Discrete Mathematics and Theoretical Computer Science DMTCS vol. **9:2**, 2007, 213–226

Tag-systems for the Hilbert curve †

Patrice Séébold‡

LIRMM, Univ. Montpellier 2, CNRS, 161 rue Ada, 34392 Montpellier, France

received 20 Dec 2004, revised 23 Apr 2005, accepted 27 Jun 2005.

Hilbert words correspond to finite approximations of the Hilbert space filling curve. The *Hilbert infinite word* H is obtained as the limit of these words. It gives a description of the Hilbert (infinite) curve. We give a uniform tag-system to generate automatically H and, by showing that it is almost cube-free, we prove that it cannot be obtained by simply iterating a morphism.

Keywords: Peano curves, Hilbert curve, Hilbert words, tag-systems, D0L-systems, cubes, repetition-free words

1 Introduction

Words without repetitions were studied since the very first work specifically dedicated to combinatorics on words by Thue (cf Thue, 1906 [16]). Also, the generation of infinite words by morphisms was the subject of many works (see, e.g., Lothaire, 2002 [9], Allouche and Shallit, 2003 [2]).

Generating a word with a HD0L-system consists of applying a morphism to an infinite word generated by another morphism. Berstel [3] gave an example showing that the power of generation of HD0L-systems is greater than the power of generation of D0L-systems (only one morphism is applied): he proved that the Arshon word (a square-free word over a 3-letter alphabet, see Arshon, 1937 [1], Séébold, 2002 [15]) is generated by a tag-system (a particular case of an HD0L-system, see Cobham, 1972 [6]) whereas it cannot be obtained with a D0L-system. Here we give a new example of this phenomenon. Studying the Hilbert word (an infinite word over a 4-letter alphabet describing the drawing of the square-filling Hilbert curve, see, e.g., Sagan, 1994 [14]) we prove that it is 4-power-free but not cube-free (as it is the case with the well known Fibonacci word, see, e.g., Berstel, 1986 [4], Allouche and Shallit, 2003 [2]) and it is generated by a uniform tag-system but not by a D0L-system (as it is the case with the Arshon word).

After some preliminaries (Section 2), we introduce in Section 3 the notion of Hilbert words and we give some simple first results. Then we show in Section 4 that they are almost cube-free, and that they are obtained by using a tag-system but not by iterating a single morphism. We conclude with some open questions.

†Part of this work has appeared in Kitaev *et al.*, 2004 [7]

‡Département Mathématiques et Informatique Appliquées, Université Paul Valéry, Route de Mende, 34199 Montpellier Cédex 5, France

2 Preliminaries

The terminology and notation are mainly those of Lothaire, 2002 [9].

Let A be a finite set called an *alphabet* and A^* the free monoid generated by A.

The elements of A are called *letters* and those of A^* are called *words*. The *empty word* ε is the identity element of A^* for the concatenation of words (the *concatenation* of two words u and v is the word uv), and we denote by A^+ the semigroup $A^* \setminus \{\varepsilon\}$.

The *length* of a word u, denoted by $|u|$, is the number of occurrences of letters in u. In particular $|\varepsilon| = 0$. The number of occurrences of a letter a in a word u is denoted by $|u|_a$.

If n is a nonnegative integer, u^n is the word obtained by concatenating n occurrences of the word u. Of course, $|u^n| = n|u|$. The cases $n = 2$, $n = 3$, and $n = 4$ deserve particular attention in what follows. A word u^2 (resp. u^3, u^4), with $u \neq \varepsilon$, is called a *square* (resp. a *cube*, a 4-*power*).

A word w is called a *factor* (resp. a *prefix*, resp. a *suffix*) of u if there exist words x, y such that $u = xwy$ (resp. $u = wy$, resp. $u = xw$). The factor (resp. the prefix, resp. the suffix) is *proper* if $xy \neq \varepsilon$ (resp. $y \neq \varepsilon$, resp. $x \neq \varepsilon$). A word u is a *subsequence* of the word v if there exist words $u_1, \dots, u_n, v_1, \dots, v_n, v_{n+1}$ such that $u = u_1 \cdots u_n$ and $v = v_1 u_1 v_2 u_2 \cdots v_n u_n v_{n+1}$.

An *infinite word* (or *sequence*) over A is an application $\mathbf{a} : \mathbb{N} \to A$. It is written $\mathbf{a} = a_0 a_1 \cdots a_i \cdots, i \in \mathbb{N}, a_i \in A$.

The notion of factor is extended to infinite words as follows: a (finite, possibly empty) word u is a *factor* (resp. *prefix*) of an infinite word $\mathbf{a}$ over A if there exist $n \in \mathbb{N}$ (resp. $n = 0$) and $m \in \mathbb{N}$ ($m = |u|$) such that $u = a_n \cdots a_{n+m-1}$ (by convention $a_n \cdots a_{n-1} = \varepsilon$).

In what follows, we will consider morphisms on A. Let B be an alphabet (often, $B = A$).

A *morphism on* A (in short a *morphism*) is an application $f : A^* \to B^*$ such that $f(uv) = f(u)f(v)$ for all $u, v \in A^*$. It is uniquely determined by its value on the alphabet A. A morphism f on A is *uniform* if there exists an integer k such that $|f(a)| = k$ for all $a \in A$; when $k = 1$, f is said to be a *literal morphism*.

