Self-shuffling Words[*]

Émilie Charlier[1], Teturo Kamae[2], Svetlana Puzynina[3,5],
and Luca Q. Zamboni[4,5]

[1] Département de Mathématique, Université de Liège, Belgium
echarlier@ulg.ac.be
[2] Advanced Mathematical Institute, Osaka City University, Japan
kamae@apost.plala.or.jp
[3] Sobolev Institute of Mathematics, Novosibirsk, Russia
svepuz@utu.fi
[4] Institut Camille Jordan, Université Lyon 1, France
lupastis@gmail.com
[5] FUNDIM, University of Turku, Finland

Abstract. In this paper we introduce and study a new property of infinite words which is invariant under the action of a morphism: We say an infinite word $x \in \mathbb{A}^{\mathbb{N}}$, defined over a finite alphabet $\mathbb{A}$, is self-shuffling if x admits factorizations: $x = \prod_{i=1}^{\infty} U_i V_i = \prod_{i=1}^{\infty} U_i = \prod_{i=1}^{\infty} V_i$ with $U_i, V_i \in \mathbb{A}^{+}$. In other words, there exists a shuffle of x with itself which reproduces x. The morphic image of any self-shuffling word is again self-shuffling. We prove that many important and well studied words are self-shuffling: This includes the Thue-Morse word and all Sturmian words (except those of the form aC where $a \in \{0, 1\}$ and C is a characteristic Sturmian word). We further establish a number of necessary conditions for a word to be self-shuffling, and show that certain other important words (including the paper-folding word and infinite Lyndon words) are not self-shuffling. In addition to its morphic invariance, which can be used to show that one word is not the morphic image of another, this new notion has other unexpected applications: For instance, as a consequence of our characterization of self-shuffling Sturmian words, we recover a number theoretic result, originally due to Yasutomi, which characterizes pure morphic Sturmian words in the orbit of the characteristic.

1   Introduction

Let $\mathbb{A}$ be a finite non-empty set. We denote by $\mathbb{A}^{*}$ the set of all finite words $u = x_1 x_2 \ldots x_n$ with $x_i \in \mathbb{A}$. The quantity n is called the length of u and is denoted $|u|$. For a letter $a \in \mathbb{A}$, by $|u|_a$ we denote the number of occurrences of

---

a in u. The empty word, denoted ε, is the unique element in $\mathbb{A}^*$ with $|\varepsilon| = 0$. We set $\mathbb{A}^+ = \mathbb{A} - \{\varepsilon\}$. We denote by $\mathbb{A}^{\mathbb{N}}$ the set of all one-sided infinite words $x = x_0 x_1 x_2 \ldots$ with $x_i \in \mathbb{A}$.

Given k finite or infinite words $x^{(1)}, x^{(2)}, \ldots, x^{(k)} \in A^* \cup \mathbb{A}^{\mathbb{N}}$ we denote by

$$\mathscr{S}(x^{(1)}, x^{(2)}, \ldots, x^{(k)}) \subset \mathbb{A}^* \cup \mathbb{A}^{\mathbb{N}}$$

the collection of all words z for which there exists a factorization

$$z = \prod_{i=0}^{\infty} U_i^{(1)} U_i^{(2)} \cdots U_i^{(k)}$$

with each $U_i^{(j)} \in \mathbb{A}^*$ and with $x^{(j)} = \prod_{i=0}^{\infty} U_i^{(j)}$ for $1 \leq j \leq k$. Intuitively, z may be obtained as a *shuffle* of the words $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$. In case $x^{(1)}, x^{(2)}, \ldots, x^{(k)} \in A^*$, each of the above products can be taken to be finite.

Finite word shuffles were extensively studied in [5]. Given $x \in \mathbb{A}^*$, it is generally a difficult problem to determine whether there exists $y \in \mathbb{A}^*$ such that $x \in \mathscr{S}(y, y)$ (see Open Problem 4 in [5]). However, in the context of infinite words, this question is essentially trivial: In fact, it is readily verified that if $x \in \mathbb{A}^{\mathbb{N}}$ is such that each $a \in \mathbb{A}$ occurring in x occurs an infinite number of times in x, then there exist infinitely many $y \in \mathbb{A}^{\mathbb{N}}$ with $x \in \mathscr{S}(y, y)$. Instead, in the framework of infinite words, a far more delicate question is the following:

Question 1. Given $x \in \mathbb{A}^{\mathbb{N}}$, does there exist an integer $k \geq 2$ such that $x \in \mathscr{S}(\underbrace{x, x, \ldots, x}_{k})$?

If such a k exists, we say x is *k-self-shuffling*.

