CHARACTERIZATION OF SETS OF LIMIT MEASURES OF A CELLULAR AUTOMATON ITERATED ON A RANDOM CONFIGURATION

BENJAMIN HELLOUIN DE MENIBUS AND MATHIEU SABLIK

ABSTRACT. The asymptotic behavior of a cellular automaton iterated on a random configuration is well described by its limit probability measure(s). In this paper, we characterize measures and sets of measures that can be reached as limit points after iterating a cellular automaton on a simple initial measure. In addition to classical topological constraints, we exhibit necessary computational obstructions. With an additional hypothesis of connectivity, we show these computability conditions are sufficient by constructing a cellular automaton realising these sets, using auxiliary states in order to perform computations. Adapting this construction, we obtain a similar characterization for the Cesàro mean convergence, a Rice theorem on the sets of limit points, and we are able to perform computation on the set of measures, i.e. the cellular automaton converges towards a set of limit points that depends on the initial measure. Last, under non-surjective hypotheses, it is possible to remove auxiliary states from the construction.

INTRODUCTION

A cellular automaton is a complex system defined by a local rule which acts synchronously and uniformly on the configuration space $\mathcal{A}^{\mathbb{Z}}$, where $\mathcal{A}$ is a finite alphabet. These simple models have a wide variety of different dynamical behaviors. We are interested in the typical asymptotic behavior starting from a random configuration, as this is usually done in simulations; different approaches stemmed from such observations. It is well-described by taking the iterated image of the initial measure under the action of the cellular automaton, and considering the limit points of this sequence in the weak* topology.

It is natural to ask which sets of measures can be obtained as limit points in this way. Obviously, any measure can be reached by iterating the identity on itself. Therefore, a more interesting approach is to start from some simple measure such as the uniform Bernoulli measure. In some sense, this corresponds to SRB measures which are "physically" relevant invariant measures obtained when starting from the Lebesgue measure in continuous dynamical systems [You02].

Formally speaking, given a simple initial measure μ, we want to characterize all reachable $\mathcal{V}(F,\mu)$, the sets of accumulation points of $(F_*^t\mu)_{t\in\mathbb{N}}$, the sequence of the images of μ under the iterated action of F, and $\mathcal{V}'(F,\mu)$, the sets of accumulation points of $\left(\frac{1}{t+1}\sum_{i=0}^{t} F_*^i\mu\right)_{t\in\mathbb{N}}$, the Cesàro mean of the previous sequence, for all possible cellular automata F.

Previous works focused on the μ-limit set, which corresponds to the union of the support of the limit measures [KM00, Ků05]. Very complex μ-limit sets can be constructed [BPT06, BDS10], and our construction is partly inspired from these works.

Describing limit measures has been done for only few concrete nontrivial examples. There are essentially two types of convergence quite well understood:

2010 *Mathematics Subject Classification.* 37B10,37B15,03D10,03D80.

Key words and phrases. Symbolic Dynamics, Cellular automata, SRB measures,Turing machines.

- convergence towards a simple measure: for example, the cyclic cellular automaton on three states introduced in [Fis90], starting from a Bernoulli measure, converges towards a linear combination of Dirac on uniform configurations [dMS11];
- randomisation phenomenon for linear cellular automata: the Cesàro mean sequence of the iteration of a linear cellular automaton on a initial measure converges to the uniform Bernoulli measure as soon as the initial measure is in a large class which contains Markov measures [FMMN00, MM98, PY02].

For any cellular automaton, starting from a Bernoulli measure or a Markov measure, we obtain after a finite number of steps a hidden Markov chain which is well understood [BP11]. If we consider a computable initial measure μ (which means that there is an algorithm that approximates at a known rate the probability that a word $u \in \mathcal{A}^*$ appears), then it is easy to see that $F_*^t\mu$ is also computable. For example, a Bernoulli or Markov measure is computable iff its parameters are computable real numbers.

The limit measure(s) are not necessarily computable since the speed of convergence is not known. Nevertheless, we show in Section 2 that there exists necessary computational obstructions. The main problem is to prove the reciprocal, in other words: given a set of measures satisfying the computational obstructions, construct a cellular automaton which, starting on any simple initial measure, reaches exactly this set asymptotically. Similar computational obstructions appear when characterizing possible topological dynamics properties of subshifts of finite type or cellular automata: possible entropies [HM10], possible growth-type invariants [Mey11], possible sub-actions [Hoc09, AS11]... However, the construction is quite different here since starting from a random configuration requires to self-organize the space, in the same spirit as the probabilistic cellular automaton of [Gác01] which corrects the random perturbations.

In Section 3, we construct a cellular automaton F such that, starting from any shift-mixing probability measure μ with full support, the limit points of the sequence of measures $(F_*^t\mu)_{t\in\mathbb{N}}$ are described as the accumulation points of a computable polygonal path of measures supported by periodic orbits. First of all the cellular automaton divides the initial configuration in segments and formats each segment using a method similar to the one developed in [DPST11]. Computation takes place in a negligible part of each segment and the result is copied periodically on the rest of the segment. In order to have an arbitrarily large area of computation, segments are merged progressively in a controlled manner. The difficulty of the construction is to synchronize all the operations to ensure the convergence.

In Section 4 we use the construction of Section 3 to solve some related problems, along with some open questions. The results are, for a fixed measure μ in a large class of measures:

- characterization of shift-invariant measures ν such that there exists a cellular automaton F which verifies $F_*^t\mu \underset{t\to\infty}{\longrightarrow} \nu$ (Corollary 1);
- characterization of connected subsets of shift-invariant measures $\mathcal{K}$ such that there exists a cellular automaton F which verifies $\mathcal{V}(F,\mu) = \mathcal{K}$ (Corollary 2);
- characterization of subsets of shift-invariant measures $\mathcal{K}'$ such that there exists a cellular automaton F which verifies $\mathcal{V}'(F,\mu) = \mathcal{K}'$ (Corollary 3);
- characterization of connected subsets of shift-invariant measures $\mathcal{K}' \subset \mathcal{K}$ such that there exists a cellular automaton F which verifies $\mathcal{V}(F,\mu) = \mathcal{K}$ and $\mathcal{V}'(F,\mu) = \mathcal{K}'$ (Corollary 4).
- Rice theorem for shift-invariant measures and connected subsets of shift-invariant measures reached by a cellular automaton (Corollaries 5, 6 and 7).

In Section 4.4, we consider the case where the set of limit points depends on the initial measure. Computational constraints appear to describe functions $\mu \longmapsto \mathcal{V}(F,\mu)$ that can be realized in this way. Indeed, it is possible to "transfer" the computational complexity of the initial measure (using it as an oracle) to the set of limit points. Modifying the construction of Section 3, we manage to build a set of limit points depending on the density of a special state; however, we do not obtain a complete characterization.

In the Section 5, we carry the previous characterizations to the case where auxiliary states are not allowed, i.e., the cellular automaton can only use the same alphabet as the limit measure(s). This is only possible, however, under some additional hypotheses on the support of the measures.

1. Definitions

1.1. Configuration space and cellular automata

Let $\mathcal{A}$ be a finite alphabet. Consider $\mathcal{A}^{\mathbb{Z}}$ the **space of configurations** which are $\mathbb{Z}$-indexed sequences in $\mathcal{A}$. If $\mathcal{A}$ is endowed with the discrete topology, $\mathcal{A}^{\mathbb{Z}}$ is compact, perfect and totally disconnected in the product topology. Moreover one can define a metric on $\mathcal{A}^{\mathbb{Z}}$ compatible with this topology:

$$\forall x, y \in \mathcal{A}^{\mathbb{Z}}, \quad d_C(x,y) = 2^{-\min\{|i| : x_i \neq y_i \ i \in \mathbb{Z}\}}.$$

Let $\mathbb{U} \subset \mathbb{Z}$. For $x \in \mathcal{A}^{\mathbb{Z}}$, denote $x_{\mathbb{U}} \in \mathcal{A}^{\mathbb{U}}$ the restriction of x to $\mathbb{U}$. Given a pattern $w \in \mathcal{A}^{\mathbb{U}}$, one defines the cylinder $[w]_{\mathbb{U}} = \{x \in \mathcal{A}^{\mathbb{Z}} : x_{\mathbb{U}} = w\}$. Denote $\mathcal{A}^* = \bigcup_n \mathcal{A}^n$ the set of all **finite words** $w = w_0 \dots w_{n-1}$; $|w| = n$ is the **length** of w. Also denote $[w]_i = [w]_{[i,i+|w|-1]}$ and $[w] = [w]_0 = [w]_{[0,|w|-1]}$. For any $u \in \mathcal{A}^*$ such that $|u| \leq |w|$, define the **frequency** of u in w as $\mathrm{Freq}(u,w) = \frac{1}{|w|-|u|+1}\mathrm{Card}\left(\left\{i \in [0,|w|-|u|] : w_{[i,i+|u|]} = u\right\}\right)$.

The **shift** map $\sigma : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ is defined by $\sigma(x)_i = x_{i+1}$ for $x = (x_m)_{m\in\mathbb{Z}} \in \mathcal{A}^{\mathbb{Z}}$ and $i \in \mathbb{Z}$. It is an homeomorphism of $\mathcal{A}^{\mathbb{Z}}$. For $w \in \mathcal{A}^*$, ${}^\infty w^\infty$ is the **σ-periodic word** defined by ${}^\infty w^\infty_{[0,|w|-1]} = w$ and $\sigma^{i+|w|}({}^\infty w^\infty) = \sigma^i({}^\infty w^\infty)$ for all $i \in \mathbb{Z}$.

A **cellular automaton** (CA) is a pair $(\mathcal{A}^{\mathbb{Z}}, F)$ where $F : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ is a continuous function that commutes with the shift ($\sigma \circ F = F \circ \sigma$). By Hedlund's theorem, it is equivalent to a function defined by $F(x)_i = \overline{F}((x_{i+u})_{u\in\mathbb{U}_F})$ for all $x \in \mathcal{A}^{\mathbb{Z}}$ and $i \in \mathbb{Z}$, where $\mathbb{U}_F \subset \mathbb{Z}$ is a finite set named **neighborhood** and $\overline{F} : \mathcal{A}^{\mathbb{U}_F} \to \mathcal{A}$ is a **local rule**.

1.2. Sets of measures on $\mathcal{A}^{\mathbb{Z}}$

1.2.1. Dynamical properties

Let $\mathfrak{B}$ be the Borel sigma-algebra of $\mathcal{A}^{\mathbb{Z}}$. Denote by $\mathcal{M}(\mathcal{A}^{\mathbb{Z}})$ the set of probability measures on $\mathcal{A}^{\mathbb{Z}}$ defined on the sigma-algebra $\mathfrak{B}$. Let $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ be the **σ-invariant probability measures** on $\mathcal{A}^{\mathbb{Z}}$, that is to say the measures $\mu \in \mathcal{M}(\mathcal{A}^{\mathbb{Z}})$ such that $\mu(\sigma^{-1}(B)) = \mu(B)$ for all $B \in \mathfrak{B}$.

Usually $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ is endowed with the **weak* topology**: a sequence $(\mu_n)_{n\in\mathbb{N}}$ of $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ converges to $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ if and only if, for all finite subsets $\mathbb{U} \subset \mathbb{Z}$ and for all patterns $u \in \mathcal{A}^{\mathbb{U}}$, one has $\lim_{n\to\infty} \mu_n([u]_{\mathbb{U}}) = \mu([u]_{\mathbb{U}})$. In the weak* topology, the set $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ is compact and metrizable. A metric is defined by

$$d_{\mathcal{M}}(\mu,\nu) = \sum_{n\in\mathbb{N}} \frac{1}{2^n} \max_{u\in\mathcal{A}^n} |\mu([u]) - \nu([u])|.$$

Define the **ball** centered on $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ of radius $\varepsilon > 0$ as

$$\mathbf{B}(\mu,\varepsilon) = \left\{\nu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}}) : d_{\mathcal{M}}(\mu,\nu) \leq \varepsilon\right\}.$$

A measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ is **σ-ergodic** if for every σ-invariant Borel subset $B \in \mathfrak{B}$ (that is to say $\sigma^{-1}(B) = B$ μ-almost everywhere), one has $\mu(B) = 0$ or 1. The set of σ-ergodic probability measures is denoted by $\mathcal{M}_{\sigma-\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}})$.

For $\mathbb{U} \subset \mathbb{Z}$ not necessarily finite, denote by $\mathfrak{B}_{\mathbb{U}}$ the σ-algebra generated by the set $\{[u]_{\mathbb{U}} : u \in \mathcal{A}^{\mathbb{U}'}, \mathbb{U}' \underset{\text{finite}}{\subset} \mathbb{U}\}$. Define the **weak mixing coefficients** of a measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ as

$$\psi_\mu(n) = \sup \left\{ \left| \frac{\mu(A \cap B)}{\mu(A)\mu(B)} - 1 \right| : A \in \mathfrak{B}_{]-\infty,0]},\, B \in \mathfrak{B}_{[n,\infty[},\, \mu(A) > 0,\, \mu(B) > 0 \right\}.$$

A measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ is **ψ-mixing** if $\psi_\mu(n) \underset{n\to\infty}{\longrightarrow} 0$. Denote $\mathcal{M}_{\sigma-\mathrm{mix}}(\mathcal{A}^{\mathbb{Z}})$ the set of ψ-mixing measures, of course $\mathcal{M}_{\sigma-\mathrm{mix}}(\mathcal{A}^{\mathbb{Z}}) \subset \mathcal{M}_{\sigma-\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}})$.

For a measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$, define $\mathrm{supp}(\mu)$, the **support** of μ, as the set of configurations of $\mathcal{A}^{\mathbb{Z}}$ such that any open neighborhood of these points have positive measure. Thus $\mu([u]) > 0$ for all $u \in \mathcal{A}^*$. Denote $\mathcal{M}^{\mathrm{full}}_{\sigma-\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}})$ the set of ergodic measures with full support, and $\mathcal{M}^{\mathrm{full}}_{\sigma-\mathrm{mix}}(\mathcal{A}^{\mathbb{Z}})$ the set of ψ-mixing measures with full support.

1.2.2. Classical examples

Let $\lambda = (\lambda_a) \in [0;1]^{\mathcal{A}}$ such that $\sum_{a\in\mathcal{A}} \lambda_a = 1$. The associated **Bernoulli measure** μ_λ is defined by

$$\mu_\lambda([u_0 \dots u_n]) = \lambda_{u_0} \cdots \lambda_{u_n} \qquad \text{for all } u_0 \dots u_n \in \mathcal{A}^*.$$

The **Dirac measure** supported by $x \in \mathcal{A}^{\mathbb{Z}}$ is defined as $\delta_x(A) = \mathbf{1}_{x\in A}$. Generally δ_x is not σ-invariant. However, if x is σ-periodic, it is possible to define the σ-invariant measure supported by x taking the mean of the measures $\delta_{\sigma^i(x)}$. Thus, for a word $w \in \mathcal{A}^*$, we define the **σ-invariant measure supported by** $^\infty w^\infty$ by

$$\widehat{\delta_w} = \frac{1}{|w|} \sum_{i\in[0,|w|-1]} \delta_{\sigma^i(^\infty w^\infty)}.$$

The set of measures $\left\{\widehat{\delta_w} : w \in \mathcal{A}^*\right\}$ is dense in $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ [Pet83].

1.2.3. Action of a cellular automaton on $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ and limit points

Let $(\mathcal{A}^{\mathbb{Z}}, F)$ be a cellular automaton and $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$. Define the **image measure** $F_*\mu$ by $F_*\mu(A) = \mu(F^{-1}(A))$ for all $A \in \mathfrak{B}$. Since F is σ-invariant, that is to say $F \circ \sigma = \sigma \circ F$, one deduces that $F_*(\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})) \subset \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ and $F_*(\mathcal{M}_{\sigma-\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}})) \subset \mathcal{M}_{\sigma-\mathrm{erg}}(\mathcal{A}^{\mathbb{Z}})$. This defines a continuous application $F_* : \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}}) \to \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$.

We consider $(F_*^t\mu)$ the sequence of iterated images of μ by F_*, and its **Cesàro mean** at time $t \in \mathbb{N}$ defined by

$$\varphi_t^F(\mu) = \frac{1}{t+1} \sum_{i=0}^{t} F_*^i\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}}).$$

For a measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$, we are interested in the asymptotic behavior of the sequences $(F_*^t\mu)_{t\in\mathbb{N}}$ and $(\varphi_t^F\mu)_{t\in\mathbb{N}}$. Define **$\mu$-limit measures set** $\mathcal{V}(F,\mu)$ as the the set of **limit points** of the sequence $(F^t\mu)_{t\in\mathbb{N}}$ and the **Cesàro mean μ-limit measures set** $\mathcal{V}'(F,\mu)$ as the set of limit points of the sequence $(\varphi_t^F\mu)_{t\in\mathbb{N}}$. Since $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ is compact, $\mathcal{V}(F,\mu)$ and $\mathcal{V}'(F,\mu)$ are nonempty. When $\mathcal{V}(F,\mu)$ is a singleton $\{\nu\}$, then $F_*^t\mu([u]) \underset{t\to\infty}{\longrightarrow} \nu([u])$.

Our main purpose is to characterize which sets of measures can be realised in this way. There are topological obstructions for these sets: $\mathcal{V}(F,\mu)$ and $\mathcal{V}'(F,\mu)$ are closed and thus compact, and $\mathcal{V}'(F,\mu)$ is connected since $d_{\mathcal{M}}(\varphi_t^F(\mu), \varphi_{t+1}^F(\mu)) \underset{t\to\infty}{\longrightarrow} 0$. In the next section, we show there are computability obstructions when the initial measure is computable.

