Infinite Words with Well Distributed Occurrences

Ľubomíra Balková[1], Michelangelo Bucci[2], Alessandro De Luca[3], and Svetlana Puzynina[2,4]

[1] Department of Mathematics, FNSPE, Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2, Czech Republic
lubomira.balkova@gmail.com
[2] Department of Mathematics, University of Turku, FI-20014 Turku, Finland
{michelangelo.bucci, svepuz}@utu.fi
[3] DIETI, Università degli Studi di Napoli Federico II via Claudio, 21, 80125 Napoli, Italy
alessandro.deluca@unina.it
[4] Sobolev Institute of Mathematics, Russia

Abstract. In this paper we introduce the *well distributed occurrences (WDO)* combinatorial property for infinite words, which guarantees good behavior (no lattice structure) in some related pseudorandom number generators. An infinite word u on a d-ary alphabet has the WDO property if, for each factor w of u, positive integer m, and vector $\mathbf{v} \in \mathbb{Z}_m^d$, there is an occurrence of w such that the Parikh vector of the prefix of u preceding such occurrence is congruent to $\mathbf{v}$ modulo m. We prove that Sturmian words, and more generally Arnoux-Rauzy words and some morphic images of them, have the WDO property.

Introduction

The combinatorial problem studied in this paper comes from random number generation. Pseudorandom number generators aim to produce random numbers using a deterministic process. No wonder they suffer from many defects. The most usual ones – linear congruential generators – are known to produce periodic sequences having a defect called lattice structure. Guimond et al. [2] proved that when two linear congruential generators are combined using infinite words coding certain classes of quasicrystals or, equivalently, of cut-and-project sets, the resulting sequence is aperiodic and has no lattice structure.

We have found a combinatorial condition – *well distributed occurrences*, or WDO for short – that guarantees absence of lattice structure if two arbitrary generators having the same output alphabet are combined using an infinite word having the WDO property. The WDO property for an infinite word u over an alphabet A means that for any integer m and any factor w of u, the set of Parikh vectors modulo m of prefixes of u preceeding the occurrences of w coincides with $\{0, 1, \ldots, m-1\}^{|A|}$ (see Definition 2.1). In other words, among Parikh vectors modulo m of such prefixes one has all possible vectors. Besides giving generators without lattice structure, the WDO property is an interesting combinatorial property of infinite words itself.

We have proved first that Sturmian words have well distributed occurrences, and then we have shown this property for Arnoux-Rauzy words. The proof for Sturmian words is based on different ideas than the one for Arnoux-Rauzy words, therefore we will provide in the sequel both of them.

In the next section, we deal with pseudorandom number generation, thus establishing the motivation for our work. Next, in Section 2, we give the basic combinatorial definitions needed for our main results, including the WDO property. Finally, in the last two sections, we prove that the property holds for Sturmian and Arnoux-Rauzy words, respectively.

1 Motivation in Pseudorandom Number Generation

For the sake of our discussion, any infinite sequence of integers can be understood as a *pseudorandom number generator (PRNG)*; see also [2].

Let $X = (x_n)_{n\in\mathbb{N}}$ and $Y = (y_n)_{n\in\mathbb{N}}$ be two PRNGs with the same output $M \subset \mathbb{N}$ and the same period $m \in \mathbb{N}$, and let $u = u_0u_1u_2\ldots$ be a binary infinite word, i.e., an infinite sequence over $\{0, 1\}$.

The PRNG

$$Z = (z_n)_{n\in\mathbb{N}} \tag{1}$$

based on u is obtained by the following algorithm:

1. Read step by step the letters of u.
2. When you read 0 for the i-th time, copy the i-th symbol from X to the end of the constructed sequence Z.
3. When you read 1 for the i-th time, copy the i-th symbol from Y to the end of the constructed sequence Z.

Of course, it is possible to generalize this construction – using infinite words over a multiliteral alphabet, one can combine more than two PRNGs.

1.1 Lattice Structure

Let $X = (x_n)_{n\in\mathbb{N}}$ be a PRNG whose output is a finite set $M \subset \mathbb{N}$. We say that X has the *lattice structure* if there exists $t \in \mathbb{N}$ such that

$$\{(x_i, x_{i+1}, \ldots, x_{i+t-1}) \mid i \in \mathbb{N}\}$$

is covered by a family of parallel equidistant hyperplanes and at the same time, this family does not cover the whole lattice

$$M^t = \{(a_1, a_2, \ldots, a_t) \mid a_i \in M \text{ for all } i \in \{1, \ldots, t\}\}.$$

It is known that all linear congruential generators have the lattice structure. Recall that a *linear congruential generator* $(x_n)_{n\in\mathbb{N}}$ is given by $a, m, c \in \mathbb{N}$ and defined by the recurrence relation $x_{n+1} = ax_n + c \mod m$. Let us mention a famous example of a PRNG with a striking lattice structure. For $t = 3$, the set of triples of RANDU, i.e., $\{(x_i, x_{i+1}, x_{i+2}) \mid i \in \mathbb{N}\}$ is covered by only 15 equidistant hyperplanes, see Figure 1.

