

## A generalization of Sturmian sequences: Combinatorial structure and transcendence

by

REBECCA N. RISLEY and LUCA Q. ZAMBONI (Denton, TX)

**I. Introduction.** Given a sequence  $\omega = \omega_0\omega_1\omega_2\dots$  on a finite alphabet, denote by  $L_n(\omega)$  the set of all subwords of  $\omega$  of length  $n$ , that is,  $L_n(\omega) = \{\omega_j\omega_{j+1}\dots\omega_{j+n-1} \mid j \geq 0\}$ . The *complexity function*  $p(n) = p_\omega(n)$  is defined as the cardinality of  $L_n(\omega)$ . A celebrated result of Morse and Hedlund states that a sequence is ultimately periodic if and only if  $p(n) \leq n$  for some  $n$  (see [27]). A binary sequence  $\omega$  is called *Sturmian* if  $p(n) = n + 1$  for all  $n \geq 1$ . Thus among all non-ultimately periodic sequences, Sturmian sequences are those having the smallest complexity. Perhaps the most well known example is the Fibonacci sequence

1211212112112121121121121121121121121121121121121121121121121121121\dots

defined as the fixed point of the morphism  $1 \mapsto 12$  and  $2 \mapsto 1$ .

The study of Sturmian sequences was originated by M. Morse and G. A. Hedlund in the 1930's. They showed that Sturmian sequences provide a symbolic coding of the orbit of a point on a circle with respect to a rotation by an irrational number  $\alpha$  (cf. [26], [27]). Sturmian sequences have since been extensively studied from many different points of view (cf. [3]–[6], [10], [11], [17], [23], [24], [29]).

We consider two natural generalizations of Sturmian sequences to alphabets of more than two letters. The condition  $p(n+1) - p(n) = 1$  implies that each word in  $L_n(\omega)$  is a prefix (suffix) of exactly one word in  $L_{n+1}(\omega)$  except for one which is a prefix (suffix) of two words of length  $n + 1$ . Thus a sequence  $\omega$  on the alphabet  $A_k = \{a_1, \dots, a_k\}$  is called Sturmian (on  $k$  letters) if  $p(n) = n + k - 1$ . As in the binary case we have  $p(n+1) - p(n) = 1$ . This notion of Sturmian was considered by S. Ferenczi and C. Mauduit in [17]. A second generalization of Sturmian, which is the focus of this paper, was originally introduced by P. Arnoux and G. Rauzy in [3] (see also [30]):

---

2000 *Mathematics Subject Classification*: Primary 68R15; Secondary 11J91.  
The second author partially supported by NSF grant INT-9726708.

DEFINITION I.1. Let  $A_k = \{a_1, \dots, a_k\}$  with  $k \geq 2$ . A sequence  $\omega$  in the alphabet  $A_k$  is called an *Arnoux–Rauzy sequence* if it satisfies the following four conditions:

- $\omega$  is uniformly recurrent, i.e., each (finite) subword  $u$  of  $\omega$  occurs in  $\omega$  with bounded gaps,
- the complexity function satisfies  $p(n) = (k - 1)n + 1$ ,
- each word in  $L_n(\omega)$  is a prefix of exactly one word in  $L_{n+1}(\omega)$  except for one which is a prefix of  $k$  words in  $L_{n+1}(\omega)$ ,
- each word in  $L_n(\omega)$  is a suffix of exactly one word in  $L_{n+1}(\omega)$  except for one which is a suffix of  $k$  words in  $L_{n+1}(\omega)$ .

In [3] the authors focused on the special case  $k = 3$ . We note that if  $k = 2$  then  $\omega$  is a (binary) Sturmian sequence. Arnoux–Rauzy sequences have been the topic of a number of recent papers including [2], [7]–[9], [17], [33], [34].

Given an Arnoux–Rauzy sequence  $\omega$  on the alphabet  $A_k = \{a_1, \dots, a_k\}$ , denote by  $X = X_\omega$  the orbit closure of  $\omega$  in  $A_k^{\mathbb{N}}$  with respect to the (left) shift map. We call  $X$  an *Arnoux–Rauzy subshift* on  $A_k$ . For each  $n$  we can write  $L_n(X) = L_n(\omega)$  because uniform recurrence implies that  $L_n(x) = L_n(\omega)$  for each  $x \in X$ . The *language* of  $X$ , denoted  $L(X)$ , is defined by  $L(X) = \bigcup_n L_n(X)$ . Thus  $p(n)$  is just the cardinality of  $L_n(X)$ . It follows from Definition I.1 that for each  $n$ , there is exactly one word in  $L_n(X)$  which is a prefix of more than one word in  $L_{n+1}(X)$ . We call such a word *right special* and denote it by  $r(n)$ . Similarly, there is exactly one word in  $L_n(X)$  which is a suffix of more than one word in  $L_{n+1}(X)$  which we call *left special* and denote it by  $l(n)$ . Thus if  $w \in L_n(X)$  is right special, then the concatenation  $wa \in L_{n+1}(X)$  for every  $a \in A_k$ , and similarly if  $w \in L_n(X)$  is left special, then  $aw \in L_{n+1}(X)$  for every  $a \in A_k$ . A word which is both left and right special is called *bispecial*.

We consider the passage from  $L_n(X)$  to  $L_{n+1}(X)$ . From the set  $L_n(X)$  both  $l(n - 1)$  and  $r(n - 1)$  are identified. In case  $l(n - 1) \neq r(n - 1)$ , the passage from  $L_n(X)$  to  $L_{n+1}(X)$  is completely determined by  $L_n(X)$ . In this case there is exactly one word in  $L_n(X)$  containing  $r(n - 1)$  as a suffix, and this word must be  $r(n)$ . Every other word of length  $n$  has a unique extension by one letter to the right which is determined by its suffix of length  $n - 1$ . The second case is when  $l(n - 1) = r(n - 1)$ , that is,  $l(n - 1)$  is bispecial. In this case there are  $k$  words of length  $n$  containing  $r(n - 1)$  as a suffix, and it is impossible to tell just from within  $L_n(X)$  which of these  $k$  words is  $r(n)$ . Thus in case  $l(n - 1)$  is bispecial, the passage from  $L_n(X)$  to  $L_{n+1}(X)$  is achieved by

- (1) specifying which of the  $k$  words  $\{a_1 r(n - 1), \dots, a_k r(n - 1)\}$  is  $r(n)$ , or equivalently the initial letter of  $r(n)$

and

- (2) specifying which letter must follow the remaining  $k - 1$  words, or equivalently the terminal letter of  $l(n)$ .

We prove that for each  $n \geq 1$ , the word  $r(n)$  is just the mirror image (or reverse) of the word  $l(n)$  (see Proposition II.5). In particular, the initial letter of  $r(n)$  is equal to the terminal letter of  $l(n)$ . Thus, if  $l(n - 1)$  is bispecial, the passage from  $L_n(X)$  to  $L_{n+1}(X)$  depends only on (1), i.e., the initial letter of  $r(n)$ . We code this information in a sequence  $I_X = (i_n)_{n=1}^\infty$  as follows: Let  $\{w_1, w_2, \dots\}$  be the set of bispecial subwords ordered so that  $1 = |w_1| < |w_2| < \dots$ . Set  $i_1 = w_1 \in A_k$ , and for  $n \geq 2$  let  $i_n \in A_k$  so that  $i_n w_{n-1}$  is right special. Then the sequence  $(i_n)$  completely determines  $L(X)$ . We observe that for each  $a \in A_k$  and each  $n \geq 1$  there exists an  $m \geq n$  so that  $al(m)$  is right special. Otherwise by minimality  $X$  would contain a periodic sequence. This implies that each  $a \in A_k$  occurs in  $(i_n)$  an infinite number of times.