2. Computability obstructions

2.1. Notion of computability

Definition 1. A **Turing machine** $\mathcal{TM} = (Q, \Gamma, \#, q_0, \delta, Q_F)$ is defined by:

- Γ a finite alphabet, with a blank symbol $\# \notin \Gamma$. Initially, a one-sided infinite memory tape is filled with $\#$, except for a finite prefix (the input), and a computing head is located on the first letter of the tape;
- Q the finite set of states of the head; $q_0 \in Q$ is the initial state;
- $\delta : Q \times \Gamma \to Q \times \Gamma \times \{\leftarrow, \cdot, \rightarrow\}$ the transition function. Given the state of the head and the letter it reads on the tape — depending on its position — the head can change state, replace the letter and move by one cell at most.
- $Q_F \subset Q$ the set of final states — when a final state is reached, the computation stops and the output is the value currently written on the tape.

A function $f : X \to Y$ with X and Y two enumerable sets is **computable** if there exists a Turing machine that, up to reasonable encoding, stops and returns $f(x)$ on any entry $x \in X$. In this paper, X and Y will be limited to $\mathbb{N}, \mathbb{Q}, \mathcal{A}^*$ and their products. Similarly, a set $K \subset X$ is computable if $\mathbf{1}_K$ is computable.

Remark. An encoding for X is simply a choice for Γ, and a surjection from a subset of Γ^* to X; strictly speaking, the computability of a function depends on the chosen encodings, but most natural choices give rise to the same result. For example, we can choose as an encoding for $\mathbb{Q}$ on $\{0, 1, |\}$ the function $\overline{p}^2|\overline{q}^2 \mapsto \frac{p}{q}$, where $\overline{p}^2$ is the binary representation of p.

2.2. Measures and computability

Definition 2. A measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$ is **computable** iff there exists $f : \mathcal{A}^* \times \mathbb{N} \to \mathbb{Q}$ computable such that

$$|\mu([u]) - f(u,n)| < 2^{-n} \qquad \text{for all } u \in \mathcal{A}^* \text{ and } n \in \mathbb{N}.$$

A sequence of measures $(\mu_i)_{i\in\mathbb{N}}$ is **computable** iff there exists $f : \mathcal{A}^* \times \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$ computable such that $|\mu_i([u]) - f(u,n,i)| < 2^{-n}$. This is a stronger statement than all μ_i are computable.

A measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$ is **semi-computable** iff there exists an computable sequence of measures $(\mu_i)_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty} \mu_i = \mu$. Equivalently there exists $f : \mathcal{A}^* \times \mathbb{N} \to \mathbb{Q}$ computable such that

$$|\mu([u]) - f(u,n)| \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{for all } u \in \mathcal{A}^*.$$

Denote $\mathcal{M}_\sigma^{\text{comp}}(\mathcal{A}^\mathbb{Z})$ the set of computable measures and $\mathcal{M}_\sigma^{\text{s-comp}}(\mathcal{A}^\mathbb{Z})$ the set of semi-computable measures. Of course $\mathcal{M}_\sigma^{\text{comp}}(\mathcal{A}^\mathbb{Z}) \subset \mathcal{M}_\sigma^{\text{s-comp}}(\mathcal{A}^\mathbb{Z})$. There exists an equivalent way to define these notions:

Proposition 1. *(i) A measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$ is computable if and only if there exists $f : \mathbb{N} \to \mathcal{A}^*$ computable such that $d_\mathcal{M}\left(\mu, \widehat{\delta_{f(n)}}\right) \le 2^{-n}$ for all $n \in \mathbb{N}$.*

(ii) A measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$ is semi-computable if and only if there exists $f : \mathbb{N} \to \mathcal{A}^*$ computable such that $\lim_{n\to\infty} \widehat{\delta_{f(n)}} = \mu$.

Proof. (i) Let $\mu \in \mathcal{M}_\sigma^{\text{comp}}(\mathcal{A}^\mathbb{Z})$. Given some $n \in \mathbb{N}$, we can enumerate words in $\mathcal{A}^*$ until we find a word $f(n)$ such that $|\mu([u]) - \widehat{\delta_{f(n)}}([u])| < 2^{-n-2}$ for all $u \in \mathcal{A}^k$ with $k \in [0, n+1]$. This is possible since the set $\left\{\widehat{\delta_w} : w \in \mathcal{A}^*\right\}$ is dense in $\mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$, and since μ and $\widehat{\delta_{f(n)}}([u])$ are computable. One has

$$d_\mathcal{M}(\mu, \widehat{\delta_{f(n)}}) = \sum_{i\in\mathbb{N}} \frac{1}{2^i} \max_{u\in\mathcal{A}^i} |\mu([u]) - \widehat{\delta_{f(n)}}([u])| \le \frac{1}{2^{n+1}} + \sum_{i\ge n+2} \frac{1}{2^i} \le \frac{1}{2^n}.$$

(ii) Let $\mu \in \mathcal{M}_\sigma^{\text{s-comp}}(\mathcal{A}^\mathbb{Z})$. There exists a computable sequence of measures $(\mu_i)_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty}\mu_i = \mu$. For each μ_n, we find a word $f(n) \in \mathcal{A}^*$ such that $d_\mathcal{M}(\mu_n, \widehat{\delta_{f(n)}}) \leq 2^{-n}$ for $n \in \mathbb{N}$. Clearly $f : \mathbb{N} \to \mathcal{A}^*$ is computable and $d_\mathcal{M}(\mu, \widehat{\delta_{f(n)}}) \leq d_\mathcal{M}(\mu, \mu_n) + d_\mathcal{M}(\mu_n, \widehat{\delta_{f(n)}}) \underset{n\to\infty}{\longrightarrow} 0$.

In both cases, the reciprocal is obvious. □

2.3. Action of a cellular automaton on computable measures

Proposition 2. *Let $(\mathcal{A}^\mathbb{Z}, F)$ be a cellular automaton. If $\mu \in \mathcal{M}_\sigma^{comp}(\mathcal{A}^\mathbb{Z})$ then $(F_*^t\mu)_{t\in\mathbb{N}}$ is a computable sequence of measures. In particular, if $F_*^t\mu \underset{t\to\infty}{\longrightarrow} \nu$ then $\nu \in \mathcal{M}_\sigma^{s\text{-}comp}(\mathcal{A}^\mathbb{Z})$.*

Proof. Suppose $|\mathcal{A}| = 2$ to simplify the proof. By definition, there is a computable function $f : \mathcal{A}^* \times \mathbb{N} \to \mathbb{Q}$ such that $|\mu([u]) - f(u,n)| \leq 2^{-n}$. Because F is defined locally, if we write $l = \min \mathbb{U}_F$ and $r = \max \mathbb{U}_F$, $F^t(x)_{[0,k]}$ will depend only on $x_{[lt,rt+k]}$. In other words, for all $u \in \mathcal{A}^k$, there is a set $\mathbf{Pred}_t(u) \subset \mathcal{A}^{[lt,rt+k]}$ such that $F^{-t}([u]) = \cup_{v\in\mathbf{Pred}_t(u)}[v]$. Now consider the function

$$f' : (u,n,t) \mapsto \sum_{v\in\mathbf{Pred}_t(u)} f(v, 2n + (r-l)t).$$

It is computable by enumerating elements of $\mathcal{A}^{k+(r-l)t}$ and checking if $F^t([v]_{-lt}) \subset [u]$ by iterating the local rule on v. Finally

$$\begin{aligned} |F_*\mu([u]) - f'(u,n,t)| &= \left| \mu\left(\bigcup_{v\in\mathbf{Pred}_t(u)} [v]\right) - \sum_{v\in\mathbf{Pred}_t(u)} f(v, 2n+(r-l)t) \right| \\ &\leq \sum_{v\in\mathbf{Pred}_t(u)} |\mu([v]) - f(v, 2n+(r-l)t)| \\ &\leq 2^{n+(r-l)t} \cdot 2^{-2n-(r-l)t} = 2^{-n} \end{aligned}$$

which means that $(F_*^t\mu)_{t\in\mathbb{N}}$ is a computable sequence of measures. □

We have obtained a computability obstruction on single limit measures. In the following section, we extend this obstruction to sets of limit points.

2.4. Compact sets in computable analysis

We introduce computability notions on compact sets. This is a part of the theory of computable analysis on metric spaces for which a standard reference book is [Wei00]. In a general metric space, we define computability by using a countable dense subset, here $(\widehat{\delta_w})_{w\in\mathcal{A}^*}$.

Definition 3. Let X, Y be two enumerable sets.

A sequence of functions $(f_i : X \to Y)_{i\in\mathbb{N}}$ is **computable** if $(i,x) \mapsto f_i(x)$ is computable.

A function $f : X \to Y$ is Σ_2**-computable** (resp. Π_2**-computable**) if $f = \sup_{i\in\mathbb{N}} \inf_{j\in\mathbb{N}} f_{i,j}$ (resp. $f = \inf_{i\in\mathbb{N}} \sup_{j\in\mathbb{N}} f_{i,j}$), where $(f_{i,j})_{(i,j)\in\mathbb{N}^2}$ is a computable sequence of functions. The computability of set $K \subset X$ is defined as the computability of its characteristic function.

Definition 4. Extending the previous definition to uncountable sets, a closed set $\mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$ is Σ_2**-computable** if the set $\left\{(w,r) \in \mathcal{A}^* \times \mathbb{Q} \ : \ \overline{\mathbf{B}(\widehat{\delta_w}, r)} \cap F = \emptyset\right\}$ is Σ_2-computable, that is to say the characteristic function of this enumerable set is Σ_2-computable.

Remark. The symmetric notions of Π_2- and Σ_2-computability comes from an analogy with the real arithmetic hierarchy [ZW01, Zie05]. These definitions extend naturally to Π_n- and Σ_n-computability.

The Σ_2-computability of a closed set can be defined in other equivalent ways. We first need to extend to notion of Σ_2-computability to functions mapping noncountable sets.

Definition 5. A sequence of functions $(f_n : \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z}) \longrightarrow \mathbb{R})_{n\in\mathbb{N}}$ is **computable** if:

- there exists $a : \mathbb{N}\times\mathbb{N}\times\mathcal{A}^* \longrightarrow \mathbb{Q}$ computable such that $\left|f_n(\widehat{\delta_w}) - a(n,m,w)\right| \leq \frac{1}{m}$ for every $w \in \mathcal{A}^*$ and $n, m \in \mathbb{N}$ (sequential computability);
- there exists $b : \mathbb{N} \longrightarrow \mathbb{Q}$ computable such that $d_\mathcal{M}(\mu,\nu) < b(m)$ implies $|f_n(\mu) - f_n(\nu)| \leq \frac{1}{m}$ for all $n, m \in \mathbb{N}$ (computable uniform equicontinuity).

A function $f : \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z}) \longrightarrow \mathbb{R}$ is **semi-computable** (or Δ_2-computable) if there exists a computable sequence of functions $(f_n : \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z}) \longrightarrow \mathbb{R})_{n\in\mathbb{N}}$ such that $f = \lim_n f_n$.
A function $f : \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z}) \longrightarrow \mathbb{R}$ is Σ_2**-computable** if there exists a computable sequence of functions $(f_{i,j} : \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z}) \longrightarrow \mathbb{R})_{(i,j)\in\mathbb{N}}$ such that $f = \sup_i \inf_j f_{i,j}$.

Proposition 3. *Let $\mathcal{K}$ be a closed set. The following are equivalent:*

(1) *the set* $\left\{(w,r) \in \mathcal{A}^* \times \mathbb{Q} \;:\; \overline{\mathbf{B}(\widehat{\delta_w}, r)} \cap \mathcal{K} = \emptyset\right\}$ *is Σ_2-computable;*
(2) *$d_\mathcal{K}$ is Σ_2-computable;*
(3) *$\mathcal{K} = f^{-1}(\{0\})$ where f is a semi-computable function.*

Proof.
($1 \Rightarrow 2$) Assume there is a computable function $f : \mathbb{N}^2 \times \mathcal{A}^* \times \mathbb{Q} \longrightarrow \{0,1\}$ such that, for every $w \in \mathcal{A}^*$ and $r \in \mathbb{Q}$, $\overline{\mathbf{B}(\widehat{\delta_w}, r)} \cap \mathcal{K} = \emptyset \Leftrightarrow \sup_i \inf_j f(i,j,w,r) = 1$.

Consider the sequence $\left(d_{i,j,w,r}(\mu) = f(i,j,w,r)\max\left(0, r - d_\mathcal{M}(\widehat{\delta_w}, \mu)\right)\right)_{(i,j,w,r)\in\mathbb{N}^2\times\mathcal{A}^*\times\mathbb{Q}}$. The function $(i,j,w,r,w') \mapsto d_{i,j,w,r}(\widehat{\delta_{w'}})$ is computable and every $d_{i,j,w,r}$ is 1-Lipschitz, hence this sequence of functions is computable. We now show that $d_\mathcal{K} = \sup_{w,r}\sup_i \inf_j d_{i,j,w,r}$.

For any (w,r) such that $\sup_i \inf_j f(i,j,w,r) \neq 0$, then $d_\mathcal{K}(\widehat{\delta_w}) > r$, and thus $\sup_i \inf_j d_{i,j,w,r}(\mu) = \max\left(0, r - d_\mathcal{M}(\widehat{\delta_w}, \mu)\right) \leq \max\left(0, d_\mathcal{K}(\widehat{\delta_w}) - d_\mathcal{M}(\widehat{\delta_w}, \mu)\right) \leq d_\mathcal{K}(\mu)$ for all $\mu \in \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$.

If $\mu \in \mathcal{K}$, we conclude that $\sup_{i,w,r} \inf_j d_{i,j,w,r}(\mu) = 0 = d_\mathcal{K}(\mu)$.

Now let $\mu \notin \mathcal{K}$. For all $\varepsilon > 0$, there exists w such that $d_\mathcal{M}(\widehat{\delta_w}, \mu) \leq \varepsilon$ and $\widehat{\delta_w} \notin \mathcal{K}$. Let $r \in \mathbb{Q}$ be such that $0 < d_\mathcal{K}(\widehat{\delta_w}) - r < \varepsilon$, which implies that $\overline{\mathbf{B}(\widehat{\delta_w}, r)} \cap \mathcal{K} = \emptyset$ and so $\sup_i \inf_j f(i,j,w,r) \neq 0$. Furthermore $d_\mathcal{K}(\mu) \leq d_\mathcal{K}(\widehat{\delta_w}) + d_\mathcal{M}(\widehat{\delta_w}, \mu) \leq r + 2\varepsilon$, and $\sup_i \inf_j d_{i,j,w,r} = r - d_\mathcal{M}(\widehat{\delta_w}, \mu) > d_\mathcal{K}(\mu) - 3\varepsilon$. The latter is true for every $\varepsilon > 0$, and we deduce that $\sup_{i,w,r} \inf_j d_{i,j,w,r} = d_\mathcal{K}(\mu)$.

($2 \Rightarrow 3$) Let $(d_{i,j} : \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z}) \longrightarrow \mathbb{R})_{(i,j)\in\mathbb{N}^2}$ be a computable sequence of functions such that $d_\mathcal{K} = \sup_{i\in\mathbb{N}} \inf_{j\in\mathbb{N}} d_{i,j}$. Denote $g_{i,n} = \inf\{d_{i,j} \;:\; j \in [0,n]\}$.

$$\begin{aligned} d_\mathcal{K}(\mu) = 0 &\Leftrightarrow \sum_{i\in\mathbb{N}} \frac{1}{2^i}\left(\inf_{j\in\mathbb{N}} d_{i,j}(\mu)\right) = 0 \\ &\Leftrightarrow \sum_{i\in\mathbb{N}} \frac{1}{2^i}\left(\lim_{n\to\infty} g_{i,n}(\mu)\right) = 0 \\ &\Leftrightarrow \lim_{n\to\infty} \sum_{i\in\mathbb{N}} \frac{1}{2^i} g_{i,n}(\mu) = 0 \end{aligned}$$

Let $f_n : \mu \mapsto \sum_{i\in\mathbb{N}} \frac{1}{2^i} g_{i,n}(\mu)$. $(f_n)_{n\in\mathbb{N}}$ is a computable sequence of functions, since computing $(n, w') \mapsto f_n(\widehat{\delta_{w'}})$ up to precision 2^{-r} only requires to compute the values of $d_{i,j}(\widehat{\delta_{w'}})$ for $i, j \in \{0,\ldots,r\}$, and the effective uniform equicontinuity of $(f_n)_{n\in\mathbb{N}}$ is a consequence of the effective uniform equicontinuity of $(d_{i,j})_{(i,j)\in\mathbb{N}^2}$. Thus $F = f^{-1}(0)$ where $f = \lim_n f_n$.