Now suppose $A = B$. A morphism is *nonerasing* if $f(a) \neq \varepsilon$ for all $a \in A$. A nonerasing morphism is *prolongable on* x_0, $x_0 \in A^+$, if there exists $u \in A^+$ such that $f(x_0) = x_0 u$. In this case, for all $n \in \mathbb{N}$ the word $f^n(x_0)$ is a proper prefix of the word $f^{n+1}(x_0)$ and this defines a unique infinite word

$$\mathbf{x} = x_0 u f(u) f^2(u) \cdots f^n(u) \cdots$$

which is the limit of the sequence $(f^n(x_0))_{n \geq 0}$. We write $\mathbf{x} = f^\omega(x_0)$ and say that $\mathbf{x}$ is *generated* by f.

A (finite or infinite) word u over A is *square-free* (resp. *cube-free*, 4-*power-free*) if none of its factors is a square (resp. a cube, a 4-power). A morphism f on A is *square-free* if the word $f(u)$ is square-free whenever u is a square-free word. The morphism f is *weakly square-free* if f generates a square-free infinite word.

A *D0L-system* is a triple $G = (A, f, u)$ where A is an alphabet, f a morphism on A and $u \in A^*$. An infinite word $\mathbf{x}$ is generated by G if $\mathbf{x} = (f^k)^\omega(u)$ for some $k \in \mathbb{N}$.

A *HD0L-system* is a quintuple $T = (A, u, f, g, B)$ where A and B are alphabets, $u \in A^+$, f is a nonerasing morphism on A, prolongable on u, and g is a morphism from A onto B. An infinite word $\mathbf{y}$ is generated by T if $\mathbf{y} = g(f^\omega(u))$. When g is a literal morphism T is called a *tag-system*.

The tag-system is *uniform* if f is a uniform morphism. The terminology of tag-system comes from the fundamental study of Cobham [6]. Chapter 5 of [13] is dedicated to a deep study of D0L-systems (see also Pansiot, 1983 [11] who used the terminology *extended tag-systems* for HD0L-systems).

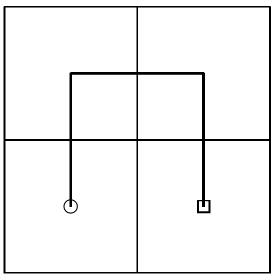
3 The Hilbert curve and the Hilbert words

Peano [12] was the first in 1890 to realize the construction of a fractal curve that fills a square without any holes. This construction is obtained by drawing, without removing the pen from the surface of the paper, an infinite succession of unit lines left, right, up, or down. Thus this succession can be represented by an infinite word over the alphabet $\Sigma = \{u, \bar{u}, r, \bar{r}\}$ where u stands for *up*, $\bar{u}$ stands for *down*, r stands for *right*, and $\bar{r}$ stands for *left* (for a description of pictures by words, see the basic study of Maurer, Rozenberg and Welzl, 1982 [10]). In 1891 Hilbert [8] defined another space filling curve. The word so obtained, called the *Hilbert infinite word*, is denoted by H (the Hilbert curve is some particular example of a space filling curve, and space filling curves are often described under the generic name of *Peano curves*. For this reason, in a previous paper [7] we improperly gave the name of Peano curve to the Hilbert curve).

Let us describe the algorithm of Hilbert. The general idea is to divide, at step n, the unit square in 4^n equal subsquares each of them containing an equal length part of the curve (except the first and the last ones which contain a part of length 1/2). The curve so obtained is then depicted by a word of length $4^n - 1$ which we will call the *n-th Hilbert word* H_n. When n tends to infinity the curve fills the unit square without any holes and the sequence of words H_n tends to the Hilbert infinite word H (see Section 4).

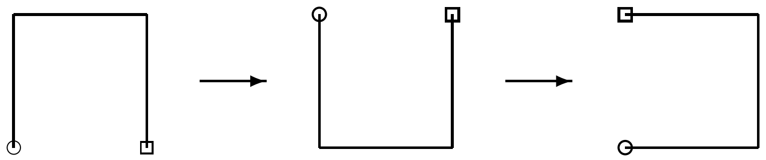
Step by step, the algorithm is the following (let us recall that the drawing of the curves is realized without removing the pen from the surface of the paper; in the following figures $\circ$ and $\square$ respectively represent the starting and the ending points of the drawing).

- At step 1 the unit square is divided in 4 equal subsquares and it contains the staple-like curve depicted by the word $H_1 = ur\bar{u}$.



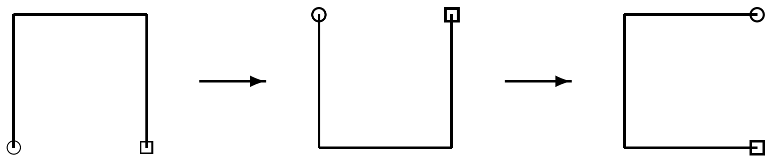
- From step n to step $n + 1$ the curve and grid sizes are decreased by a factor two and four copies are put together to form a new square.

– The first copy is the left lower one, and it is obtained as follows: first perform a vertical flip, then rotate a quarter turn left.

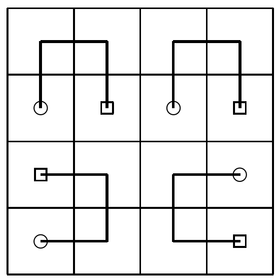


– The second and the third copies are the upper two ones: they are placed as they are.

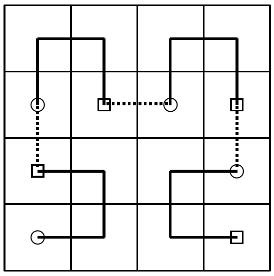
– The fourth copy is the right lower one: first perform a vertical flip, then rotate a quarter turn right.