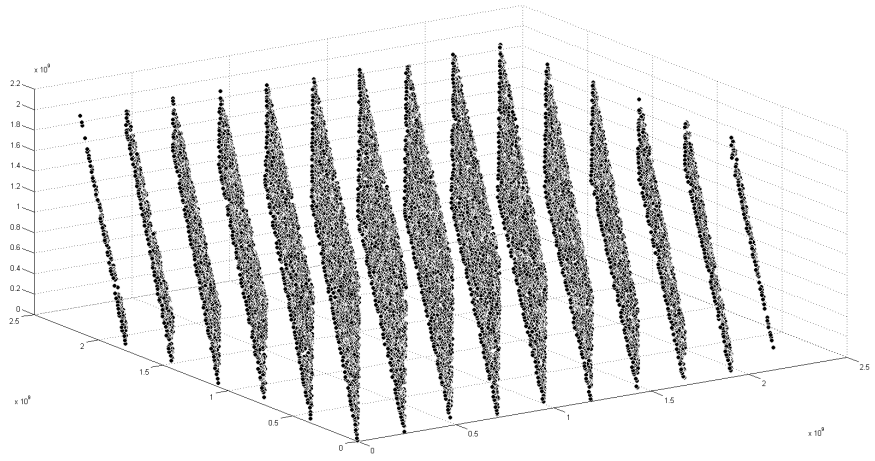

Fig. 1. The triples of RANDU – the linear congruential generator with $a = (2^{16} + 3), m = 2^{31}, c = 0$ – are covered by as few as 15 parallel equidistant planes.

1.2 Combinatorial Condition on Absence of the Lattice Structure

Guimond et al. in [2] have shown that PRNGs based on infinite words coding a certain class of cut-and-project sets have no lattice structure. A crucial part of their proof is the following lemma.

Lemma 1.1. *Let Z be the PRNG from* (1) *based on an aperiodic infinite word. If there exist for any $a, b \in M$ and for any $\ell \in \mathbb{N}$ an ℓ-tuple z such that both za and zb are $(\ell+1)$-tuples of the sequence Z, then Z does not have the lattice structure.*

We have found the following combinatorial condition on binary infinite words guaranteeing that the assumptions of the previous lemma are met: we say that a binary aperiodic infinite word u over the alphabet $\{0, 1\}$ has *well distributed occurrences* (or has *the WDO property*) if u satisfies for any $m \in \mathbb{N}$ and any factor w of u the following condition. If we denote $i_0, i_1, \ldots$ the occurrences of w in u, then

$$\left\{ \left(|u_0 u_1 \cdots u_{i_j - 1}|_0, |u_0 u_1 \cdots u_{i_j - 1}|_1 \right) \bmod m \mid j \in \mathbb{N} \right\} = \mathbb{Z}_m^2 \,,$$

where $\bmod m$ is applied elementwise.

See the next section for the definition of aperiodicity, factor occurrences, and the WDO property for general alphabets.

The WDO property for binary words thus ensures no lattice structure for PRNGs defined in (1).

Theorem 1.2. *Let Z be the PRNG from* (1) *based on a binary aperiodic infinite word having the WDO property. Then Z has no lattice structure.*

We omit the proof of this theorem for the sake of brevity.

Moreover, we have shown that the class of infinite words satisfying the WDO property for binary words is larger than the class described in [2] (see Section 3).

2 Combinatorics on Words and the WDO Property

By A we denote a finite set of symbols called *letters*; the set A is therefore called an *alphabet*. A finite string $w = w_1 w_2 \dots w_n$ of letters from A is said to be a *finite word*, its length is denoted by $|w| = n$ and $|w|_a$ denotes the number of occurrences of $a \in A$ contained in w. The empty word, a neutral element for concatenation of finite words, is denoted ε and it is of zero length.

Under an *infinite word* we understand an infinite sequence $u = u_0 u_1 u_2 \dots$ of letters from A. A finite word w is a *factor* of a word v (finite or infinite) if there exist words p and s such that $v = pws$. If $p = \varepsilon$, then w is said to be a *prefix* of v; if $s = \varepsilon$, then w is a *suffix* of v. The set of factors and prefixes of v are denoted by $\mathrm{Fact}(v)$ and $\mathrm{Pref}(v)$, respectively. If $v = ps$ for finite words v, p, s, then we write $p = vs^{-1}$ and $s = p^{-1}v$.

An infinite word u over the alphabet A is called *eventually periodic* if it is of the form $u = vw^\omega$, where v, w are finite words over A and ω denotes an infinite repetition. An infinite word is called *aperiodic* if it is not eventually periodic.

