of w is bounded by $K|w|$ see [31]). He gives also a link between bounded remainder sets that appear on the works of Rauzy and on the famous Rauzy fractal [59, 58] and generalized balance property.

4 Balance property for an alphabet of more than two letters

The preceding generalization of the balance property is defined on Sturmian words and then gives a definition for a two-letter alphabet. A natural way to extend the balance property is to consider words on an alphabet of more than two letters. By definition a word is balanced on each letter if for all letters of the alphabet $\mathcal{A} = L_1(x)$ we have

$$\forall a \in L_1(x), \forall n \in \mathbb{N}, \forall w, w' \in L_n(x), ||w|_a - |w'|_a| \leq 1.$$

A result of Graham [40] on covering of integers by sequences and a result of Hubert [42] using combinatorics on words show that infinite non-periodical words balanced on each letter are constructed by a modification of Sturmian words. The main idea is to periodically replace the occurrences of the letter a of the Sturmian word by a periodic word on an alphabet A and to replace the occurrences of b by a periodic word on another alphabet B. The conditions in order to have non-periodic words balanced on each letter are firstly that A and B are disjoint and secondly that the words $(a_1, a_2, \cdots a_{k_1})^\omega$ with $a_i \in A$ and $(b_1, b_2, \cdots b_{k_2})^\omega$ with $b_i \in B$ are with constant gaps. An infinite periodic word w^ω has *constant gaps* if the number of letters between two occurrences of successive letter a_i of w^ω is constant for each i). For example $(abac)^\omega$ is a constant gap word and $(abaac)^\omega$ is not with constant gaps.

As an example we build a non-periodic word on a four-letter alphabet by modification of the Fibonacci word:

$$x = abaababaabaababaababaabaabab \cdots$$

To do this, we replace periodically the occurrences of the letter a by the constant gap word $(cdce)^\omega$:

$$cbdcbebcdbcebcbdcbebcdbcebcb \cdots$$

By construction, the word is balanced on the letter b, because the Fibonacci word is balanced. It remains as an exercise for the reader to check that the word is balanced on the letter c, d and e.

In fact, Graham presents is result using covering of integers by Beatty of the form $\lceil \alpha n + \beta \rceil$. He extends the following Theorem [65, 36, 68]:

Theorem 4 (Skolem-Fraenkel). *The Beatty sequences $\lceil \alpha_1 n + \beta_1 \rceil$ and $\lceil \alpha_2 n + \beta_2 \rceil$ cover the integer if and only if*

$$\frac{1}{\alpha_1} + \frac{1}{\alpha_2} = 1 \textit{ and } \frac{\beta_1}{\alpha_1} + \frac{\beta_2}{\alpha_2} \in \mathbb{Z}.$$

Skolem proves the theorem for α_1 irrational (for non-periodic balanced words) and Fraenkel for α_1 rational (periodic balanced words). The link between Beatty sequences and Sturmian words is simple. Indeed, the first Beatty sequence gives the indices of occurrence of the letter a and the second Beatty sequence gives the indices of occurrence of the letter b in the associated Sturmian word. For example, the Fibonacci word is given by

$$x = abaababaabaababaababaabaabab \cdots ,$$

and the first Beatty sequence gives the following set of indices

$$\{1, 3, 4, 6, 8, 9, 11, 12, 14, 16, 17, 19 \cdots\}$$

and the second

$$\{2, 5, 7, 10, 13, 15, 18 \cdots\}.$$

The union of the two sets is the positive integers $\mathbb{N}^*$. That is the two Beatty sequences cover the integers. Note that the theorem of Skolem-Fraenkel gives Beatty sequences associated with Sturmian words if α_1 is irrational and associated with periodic balanced words if α_1 is rational.

Now, if we cover the integers by three or more Beatty sequences $\lceil \alpha_i n + \beta_i \rceil$ with $i = 1, 2, \cdots k, k \geq 3$ with all distinct frequencies of letters (i.e. α_i two by two distinct), then Graham shows that the coefficients α_i remains rational. This implies in particular that the associated infinite word is periodic and balanced.

In computer science, this kind of balance property appears in optimization problems and in particular for job-shops. If we consider k tasks that share the same machine, we can find a solution in order to minimize the utilization of the machine by using a periodic word on k letters to dispatch the information on the k tasks. Gaujal [39] studies this problem by using a discrete version of convexity, namely the multi-modularity property.

A conjecture of Fraenkel [36, 68, 69, 39] says that this problem is very constrained. First we define the *frequency* of the letter $a \in \mathcal{A}$ in the infinite word x is defined by $\lim_{n\to\infty} |x_1 x_2 \cdots x_n|_a / n$ when the limit exists.