The sequence  $(i_n)$  defined above coincides with the sequence  $(i_n)$  defined by P. Arnoux and G. Rauzy in [3] for  $k = 2, 3$  (see Proposition on page 206 for  $k = 2$  and Proposition on page 208 for  $k = 3$ ). Let  $\mathbf{1}_X \in X$  denote the unique accumulation point of the set of all left special words, i.e., each initial subword of  $\mathbf{1}_X$  is left special. In this paper we give two combinatorial algorithms for constructing the (characteristic) sequence  $\mathbf{1}_X$  from the sequence  $(i_n)$ . (See [3], [16], [23], [24], [32] and [31] for examples of algorithms for constructing sequences of specified complexity.)

The first method (see Theorem III.5), which is the central idea of the paper, involves a simple combinatorial algorithm for constructing all bispecial words. Applied in the binary case, this algorithm provides a new method of generating characteristic Sturmian sequences.

The second method is an  $S$ -adic description of the characteristic sequence dual to that given by Arnoux and Rauzy for  $k \in \{2, 3\}$  in [3]. We use it to show that Arnoux–Rauzy subshifts arising from fixed points of primitive morphisms are characterized by the following underlying periodic structure <sup>(1)</sup>:

**THEOREM I.2.** *An Arnoux–Rauzy subshift  $X$  is primitive substitutive (i.e.,  $X$  contains the image, under a letter-to-letter morphism, of a fixed point of a primitive substitution) if and only if the associated sequence  $(i_n)$  is eventually periodic <sup>(2)</sup>.*

---

<sup>(1)</sup> In the Sturmian case, F. Durand and B. Host [14] have an independent proof of Theorem I.2 using dimension groups.

<sup>(2)</sup> In the binary case, the connection between Sturmian sequences and continued fractions implies that the periodicity condition given in Theorem I.2 is equivalent to the frequencies of the letters being quadratic irrationals.

The algorithm described in Theorem III.5 imposes a “rigid” combinatorial structure on the characteristic sequence  $1_X \in X$ , partially shared by all sequences in  $X$ . We show that the language  $L(X)$  contains arbitrarily large blocks of the type  $V^{2+\epsilon}$ .

**PROPOSITION I.3.** *For each  $k = 2, 3, \dots$  there is a positive number  $\epsilon = \epsilon(k)$  such that if  $\omega$  is an Arnoux–Rauzy sequence in  $A_k$ , then  $\omega$  begins in an infinite number of blocks of the form  $UVVV'$  with  $V'$  a prefix of  $V$ , and  $\min\{|V'|/|V|, |V|/|U|\} > \epsilon$ .*

In case  $k \in \{2, 3\}$ , the result of Proposition I.3 was already established by Ferenczi and Mauduit in [17]. In the binary Sturmian case we can say more:

**PROPOSITION I.4.** *Let  $\omega$  be a binary Sturmian sequence. Then  $\omega$  contains arbitrarily large subwords of the form  $V^{(5+\sqrt{5})/2}$ .*

Proposition I.4 extends a result of F. Mignosi and G. Pirillo [25] on the Fibonacci sequence to all Sturmian sequences.

Although Mignosi [24] proved that the Fibonacci sequence does not contain any 4th powers, in the proof of Proposition I.4 we show that unless the associated sequence  $(i_n)$  of  $\omega$  admits the periodic sequence  $12121212\dots$  as a tail (in which case  $\omega$  is the morphic image of the Fibonacci word), the sequence  $\omega$  contains arbitrarily large 4th powers.

Combining Proposition I.3 with a recent combinatorial interpretation of Ridout’s Theorem due to S. Ferenczi and C. Mauduit in [17], we obtain the following extension to  $k \geq 4$  of Proposition 2 and Proposition 4 in [17]:

**THEOREM I.5.** *If for some base  $b \geq 2$  the digit expansion of an irrational number  $\theta$  is an Arnoux–Rauzy sequence on  $k$  letters, then  $\theta$  is a transcendental number.*

We wish to thank the referee of the paper for his many useful comments and suggestions which helped improve both the content and presentation of the paper.

**II. Combinatorial structure of bispecial words.** Let  $\omega$  be an Arnoux–Rauzy sequence on the alphabet  $A_k = \{a_1, \dots, a_k\}$  and  $X = X_\omega$  the associated Arnoux–Rauzy subshift. We denote the length of a word  $w$  by  $|w|$ . We regard the empty word, denoted by  $\varepsilon$ , as the unique word in  $L(X)$  of length zero.

Given a non-empty word  $w = w_1 \dots w_n$  with  $w_i \in A_k$  we define the reverse word  $w_{\text{rev}}$  by  $w_{\text{rev}} = w_n w_{n-1} \dots w_2 w_1$ . If  $u$  and  $v$  are non-empty words in  $L(X)$  we write  $u \vdash uv$  to mean that for each word  $w \in L(X)$  with  $|w| = |u| + |v|$  if  $w$  begins in  $u$  then  $w = uv$ . If it is not the case that  $u \vdash uv$  then we write  $\neg(u \vdash uv)$ . Similarly we write  $vu \dashv u$  to mean that for each

word  $w \in L(X)$  with  $|w| = |u| + |v|$  if  $w$  ends in  $u$  then  $w = vu$ . Otherwise we write  $\neg(vu \dashv u)$ .

LEMMA II.1. *Suppose  $w \in L(X)$  is either the empty word or bispecial and  $a \in A_k$ . Then  $aw$  is right special if and only if  $wa$  is left special.*

PROOF. Because for each  $n \geq 1$  there is exactly one left special word and one right special word in  $L_n(X)$  it suffices to show that if  $aw$  is right special then  $wa$  is left special. We proceed by induction on  $|w|$ . If  $w$  is the empty word, then  $a$  is right special. There is a unique letter  $x \in A_k$  such that for all  $y \in A_k \setminus \{a\}$  we have  $y \vdash yx$ . We claim  $x = a$ . Otherwise, if  $x \neq a$ , then  $x \vdash xx$ , a contradiction. Thus  $a$  is also left special.

Next suppose that  $|w| \geq 1$ . Again there is a unique letter  $x \in A_k$  such that for all  $y \in A_k \setminus \{a\}$  we have  $yw \vdash ywx$ . We claim  $x = a$ . Suppose to the contrary that  $x \neq a$ . Then  $xw \vdash xwx$ . If no prefix of  $xw$  is right special then  $x \vdash xw \vdash xwx$ , which would imply that  $X$  contains the periodic sequence  $xwxwxwxwxw \dots$ , a contradiction. Let  $v$  (possibly the empty word) be the longest prefix of  $w$  such that  $xv$  is right special. Since we are assuming that  $x \neq a$ , it follows that  $|v| < |w|$ . Equivalently, we can write  $xw = xv u$  where  $u$  is not the empty word. By induction hypothesis, the first letter of  $u$  must be  $x$ . Set  $u = xv'$ . It follows by maximality of  $v$  that  $xvx \vdash xv'xv'$ . Since  $xv$  and  $aw$  are both right special and  $|v| < |w|$ , we see that  $xv$  is a proper suffix of  $aw$  and hence a suffix of  $w$ . Thus  $xw \vdash xwx \vdash xv'xv' = xv'u$ . But  $xw$  is a suffix of  $xv'u$  since  $xw = xv'u$  and  $xv$  is a suffix of  $w$ . This implies that  $xwu^n$  is a suffix of  $xv'u^{n+1}$  for each natural number  $n$ . Thus we obtain

$$xw \vdash xv'u \vdash xv'u^2 \vdash xv'u^3 \vdash xv'u^4 \vdash \dots,$$

a contradiction. Hence  $x = a$  and  $wa$  is left special. ■

LEMMA II.2. *Let  $w \in L(X)$  be bispecial and  $a \in A_k$ .*

(1) *Suppose  $aw$  is right special and  $a \vdash aw$ . Then  $wa \vdash waw$  and  $waw$  is bispecial.*