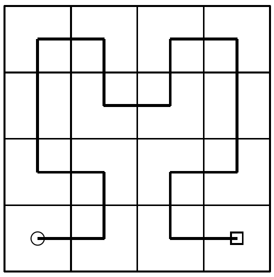
This gives the following.



Then, the curve is made continuous by connecting the ending point of the first (resp. the second, the third) copy to the starting point of the second (resp. the third, the fourth) one with three unit segments respectively corresponding to a move up (u), a move right (r), and a move down ($\bar{u}$) – the dashes lines in the following diagram.



To end, all the starting and ending points are removed, except the starting point of the first copy and the ending point of the fourth copy.



For more about the construction of space filling curves, see, for example, Sagan, 1994 [14].

Now, let us define on Σ three literal morphisms f, r_g, and r_d by

$$f(u) = \bar{u}, f(\bar{u}) = u, f(r) = r, f(\bar{r}) = \bar{r},$$

$$r_g(u) = \bar{r}, r_g(\bar{u}) = r, r_g(r) = u, r_g(\bar{r}) = \bar{u},$$

$$r_d(u) = r, r_d(\bar{u}) = \bar{r}, r_d(r) = \bar{u}, r_d(\bar{r}) = u.$$

These three morphisms respectively represent a vertical flip, a quarter turn left rotation, and a quarter turn right rotation.

From the construction we have that the Hilbert word H_{n+1} (which represents the drawing without removing the pen from the surface of the paper of the Hilbert curve at step $n+1$) is obtained from H_n by

$$H_{n+1} = \rho(H_n)\ u\ H_n\ r\ H_n\ \bar{u}\ \lambda(H_n), \tag{1}$$

where $\rho = r_g \circ f$ and $\lambda = r_d \circ f$ ($\rho : u \mapsto r \mapsto u, \bar{u} \mapsto \bar{r} \mapsto \bar{u}$; $\lambda : u \mapsto \bar{r} \mapsto u, \bar{u} \mapsto r \mapsto \bar{u}$).

$$\begin{aligned}\text{One has } H_2 &= r u \bar{r}\ u\ u r \bar{u}\ r\ u r \bar{u}\ \bar{u}\ \bar{r} \bar{u} r \\ \text{and } H_3 &= u r \bar{u} r r u \bar{r} u r u \bar{r} \bar{r} \bar{u} \bar{r} u \\ & \quad u \\ & \quad r u \bar{r} u u r \bar{u} r u r \bar{u} \bar{u} \bar{r} \bar{u} r \\ & \quad r \\ & \quad r u \bar{r} u u r \bar{u} r u r \bar{u} \bar{u} \bar{r} \bar{u} r \\ & \quad \bar{u} \\ & \quad \bar{u} \bar{r} u \bar{r} \bar{r} \bar{u} r \bar{u} \bar{r} \bar{u} r r u r \bar{u}\end{aligned}$$

Now let w be a word over Σ. The word $\bar{w}$ is obtained from w by replacing each occurrence of u, r, $\bar{u}$, $\bar{r}$ respectively by $\bar{u}$, $\bar{r}$, u, r ($\bar{\varepsilon} = \varepsilon$). It is clear that ρ and λ are the literal morphisms defined on Σ by

$$\rho(u) = r \text{ and, for any } x \in \Sigma,\ \rho^2(x) = x \text{ and } \rho(\bar{x}) = \overline{\rho(x)} \tag{2}$$

$$\lambda(u) = \bar{r} \text{ and, for any } x \in \Sigma,\ \lambda^2(x) = x \text{ and } \lambda(\bar{x}) = \overline{\lambda(x)} \tag{3}$$

Let g be the literal morphism defined on Σ by $g(u) = g(\bar{u}) = u$, $g(r) = g(\bar{r}) = r$, and let us recall that if w is a word over Σ with $w = w_1 \cdots w_n$, $w_i \in \Sigma$, then $\tilde{w} = w_n \cdots w_1$ ($\tilde{\varepsilon} = \varepsilon$).

Together with the literal morphisms ρ, λ, f, and g, the Hilbert words H_n have the following straightforward properties.

Property 1

1. $\rho \circ \lambda = \lambda \circ \rho$.
2. *For any* $w \in \Sigma^*$, $\lambda(w) = \overline{\rho(w)}$.
3. $g \circ \rho = g \circ \lambda$.
4. *For any* $n \geq 1$,
   - $f(H_n) = \widetilde{H_n}$,
   - $g(H_n) = g(\widetilde{H_n}) = \widetilde{g(H_n)}$,
   - $\rho(H_n) = f(\widetilde{\lambda(H_n)})$.
5. *For any* $n \geq 1$, *the Hilbert word* H_n *is irreducible, that is, it does not contain any factor* $u\bar{u}$, $\bar{u}u$, $r\bar{r}$, *or* $\bar{r}r$.

From the construction it is easy to see that, for any positive integer n, $|H_n| = 4^n - 1$.

Moreover, we have the following more precise counting of each letter in H_n.

Lemma 2 *For any* $n \in \mathbb{N} \setminus \{0\}$, *one has*
$|H_n|_u = |H_n|_{\bar{u}} = 4^{n-1}$,
$|H_n|_r = 4^{n-1} + 2^{n-1} - 1$,
$|H_n|_{\bar{r}} = 4^{n-1} - 2^{n-1}$.