Given $x = x_0 x_1 x_2 \ldots \in \mathbb{A}^{\mathbb{N}}$ and an infinite subset $N = \{N_0 < N_1 < N_2 < \ldots\} \subseteq \mathbb{N}$, we put $x[N] = x_{N_0} x_{N_1} x_{N_2} \ldots \in \mathbb{A}^{\mathbb{N}}$. Alternatively,

Definition 1. For $x \in \mathbb{A}^{\mathbb{N}}$ and $k = 2, 3, \ldots$, we say x is *k-self-shuffling* if there exists a k-element partition $\mathbb{N} = \bigcup_{i=1}^{k} N^i$ with $x[N^i] = x$ for each $i = 1, \ldots, k$.

In case $k = 2$, we say simply x is *self-shuffling*. We note that if x is k-self-shuffling, then x is ℓ-self-shuffling for each $\ell \geq k$ but not conversely (see §2), whence each self-shuffling word is k-self-shuffling for all $k \geq 2$. In this paper we are primarily interested in self-shuffling words, however, many of the results presented here extend to general k. Thus $x \in \mathbb{A}^{\mathbb{N}}$ is self-shuffling if and only if x admits factorizations

$$x = \prod_{i=1}^{\infty} U_i V_i = \prod_{i=1}^{\infty} U_i = \prod_{i=1}^{\infty} V_i$$

with $U_i, V_i \in \mathbb{A}^+$.

The property of being self-shuffling is an intrinsic property of the word (and not of the associated language) and seems largely independent of its complexity

(examples exist from the lowest to the highest possible complexity). The simplest class of self-shuffling words consists of all (purely) periodic words $x = u^\omega$. It is clear that if x is self-shuffling, then every letter $a \in \mathbb{A}$ occurring in x must occur an infinite number of times. Thus for instance, the ultimately periodic word 01^ω is not self-shuffling. As we shall see, many well-known words which are of interest in both combinatorics on words and symbolic dynamics, are self-shuffling. This includes for instance the famous *Thue-Morse* word

$$\mathbf{T} = 0110100110010110100101100110100110010110\ldots$$

whose origins go back to the beginning of the last century with the works of the Norwegian mathematician Axel Thue [9]. The nth entry t_n of $\mathbf{T}$ is defined as the sum modulo 2 of the digits in the binary expansion of n. While the Thue-Morse word appears naturally in many different areas of mathematics (from discrete mathematics to number theory to differential geometry-see [1] or [2]), proving that Thue-Morse is self-shuffling is somewhat more involved than expected.

Sturmian words constitute another important class of aperiodic self-shuffling words. Sturmian words are infinite words over a binary alphabet having exactly $n+1$ factors of length n for each $n \geq 0$ [7]. Their origin can be traced back to the astronomer J. Bernoulli III in 1772. They arise naturally in many different areas of mathematics including combinatorics, algebra, number theory, ergodic theory, dynamical systems and differential equations. Sturmian words are also of great importance in theoretical physics and in theoretical computer science and are used in computer graphics as digital approximation of straight lines. We show that all Sturmian words are self-shuffling except those of the form aC where $a \in \{0,1\}$ and C is a characteristic Sturmian word. Thus for every irrational number α, all (uncountably many) Sturmian words of slope α are self-shuffling except for two. Our proof relies on a geometric characterization of Sturmian words via irrational rotations on the circle.

So while there are many natural examples of aperiodic self-shuffling words, the property of being self-shuffling is nevertheless quite restrictive. We obtain a number of necessary (and in some cases sufficient) conditions for a word to be self-shuffling. For instance, if a word x is self-shuffling, then x begins in only finitely many Abelian border-free words. As an application of this we show that the well-known *paper folding word* is not self-shuffling. Infinite Lyndon words (i.e., infinite words which are lexicographically smaller than each of its suffixes) are also shown not to be self-shuffling.

One important feature of self-shuffling words stems from its invariance under the action of a morphism: The morphic image of a self-shuffling word is again self-shuffling. In some instances this provides a useful tool for showing that one word is not the morphic image of another. So for instance, the paper folding word is not the morphic image of any self-shuffling word. However this application requires knowing a priori whether a given word is or is not self-shuffling. In general, to show that a word is self-shuffling, one must actually exhibit a shuffle. Self-shuffling words have other unexpected applications particularly in the study of fixed points of substitutions. For instance, as an almost immediate consequence

of our characterization of self-shuffling Sturmian words, we recover a result, first proved by Yasutomi via number theoretic methods, which characterizes pure morphic Sturmian words in the orbit of the characteristic.

2   Examples and Non-examples

In this section we list some examples and non-examples of self-shuffling words. As usual in combinatorics on words, we follow notation from [7].

