Balanced Labeled Trees: Density, Complexity and Mechanicity

Nicolas Gast Bruno Gaujal

September 4, 2007

Abstract

Sturmian words are very particular infinite words with many equivalent definitions: minimal complexity of aperiodic sequences, balanced sequences and mechanical words. One natural way to generalized this definition to trees is to consider trees of complexity $n + 1$ but then one looses the two other properties.

In this paper we will study non-planar balanced tree and show that they are exactly mechanical trees. Moreover we will see that they are also aperiodic trees of minimal complexity.

1 Introduction

Sturmian words are infinite words over a binary alphabet, say $\{0, 1\}$, that have exactly $n + 1$ factors of length n. They also admit other equivalent definitions (see [**?**], for a rather exhaustive presentation of Sturmian words).

Definition 1.1. *A word $w \in \{0, 1\}^{\mathbb{N}}$ is Sturmian word if it verifies of the three equivalent properties.*
(i) For all $n \geq 0$: w has exactly $n + 1$ factor of length n.
(ii) w is balanced and aperiodic: if x and y are two factors of length n and if we denote by $|x|_1$ the number of 1 in x, then $\Big||x|_1 - |y|_1\Big| \leq 1$.
(iii) w is a mechanical word with an irrational slope: there exist $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and $\phi \in [0; 1[$ such that: for all i, $w_i = \lfloor (i + 1)\alpha + \phi \rfloor - \lfloor i\alpha + \phi \rfloor$ *or* $w_i = \lceil (i + 1)\alpha + \phi \rceil - \lceil i\alpha + \phi \rceil$.

In [**?**], Berstel and al. generalized this notion to Sturmian trees which are planar binary trees with complexity $n+1$, that is irrational trees with minimal complexity. But contrary to the case of words, it seams that there is no simple equivalent definition. In particular, there is no link with the balance property or with the mechanical construction.

In this document, we will focus on non-planar balanced tree, showing that they provide several existence and equivalent definitions in the flavor of Sturmian words. A longer version of this paper with detailed proofs is available on request to any of the two authors.

1.1 Definitions

Throughout this document, we will focus on trees which are rooted, binary, infinite, labelled by $\{0, 1\}$ and non planar (*i.e.* there is no distinction between the "left" or "right" sub-tree). More precisely:

Definition 1.2 (Infinite tree)**.** *A tree is a triplet $(\mathfrak{E}, \mathbf{P}, f)$ where:*

1. $\mathfrak{E} \subset \mathbb{N}$ *is the set of nodes.*

2. $\mathbf{P} : \mathfrak{E} \to \mathfrak{E}$ *has the following properties:*

   - $\exists! r$ *such that* $\mathbf{P}(r) = r$ *(r is the root of the tree)*
   - $\forall n \neq r : \ \mathbf{P}(n) < n$
   - $\forall n \neq r : \mathrm{card}(\{x/\mathbf{P}(x) = n\}) = 2$ *(the tree is binary).*
   - $\mathrm{card}(\{x/\mathbf{P}(x) = r\}) = 3.$

3. $f : \mathfrak{E} \to \{0; 1\}$*.* $f(n)$ *is the label of* n*.*

Definition 1.3 (Rooted sub-tree, non-rooted sub-tree)**.** *Let* $\mathcal{A} = (\mathfrak{E}, \mathbf{P}, f)$ *be a tree.*

We call sub-tree of root n *and height* $k \in \mathbb{N} \cup \{+\infty\}$ *and we denote by* $\mathcal{A}_{[n,k]} = (\mathfrak{E}_{[n,k]}, \mathbf{P}_{[n,k]}, f_{[n,k]})$*, the tree defined by:*

- $\mathfrak{E}_{[n,k]} = \left\{x/\exists q < k \text{ such that } \mathbf{P}^q(x) = n\right\}$
- $\mathbf{P}_{[n,k]}$ *is the restriction of* $\mathbf{P}$ *to* $\mathfrak{E}_{[n,k]}$ *with* $\mathbf{P}_{[n,k]}(n) = n$*.*
- $f_{[n,k]}$ *is the restriction of* f *to* $\mathfrak{E}_{[n,k]}$*.*

A non-rooted sub-tree of height k *and wide* 2^q *(q¡k) is the restriction of a tree to the subset* $\mathfrak{E}_{[n,k]} \setminus \mathfrak{E}_{[n,q]}$*.*