(2) *Suppose  $wa$  is left special and  $wa \dashv a$ . Then  $waw \dashv aw$  and  $waw$  is bispecial.*

PROOF. The proof of this lemma relies on the previous lemma. In view of the completely symmetric nature of Lemma II.1 and of the statements (1) and (2), we will prove only (1). Assume that  $aw$  is right special and  $a \vdash aw$ . Clearly  $wa \vdash waw$ . We show that  $waw$  is bispecial. Lemma II.1 implies that  $wa$  is left special; since  $wa \vdash waw$ ,  $waw$  is also left special. It remains to show that  $waw$  is also right special. For this it will suffice to show that  $wa \dashv a$  since we already know that  $aw$  is right special. Suppose to the contrary that there is a proper suffix  $v$  (possibly empty) of  $w$  such that

$va$  is left special. Then by Lemma II.1,  $av$  is a right special proper prefix of  $aw$ , contradicting our assumption that  $a \vdash aw$ . ■

LEMMA II.3. *Let  $w \in L(X)$  be either empty or bispecial and  $a \in A_k$ .*

(1) *Suppose  $aw$  is right special and  $\neg(a \vdash aw)$ . Let  $v$  (possibly empty) be the longest proper prefix of  $w$  with the property that  $av$  is right special. By Lemma II.1 we have  $w = vau$  for some  $u \in L(X)$ . Then  $wa \vdash wau$  and  $wau$  is bispecial.*

(2) *Suppose  $wa$  is left special and  $\neg(wa \dashv a)$ . Let  $v$  (possibly empty) be the longest proper suffix of  $w$  with the property that  $va$  is left special. By Lemma II.1 we have  $w = uav$  for some  $u \in L(X)$ . Then  $uaw \dashv aw$  and  $uaw$  is bispecial.*

PROOF. Again by symmetry it suffices to establish (1). We suppose  $aw$  is right special, and  $v$  (possibly empty) is the longest proper prefix of  $w$  with the property that  $av$  is right special. Since  $v$  is either empty or bispecial it follows by Lemma II.1 that  $va$  is left special and hence a prefix of  $w$ . That is, we can write  $w = vau$  for some  $u \in L(X)$ . The maximality of the length of  $v$  implies that  $ava \vdash avau$ . But since  $av$  is right special and  $|av| \leq |w|$  it follows that  $av$  is a suffix of  $w$  and hence  $ava$  a suffix of  $wa$ . Thus  $wa \vdash wau$ . We now show  $wau$  is bispecial. By Lemma II.1 we have  $wa$  is left special and since  $wa \vdash wau$  we have  $wau$  is also left special. Since  $av$  is a suffix of  $w$  it follows that  $avau$  is a suffix of  $wau$ . But  $avau = aw$ , which is right special. Thus to see that  $wau$  is right special it suffices to show that  $wa \dashv avau$ . Suppose that some suffix of  $wa$  of the form  $zava$  ( $z \in L(X)$  possibly empty) were left special. Then  $zav$  would be a bispecial suffix of  $w$  of length strictly greater than  $|v|$ , and  $zava$  is left special, and hence by Lemma II.1,  $azav$  is right special. It follows from the maximality of the length of  $v$  that  $wa = zava$ . ■

REMARK II.4. It follows from the previous lemmas that if  $w \in L(X)$  is bispecial then there is a shortest bispecial word  $W$  properly containing  $w$  as a prefix. Moreover  $W$  is of the form  $waw'$  where  $a \in A_k$  is the unique letter for which  $aw$  is right special and  $w'$  (possibly the empty word) is a suffix of  $w$ . We also remark that  $W$  is also the shortest bispecial word containing  $w$  as a suffix and hence  $W$  is also of the form  $v'aw$  where  $a$  is as above and  $v'$  (possibly the empty word) is a prefix of  $w$ .

PROPOSITION II.5. *A non-empty word  $v \in L(X)$  is right special if and only if  $v_{\text{rev}}$  is left special. In particular if  $v$  is bispecial then  $v = v_{\text{rev}}$ .*

PROOF. We proceed by induction on  $|v|$ . It suffices to show that if  $v$  is right special then  $v_{\text{rev}}$  is left special. We saw at the beginning of the proof of Lemma II.1 that if  $|v| = 1$  and  $v$  is right special, then  $v = v_{\text{rev}}$  is also

left special. Next suppose that  $|v| > 1$  and  $v$  is right special. Let  $w$  be the unique left special word of length  $|v|$ . We show that  $w = v_{\text{rev}}$ . Let  $v'$  denote the longest proper suffix of  $v$  which is bispecial, and  $w'$  the longest proper prefix of  $w$  which is bispecial. Since  $v'$  is left special,  $v'$  is a prefix of  $w$ . Since  $w'$  is chosen to be the longest bispecial prefix of  $w$  we have  $|w'| \geq |v'|$ . A similar argument shows that  $|v'| \geq |w'|$  and hence  $v' = w'$ . Also by induction hypothesis we have  $v'_{\text{rev}} = w' = w'_{\text{rev}}$ . We write  $w = w'aw''$  and  $v = v''bv'$  where  $a, b \in A_k$  and  $|v''| = |w''|$ . Since  $v' = w'$  and  $bv'$  is right special it follows from Lemma II.1 that  $a = b$ . So we have  $w = w'aw''$  and  $v = v''aw'$ . If  $|w''| = 0$  then  $v_{\text{rev}} = w'_{\text{rev}}a = w'a = w$ . So we can assume that  $|w''| > 0$ . Let  $W$  be the shortest bispecial word containing  $w'$  as a prefix and hence containing  $w'$  as a suffix. In view of Remark II.4 we can write  $W = w'aw''' = v'''aw'$  with  $w'''$  a suffix of  $w'$  and  $v'''$  a prefix of  $w'$ . The maximality of  $w'$  implies that  $w''$  is a prefix of  $w'''$  and  $v''$  is a suffix of  $v'''$ . It suffices to show that  $v'''_{\text{rev}} = w'''$ . But  $v'''$  is a prefix of  $w'$  and so  $v'''_{\text{rev}}$  is a suffix of  $w'_{\text{rev}} = w'$ . Since  $|v'''_{\text{rev}}| = |w'''|$  and both are a suffix of  $w'$  it follows that  $v'''_{\text{rev}} = w'''$  as required. ■

**COROLLARY II.6.** *For each non-empty word  $v$  we have  $v \in L(X)$  if and only if  $v_{\text{rev}} \in L(X)$ .*

**Proof.** This follows immediately from the above lemma since for each word  $v \in L(X)$  there is a bispecial word which contains  $v$  as a subword. ■

**III. Constructing the characteristic sequence.** Let  $X$  be an Arnoux–Rauzy subshift on  $A_k = \{a_1, \dots, a_k\}$ .

**DEFINITION III.1.** Let  $\{\varepsilon = w_1, w_2, \dots\}$  be the set of all bispecial words in  $X$  ordered so that  $0 = |w_1| < |w_2| < \dots$ . Let  $I_X = (i_n)_{n=1}^\infty \in A_k^\mathbb{N}$  be the sequence defined by  $i_n \in A_k$  so that  $i_n w_n$  is right special.

We saw in Section I that the sequence  $I_X$  completely determines the language  $L(X)$ . Hence two Arnoux–Rauzy subshifts  $X$  and  $Y$  are equal if and only if  $I_X = I_Y$ .