Fibonacci Word: The Fibonacci infinite word

$$x = 0100101001001010010100\ldots$$

is defined as the fixed point of the morphism φ given by $0 \mapsto 01$, $1 \mapsto 0$. It is readily verified that $\varphi^2(a) = \varphi(a)a$ for each $a \in \{0,1\}$. Whence, writing $x = x_0 x_1 x_2 \ldots$ with each $x_i \in \{0,1\}$ we obtain

$$x = x_0 x_1 x_2 \ldots = \varphi(x_0)\varphi(x_1)\varphi(x_2)\ldots = \varphi^2(x_0)\varphi^2(x_1)\varphi^2(x_2)\ldots = $$
$$\varphi(x_0)x_0\varphi(x_1)x_1\varphi(x_2)x_2 \ldots$$

which shows that x is self-shuffling. In contrast, the word $y = 0x$ is not self-shuffling. The word y starts with infinitely many prefixes of the form $0B1$ with B a palindrome. It follows that $0B1$ is *Abelian border-free* (i.e., no proper suffix of $0B1$ is Abelian equivalent to a proper prefix of $0B1$). By Proposition 3 the word y is not self-shuffling.

Paper-folding Word: The paper-folding word

$$x = 00100110001101100010\ldots$$

is a Toeplitz word generated by the pattern $u = 0?1?$ (see, e.g., [4]). It is readily verified that x begins in arbitrarily long Abelian border-free words and hence by Proposition 3 is not self-shuffling. More precisely, the prefixes u_j of x of length $n_j = 2^j - 1$ are Abelian border-free. Indeed, it is verified that for each $k < n_j$, we have $|\mathrm{pref}_k(u_j)|_0 > k/2$ while $|\mathrm{suff}_k(u_j)|_0 \leq k/2$. Here $\mathrm{pref}_k(u)$ (resp., $\mathrm{suff}_k(u)$) denotes the prefix (resp., suffix) of length k of a word u.

A 3-Self-shuffling Word Which Is Not Self-shuffling: Let y denote the fixed point of the morphism $\sigma : 0 \mapsto 0001$ and $1 \mapsto 0101$, and put

$$x = 0^{-2}y = 0100010001010100010001000101010001000100010101010001010100\ldots,$$

where the notation $w = v^{-k}u$ means that $u = v^k w$. Then for each prefix u_j of x of length $4^j - 2$, the longest Abelian border of u_j of length less than or equal to $(4^j - 2)/2$ has length 2. Hence x is not self-shuffling (see Proposition 3). The 3-shuffle is given by the following:

$$U_0 = 0100, U_1 = 01, \quad \ldots, \ U_{4i+2} = \varepsilon, \qquad\qquad U_{4i+3} = \sigma^{i+1}(0100),$$
$$U_{4i+4} = \sigma(0), \qquad\qquad U_{4i+5} = (\sigma(0))^{-1}\sigma^{i+1}(01),$$
$$V_0 = 0100, V_1 = 01, \quad \ldots, \ V_{4i+2} = (\sigma(0))^{-1}\sigma^{i+1}(0), \quad V_{4i+3} = \sigma(0),$$
$$V_{4i+4} = (\sigma(0))^{-1}\sigma^{i+1}(01)\sigma(0), \ \ V_{4i+5} = \varepsilon,$$
$$W_0 = 01, \ \ W_1 = (\sigma(0))^2, \ldots, \ W_{4i+2} = \varepsilon, \qquad\qquad W_{4i+3} = (\sigma(0))^{-1}\sigma^{i+1}(01),$$
$$W_{4i+4} = \varepsilon, \qquad\qquad W_{4i+5} = \sigma^{i+2}(0)\sigma(0).$$

It is then verified that

$$x = \prod_{i=0}^{\infty} U_i V_i W_i = \prod_{i=0}^{\infty} U_i = \prod_{i=0}^{\infty} V_i = \prod_{i=0}^{\infty} W_i,$$

from which it follows that x is 3-self-shuffling.

A Recurrent Binary Self-shuffling Word with Full Complexity: For each positive integer n, let z_n denote the concatenation of all words of length n in increasing lexicographic order. For example, $z_2 = 00011011$. For $i \geq 0$ put

$$v_i = \begin{cases} z_n, & \text{if } i = n2^{n-1} \text{ for some } n, \\ 0^i 1^i, & \text{otherwise,} \end{cases}$$

and define

$$x = \prod_{i=0}^{\infty} X_i = 010100110^3 011^3 0^4 010^2 1^2 011^4 \ldots,$$

where $X_0 = X_1 = 01$, $X_2 = 0011$, and for $i \geq 3$, $X_i = 0^i y_{i-2} 1^i$, where $y_{i-2} = y_{i-3} v_{i-2} y_{i-3}$, and $y_0 = \varepsilon$. We note that x is recurrent (i.e., each prefix occurs twice) and has full complexity (since it contains z_n as a factor for every n).

To show that the word x is self-shuffling, we first show that $X_{i+1} \in \mathcal{S}(X_i, X_i)$. Take $N_i = \{0, \ldots, i-1, i+1, \ldots, 2^i - i, 2^i - i + v_{i-1}|_1, 2^{i+1} - i - 1\}$, where $u|_1$ denotes the positions j of a word u in which the j-th letter u_j of u is equal to 1. Then it is straightforward to see that $X_i = X_{i+1}[N_i] = X_{i+1}[\{1, \ldots, 2^{i+1}\} \backslash N_i]$. The self-shuffle of x is built in a natural way concatenating shuffles of X_i starting with $U_0 = V_0 = 01$, so that $X_0 \ldots X_{i+1} \in \mathcal{S}(X_0 \ldots X_i, X_0 \ldots X_i)$.