($3 \Rightarrow 1$) Let $(f_n : \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z}) \longrightarrow \mathbb{R})_{n\in\mathbb{N}}$ be a computable sequence of functions such that $f = \lim_n f_n$. For any $q \in \mathbb{Q}, w' \in \mathcal{A}^*$ and $i \in \mathbb{N}$, let $F_{q,w',i} = \left\{(w,r) \in \mathcal{A}^* \times \mathbb{Q} \ : \ d_\mathcal{M}(\widehat{\delta_w}, \widehat{\delta_{w'}}) \le r \Rightarrow |f_i(\widehat{\delta_{w'}})| > q\right\}$. The functions $(q, w', i, w, r) \mapsto \mathbf{1}_{(w,r)\in F_{q,w',i}}$ are computable and thus the characteristic functions $\mathbf{1}_{F_{q,w',i}}$ are sequentially computable. Additionally:

$$F = \bigcup_{\substack{q\in\mathbb{Q}^+\\ n\in\mathbb{N}}} \bigcap_{\substack{w'\in\mathcal{A}^*\\ i\ge n}} F_{q,w',i} \qquad \text{and thus} \qquad \mathbf{1}_F = \sup_{\substack{q\in\mathbb{Q}^+\\ n\in\mathbb{N}}} \inf_{\substack{w'\in\mathcal{A}^*\\ i\ge n}} \mathbf{1}_{F_{q,w',i}}.$$

Indeed, let (w, r) be such that $\overline{\mathbf{B}(\widehat{\delta_w}, r)} \cap \mathcal{K} = \emptyset$. Let $\varepsilon = \min\{|f(\mu)| \ : \ \mu \in \overline{\mathbf{B}(\widehat{\delta_w}, r)}\} > 0$. For any $\mu \in \overline{\mathbf{B}(\widehat{\delta_w}, r)}$, there is a rank $n_\varepsilon(\mu)$ such that for all $n \ge n_\varepsilon(\mu)$, $f_n(\mu) > \frac{3\varepsilon}{4}$. By taking $r_\varepsilon \in \mathbb{N}$ such that $b(r_\varepsilon) < \frac{\varepsilon}{4}$, where b is given in the definition of the computable uniform equicontinuity of $(f_n)_{n\in\mathbb{N}}$, we have $f_n(\nu) > \frac{\varepsilon}{2}$ for all $\nu \in \overline{\mathbf{B}(\mu, b(r_\varepsilon))}$ and all $n \ge n_\varepsilon(\mu)$. Since $\overline{\mathbf{B}(\widehat{\delta_w}, r)}$ is compact, it can be covered by a finite number of balls of radius r_ε, and we take n_ε the maximal value of $n_\varepsilon(\mu)$ on all the ball centers.

The previous paragraph shows that for all w' such that $\widehat{\delta_{w'}} \in \overline{\mathbf{B}(\widehat{\delta_w}, r)}$, $|f_i(\widehat{\delta_{w'}})| > \frac{\varepsilon}{2}$ for all $i \ge n_\varepsilon$. Thus $(w, r) \in F$ by taking any $q \le \frac{\varepsilon}{2}$. The converse is clear.

□

Proposition 4. *Let $(\mathcal{A}^\mathbb{Z}, F)$ be a cellular automaton and $\mu \in \mathcal{M}_\sigma^{comp}(\mathcal{A}^\mathbb{Z})$. Then $\mathcal{V}(F, \mu)$ and $\mathcal{V}'(F, \mu)$ are Σ_2-computable compact sets.*

Proof. Let $f_n : \nu \mapsto d_\mathcal{M}(F_*^n\mu, \nu)$. Since μ is computable, $(f_n)_{n\in\mathbb{N}}$ is sequentially computable. Moreover $|f_n(\nu) - f_n(\nu')| = |d_\mathcal{M}(F_*^n\mu, \nu) - d_\mathcal{M}(F_*^n\mu, \nu')| \le d_\mathcal{M}(\nu, \nu')$ so $(f_n)_{n\in\mathbb{N}}$ is computably uniformly continuous. The result follows from the fact that $d_{\mathcal{V}(F,\mu)}(\nu) = \liminf_{n\to\infty} d_\mathcal{M}(F_*^n\mu, \nu) = \sup_m \inf_{n>m} d_\mathcal{M}(F_*^n\mu, \nu)$.

The same reasoning holds for $\mathcal{V}'(F, \mu)$. □

When the initial measure is not computable, it can be used as an oracle. These obstructions will be generalized accordingly in Section 4.4.

2.5. Some examples

These computable obstructions are not restrictive and it is possible to exhibit a wide variety of computable measures, semi-computable measures or Σ_2-computable compact sets of measures:

- a Bernoulli measure or a Markov measure with computable (resp. semi-computable) parameters are computable (semi-computable);
- an unique ergodic subshift which is effective has a semi-computable measure; this is the case for any subshift obtained by a primitive substitution or a Sturmian subshift where the slope is computable;
- the set of shift-invariant measures and the set of measures of maximal entropy for any effective subshift are Σ_2-computable compact sets;
- denote λ_p the Bernoulli measure on $\{0,1\}^\mathbb{Z}$ such that $\mu([0]) = p$. The set $\{\lambda_p : p \in F\}$, where F is a Σ_2-computable closed subset of $[0, 1]$, is a Σ_2-computable compact set of $\mathcal{M}_\sigma(\{0,1\}^\mathbb{Z})$ but is connected only if F is. However $\{\alpha\lambda_p + (1-\alpha)\lambda_q : p, q \in F \text{ and } \alpha \in [0, 1]\}$ is a Σ_2-computable compact connected set of $\{0,1\}^\mathbb{Z}$;

- denote $\mu_p \in \mathcal{M}_\sigma(\{0,1\}^{\mathbb{Z}})$ the measure supported by the Sturmian subshift of slope α. The set $\{\mu_p : p \in F\}$, where F is a Σ_2-computable closed subset of $[0,1]$, is a Σ_2-computable compact set of $\mathcal{M}_\sigma(\{0,1\}^{\mathbb{Z}})$ but is not connected if F it is not.

As we will see in Section 4.1, these sets and many others can be realized as the μ-limit measures set of a cellular automaton.

2.6. Technical characterization of Σ_2-computable compact connected sets

To build a cellular automaton reaching a Σ_2-computable compact set of measures as its μ-limit measures set, we need a recursive enumeration of words $(w_n)_{n\in\mathbb{N}}$ which describes it in a certain way. For technical reasons, the μ-limit measures set of the construction presented in Section 3 is a connected set, because it builds an infinite polygonal path composed of segments of the form $\left[\widehat{\delta_u}, \widehat{\delta_v}\right] = \left\{t\widehat{\delta_u} + (1-t)\widehat{\delta_v} : t \in [0,1]\right\} \subset \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ where $u, v \in \mathcal{A}^*$. The following proposition describes how such connected sets can be covered by a polygonal path.

Definition 6. Let $(w_n)_{n\in\mathbb{N}}$ be a sequence of words of A^*. Denote $\mathcal{V}((w_n)_{n\in\mathbb{N}})$ the **limit points of the polygonal path** defined by the sequence of measures $(\widehat{\delta_{w_n}})_{n\in\mathbb{N}}$

$$\mathcal{V}((w_n)_{n\in\mathbb{N}}) = \bigcap_{N>0} \overline{\bigcup_{n\geq N} \left[\widehat{\delta_{w_n}}, \widehat{\delta_{w_{n+1}}}\right]}.$$

Proposition 5. *Let $\mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ be a non-empty Σ_2-computable, compact, connected set (Σ_2-CCC for short). Then there exists a computable sequence of words $(w_n)_{n\in\mathbb{N}}$ such that $\mathcal{K} = \mathcal{V}((w_n)_{n\in\mathbb{N}})$.*

Proof. By Proposition 3 there is a computable sequence of functions $(f_n)_{n\in\mathbb{N}}$ satisfying $\mathcal{K} = f^{-1}(\{0\})$ where $f = \lim_{n\in\mathbb{N}} f_n$, and let $a : \mathbb{N}\times\mathbb{N}\times\mathcal{A}^* \to \mathbb{Q}$ and $b : \mathbb{N} \to \mathbb{Q}$ be the computable functions given by Definition 5. Without loss of generality,we can assume that b is a strictly decreasing function and $b(i) \underset{i\to\infty}{\longrightarrow} 0$. For $k \in \mathbb{N}$, define $\alpha_k = \min\left\{l \in \mathbb{N} \ : \ \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}}) = \bigcup_{u\in\mathcal{A}^{\leq l}} \mathbf{B}(\widehat{\delta_u}, b(k))\right\}$.

Define:

$$\mathbf{V}_k = \left\{w \in \mathcal{A}^{\leq \alpha_k} : \exists n \geq k \text{ such that } a(n, 2k, w) < \frac{2}{k}\right\}$$

$$\mathbf{V}_k^t = \left\{w \in \mathcal{A}^{\leq l} : \begin{array}{l} \exists n \in [k,t] \text{ such that } a(n,2k,w) < \frac{2}{k} \\ l = \min\{i \leq t \ : \ \forall u \in \mathcal{A}^{\leq t}, w \in \mathcal{A}^{\leq i}, d_{\mathcal{M}}(\widehat{\delta_u}, \widehat{\delta_w}) \leq b(k)\} \end{array}\right\}$$

Claim 1: *$\mathbf{V}_k^t$ is increasing w.r.t. t and $\exists T_k, \mathbf{V}_k = \mathbf{V}_k^{T_k}$. Furthermore, the function $(k,t) \to \mathbf{V}_k^t$ is computable.*

Proof. For all k and t, it is clear that $\mathbf{V}_k^t \subset \mathbf{V}_k^{t+1}$. The conditions for being included in $\mathbf{V}_k^t$ can be checked by computing computable functions over a finite range of values, so $(k,t) \mapsto \mathbf{V}_k^t$ is computable.

By definition, we have $l \leq \alpha_k$. Because the periodic measures are dense in $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$, we actually have $l = \alpha_k$ when t is large enough. Furthermore, if $w \in \mathbf{V}_k$, then w satisfies the first condition in $\mathbf{V}_k^t$ for t large enough. Therefore, there is a T_k such that $\mathbf{V}_k = \mathbf{V}_k^{T_k}$. $\diamond$ Claim 1

Notice that T_k is not necessarily computable, which means that even though $\mathbf{V}_k$ is finite, we do not know when the computation is finished. The algorithm for computing the sequence $(w_n)_{n\in\mathbb{N}}$ is the following:

Algorithm.

- Compute each $\mathbf{V}_k^t$ for $k \leq t$, for increasing values of t;
- Assume $w_0 \dots w_n$ have already been computed and a new element $w \in \mathbf{V}_k^{t+1} \backslash \mathbf{V}_k^t$ is computed.

– Find the largest $i \leq k$ such that one can find a path $w_n = u_0, u_1, \ldots, u_l = w$ with $u_1, \ldots u_{l-1} \in V_i^t$ and $d_{\mathcal{M}}(u_k, u_{k+1}) \leq 4b(i)$.
– This path is added to the sequence (if no such path is found, w alone is added to the sequence).

Now we will prove the correctness of this algorithm.

Claim 2: *If $\mu \in \mathcal{K}$, then $\mu \in \mathcal{V}((w_n)_{n\in\mathbb{N}})$.*

Proof. There is a sequence of words $(u_n)_{n\in\mathbb{N}}$ such that $u_n \in \mathcal{A}^{\alpha_n}$ and $d_{\mathcal{M}}(\widehat{\delta_{u_n}}, \mu) < b(n)$ for all $n \in \mathbb{N}$; by equicontinuity one has $|f(\widehat{\delta_{u_n}}) - f(\mu)| < \frac{1}{n}$ so $f(\widehat{\delta_{u_n}}) < \frac{1}{n}$. Thus, there is a $t > |u_n|$ such that $f_t(\widehat{\delta_{u_n}}) < \frac{3}{2n}$. One deduces that $a(t, 2n, u_n) \leq f_t(\widehat{\delta_{u_n}}) + \frac{1}{2n} < \frac{2}{n}$, which means that each $u_n \in \mathbf{V}_n$ for every n, and by construction it appears at some point in the sequence $(w_n)_{n\in\mathbb{N}}$. ◇ Claim 2

Claim 3: $\forall\varepsilon > 0, \exists k_\varepsilon, \forall k > k_\varepsilon, w \in \mathbf{V}_k \Rightarrow d_{\mathcal{M}}(\widehat{\delta_w}, \mathcal{K}) \leq \varepsilon$.

Proof. By compacity, there exists a $\delta_\varepsilon > 0$ such that $f(\widehat{\delta_w}) \leq \delta_\varepsilon \Rightarrow d_{\mathcal{M}}(\widehat{\delta_w}, \mathcal{K}) \leq \varepsilon$.

Now let $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ be any measure such that $f(\mu) \geq \delta_\varepsilon$. $\exists n_\varepsilon(\mu), \forall n \geq n_\varepsilon(\mu), f_n(\mu) > \frac{2\delta_\varepsilon}{3}$. By taking $r_\varepsilon \in \mathbb{N}$ such that $\frac{1}{r_\varepsilon} < \frac{\delta_\varepsilon}{3}$, we have by computable uniform equicontinuity of $(f_n)_{n\in\mathbb{N}}$ $f_n(\nu) > \frac{\delta_\varepsilon}{3}$ for all $\nu \in \mathbf{B}(\mu, b(r_\varepsilon))$ and all $n \geq n_\varepsilon(\mu)$.

Since $\{\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}}) : f(\mu) \geq \delta_\varepsilon\}$ is compact, we can cover it with a finite number of balls of radius $b(r_\varepsilon)$, and we define n_ε the maximum value of $n_\varepsilon(\mu)$ on ball centers. Thus, $\forall n > n_\varepsilon, \forall\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}}), f(\mu) > \delta_\varepsilon \Rightarrow f_n(\mu) > \frac{\delta_\varepsilon}{3}$.

To conclude, taking $k_\varepsilon \geq \max(n_\varepsilon, \frac{9}{\delta_\varepsilon})$, we have for all $k \geq k_\varepsilon$: $w \in \mathbf{V}_k \Rightarrow f_k(\widehat{\delta_w}) \leq \frac{2}{k} + \frac{1}{2k} \leq \frac{\delta_\varepsilon}{3} \Rightarrow f(\widehat{\delta_w}) \leq \delta_\varepsilon \Rightarrow d_{\mathcal{M}}(\widehat{\delta_w}, \mathcal{V}) \leq \varepsilon$. ◇ Claim 3

Claim 4: *For every $\varepsilon > 0$, there exists a t_ε such that, for every $t' \geq t \geq t_\varepsilon$, if $w_n \in \mathbf{V}_{t+1}^k \backslash \mathbf{V}_t^k$ and $w \in \mathbf{V}_{t'+1}^{k'} \backslash \mathbf{V}_{t'}^{k'}$, then the path $u_0, \ldots u_l$ built in the corresponding step of the algorithm satisfies $\forall\nu \in \bigcup_{0\leq i<l} [\widehat{\delta_{u_i}}, \widehat{\delta_{u_{i+1}}}], d_{\mathcal{M}}(\nu, \mathcal{K}) \leq \varepsilon$.*

Proof. Let K_1 be large enough such that $b(i) \leq \frac{\varepsilon}{4}$ for any $i \geq K_1$ and $K_1 \geq k_{\frac{\varepsilon}{2}}$, and put $K_2 = k_{b(K_1)}$ as defined in the previous claim. Let $t_\varepsilon = \max_{0\leq i\leq K_2}(T_i)$ and assume $w_n \in \mathbf{V}_k^{t+1} \backslash \mathbf{V}_k^t$ and $w \in \mathbf{V}_{k'}^{t'+1} \backslash \mathbf{V}_{k'}^{t'}$ with $t' \geq t \geq t_\varepsilon$. Then $k \geq K_2$ and $k' \geq K_2$.

For each element $\mu \in \mathcal{K}$ there is an element $u_{K_1} \in \mathcal{A}^{\leq\alpha_{K_1}}$ such that $d_{\mathcal{M}}(\mu, \widehat{\delta_{u_{K_1}}}) < b(K_1)$, and therefore $f(\widehat{\delta_{u_{K_1}}}) < \frac{1}{K_1}$ so $u_{K_1} \in \mathbf{V}_{K_1}$. In other words, $\mathcal{K} \subset \bigcup_{u\in\mathbf{V}_{K_1}} \mathbf{B}\left(\widehat{\delta_u}, b(K_1)\right)$.

Since $w_n \in \mathbf{V}_k$ with $k \geq K_2 = k_{b(K_1)}, d_{\mathcal{M}}(\widehat{\delta_{w_n}}, \mathcal{K}) \leq b(K_1)$ and the same is true for w. Therefore $\bigcup_{u\in\mathbf{V}_{K_1}} \mathbf{B}\left(\widehat{\delta_u}, 2b(K_1)\right)$ contains $\widehat{\delta_{w_n}}$ and $\widehat{\delta_w}$ as well as $\mathcal{K}$ is a single connected component, since $\mathcal{K}$ is connected. This means that the i chosen in this step of the algorithm satisfies $i \geq K_1$. Since the path is entirely included in $\bigcup_{u\in\mathbf{V}_{K_1}} \mathbf{B}\left(\widehat{\delta_u}, 2b(i)\right)$ with $b(i) \leq \frac{\varepsilon}{4}$, and since $u \in \mathbf{V}_{K_1} \Rightarrow d(\widehat{\delta_u}, \mathcal{K}) \leq \frac{\varepsilon}{2}$, the result follows. ◇ Claim 4

Claim 5: *If $\mu \notin \mathcal{K}$, then $\mu \notin \mathcal{V}((w_n)_{n\in\mathbb{N}})$.*

Proof. This is a direct consequence of Claim 4. ◇ Claim 5

□

3. Construction of a cellular automaton realising a given set of measures

We want to prove a reciprocal to Proposition 2 and a partial reciprocal to Proposition 4 using Proposition 5. Given a computable sequence of words $(w_n)_{n\in\mathbb{N}}$ in $\mathcal{B}^*$, we construct a cellular automaton realising $\mathcal{V}((w_n)_{n\in\mathbb{N}})$ as its μ-limit measures set.