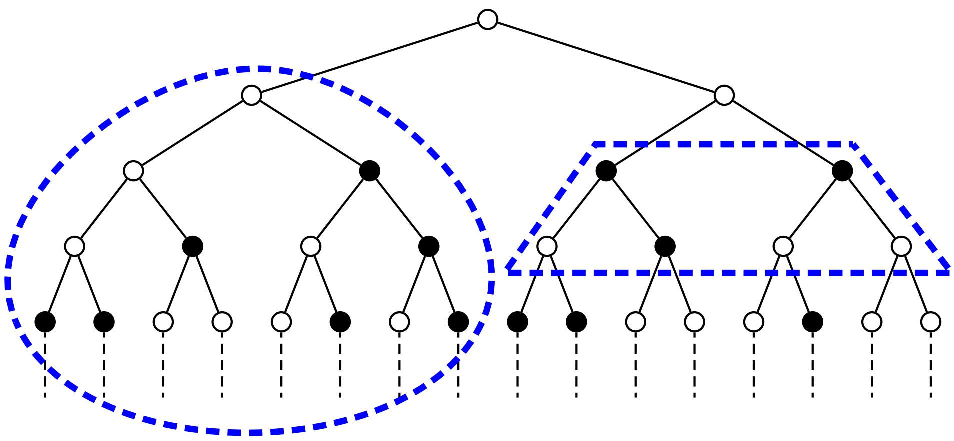


Figure 1: The left sub-tree is a rooted sub-tree of height 4, the right sub-tree is a non-rooted sub-tree of height 2 and width 2

In the rest of the document, we will call tree either an infinite tree or a sub-tree of an infinite tree. We can also define the height of tree $\mathcal{B}$ by ∞ if $\mathcal{B}$ is an infinite tree and by k if $\mathcal{B} = \mathcal{A}_{[n,k]}$ for a tree $\mathcal{A}$.

The number of node of a tree $\mathcal{B}$ is $\mathrm{card}(\mathbb{N})$ if $\mathcal{B}$ infinite and $2^k - 1$ if $\mathcal{B} = \mathcal{A}_{[n,k]}$.

Canonical representation

Two trees $(\mathfrak{E}, \mathbf{P}, f)$, $(\mathfrak{E}', \mathbf{P}', f')$ are equivalent if there exists a bijection $b : \mathfrak{E} \mapsto \mathfrak{E}'$ such that: $b(r) = r'$, $f(n) = f'(b(n))$ and $b(\mathbf{P}(n)) = \mathbf{P}'(b(n))$

Let $\mathcal{A}$ be a tree, we can choose a canonical representation of $\mathcal{A}$ by a word on $\{0,1\}$: if we call $\{\mathcal{B}\}$ the set of the trees equivalent to $\mathcal{A}$ such that $\forall k \forall 0 \leq i \leq k-1 : \mathbf{P}_{\mathcal{B}}(2^{k+1}+2i) = 2^k+i$, we can define for each $\mathcal{B}$ a word $u_{\mathcal{B}}$ by $u_i = f_{\mathcal{B}}(i)$. The word associated with $\mathcal{A}$ is the minimal word (considering the lexical order) of these $u_{\mathcal{B}}$.

Density

Let $\mathcal{A}$ be an infinite tree, n a node and $k \geq 0$. We define $h(\mathcal{A}_{[n,k]})$ to be the number of nodes labeled by 1 is the sub-tree of root n and height k:

$$h(\mathcal{A}_{[n,k]}) = \sum_{i \in \mathfrak{E}_{[n,k]}} f(i)$$

The density of a sub-tree $\Pi(\mathcal{A}_{[n,k]})$ by the number of 1 on its total number of nodes:

$$\Pi(\mathcal{A}_{[n,k]}) = \frac{h(\mathcal{A}_{[n,k]})}{2^k - 1}$$

We want to call density of an infinite tree $\mathcal{A}$, the average number of 1 in the tree. Actually, this quantity is the limit of the density of sub-trees of height n and it may or may not exists. We say that the density of a tree is α *iff*:

$$\lim_{k \to \infty} \Pi(\mathcal{A}_{[n,k]}) = \alpha$$

1.2 Rational trees: complexity and density

For words, there is a simple definition of what a periodic word is: it is a finite factor that *repeats* itself. The case of trees is a little more complicated. The property captured in the definition below is that a finite number of finite paterns will define the infinite tree.