3   General Properties

In this section we develop several fundamental properties of self-shuffling words. The next two propositions show the invariance of self-shuffling words with respect to the action of a morphism:

Proposition 1. *Let $\mathbb{A}$ and $\mathbb{B}$ be finite non-empty sets and $\tau : \mathbb{A} \to \mathbb{B}^*$ a morphism. If $x \in \mathbb{A}^{\mathbb{N}}$ is self-shuffling, then so is $\tau(x) \in \mathbb{B}^{\mathbb{N}}$.*

Proof. If $x \in \mathscr{S}(x, x)$, then we can write $x = \prod_{i=1}^{\infty} U_i V_i = \prod_{i=1}^{\infty} U_i = \prod_{i=1}^{\infty} V_i$. Whence $\tau(x) = \prod_{i=1}^{\infty} \tau(U_i V_i) = \prod_{i=1}^{\infty} \tau(U_i) \tau(V_i) = \prod_{i=1}^{\infty} \tau(U_i) = \prod_{i=1}^{\infty} \tau(V_i)$ as required.

Proposition 2. *Let $\tau : \mathbb{A} \to \mathbb{A}^*$ be a morphism, and $x \in \mathbb{A}^{\mathbb{N}}$ be a fixed point of τ.*

1. *Let u be a prefix of x and k be a positive integer such that $\tau^k(a)$ begins in u for each $a \in \mathbb{A}$. Then if x is self-shuffling, then so is $u^{-1}x$.*
2. *Let $u \in \mathbb{A}^*$, and let k be a positive integer such that $\tau^k(a)$ ends in u for each $a \in \mathbb{A}$. Then if x is self-shuffling, then so is ux.*

Proof. We prove only item (1) since the proof of (2) is essentially identical. Suppose $x = \prod_{i=1}^{\infty} U_i V_i = \prod_{i=1}^{\infty} U_i = \prod_{i=1}^{\infty} V_i$. Then by assumption, for each $i \geq 1$ we can write $\tau^k(U_i) = uU_i'$ and $\tau^k(V_i) = uV_i'$ for some $U_i', V_i' \in \mathbb{A}^*$. Put $X_i = U_i'u$ and $Y_i = V_i'u$. Then since

$$x = \tau^k(x) = \prod_{i=1}^{\infty} \tau^k(U_i V_i) = \prod_{i=1}^{\infty} \tau^k(U_i)\tau^k(V_i) = \prod_{i=1}^{\infty} \tau^k(U_i) = \prod_{i=1}^{\infty} \tau^k(V_i),$$

we deduce that

$$u^{-1}x = \prod_{i=1}^{\infty} X_i Y_i = \prod_{i=1}^{\infty} X_i = \prod_{i=1}^{\infty} Y_i.$$

Corollary 1. *Let $\tau : \mathbb{A} \to \mathbb{A}^*$ be a primitive morphism, and $a \in \mathbb{A}$. Suppose $\tau(b)$ begins (respectively ends) in a for each letter $b \in \mathbb{A}$. Suppose further that the fixed point $\tau^{\infty}(a)$ is self-shuffling. Then every right shift (respectively left shift) of $\tau^{\infty}(a)$ is self-shuffling.*

Remark 1. Since the Fibonacci word is self-shuffling and is fixed by the primitive morphism $0 \mapsto 01$, $1 \mapsto 0$, it follows from Corollary 1 that every tail of the Fibonacci word is self-shuffling.

There are a number of necessary conditions that a self-shuffling word must satisfy, which may be used to deduce that a given word is not self shuffling. For instance:

Proposition 3. *If $x \in \mathbb{A}^{\mathbb{N}}$ is self-shuffling, then for each positive integer N there exists a positive integer M such that every prefix u of x with $|u| \geq M$ has an Abelian border v with $|u|/2 \geq |v| \geq N$. In particular, x must begin in only a finite number of Abelian border-free words.*

Proof. Suppose to the contrary that there exist factorizations $x = \prod_{i=0}^{\infty} U_i V_i = \prod_{i=0}^{\infty} U_i = \prod_{i=0}^{\infty} V_i$ with $U_i, V_i \in \mathbb{A}^+$, and there exists N such that for every M there exists a prefix u of x with $|u| \geq M$ which has no Abelian borders of length between N and $|u|/2$. Take $M = |\prod_{i=0}^{N} U_i V_i|$ and a prefix u satisfying these conditions. Then there exist non-empty proper prefixes U' and V' of u such that $u \in \mathscr{S}(U', V')$ with $|U'|, |V'| > N$. Writing $u = U'U''$ it follows that U'' and V' are Abelian equivalent. This contradicts that u has no Abelian borders of length between N and $|u|/2$.