Theorem 1. *Let $(w_n)_{n\in\mathbb{N}}$ be a computable sequence of words of $\mathcal{B}^*$, where $\mathcal{B}$ is a finite alphabet. Then there is a finite alphabet $\mathcal{A}\supset\mathcal{B}$ and a cellular automaton $F:\mathcal{A}^{\mathbb{Z}}\to\mathcal{A}^{\mathbb{Z}}$ such that:*

- *for any measure $\mu\in\mathcal{M}^{full}_{\sigma-mix}(\mathcal{A}^{\mathbb{Z}})$, $\mathcal{V}(F,\mu)=\mathcal{V}((w_n)_{n\in\mathbb{N}})$.*
- *if $\mathcal{V}((w_n)_{n\in\mathbb{N}})=\{\nu\}$, then for any measure $\mu\in\mathcal{M}^{full}_{\sigma-erg}(\mathcal{A}^{\mathbb{Z}})$, $F^t_*\mu \underset{t\to\infty}{\longrightarrow} \nu$.*

Furthermore we get an explicit bound for the convergence rate in the first point of the theorem. If w_n is computable in space $S(n)$, assuming w.l.o.g. that $S(n)$ is an increasing sequence, define $S^{-1}(k)=\max\{n: S(n)\leq k\}$.

$$d_{\mathcal{M}}(F^t_*\mu,\mathcal{V}((w_n)_{n\in\mathbb{N}}))\leq O\left(\frac{1}{\log(t)}\right)+\sup\left\{d_{\mathcal{M}}(\nu,\mathcal{V}((w_n)_{n\in\mathbb{N}})):\nu\in\bigcup_{n\geq n(t)}[\widehat{\delta_{w_{S^{-1}(\sqrt{n})}}},\widehat{\delta_{w_{S^{-1}(\sqrt{n})+1}}}]\right\},$$

where $n(t)=\Theta(\log(t)^2)$. The first term of the upper bound corresponds to the intrisic limitations of the construction, the second term depends on the speed of convergence of the polygonal path defined by $\widehat{\delta_{w_n}}$, $n>n(t)$ to its limit $\mathcal{V}((w_n)_{n\in\mathbb{N}}$, when the sequence is “slowed down” by repeating elements so that computational space does not exceed $\sqrt{n}$.

In the rest of the section, we detail the construction of this cellular automaton and prove this theorem.

3.1. Sketch of the construction

In this section, we present a sketch of the construction of the alphabet $\mathcal{A}$ and the cellular automaton F. Our goal is to compute each w_n successively and write concatenated copies of it on the whole configuration to approach the measure $\widehat{\delta_{w_n}}$. $\mathcal{A}$ will contain a symbol $\boxed{\mathrm{W}}$ (for **wall**) persisting in time, except under special circumstances; w_n will be computed and then copied repeatedly, in each area between two subsequent walls, in an independant manner.

A main issue is to initialize the computation synchronously for each wall, even though we have no control over what cells appear at time 0. To do this, we define another symbol $\boxed{\mathrm{I}}$ (**init**), which appears only in the initial configuration, creating a wall while erasing the contents of neighboring cells and initializing different processes defined below. This process is detailed in Section 3.2.1 The resulting wall is said to be **initialized**.

Definition 7. Let $x\in\mathcal{A}^{\mathbb{Z}}$. $[i,j]$ is a **segment at time 0** if x_i and x_j are two consecutive $\boxed{\mathrm{I}}$ symbols, and a **segment at time** t if $F^t(x)_i$ and $F^t(x)_j$ are two consecutive initialized walls $\boxed{\mathrm{W}}$. Define the **length** of this segment as $i-j-1$.

Computation on each segment will be performed independantly. Apart from $\boxed{\mathrm{I}}$ and $\boxed{\mathrm{W}}$, the new alphabet $\mathcal{A}$ will be divided in different layers: the **main layer** where the words w_n will be output and recopied, and **auxiliary layers** where computation and other processes will take place. Since we have no control over the initial contents of each segment, we first want to erase non-initialized walls and anything on the auxiliary layers not issued from an $\boxed{\mathrm{I}}$ symbol (**sweeping** the segment), to guarantee that synchroneous computation takes place everywhere.

To do that, each initialized wall keeps on its left the value of the current time under the form of a binary counter incrementing at each step on one layer (**time counter** - see Section 3.2.3), and sends another incrementing counter to its right progressing at speed one on another layer (**sweeping counter** - see Section 3.2.4). Sweeping counters will sweep the segment as they progress, using the following method.

Time and sweeping counters already present in the initial configuration (not initialized) have a positive value at time 0, whereas those created by an $\boxed{\mathrm{I}}$ symbol (initialized) have value 0 at time 1, and they increment at the same rate. Thus, non initialized walls have older time counters, and by comparing time counters and sweeping counters as they cross, we can erase older counters and non-initialized walls as well. Figure 1 is an overview of those processes.

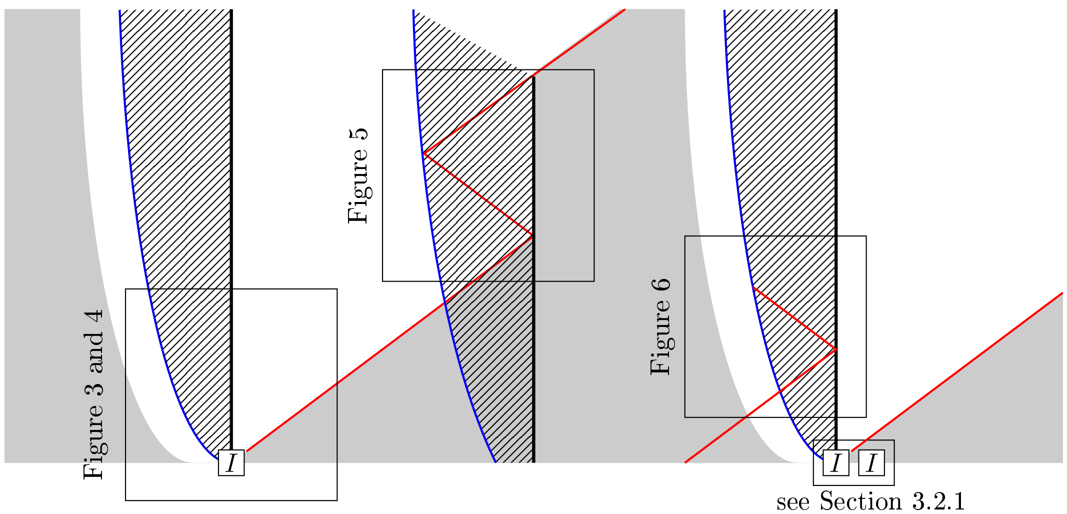

FIGURE 1. Sketch of the bootstrapping and sweeping processes. Vertical lines are walls. Dashed parts contain time counters (section 3.2.3) and Turing machines (section 3.3.2). Slanted lines are sweeping counters (section 3.2.4), and white areas and grey areas are swept and non-swept, respectively.

Meanwhile, a Turing machine is simulated on another layer in the space delimited by the time counter. This machine will successively compute each w_n (see Section 3.3.2) and copy it on the main layer of the segment to its left (see Section 3.3.3). For each w_n, this copy happens synchronously on the whole configuration, at some time T_n that we will fix later. At the same time T_n, segments of length n are merged with their left neighbour in order to enlarge computational space and decrease the density of cells with nonempty auxiliary layers (see Section 3.4). To determine the length of its right segment, each wall sends on a dedicated layer a signal to the right that bounces off the next wall and counts the return time. The Figure 2 is an overview of copy and merging processes.

Thus the enlarged alphabet can be written as $\mathcal{A} = \{\boxed{\mathrm{I}}, \boxed{\mathrm{W}}\} \cup \mathcal{A}_{\texttt{main}} \times \mathcal{A}_{\texttt{comp}} \times \mathcal{A}_{\texttt{time}} \times \mathcal{A}_{\texttt{sweeping}} \times \mathcal{A}_{\texttt{copy}} \times \mathcal{A}_{\texttt{merge}}$, where:

- $\boxed{\mathrm{I}}$ and $\boxed{\mathrm{W}}$ are the two above-mentioned symbols;
- $\mathcal{A}_{\texttt{main}} = \mathcal{B} \cup \{\#\}$ is the layer on which w_n is output and then recopied;
- $\mathcal{A}_{\texttt{comp}}$ is the layer where computing takes place by simulating Turing machines;
- $\mathcal{A}_{\texttt{time}}$ is the layer on which time counters are incremented;
- $\mathcal{A}_{\texttt{sweeping}}$ is the layer on which sweeping counters move and are incremented, and where comparisons are done;
- $\mathcal{A}_{\texttt{copy}}$ is an auxiliary layer used in the process of writing copies of the output on the main layer;
- $\mathcal{A}_{\texttt{merge}}$ is an auxiliary layer used in the process of merging two segments.

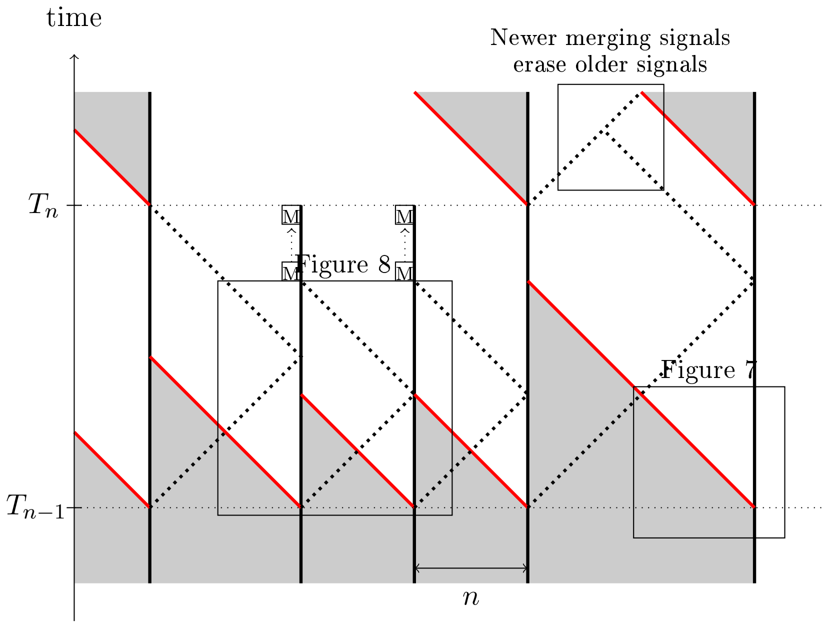

FIGURE 2. Sketch of the copying and merging processes. We suppose all walls are initialized. Slanted thick lines are copy processes (see Section 3.3.3), slanted dotted lines are merging signals (see Section 3.4).

All those alphabets contain a symbol # (blank) representing the absence of information. If $u \in \mathcal{A}$, note $main(u)$, resp. $comp(u)$, $time(u)$... the projections on each layer (the result being # on $\boxed{\mathrm{I}}$ and $\boxed{\mathrm{W}}$). We have $\mathcal{B} \subset \mathcal{A}$ up to the identification $b \mapsto (b, \#, \#, \#, \#, \#)$.

We shall detail the different alphabets in the following sections. As we will see, our construction needs interactions at a distance at most three, so we can take $\mathbb{U}_F = \{-3, \ldots, 3\}$ as the neighbourhood of $\overline{F}$.

3.2. Formatting the segments

3.2.1. Bootstrapping

If two symbols $\boxed{\mathrm{I}}$ are separated by two cells or less, the rightmost one is destroyed. Otherwise, any $\boxed{\mathrm{I}}$ symbol turns into a $\boxed{\mathrm{W}}$, erasing the contents of three cells to its right and left (including walls), initializing on its left a computation and a time counter, and on its right a sweeping counter. No more $\boxed{\mathrm{I}}$ or $\boxed{\mathrm{W}}$ symbols can be created.

Walls, counters and computing areas created in this way are **initialized**, by opposition to those already present at time 0. Walls persist over time and are only destroyed under two circumstances:

- when it is in a situation such that it is impossible that it is initialized (e.g. without a time counter to its left);
- at time T_n, if it is the left bound of a segment of length n.

If a segment is of length three at time 0, then the time counter of the rightmost wall is erased at time 1 and the wall itself is destroyed at time 2. Thus segments have minimum length four from time 2 onwards.

3.2.2. **Counters**

All counters are binary in a redundant basis, so that they can be incremented by one at each step (keeping track of current time) in a local manner.

Definition 8 (Redundant binary)**.** Let $u = u_{n-1} \dots u_0 \in \{0,1,2\}^*$. The **value** of u is

$$val(u) = \sum_{i=1}^{n} u_i 2^i.$$

Since the basis is redundant, different counters can have the same value.

Definition 9 (Incrementation)**.** The incrementation operation $inc : \{0,1,2\}^* \mapsto \{0,1,2\}^*$ is defined in the following way. If $u_{|u|-1} = 2$, then $|inc(u)| = |u| + 1$, $|u|$ otherwise, and:

$$inc(u)_i = \begin{cases} 1 & \text{si } i = |u| + 1 \text{ and } u_{|u|-1} = 2; \\ u_i \mod 2 + 1 & \text{if } i = 0 \text{ or } u_{i-1} = 2; \\ u_i \mod 2 & \text{otherwise.} \end{cases}$$

Intuitively, the counter is increased by one at the rightmost bit and 2 behaves as a carry propagating along the counter. When the leftmost bit is a carry, the length of the counter is increased by one. Thus:

Fact 1. $val(inc(u)) = val(u) + 1$.

This operation is defined locally and can be seen as the local rule of a cellular automaton.

3.2.3. **Time**

We use the alphabet $\mathcal{A}_{\texttt{time}} = \{0, 1, 2, \#\}$. In a configuration, a time counter is a word of maximal length containing no # in the time layer. A time counter is **attached** if it is bounded on its right by a wall $\boxed{W}$, **detached** otherwise.

#	#	#	#	#	#	#	1	0	2	W
#	#	#	#	#	#	#	#	2	1	W
#	#	#	#	#	#	#	#	1	2	W
#	1	#	#	#	#	#	#	1	1	W
#	1	0	#	#	#	#	#	#	2	W
#	1	0	0	#	#	#	#	#	1	W
#	#	2	0	0	#	?	#	#	0	W
?	#	1	2	0	1	#	?	?	?	I

FIGURE 3. A detached time counter, and a time counter attached to an initialized wall. Only the time layer is represented. ? cells have arbitrary values.

At each step, attached counters are incremented by one while detached counters have their rightmost bit deleted (see Figure 3). Indeed, detached counters cannot be initialized and can be safely deleted. Formally,

- if $u_1 = \boxed{\mathrm{W}}$, then $time(F(u)_0) = time(u_0) \mod 2 + 1$;
- if $time(u_1) = \#$, then $time(F(u)_0) = \#$;
- otherwise, follow the incrementation definition (Definition 9).

When a counter increases in length, it can erase a wall. However, this is not a problem, as we shall see in Facts 2 and 6.

Fact 2. *An initialized wall cannot be erased by a detached time counter.*

Proof. A detached counter is not incremented and can extend by one cell at most because of the carries initially present in the word. But $\boxed{\mathrm{I}}$ symbols erase two cells to their right at initialization. □

Fact 3. *Each attached time counter u in $F^t(a)$ satisfies $val(u) \geq t-1$, the equality being attained if this counter is attached to an initialized wall.*

Proof. No time counter is created except at $t = 1$ (by $\boxed{\mathrm{I}}$). Therefore such a counter was present either in the initial configuration (with a nonnegative value), or was created at $t = 1$ by a $\boxed{\mathrm{I}}$ symbol. It is incremented by one at each step in both cases. □

Thus we can use time counters to tell apart initialized walls from non-initialized walls, which will be the object of the next section.

3.2.4. **Sweeping and comparisons**

Sweeping counters are defined and incremented at each step in a similar way as time counters, but they have a range of different behaviors. The sweeping layer is decomposed into two layers $\mathcal{A}_{\texttt{state}}$ and $\mathcal{A}_{\texttt{value}}$. A sweeping counter is a word of maximal length of state different than #. The possible states of the counter are:

"Go" state: The counter progresses at speed one to the right.

"Stop" state: Once a wall is encountered, the counter progressively (right to left) stops.

Comparison states: Once the whole counter has stopped, we locally compare the sweeping counter and the time counter, left to right, with a method we will describe later.

The wall is destroyed if the sweeping counter is strictly younger, and the sweeping counter is destroyed otherwise (see Figures 5 and 6). In the former case, the counter progressively returns to the "Go" state.

Changing state takes some time to propagate the information along the counter. Therefore, counters passing from a "Go" state to a "Stop" state are temporarily in a situation where the left part of the counter progresses whereas the right part does not. To avoid erasing information, counters in a "Go" state have **buffers**, i.e. the value of the counter is only written on half the cells, the other being erased when changing state (see Figure 4).

When its length increase, a counter will never merge with another counter, instead erasing bits from the right-hand counter to avoid merging: we say the right-hand counter is **dominated**. Notice that it is impossible for a counter located at the right of another counter to be initialized, and so it is safe to erase bits of it.

Fact 4. *Any non-dominated sweeping counter u of $F^t(x)$ satisfies $val(u) \geq t-1$, the equality being attained if the counter is initialized (issued from a $\boxed{I}$ symbol).*

Proof. Similar to Fact 3. □

Thus, we guarantee that an initialized (hence non-dominated) sweeping counter is strictly younger than any non-initialized wall, and symmetrically. As for dominated counters, whose value is arbitrary, we will see that they are erased before any comparison takes place.