Definition 1.4. *Let $\mathcal{A}$ be a infinite tree. We call $S_n(\mathcal{A})$ the set of the equivalence class of sub-trees of $\mathcal{A}$ of height n.*

We denote by $P(\mathcal{A}, n)$ the number of this class:

$$P(\mathcal{A}, n) = \text{card}(S_n(\mathcal{A}))$$

A tree has always a finite number of sub-trees of height k (bounded by $2^k - 1$). Let $n \geq 0$, we call *factor graph of order n* of $\mathcal{A}$ the graph $G_n = (S_n, E_n)$ defined by:

- S_n is the set of sub-trees of $\mathcal{A}$ defined above.
- a tuple (F, C_1, C_2) belongs to $E_n \subset S_n \times (S_n \times S_n)$ *iff* there exists three nodes f, c_1, c_2 such that c_1 and c_2 are the two children of f and f, c_1, c_2 are roots of sub-trees respectively equivalent to F, C_1, C_2. In that case, we say that there is an edge from F to $\{C_1, C_2\}$.

An example of such a graph is shown figure 2

Let us consider an infinite tree and let u be a node. The signification of this graph is that if the sub-tree of size n corresponding to u is F, its two children will be in the set $\{\{C_1, C_2\}/(F, C_1, C_2) \in \mathfrak{E}_n\}$. For example if the graph has exactly one outgoing edge for each vertices F, the tree is fixed by the graph and its first sub-tree.

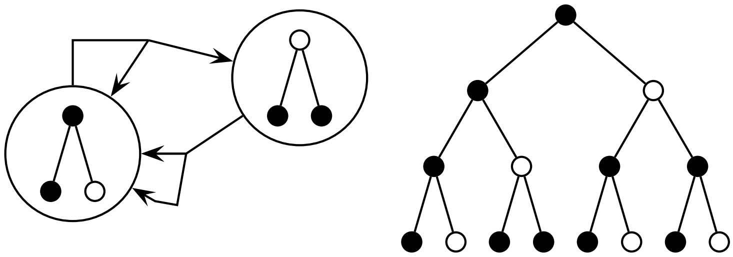


Figure 2: A rational tree and its factor graph

Definition 1.5 (Rational tree)**.** *Let $\mathcal{A}$ be an infinite tree labeled with an alphabet with k elements. We said that $\mathcal{A}$ is rational if it satisfy one of the three equivalent properties:*

(i) $\{P(\mathcal{A}, n)\}_n$ *is bounded*

(ii) $P(\mathcal{A}, n) = P(\mathcal{A}, n+1)$ *for one n.*

(iii) $P(\mathcal{A}, n) < n+k-1$ *for one n where k is the number of different letters appearing in $\mathcal{A}$.*

Proof. (i) implies (iii) is clear.

(iii) implies (ii): if $P(\mathcal{A}, i) < P(\mathcal{A}, i+1)$ for $i < n$, then $P(\mathcal{A}, n) \geq n - 1 + P(\mathcal{A}, 0) = n - 1 + k$, which contradicts (iii).

(ii) implies (i): for all n, we can define the factor graph of $\mathcal{A}$. As each node of this graph is a factor of the tree there is at least one edge going out of each node. Moreover the number of outgoing edges is the number of factor of length $n+1$ so it is $P(\mathcal{A}, n+1) = P(\mathcal{A}, n)$. Thus there is one and only one edge going out of each node. That means that if we start from a vertex of the graph, we will follow a deterministic path so there is exactly $P(\mathcal{A}, n)$ factors of length $k \leq n$ □

Proposition 1.6. *Let $\mathcal{A}$ be a rational tree. If $\mathcal{A}$ has a density α, then α is rational.*

Proof. (sketch) As $\mathcal{A}$ is rational, there exists k such that the tree is completely defined by its sub-trees $(A_1, \ldots, A_k)$ of height k. As the tree is rational, each sub-tree A_i has two children (A_{i_1}, A_{i_2}). We consider a Markov chain (X_n) on the set $\{A_1, \ldots, A_k\}$:

$$\mathbf{P}(X_{n+1} = A_{i_1} | X_n = A_i) = \mathbf{P}(X_{n+1} = A_{i_2} | X_n = A_i) = \frac{1}{2}$$

A sequence $X_0, \ldots, X_n, \ldots$ defines a unique path in the tree corresponding to a word w where w_i is the label of the root of A_i. We call density of this path the density of the word w if it exists.