An extension of this argument gives both a necessary and sufficient condition for self-shuffling in terms of Abelian borders (which is however difficult to check in practice). For $u \in \mathbb{A}^*$ let $\Psi(u)$ denote the *Parikh vector* of u, i. e., $\Psi(u) = (|u|_a)_{a \in \mathbb{A}}$.

Definition 2. Given $x \in \mathbb{A}^{\mathbb{N}}$, we define a directed graph $G_x = (V_x,\ E_x)$ with vertex set

$$V_x = \{(n, m) \in \mathbb{N}^2 \mid \Psi(\mathrm{pref}_n x) + \Psi(\mathrm{pref}_m x) = \Psi(\mathrm{pref}_{n+m} x)\}$$

and the edge set

$$E_x = \{\, ((n, m),\ (n', m')) \in V_x \times V_x \mid$$
$$n' = n + 1 \text{ and } m' = m \text{ or } m' = m + 1 \text{ and } n' = n\}.$$

We say that G_x *connects* $\mathbf{0}$ *to* $\boldsymbol{\infty}$ if there exists an infinite path $\prod_{j=1}^{\infty}(n_j, m_j)$ in G_x such that $(n_0, m_0) = (0,0)$ and $n_j, m_j \to \infty$ as $j \to \infty$.

Theorem 1. *A word $x \in \mathbb{A}^{\mathbb{N}}$ is self-shuffling if and only if the graph G_x connects* $\mathbf{0}$ *to* $\boldsymbol{\infty}$.

The theorem gives a constructive necessary and sufficient condition for self-shuffling since a path to infinity defines a self-shuffle.

As we shall now see, lexicographically extremal words are never self-shuffling. Let $(\mathbb{A}, \leq)$ be a finite linearly ordered set. Then $\leq$ induces the lexicographic ordering $\leq_{\mathrm{lex}}$ on $\mathbb{A}^+$ and $\mathbb{A}^{\mathbb{N}}$ defined as follows: If $u, v \in \mathbb{A}^+$ (or $\mathbb{A}^{\mathbb{N}}$) we write $u \leq_{\mathrm{lex}} v$ if either $u = v$ or if u is lexicographically smaller than v. In the latter case we write $u <_{\mathrm{lex}} v$.

Let $x \in \mathbb{A}^{\mathbb{N}}$. A factor u of x is called *minimal* (in x) if $u \leq_{\mathrm{lex}} v$ for all factors v of x with $|v| = |u|$. An infinite word y in the shift orbit closure S_x of x is called *Lyndon* (in S_x) if every prefix of y is minimal in x. The proof of the following result is omitted for space considerations:

Theorem 2. *Let $(A, \leq)$ be a linearly ordered finite set and let $x \in A^{\mathbb{N}}$. Let $y, z \in S_x$ with y Lyndon and aperiodic. Then for each $w \in \mathscr{S}(y, z)$, we have $w <_{\mathrm{lex}} z$. In particular, taking $z = y$ we deduce that y is not self-shuffling.*

Let $\mathbb{A}$ be a finite non-empty set. We say $x \in \mathbb{A}^{\mathbb{N}}$ is *extremal* if there exists a linear ordering $\leq$ on $\mathbb{A}$ with respect to which x is Lyndon. As an immediate consequence of Theorem 2 we obtain:

Corollary 2. *Let $\mathbb{A}$ be a finite non-empty set and $x \in \mathbb{A}^{\mathbb{N}}$ be an aperiodic extremal word. Then x is not self-shuffling.*

Remark 2. Let $x = 11010011001011010010110\ldots$ denote the first shift of the Thue-Morse infinite word. It is easily checked that x is extremal and hence is not self-shuffling; yet it can be verified that x begins in only a finite number of Abelian border-free words.

4   The Thue-Morse Word Is Self-shuffling

Theorem 3. *The Thue-Morse word* $\mathbf{T} = 011010011001\ldots$ *fixed by the morphism* τ *mapping* $0 \mapsto 01$ *and* $1 \mapsto 10$ *is self-shuffling.*

Proof. For $u \in \{0,1\}^*$ we denote by $\bar{u}$ the word obtained from u by exchanging 0s and 1s. Let $\sigma : \{1,2,3,4\} \rightarrow \{1,2,3,4\}^*$ be the morphism defined by

$$\sigma(1) = 12, \quad \sigma(2) = 31, \quad \sigma(3) = 34, \quad \sigma(4) = 13.$$

Set $u = 01101$ and $v = 001$; note that uv is a prefix of $\mathbf{T}$. Also define morphisms $g, h : \{1,2,3,4\} \rightarrow \{0,1\}^*$ by

$$g(1) = v\bar{u}, \quad g(2) = \bar{v}\bar{u}, \quad g(3) = \bar{v}u, \quad g(4) = vu$$

and

$$h(1) = uv, \quad h(2) = \bar{u}\bar{v}, \quad h(3) = \bar{u}\bar{v}, \quad h(4) = uv$$

We will make use of the following lemmas:

Lemma 1. $g(\sigma(a)) \in \mathscr{S}(g(a), h(a))$ *for each* $a \in \{1,2,3,4\}$. *In particular* $ug(\sigma(1)) \in \mathscr{S}(ug(1), h(1))$.