For any factor w of an infinite word u, every index i such that w is a prefix of the infinite word $u_i u_{i+1} u_{i+2} \dots$ is called an *occurrence* of w in u.

The *factor complexity* of an infinite word u is a map $\mathcal{C}_u : \mathbb{N} \mapsto \mathbb{N}$ defined by $\mathcal{C}_u(n) :=$ the number of factors of length n contained in u. The factor complexity of eventually periodic words is bounded, while the factor complexity of an aperiodic word u satisfies $\mathcal{C}_u(n) \geq n + 1$ for all $n \in \mathbb{N}$. A *right extension* of a factor w of u over the alphabet A is any letter $a \in A$ such that wa is a factor u. Of course, any factor of u has at least one right extension. A factor w is called *right special* if w has at least two right extensions. Similarly, one can define a *left extension* and a *left special* factor. A factor is *bispecial* if it is both right and left special. An aperiodic word contains right special factors of any length.

The *Parikh vector* of a finite word w over an alphabet $\{0, 1, \dots, d-1\}$ is defined as $(|w|_0, |w|_1, \dots, |w|_{d-1})$. For a finite or infinite word $u = u_0 u_1 u_2 \dots$, we denote by $\mathrm{Pref}_n u$ the prefix of length n of u, i.e., $\mathrm{Pref}_n u = u_0 u_1 \dots u_{n-1}$.

Let us generalize the combinatorial condition on infinite words that guarantees no lattice structure for pseudorandom number generators from binary to multiliteral alphabets.

Definition 2.1 (The WDO property). *We say that an aperiodic infinite word u over the alphabet $\{0, 1, \dots, d-1\}$ has* well distributed occurrences *(or has* the WDO property*) if u satisfies for any $m \in \mathbb{N}$ and any factor w of u the following condition. If we denote $i_0, i_1, \dots$ the occurrences of w in u, then*

$$\left\{ \left(|u_0 u_1 \cdots u_{i_j - 1}|_0, \dots, |u_0 u_1 \cdots u_{i_j - 1}|_{d-1} \right) \bmod m \mid j \in \mathbb{N} \right\} = \mathbb{Z}_m^d \,;$$

that is, the Parikh vectors of $\mathrm{Pref}_{i_j}(u)$ for $j \in \mathbb{N}$, when reduced modulo m, give the whole $\mathbb{Z}_m^d$.

We define the WDO property for aperiodic words since it clearly never holds for periodic ones.

With the above notation, it is easy to see that if a recurrent infinite word u has the WDO property, then for every vector $\mathbf{v} \in \mathbb{Z}_m^d$ there are infinitely many values of j such that the Parikh vector of $\mathrm{Pref}_{i_j}(u)$ is congruent to $\mathbf{v}$ modulo m.

Example 2.2. The Thue-Morse word $t = 0110100110010110\cdots$, which is a fixed point of the morphism $0 \mapsto 01$, $1 \mapsto 10$, does not satisfy the WDO property. Indeed, take $m = 2$ and $w = 00$, then w occurs only in odd positions i_j so that $(|t_0 \cdots t_{i_j-1}|_0 + |t_0 \cdots t_{i_j-1}|_1) = i_j$ is odd. Thus, e.g., $(|t_0 \cdots t_{i_j-1}|_0, |t_0 \cdots t_{i_j-1}|_1) \bmod 2 \neq (0,0)$, and hence $\{(|t_0 \cdots t_{i_j-1}|_0, |t_0 \cdots t_{i_j-1}|_1) \bmod 2 \mid j \in \mathbb{N}\} \neq \mathbb{Z}_2^2$.

Example 2.3. We say that an infinite word u over an alphabet A, $|A| = d$, is *universal* if it contains all finite words over A as its factors. It is easy to see that any universal word satisfies the WDO property. Indeed, for any word $w \in A^*$ and any m there exists a finite word v such that if we denote $i_0, i_1, \ldots, i_k$ the occurrences of w in v, then

$$\left\{\left(|\mathrm{Pref}_{i_j} v|_0, \ldots, |\mathrm{Pref}_{i_j} v|_{d-1}\right) \bmod m \mid j \in \{0,1,\ldots,k\}\right\} = \mathbb{Z}_m^d\,.$$

Since u is universal, v is a factor of u. Denoting by i an occurrence of v in u, one gets that the positions $i + i_j$ are occurrences of w in u. Hence

$$\begin{aligned}&\left\{\left(|\mathrm{Pref}_{i+i_j} u|_0, \ldots, |\mathrm{Pref}_{i+i_j} u|_{d-1}\right) \bmod m \mid j \in \{0,1,\ldots,k\}\right\} = \\ &= \left(|\mathrm{Pref}_i u|_0, \ldots, |\mathrm{Pref}_i u|_{d-1}\right) + \\ &+ \left\{\left(|\mathrm{Pref}_{i_j} v|_0, \ldots, |\mathrm{Pref}_{i_j} v|_{d-1}\right) \bmod m \mid j \in \{0,1,\ldots,k\}\right\} = \mathbb{Z}_m^d\,.\end{aligned}$$

Therefore, u satisfies the WDO property.