**DEFINITION III.2.** The *characteristic sequence* of  $X$ , denoted by  $\mathbf{l}_X = l_1 l_2 \dots$  with  $l_i \in A_k$ , is the unique accumulation point in  $X$  of the set of all left special words in  $L(X)$ .

Since  $i_n w_n$  is right special is equivalent to  $w_n i_n$  is left special (Lemma II.1), the sequence  $(i_n)$  is a subsequence of the sequence  $\mathbf{l}_X$ . In fact, for each  $n$  we have  $i_n = l_{|w_n|+1}$ . Similarly there is a unique sequence  $\mathbf{r}_X = \dots r_{-3} r_{-2} r_{-1}$  indexed by the negative integers with the property that  $r_{-n} r_{-n+1} \dots r_{-2} r_{-1}$  is right special for each  $n \geq 1$ . In view of Propo-

sition II.5 the sequences  $\mathbf{l}$  and  $\mathbf{r}$  are mirror images of one another, that is,  $r_{-n} = l_n$  for each  $n \geq 1$ .

We now give two combinatorial algorithms for building the sequence  $\mathbf{l}_X$  from the sequence  $(i_n)$  (see Theorem III.5 and Proposition III.7). We begin with a combinatorial construction used in Theorem III.5.

Define a function  $F : A_k^{\mathbb{N}} \rightarrow A_k^{\mathbb{N}}$  as follows: Set

$$A'_k = \{a_1, \dots, a_k, \hat{a}_1, \dots, \hat{a}_k\}$$

and let  $\phi$  denote the morphism  $\phi : A'_k \rightarrow A_k$  defined by  $\phi(a_i) = \phi(\hat{a}_i) = a_i$  for each  $1 \leq i \leq k$ . The morphism  $\phi$  extends to a morphism (also denoted by  $\phi$ ) from words in  $A'_k$  to words in  $A_k$  and from sequences in  $A'_k$  to sequences in  $A_k$ . With each sequence  $\mathbf{x} = \{x_n\}_{n=1}^{\infty}$  in  $A_k^{\mathbb{N}}$  we associate a sequence of words  $\{B_n\}_{n=1}^{\infty}$  in the alphabet  $A'_k$  as follows:  $B_1 = \hat{x}_1$  and for  $n > 1$ ,  $B_n$  is obtained from  $B_{n-1}$  according to the following rule: If  $\hat{x}_n$  does not occur in  $B_{n-1}$  then  $B_n = B_{n-1}\hat{x}_n\phi(B_{n-1})$ . Otherwise if  $\hat{x}_n$  occurs in  $B_{n-1}$ , then we can write  $B_{n-1} = v'\hat{x}_nu'$  where  $v'$  and  $u'$  are words in  $A'_k$  (possibly empty) and  $\hat{x}_n$  does not occur in  $u'$ . In this case we set  $B_n = B_{n-1}\hat{x}_n\phi(u')$ . The sequence  $\{B_n\}_{n=1}^{\infty}$  of words converges to a unique sequence  $B$  in the alphabet  $A'_k$ . We set  $F(\mathbf{x}) = \phi(B)$ .

EXAMPLE III.3. Let  $\mathbf{x} = 1212121212\dots$ . Then

$$\begin{aligned} B_1 &= \hat{1}, \\ B_2 &= \hat{1}\hat{2}1, \\ B_3 &= \hat{1}\hat{2}1\hat{1}\hat{2}1, \\ B_4 &= \hat{1}\hat{2}1\hat{1}\hat{2}1\hat{2}1\hat{1}\hat{2}1, \\ B_5 &= \hat{1}\hat{2}1\hat{1}\hat{2}1\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1, \\ B_6 &= \hat{1}\hat{2}1\hat{1}\hat{2}1\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1, \dots \end{aligned}$$

Then

$$B = \hat{1}\hat{2}1\hat{1}\hat{2}1\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\hat{1}\hat{2}1\dots$$

and

$$F(\mathbf{x}) = 121121211211212112121121121211212112121121121211212112112121121\dots$$

is the Fibonacci sequence. In general the periodic sequence

$$\mathbf{x} = 1^n 2^n 1^n 2^n 1^n 2^n \dots$$

gives rise to the fixed point of the morphism

$$1 \mapsto 1^n 2, \quad 2 \mapsto 1.$$



EXAMPLE III.4. Let  $\mathbf{x} = 123123123123\dots$ . Then

$$\begin{aligned} B_1 &= \hat{1}, \\ B_2 &= \hat{1}\hat{2}1, \\ B_3 &= \hat{1}\hat{2}1\hat{3}121, \\ B_4 &= \hat{1}\hat{2}1\hat{3}121\hat{1}\hat{2}13121, \\ B_5 &= \hat{1}\hat{2}1\hat{3}121\hat{1}\hat{2}13121\hat{2}\hat{1}31211\hat{2}13121, \dots \end{aligned}$$

Then

$$B = \hat{1}\hat{2}1\hat{3}121\hat{1}\hat{2}13121\hat{2}\hat{1}31211\hat{2}13121\hat{3}1211\hat{2}131212131211\hat{2}13121\dots$$

and

$$F(\mathbf{x}) = 121312112131212131211213121312112131212131211213121\dots$$

is the fixed point of the morphism

$$1 \mapsto 12, \quad 2 \mapsto 13, \quad 3 \mapsto 1.$$

In general if  $\mathbf{x}$  is the periodic sequence

$$\mathbf{x} = a_1^n a_2^n \dots a_k^n a_1^n a_2^n \dots a_k^n a_1^n a_2^n \dots a_k^n \dots$$

then  $F(\mathbf{x})$  is the fixed point of the morphism <sup>(3)</sup>

$$\begin{aligned} a_1 &\mapsto a_1^n a_2, \\ a_2 &\mapsto a_1^n a_3, \\ a_3 &\mapsto a_1^n a_4, \\ &\vdots \\ a_{k-1} &\mapsto a_1^n a_k, \\ a_k &\mapsto a_1. \end{aligned}$$

THEOREM III.5. *Let  $X$  be an Arnoux–Rauzy subshift on  $A_k$ . Let  $\mathbf{l}_X \in X$  denote the characteristic sequence of  $X$  and  $I_X = (i_n)$  the sequence in Definition III.1. Then each  $a \in A_k$  occurs in  $(i_n)$  an infinite number of times and  $\mathbf{l}_X = F(I_X)$ . Conversely, if  $\mathbf{x} = \{x_n\}_{n=1}^\infty$  is a sequence in  $A_k$  such that each letter  $a \in A_k$  occurs infinitely often in  $\mathbf{x}$ , then  $F(\mathbf{x})$  is the characteristic sequence of an Arnoux–Rauzy subshift.*

PROOF. Let  $X$  be an Arnoux–Rauzy subshift and  $I_X = (i_n)$  be as in Definition III.1. We saw in Section I that each  $a \in A_k$  occurs in  $(i_n)$  an infinite number of times. Let  $\{B_n\}_{n=1}^\infty$  be the sequence of words defined above. Then Lemmas II.1–II.3 imply that  $\{\phi(B_n)\}_{n=1}^\infty$  is precisely the set of all bispecial words. We prove this by induction on  $n$ .

---

<sup>(3)</sup> Substitutions of this type were investigated by J.-I. Tamura in [32].

We show that for each  $n$ ,  $\phi(B_n)$  is the shortest bispecial word containing  $\phi(B_{n-1})$  as a proper prefix. We take  $B_0 = \varepsilon$  the empty word. For  $n = 1$  we see that  $\phi(B_1) = i_1$  defined as the unique bispecial word of length one. Next suppose that  $\{\phi(B_1), \dots, \phi(B_{n-1})\}$  are the  $n - 1$  shortest bispecial words. Recall that  $i_n$  was defined as the unique letter in  $A_k$  such that  $i_n\phi(B_{n-1})$  is right special. By Lemma II.1,  $\phi(B_{n-1})i_n$  is left special.