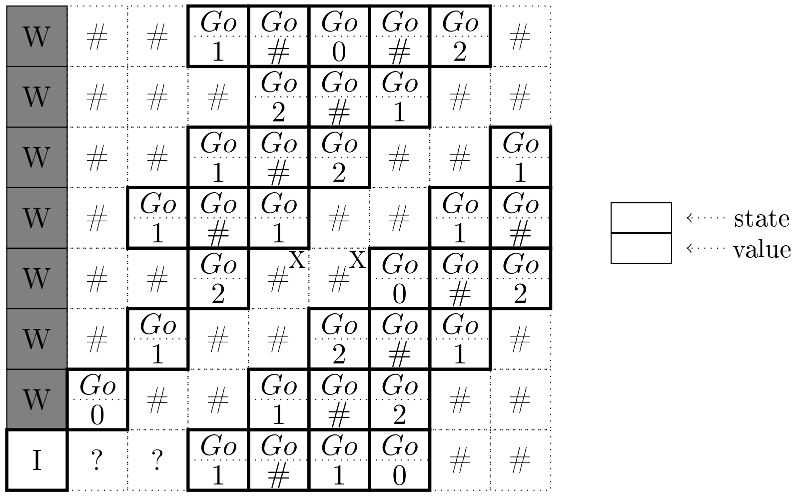

FIGURE 4. One initialized and one uninitialized sweeping counter. X symbols mark the cells where values are prevented to appear to avoid merging: the right counter is dominated. Only the sweeping layer is represented.

Definition 10 (Comparison method)**.** Let $u = u_0u_1\dots$ and $v = v_0v_1\dots$ be two counters in redundant binary basis (adding zeroes so that $|u| = |v|$). Let us note $sign(u - v)$ the result of the comparison between u and v, that is, $+, 0$ or $-$.

Case 1: if $|u| = |v| = 1$, $sign(u - v) = sign(u_0 - v_0)$;
Case 2: if $u_0 + \lfloor u_1/2 \rfloor > v_0 + \lfloor v_1/2 \rfloor$, then $sign(u - v) = +$, and symmetrically;
Case 3: if $u_0 + \lfloor u_1/2 \rfloor = v_0 + \lfloor v_1/2 \rfloor$, then $sign(u - v) = sign(u'_1 u_2 \dots - v'_1 v_2 \dots)$, where $u'_1 = u_1 \mod 2$ and $v'_1 = v_1 \mod 2$.

In other words, we do a bit-by-bit comparison starting from the leftmost bit, considering that $\# = 0$, and taking into account the carry propagation "in advance", so that the incrementation and carry propagation can continue during the comparison. If the result can be determined locally (cases 1 and 2), the state is changed to $+$ or $-$, and it will propagate to the right along the counter. Otherwise (case 3), the state changes to $=$, which means future bit comparisons will decide the result in the same way (see Figure 6).

After the comparison, two cases are possible:

- if the state of the rightmost bit is $-$, the wall is destroyed and the state of the rightmost bit becomes "Go". The counter then progressively returns to the "Go" state.
- if the state of the rightmost bit is $+$ or $=$, it is erased. The remaining bits are progressively erased similarly to detached time counters.

Notice that if the counter is dominated, then its leftmost bit is erased at each step, preventing the comparison to start, until the counter is entirely erased.

Finally, we have $\mathcal{A}_{\texttt{sweeping}} = \{\#\} \cup \{Go\} \times \{0, 1, 2, \#\} \cup \{Stop, +, -, =\} \times \{0, 1, 2\}$.
When a sweeping counter reaches the right wall of the segment, the segment is said to be **swept**. This implies all walls and auxiliary states remaining in the segment are initialized.

Fact 5. *At time $k(1 + \lceil \log k \rceil)$, all segments of length k are swept.*

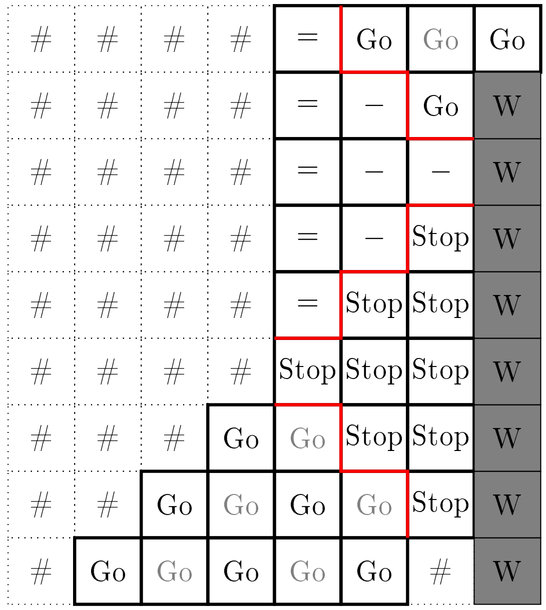

FIGURE 5. A younger sweeping counter encountering an older wall. Only the state layer of $\mathcal{A}_{\texttt{sweeping}}$ is represented, with greyed words for buffers.

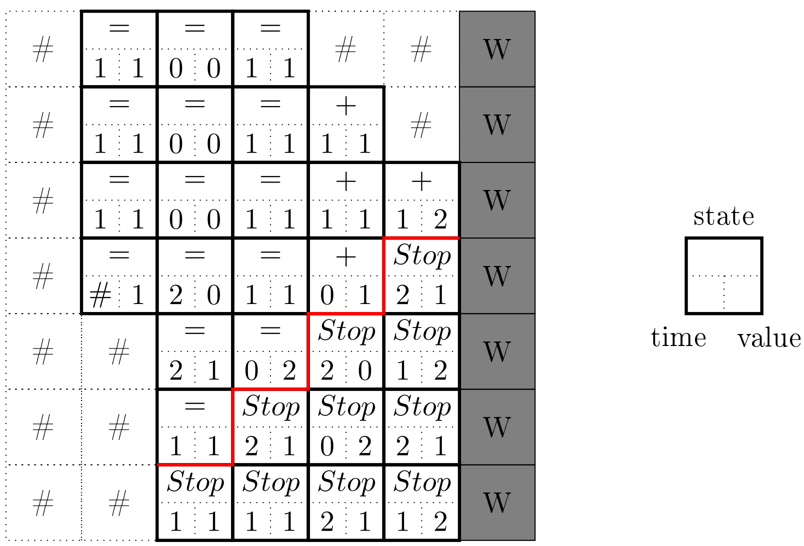

FIGURE 6. The comparison process in detail. Here the sweeping counter is older than the wall and is destroyed. Only the layer $\mathcal{A}_{\texttt{sweeping}}$ is represented.

Proof. As long as $t \leq k(1 + \lceil \log k \rceil)$, any initialized sweeping counter has less than $2\lceil \log k \rceil$ cells containing a value. The counter progresses at speed one except when it meets another wall. Each comparison takes a time equal to twice the current length of the counter. Furthemore, two consecutive walls are separated by three cells at least (cf. Section 3.2.1). Thus, the segment is swept in less than $k + \frac{k}{4} \cdot 2 \cdot 2\lceil \log k \rceil$ steps, and we can check that $t \leq k(1 + \lceil \log k \rceil)$. $\square$

Fact 6. *An initialized wall cannot be erased by a time counter attached to a non-initialized wall.*

Proof. Consider two walls separated by $k \geq 3$ cells, the left being initialized and the right non-initialized. The value of the time counter attached to the right wall cannot exceed 2^{k-3} at $t = 1$ (since $\boxed{\mathrm{I}}$ erases three cells to its right), it will take more than $2^k - 2^{k-3}$ steps before the left wall is erased. According to Fact 5, the right wall will be destroyed in less than $k(1 + \lceil \log k \rceil)$ steps, and the time counter will take at most k more steps to be erased.

For $k \geq 5$, $k(1 + \log k) + k \leq 2^k - 2^{k-3}$, so the counter is erased before it reaches the left wall. For $k = 4$, there cannot be another wall between them, so the destruction time is actually less than $k + 2\log k + k \leq 2^k - 2^{k-3}$. For $k = 3$, the left $\boxed{\mathrm{I}}$ symbol present at time 0 erases the contents of the three cells, which includes the time counter of the right wall. As explained in Section 3.2.1, the right wall is then immediately destroyed. □

3.3. Computation and copy

3.3.1. Simulating a Turing machine in a cellular automaton

Let $\mathcal{TM} = (Q, \Gamma, \#, q_0, \delta, Q_F)$ be a Turing machine. We will show how to simulate this machine in a cellular automaton F on the alphabet $\Gamma \times (Q \cup \#)$. The left part contains the content of the tape, and the right part contains the state of the machine when the head is located on this cell, and # everywhere else.

The local rule of F is governed by the rules of the machine, i.e., for all $u \in \mathcal{A}^{\mathbb{Z}}$:

- if the head is on u_0 and $\delta(u_0) = (q, \gamma, \cdot)$, then $F(u)_0 = (q, \gamma)$;
- if the head is on u_1, $\delta(u_1) = (q, \gamma, \leftarrow)$ and $u_0 = (\#, \gamma')$, then $F(u)_1 = (\#, \gamma)$ and $F(u)_0 = (q, \gamma')$;
- similarly if the head is on u_{-1} and $\delta(u_{-1}) = (q, \gamma, \rightarrow)$;
- otherwise, $F(u)_0 = u_0$.

When starting from a configuration filled with # everywhere except for a finite window with only one head, the time evolution of the cellular automaton matches the one of the Turing machine. When the machine has stopped (the state being in Q_F), the local rule is the identity function.

3.3.2. Computation

Computation takes place in the area delimited by the time counter attached to the right wall. $\mathcal{A}_{\texttt{comp}}$ is divided into three layers, on which three Turing machines are simulated. We use the alphabet $\mathcal{A}_{\texttt{comp}} = \bigotimes_{i=1}^{3} \Gamma_i \times (Q_i \cup \#)$. Compared with the previous subsection, the Turing machines have access to a limited space delimited by the time counter, and can read input from or write output to another layer (when indicated).

We now describe the operations expected to be performed during the time interval $[T_{n-1}, T_n]$. At time T_{n-1}, n is already written on the layer 1. The machines:

- replace n by $n+1$ on layer 1 and stops;
- compute w_n on layer 2, outputting it on the main layer, and stops;
- compute T_n on layer 3, and stops;
- when $t = T_n$ (t being read from the time layer), the copying process triggers and the next computation starts, except when merging occurs; see next subsections.

All these operations must be performed in less than $T_n - T_{n-1}$ steps.

First we suppose that each w_n can be computed in space $\sqrt{n}$, defining if necessary a new sequence where each w_n is repeated as long as there is not enough space to compute w_{n+1}. Now fix $T_n - T_{n-1} = q^{\lfloor \sqrt{n} \rfloor}$, taking the smallest q such that $q^{\lfloor \sqrt{n} \rfloor} \geq \mathrm{Card}(\Gamma_2)^{\sqrt{n}} \times \sqrt{n} \times \mathrm{Card}(Q_2)$ where Γ_2 and Q_2 correspond to the Turing machine of layer 2. Indeed, this is the maximum time needed for any computation using only space $\sqrt{n}$ and these alphabets.

Moreover, at time T_{n-1} the time counter is longer than $\log_2(T_{n-1}) - 1 \geq \sqrt{n}$ for $q \geq 5$. For layers 1 and 3, the time and space bounds are verified asymptotically, i.e. there are machines satisfying these bounds for $n > N$. Let t_N be the maximal time necessary for those machines to perform those operations when $n < N$; we can fix $T_{n+1} - T_n = t_N$ when $n < N$, which has no influence on the asymptotic behavior of T_n and ensures that the machines satisfies the time bound for any n. For the space bound, it is always possible to compress the space by a constant factor (by grouping tape cells) so that the space bound is satisfied for $n < N$, with no impact on the computing.

Remark. We fix T_n to have a computation space of size $\sqrt{n}$ at time T_n, so that it constitutes an asymptotically negligible fraction of its segment. We could choose instead of $\sqrt{n}$ any other easily computable function which is $o(n)$.

3.3.3. Copying

At time T_{n+1} ($n \geq 0$), w_n has been output on the main layer. If the segment is not merging with its right segment, the Turing machine triggers the copying process by copying the rightmost letter of w_n from the main layer to the copy layer.

First phase: Inside the time counter, the word on the copy layer progresses at speed -2, and a letter at each step is copied from the main layer to the tail of the word;

Second phase: When the head is out of the time counter, the word keeps progressing at speed -2 but the head loses one letter at each step and copies it on the main layer. The tail keeps copying letters from the main layer.

Intuitively, the cellular automaton performs a catterpillar-like movement between the copy and main layers (see Figure 7 for an example). The process ends when it meets a wall or a sweeping counter to its left. Thus, $\mathcal{A}_{\texttt{copy}} = \mathcal{B} \cup \{\#\}$.

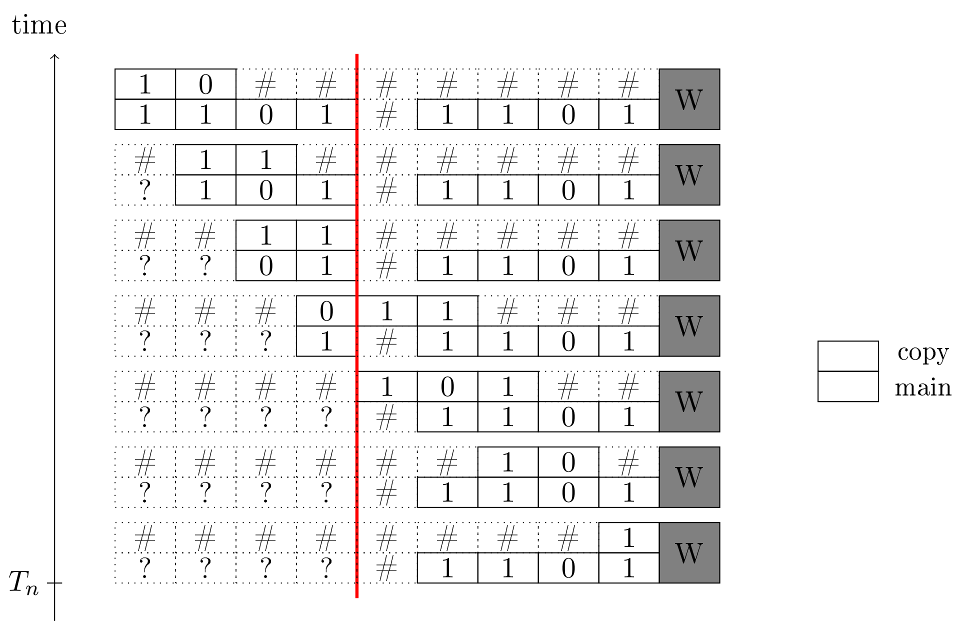

FIGURE 7. Beginning of the copying process, with $w_n = 1101$. Only the layers $\mathcal{A}_{\texttt{copy}}$ and $\mathcal{A}_{\texttt{main}}$ are represented. The thick line is the limit of the time counter.

3.4. Merging of segments

At time T_n, all segments of length n merge with their left neighbor, so that the density of walls tend to 0. To determine the length of each segment, a signal is sent to the right and bounces off the right wall, and its return time is measured.

To do so, a **merging counter** of value $2n$ is initilized at time T_{n-1} on the merge layer. The value of n is copied from the first computing layer to the merge layer (with an additional 0 at the end), using an auxiliary state $\boxed{\mathrm{C}}$ (**copy**). This counter is decrementing at each step, similarly to incrementing counters except it uses -1 as negative carry.

If the signal returns at the end of the decrementation, a symbol $\boxed{\mathrm{M}}$ (**merge**) is created on the merge layer, to indicate the wall will be destroyed at next T_n; otherwise, the output will be copied in the main layer as described above. Thus $\mathcal{A}_{\texttt{merge}} = \{-1, 0, 1, \boxed{\mathrm{M}}, \boxed{\mathrm{C}}\} \times \{\rightarrow, \leftarrow\} \cup \{\#\}$, see Figure 8 for an example of this process.

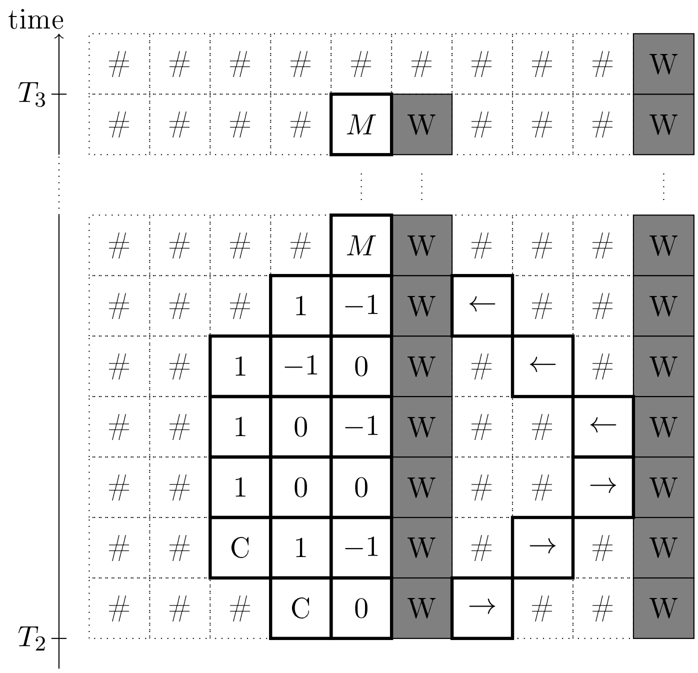

FIGURE 8. Determination of the length of the segment. Here the right segment is of length 3 and the wall merges at time T_3. The counter of the right segment has been omitted for clarity.