Conjecture 1 (Fraenkel). *The unique solution (up to a permutation of letters) of balanced word on each the $k \geq 3$ letters with all distinct frequencies of letters is $(FR_k)^\omega = (FR_{k-1} k FR_{k-1})^\omega$ where $FR_3 = 1213121$.*

This conjecture is true for $k = 3, 4, 5, 6$ [68, 69, 39]. In particular for $k = 3$ the unique word is $(1213121)^\omega$ and the frequencies are all different, the frequency of the letter 1 is $\frac{4}{7}$ because there are 4 occurrences of the letter 1 in the period, the frequency of the letter 2 is $\frac{2}{7}$ and the frequency of the letter 3 is $\frac{1}{7}$.

5 Generalizations of balance property

5.1 Infinite words with complexity $p(n) = 2n + 1$

The infinite words with complexity $p(n) = 2n + 1$ are a well studied generalization of Sturmian words [10]. Numerous constructions exist for building these words by coding of rotations [60, 35, 13], interval exchange transformations [5, 47, 34, 57], combinatorics on words [27, 30, 26] or by iterated morphisms [5, 54]. Remark that to find a complete characterization of infinite words with complexity $p(n) = 2n + 1, \forall n \geq 0$ is still an open problem.

A Sturmian word has complexity $p(n) = n+1$ and then the first difference of the complexity function $p(n+1)-p(n)$ is equal to 1 for all n. In terms of combinatorics on words, for each length n there exists a unique word with two right prolongations i.e. for all n there exists w with $|w| = n$ such that wa and wb are factors of x and a unique word with two left prolongations i.e. for all n there exists w' with $|w'| = n$ such that aw' and bw' are factors of x.

The Arnoux-Rauzy words are built using this property on a three-letter alphabet: for each length there exists a unique word with three right prolongations and a unique word with three left prolongations. Thus the complexity function for Arnoux-Rauzy words is $p(n) = 2n + 1$. This generalization can also be seen as a coding of interval exchange transformation on 6 intervals [5].

The prototypical Arnoux-Rauzy word is given by the following iterated morphism: $\sigma(a) = ab, \sigma(b) = ac, \sigma(c) = a$ with the beginning

$$x = abacabaabacababacabaabacabacaba\cdots$$

We can check that this word is not balanced but is 2-balanced on each letter.

Definition 2. *An infinite word x is a c-balanced word on each letter if* $\forall a \in \mathcal{A}, \forall n \in \mathbb{N}, \forall w, w' \in L_n(x)$ $||w|_a - |w'|_a| \leq c$.

Unfortunately, one can find Arnoux-Rauzy words which are not c-balanced for any c.

Theorem 5 (Cassaigne, Ferenczi, Zamboni). *There exist Arnoux-Rauzy words that are not c-balanced, that is for each integer c, there exist factors w and w' with same length of an Arnoux-Rauzy word x such that $||w|_a - |w'|_a| \geq c$ for some letter a.*

Furthermore, if the Arnoux-Rauzy word x is linearly recurrent (if there exists an integer k such that for any factors w the difference between the indices of two successive occurrences of w in x is bounded by $k|w|$) then x is c-balanced for some c. Justin and his co-authors also give a combinatorial study of Arnoux-Rauzy words and of a natural generalization of these words, namely the Episturmian words [30, 45, 46].

5.2 Billiards

Now, we investigate the world of billiards [70]. Indeed, Sturmian words are characterized by coding of square billiard words with irrational slope. Let us consider a square billiard table and a trajectory of a point along an irrational slope. Each time that the point touches the border of the billiard table , it bounces according to the reflection laws (the angles of reflection of the trajectory with the normal at the border are equal before and after the reflection). Notice that with this definition the trajectory is not defined at corner points. Now, if we code by a (resp. by b) when the point touches the horizontal (resp. vertical) sides then the infinite word given by the coding of the trajectory is a Sturmian word.

A natural generalization of square billiards is cubic billiards. We play billiard on a cube with trajectory along a totally irrational direction (a direction (a_1, a_2, a_3)

such that for b_j integers we have $b_1a_1+b_2a_2+b_3a_3=0$ if and only if $b_1=b_2=b_3=0$). We code by i when the trajectory bounces on the side normal to e_i. By this coding we build cubic billiard words [4].

Theorem 6 (Arnoux, Mauduit, Shiokawa, Tamura). *The complexity function of a billiard word is* $p(n)=n^2+n+1$.

Each cubic billiard word is a mix of three Sturmian words and also a cutting word in the unit cubic grid. Indeed, cubic billiard words are words on a three-letter alphabet $\{1,2,3\}$. Now, if we erase all letters 1 on a cubic billiard word (or letter 2 or 3), then it remains a Sturmian word (geometrically, a trajectory on a cubic billiard seen in a direction normal to one side is a square billiard trajectory).