Proof. For $a = 1$ we note that

$$g(\sigma(1)) = g(12) = v\bar{u}\bar{v}\bar{u} = 0011001011010010.$$

Factoring $0011001011010010 = 0 \cdot 011 \cdot 0 \cdot 010 \cdot 11 \cdot 01 \cdot 0010$ we obtain

$$g(\sigma(1)) \in \mathscr{S}(00110010, 01101001) = \mathscr{S}(v\bar{u}, uv) = \mathscr{S}(g(1), h(1)).$$

Similarly, for $a = 2$ we have

$$g(\sigma(2)) = g(31) = \bar{v}uv\bar{u} = 1100110100110010.$$

Factoring $1100110100110010 = 1 \cdot 100 \cdot 1 \cdot 1 \cdot 010 \cdot 0110 \cdot 010$ we obtain

$$g(\sigma(2)) \in \mathscr{S}(11010010, 10010110) = \mathscr{S}(\bar{v}\bar{u}, \bar{u}\bar{v}) = \mathscr{S}(g(2), h(2)).$$

Exchanging 0s and 1s in the previous two shuffles yields

$$g(\sigma(3)) = g(34) = \bar{v}uvu \in \mathscr{S}(\bar{v}u, \bar{u}\bar{v}) = \mathscr{S}(g(3), h(3))$$

and

$$g(\sigma(4)) = g(13) = v\bar{u}\bar{v}u \in \mathscr{S}(vu, uv) = \mathscr{S}(g(4), h(4)).$$

It is readily verified that

Lemma 2. $h(\sigma(a)) = \tau(h(a))$ *for each* $a \in \{1,2,3,4\}$.

Let $w = w_0 w_1 w_2 w_3 \ldots$ with $w_i \in \{1,2,3,4\}$ denote the fixed point of σ beginning in 1. As a consequence of the previous lemma we deduce that

Lemma 3. T $= h(w)$.

Proof. In fact $\tau(h(w)) = h(\sigma(w)) = h(w)$ from which it follows that $h(w)$ is one of the two fixed points of τ. Since $h(w)$ begins in $h(1)$ which in turn begins in 0, it follows that $\mathbf{T} = h(w)$.

Lemma 4. T $= ug(w)$.

Proof. It is readily verified that:

$$ug(1) = h(1)\bar{u}$$

$$\bar{u}g(2) = h(2)\bar{u}$$

$$\bar{u}g(3) = h(3)u$$

$$ug(4) = h(4)u.$$

Moreover, each occurrence of $g(1)$ and $g(4)$ in $ug(w)$ is preceded by u while each occurrence of $g(2)$ and $g(3)$ in $ug(w)$ is preceded by $\bar{u}$. It follows that $ug(w) = h(w)$ which by the preceding lemma equals $\mathbf{T}$.

Set

$$A_0 = ug(\sigma(w_0)) \quad \text{and} \quad A_i = g(\sigma(w_i)), \quad \text{for} \ \ i \geq 1$$

$$B_0 = ug(w_0)) \quad \text{and} \quad B_i = g(w_i)), \quad \text{for} \ \ i \geq 1$$

and

$$C_i = h(w_i) \quad \text{for} \ \ i \geq 0.$$

It follows from Lemma 3 and Lemma 4 that

$$\mathbf{T} = \prod_{i=0}^{\infty} A_i = \prod_{i=0}^{\infty} B_i = \prod_{i=0}^{\infty} C_i$$

and it follows from Lemma 1 that $A_i \in \mathscr{S}(B_i, C_i)$ for each $i \geq 0$. Hence $\mathbf{T} \in \mathscr{S}(\mathbf{T}, \mathbf{T})$ as required.

5   Self-shuffling Sturmian Words

In this section we characterize self-shuffling Sturmian words. Sturmian words admit various types of characterizations of geometric and combinatorial nature, e.g., they can be defined via balance, complexity, morphisms, etc. (see Chapter 2 in [7]). In [8], Morse and Hedlund showed that each Sturmian word may be realized geometrically by an irrational rotation on the circle. More precisely, every Sturmian word x is obtained by coding the symbolic orbit of a point $\rho(x)$ on the circle (of circumference one) under a rotation by an irrational angle α where the circle is partitioned into two complementary intervals, one of length α (labeled 1) and the other of length $1 - \alpha$ (labeled 0) (see Fig. 1). And conversely each such coding gives rise to a Sturmian word. The irrational α is called the *slope*

and the point $\rho(x)$ is called the *intercept* of the Sturmian word x. A Sturmian word x of slope α with $\rho(x) = \alpha$ is called a *characteristic Sturmian word*. It is well known that every prefix u of a characteristic Sturmian word is *left special*, i.e., both $0u$ and $1u$ are factors of x [7]. Thus if x is a characteristic Sturmian word of slope α, then both $0x$ and $1x$ are Sturmian words of slope α and $\rho(0x) = \rho(1x) = 0$. The fact that ρ is not one-to-one stems from the ambiguity of the coding of the boundary points 0 and $1 - \alpha$.
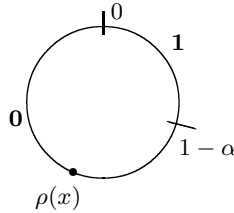