Fact 7. *All left walls of segments of length k are erased simultaneously at time* $\min(T_k, 2^k + k)$.

Proof. Except for the situation described above, the only other way for an initialized wall to be erased is a time counter attached to an initialized wall, see Facts 2 and 6. A redundant binary counter whose initial value is 0 reaches length k at time $2^k + k$ (carry propagation).

□

We will consider from now on that n is large enough so that $2^n + n > T_n$.

3.5. Correctness of the cellular automaton

The operations described in the previous section have to be performed between time T_n and time T_{n+1} with high probability, which requires that the segments are not too large. In this section, we control the length of segments at time T_n.

Proposition 6. $T_n = \Theta(\lfloor\sqrt{n}\rfloor q^{\lfloor\sqrt{n}\rfloor})$ *where q is defined in Section 3.3.2.*

Proof. $T_n = \sum_{k=1}^{n} T_k - T_{k-1}$. Since asymptotically $T_{k+1} - T_k = q^{\lfloor\sqrt{k}\rfloor}$, and:

$$(2\lfloor\sqrt{n}\rfloor - 1)q^{\lfloor\sqrt{n}\rfloor-1} \leq \sum_{k=1}^{\lfloor\sqrt{n}\rfloor-1}(2k+1)q^k \leq \sum_{k=1}^{n} q^{\lfloor\sqrt{k}\rfloor} \leq \sum_{k=1}^{\lfloor\sqrt{n}\rfloor}(2k+1)q^k \leq (2\lfloor\sqrt{n}\rfloor + 1)q^{\lfloor\sqrt{n}\rfloor+1}.$$

the proposition follows. ☐

3.5.1. **Acceptable segments**

Definition 11. Denote

$$\Gamma^t_{l,k} = \left\{x \in \mathcal{A}^{\mathbb{Z}} : [0,l] \text{ is included in a segment of } F^t(x) \text{ of length } k\right\}$$

$$\Gamma^t_{l,\geq k} = \bigcup_{i\geq k} \Gamma^t_{l,i} \quad \text{and} \quad \Gamma^t_l = \Gamma^t_{l,\geq 1}$$

Proposition 7 (Lower bound)**.** *Let $\mu \in \mathcal{M}^{full}_{\sigma-erg}(\mathcal{A}^{\mathbb{Z}})$. For all $l \in \mathbb{N}$, one has $\mu(\Gamma^{T_n}_{l,\geq n}) \underset{n\to\infty}{\longrightarrow} 1$.*

Proof. At time T_n, no configuration can contain a segment smaller than n. Since μ has full support, $\mu\left([\boxed{\mathrm{I}}] \bigcap_{i\in[1,n]} \sigma^i\left(\overline{[\boxed{\mathrm{I}}]}\right)\right) \neq 0$. By σ-ergodicity, these segments of length larger than n exist for μ-almost all configurations at $t = 0$, and those segments survive up to time T_n by construction.

Therefore, the cell 0 is μ-almost surely included in a segment at time T_n. Since this segment has length larger than n and by σ-invariance, the probability that $[0,l]$ crosses a border of the segment tends to 0 as n tends to infinity. ☐

Definition 12. Let $x \in \mathcal{A}^{\mathbb{Z}}$, $[i,j]$ a segment at time $t \in [T_n, T_{n+1}]$. It is **acceptable** if $j - i - 1 \leq K_n = \sqrt{T_{n+1} - T_n}$. For n large enough $K_n = q^{\frac{\lfloor\sqrt{n}\rfloor}{2}}$.

Proposition 8 (Upper bound)**.** *Let $\mu \in \mathcal{M}^{full}_{\sigma-mix}(\mathcal{A}^{\mathbb{Z}})$. One has $\mu(\Gamma^{T_n}_{l,\geq K_n}) \underset{n\to\infty}{\longrightarrow} 0$, that is to say:*

$$\mu(\{x \in \mathcal{A}^{\mathbb{Z}} : [0,l] \textit{ is in an acceptable segment of } F^t(x)\}) \underset{t\to\infty}{\longrightarrow} 1$$

and the rate of convergence is exponential.

Proof. Any segment at time T_n corresponds, at time T_{n-1}, to a segment of arbitrary size plus an arbitrary number of segments of size n (see Figure 8 for an illustration of this decomposition). For $l \leq n$, define

$$\Delta^t_{n,\alpha} = \{x \in \mathcal{A}^{\mathbb{Z}} : \text{starting from 0 there is a strip of } \alpha \text{ consecutive segments of size } n \text{ in } F^t(x)\}.$$

Suppose $[0,l]$ is included in a segment longer than k at time T_n. Take $L > 2n$ and distinguish the two following cases:

- There were less than $\lfloor\frac{L}{n}\rfloor$ segments of length n: then the other segment is larger than $k - L$. By shifting the configuration by $L - l$ cells at most, we can ensure that $[0,l]$ is included in this segment. at time T_{n-1}.
- There were more than $\lfloor\frac{L}{n}\rfloor$ segments of length n. Therefore there is a strip of $\lfloor\frac{L}{n}\rfloor$ segments of length n starting somewhere in $[-k,k]$.

In other words,

$$\Gamma^{T_n}_{l,\geq k} \subset \bigcup_{i=-L+l}^{0} \sigma^i\left(\Gamma^{T_{n-1}}_{l,\geq k-L}\right) \cup \bigcup_{j=-k+1}^{k-1} \sigma^j\left(\Delta^{T_{n-1}}_{n,\lfloor\frac{L}{n}\rfloor}\right)$$

$$\mu\left(\Gamma^{T_n}_{l,\geq k}\right) \leq L\mu\left(\Gamma^{T_{n-1}}_{l,\geq k-L}\right) + 2k\mu\left(\Delta^{T_{n-1}}_{n,\lfloor\frac{L}{n}\rfloor}\right) \tag{1}$$

Thus we try to bound the value of $\mu(\Delta^t_{n,\alpha})$. If $x \in \Delta^t_{n,\alpha}$ then for all $i \in [0,\alpha]$ one has $x_{in} = \boxed{\mathrm{I}}$ (corresponding to initialized walls at time t). For any $m > 0$, by considering one symbol out of every m:

$$\begin{aligned}
\mu\left(\Delta^t_{n,\alpha}\right) &\leq \mu\left(\bigcap_{i\in[0,\alpha]} \sigma^{in}\left([\boxed{\mathrm{I}}]\right)\right) \\
&\leq \mu\left(\bigcap_{i\in\left[0,\lfloor\frac{\alpha}{m}\rfloor\right]} \sigma^{in\cdot m}\left([\boxed{\mathrm{I}}]\right)\right) \\
&\leq (1+\psi_\mu(mn))^{\lfloor\frac{\alpha}{m}\rfloor}\mu\left([\boxed{\mathrm{I}}]\right)^{\lfloor\frac{\alpha}{m}\rfloor+1}. \qquad (2)
\end{aligned}$$

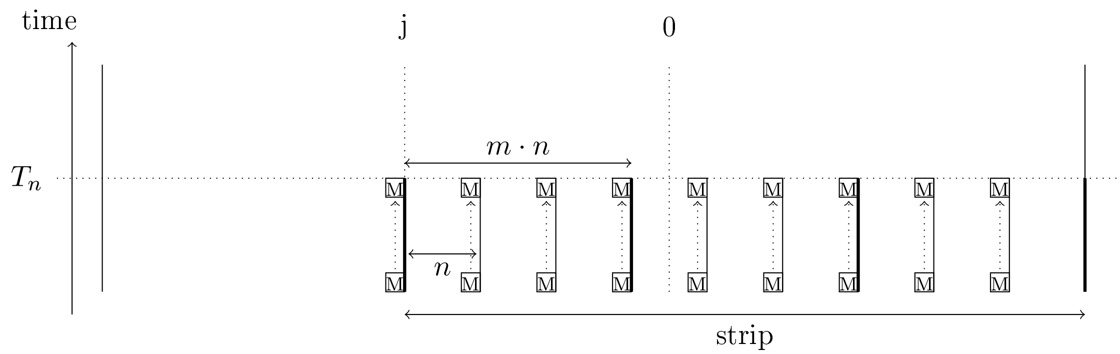

FIGURE 9. Illustration of the proof of Proposition 8 with $\alpha = 9$ and $m = 3$.

Now take any $M > n$. Using (2) with $m = \lceil \frac{M}{n} \rceil$ inside equation (1):

$$\begin{aligned}
\mu\left(\Gamma^{T_n}_{l,\geq k}\right) &\leq L\mu\left(\Gamma^{T_{n-1}}_{l,\geq k-L}\right) + 2k\left[1+\psi_\mu\left(n\cdot\left\lceil\frac{M}{n}\right\rceil\right)\right]^{\frac{L}{M}}\mu\left([\boxed{\mathrm{I}}]\right)^{\frac{L}{M}+1} \\
&\leq L\mu\left(\Gamma^{T_{n-1}}_{l,\geq k-L}\right) + 2k\left[(1+\psi_\mu(M))\mu\left([\boxed{\mathrm{I}}]\right)\right]^{\frac{L}{M}}
\end{aligned}$$

Now, if $k \geq nL$, we obtain by induction:

$$\mu\left(\Gamma^{T_n}_{l,\geq k}\right) \leq L^n\mu\left(\Gamma^0_{l,\geq k-nL}\right) + 2kL^n\left[(1+\psi_\mu(M))\mu\left([\boxed{\mathrm{I}}]\right)\right]^{\frac{L}{M}} \qquad (3)$$

For the left-hand term, we have:

$$\begin{aligned}
\mu\left(\Gamma^0_{l,\geq k-nL}(x)\right) &\leq \mu\left(\bigcup_{j\in[-k+nL,-1]}\ \bigcap_{i\in[0,k-nL]} \sigma^{j+i}\left(\overline{[\boxed{\mathrm{I}}]}\right)\right) \\
&\leq \mu\left(\bigcup_{j\in[-k+nL,-1]}\ \bigcap_{i\in[0,\lfloor\frac{k-nL}{n}\rfloor]} \sigma^{j+in}\left(\overline{[\boxed{\mathrm{I}}]}\right)\right) \\
&\leq (k-nL)(1+\psi_\mu(n))^{\lfloor\frac{k-nL}{n}\rfloor}\mu\left(\overline{[\boxed{\mathrm{I}}]}\right)^{\lfloor\frac{k-nL}{n}\rfloor+1}
\end{aligned}$$

the second line being obtained by considering one symbol out of every n. Putting $M = n$, $L = n^2\sqrt{n}$, and $k = K_n$ in (3) then since $\psi_\mu(n) \to 0$, we have $\mu(\Gamma^{T_n}_{\geq K_n}) \underset{n\to\infty}{\longrightarrow} 0$ and the rate of convergence is exponential. $\square$

Remark. Remark that it is possible to take any value for K_n as soon as $K_n = \omega(n^2\sqrt{n})$.

3.5.2. **Density of auxiliary states**

Proposition 9. *For t large enough, an acceptable segment is swept.*

Proof. When $T_n \leq t < T_{n+1}$, for an acceptable segment of length k, we have $k(1 + \log k) \leq K_n(1 + \log(K_n)) = o(T_n)$ by Proposition 6. Taking n large enough, we conclude by Fact 5. $\square$

Proposition 10. *Let $\mu \in \mathcal{M}^{full}_{\sigma-erg}(\mathcal{A}^{\mathbb{Z}})$ and $u \in \mathcal{B}^l$ for some fixed l. For a given segment length k such that $n + 1 \leq k \leq K_n$ one has:*

- *If $t \in [T_n + k, T_{n+1}]$,*
$$\left|\mu\left(F^{-t}([u])|\Gamma^{T_n}_{l,k}\right) - \widehat{\delta_{w_n}}([u])\right| = O\left(\frac{1}{\sqrt{n}}\right);$$
- *If $t \in [T_n, T_n + k]$ one has*
$$\left|\mu\left(F^{-t}([u])|\Gamma^{T_n}_{l,k}\right) - \left(\frac{k-(t-T_n)}{k}\widehat{\delta_{w_{n-1}}}([u]) + \frac{t-T_n}{k}\widehat{\delta_{w_n}}([u])\right)\right| = O\left(\frac{1}{\sqrt{n}}\right).$$

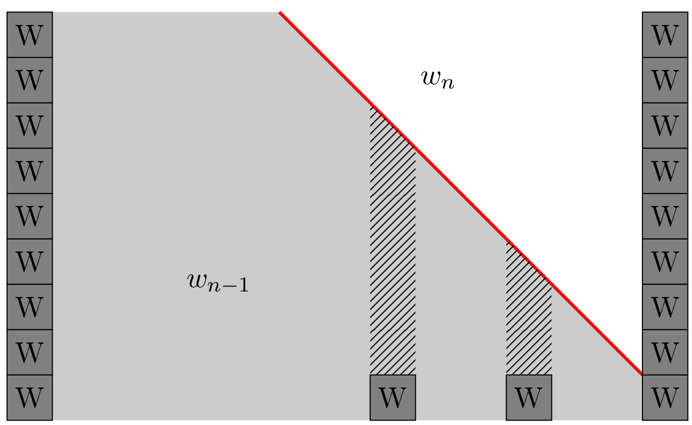

FIGURE 10. Illustration of Proposition 10. The output is not correctly written in dashed areas because of the destruction of a wall.

Proof. We write $\Gamma^{T_n}_{[i,i+k]} = \left\{x \in \mathcal{A}^{\mathbb{Z}} \mid [i, i+k] \text{ is a segment at time } T_n\right\}$, so that

$$\Gamma^{T_n}_{l,k} = \bigsqcup_{i=-k+l}^{-1} \Gamma^{T_n}_{[i,i+k+1]} = \bigsqcup_{i=-k+l}^{-1} \sigma^i(\Gamma^{T_n}_{[-1,k]}) \qquad \text{(disjoint union)}.$$

Suppose $x \in \Gamma^{T_n}_{[-1,k]}$. Since such a segment is acceptable, it is swept, and any non-initialized counter or wall has been destroyed. Since $|w_n| \leq \sqrt{n}$ (smaller than the computing space), the copying process will use less than $\sqrt{n}$ auxiliary cells.

First point: The tail of the copying process progresses at speed one, so at time $T_n + k$ the copy of the word is finished (since $T_n + k \leq T_{n+1}$), and the segment is constituted only by copies of w_n except for the time counter and computation area ($O(\sqrt{n})$ cells) and a merging signal (one cell).

Therefore for all $x \in \Gamma^{T_n}_{[-1,k]}$, one has $\left|\mathrm{Freq}(u, F^t(x)_{[0,k-1]}) - \widehat{\delta_{w_n}}([u])\right| = \frac{O(\sqrt{n})}{k} = O\left(\frac{1}{\sqrt{n}}\right)$, taking into account the last copy of w_n in the segment which can be incomplete ($|w_n| \leq \sqrt{n}$), and since $k \geq n$. Thus we have

$$\left|\frac{1}{k}\sum_{i=0}^{k-1}\mu\left(F^{-t}([u]_i) \mid \Gamma^{T_n}_{[-1,k]}\right) - \widehat{\delta_{w_n}}([u])\right| = O\left(\frac{1}{\sqrt{n}}\right).$$

Since μ is σ-invariant, $\mu\left(F^{-t}([u]_i) \mid \Gamma^{T_n}_{[-1,k]}\right) = \mu\left(F^{-t}([u]_0) \mid \Gamma^{T_n}_{[i,i+k+1]}\right)$. So:

$$\begin{aligned}\mu\left(F^{-t}([u]_0) \mid \Gamma^{T_n}_{l,k}\right) &= \sum_{i=-k+l}^{-1}\mu\left(F^{-t}([u]_0) \mid \Gamma^{T_n}_{[i,i+k+1]}\right)\cdot\mu\left(\Gamma^{T_n}_{[i,i+k+1]} \mid \Gamma^{T_n}_{l,k}\right)\\ &= \frac{1}{k-l}\sum_{i=-k+l}^{-1}\mu\left(F^{-t}([u]_0) \mid \Gamma^{T_n}_{[i,i+k+1]}\right)\end{aligned}$$

by σ-invariance and disjoint union of $\Gamma^{T_n}_{l,k}$. The result follows.