For example, this is the beginning of a billiard sequence:

$$231232132312321322312312321323123213223123\cdots$$

And here you find the three associated Sturmian words:

$$2323232323232232323232323232322323\cdots$$

$$31313313133131313313133 13\cdots$$

$$21221212212212121221212212212\cdots$$

Conversely, if we consider three Sturmian words S_1 in the alphabet $\{2,3\}$, S_2 in the alphabet $\{1,3\}$ and S_3 in the alphabet $\{1,2\}$ there is an easy pseudo algorithm to check and to construct the associated cubic billiard word C. At each step we choose the two letters equal in two of the Sturmian words and we add this letter to the cubic billiard word C. Here you find the pseudo algorithm that is to have a cubic billiard word, it must stay infinitely many time on the repeat-until loop. Indeed, as we have three Sturmian words, S_i is the coding of the line D_i for $i=1,2,3$. The Sturmian word S_1 (resp. S_2, S_3) is the coding of a geometrical object on the plane D_1Oz (resp. D_2Ox ,D_3Oy). If the algorithm does not stop then the remaining geometrical object is at the intersection of the three planes D_1Oz , D_2Ox and D_3Oy and must be a line. Thus this is a cutting word in dimension 3 with totally irrational direction and then a cubic billiard.

$i:=1, j:=1, k:=1, \ell:=1$;
repeat
  if $S_1(i)==S_2(j)$ **then** $C(\ell):=3, i:=i+1, j:=j+1, \ell:=\ell+1$;
  if $S_1(i)==S_3(k)$ **then** $C(\ell):=2, i:=i+1, k:=k+1, \ell:=\ell+1$;
  if $S_2(j)==S_3(k)$ **then** $C(\ell):=1, j:=j+1, k:=k+1, \ell:=\ell+1$;
until $S_1(i)\neq S_2(j)$ **and** $S_1(i)\neq S_3(k)$ **and** $S_2(j)\neq S_3(k)$;
failure this is not a cubic billiard word.

In the article [4], the authors conjecture an elegant formula. Baryshnikov shows that this conjecture is true in any dimension and for any length of words [6].

Theorem 7 (Baryshnikov). *The complexity function of hypercubic billiard words of dimension $s+1$ and for factors of length n is:*

$$p(n,s) = \sum_{i=0}^{\inf(n,s)} \frac{n!s!}{(n-i)!i!(s-i)!}.$$

We have to notice the symmetric role in the formula between the dimension of the space and the length of the words. An open problem proposed by Mauduit is to find a combinatorial proof of this fact $p(n,s) = p(s,n), \forall (n,s) \in \mathbb{N}^2$.

6 Generalized balance property of hypercubic billiards

We now prove a new balance property for hypercubic billiard words.

A way to define hypercubic billiards is to unfold the trajectory and to consider cutting words on a hypercubic grid. We define a word x associated with hypercubic billiard in dimension d of angle vector $\alpha = (\alpha_1, \alpha_2, \cdots \alpha_d)$ with rationally independent values (i.e. the unique solution of $b_1\alpha_1 + b_2\alpha_2 + \cdots + b_d\alpha_d = 0$ with b_i integers is $b_1 = b_2 = \cdots = b_d = 0$) and of starting point $\beta = (\beta_1, \beta_2, \cdots, \beta_d)$ where $0 \leq \beta_i < \alpha_i, \forall_i$.

A hypercubic billiard word x is an infinite word on the alphabet $\mathcal{A} = \{1, 2, \cdots, d\}$ given by

$$\cup_{i=1}^{d} \{n\alpha_i + \beta_i | n \in \mathbb{N}\}$$

in its natural order and the corresponding sequence of labels read from the origin. Note that for $d = 2$ we have Sturmian words (see Figure 2) and for $d = 3$ we have cubic billiard words.

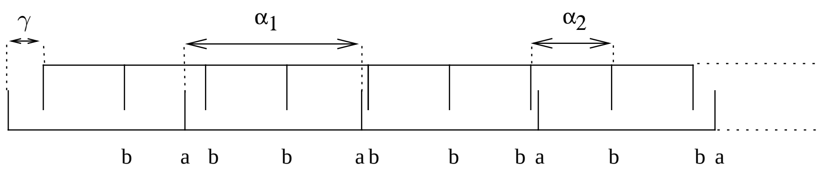

Figure 2: Cutting word.

Theorem 8. *Let x be an infinite word associated with hypercubic billiards in dimension d of angle α and starting point β. Then the infinite word x is $(d-1)$-balanced on each letter (i.e. $\forall i \in \mathcal{A}, \forall n \in \mathbb{N}, \forall w, w' \in L_n(x), ||w|_i - |w'|_i| \leq d-1$).*

Proof. Let bal be a positive integer. Let (w, w') a pair of words such that w and w' are factors of same length of x and such that $|w|_i - |w'|_i = bal$.

We suppose also that the length of w and w' is the smallest integer such that a pair of words of same length is bal-balanced.

Without loss of generality, we can suppose that $w = izi$ and $w' = cz'd$ where c, d are letters of the alphabet $\mathcal{A}$ and $c, d \neq i$. Indeed, in general form $w = azb, w' = cz'd$ and since n is the smallest integer for the balance property then $|zb|_i - |z'd|_i = bal - 1$ implies $a = i, c \neq i$ and $|az|_i - |cz'|_i = bal - 1$ implies $b = i, d \neq i$.