We consider the two cases in the recursive definition of  $B_n$  separately. In case there is no occurrence of  $\widehat{i}_n$  in  $B_{n-1}$ , then  $\phi(B_{n-1})$  satisfies the hypothesis of Lemma II.2 with  $w = \phi(B_{n-1})$  and  $a = i_n$ . It follows from Lemma II.2 that  $\phi(B_{n-1})i_n \vdash \phi(B_{n-1})i_n\phi(B_{n-1})$  and  $\phi(B_{n-1})i_n\phi(B_{n-1}) = \phi(B_n)$ . Thus  $\phi(B_n)$  is the shortest bispecial word containing  $\phi(B_{n-1})$  as a proper prefix.

In case  $\widehat{i}_n$  occurs in  $B_{n-1}$  we write  $B_{n-1} = v'\widehat{i}_nu'$  where  $\widehat{i}_n$  does not occur in  $u'$ . In this case  $\phi(B_{n-1})$  satisfies the hypothesis of Lemma II.3 with  $w = \phi(B_{n-1})$ ,  $a = i_n$ ,  $v = \phi(v')$  and  $u = \phi(u')$ . By Lemma II.3 we have  $\phi(B_{n-1})i_n \vdash \phi(B_{n-1})i_n\phi(u')$  and  $\phi(B_{n-1})i_n\phi(u') = \phi(B_n)$ . Thus  $\phi(B_n)$  is the shortest bispecial word containing  $\phi(B_{n-1})$  as a proper prefix.

Having established that  $\phi(B_n)$  is bispecial for each  $n$  it follows that  $F(I_X)$  defined to be the limit of  $\{\phi(B_n)\}_{n=1}^\infty$  is equal to the characteristic sequence  $\mathbf{1}_X$ .

The converse follows from our discussion in Section I concerning the obstruction in building the language  $L(X)$  of an Arnoux–Rauzy subshift. We saw that the obstruction is coded in a sequence  $(i_n)$  with the property that each  $a \in A_k$  occurs in  $(i_n)$  an infinite number of times. ■

As an immediate consequence of the above construction we have the following generalization of a theorem of F. Mignosi [23] which states that if the sequence of partial quotients in the continued fraction expansion of the slope of a (binary) Sturmian sequence  $\omega$  is unbounded then  $\omega$  contains arbitrarily large powers of words:

**COROLLARY III.6.** *If the sequence  $I_X = (i_n)$  contains arbitrarily large blocks of the form  $a^m$  for some  $a \in A_k$ , then for each  $N \geq 1$  and  $M \geq 1$  the language  $L(X)$  contains a block of the form  $V^M$  where  $V$  is a word of length  $\geq N$ .*

We now establish the following alternative description of the sequence  $\mathbf{1}_X$ :

**PROPOSITION III.7.** *Let  $X$  be an Arnoux–Rauzy subshift on  $A_k$  and  $I_X = (i_n)$  the sequence in Definition III.1. For each  $a \in A_k$  define the morphism  $\tau_a$  by  $\tau_a(a) = a$  and  $\tau_a(b) = ab$  for each  $b \in A_k \setminus \{a\}$ . Then for each  $x \in A_k$  the characteristic sequence  $\mathbf{1}_X$  is given by*

$$\lim_{n \rightarrow \infty} \tau_{i_1} \circ \dots \circ \tau_{i_n}(x).$$

Proposition III.7 is a “dual” reformulation of the following result due to Arnoux and Rauzy in [3]:

THEOREM III.8 (P. Arnoux and G. Rauzy, [3]). *For each  $n \geq 1$  and for each  $x \in A_k$*

$$U_{(n,x)} = \sigma_{i_1} \circ \dots \circ \sigma_{i_n}(x)$$

where for each  $a \in A_k$  the morphism  $\sigma_a$  is defined by  $\sigma_a(a) = a$  and  $\sigma_a(b) = ba$  for  $b \in A_k \setminus \{a\}$ .

Although Arnoux and Rauzy only prove Theorem III.8 for  $k = 2$  (Proposition on page 206) and  $k = 3$  (Proposition on page 208), they point out (in Remarque 2 on page 202) that the results in [3] extend to all  $k \geq 2$ .

*Proof of Proposition III.7.* For each positive integer  $n$  let  $w_{n+1}$  denote the  $n$ th bispecial word (see Definition III.1). Following [3], for  $x \in A_k$ , the word  $U_{(n,x)}$  is obtained by concatenating the labels of the edges along the simple closed directed path in the Rauzy graph  $\Gamma_{|w_{n+1}|}$  beginning and ending at  $w_{n+1}$  whose first directed edge is labeled  $x$ . It is readily verified that for each  $n \geq 1$  and  $x \in A_k$ ,

$$\tau_{i_1} \circ \dots \circ \tau_{i_n}(x) = (\sigma_{i_1} \circ \dots \circ \sigma_{i_n}(x))_{\text{rev}},$$

and hence by Theorem III.8,

$$\tau_{i_1} \circ \dots \circ \tau_{i_n}(x) = (U_{(n,x)})_{\text{rev}}.$$

Set  $\mu(n, x) = \min\{|U_{(n,x)}|, |w_{n+1}|\}$ . Since  $U_{(n,x)}$  is a loop based at the  $n$ th bispecial word  $w_{n+1}$  it follows that for each  $1 \leq k \leq \mu(n, x)$  the  $k$ th letter of  $(U_{(n,x)})_{\text{rev}}$  is the  $k$ th letter of  $w_{n+1}$ . In other words  $(U_{(n,x)})_{\text{rev}}$  and  $w_{n+1}$  have a common prefix of length  $\mu(n, x)$ . Since  $\mu(n, x) \rightarrow \infty$  as  $n \rightarrow \infty$  the sequence of compositions  $\{\tau_{i_1} \circ \dots \circ \tau_{i_n}(x)\}_{n=1}^\infty$  converges to the accumulation point of the sequence  $\{w_n\}_{n=1}^\infty$  which is  $1_X$  as required. ■

COROLLARY III.9 <sup>(4)</sup>. *Each letter  $a \in A_k$  occurs in  $(i_n)$  with bounded gaps if and only if  $X$  is linearly recurrent in the sense of [15] or [13].*

PROOF. By Proposition III.7, each  $a \in A_k$  occurs in  $(i_n)$  with bounded gaps if and only if  $X$  is a primitive  $S$ -adic subshift (see §2.5 in [13]). The result now follows from Proposition 5 in [13]. ■

**IV. Primitive substitutive subshifts.** In this section we use Proposition III.7 to prove Theorem I.2.

DEFINITION IV.1. A sequence  $\omega$  in a finite alphabet  $A$  is called *primitive substitutive* if it is the image (under a letter-to-letter morphism) of a fixed

---

<sup>(4)</sup> Compare with Proposition 9 in [13].

point of a primitive substitution. A minimal symbolic space  $X$  is said to be *primitive substitutive* if it contains a primitive substitutive sequence.