If the Markov chain is irreducible and aperiodic, there exists α such that all paths w have almost surely a density α corresponding to the stationary distribution of the chain, which is a solution of a rational linear system and is therefore rational. The rest of the proof is skipped and is available in the long version of the paper.

□

2 Balanced trees

An infinite word w is said to be balanced if for all $n > 0$, two subsequences of w of length n will have almost the same density of 1. More precisely, if we call $h(x)$ the number of 1 in the word x, w is balanced *iif* for all $n > 0$ and for all x, y of length n:

$$|h(x) - h(y)| \leq 1$$

A balanced word has many properties. In particular, a balanced word always admits a density (which means $\lim_{n\to\infty} \frac{h(w_1 \ldots w_n)}{n}$ exists) and an aperiodic balanced word is Sturmian.

In this document, we will study two generalizations to trees of this definition. The first one and probably the most natural one, is what we call a *balanced tree*: a tree is said to be balanced if all of its rooted sub-trees of height k have almost the same number of 1:

Definition 2.1 (Balanced tree)**.** *A infinite tree $\mathcal{A}$ is* balanced *if for all node n and for all $k \geq 0$:*

$$|h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n',k]})| \leq 1$$

Let us also introduce a stronger definition: a tree is *strongly balanced* if all of its non-rooted sub-tree of height k and wide w have (almost) the same number of 1:

Definition 2.2 (Strongly balanced tree)**.** *A tree $\mathcal{A}$ is said to be* strongly balanced *if for all $k, q \geq 0$ and for all nodes n, n':*

$$\Big| |h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n,q]})| - |h(\mathcal{A}_{[n',k]}) - h(\mathcal{A}_{[n',q]})| \Big| \leq 1$$

This latest definition is clearly stronger than the standard one since taking $q = 0$ implies $\Big| h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n',k]}) \Big| \leq 1$. We will see in the next part that this definition is strictly stronger since there exist trees that are balanced but not strongly balanced.

Although the definition of a balanced tree is weaker and seams more natural for a generalization from words, we will see that strongly balanced tree have almost the same properties than its counterpart on words.

2.1 Density of a balanced tree

In this part, $\mathcal{A}$ denotes a balanced tree. As for the case of balanced sequences, we can define the density of a balanced tree.

For all n, $|h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n',k]}| \leq 1$ thus for all k we can define m_k which as the minimum of $h(\mathcal{A}_{[n,k]})$ over all n.

$$m_k \leq h(\mathcal{A}_{[n,k]}) \leq m_k + 1$$

A sub-tree of height $q + k$ is composed of a tree of size q which has between m_q and $m_q + 1$ ones with 2^q childs of size k which have between m_k and $m_k + 1$ ones. Thus:

$$2^q m_k + m_q \leq m_{q+k} \leq 2^q(m_k + 1) + m_q$$

Proposition 2.3 (Density of balanced tree). *Let $\mathcal{A}$ be a balanced tree.*

There exists a unique α, called the density *of $\mathcal{A}$ such that for all n:*

$$\lim_{k\to\infty} \Pi(\mathcal{A}_{[n,k]}) = \alpha$$

Moreover

$$\forall n, k : |h(\mathcal{A}_{[n,k]}) - \lfloor (2^k - 1)\alpha \rfloor| \leq 1$$

Proof. (sketch) The proof simply uses the definition of m_q to bound $\Pi(\mathcal{A}_{[n,k+q]}) - \Pi(\mathcal{A}_{[n,k]})$ on both sides. □

If $\mathcal{A}$ is a strongly balanced tree, there is a very similar formula linking its density and the number of 1 in one of its sub-tree: let us assume that $k > q > 0$, then for all node n, we have:

$$\left| h(\mathcal{A}_{[n,k]}) - h(\mathcal{A}_{[n,q]}) - (2^k - 2^q)\alpha \right| \leq 1 \tag{1}$$

The remarkable property about density it that it does not depends on the sub-tree chosen: there exists α such that the density of any sequence of sub-trees of increasing height will converge to α. Naturally, many trees do not have a density. For example the tree which line i is labeled by $i \bmod 2$ does not have a density.