The number of i in the finite word w is equal to $|izi|_i = n_i \geq 2$. The distance on the line between the position of the point labelled by the first i and the position of the point labelled by the last i of the word izi is equal to

$$d_i = (n_i - 1)\alpha_i.$$

For the angle α_j with $j \neq i$ the number of points labelled by the letter a_j on a half open interval of length d_i takes two consecutive values:

$$\underline{n_j} = \left\lfloor \frac{d_i}{\alpha_j} \right\rfloor \text{ and } \overline{n_j} = \left\lceil \frac{d_i}{\alpha_j} \right\rceil.$$

Obviously $\underline{n_j} + 1 = \overline{n_j}$. As α_i and α_j are rationally independent then both values are taken. On a word beginning and ending by i with $|izi|_i = n_i$ the minimal number of letters distinct from i is $\sum_{j\neq i} \underline{n_j}$(such word is called w_{min} and has length $minlength = n_i + \sum_{j\neq i} \underline{n_j}$) and the maximal number of letters distinct from i is $\sum_{j\neq i} \overline{n_j}$ (such word is called w_{max} and has length $maxlength = n_i + \sum_{j\neq i} \overline{n_j}$). Then the maximal difference is

$$\sum_{j\neq i} \overline{n_j} - \sum_{j\neq i} \underline{n_j} = \sum_{j\neq i} (\underline{n_j} + 1) - \sum_{j\neq i} \underline{n_j} = d - 1.$$

In other words in the best case if w_{max} contains a factor w' of length $|w_{min}| = minlength$ with exactly $n_i - d + 1$ letters i then the difference $|w_{min}|_i - |w'|_i$ is $d-1$, otherwise the difference $|w_{min}|_i - |w'|_i$ for w' factor of w_{max} and $|w'| = |w_{min}|$is less than $d - 1$. Thus $0 \leq bal \leq d - 1$.

6.1 Hypercubic billiards with maximal generalized balance property

We now construct an example where the bound for the $d - 1$-balance property is reached.

Take the angles $\alpha_j, j = 1, \cdots, d$ with the order $\alpha_j < \alpha_{j'} \Leftrightarrow j < j'$ and $\alpha_1 < \frac{\alpha_2}{d}$. We construct a hypercubic billiard word in dimension d with maximal balance property. Let x be a hypercubic billiard word in dimension d such that $\beta_i = 0, \forall i = 1, \cdots, d$ then by definition of cutting words we can choose that the infinite word begins by the finite word $12 \cdots d$. As the infinite word is recurrent (i.e. each factor appears infinitely many times because the trajectory is dense on the hypercubic billiard, indeed the vector α is totally irrational) the word $12 \cdots d$ appears infinitely many times in x. Furthermore the condition $\alpha_1 < \frac{\alpha_j}{d}$ for each $j \neq 1$ implies that between two consecutive occurrences of the letter j in x the number of 1's is at least d. In particular words $12 \cdots dw12 \cdots d \in L(x)$ can be continued by 1^d. Then $12 \cdots dw12 \cdots d1^d$ is a factor of x. Thus $|w12 \cdots d1^d| = |12 \cdots dw12 \cdots d|$ and the balance on the letter 1 is

$$|w12 \cdots d1^d|_1 - |12 \cdots dw12 \cdots d|_1 = d - 1.$$

To summarize for each length $|12 \cdots dw12 \cdots d|$ with $12 \cdots dw12 \cdots d \in L(x)$ the maximal balance is attained.

6.2 Balance property in double sequences

Another generalization of Sturmian words is to consider bidimensional words on a finite alphabet $\mathcal{A} = \{0, 1\}$ with various Sturmian like properties [17, 18, 19, 24, 25, 32, 61, 62, 63, 72].

Berthé and Tijdeman [16] show that balanced double sequences on a two-letter alphabet are very rare. A double sequence is a sequence indexed by $\mathbb{Z}^2$. The set of rectangular factors of length m and height n of a double sequence x is noted $L_{m,n}(x)$. A double sequence x is balanced if $\forall m, n \in \mathbb{N}, \forall w, w' \in L_{m,n}(x)\ ||w|_1 - |w'|_1| \leq 1$. Define also the frequency α of the letter 1 to be the limit (if it exists) of the number of ones in an $m \times n$ word centered at the origin divided by mn.

Theorem 9 (Berthé, Tijdeman). *For balanced double sequences on a two-letter alphabet the frequency α of the letter 1 is an element of the following finite set:*

$$\alpha \in \left\{0, \frac{1}{5}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{4}{5}, 1\right\}.$$

In the article, they give some examples of double sequences that illustrate the following conjecture:

Conjecture 2 (Berthé, Tijdeman). *Let U be a double sequence on the alphabet $\{0, 1\}$ such that the frequencies of letters exist and are irrational ($f(1) = \alpha \notin \mathbb{Q}$) then the double sequence U is imbalanced $\forall N \exists n \exists w, w' \in L(n, n) s.t. ||w|_1 - |w'|_1| > N$.*

The balanced property drastically restricts the number of solutions in two dimensions than in one dimension. An alternative definition is to impose very restricted conditions on bidimensional words not for each length and height but for certain $m \times n$ words.