3 Sturmian Words

In this section, we show that Sturmian words have well distributed occurrences.

Definition 3.1. *An aperiodic infinite word u is called* Sturmian *if its factor complexity satisfies $\mathcal{C}_u(n) = n + 1$ for all $n \in \mathbb{N}$.*

So, Sturmian words are by definition binary and they have the lowest possible factor complexity among aperiodic infinite words. Sturmian words admit various types of characterizations of geometric and combinatorial nature. One of such characterizations is via irrational rotations on the unit circle. In [4] Hedlund and Morse showed that each Sturmian word may be realized measure-theoretically by an irrational rotation on the circle. That is, every Sturmian word is obtained by coding the symbolic orbit of a point on the circle of circumference one under a rotation R_α by an irrational angle[5] α, $0 < \alpha < 1$, where the circle is partitioned into two complementary intervals, one of length α and the other of length $1 - \alpha$. And conversely each such coding gives rise to a Sturmian word.

[5] Measured by arc length (thus equivalent to $2\pi\alpha$ radians).

Definition 3.2. *The* rotation *by angle α is the mapping R_α from $[0,1)$ (identified with the unit circle) to itself defined by $R_\alpha(x) = \{x+\alpha\}$, where $\{x\} = x - \lfloor x \rfloor$ is the fractional part of x. Considering a partition of $[0,1)$ into $I_0 = [0, 1-\alpha)$, $I_1 = [1-\alpha, 1)$, define a word*

$$s_{\alpha,\rho}(n) = \begin{cases} 0, & \text{if } R_\alpha^n(\rho) = \{\rho + n\alpha\} \in I_0, \\ 1, & \text{if } R_\alpha^n(\rho) = \{\rho + n\alpha\} \in I_1. \end{cases}$$

One can also define $I_0' = (0, 1-\alpha]$, $I_1' = (1-\alpha, 1]$, the corresponding word is denoted by $s'_{\alpha,\rho}$.

For more information on Sturmian words we refer to [3, Chapter 2].

Theorem 3.3. *Let u be a Sturmian word on $\{0,1\}$. Then u has Property WDO.*

Proof. In the proof we use the definition of Sturmian word via rotation. The main idea is controlling the number of 1's modulo m by taking circle of length m, and controlling the length taking the rotation by $m\alpha$.

For the proof we will use an equivalent reformulation of the theorem:

Let u be a Sturmian word on $\{0,1\}$, for any natural number m and any factor w of u let us denote $i_0, i_1, \dots$ the occurrences of w in u. Then

$$\left\{ \left(i_j, |u_0 u_1 \cdots u_{i_j - 1}|_1\right) \bmod m \mid j \in \mathbb{N} \right\} = \{0, 1, ..., m-1\}^2.$$

That is, we will control the number of 1's and the length instead the number of 0's.

Since a Sturmian word can be defined via rotations by an irrational angle on a unit circle, without loss of generality we may assume that $u = s_{\alpha,\rho}$ for some $0 < \alpha < 1$, $0 \le \rho < 1$, α irrational (see Definition 3.2). Equivalently, we can consider m copies of the circle connected into one circle of length m with m intervals $I_1^i = [i-\alpha, i)$ of length α corresponding to 1. The Sturmian word is obtained by rotation by α on this circle of length m (see Fig. 2).

Namely, we define the rotation $R_{\alpha,m}$ as the mapping from $[0,m)$ (identified with the circle of length m) to itself defined by $R_{\alpha,m}(x) = \{x+\alpha\}_m$, where $\{x\}_m = x - \lfloor x/m \rfloor m$ and for $m = 1$ coincides with the fractional part of x. A partition of $[0,m)$ into $2m$ intervals $I_0^i = [i, i+1-\alpha)$, $I_1^i = [i+1-\alpha, i+1)$, $i = 0, \dots, m-1$ defines the Sturmian word $u = s_{\alpha,\rho}$:

$$s_{\alpha,\rho}(n) = \begin{cases} 0, & \text{if } R_{\alpha,m}^n(\rho) = \{\rho + n\alpha\} \in I_0^i \text{ for some } i = 0, \dots, m-1, \\ 1, & \text{if } R_{\alpha,m}^n(\rho) = \{\rho + n\alpha\} \in I_1^i \text{ for some } i = 0, \dots, m-1. \end{cases}$$

It is well known that any factor $w = w_0 \cdots w_{k-1}$ of u corresponds to an interval I_w in $[0,1)$, so that whenever you start rotating from the interval I_w, you obtain w. Namely, $x \in I_w$ if and only if $x \in I_{w_0}, R_\alpha(x) \in I_{w_1}, \dots, R_\alpha^{|w|-1}(x) \in I_{w_{|w|-1}}$.