Also, deciding if a finite tree is balanced can be done in linear time in the number of nodes of the tree.

Similarly, testing if a tree of height k is strongly balanced can be done in time $O(2^k)$ and size $O(k^2)$.

3 Mechanical trees

Let us recall the definition of a mechanical words: w is a mechanical word with slope α if there exist $\phi \in [0, 1)$ such that:

$$\text{for all } i : w_i = \lfloor (i+1)\alpha + \phi \rfloor - \lfloor i\alpha + \phi \rfloor$$

Sturmian words can be also defined as aperiodic balanced words or as mechanical words of irrational slope. In this part, we will see that we have the same equivalence properties between strongly balanced trees and mechanical trees.

Definition 3.1 (Mechanical tree). *A tree $\mathcal{A}$ is said to be mechanical of density α if for all node n:*

$$\exists \phi_n \in [0, 1[\ \forall k : \ h(\mathcal{A}_{[n,k]}) = \lfloor (2^k - 1)\alpha + \phi_n \rfloor \text{ or } \exists \phi_n \in [0, 1[\ \forall k : \ h(\mathcal{A}_{[n,k]}) = \lceil (2^k - 1)\alpha - \phi_n \rceil$$

If the node n of an mechanical tree verifies the relation $\forall k : \ h(\mathcal{A}_{[n,k]}) = \lfloor (2^k - 1)\alpha + \phi \rfloor$, this node is said inferior of phase ϕ. In fact it is an abuse of notation to say so because there could exist ϕ_1 and ϕ_2 such that for all k: $h(\mathcal{A}_{[n,k]}) = \lfloor (2^k - 1)\alpha + \phi_1 \rfloor = \lfloor (2^k - 1)\alpha + \phi_2 \rfloor$. However to simplify the notation, when the phase of a sub-tree is said to be ϕ, it could be any ϕ that works.

Let us call $\operatorname{frac}(x)$ the fractional part of a real number x and let us look at the sequence $\left(\operatorname{frac}(2^k\alpha - \alpha + \phi)\right)_k$. If this sequence can be arbitrary close to 0, this means that if $\psi > \phi$, there exists k such that $\lfloor (2^k-1)\alpha+\psi \rfloor > \lfloor (2^k-1)\alpha+\phi \rfloor$. On the other hand, if this sequence can be arbitrary close to 1, if $\psi < \phi$, then there exists k such that $\lfloor (2^k-1)\alpha+\psi \rfloor < \lfloor (2^k-1)\alpha+\phi \rfloor$.

Therefore, a phase ϕ is unique *iff* $\big(\text{frac}((2^k-1)\alpha+\phi)\big)_k$ is arbitrary close to 0 and 1.

Now let us call $\alpha_1, \alpha_2, \ldots, \alpha_n, \cdots \in \{0;1\}^{\mathbb{N}}$ the sequence of digits of α in base 2 and $x_1, \ldots, x_n, \ldots$ the sequence of digits of $\alpha - \phi$ in base 2. A multiplication by 2^k corresponds to a shift of the digits by k. If $x_1, \ldots, x_k$ does not end with neither an infinite number of 0 nor an infinite sequence of 1[1], $\big(\text{frac}((2^k-1)\alpha+\phi)\big)_k$ is arbitrary close to 0 or 1 means that for all n, there exists k such that $\alpha_k, \ldots, \alpha_{k+n-1} = x_1, \ldots, x_n$. More precisely, if for all n, there exists $m > n$ and $k > 0$ such that $x_m = 0$ (resp. 1) and $\alpha_k, \ldots, \alpha_{k+m-1} = x_1, \ldots, x_{m-1}, 1$ (resp. $x_1, \ldots, x_{m-1}, 0$), then $\big(\text{frac}((2^k-1)\alpha+\phi)\big)_k$ is arbitrary close to 0.