Nivat defines a double sequence on a finite alphabet $\mathcal{A} = \{0, 1\}$ to be k-homogeneous for a $m \times n$ window F if whatever the position of the window in the double sequence exactly k ones appear in the window. In other words, $\forall w \in L_{m,n}(x), |w|_1 = k$ and in a balanced form $\forall w, w' \in L_{m,n}(x), ||w|_1 - |w'|_1| = 0$. There is an interesting decomposition theorem (see [53]):

Theorem 10 (Nivat). *Each k-homogeneous double sequence is the disjoint union of k 1-homogeneous double sequences.*

In addition to that, there is a remarkable link between 1-homogeneous double sequences and tilings of the plane [7]. Let F be a finite subset of $\mathbb{Z}^2$. A double sequence is 1-homogeneous for the window F if what ever the position of the window there is exactly 1 ones on it. We have the following theorem (see [53]):

Theorem 11 (Nivat). *A double sequence is 1-homogeneous for F if and only if F tiles the plane.*

7 Application to optimization problems

This model is based on pieces piling up according to a Tetris game. The main problem is to study the behavior of a class of discrete event systems. In particular, this model allows one to build a heap of pieces associated with certain Petri nets. From a theoretical computer science point of view this domain is at the confluence between trace monoid, Petri nets and discrete event systems [37]. The works of Bacelli, Mairesse, Gaubert and Gaujal show the wealth of this domain by using discrete, probabilistic or dynamical system approaches [39]. This area is also close to heap of pieces of Viennot [71].

Heap of pieces allow us to solve particular optimization problems. Let us consider two tasks sharing the same machine. The tasks can be done in parallel under temporal constraints. The main goal is to minimize the occupation time of the machine. To do that a Gantt diagram shows the execution time of each task and can be seen after rotation as a heap of pieces [37]. In this heap the pieces represent the execution time of each task. The heap of pieces is constructed by piling up each piece according to finite or infinite words (see Figure 3). In fact, the minimization of the execution time of the machine is equivalent to constructing the more compact heap of pieces with given pieces.

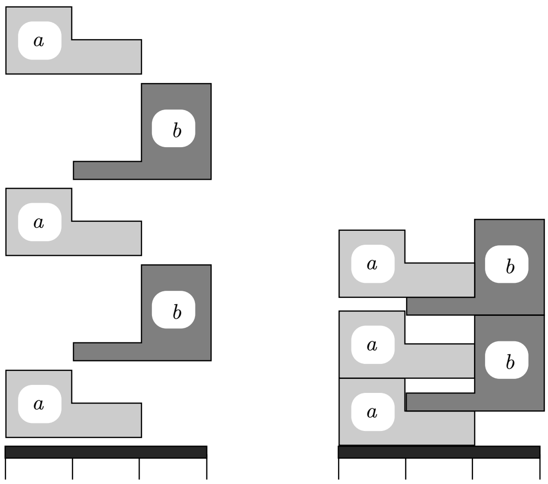

Figure 3: Heap of pieces.

In mathematical terms, we study Liapunov exponents associated with the growth of the heap. More precisely, if we note by $y(w)$ the height of the heap of pieces associated with the finite word w then the optimal growth is given by the following limit $\rho_{\min} = \liminf_n \min_{w \in \mathcal{A}^n} \frac{y(w)}{n}$. An infinite word u will be optimal if $\lim_n \frac{y(u[n])}{n}$ where $u[n]$ is the prefix of length n of u, is equal to $\rho_{\min}$.

Gaujal studies the following job-shop problem : Consider two different tasks, the first one uses a time $\alpha_1 + \alpha_2$ to be done and the second one a time $\beta_1 + \beta_2$. The constraint is that if we do the task 1, we have to wait a time α_1 before beginning the task 2 and a time $\alpha_1 + \alpha_2$ before beginning the task 1. Conversely, if we do the task 2, we have to wait a time β_1 before beginning the task 1 and a time $\beta_1 + \beta_2$ before beginning the task 2. These constraints are given by a Petri net and can be coded on pieces (see the Figure 4). Gaujal shows that the optimal schedule is obtained by pilling up the pieces according to a balanced word (see [38]). We can also construct a Sturmian word (see the Figure 5) by taking α_1 and β_1 rationally independent and $\alpha_2 = \beta_2 = 0$.

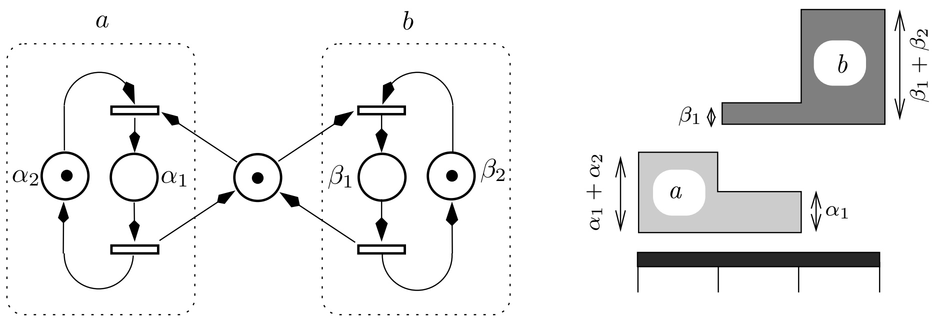

Figure 4: Petri net and pieces.