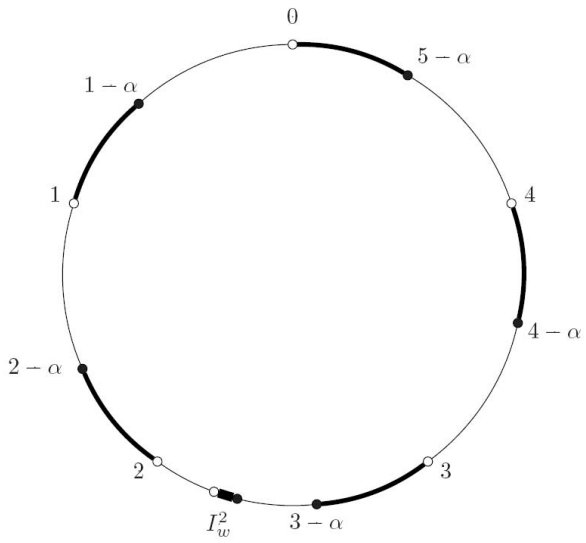

Fig. 2. Illustration to the proof of Theorem 3.3: the example for $m = 5$.

Similarly, we can define m intervals corresponding to w in $[0, m)$ (circle of length m), so that if $I_w = [x_1, x_2)$, then $I_w^i = [x_1 + i, x_2 + i)$, $i = 0, \dots, m-1$.

Fix a factor w of u, take arbitrary $(j, i) \in \{0, 1, \dots, m-1\}^2$. Now we will organize (j, i) among the occurrences of w, i.e., find l such that $u_l \dots u_{l+|w|-1} = w$, $l \mod m = j$ and $|\mathrm{Pref}_l u|_1 \mod m = i$.

Consider rotation $R_{m\alpha,m}(x)$ by $m\alpha$ instead of rotation by α, and start m-rotating from $j\alpha + \rho$. Formally, $R_{m\alpha,m}(x) = \{x + m\alpha\}_m$, where, as above, $\{x\}_m = x - [x/m]m$. This rotation will put us to positions $mk + j$, $k \in \mathbb{N}$ in the Sturmian word: for $a \in \{0, 1\}$ one has $s_{\alpha,\rho}(mk + j) = a$ if $R^k_{m\alpha,m}(j\alpha + \rho) = \{j\alpha + \rho + km\alpha\}_m \in I_a^i$ for some $i = 0, \dots, m-1$.

Remark that the points in the orbit of an m-rotation of a point on the m-circle are dense, and hence the rotation comes infinitely often to each interval. So pick k when $j\alpha + mk\alpha + \rho \in I_w^i \subset [i, i+1)$ (and actually there exist infinitely many such k). Then the length l of the corresponding prefix is equal to $km + j$, and the number of 1's in it is $i + mp$, where p is the number of complete circles you made, i.e., $p = [(j\alpha + mk\alpha + \rho)/m]$. □

Remark 3.4. In the next section we will show that Arnoux-Rauzy words [1], which are natural extensions of Sturmian words to larger alphabets, also satisfy the WDO property. Note that the proof above cannot be generalized to Arnoux-Rauzy words, because it is based on the geometric interpretation of Sturmian

words via rotations, while this interpretation does not extend to Arnoux-Rauzy words.

4 Arnoux-Rauzy Words

4.1 Basic Definitions

Definition 4.1. *Let A be a finite alphabet. The* reversal operator *is the operator $\sim\colon A^* \mapsto A^*$ defined by recurrence in the following way:*

$$\tilde{\varepsilon} = \varepsilon, \quad \widetilde{va} = a\tilde{v}$$

for all $v \in A^*$ and $a \in A$. The fixed points of the reversal operator are called palindromes.

Definition 4.2. *Let $u \in A^*$ be a finite word over the alphabet A. We define the* right palindromic closure *of u, and we denote it by $u^{(+)}$ as the shortest palindrome that has u as a prefix. It is readily verified that if p is the longest palindromic suffix of $u = vp$, then $u^{(+)} = vp\tilde{v}$.*

Definition 4.3. *We call the* iterated (right) palindromic closure operator *the operator ψ recurrently defined by the following rules:*

$$\psi(\varepsilon) = \varepsilon, \quad \psi(va) = (\psi(v)a)^{(+)}$$

for all $v \in A^*$ and $a \in A$. The definition of ψ may be extended to infinite words u over A as $\psi(u) = \lim_n \psi(\mathrm{Pref}_n\, u)$, i.e., $\psi(u)$ is the infinite word having $\psi(\mathrm{Pref}_n\, u)$ as its prefix for every $n \in \mathbb{N}$.