Proof: The result is obvious for $n = 1$.

From (1) we get, for $x \in \Sigma$,

$$|H_{n+1}|_x = 2|H_n|_x + |\rho(H_n)|_x + |\lambda(H_n)|_x + \begin{cases} 1, & \text{if } x = u, r, \bar{u} \\ 0, & \text{if } x = \bar{r}. \end{cases}$$

Then, by the definition of ρ and λ, we obtain

$$\begin{array}{lcl} |H_{n+1}|_u &=& 2|H_n|_u + |H_n|_r + |H_n|_{\bar{r}} + 1 \\ |H_{n+1}|_{\bar{u}} &=& 2|H_n|_{\bar{u}} + |H_n|_{\bar{r}} + |H_n|_r + 1 \\ |H_{n+1}|_r &=& 2|H_n|_r + |H_n|_u + |H_n|_{\bar{u}} + 1 \\ |H_{n+1}|_{\bar{r}} &=& 2|H_n|_{\bar{r}} + |H_n|_{\bar{u}} + |H_n|_u \end{array}$$

and the result follows by induction. □

As an immediate corollary one has the following.

Corollary 3 *For any $n \geq 1$,*

- $|g(H_n)|_u = |g(H_n)|_r + 1$;
- $|g(\rho(H_n))|_u = |g(\lambda(H_n))|_u = |g(\rho(H_n))|_r - 1 = |g(\lambda(H_n))|_r - 1.$

4 Generating the Hilbert infinite word

Preliminary remark. The sequence $(H_n)_{n\geq 1}$ has two limits according to whether n is even or odd. An equivalent construction (equivalent in the sense that it provides a curve drawn without removing the pen from the surface of the paper and filling the unit square without any holes) can be obtained with no distinction between the even case and the odd one: it is enough at each even step, before computing the corresponding Hilbert word, to apply to the whole picture a vertical flip followed by a quarter turn left rotation and then each H_n is a prefix of H_{n+1}. But the limit for the odd indices is the same in the two cases so, because the properties of the Hilbert words H_n are more interesting with our first construction, we keep the definition of the Hilbert words H_n given in Section 3 and define the Hilbert infinite word H as the limit of odd rank Hilbert words, that is, $H = \lim_{n\to\infty} H_{2n+1}$.

In this section, we prove that the Hilbert infinite word H is generated by a HD0L-system, it contains no cube except those of only one letter, and it cannot be generated by a single morphism (it is not even generated by a D0L-system). Then, applying a deep result of Pansiot [11] (see also Allouche and Shallit, 2003 [2, Chapter 7]), we obtain an equivalent uniform tag-system generating H.

Let Ω be the eight-letter alphabet $\Omega = \{A, B, C, D, a, b, c, d\}$, and let γ and h be the following morphisms.

$$\begin{array}{rcl} \gamma : \Omega^* & \to & \Omega^* \\ A & \mapsto & BaAbAcD \\ B & \mapsto & AbBaBdC \\ C & \mapsto & DcCdCaB \\ D & \mapsto & CdDcDbA \\ a & \mapsto & a \\ b & \mapsto & b \\ c & \mapsto & c \\ d & \mapsto & d \end{array} \qquad \begin{array}{rcl} h : \Omega^* & \to & \Sigma^* \\ A & \mapsto & ur\bar{u} \\ B & \mapsto & ru\bar{r} \\ C & \mapsto & \bar{u}\bar{r}u \\ D & \mapsto & \bar{r}\bar{u}r \\ a & \mapsto & u \\ b & \mapsto & r \\ c & \mapsto & \bar{u} \\ d & \mapsto & \bar{r} \end{array}$$

Theorem 4 *H is the infinite word generated by the HD0L-system $(\Omega, A, \gamma^2, h, \Sigma)$, that is, $H = h((\gamma^2)^\omega(A))$.*

The proof of this result will use the following lemma which expresses the geometrical meaning of the morphisms γ and h.

Lemma 5 *For every $n \in \mathbb{N}$,*

- $h(\gamma^n(A)) = \rho(h(\gamma^n(B))) = \lambda(h(\gamma^n(D)))$,
- $h(\gamma^n(B)) = \rho(h(\gamma^n(A))) = \lambda(h(\gamma^n(C)))$,
- $h(\gamma^n(C)) = \rho(h(\gamma^n(D))) = \lambda(h(\gamma^n(B)))$,
- $h(\gamma^n(D)) = \rho(h(\gamma^n(C))) = \lambda(h(\gamma^n(A)))$.

Proof: The eight equalities are obviously true if $n = 0$.

Now, let us prove for any integer $n \geq 0$ that if the eight equalities are true for n then they are also true for $n + 1$. One has

$$\begin{aligned} h(\gamma^{n+1}(A)) &= h(\gamma^n(BaAbAcD)) \\ &= h(\gamma^n(B))h(a)h(\gamma^n(A))h(b)h(\gamma^n(A))h(c)h(\gamma^n(D)) \\ &= \rho(h(\gamma^n(A)))uh(\gamma^n(A))rh(\gamma^n(A))\bar{u}\lambda(h(\gamma^n(A))) \text{ (by induction)} \\ &= \rho(h(\gamma^n(A)))\rho(r)h(\gamma^n(A))\rho(u)h(\gamma^n(A))\rho(\bar{r})\lambda(h(\gamma^n(A))) \\ &= \rho(h(\gamma^n(A)))\rho(h(b))\rho(h(\gamma^n(B)))\rho(h(a))\rho(h(\gamma^n(B)))\rho(h(d))\rho(h(\gamma^n(C))) \\ &= \rho(h(\gamma^n(AbBaBdC))) \\ &= \rho(h(\gamma^{n+1}(B))). \end{aligned}$$

This proves the first equality. The seven others are verified in the same way. □

Proof of Theorem 4. We will prove by induction that, for any $n \in \mathbb{N}$, $H_{n+1} = h(\gamma^n(A))$. The result follows because $H = \lim_{n\to\infty} H_{2n+1} = \lim_{n\to\infty} h(\gamma^{2n}(A))$.