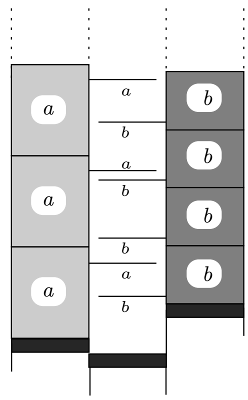

Figure 5: Pieces and Sturmian word.

In fact the result is true for general type of pieces [51].

Theorem 12 (Mairesse, Vuillon). *Let consider a heap of pieces with two pieces. An optimal schedule is obtained by piling up the pieces either by a balanced periodic word or by a balanced non-periodic word.*

Bousch and Mairesse in [21] give another demonstration of this result by using topical forms and Sturmian measures. Furthermore, a result of Gaubert and Mairesse shows that if the pieces are polyominoes with rational heights then the optimal schedule is always periodic (see [37]).

The heaps of pieces are more than the mathematical study of the Tetris game. It can model job-shop problems and also gives geometrical intuition for infinite product of matrices in the (max, +) algebra. In the same spirit the mathematical techniques of Bousch and Mairesse [21] are powerful and led to the refutation of the Lagarias and Wang conjecture [48] which claims that the spectral radius of a usual product of matrices is always reached by the spectral radius of a periodic product of matrices. Once again

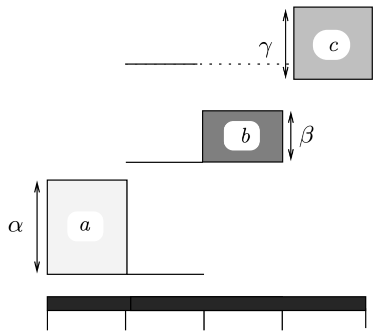

Figure 6: Pieces for cubic billiard.

the existence of the spectral radius of a product of matrices according to a non-periodic balanced word has been the crucial element to refute the conjecture [21].

It remains many works to understand heap of pieces with more than three pieces. For instance, we can build cubic billiard words by piling the pieces in order to have the more compact heap of pieces with the following pieces in Figure 6 and with α, β and γ rationally independent. But the general problem of the behavior of optimal schedule of heap with three pieces is still open.

8 Acknowledgements

We would like to thank O. Jenkinson who read carefully a previous version of this article.

References

[1] B. Adamczewski, *Balances for fixed points of primitive substitutions*, Theoret. Comput. Sci., **307**, (2003), 47–75.

[2] P. Alessandri, V. Berthé, *Three distance theorems and combinatorics on words*, Enseig. Math., **44** (1998), 103–132.

[3] J.-P. Allouche, *Sur la complexité des suites infinies*, Bull. Belg. Math. Soc. **1** (1994), 133–143.

[4] P. Arnoux, C. Mauduit, I. Shiokawa and J.-I. Tamura, *Complexity of sequences defined by billiards in the cube*, Bull. Soc. Math. France, **122** (1994), 1–12.

[5] P. Arnoux and G. Rauzy, *Représentation géométrique de suites de complexité* $2n+1$*,* Bull. Soc. Math. France, **119** (1991), 199–215.

[6] Y. Baryshnikov, *Complexity of trajectories in rectangular billiards*, Comm. Math. Phys., **174** (1995), 43–56.

[7] D. Beauquier and M. Nivat, *On Translating One Polyomino To Tile the Plane,* Discrete Comput. Geom.,**6** (1991), 575–592.

[8] J. Berstel, *Fibonacci words - a survey,* In The book of L, Springer-Verlag, 1986.

[9] J. Berstel, *Tracé de droites, fractions continues et morphismes itérés* dans Mots, Lang. Raison. Calc., Editions Hermès, Paris, (1990), 298–309.

[10] J. Berstel, *Recent results on extensions of sturmian words*, International Journal of Algebra and Computation, Vol 12, Nos 1 and 2 (2002), 371–385.

[11] J. Berstel and P. Séébold, *Morphismes de Sturm,* Bull. Belg. Math. Soc., **1** (1994), 175–189.

[12] J. Berstel and P. Séébold, *Sturmian words,* In M. Lothaire, Algebraic combinatorics on Words, Cambridge University Press, Cambridge U.K., 2002.

[13] J. Berstel and L. Vuillon, *Coding rotations on intervals*, Theoret. Comput. Sci., **281**, (2002), 99–107.

[14] V. Berthé, *Fréquences des facteurs des suites sturmiennes*, Theoret. Comput. Sci., **165** (1996), 295–309.

[15] V. Berthé, *Autour du système de numération d'Ostrowski*, Bull. Belg. Math. Soc. **8** (2001), 209–239.