Definition 4.4. *Let Δ be an infinite word on the alphabet A such that every letter occurs infinitely often in Δ. The word $c = \psi(\Delta)$ is then called a* characteristic (or standard) Arnoux-Rauzy word *and Δ is called the* directive sequence *of c. An infinite word u is called an Arnoux-Rauzy word if it has the same set of factors as a (unique) characteristic Arnoux-Rauzy word, which is called the characteristic word of u. The directive sequence of an Arnoux-Rauzy word is the directive sequence of its characteristic word.*

Let us also recall the following well-known characterization:

Theorem 4.5. *Let u be an aperiodic infinite word over the alphabet A. Then u is a standard Arnoux-Rauzy word if and only if the following hold:*

1. $\mathrm{Fact}(u)$ *is closed under reversal (that is, if v is a factor of u so is $\tilde{v}$).*
2. *Every left special factor of u is also a prefix.*
3. *If v is a right special factor of u then va is a factor of u for every $a \in A$.*

From the preceding theorem, it can be easily verified that the bispecial factors of a standard Arnoux-Rauzy correspond to its palindromic prefixes (including the empty word), and hence to the iterated palindromic closure of the prefixes of its directive sequence. That is, if

$$\varepsilon = b_0, b_1, b_2, \dots$$

is the sequence, ordered by length, of bispecial factors of the standard Arnoux-Rauzy word u, $\Delta = \Delta_0\Delta_1\cdots$ its directive sequence (with $\Delta_i \in A$ for every i), we have $b_{i+1} = (b_i\Delta_i)^{(+)}$.

A direct consequence of this, together with the preceding definitions, is the following statement, which will be used in the sequel.

Lemma 4.6. *Let u be a characteristic Arnoux-Rauzy word and let Δ and $(b_i)_{i\geq 0}$ be defined as above. If Δ_i does not occur in b_i, then $b_{i+1} = b_i\Delta_i b_i$. Otherwise let $j < i$ be the largest integer such that $\Delta_j = \Delta_i$. Then $b_{i+1} = b_i b_j^{-1} b_i$.*

4.2 Parikh Vectors and Arnoux-Rauzy Factors

Where no confusion arises, given an Arnoux-Rauzy word u, we will denote by

$$\varepsilon = b_0, b_1, \dots, b_n, \dots$$

the sequence of bispecial factors of u ordered by length and we will set for any $i \in \mathbb{N}$, B_i as the Parikh vector of b_i.

Remark 4.7. By the pigeonhole principle, it is clear that for every $m \in \mathbb{N}$ there exists an integer $N \in \mathbb{N}$ such that, for every $i \geq N$, the set $\{j > i \mid B_j \equiv_m B_i\}$ is infinite. Where no confusion arises and with a slight abuse of notation, fixed m, we will always denote by N the smallest of such integers.

Lemma 4.8. *Let u be a characteristic Arnoux-Rauzy word and let $m \in \mathbb{N}$. Let*

$$\alpha_1 B_{j_1} + \cdots + \alpha_k B_{j_k} \equiv_m \bar{\mathbf{v}} \in \mathbb{Z}_m^d$$

be a linear combination of Parikh vectors such that $\sum_{i=1}^{k} \alpha_i = 0$, with $j_i \geq N$ and $\alpha_i \in \mathbb{Z}$ for all $i \in \{1, \dots k\}$. Then, for any $\ell \in \mathbb{N}$, there exists a prefix v of u such that the Parikh vector of v is congruent to $\bar{\mathbf{v}}$ modulo m and vb_ℓ is also a prefix of u.

Proof. Without loss of generality, we can assume $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_k$, hence there exists k' such that

$$\alpha_1 \geq \alpha_{k'} \geq 0 \geq \alpha_{k'+1} \geq \alpha_k.$$

We will prove the result by induction on $\beta = \sum_{j=1}^{k'} \alpha_j$. If $\beta = 0$, trivially, we can take $v = \varepsilon$ and the statement is clearly verified. Let us assume the statement true for all $0 \leq \beta < M$ and let us prove it for $\beta = M$. By the remark preceding

this lemma, for every ℓ we can choose $i' > j' > \ell$ such that $B_{j_1} \equiv_m B_{i'}$ and $B_{j_k} \equiv_m B_{j'}$. Since every bispecial factor is a prefix and suffix of all the bigger ones, in particular we have that $b_{j'}$ is a suffix of $b_{i'}$, and b_ℓ is a prefix of $b_{j'}$; this implies that $b_{i'}b_{j'}^{-1}b_\ell$ is actually a prefix of $b_{i'}$. By assumption, the Parikh vector of $b_{i'}b_{j'}^{-1}$ is clearly $B_{i'} - B_{j'} \equiv_m B_{j_1} - B_{j_k}$. Since $\alpha_1 \geq 1$ implies $\alpha_k \leq -1$, we have, by induction hypothesis, that there exists a prefix v of u such that the Parikh vector of v is congruent modulo m to