The equality is of course true if $n = 0$ since $H_1 = ur\bar{u} = h(A)$. Now,

$$\begin{aligned} H_{n+2} &= \rho(H_{n+1})uH_{n+1}rH_{n+1}\bar{u}\lambda(H_{n+1}) \\ &= \rho(h(\gamma^n(A)))h(a)h(\gamma^n(A))h(b)h(\gamma^n(A))h(c)\lambda(h(\gamma^n(A))) \text{ (by induction)} \\ &= h(\gamma^n(B))h(\gamma^n(a))h(\gamma^n(A))h(\gamma^n(b))h(\gamma^n(A))h(\gamma^n(c))h(\gamma^n(D)) \text{ (Lemma 5)} \\ &= h(\gamma^n(BaAbAcD)) \\ &= h(\gamma^{n+1}(A)). \end{aligned}$$

□

Now, to prove that the Hilbert infinite word H contains no cube except x^3, $x \in \Sigma$, we need an intermediate lemma. First remark that the morphism γ is clearly not a square-free morphism (for example, $\gamma(CA)$ contains BB as a factor). It is even not weakly square-free (it does not generate a square-free word because, for example, $\gamma^4(A)$ contains $bAbAb$ as a factor). But we have the following.

Lemma 6 *For any $n \in \mathbb{N}$, $\gamma^n(A)$ does not contain any factor $YwYwY$ with $Y \in \{A, B, C, D\}$ and $w \in \Omega^*$.*

Proof: The property is straightforward if $n = 0$ or $n = 1$.

In order to get a contradiction, let us suppose that, for some integer $n \geq 2$, $\gamma^n(A)$ contains a factor $YwYwY$, $Y \in \{A, B, C, D\}$, $w \in \Omega^*$, when $\gamma^{n-1}(A)$ does not contain any such factor. Moreover, let us suppose that $Y = A$ (the three other cases are symmetrical by definition of γ).

Let $u, v \in \Omega^*$ be such that $\gamma^n(A) = uAwAwAv$.

By definition of γ, four cases are possible for u: $u = \gamma(w_1)$, $u = \gamma(w_1)Ba$, $u = \gamma(w_1)BaAb$, or $u = \gamma(w_1)CdDcDb$, for some prefix w_1 of $\gamma^{n-1}(A)$.

Before continuing, we remark that the symbols of $\gamma^n(A)$ alternate between lower- and upper-case letters.

1. $u = \gamma(w_1)$

   In this case, the first occurrence of A following u is necessarily the first letter of $\gamma(B)$. This implies that Aw starts with $\gamma(B) = AbBaBdC$ and, since this last factor can only appear, in $\gamma^n(A)$, as an occurrence of $\gamma(B)$, it follows that the second occurrence of Aw also starts with $\gamma(B)$, that is, there exists $W' \in \Omega^*$ such that $AwAw = \gamma(BW'BW')$ and $\gamma^{n-1}(A)$ starts with $w_1BW'BW'$. But since $AwAw$ is followed, in $\gamma^n(A)$, by the letter A, the next letter in $\gamma^{n-1}(A)$ is a B, which implies that $\gamma^{n-1}(A)$ contains $BW'BW'B$ as a factor, a contradiction.

2. $u = \gamma(w_1)Ba$

   In this case, the first occurrence of A after u is followed by $bAcD$. This implies that Aw starts with $AbAcD$ and, since this last factor can only appear, in $\gamma^n(A)$, in an occurrence of $\gamma(A)$, it follows that $Aw = AbAcD\gamma(W')Ba$ for some $W' \in \Omega^*$ where $AbAcD$ and Ba are respectively the suffix and the prefix of $\gamma(A)$. This means $\gamma^{n-1}(A)$ starts with $w_1AW'AW'A$, a contradiction.

3. $u = \gamma(w_1)BaAb$

   In this case, the first occurrence of A after u is followed by cD. This implies that Aw starts with AcD. Here two cases are possible.

   Either w ends with $BaAb$ and, as in the previous case, $\gamma^{n-1}(A)$ contains a factor $AW'AW'A$, a contradiction.

   Or this factor AcD is the central part of some $\gamma(DcC)$. But in this case AcD is followed, in Aw, by $cCdCaB$ and, since $\gamma^n(A)$ starts with uAw, this factor $cCdCaB$ should be the beginning of some $\gamma(W')$ in $\gamma^n(A)$. This is impossible.

4. $u = \gamma(w_1)CdDcDb$

   In this case, the first occurrence of A after u is the last letter of $\gamma(D)$ and w starts in the same manner as some $\gamma(W')$.

   Let us consider the letter A at the beginning of the second occurrence of Aw.