[16] V. Berthé and R. Tijdeman, *Balance properties of multi-dimensional words*, Theoret. Comput. Sci., **273**, (2002), 197–224.

[17] V. Berthé and L. Vuillon, *A two-dimensional generalization of Sturmian sequences: tilings and rotations*, Discrete Math., **223** (2000), 27–53.

[18] V. Berthé and L. Vuillon, *Suites doubles de basse complexité*, J. Théor. des Nombres de Bordeaux, **12** (2000), 179–208.

[19] V. Berthé and L. Vuillon, *Palindromes and two-dimensional Sturmian sequences*, J. Autom. Lang. Comb., **6** (2001), 121–138.

[20] *Substitutions in Dynamics, Arithmetics and Combinatorics*, V. Berthé, S. Férenczi, C. Mauduit, A. Siégel eds, Lecture Notes in Mathematics Lecture Notes in Mathematics **1794**, Springer Verlag, 2002.

[21] T. Bousch and J. Mairesse, *Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture*, Journal of the American Mathematical Society, **15** number 1, (2001), 77–111.

[22] M. J. Bresenham, *Algorithm for computer control of a digital plotter*, IBM System Journal, vol. 4, (1965), 25–30.

[23] K. O'Bryant, *Sturmian words and the permutation that orders fractional parts*, to appear in Journal of Algebraic Combinatorics.

[24] J. Cassaigne, *Double sequences with complexity* $mn+1$, J. Auto. Lang. Comb., **3**, 153–170, (1999).

[25] J. Cassaigne, *Subword complexity and periodicity in two or more dimensions*, Developments in Language Theory. Foundations, Applications and Perspectives (DLT'99), Aachen (Allemagne), World Scientific, 14–21, (2000).

[26] J. Cassaigne, S. Ferenczi and L. Q. Zamboni, *Imbalances in Arnoux-Rauzy sequences,* Ann. Inst. Fourier, **50** (2000), no. 4, 1265–1276.

[27] M.G. Castelli, F. Mignosi and A. Restivo, *Fine and Wilf's theorem for three periods and a generalization of Sturmian words,* Theoret. Comp. Sci., **218** (1997), 45–82.

[28] D. Crisp, W. Moran, A. Pollington and P. Shiue, *Substitution invariant cutting sequences* J. Théor. des Nombres de Bordeaux, **5** (1993), 123– 137.

[29] K. Culik II and A. Salomaa, *On infinite words obtained by iterating morphisms*, Theoret. Compt. Sci., **19** (1982), 29–38.

[30] X. Droubay, J. Justin and G. Pirillo, *Episturmian words and some constructions of de Luca and Rauzy*, Theoret. Comp. Sci., **255** (2001), 539–553.

[31] F. Durand, *A charaterization of substitutive sequences using return words*, Discrete Math., **179** (1998), 89–101.

[32] C. Epifanio, P. Mignosi and M. Koskas, *On a conjecture on bidimensional words*, Theoret. Comput. Sci., **299**, (2003), 123–150.

[33] I. Fagnot and L. Vuillon, *Generalized balances in Sturmian words*, Discrete Applied Mathematics **121**, Issues 1-3, (2002), 83–101.

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

[35] S. Ferenczi, *Complexity of sequences and dynamical systems*, Discrete Math., **206** (1999), 145–154.

[36] A.S. Fraenkel, *Complementing and exactly covering sequences*, J. Combin. Theory, Ser. A, **14** (1973), 8–20.

[37] S. Gaubert and J. Mairesse, *Performance evaluation of times Petri nets using heaps of pieces*, In P. Bucholz and M. Silva editors, Petri Nets and Performance Models (PNPM'99), IEEE Computer Society, (1999), 158–169,

[38] B. Gaujal, *Optimal allocation sequences of two processes sharing a resource*, Discrete Event Dynamics Systems, **7** (1997), 327–354.

[39] B. Gaujal, *Convexité discrète et régularité; application au contrle des systèmes à événements discrets*, Habilitation à diriger des recherches, LORIA et INRIA, 2001.

[40] R. L. Graham, *Covering the positive integers by disjoint sets of the form* $\{[n\alpha+\beta] : n = 1, 2, \cdots\}$, J. Combin. Theory Ser. A, **15** (1973), 354-358.

[41] C. Holton and L. Q. Zamboni, *Descendants of primitive substitutions,* Theory Comput. Systems 32, (1999), 133–157.

[42] P. Hubert, *Well balanced sequences*, Theoret. Comput. Sci., **242** (2000), 91-108.

[43] P. Hubert and L. Vuillon, *Complexity of cutting words on regular tilings*, Preprint LIAFA 2002.

[44] O. Jenkinson and L. Zamboni, *Characterisations of balanced words via orderings*, Theoret. Comput. Sci., to appear.

[45] J. Justin and G. Pirillo, *Episturmian words and episturmian morphisms*, Theoret. Comput. Sci., **276**, (2002), 281–313.