Recall that a morphism  $\tau$  on a finite alphabet  $A$  is called *primitive* if there is a positive integer  $N$  such that for all  $a, b \in A$  the composition  $\tau^N(a)$  contains an occurrence of  $b$  (see [28]). We will need the following useful characterization of primitive substitutive sequences due to F. Durand <sup>(5)</sup>:

**THEOREM IV.2** (F. Durand, [12]). *A sequence  $\omega$  is primitive substitutive if and only if it has a finite number of derived sequences.*

*Proof of Theorem I.2.* In case  $(i_n)$  is eventually periodic, then by Proposition III.7 there exist words  $v$  and  $w$  in  $A_k$  such that for each  $x \in A_k$ ,

$$\mathbf{1}_X = \tau_v \circ \tau_w \circ \tau_w \circ \tau_w \circ \dots(x).$$

If  $u = u_1 \dots u_n$  with  $u_i \in A_k$ , then  $\tau_u$  denotes the composition  $\tau_{u_1} \circ \dots \circ \tau_{u_n}$ . Since each letter in  $A_k$  must occur in  $(i_n)$  an infinite number of times, the word  $w$  contains each letter in  $A_k$ , and hence  $\tau_w$  is a primitive morphism. Thus  $\mathbf{1}_X$  is the image (under the morphism  $\tau_v$ ) of the fixed point of the primitive morphism  $\tau_w$ . Using Proposition 3.1 in [12] we conclude that  $\mathbf{1}_X$  is primitive substitutive.

Conversely, suppose  $X$  is primitive substitutive. We use the following lemma proved in [19]:

**LEMMA IV.3** (C. Holton and L. Q. Zamboni, [19]). *Let  $Y$  be a primitive substitutive subshift. Then any point  $y \in Y$  having more than one backward extension is primitive substitutive.*

The above lemma implies that  $\mathbf{1}_X$  is a primitive substitutive sequence. By Proposition III.7,

$$\omega = \mathbf{1}_X = \lim_{n \rightarrow \infty} \tau_{i_1} \circ \dots \circ \tau_{i_n}(x).$$

For each  $m \geq 1$  set

$$\omega(m) = \lim_{n \rightarrow \infty} \tau_{i_m} \circ \tau_{i_{m+1}} \circ \dots \circ \tau_{i_{m+n}}(x).$$

Then for each  $m \geq 1$ ,

$$\omega(m+1) = \lim_{n \rightarrow \infty} \tau_{i_{m+1}} \circ \tau_{i_{m+2}} \circ \dots \circ \tau_{i_{m+n}}(x)$$

is (up to a bijection between  $A_k$  and  $\{1, \dots, k\}$ ) a derived sequence of  $\omega$ . In fact  $\omega(m+1)$  is (up to a bijection) the derived sequence of

$$\omega(m) = \lim_{n \rightarrow \infty} \tau_{i_m} \circ \tau_{i_{m+1}} \circ \dots \circ \tau_{i_n}(x)$$

---

<sup>(5)</sup> A slightly more general characterization of primitive substitutive sequences together with a characterization of primitive substitutive subshifts is given in [18].

with respect to the initial letter of  $\omega(m)$  (cf. [12] and [18]). More precisely, if  $a \in A_k$  denotes the initial letter of  $\omega(m)$ , then  $a$  is the unique bispecial word of length 1 in  $L(\omega(m))$  and therefore the return words to  $a$  are given by  $\{a\} \cup \{ab \mid b \in A_k \setminus \{a\}\}$  (cf. [12] or [18] for a precise definition of return words). The sequence  $\omega(m+1)$  is obtained from  $\omega(m)$  by coding the return words to  $a$  as follows <sup>(6)</sup>: the return word  $a$  is coded by the letter  $a$  while for each  $b \in A_k \setminus \{a\}$ , the return word  $ab$  is coded by the letter  $b$ . Since  $\omega$  has only finitely many induced sequences, the sequence  $(i_n)$  must be eventually periodic. ■

**V. Powers of words.** In this section we prove Propositions I.3 and I.4. We assume  $X$  is an Arnoux–Rauzy subshift on  $A_k$ .

PROPOSITION V.1. *Let  $X$  be an Arnoux–Rauzy subshift on  $A_k$ . There exists a positive number  $\epsilon_0 = \epsilon_0(k)$  and an infinite number of bispecial words of the form  $UVVV'$  with*

- $V'$  a prefix of  $V$ ,
- $\min\{|V'|/|V|, |V|/|U|\} \geq \epsilon_0$ .

COROLLARY V.2. *The characteristic sequence  $\mathbf{I}_X$  begins in an infinite number of words of the form  $WWw$  where  $w$  is a prefix of  $W$ .*

PROOF. Proposition V.1 implies that  $\mathbf{I}_X$  begins in an infinite number of bispecial blocks of the form  $UVVV'$  where  $V'$  is a prefix of  $V$ . Writing  $V = V'V''$  we obtain

$$UVVV' = (UVVV')_{\text{rev}} = (V'_{\text{rev}}V''_{\text{rev}})(V'_{\text{rev}}V''_{\text{rev}})V'_{\text{rev}}U_{\text{rev}}. \blacksquare$$

*Proof of Proposition V.1.* We begin with a series of lemmas. Let  $\mathbf{I}_X = (i_n) \in A_k^{\mathbb{N}}$  be the sequence defined in Definition III.1.

LEMMA V.3. *Let  $\{w_1, w_2, \dots\}$  be the set of bispecial words in  $L(X)$  ordered so that  $0 = |w_1| < |w_2| < \dots$ . For each  $n \geq 1$  we can write  $w_{n+1} = w_n i_n v_n$  for some suffix  $v_n$  of  $w_n$ . Moreover  $|v_n| \rightarrow \infty$  as  $n \rightarrow \infty$ .*

PROOF. This follows immediately from the definition of the function  $F$  defined in the previous section, the proof of Theorem III.5, and the fact that for each letter  $a \in A_k$ ,  $i_n = a$  for infinitely many  $n$ . In fact, if  $i_n \neq i_{n+1}$ , then  $|v_n| < |v_{n+1}|$ . ■

LEMMA V.4. *There exist  $a \in A_k$  and a word  $W$  (possibly empty) of length  $\leq k - 1$  such that the block  $aWa$  occurs an infinite number of times in  $(i_n)$ .*

---

<sup>(6)</sup> In [12] and [18] return words are coded by the “derived alphabet”  $\{1, \dots, k\}$  in order of first appearance.

PROOF. In fact for each block  $B$  of length  $k + 1$  in  $(i_n)$ , there is a letter in  $A_k$  which occurs twice in  $B$ . ■

LEMMA V.5. *Let  $a \in A_k$  be as in Lemma V.4. There exist  $K_1 > 0$  and an infinite number of words of the form  $wauau \in L(X)$  with  $w$ ,  $wau$ , and  $wauau$  each bispecial and  $|u| \leq K_1|w|$ .*

PROOF. Let  $a$  and  $W$  be as in Lemma V.4. In view of Lemma V.3 (applied  $|W|$  times), for each occurrence of  $aWa$  in  $(i_n)$  (say  $aWa = i_n i_{n+1} \dots i_{n+|W|} i_{n+|W|+1}$ ) we can write  $w_{n+|W|+1} = w_n a u_n$  and  $w_{n+|W|+2} = w_n a u_n a u_n$  for some word  $u_n$ . Moreover, since  $|W| \leq k - 1$ , it follows from the proof of Theorem III.5 that

$$|u_n| \leq |w_n| + (2|w_n| + 2) + (4|w_n| + 4) + \dots + (2^{k-1}|w_n| + 2^{k-1}).$$

Because for each  $n \leq j \leq n + |W|$ , we have  $|w_{j+1}| \leq 2|w_j| + 1$ . Hence there is a constant  $K_1 > 0$  (depending only on  $k$ ) so that  $|u_n| \leq K_1|w_n|$ . ■

We now return to the proof of Proposition V.1. By Lemma V.5 there exists  $K_1 > 0$  and an infinite number of words of the form  $wauau \in L(X)$  with  $w$ ,  $wau$ , and  $wauau$  each bispecial and  $|u| \leq K_1|w|$ . We consider two cases:

CASE 1:  $|w| \leq |u|$ . In this case  $w$  is a suffix of  $u$  and we can write  $u = vw$  for some word  $v$ . This gives the decomposition  $wauau = (wav)(wav)w$ . In this case we take  $V = wav$ ,  $V' = w$  and  $U = \varepsilon$  (the empty word). Then

$$\frac{|V'|}{|V|} = \frac{|w|}{|wav|} = \frac{|w|}{|u| + 1} \geq \frac{|w|}{2|u|} \geq \frac{1}{2K_1}.$$

CASE 2:  $|w| > |u|$ . We consider two subcases: In case  $|w| \geq 3(|u| + 1)$ , let  $K_2 \geq 3$  be the largest positive integer such that  $|w| \geq K_2(|u| + 1)$ . We write  $K_2 = 3r + s$  for some integer  $r \geq 1$  and  $s \in \{0, 1, 2\}$ . Since  $w$  is a suffix of  $wau$ , the defining condition of  $K_2$  implies that  $w = w'(au)^r(au)^r(au)^r$  for some word  $w'$  with  $|w'| < 3(|u| + 1)$ . In this case we take  $U = w'$  and  $V = V' = (au)^r$ . Then

$$\frac{|V|}{|U|} = \frac{|(au)^r|}{|w'|} = \frac{r(|u| + 1)}{|w'|} \geq \frac{r}{3} \geq \frac{1}{3}.$$

It remains to consider the case in which  $|u| < |w| < 3(|u| + 1)$ . Since  $u$  is a suffix of  $w$  we can write  $w = zu$  for some word  $z$ . This gives the decomposition  $wauau = z(ua)(ua)u$ . In this case we take  $U = z$ ,  $V = ua$ , and  $V' = u$ . Then

$$\frac{|V|}{|U|} = \frac{|ua|}{|z|} \geq \frac{|u| + 1}{|w|} > \frac{1}{3}$$

and

$$\frac{|V'|}{|V|} = \frac{|u|}{|u| + 1} \geq \frac{1}{2}.$$

To complete the proof of Proposition V.1 we have only to take  $\epsilon_0$  as the minimum of the two numbers  $\{1/(2K_1), 1/3\}$ . ■

*Proof of Proposition I.3.* We begin with a definition:

DEFINITION V.6. Given two occurrences of a subword  $w$  in a sequence  $x$ , say  $x = uw\dots = uvw\dots$  the word  $v$  is called the *offset* between these two occurrences of  $w$ .

LEMMA V.7. Let  $u$  and  $w$  be subwords of  $\mathbf{1}_X$  with  $uw$  bispecial. Then the offset between any two consecutive occurrences of  $w$  in  $\mathbf{1}_X$  has length at most  $|u| + |w| + 1$ .

PROOF. Let  $\{w_1, w_2, \dots\}$  be the set of all bispecial words in  $L(X)$  ordered as in Lemma V.3. Fix  $N$  so that  $w_N = uw$ . By Theorem III.5 for each  $n \geq N$  either  $w_{n+1} = w_n i_n w_n$  (case 1) or  $w_{n+1} = w_n i_n v_n$  for some suffix  $i_n v_n$  of  $w_n$  (case 2). In case 2 we can write  $w_n = w_j i_n v_n$  for some  $j < n$ . We divide case 2 into two subcases:  $|w_j| < |w|$  (case 2a) and  $|w_j| \geq |w|$  (case 2b). We first observe that in all cases  $w$  is a suffix of  $w_n$  since we are assuming  $n \geq N$ . If  $n = N$  then the offset between any two consecutive occurrences of  $w$  in  $w_n = w_N = uw$  has length at most  $|u| < |u| + |w| + 1$ . In case 2b we see that  $w$  is a suffix of both  $w_n$  and  $w_j$  and hence the length of the offset between any two occurrences of  $w$  does not increase in passing from  $w_n$  to  $w_{n+1}$ . On the other hand, in case 1 or case 2a the length of the offset between the last occurrence of  $w$  in  $w_n$  (viewed as an occurrence of  $w$  in  $w_{n+1}$  by regarding  $w_n$  as a prefix of  $w_{n+1}$ ) and the next occurrence of  $w$  in  $w_{n+1}$  is at most  $|u| + |w| + 1$ . Thus for all  $n \geq N$ , the offset between any two consecutive occurrences of  $w$  in  $w_n$  has length at most  $|u| + |w| + 1$  as required. ■

COROLLARY V.8. Let  $u$  and  $w$  be as in Lemma V.7. Then each sequence  $x \in X$  begins in a block of the form  $u'w$  where  $|u'| \leq |u| + |w| + 1$ .

Proposition I.3 now follows by combining Proposition V.1 and Corollary V.8. ■

*Proof of Proposition I.4.* We can suppose that  $\omega$  is a Sturmian sequence on the alphabet  $\{1, 2\}$ . Let  $X = X_\omega$  be the associated subshift and  $I_X = (i_n)$  as in Definition III.1. We will prove that unless  $I_X$  contains a periodic tail of the form  $12121212\dots$ , the sequence  $\omega$  contains arbitrarily large subwords of the form  $V^4$ . Let  $\{w_n\}$  be the set of all bispecial words in  $L(X)$  ordered so that  $0 = |w_1| < |w_2| < \dots$ . Recall that  $i_n w_n$  is right special for each  $n \geq 1$ . We consider three cases. First suppose that  $(i_n)$  contains infinitely many occurrences of either 111 or 222. Without loss of generality we can

suppose 111 occurs an infinite number of times in  $(i_n)$ . For all  $n$  sufficiently large, if  $i_{n+j} = 1$  for  $0 \leq j \leq 2$  then we can write  $w_n = u_n 1 v_n$  and  $w_{n+3} = w_n 1 v_n 1 v_n 1 v_n = u_n 1 v_n 1 v_n 1 v_n 1 v_n$  for some  $u_n, v_n \in L(X)$ .

In the second case we suppose that  $(i_n)$  contains finitely many occurrences of both 111 and 222 but infinitely many occurrences of either 11 or 22. Without loss of generality we can assume that 11 occurs an infinite number of times in  $(i_n)$ . Thus one of 121121, 1211221, 1221121, or 12211221 must occur an infinite number of times in  $(i_n)$ . Since each of these words begin in 12 we can write  $w_n = u_n \hat{1} v_n \hat{2} u'_n 1 v_n$  where  $u'_n$  and  $v_n$  are each a suffix of  $u_n$  and where the  $\hat{1}$  and  $\hat{2}$  correspond to the prefix 12 in each of the above words. If this occurrence of 12 in  $(i_n)$  is followed by 1121 or 11221 then it is easily verified using the algorithm in Theorem III.5 that  $w_{n+5}$  contains the subword  $(1v_n \hat{2} u'_n 1 v_n)^4$ . On the other hand if this occurrence of 12 in  $(i_n)$  is followed by 21121 or 211221 then  $w_{n+6}$  contains the subword  $(1v_n \hat{2} u'_n 1 v_n \hat{2} u'_n 1 v_n)^4$ .

In the final case we can suppose that  $(i_n)$  has finitely many occurrences of both 11 and 22. Then  $(i_n)$  is ultimately equal to the periodic sequence 1212121212... In this case  $\omega$  is the morphic image of the Fibonacci word; the result now follows by a theorem of Mignosi and Pirillo in [25] which states that the Fibonacci word contains arbitrarily large subwords of the form  $V^{(5+\sqrt{5})/2}$ . This completes the proof of Proposition I.4. ■

**VI. A connection with transcendental numbers.** It is generally believed that the digit expansion of an algebraic irrational number  $\alpha$  is very random <sup>(7)</sup> and cannot be generated by a simple algorithm. For instance the Champernowne number  $x = .1234567891011121314\dots$ , obtained by concatenating the decimal expansions of the consecutive integers, was shown to be transcendental by K. Mahler in [22]. J. Loxton and A. van der Poorten [20] stated that the digits in the  $k$ -ary expansion of an algebraic irrational could not be generated by a finite automaton, i.e., a deterministic machine having a finite number of allowable states.