   - It is impossible that this A is the first letter of $\gamma(B)$ because $\gamma(W')$ (and thus w) cannot start with $bBaBdC$.
   - It is impossible that this A is the first A in $\gamma(A)$ because $\gamma(W')$ (and thus w) cannot start with $bAcD$.
   - It is impossible that this A is the second A in $\gamma(A)$. Indeed otherwise w starts with cD and since w starts in the same manner as some $\gamma(W')$, w starts with $cDcCdCaB$. But this would imply that $CdCaB$ is the beginning of some $\gamma(Z)$ which is impossible.
   - Thus this A is again the last letter of $\gamma(D)$. This is also the case for the last A of $AwAwA$, which implies that $\gamma^{n-1}(A)$ starts with $w_1DW'DW'D$, a contradiction. $\square$

Now, we are ready to prove a second noteworthy result.

Theorem 7 *The infinite word H does not contain any factor $xyWxyWxy$ with x, y letters and W a word. In particular, the only cubes in H are x^3 with x a single letter. Moreover, H is* 4*-power-free.*

Proof: We first suppose that H does not contain any factor $xyWxyWxy$ with x, y letters and W a word. If H contains a cube X then necessarily $X = x^3$ where x is a letter. That H effectively contains all the cubes x^3 for $x \in \Sigma$, and also that it contains no factor x^4, comes from the following. A factor u^3 (resp. r^3, $\bar{u}^3$, $\bar{r}^3$) can only be found in $h(CaA)$ (resp. $h(DbB)$, $h(AcC)$, $h(BdD)$); a factor CaA (resp. DbB, AcC, BdD) can only be found as the central part of $\gamma(BaB)$ (resp. $\gamma(AbA)$, $\gamma(DcD)$, $\gamma(CdC)$); finally BaB (resp. AbA, DcD, CdC) is the central factor of $\gamma(B)$ (resp. $\gamma(A)$, $\gamma(D)$, $\gamma(C)$).

Since $\gamma^2(A)$ contains A, B, C, and D, $\gamma^3(A)$ contains BaB, AbA, DcD, and CdC, thus $\gamma^4(A)$ contains CaA, DbB, AcC, and BdD, that is, H_5 contains u^3, r^3, $\bar{u}^3$, and $\bar{r}^3$.

It is interesting to remark here that, starting from the index 0, we find in the infinite word H the first occurrence of r^3 (resp. u^3, $\bar{r}^3$, $\bar{u}^3$) at the index 30 (resp. 94, 222, 478).

Now we prove the first part of the theorem.

Suppose that H contains a factor $xyWxyWxy$ with x, y letters and W a word, and let T be such that $TxyWxyWxy$ is a prefix of H, i.e., $TxyWxyWxy$ is a prefix of $h(\gamma^{2n}(A))$ for some $n \in \mathbb{N}$.

There are four possible cases depending on the value of $|T| \bmod 4$.

- If $|T| \bmod 4 = 0$ then x is the first letter of $h(X)$ for some $X \in \{A, B, C, D\}$. We suppose $X = A$ (the other cases are symmetrical).

  Then $x = u$, $y = r$ and W starts with $\bar{u}$: $W = \bar{u}W'$. By definition of h and γ, the factor $ur\bar{u}$ can only appear, in H, as $h(A)$. This implies that there exists a word $w' \in \Omega^*$ such that $TxyWxyWxy = Th(Aw'Aw')ur$. But, by construction, w' ends with a lower-case letter thus, in $\gamma^{2n}(A)$, $Aw'Aw'$ is followed by an upper-case letter. Since the image of this letter by h starts with ur, this letter is necessarily A. Thus $\gamma^{2n}(A)$ contains $Aw'Aw'A$, a contradiction with Lemma 6.

- If $|T| \bmod 4 = 1$ then x is the second letter of $h(X)$, $X \in \{A, B, C, D\}$, y is the third letter of $h(X)$, and W ends with the first letter of $h(X)$. As in the previous case, we obtain a contradiction with Lemma 6.

- If $|T| \bmod 4 = 2$ then x is the third letter of $h(X)$, $X \in \{A, B, C, D\}$. We suppose $X = A$ which implies $x = \bar{u}$. Since, in $\gamma^{2n}(A)$, A is necessarily followed by b or c, we have $y = r$ or $y = \bar{u}$.

  First suppose $|Wx| \bmod 4 \neq 3$. Then if $y = r$ the only possibility is $|W| \bmod 4 = 1$ (that is, after T the second occurrence of $xy = \bar{u}r$ is at the end of $h(D)$) and if $y = \bar{u}$ the only possibility is $|W| \bmod 4 = 3$ (that is, after T the second occurrence of $xy = \bar{u}\bar{u}$ is such that x is the image by h of a lower-case letter and y is the beginning of the image of an upper-case letter.) In the two cases $|TxyWxyW| \bmod 4 = 0$ which implies that the third occurrence of xy is the beginning of some $h(Y)$ with $Y \in \{A, B, C, D\}$: this is impossible because $xy = \bar{u}r$ or $xy = \bar{u}\bar{u}$.

  Thus $|Wx| \bmod 4 = 3$ which implies that Wx ends with the image by h of an upper-case letter. Since $x = \bar{u}$ this letter is A and, as previously, we obtain that $\gamma^{2n}(A)$ contains $Aw'Aw'A$ for some word $w' \in \Omega^*$, a contradiction with Lemma 6.

- If $|T| \bmod 4 = 3$ then x is the image by h of a lower-case letter. Then y is the first letter of $h(X)$ for some $X \in \{A, B, C, D\}$.

  This implies that $TxyWxyWxy$ is a prefix of $Txh(X)W'xh(X)W'xh(X)$ where W' is the word such that $yW = h(X)W'$. Again, this means that $\gamma^{2n}(A)$ contains $Xw'Xw'X$, a contradiction with Lemma 6. □

A direct corollary is the following.