Thus we have to distinguish three cases:

- If α is a number such that all finite binary sequences appear in the binary expansion of α, then for all phase ϕ, then ϕ is unique. In particular, all *normal numbers* verify this property and we know that almost every number in $[0,1]$ is normal, see [?] (a number is normal if all sequences of length k appear with rate 2^{-k} in the binary expansion of α).
- If $\alpha \in \mathbb{Q}$, then the sequence $\text{frac}((2^k-1)\alpha+\phi)\big)$ is periodic and there are no phase ϕ such that ϕ is unique.
- If α is neither rational nor has the property that all binary sequences appear in α, then some ϕ can be unique and some others may not. For example, if α is the number:

  $$\alpha = 0.1011001110001111000011111100000\ldots,$$

  then if $\text{frac}(\alpha-\phi) = 0$, ϕ is unique, while ϕ_1 and ϕ_2 such that $\text{frac}(\alpha-\phi_1) = 0.10100$ and $\text{frac}(\alpha-\phi_2) = 0.1010$ are equivalent.

If one fixes the density α and the phase ϕ_n of each node, there is at most one matching mechanical tree.

Proposition 3.2. *Let $\alpha \in [0,1], \phi \in [0,1[$.*
There exists a unique mechanical tree $\mathcal{A}$ of density α and initial phase ϕ_0.

Proof. (sketch) It is based on some relations on rounding functions that will prove that if a node has a phase ϕ, the phases of its two children are fixed. □

3.1 Strongly balanced and mechanical are the same

In this section, we will show that if the density is irrational, a strongly balanced tree is a mechanical tree. This result can be expressed by the following proposition.

Proposition 3.3. *Let $\mathcal{A}$ be a infinite binary tree.*
(i) If $\mathcal{A}$ is mechanical then $\mathcal{A}$ is strongly balanced.
(ii) If $\mathcal{A}$ is strongly balanced and not rational, then $\mathcal{A}$ is a mechanical tree.
(iii) If $\mathcal{A}$ is strongly balanced and rational, $\mathcal{A}$ is ultimately mechanical.

[1] The case where x ends with an infinite number of 0 (or 1) is quite similar: if $x_1, x_2, \ldots, = x_1, x_l, 1, 0, 0, 0, \ldots,$ that means that $\big(\text{frac}((2^k-1)\alpha+\phi)\big)_k$ is arbitrarily close to 1 if for all m, there exists k such that $\alpha_k, \ldots, \alpha_{k+n-1} = x_1, \ldots, x_l, 0, 0, 0, \ldots, 0$

Proof. (sketch) The first point is a rather direct consequence of the definition of mechanical trees. As for the second point, let τ be a real number and n a node. At least one of the two following properties is true:

- For all k: $h(\mathcal{A}_{[n,k]}) \leq \lfloor (2^k - 1)\alpha + \tau \rfloor$
- For all k: $h(\mathcal{A}_{[n,k]}) \geq \lfloor (2^k - 1)\alpha + \tau \rfloor$

To prove this, assume that it is not true. Then there exists k, q such that $h(\mathcal{A}_{[n,k]}) < \lfloor (2^k - 1)\alpha + \tau \rfloor$ and $h(\mathcal{A}_{[n,k+q]}) > \lfloor (2^{k+q} - 1)\alpha + \tau \rfloor$ (or the opposite).

In that case: $h(\mathcal{A}_{[n,k+q]}) - h(\mathcal{A}_{[n,k]}) \leq 2 + \lfloor (2^{k+q}-1)\alpha + \phi \rfloor - \lfloor (2^k - 1)\alpha + \phi \rfloor > 1 + (2^{k+q} - 2^k)\alpha$ which violates the formula obtained in (1).

Let us define the number ϕ as follows:

$$\phi = \inf_{\tau} \left\{ \text{For all } k : h(\mathcal{A}_{[n,k]}) \leq \lfloor (2^k - 1)\alpha + \tau \rfloor \right\}$$

Then we have for all k:

$$h(\mathcal{A}_{[n,k]}) \leq (2^k - 1)\alpha + \phi \leq h(\mathcal{A}_{[n,k]}) + 1$$

If $\alpha \notin \{\frac{p}{2^k - 2^q}, p, k, q \in \mathbb{N}\}$, then $(2^k - 1)\alpha + \phi$ is an integer for at most one k_0. Depending on the value in k_0: if $h(\mathcal{A}_{[n,k_0]}) = (2^{k_0} - 1)\alpha + \phi$, then for all k: $h(\mathcal{A}_{[n,k]}) = \lfloor (2^k - 1)\alpha + \phi \rfloor$. Otherwise for all k: $h(\mathcal{A}_{[n,k]}) = \lceil (2^k - 1)\alpha + \phi - 1 \rceil$.