In [17] Ferenczi and Mauduit derive the following combinatorial translation of a well known theorem of Ridout [21]:

**THEOREM VI.1** (S. Ferenczi and C. Mauduit, [17]). *Let  $\theta$  be an irrational number such that its  $k$ -ary expansion begins, for every integer  $n \in \mathbb{N}$ , in  $0.U_n V_n V_n V'_n$ , where  $U_n$  is a possibly empty word and where  $V_n$  is a non-empty word admitting  $V'_n$  as a prefix. If  $|V_n|$  tends to infinity,  $\limsup |U_n|/|V_n| < \infty$ , and  $\liminf |V'_n|/|V_n| > 0$ , then  $\theta$  is a transcendental number.*

---

<sup>(7)</sup> It is conjectured that an algebraic irrational number is normal in each base  $b \geq 1$ .



Theorem I.5 follows immediately from Theorem VI.1 and Proposition I.3. Theorem I.5 yields a class of transcendental numbers of complexity  $p(n) = (k-1)n+1$  for every integer  $k \geq 2$ . In the special case  $k \in \{2, 3\}$ , this result was proved by Ferenczi and Mauduit (see Propositions 2 and 4 in [17]) also by way of Theorem VI.1. Theorem VI.1 was also used by J.-P. Allouche and L. Q. Zamboni in [1] to show that a real number whose base  $b$  digit expansion is a fixed point of a binary morphism (either of constant length  $\geq 2$  or primitive) is either rational (if and only if the sequence is eventually periodic) or transcendental. Together with the theorem of Loxton and van der Poorten in [20] mentioned above, these are all results establishing the existence of transcendental numbers of complexity  $p(n)$  with  $kn < p(n) < k'n$  for any given  $k$  and some  $k'$ .

### References

- [1] J.-P. Allouche and L. Q. Zamboni, *Algebraic irrational binary numbers cannot be fixed points of non-trivial constant length or primitive morphisms*, J. Number Theory 69 (1998), 119–124.
- [2] P. Arnoux and S. Ito, *Pisot substitutions and Rauzy fractals*, preprint, 1999.
- [3] P. Arnoux et G. Rauzy, *Représentation géométrique de suites de complexité  $2n+1$* , Bull. Soc. Math. France 119 (1991), 199–215.
- [4] J. Berstel, *Mots de Fibonacci*, Séminaire d'Informatique Théorique, LITP, Universités Paris 6-7 (1980-1981), 57–78.
- [5] J. Berstel et P. Séébold, *Morphismes de Sturm*, Bull. Belg. Math. Soc. 1 (1994), 175–189.
- [6] V. Berthé, *Fréquences des facteurs des suites sturmiennes*, Theoret. Comput. Sci. 165 (1996), 295–309.
- [7] M. G. Castelli, F. Mignosi and A. Restivo, *Fine and Wilf's theorem for three periods and a generalization of sturmian words*, *ibid.* 218 (1999), 83–94.
- [8] N. Chekhova, *Les suites d'Arnoux–Rauzy : algorithme d'approximation et propriétés ergodiques*, preprint, 1998.
- [9] N. Chekhova, P. Hubert et A. Messaoudi, *Propriétés combinatoires, ergodiques et arithmétiques de la substitution de Tribonacci*, J. Théor. Nombres Bordeaux, to appear.
- [10] E. M. Coven and G. A. Hedlund, *Sequences with minimal block growth*, Math. Systems Theory 7 (1973), 138–153.
- [11] A. de Luca and F. Mignosi, *Some combinatorial properties of Sturmian words*, Theoret. Comput. Sci. 136 (1994), 361–385.
- [12] F. Durand, *A characterization of substitutive sequences using return words*, Discrete Math. 179 (1998), 89–101.
- [13] —, *Linearly recurrent subshifts*, Ergod. Theory Dynam. Systems, to appear.
- [14] F. Durand and B. Host, private communication.
- [15] F. Durand, B. Host and C. Skau, *Substitution dynamical systems, Bratteli diagrams and dimension groups*, Ergod. Theory Dynam. Systems 19 (1999), 953–993.

- [16] S. Ferenczi, *Les transformations de Chacon : combinatoire, structure géométrique, lien avec les systèmes de complexité  $2n + 1$* , Bull. Soc. Math. France 123 (1995), 271–292.
- [17] S. Ferenczi and C. Mauduit, *Transcendence of numbers with a low complexity expansion*, J. Number Theory 67 (1997), 146–161.
- [18] C. Holton and L. Q. Zamboni, *Descendants of primitive substitutions*, Theory Comput. Syst. 32 (1998), 133–157.
- [19] —, —, *Directed graphs and substitutions*, in: From Crystals to Chaos, P. Hubert, R. Lima and S. Vaienti (eds.), World Sci., 1999, to appear.
- [20] J. H. Loxton and A. van der Poorten, *Arithmetic properties of automata: regular sequences*, J. Reine Angew. Math. 392 (1988), 57–69.
- [21] K. Mahler, *Lectures on Diophantine Approximations*, Part I: *g-adic Numbers and Roth's Theorem*, Univ. of Notre Dame, 1961.
- [22] —, *Arithmetische Eigenschaften einer Klasse von Dezimalbrüchen*, Proc. Konink. Nederl. Akad. Wetensch. Ser. A 40 (1937), 421–428.
- [23] F. Mignosi, *Infinite words with linear subword complexity*, Theoret. Comput. Sci. 65 (1989), 221–242.
- [24] —, *On the number of factors of Sturmian words*, *ibid.* 82 (1991), 71–84.
- [25] F. Mignosi and G. Pirillo, *Repetitions in the Fibonacci infinite word*, RAIRO Inform. Théor. Appl. 26 (1992), 199–204.
- [26] M. Morse and G. A. Hedlund, *Symbolic dynamics*, Amer. J. Math. 60 (1938), 815–866.
- [27] —, —, *Symbolic dynamics II: Sturmian sequences*, *ibid.* 62 (1940), 1–42.
- [28] M. Queffélec, *Substitution Dynamical Systems—Spectral Analysis*, Lecture Notes in Math. 1294, Springer, 1987.
- [29] G. Rauzy, *Mots infinis en arithmétique*, in: Automata on Infinite Words, M. Nivat and D. Perrin (eds.), Lecture Notes in Comput. Sci. 192, Springer, Berlin, 1985, 165–171.
- [30] —, *Nombres algébriques et substitutions*, Bull. Soc. Math. France 110 (1982), 147–178.
- [31] G. Rote, *Sequences with subword complexity  $2n$* , J. Number Theory 46 (1994), 196–213.
- [32] J.-I. Tamura, *A class of transcendental numbers having explicit g-adic and Jacobi–Perron expansions of arbitrary dimension*, Acta Arith. 71 (1995), 301–329.
- [33] N. Wozny and L. Q. Zamboni, *Frequencies of factors in Arnoux–Rauzy sequences*, Acta Arith., to appear.
- [34] L. Q. Zamboni, *Une généralisation du théorème de Lagrange sur le développement en fraction continue*, C. R. Acad. Sci. Paris Sér. I 327 (1998), 527–530.

Department of Mathematics  
 University of North Texas  
 Denton, TX 76203-5116, U.S.A.  
 E-mail: rnr0002@jove.acs.unt.edu  
 luca@unt.edu

*Received on 23.12.1998  
 and in revised form on 24.5.1999*

(3532)