Corollary 8 *The infinite word H cannot be generated by a D0L-system.*

Proof: If H were generated by a D0L-system (f, Σ, v) then $H = f^n(H)$ for some $n \in \mathbb{N} \setminus \{0\}$. Consequently $f(u^3)$, $f(r^3)$, $f(\bar{u}^3)$, and $f(\bar{r}^3)$ are factors of H. Since H does not contain any cube except u^3, r^3, $\bar{u}^3$, and $\bar{r}^3$, this implies that $|f(x)| \leq 1$ for any $x \in \Sigma$: a contradiction because, to generate an infinite word, f must be prolongable on at least one letter. □

To end this part let us remark that, following a deep result of Cobham [5] (see also Pansiot, 1983 [11]), we know that there exists a tag-system generating the infinite word H. However a direct application of Pansiot's algorithm gives a non-uniform tag-system with a 16-letter alphabet and a morphism whose image length is 21 for almost all the letters. Here, we obtain below a better result by proving that H is generated by a uniform tag-system; indeed it is known that there exist words generated by tag-system but not by a uniform one, see, e.g., Pansiot, 1983 [11].

Moreover the tag-system below has a 8-letter alphabet and a morphism whose image length is 16.

Let $\Omega' = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}$ be a new alphabet, and let γ' and h' be the following morphisms.

$$\begin{array}{rcll} \gamma' : & \Omega'^* & \to & \Omega'^* \\ & \omega_1 & \mapsto & \omega_3\omega_1\omega_2\omega_7 \\ & \omega_2 & \mapsto & \omega_3\omega_1\omega_2\omega_8 \\ & \omega_3 & \mapsto & \omega_1\omega_3\omega_4\omega_5 \\ & \omega_4 & \mapsto & \omega_1\omega_3\omega_4\omega_6 \\ & \omega_5 & \mapsto & \omega_8\omega_6\omega_5\omega_3 \\ & \omega_6 & \mapsto & \omega_8\omega_6\omega_5\omega_4 \\ & \omega_7 & \mapsto & \omega_6\omega_8\omega_7\omega_1 \\ & \omega_8 & \mapsto & \omega_6\omega_8\omega_7\omega_2 \end{array} \qquad \begin{array}{rcl} h' : \Omega'^* & \to & \Sigma^* \\ \omega_1 & \mapsto & u \\ \omega_2 & \mapsto & \bar{r} \\ \omega_3 & \mapsto & r \\ \omega_4 & \mapsto & \bar{u} \\ \omega_5 & \mapsto & r \\ \omega_6 & \mapsto & \bar{u} \\ \omega_7 & \mapsto & u \\ \omega_8 & \mapsto & \bar{r} \end{array}$$

Lemma 9 *For every integer $n \geq 1$,*

$$\begin{array}{rclrcl} h'(\gamma'^{2n}(\omega_1)) & = & h(\gamma^{2n-1}(Ba)), & h'(\gamma'^{2n}(\omega_2)) & = & h(\gamma^{2n-1}(Bd)), \\ h'(\gamma'^{2n}(\omega_3)) & = & h(\gamma^{2n-1}(Ab)), & h'(\gamma'^{2n}(\omega_4)) & = & h(\gamma^{2n-1}(Ac)), \\ h'(\gamma'^{2n}(\omega_5)) & = & h(\gamma^{2n-1}(Db)), & h'(\gamma'^{2n}(\omega_6)) & = & h(\gamma^{2n-1}(Dc)), \\ h'(\gamma'^{2n}(\omega_7)) & = & h(\gamma^{2n-1}(Ca)), & h'(\gamma'^{2n}(\omega_8)) & = & h(\gamma^{2n-1}(Cd)). \end{array}$$

Proof: It is a little tedious but not difficult task to verify the result for $n = 1$.

Now, let us prove for any integer $n \geq 1$ that if the eight equalities are true for $2n$ then they are also true for $2n + 2$.

One has

$$\begin{array}{rl} & h'(\gamma'^{2n+2}(\omega_3)) \\ = & h'(\gamma'^{2n}(\gamma'^2(\omega_3))) \\ = & h'(\gamma'^{2n}(\gamma'(\omega_1\omega_3\omega_4\omega_5))) \\ = & h'(\gamma'^{2n}(\omega_3\omega_1\omega_2\omega_7\omega_1\omega_3\omega_4\omega_5\omega_1\omega_3\omega_4\omega_6\omega_8\omega_6\omega_5\omega_3)) \\ = & h(\gamma^{2n-1}(Ab)\gamma^{2n-1}(Ba)\gamma^{2n-1}(Bd)\gamma^{2n-1}(Ca)\gamma^{2n-1}(Ba)\gamma^{2n-1}(Ab)\gamma^{2n-1}(Ac) \\ & \gamma^{2n-1}(Db)\gamma^{2n-1}(Ba)\gamma^{2n-1}(Ab)\gamma^{2n-1}(Ac)\gamma^{2n-1}(Dc)\gamma^{2n-1}(Cd)\gamma^{2n-1}(Dc) \\ & \gamma^{2n-1}(Db)\gamma^{2n-1}(Ab)) \text{ (by induction)} \\ = & h(\gamma^{2n-1}(AbBaBdCaBaAbAcDbBaAbAcDcCdDcDbAb)) \\ = & h(\gamma^{2n-1}(\gamma^2(Ab))) \\ = & h(\gamma^{2n+1}(Ab)). \end{array}$$

This proves the third equality. The seven others are verified in the same way. □

Theorem 10 *H is the infinite word generated by the tag-system $(\Omega', \omega_1, \gamma'^2, h', \Sigma)$, that is, $H = h'((\gamma'^2)^\omega(\omega_1))$.*

Proof: Since $\gamma(A)$ begins with Ba and $\gamma(B)$ begins with A we have that, for every integer $n \geq 1$, $\gamma^{2n-2}(A)$ is a prefix of $\gamma^{2n-1}(Ba)$ which is itself a prefix of $\gamma^{2n}(A)$.