As for the last assertion, the same as for the previous point holds until the fact that there is at most integer k_0 such that $(2^k_0 - 1)\alpha + \phi \in \mathbb{N}$

Now we assume that there exists k_0, k_1 such that $(2^{k_i} - 1)\alpha + \phi \in \mathbb{N}$ (otherwise we go to the end of the proof of (ii)).

A direct computation leads to $\alpha \notin \{\frac{p}{2^k - 2^q}, p, k, q \in \mathbb{N}\}$ and thus there exist p and $q_0 < q_1$ such that

$$\alpha = \frac{p}{2^{q_0} - 2^{q_1}}, \mathrm{GCD}(p, 2^{q_0} - 2^{q_1}) = 1$$

According to the formula 1, this means that $h(\mathcal{A}_{[n,q_1]}) - h(\mathcal{A}_{[n,q_1]})$ equals $p - 1$,p or $p + 1$. But as the tree is strongly balanced, there can not be in the tree two sub-trees with values $p - 1$ and $p + 1$. Assume there is $q_1 - q_0$ other pairwise disjoint sub-trees with value $p + 1$, then there would be two of them for which the roots will be at height h and $h + (q_1 - q_0) * n$ and the minimal sub-tree that contains this two sub-trees would violate the formula 1.

Thus ultimately, all sub-trees corresponding to $h(\mathcal{A}_{[n,q_1]}) - h(\mathcal{A}_{[n,q_1]})$ will take the value p. For the rest of the proof, we assume that our tree verifies this property.

Now, let us recall that for all k:

$$h(\mathcal{A}_{[n,k]}) \leq (2^k - 1)\alpha + \phi \leq h(\mathcal{A}_{[n,k]}) + 1$$

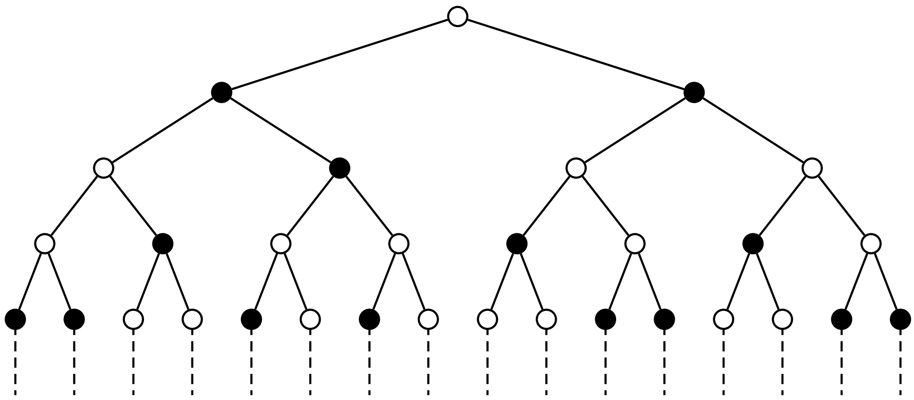
If there are two integers k_0, k_1 such that $(2^{k_i} - 1)\alpha + \phi \in \mathbb{N}$ we have $\frac{2^{k_1} - 2^{k_0}}{2^{q_0} - 2^{q_1}} p \in \mathbb{N}$. Thus $\frac{2^{k_1} - 2^{k_0}}{2^{q_0} - 2^{q_1}} \in \mathbb{N}$. If we write $x = q_1 - q_0$ and $y = k_1 - k_0$, $y = px + r$ $(0 \leq r < x)$, we have: $\frac{2^{k_1} - 2^{k_0}}{2^{q_0} - 2^{q_1}} = 2^{k_0 - q_0} \frac{2^x - 1}{2^y - 1}$ which implies $\frac{2^x - 1}{2^y - 1} \in \mathbb{N}$, $\frac{2^x - 1}{2^y - 1} = 2^r \frac{2^{px} - 1}{2^x - 1} + \frac{2^r - 1}{2^x - 1} \in \mathbb{N}$, so $r = 0$ and $q_0 - q_1$ divides $k_1 - k_0$.