Fig. 1. Geometric picture of a Sturmian word of slope α

Theorem 4. *Let S, M and L be Sturmian words of the same slope α, $0 < \alpha < 1$, satisfying $S \leq_{\text{lex}} M \leq_{\text{lex}} L$. Then $M \in \mathscr{S}(S,L)$ if and only if the following conditions hold: If $\rho(M) = \rho(S)$ (respectively, $\rho(M) = \rho(L)$), then $\rho(L) \neq 0$ (respectively $\rho(S) \neq 0$).*

In particular (taking $S = M = L$), we obtain

Corollary 3. *A Sturmian word $x \in \{0,1\}^{\mathbb{N}}$ is self-shuffling if and only if $\rho(x) \neq 0$, or equivalently, x is not of the form aC where $a \in \{0,1\}$ and C is a characteristic Sturmian word.*

Our proof explicitly describes an algorithm for shuffling S and L so as to produce M. It is formulated in terms of the circle rotation description of Sturmian words. Geometrically speaking, points $\rho(S)$ and $\rho(L)$ will take turns following the trajectory of $\rho(M)$ so that the respective codings agree; as one follows the other waits its turn (remains neutral). The algorithm specifies this following rule depending on the relative positions of the trajectories of all three points and is broken down into several cases. The proof can be summarized by the directed graph in Fig. 2 in which each state n corresponds to "case n" in the proof.

We let s, m, and ℓ denote the current tail of the words S, M, and L. They are initialized as

$$s := S, \ \ell := L, \text{ and } m := M.$$

While m is always a tail of M, the letters s and ℓ may be tails of S or L, depending on which is the current lexicographically largest[1]. Each directed edge

---

[1] The choice of the letter s, m, and ℓ is intended to refer to *small, medium,* and *large* respectively.
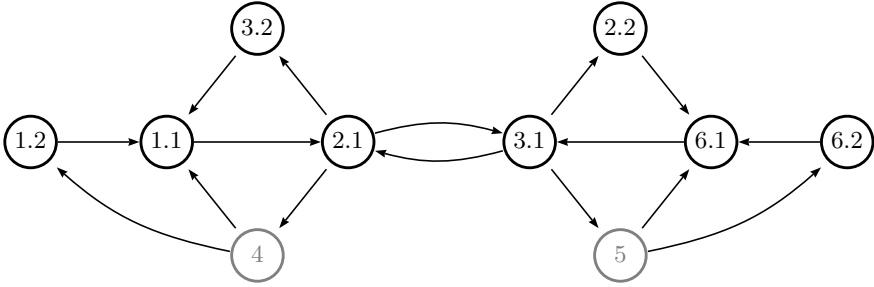
Fig. 2. Graphical depiction of the proof of Theorem 4

corresponds to a precise set of instructions which specify which of s or ℓ is neutral, which of s or ℓ follows m and for how long, and in the end a possible relabeling of the variables s and ℓ. In each case the outcome leads to a new case in which there is a switch in the follower. In other words, if there is an edge from case i to case j in the graph, then either the instructions for case i and case j specify different followers (as is the case for cases 1.1 and 2.1) in which case the passage from i to j leaves the labeling of s and ℓ unchanged, or the instructions for case i and case j specify the same follower (as is the case for cases 1.2 and 1.1) in which case the passage from i to j exchanges the labeling of s and ℓ. The proof of Theorem 4 amounts to showing that for each state n in the graph, the specified instructions will take n to an adjacent state in the graph.

As an almost immediate application of Corollary 3 we recover the following result originally proved by Yasutomi in [10] and later reproved by Berthé, Ei, Ito and Rao in [3] and independently by Fagnot in [6]. We say an infinite word is *pure morphic* if it is a fixed point of some morphism different from the identity.

Theorem 5 (Yasutomi [10]). *Let $x \in \{0,1\}^{\mathbb{N}}$ be a characteristic Sturmian word. If y is a pure morphic word in the orbit of x, then $y \in \{x, 0x, 1x, 01x, 10x\}$.*

Proof. We begin with some preliminary observations. Let $\Omega(x)$ denote the set of all left and right infinite words y such that $\mathcal{F}(x) = \mathcal{F}(y)$ where $\mathcal{F}(x)$ and $\mathcal{F}(y)$ denote the set of all factors of x and y respectively. If $y \in \Omega(x)$ is a right infinite word, and $0y, 1y \in \Omega(x)$, then $y = x$. This is because every prefix of y is a left special factor and hence also a prefix of the characteristic word x. Similarly if y is a left infinite word and $y0, y1 \in \Omega(x)$, then y is equal to the reversal of x. If τ is a morphism fixing some point $y \in \Omega(x)$, then $\tau(z) \in \Omega(x)$ for all $z \in \Omega(x)$.