$$(\alpha_1 - 1)B_{j_1} + \cdots + (\alpha_k + 1)B_{j_k}$$

and $vb_{i'}$ is a prefix of u. Hence $vb_{i'}b_{j'}^{-1}b_\ell$ is also a prefix of u and, by simple computation, the Parikh vector of $vb_{i'}b_{j'}^{-1}$ is congruent modulo m to $\bar{\mathbf{v}}$. □

Definition 4.9. *Let $n \in \mathbb{Z}$. We will say that an integer linear combination of integer vectors is a* n-combination *if the sum of all the coefficients equals n.*

Lemma 4.10. *Let u be a characteristic Arnoux-Rauzy word and let $n \in \mathbb{N}$. Every n-combination of Parikh vectors of bispecial factors can be expressed as a n-combination of Parikh vectors of arbitrarily large bispecials. In particular, for every $K, M \in \mathbb{N}$, it is possible to find a finite number of integers $\alpha_1, \ldots, \alpha_k$ such that $B_K = \alpha_1 B_{j_1} + \cdots + \alpha_k B_{j_k}$ with $j_i > M$ for every i and $\alpha_1 + \cdots + \alpha_k = 1$.*

Proof. A direct consequence of Lemma 4.6 is that for every i such that Δ_i appears in b_i, we have $B_{i+1} = 2B_i - B_j$, where $j < i$ is the largest such that $\Delta_j = \Delta_i$. This in turn (since every letter in Δ appears infinitely many times from the definition of Arnoux-Rauzy word) implies that *for every* non-negative integer j, there exists a positive k such that $B_j = 2B_{j+k} - B_{j+k+1}$, that is, we can substitute each Parikh vector of a bispecial with a 1-combination of Parikh vectors of strictly larger bispecials. Simply iterating the process, we obtain the statement. □

In the following we will assume the set A to be a finite alphabet of cardinality d. For every set $X \subseteq A^*$ of finite words, we will denote by $\mathrm{PV}(X) \subseteq \mathbb{Z}^d$ the set of Parikh vectors of elements of X and for every $m \in \mathbb{N}$ we will denote by $\mathrm{PV}_m(X) \subseteq \mathbb{Z}_m^d$ the set of elements of $\mathrm{PV}(X)$ reduced modulo m.

Let u be an infinite word over A and let v be a factor of u. We denote by $S_v(u)$ the set of all prefixes of u followed by an occurrence of v. In other words,

$$S_v(u) = \{p \in \mathrm{Pref}(u) \mid pv \in \mathrm{Pref}(u)\}.$$

Definition 4.11. *For any set of finite words $X \subseteq A^*$, we will say that u* has the property $\mathcal{P}_X$ *(or, for short, that u has $\mathcal{P}_X$) if, for every $m \in \mathbb{N}$ and for every $v \in X$ we have that*

$$\mathrm{PV}_m(S_v(u)) = \mathbb{Z}_m^d.$$

That is to say, for every vector $\mathbf{v} \in \mathbb{Z}_m^d$ there exists a word $w \in S_v(u)$ such that the Parikh vector of w is congruent to $\mathbf{v}$ modulo m.

With this notation, an infinite word u has the WDO property if and only if it has property $\mathcal{P}_{\mathrm{Fact}(u)}$.

Proposition 4.12. *Let u be a characteristic Arnoux-Rauzy word over the d-letter alphabet A. Then u has the property $\mathcal{P}_{\mathrm{Pref}(u)}$.*