As for all n': $h(\mathcal{A}_{[n',q_1]}) - h(\mathcal{A}_{[n',q_0]}) = p$, if $h(\mathcal{A}_{[n,k_0]}) = (2^{k_0} - 1)\alpha + \phi$ then $h(\mathcal{A}_{[n,k_1]}) = (2^{k_1} - 1)\alpha + \phi$ and then for all k:

$$h(\mathcal{A}_{[n,k]}) = \lfloor (2^k - 1)\alpha + \phi \rfloor$$

If $h(\mathcal{A}_{[n,k_0]}) = (2^{k_0} - 1)\alpha + \phi - 1$ then $h(\mathcal{A}_{[n,k_1]}) = (2^{k_1} - 1)\alpha + \phi - 1$ and then for all k: $h(\mathcal{A}_{[n,k]}) = \lceil (2^k - 1)\alpha + \phi \rceil$ □

Although we can characterize strongly balanced trees, it is still an open question to characterize balanced trees. The first answer is that there exist some trees that are balanced but not strongly balanced. We give an example with figure 3 where this picture is the beginning of a rational tree with density 3/7.



This tree is rational: it has 3 sub-trees of height 3 (the 3 written on the figure).

Figure 3: Beginning of a balanced tree of density 3/7 which is not strongly balanced (and thus non mechanical). This tree is not strongly balanced since in the picture $h(A) = 2$ and $h(B) = 0$. As this tree is rational, it is neither ultimately strongly balanced since the motifs A and B are repeated an infinite times.

The previous example shows a rational tree which is balanced but not strongly balanced but as we have seen, rational trees admit more exceptions than irrational ones. However, there are multiple examples of trees that are irrational but not strongly balanced. Such a construction is more involved and is given in the long version of the paper.

3.2 Strongly balanced tree are Sturmian

Proposition 3.4. *Let $\mathcal{A}$ be a mechanical tree and $k \geq 0$.*
(i) There exists at most $k+1$ sub-tree of height k
(ii) Moreover, if α is irrational, then the number of sub-trees is exactly $k+1$.

Proof. Let $\mathcal{A}$ be a mechanical tree of density α and let $k \leq 0$. According to the proposition 3.2, the sub-tree $\mathcal{A}_{[n,k]}$ depends only on its phase ϕ_n. In fact, this sub-tree depends only on the values $\lfloor (2^i - 1)\alpha + \phi_n \rfloor$.

For all $i \leq k$ and $\phi \leq 1$, we can define functions $h_i(.)$

$$h_i(\phi) = \lfloor (2^i - 1)\alpha + \phi \rfloor$$

These are increasing functions taking integer values and $h_i(1) - h_i(0) = 1$.

Thus the $k-$tuple $(h_1(\phi), \dots, h_k(\phi))$ can take at most $k+1$ values and there are at most $k+1$ sub-trees.

Moreover if α is irrational, the tree has an irrational density and then it is not rational so there are at least $k+1$ trees of height k which means that the tree has exactly $k+1$ sub-trees of height k. □

As shown in the previous section, if a tree has k factors of size k, it is rational. Thus strongly balanced trees of irrational density are trees with minimal complexity. We know that in the case of words, aperiodic balanced word are exactly words with complexity $n+1$ however this is not the case here. [**?**] gives some examples of planar Sturmian trees, although some of theses examples do not work in the non-planar case, many do.

- *Uniform trees*: considering a word w, define the uniform for w as the tree where a node with height k is labeled by w_k. If w is Sturmian, this tree is also Sturmian.
- *Left branch trees*: considering a word w_k, this definition works in the planar case: the label of a node n is w_k where k is the number of time with have to go left on a path from the root. Even if we consider the non-planar version of this tree, it is clearly Sturmian *iff* w is Sturmian.
- *Dyck tree*: recall that in the planar case, a node can be represented by a word on $\{0, 1\}$ representing the path from the root. The Dyck tree is the tree where the label of a node is 1 *iff* its representing word belongs to the Dyck language. The non-planar version of this tree is also Sturmian.

To end this part about complexity, notice that two mechanical trees with the same density are very close. Also, two mechanical trees with different densities just have a finite number of factors in common.

Proposition 3.5. *Let $\mathcal{A}$ and $\mathcal{B}$ be two mechanical trees. Let $S(\mathcal{A})$ and $S(\mathcal{B})$ be the set of their respective sub-trees.*
(i) If the densities of $\mathcal{A}$ and $\mathcal{B}$ are the same, then $S(\mathcal{A}) = S(\mathcal{B})$.
(ii) If the densities are different, then $S(\mathcal{A}) \cap S(\mathcal{B})$ is finite.