Suppose to the contrary that $\tau \neq id$ is a morphism fixing a proper tail y of x. Then y is self-shuffling by Corollary 3. Put $x = uy$ with $u \in \{0,1\}^{+}$. Using the characterization of Sturmian morphisms (see Theorem 2.3.7 & Lemma 2.3.13 in [7]) we deduce that τ must be primitive. Thus we can assume that $|\tau(a)| > 1$ for each $a \in \{0,1\}$. If $\tau(0)$ and $\tau(1)$ end in distinct letters, then as both $0\tau(x), 1\tau(x) \in \Omega(x)$, it follows that $\tau(x) = x$. Since also $\tau(y) = y$ and $|\tau(u)| > |u|$, it follows that y is a proper tail of itself, a contradiction since x is aperiodic. Thus $\tau(0)$ and $\tau(1)$ must end in the same letter. Whence by Corollary 1 it follows

that every left extension of y is self-shuffling, which is again a contradiction since $0x$ and $1x$ are not self-shuffling.

Next suppose $\tau \neq id$ is a morphism fixing a point $y = uabx \in \Omega(x)$ where $u \in \{0,1\}^+$ and $\{a,b\} = \{0,1\}$. Again we can suppose τ is primitive and $|\tau(0)| > 1$ and $|\tau(1)| > 1$. If $\tau(0)$ and $\tau(1)$ begin in distinct letters, then $\tau(\tilde{x})0, \tau(\tilde{x})1 \in \Omega(x)$ where $\tilde{x}$ denotes the reverse of x. Thus $\tau(\tilde{x}) = \tilde{x}$. Thus for each prefix v of abx we have $\tau(\tilde{x}v) = \tilde{x}\tau(v)$ whence $\tau(v)$ is also a prefix of abx. Hence $\tau(abx) = abx$. As before this implies that abx is a proper tail of itself which is a contradiction. Thus $\tau(0)$ and $\tau(1)$ begin in the same letter. Whence by Corollary 1 it follows that every tail of y is self-shuffling, which is again a contradiction since $0x$ and $1x$ are not self-shuffling.

Remark 3. In the case of the Fibonacci infinite word x, each of $\{x, 0x, 1x, 01x, 10x\}$ is pure morphic. For a general characteristic word x, since every point in the orbit of x except for $0x$ and $1x$ is self-shuffling, it follows that if τ is a morphism fixing x (respectively $01x$ or $10x$), then $\tau(0)$ and $\tau(1)$ must end (respectively begin) in distinct letters.

References

1. Allouche, J.-P., Shallit, J.: The ubiquitous Prouhet-Thue-Morse sequence. In: Ding, C., Helleseth, T., Niederreiter, H. (eds.) Proceedings of Sequences and Their Applications, SETA 1998, pp. 1–16. Springer (1999)
2. Allouche, J.-P., Shallit, J.: Automatic sequences. In: Theory, Applications, Generalizations. Cambridge University Press (2003)
3. Berthé, V., Ei, H., Ito, S., Rao, H.: On substitution invariant Sturmian words: an application of Rauzy fractals. Theor. Inform. Appl. 41, 329–349 (2007)
4. Cassaigne, J., Karhumäki, J.: Toeplitz Words, Generalized Periodicity and Periodically Iterated Morphisms. European J. Combin. 18, 497–510 (1997)
5. Henshall, D., Rampersad, N., Shallit, J.: Shuffling and unshuffling. Bull. EATCS 107, 131–142 (2012)
6. Fagnot, I.: A little more about morphic Sturmian words. Theor. Inform. Appl. 40, 511–518 (2006)
7. Lothaire, M.: Algebraic Combinatorics on Words. Encyclopedia of Mathematics and its Applications, vol. 90. Cambridge University Press, U.K (2002)
8. Morse, M., Hedlund, G.A.: Symbolic dynamics II: Sturmian sequences. Amer. J. Math. 62, 1–42 (1940)
9. Thue, A.: Über unendliche Zeichenreihen. Norske Vid. Selsk. Skr. I Math-Nat. Kl. 7, 1–22 (1906)
10. Yasutomi, S.-I.: On sturmian sequences which are invariant under some substitutions. In: Kanemitsu, S., et al. (eds.) Number Theory and Its Applications, Proceedings of the Conference held at the RIMS, Kyoto, Japan, November 10-14, 1997, pp. 347–373. Kluwer Acad. Publ., Dordrecht (1999)