Second point: When $t \in [T_n, T_n + k]$, the copy is still taking place, with $t - T_n$ cells containing copies of w_n and the rest containing copies of w_{n-1}, except for: the computation part, the copy auxiliary states, the merging signal, and possibly defects when a wall has been destroyed at time T_n (there are at most $\frac{k}{n}$ of them). Therefore

$$\left|\mathrm{Freq}(u, F^t(x)_{[0,k-1]}) - \left(\frac{k-(t-T_n)}{k}\widehat{\delta_{w_{n-1}}}([u]) + \frac{t-T_n}{k}\widehat{\delta_{w_n}}([u])\right)\right| = \frac{1}{k}O(\sqrt{n})\cdot\frac{k}{n}$$

$\frac{1}{k}O(\sqrt{n})\cdot\frac{k}{n} = O\left(\frac{1}{\sqrt{n}}\right)$ since $k \geq n$. Using the same reasoning as the first point, we conclude. $\square$

3.5.3. Proof of Theorem 1 - first point

Let $\mu \in \mathcal{M}^{\mathrm{full}}_{\sigma-\mathrm{mix}}(\mathcal{A}^{\mathbb{Z}})$ and $u \in \mathcal{B}^{l+1}$. Since at time T_n there are no segment of length less that n, and by Propositions 7 and 8, one has $\max\left\{\mu\left(\bigcup_{n\leq k\leq K_n}\Gamma^t_{l,k}\right) : T_n \leq t < T_{n+1}\right\} \underset{n\to\infty}{\longrightarrow} 1$ exponentially fast. Therefore:

$$\max_{T_n\leq t<T_{n+1}} F^t_*\mu([u]) - \sum_{k=n}^{K_n}\mu\left(F^{-t}([u])|\Gamma^t_{l,k}\right)\mu\left(\Gamma^t_{l,k}\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

and $\Gamma^t_{l,k} = \Gamma^{T_n}_{l,k}$ since no segment is destroyed between T_n and T_{n+1}. By Proposition 10,

$$\begin{aligned}\max_{T_n\leq t<T_{n+1}}\Bigg|F^t_*\mu([u]) - \sum_{k=n}^{K_n}\mu(\Gamma^{T_n}_{l,k})\Bigg(&\max\left(0, \frac{k-(t-T_n)}{k}\right)\widehat{\delta_{w_n}}([u])\\ &+\min\left(1, \frac{t-T_n}{k}\right)\widehat{\delta_{w_{n-1}}}([u])\Bigg)\Bigg| = O\left(\frac{1}{\sqrt{n}}\right)\end{aligned}$$

$$\max_{T_n\leq t<T_{n+1}}\left|F^t_*\mu([u]) - \left(f_n(t)\widehat{\delta_{w_n}}([u]) + (1-f_n(t))\widehat{\delta_{w_{n-1}}}([u])\right)\right| = O\left(\frac{1}{\sqrt{n}}\right).$$

where f_n is the piecewise affine function defined by

$$\begin{array}{rcl} f_n : \quad [T_n, T_{n+1}] & \longrightarrow & [0,1] \\ t & \longmapsto & \sum_{k=n}^{K_n} \max\left(0, \frac{k-(t-T_n)}{k}\right) \mu\left(\Gamma_{l,k}^{T_n}\right) + \frac{t-T_n}{T_{n+1}-T_n}\mu\left(\Gamma_{l,>K_n}^{T_n}\right). \end{array}$$

The second term is chosen so that $f_n(T_n) = 0$ and $f_n(T_{n+1}) = 1$, but it converges to 0 exponentially and thus does not affect the equation. Therefore

$$\max_{T_n \leq t < T_{n+1}} d_{\mathcal{M}}\left(F_*^t\mu, \left[\widehat{\delta_{w_n}}, \widehat{\delta_{w_{n+1}}}\right]\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

Since f_n is $\frac{1}{n}$ Lipschitz on $[T_n, T_{n+1}]$, we deduce that

$$\max_{\nu \in \left[\widehat{\delta_{w_n}}, \widehat{\delta_{w_{n+1}}}\right]} d_{\mathcal{M}}\left(\nu, \left\{F_*^t\mu \mid T_n \leq t < T_{n+1}\right\}\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

We conclude that $\mathcal{V}(F,\mu) = \mathcal{V}((w_n)_{n\in\mathbb{N}})$.

When w_n is computable in space $\sqrt{n}$, by Proposition 6 we find that the rate of convergence is

$$d_{\mathcal{M}}\left(F_*^t\mu, \mathcal{V}\left((w_n)_{n\in\mathbb{N}}\right)\right) \leq O\left(\frac{1}{\log(t)}\right) + \sup\left\{d_{\mathcal{M}}\left(\nu, \mathcal{V}\left((w_n)_{n\in\mathbb{N}}\right)\right) : \nu \in \bigcup_{n \geq n(t)} \left[\widehat{\delta_{w_n}}, \widehat{\delta_{w_{n+1}}}\right]\right\},$$

where $n(t) = \Theta(\log(t)^2)$. We recall that w_n can always be computed in space $\sqrt{n}$ by repeating elements.

3.5.4. **Proof of Theorem 1 - second point**

Assume that $\mathcal{V}((w_i)_{i\in\mathbb{N}}) = \{\nu\}$, let F be the cellular automaton associated with this sequence as described above, and consider $\mu \in \mathcal{M}_{\sigma-\text{erg}}^{\text{full}}(\mathcal{A}^{\mathbb{Z}})$. Since μ is not assumed to be ψ-mixing, Proposition 8 does not apply, and there is no guarantee most segments are acceptable. However, since μ is ergodic, so is $F_*^t\mu$ for all t, and $\mu(\Gamma_{l,\geq k}^t) \underset{k\to\infty}{\longrightarrow} 0$.

Claim 1: $\mu(F^t(x)_0 \in \mathcal{A}\backslash\mathcal{B}) \underset{t\to\infty}{\longrightarrow} 0$, *i.e., the density of auxiliary states tends to 0.*

Proof. Suppose we are in an initial segment of length k. Detached time counters, Turing machines and merging counters initially present are destroyed in less than k steps. Similarly, left merging signals and copy auxiliary states initially present progress at speed -1, so they are destroyed before time k. An uninitialized wall is destroyed after $k(1+\log k)$ steps at most, and any counter attached to it are destroyed after less than k more steps. For all those states, the probability of apparition after time $t = k(2+\log k)$ is less than $\mu(\Gamma_{\geq k}^0) \underset{k\to\infty}{\longrightarrow} 0$.

At time T_n, all segments are longer than n, so the density of initialized walls and of auxiliary states that have been generated by them inside each segment is $O\left(\frac{1}{\sqrt{n}}\right)$.

Only uninitialized sweeping counters and right merging signals remain. Inside each segment, call **non-swept area** the interval between the initialized sweeping counter of the left wall and the rightmost cell containing one of those two states. At each step, this area decreases by one cell to its right but may grow by one cell to its left. Notice that merging with other segments cannot increase this area since segments of length n at time T_n are swept (see Figure 11).

At time T_n, a segment can contain a non-swept area longer than $\sqrt{n}$ only if it is issued from a segment longer than $\sqrt{n}$ initially, and the non-swept area of other segments have a density smaller than $\frac{1}{\sqrt{n}}$. By σ-invariance, $\mu(\{x \in \mathcal{A}^{\mathbb{Z}} \mid x_0 \text{ is in a non-swept area}\}) \leq \frac{1}{\sqrt{n}} + \mu(\Gamma_{\geq\sqrt{n}}^0) \underset{n\to\infty}{\longrightarrow} 0$.

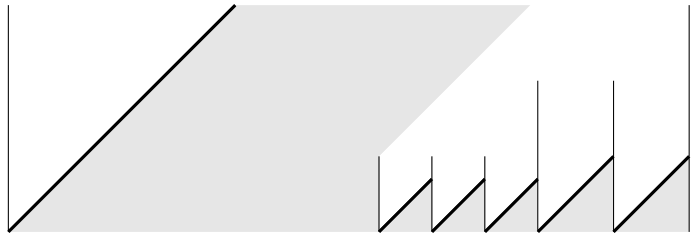

FIGURE 11. Illustration of the last part of the proof of Claim 1. Slanted lines are sweeping counters and grey areas are potentially non-swept.

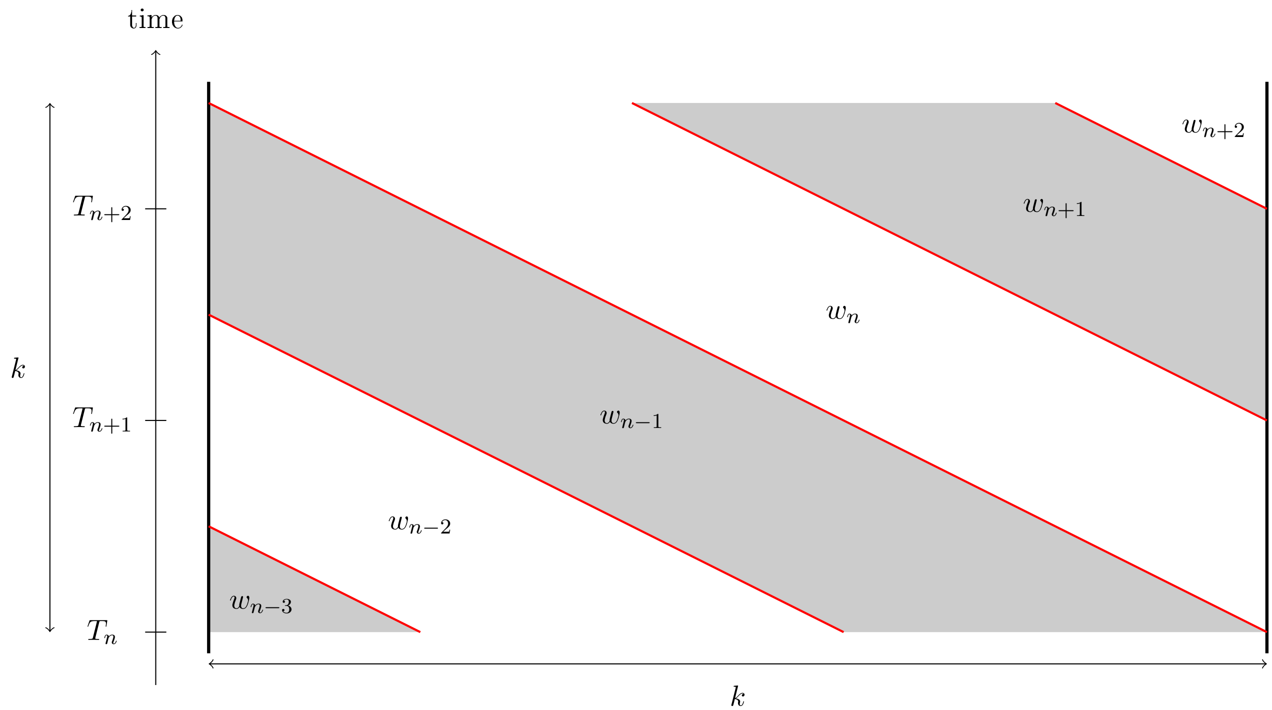

FIGURE 12. Illustration of Claim 2. When $t > T_n + k$, a segment of length k is a succession of stripes containing $w_n, w_{n+1}, \ldots$ plus a negligible part of auxiliary states and defects.

Therefore, for $a \in \mathcal{A}\backslash\mathcal{B}$, we have $F_*^t\mu([a]) \underset{t\to\infty}{\to} 0$. $\diamond$ `Claim 1`

CLAIM 2: *For any $n \in \mathbb{N}$, we have for t large enough $d_{\mathcal{M}}\left(F_*^t\mu, Conv\left((\widehat{\delta_{w_i}})_{i\geq n}\right)\right) \underset{t\to\infty}{\longrightarrow} 0$, where $Conv(X)$ is the convex hull of the set X.*

Proof. Consider a segment of length k at time T_n. At time $T_n + k$ the copying process for w_n will be finished, but since the segment is not necessarily acceptable, other copying processes may have started in the same segment. Therefore, the segment will be constituted by:

- a negligible number of auxiliary states;
- strips containing repeated copies of w_n, then $w_{n+1}, w_{n+2}\ldots$ separated by ongoing copy processes (the number of auxiliary copy states being negligible). See Figure 12.

Since the density of auxiliary states tends to 0, and $\mu(\Gamma^{T_n}_{l,\geq k}) \underset{k\to\infty}{\longrightarrow} 0$, for all $\varepsilon > 0$ it is possible to take k large enough so that $d_{\mathcal{M}}\left(F_*^{T_n+k}\mu, Conv((w_i)_{i\geq n})\right) < \varepsilon$. $\diamondsuit$ `Claim 2`

The second point of the Theorem 1 follows easily from Claim 2.

Remark. It does not follow from the last claim that the sequence $(F_*^t\mu)$ is close to any of the $\widehat{\delta_{w_i}}$ at any point, which is the reason why the result holds only for a single measure. This is why we control the length of the segments in the proof of the first point, which requires ψ-mixing.

4. Related problems solved with this construction

In this section, we use the construction developed in Theorem 1 in view to solve natural problems concerning accumulation point of the iteration of a cellular automaton on an initial measure.

4.1. Characterization of the μ-limit measures set

4.1.1. The connected case

Reciprocals of the computable obstructions described in Section 2 follow directly from Theorem 1.

Corollary 1. *Let $\nu \in \mathcal{M}_\sigma^{s\text{-}comp}(\mathcal{B}^{\mathbb{Z}})$ be a semi-computable measure. There is an alphabet $\mathcal{A} \supset \mathcal{B}$ and a cellular automaton $F : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ such that for any $\mu \in \mathcal{M}_{\sigma-erg}^{full}(\mathcal{A}^{\mathbb{Z}})$, one has $\lim_{n\to\mathbb{N}} F_*\mu = \nu$.*

This is in particular a full characterization of limit measures that are reachable from a computable initial measure $\mu \in \mathcal{M}_{\sigma-\mathrm{erg}}^{\mathrm{full}}(\mathcal{A}^{\mathbb{Z}})$.

Proof. Combine Proposition 1 with the first point of Theorem 1. $\square$

Corollary 2. *Let $\mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}})$ be a compact, Σ_2-computable and connected (Σ_2-CCC) subset of $\mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}})$. There is an alphabet $\mathcal{A} \supset \mathcal{B}$ and a cellular automaton $F : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ such that for any $\mu \in \mathcal{M}_{\sigma-mix}^{full}(\mathcal{A}^{\mathbb{Z}})$, one has $\mathcal{V}(F,\mu) = \mathcal{K}$.*

This is in particular a full characterization of connected μ-limit measures set that are reachable from a computable initial measure $\mu \in \mathcal{M}_{\sigma-\mathrm{erg}}^{\mathrm{full}}(\mathcal{A}^{\mathbb{Z}})$.

Proof. Combine Proposition 5 with Theorem 1. $\square$

Open question 1. *Is it possible to improve the speed of convergence?*

4.1.2. Towards the non-connected case

In Corollary 2 it is assumed that the set is connected. It is due to the fact that in the construction of Theorem 1, the words $(w_n)_{n\in\mathbb{N}}$ are copied progressively and not instantaneously on each segment, so that we get the closure of an infinite polygonal path, which is connected. However, we get topological obstructions even if we consider a non-connected μ-limit measures set. For example, if $\mathcal{V}(F,\mu)$ is finite one has the following proposition.

Proposition 11. *Let $F : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ be a cellular automaton and $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ such that $\mathcal{V}(F,\mu)$ is finite. Then F_* induces a cycle on $\mathcal{V}(F,\mu)$.*

Proof. Let $d = \min\{d_{\mathcal{M}}(\nu,\nu') : \nu,\nu' \in \mathcal{V}(F,\mu) \text{ with } \nu \neq \nu'\}$ and consider $\nu \in \mathcal{V}(F,\mu)$. It is possible to extract a sequence $(n_i)_{i\in\mathbb{N}}$ such that $d_{\mathcal{M}}(F_*^{n_i}\mu,\nu) < \frac{d}{3}$ and $d_{\mathcal{M}}(F_*^{n_i+1}\mu,\nu) > \frac{2d}{3}$. Since $d_{\mathcal{M}}(F_*^{n}\mu,\mathcal{V}(F,\mu)) \underset{n\to\infty}{\longrightarrow} 0$, we have $d_{\mathcal{M}}(F_*^{n_i}\mu,\nu) \underset{i\to\infty}{\longrightarrow} 0$. By continuity of F_*, $d_{\mathcal{M}}(F_*^{n_i+1}\mu,F_*\nu) \underset{i\to\infty}{\longrightarrow} 0$.

One deduces that for all $\nu \in \mathcal{V}(F,\mu)$ there exists $\nu' \in \mathcal{V}(F,\mu)$ such that $F_*\nu = \nu'$. So there is $k \in \mathbb{N}$ such that $\mathcal{V}(F,\mu) = \{\nu_0,\ldots,\nu_{k-1}\}$ and $F_*\nu_i = \nu_{i+1}$ where the addition is modulo k. $\square$

We exhibit some examples of more sophisticated behaviors based on the construction in Theorem 1. The first one is a family of cellular automata where $\mathcal{V}(F,\mu)$ is a finite set of connected components, which is a partial reciprocal of Proposition 11. The second one is a family of cellular automata where $\mathcal{V}(F,\mu)$ has an infinite number of connected components. However these are not total characterizations of the possible μ-limit measures sets.

Example 1 (Finite set of connected components)**.** Suppose $\mathcal{K}=\{\nu_0,\dots,\nu_{k-1}\}\subset\mathcal{M}_\sigma(\mathcal{B}^\mathbb{Z})$ is a finite set of σ-invariant semi-computable measures such that $G\nu_i=\nu_{i+1}$ for some periodic cellular automaton $G:\mathcal{B}^\mathbb{Z}\to\mathcal{B}^\mathbb{Z}$ ($G^k=Id$). Then there is an alphabet $\mathcal{A}\supset\mathcal{B}$ and a cellular automaton $F:\mathcal{A}^\mathbb{Z}\to\mathcal{A}^\mathbb{Z}$ such that $\mathcal{V}(F,\mu)=\mathcal{K}$ for $\mu\in\mathcal{M}^{\mathrm{full}}_{\sigma-\mathrm{erg}}(\mathcal{A}^\mathbb{Z})$. Indeed, let F be the cellular automaton satisfying $F^t_*\mu\to\nu_0$ obtained by Theorem 1. consider the cellular automaton that applies G on the main layer and applies the local rule of F once every k steps if an auxiliary state appears.

The same idea holds if $\mathcal{K}$ is a finite union of Σ_2-CCC sets which are mapped by a periodic cellular automaton $G:\mathcal{B}^\mathbb{Z}\to\mathcal{B}^\mathbb{Z}$.