Thus, from Lemma 9, we deduce that for every integer $n \geq 1$, $h(\gamma^{2n-2}(A))$ is a prefix of $h'(\gamma'^{2n}(\omega_1))$ which is itself a prefix of $h(\gamma^{2n}(A))$.

The result then follows from $H = \lim_{n\to\infty} h(\gamma^{2n}(A))$. □

5 Further questions

In this paper we have started the study of the Hilbert infinite word H. Many other questions could be looked for. Here are some of them.

The *subword complexity* of an infinite word $\mathbf{w}$ is the function counting the number of distinct length-n subwords of $\mathbf{w}$. It is reasonable to hope that, using either the HD0L-system or the tag-system generating H, the complexity function of this word could be obtained.

The *critical exponent* of an infinite word $\mathbf{w}$ is a number e such that $\mathbf{w}$ contains α-powers for some $\alpha < e$, but has no α-powers for $\alpha > e$ (it may or may not have e-powers). Here α is a rational number and e may be rational or real. Since H contains cubes and the only cubes it contains are one-letter cubes, the critical exponent of H is 3. However there is another interesting notion connected to the notion of critical exponent. Let us call *super-critical exponent* of an infinite word $\mathbf{w}$ the number e_s such that $\mathbf{w}$ contains α-powers for every rational number $\alpha < e_s$, but has no α-powers for $\alpha > e_s$ (again, it may or may not have e_s-powers). Of course $e_s \leq e$. In the case of some classical infinite words, as the Thue-Morse word or the Fibonacci word, the critical exponent and the super-critical exponent have the same value. But, in the present case the super-critical exponent is undoubtedly less than 3 because, from Theorem 7, H cannot contain all rational powers less than 3.

One of the referees proposed to replace absolute directions (left, right, up, down) by relative directions (left, right, straight). The infinite word so obtained on a 3-letter alphabet, which is of course different from H, is generated by a tag-system with a 7-letter alphabet and a uniform morphism whose image length is only 4. Comparing the properties of this word with those of H could be interesting.

Acknowledgements

Discussions with Valérie Berthé and Julien Cassaigne were helpful. Referees comments and suggestions considerably improved the paper.

References

[1] S. ARSHON, Démonstration de l'existence de suites asymétriques infinies, *Mat. Sb.* **44** (1937), 769–777 (in Russian), 777–779 (French summary).

[2] J.-P. ALLOUCHE, J. SHALLIT, *Automatic sequences: theory, applications, generalizations*, Cambridge University Press, Cambridge, UK, 2003.

[3] J. BERSTEL, Mots sans carré et morphismes itérés, *Discrete Math.* **29** (1980), 235–244.

[4] J. BERSTEL, Fibonacci words – a survey, *in*: G. Rozenberg, A. Salomaa (ed), *The Book of L*, Springer-Verlag 1986, 13–27.

[5] A. COBHAM, On the Hartmanis-Stearns problem for a class of tag machines, *in IEEE Conference Record of 1968 Ninth Annual Symposium on Switching and Automata Theory*, 1968, 51–60.

[6] A. COBHAM, Uniform tag sequences, *Math. Systems Th.* **6** (1972), 164–192.

[7] S. KITAEV, T. MANSOUR, P. SÉÉBOLD, Generating the Peano curve and counting occurrences of some patterns, *Journal of Automata, Languages and Combinatorics* **9** (2004), 439–455.

[8] D. HILBERT, Uber die stetige Abbildung einer Linie auf ein FlÂňachenstÂňuck, *Mathematische Annalen* **38** (1891), 459–460. Gesammelte Abhandlungen Vol III, 1-2, Springer-Verlag, Berlin, 1935.

[9] M. LOTHAIRE, *Algebraic Combinatorics on Words*, vol. 90 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge, UK, 2002.

[10] H. A. MAURER, G. ROZENBERG, E. WELZL, Using String Languages to Describe Picture Languages, *Inform. and Control* **54** (1982), 155–185.

[11] J.-J. PANSIOT, Hiérarchie et fermeture de certaines classes de tag-systèmes, *Acta Informatica* **20** (1983), 179–196.

[12] G. PEANO, Sur une courbe qui remplit toute une aire plane, *Mathemmatische Annalen* **36** (1890), 157–160.

[13] G. ROZENBERG, A. SALOMAA (Eds), *Handbook of Formal Languages* Vol. 1, Springer, 1997.

[14] H.SAGAN, *Space-Filling curves*, Springer-Verlag, New York, 1994.

[15] P. SÉÉBOLD, About some overlap-free morphisms on a n-letter alphabet, *Journal of Automata, Languages and Combinatorics* **7** (2002), 579–597.

[16] A. THUE, Über unendliche Zeichenreihen, *Christiania Vidensk.-Selsk. Skrifter. I. Mat. Nat. Kl.* **7** (1906), 1–22.
Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, 139-158.