[46] J. Justin and L. Vuillon, *Return words in Sturmian and episturmian words,* RAIRO Theoret. Inform. Appl., **34** (2000), 343–356.

[47] M. Keane, *Interval exchange transformations*, Math. Zeit., **141** (1975), 25–31.

[48] J. C. Lagarias and Y. Wang, *The finiteness conjecture for the generalized spectral radius of a set of matrices*, Linear Algebra Appl., **214** (1995), 17–42.

[49] M. Lothaire, *Combinatorics on words*, Encyclopedia of Mathematics and its Applications, Addison-Wesley, 1983.

[50] A. de Luca, *Sturmian words: structure, combinatorics and their arithmetics*, Theoret. Comput. Sci., **183** (1997), 45–82.

[51] J. Mairesse and L. Vuillon, *Asymptotic behavior in a heap model with two pieces*, Theoret. Comput. Sci., **270**, (2002), 525–560.

[52] M. Morse and G. A. Hedlund , *Symbolic dynamics II: Sturmian trajectories*, Amer. J. Math., **62** (1940), 1–42.

[53] M. Nivat, *Sous-ensemble homogènes de* $\mathbb{Z}^2$ *et pavages du plan*, C. R. Math. Acad. Sci. Paris **335** (2002), no. 1, 83–86.

[54] J.-J. Pansiot, *Complexité des facteurs des mots infinis engendrés par morphismes itérés*, in Automata, languages and Programming (Antwerp, 1984), Lecture Notes in Comput. Sci., **172** (1984), 380–389.

[55] M. Queffélec, *Subtitution dynamical systems - spectral analysis*, Lecture Notes Math., **1294**, Springer-Verlag, 1987.

[56] G. Rauzy, *Une généralisation du développement en fraction continue,* Sémin. Delange-Pisot-Poitou, 18e Année **1** (1977), 1501–1515.

[57] G. Rauzy, *échange d'intervalles et transformations induites,* Acta. Arith., **34** (1979), 315–328.

[58] G. Rauzy, *Nombres algébriques et substitutions*, Bull. Soc. Math. France, **110** (1982), 147–178.

[59] G. Rauzy, *Ensembles restes bornés*, Séminaire de Théorie des Nombres de Bordeaux, Exposé 24, (1983-1984).

[60] G. Rote, *Sequences with subword complexity* $2n$, J. Number Theory, **46** (1994), 196–213.

[61] J. W. Sander and R. Tijdeman, *Low complexity functions and convex sets in* $\mathbb{Z}^k$, Math. Z., **233** (2000), 205–218.

[62] J. W. Sander and R. Tijdeman, *The complexity of functions on lattices*, Theoret. Comput. Sci., **246**, (2000), 195–225.

[63] J. W. Sander and R. Tijdeman, *The rectangle complexity of functions on two-dimensional lattices*, Theoret. Comput. Sci., **246**(2000), 195–225.

[64] P. Séébold, *Fibonacci morphism and Sturmian words*, Theoret. Comput. Sci., **88** (1991), 367–384.

[65] Th. Skolem, *Über einige eigenschafen der zahlenmengen* $\lceil \alpha n + \beta \rceil$ *bei irrationalem* α *mit einleitenden bemerkungen über einige kombinatorische probleme*, Norske Vid. Selsk. Forh. (Trondheim), **30** (1957), 118–125.

[66] V. Sós, *On the distribution mod 1 of the sequence* $n\alpha$, Ann. Univ. Sci. Budapest, Eötvös Sect. Math. **1** (1958), 127–134.

[67] R. Tijdeman, *Decomposition of the integers as a direct sum of two subsets*, In Number Theory Seminar Paris 1992-1993 S. David Ed. Cambridge Univ. Press, Cambridge, 1995, 261–276.

[68] R. Tijdeman, *Exact covers of balanced sequences and Fraenkel's conjecture*, Algebraic Number Theory and Diophantine Analysis, Proc. Intern. Conf. Graz, Austria, Ed. by F. Halter-Koch and R.F. Tichy, Walter de Gruyter, Berlin etc., (2000), 467–483.

[69] R. Tijdeman, *Fraenkel's conjecture for six sequences*, Discrete Math., **222** (2000), 223–234.

[70] S. Tabachnikov, *Billiards*, Panoramas et synthèse, SMF, Numéro 1, 1995.

[71] G. Viennot, *Problèmes combinatoires posés par la physique statistique*, Astérisque **121–122**, Société Mathématique de France, 1985.

[72] L. Vuillon, *Combinatoire des motifs d'une suite sturmienne bidimensionnelle,* Theoret. Comput. Sci., **209** (1998), 261–285.

[73] L. Vuillon, *Local configurations in discrete planes*, Bull. Belg. Math. Soc., **6** (1999), p. 625–636.

[74] L. Vuillon, *A characterization of Sturmian words by return words,* European J. Combin., **22** (2001), 263–275.

Laboratoire de mathématiques, CNRS UMR 5127,
Université de Savoie, 73376 Le Bourget-du-lac cedex, France,
Laurent.Vuillon@univ-savoie.fr