Proof. Let us fix an arbitrary $m \in \mathbb{N}$. We want to show that, for every $v \in \mathrm{Pref}(u)$, $\mathrm{PV}_m(S_v(u)) = \mathbb{Z}_m^d$. Let then $\bar{\mathbf{v}} \in \mathbb{Z}^d$ and ℓ be the smallest number such that v is a prefix of b_ℓ. Let $i_1 < i_2 < \cdots < i_d$ be such that Δ_{i_j} does not appear in b_{i_j}, where Δ is the directive word of u. Without loss of generality, we can rearrange the letters so that each Δ_{i_j} is lexicographically smaller than $\Delta_{i_{j+1}}$. With this assumption if, for every j, we set $\bar{\mathbf{v}}_j$ as the Parikh vector of b_{i_j+1}, which, by the first part of Lemma 4.6, equals $b_{i_j}\Delta_{i_j}b_{i_j}$, we can find $j-1$ positive integers $\mu_1, \ldots, \mu_{j-1}$ such that $\bar{\mathbf{v}}_j = (\mu_1, \mu_2, \ldots, \mu_{j-1}, 1, 0, \ldots, 0)$. It is easy to show, then, that the set $V = \{\bar{\mathbf{v}}_1, \ldots, \bar{\mathbf{v}}_d\}$ generates $\mathbb{Z}^d$, hence there exists an integer n such that $\bar{\mathbf{v}}$ can be expressed as an n-combination of elements of V (which are Parikh vectors of bispecial factors of u). Trivially, then, $\bar{\mathbf{v}} = \bar{\mathbf{v}} - n\bar{\mathbf{0}} = \bar{\mathbf{v}} - nB_0$; thus, it is possible to express $\bar{\mathbf{v}}$ as a 0-combination of Parikh vectors of (by the previous Lemma 4.10) arbitrarily large bispecial factors of u. By Lemma 4.8, then there exists a prefix p of u with Parikh vector $\bar{\mathbf{p}}$ such that $\bar{\mathbf{p}} \equiv_m \bar{\mathbf{v}}$ and pb_ℓ is a prefix of u. Since we picked ℓ such that v is a prefix of b_ℓ, we have that $p \in S_v(u)$. From the arbitrariness of v, $\bar{\mathbf{v}}$ and m, we obtain the statement. □

As a corollary of Proposition 4.12, we obtain the main result of this section.

Theorem 4.13. *Let u be an Arnoux-Rauzy word over the d-letter alphabet A. Then u has the property $\mathcal{P}_{\mathrm{Fact}(u)}$.*

Proof. Let m be a positive integer and let c be the characteristic word of u. Let v be a factor of u and xvy be the smallest bispecial containing v. By Proposition 4.12, we have that $\mathrm{PV}_m(S_{xv}(c)) = \mathbb{Z}_m^d$ and, since the set is finite, we can find a prefix p of c such that $\mathrm{PV}_m(S_{xv}(p)) = \mathbb{Z}_m^d$. Let w be a prefix of u such that wp is a prefix of u. If $\bar{\mathbf{x}}$ and $\bar{\mathbf{w}}$ are the Parikh vectors of, respectively, x and w, it is easy to see that

$$\bar{\mathbf{w}} + \bar{\mathbf{x}} + \mathrm{PV}(S_{xv}(p)) \subseteq \bar{\mathbf{w}} + \mathrm{PV}(S_v(p)) \subseteq \mathrm{PV}(S_v(u))$$

Since we have chosen p such that $\mathrm{PV}_m(S_{xv}(p)) = \mathbb{Z}_m^d$, we clearly obtain that $\mathrm{PV}_m(S_v(u)) = \mathbb{Z}_m^d$ and hence, by the arbitrariness of v and m, the statement. □

Remark 4.14. Actually, Theorem 4.13 implies Theorem 3.3.

Remark 4.15. Note the following simple method of obtaining words satisfying the WDO property. Take a word u with the WDO property over an alphabet $\{0, 1, \ldots, d-1\}$, $d > 2$, apply a morphism $\varphi : d-1 \mapsto 0, i \mapsto i$ for $i = 0, \ldots, d-2$, i. e., φ joins two letters into one. It is straightforward that $\varphi(u)$ has WDO property. So, taking Arnoux-Rauzy words and joining some letters, we obtain other words than Sturmian and Arnoux-Rauzy satisfying the WDO property.

Acknowledgements

The first author was supported by the Czech Science Foundation grant GAČR 13-03538S, and thanks L'Oréal Czech Republic for the Fellowship Women in Science. The third author was partially supported by the Italian Ministry of Education (MIUR), under the PRIN 2010–11 project "Automi e Linguaggi Formali: Aspetti Matematici e Applicativi". The fourth author is supported in part by the Academy of Finland under grant 251371 and by Russian Foundation of Basic Research (grants 12-01-00089 and 12-01-00448).

We would like to acknowledge statistical testing of the pseudorandom number generators based on Sturmian and Arnoux-Rauzy words made by Jiří Hladký. He has shown using the Diehard and U01 tests that not only the lattice structure is absent, but also other important properties of PRNGs are improved when LCGs are combined using infinite words having the WDO property.

References

1. P. Arnoux, G. Rauzy, *Représentation géométrique de suites de complexité* $2n + 1$, Bull. Soc. Math. France **119** (1991), 199–215.
2. L.-S. Guimond, Jan Patera, Jiří Patera, *Statistical properties and implementation of aperiodic pseudorandom number generators*, Applied Numerical Mathematics **46(3-4)** (2003), 295–318.
3. M. Lothaire, *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications 90, Cambridge University Press, 2002.
4. M. Morse and G.A. Hedlund, *Symbolic Dynamics II: Sturmian trajectories*, Amer. J. Math. **62 (1)** (1940), 1–42.