Example 2 (Infinite set of connected components)**.** We give some informal elements to modify the construction of Theorem 1 to get examples of cellular automata where $\mathcal{V}(F,\mu)$ has an infinite number of connected components. The construction uses the firing squad cellular automaton $(\mathcal{B}_{\mathtt{FS}},F_{\mathtt{FS}})$ which has the following properties: there exists four states $\{\boxed{F},\blacksquare,\square,\boxed{O}\}\subset\mathcal{B}_{\mathtt{FS}}$ such that if $x_{[0,n]}=\boxed{O}\,\square^{n-1}\,\blacksquare$ then $F^{2n}_{\mathtt{FS}}(x)_{[0,n]}=\boxed{F}^{n+1}$ and the state $\boxed{F}$ does not appear in $(F^i_{\mathtt{FS}}(x)_j)_{(i,j)\in[0,n]\times[0,2n-1]}$[Maz96].

Consider a computable family $(\mathcal{K}_i)_{i\in\mathbb{N}}$ of Σ_2-CCC subsets of $\mathcal{M}_\sigma(\mathcal{B}^\mathbb{Z})$ and assume that $\mathcal{K}_i\cap\mathcal{K}_j=\emptyset$ for all $i,j\in\mathbb{N}$. There is a computable sequence of words $(w_n)_{n\in\mathbb{N}}$ such that $\mathcal{V}((w_n)_{n\in\mathbb{N}}=\bigcup_{i\in\mathbb{N}}\mathcal{K}_i$. Define $w'_n=w_n\times\square^{|w_n|}$ and consider the cellular automaton $(\mathcal{A}^\mathbb{Z},F)$ given by Theorem 1 which produces $\mathcal{V}((w'_n)_{n\in\mathbb{N}})$, with $\mathcal{A}\supset\mathcal{B}\times\mathcal{B}_{\mathtt{FS}}$. We modify F to obtain $\widetilde{F}$ in the following way.

- at time T_n, when the copy of w_n is initiated, we initialize a counter on another layer to count the length k of the segment;
- at time $t=T_{n+1}-2k$, the state $\boxed{O}$ appears on the left border of the segments (remember that the time counter keeps track of time);
- All $\boxed{F}$ symbols are immediatly transformed into $\square$ symbols.

This requires the segments to be short enough, but the probability that $[0,l]$ belongs to such a segment tends to 1 as time tends to infinity (see Remark 3.5.1). In those segments, $\widetilde{F_*}\mu$ approximates the measure $\widehat{\delta_{w_n}}\times\widehat{\delta_{\boxed{F}}}$ at time T_{n+1} and the measure $\widehat{\delta_{w_n}}\times\widehat{\delta_{\square}}$ at time $T_{n+1}+1$. The state $\boxed{F}$ appears only at times $(T_n)_{n\in\mathbb{N}}$.

For an initial measure $\mu\in\mathcal{M}^{\mathrm{full}}_{\sigma-\mathrm{mix}}(\mathcal{A}^\mathbb{Z})$, one has $\mathcal{V}(\widetilde{F},\mu)=\mathcal{V}\times\widehat{\delta_{\boxed{F}}}\cup\mathcal{K}'$ with $\mathcal{K}'\subset\mathcal{M}_\sigma\left(\left(\mathcal{B}_{\mathtt{FS}}\setminus\{\boxed{F}\}\right)^\mathbb{Z}\right)$, which means it has an infinite number of connected components.

Open question 2. *Is it possible to characterize all compact subsets of $\mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$ that can be reached as μ-limit measures set of some cellular automaton when μ is computable?*

4.2. Cesàro mean

In this section, by adapting the enumeration (w_n), we are able to get some control over the set $\mathcal{V}'(F,\mu)$ of limit points for the Cesàro mean sequence.

Corollary 3. *Let $\mathcal{B}$ be a finite alphabet and $\mathcal{K}'\subset\mathcal{M}_\sigma(\mathcal{B}^\mathbb{Z})$ a Σ_2-CCC set. There exists an alphabet $\mathcal{A}\supset\mathcal{B}$, and a cellular automaton $F:\mathcal{A}\to\mathcal{A}$ such that for any $\mu\in\mathcal{M}^{full}_{\sigma-mix}(\mathcal{A}^\mathbb{Z})$, one has $\mathcal{V}'(F,\mu)=\mathcal{K}'$.*

Remind that the latter is the set of limit points of the sequence $\varphi_t^F(\mu) = \frac{1}{t+1}\sum_{i=0}^t F_*^i\mu$. $\mathcal{V}'(F,\mu)$ is necessarily connected (because $\mathcal{M}(\mathcal{A}^{\mathbb{Z}})$ is metric and compact), and if we suppose that the initial measure μ is computable, we obtain a full characterization of reachable subsets $\mathcal{K}'$.

This corollary is a consequence of the following stronger result, where we have control over both $\mathcal{V}(F,\mu)$ and $\mathcal{V}'(F,\mu)$.

Corollary 4. *Let $\mathcal{B}$ be a finite alphabet and $\mathcal{K}' \subset \mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}})$ two Σ_2-CCC sets. There exist an alphabet $\mathcal{A} \supset \mathcal{B}$, and a cellular automaton $F : \mathcal{A} \to \mathcal{A}$ such that for any $\mu \in \mathcal{M}_{\sigma-mix}^{full}(\mathcal{A}^{\mathbb{Z}})$, one has*

- $\mathcal{V}(F,\mu) = \mathcal{K}$;
- $\mathcal{V}'(F,\mu) = \mathcal{K}'$.

$\mathcal{V}'(F,\mu)$ is necessarily included in the convex hull of $\mathcal{V}(F,\mu)$. Here we need a stronger hypothesis, namely, that it is included in $\mathcal{V}(F,\mu)$. Therefore, if we suppose the initial measure is computable, this is a characterization of reachable pairs of connected subsets $(\mathcal{K},\mathcal{K}')$ such that $\mathcal{K}' \subset \mathcal{K}$.

Proof. We will use notations from the proof of Proposition 5. Notably $(w_n)_{n\in\mathbb{N}}$ and $(w'_n)_{n\in\mathbb{N}}$ are the computable sequences of words associated to $\mathcal{K}$ and $\mathcal{K}'$, respectively, and $\mathbf{V}_k$ and $\mathbf{V}_k^t$ are defined with regard to $\mathcal{K}$. Without loss of generality, suppose that $\max(|w_n|,|w'_n|) \leq \sqrt{n}$ for all n (repeating some words if necessary).

We will define a new sequence of words $(w''_n)_{n\in\mathbb{N}}$ in the following manner, using a similar method as Proposition 5. For $n \in \mathbb{N}$, let i_n be the maximal value such that one can find a path $w_n = u_0, u_1, \ldots, u_l = w'_n, u_{l+1}, \ldots, u_{l'} = w_{n+1}$ with $u_1, \ldots u_{l-1}, u_{l+1}, \ldots, u_{l'} \in V_{i_n}^t$ and $d_{\mathcal{M}}(u_k, u_{k+1}) \leq 4b(i_n)$.

Let $P_n : [0,p_n] \to \mathbf{V}_{i_n}^t$ such a path. Since there are less than $|A|^{i_n}$ elements in $\mathbf{V}_{i_n}^t$, this path is of length $p_n \leq 2|A|^{i_n} \leq 2|A|^{|w_n|} < 2|A|^n$.

For $n \in [|\mathcal{A}|^{i^2}, |\mathcal{A}|^{(i+1)^2}]$, we define:

- if $n < |\mathcal{A}|^{i^2} + p_i$, $w''_n = P_i(n - |\mathcal{A}|^{i^2})$;
- otherwise, $w''_n = w'_{i+1}$.

and let F be the CA defined as in Theorem 1. Since all elements of $(w_n)_{n\in\mathbb{N}}$ are enumerated as in Proposition 5, we have $\mathcal{V}(F,\mu) = \mathcal{V}((w''_n)_{n\in\mathbb{N}}) = \mathcal{K}$.

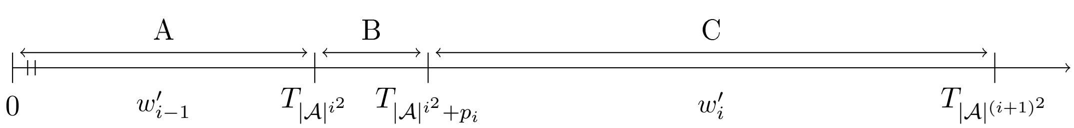

FIGURE 13. Intuitively, we prove $A + B \ll C$, then $B \ll A$.

We have

$$\frac{|\mathcal{A}|^{n^2}+p_n}{|\mathcal{A}|^{(n+1)^2} - (|\mathcal{A}|^{n^2}+p_n)} < \frac{|\mathcal{A}|^{n^2+1}}{|\mathcal{A}|^{(n+1)^2} - |\mathcal{A}|^{n^2+1}} \underset{n\to\infty}{\longrightarrow} 0.$$

In other words, the subset $[0, |\mathcal{A}|^{n^2}+p_n]$ is (asymptotically) of negligible density in $[0, |\mathcal{A}|^{(n+1)^2}]$. Since $T_{i+1}-T_i$ is an increasing sequence, the subset $[0, T_{|\mathcal{A}|^{n^2}+p_n}]$ is of negligible density in $[0, T_{|\mathcal{A}|^{(n+1)^2}}]$. This means that, putting $t_n = T_{|\mathcal{A}|^{(n+1)^2}}$, $d(\varphi_{t_n}^F(\mu), \widehat{\delta_{w'_{n+1}}}) \underset{n\to\infty}{\longrightarrow} 0$.

Furthermore, notice that for $x, y \in \mathbb{R}_+$, when $y \leq \sqrt{x}$, we have $\lfloor\sqrt{x+y}\rfloor \leq \lfloor\sqrt{x}\rfloor + 1$ and $\lfloor\sqrt{x-y}\rfloor \geq \lfloor\sqrt{x}\rfloor - 1$. Thus :

$$T_{|\mathcal{A}|^{n^2}+p_n} - T_{|\mathcal{A}|^{n^2}} < q^{|\mathcal{A}|^{\frac{n^2}{2}+1}} \cdot |\mathcal{A}|^n.$$

$$T_{|\mathcal{A}|^{n^2}} > T_{|\mathcal{A}|^{n^2}} - T_{|\mathcal{A}|^{n^2}-|\mathcal{A}|^{\frac{n^2}{2}}} > q^{|\mathcal{A}|^{\frac{n^2}{2}-1}} \cdot |\mathcal{A}|^{\frac{n^2}{2}}$$

where q is defined in Section 3.3.2 and therefore

$$\frac{T_{|\mathcal{A}|^{n^2}+p_n} - T_{|\mathcal{A}|^{n^2}}}{T_{|\mathcal{A}|^{n^2}+p_i}} \underset{n\to\infty}{\longrightarrow} 0.$$

This means that, when $t'_n = T_{|\mathcal{A}|^{n^2}+p_n}$, $d(\varphi^F_{t'_n}(\mu), \widehat{\delta_{w'_n}}) \underset{n\to\infty}{\longrightarrow} 0$.

The Cesàro mean sequence $\varphi^F_t(\mu)$ is (asymptotically) close to $\widehat{\delta_{w'_n}}$ between times t_n and t'_n, and is close to $\widehat{\delta_{w'_{n+1}}}$ at time t_{n+1}. Therefore, it is close to the segment $[\widehat{\delta_{w'_n}}, \widehat{\delta_{w'_{n+1}}}]$ between times t_n and t_{n+1}. We conclude that asymptotically, the sequence is close to $\mathcal{V}((w'_n))$, and thus its set of limit points is $\mathcal{K}'$. □

Open question 3. *Is it possible to extend Corollary 4 when $\mathcal{K}'$ is not included in $\mathcal{K}$?*

Using Example 1 we can only provide some examples where $\mathcal{V}(F,\mu) \cap \mathcal{V}'(F,\mu) = \emptyset$.

4.3. **Decidability consequences**

We give an undecidability result extending a result of Delacourt on μ-limit sets [Del11].

Corollary 5 (Rice theorem on μ-limit measures sets)**.** *Let P be a nontrivial property on non-empty Σ_2-CCC sets of $\mathcal{M}_\sigma(\mathcal{B}^\mathbb{Z})$ (i.e. not always or never true). Then it is undecidable, given an alphabet $\mathcal{A}$ and a CA $F : \mathcal{A}^\mathbb{Z} \to \mathcal{A}^\mathbb{Z}$, whether $\mathcal{V}(F,\mu)$ satisfies P for $\mu \in \mathcal{M}^{full}_{\sigma-mix}(\mathcal{B}^\mathbb{Z})$.*

Proof. We proceed by reduction to the halting problem. Since P is nontrivial, let $\mathcal{K}_1$ and $\mathcal{K}_2$ be two Σ_2-CCC sets that satisfies and does not satisfy P, respectively. By Proposition 5, there exists two computable sequences of words $(w_n)_{n\in\mathbb{N}}, (w'_n)_{n\in\mathbb{N}} \in (\mathcal{A}^*)^\mathbb{N}$ such that $\mathcal{K}_1 = \mathcal{V}((w_n)_{n\in\mathbb{N}}), \mathcal{K}_2 = \mathcal{V}((w'_n)_{n\in\mathbb{N}})$.

Now let $\mathcal{TM}$ be a Turing machine. Define the sequence $(w''_n)_{n\in\mathbb{N}}$ in the following way.

- If $\mathcal{TM}$ halts on the empty input in less than n steps, $w''_n = w_n$.
- Otherwise, $w''_n = w'_n$.

This sequence is computable by simulating n steps of the Turing machine and computing the corresponding sequence. Therefore, we can use the previous construction to build a CA F such that $\mathcal{V}(F,\mu) = \mathcal{V}((w''_n)_{n\in\mathbb{N}})$. If $\mathcal{TM}$ halts on the empty input, then $w''_n = w_n$ for n large enough; otherwise, $w''_n = w'_n$ for n large enough. Thus, $\mathcal{V}(F,\mu)$ satisfies P if and only if $\mathcal{TM}$ halts. □

The same reasoning holds for a single limit and the Cesàro mean sequence.

Corollary 6 (Rice theorem on single limit measures)**.** *Let P be a nontrivial property on $\mathcal{M}^{s\text{-}comp}_\sigma(\mathcal{B}^\mathbb{Z})$. Then it is undecidable, given an alphabet $\mathcal{A}$ and a CA $F : \mathcal{A}^\mathbb{Z} \to \mathcal{A}^\mathbb{Z}$, whether $F^t_*\mu \to \nu$ where ν satisfies P for $\mu \in \mathcal{M}^{full}_{\sigma-erg}(\mathcal{B}^\mathbb{Z})$.*

Corollary 7 (Rice theorem on Cesàro mean μ-limit measures sets)**.** *Let P be a nontrivial property on non-empty Σ_2-CCC sets of $\mathcal{M}_\sigma(\mathcal{B}^\mathbb{Z})$. Then it is undecidable, given an alphabet $\mathcal{A}$ and a CA $F : \mathcal{A}^\mathbb{Z} \to \mathcal{A}^\mathbb{Z}$, whether $\mathcal{V}'(F,\mu)$ satisfies P for $\mu \in \mathcal{M}^{full}_{\sigma-mix}(\mathcal{B}^\mathbb{Z})$.*

4.4. **Computation on the set of measures**

In this section the construction developed in Section 3 is modified to perform computation on the space of probability measures. In other words, we want the μ-limit measures set to be a function of the initial measure.

4.4.1. **Computation with oracle**

The obstructions of Section 2 can be generalized to the case where the initial measure is not computable, by considering computability with an oracle $\mu \in \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$.

A **Turing machine with oracle** in $\mathcal{M} \subset \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$ has the same behavior as a classical Turing machine, except that an oracle $\mu \in \mathcal{M}$ is fixed prior to computation. It can query the oracle during the computation by writing $u \in \mathcal{A}^*$ and $n \in \mathbb{N}$ on an additional **oracle tape** and entering a special **oracle state**. After one step, the oracle returns an approximation of $\mu([u])$ up to an error 2^{-n} and the computation resumes.

Let $\mathcal{M} \subset \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$ and X, Y two enumerable sets. A function $f : \mathcal{M} \times X \to Y$ is **computable with oracles in** $\mathcal{M}$ if there exists a Turing machine with oracle in $\mathcal{M}$ which takes as input $x \in X$ and returns $y = f(\mu, x) \in Y$, up to reasonable encoding.

Definition 13. Let $\mathcal{M} \subset \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$.

A function $\varphi : \mathcal{M} \longrightarrow \mathcal{M}_\sigma(\mathcal{B}^\mathbb{Z})$ is **computable in** $\mathcal{M}$ if there exists $f : \mathcal{M} \times \mathbb{N} \longrightarrow \mathcal{B}^*$ a computable function with oracle in $\mathcal{M}$ such that $|\varphi(\mu) - \widehat{\delta_{f(\mu,n)}}| \leq 2^{-n}$. This is an extension of the previous definition where the image is not countable, hence the abuse of notation.

A function $\varphi : \mathcal{M} \longrightarrow \mathcal{M}_\sigma(\mathcal{B}^\mathbb{Z})$ is **semi-computable with oracle** in $\mathcal{M}$ if there exists $f : \mathcal{M} \times \mathbb{N} \longrightarrow \mathcal{B}^*$ a computable function with oracle in $\mathcal{M}$ such that $\widehat{\delta_{f(\mu,n)}} \underset{n\to\infty}{\longrightarrow} \varphi(\mu)$.

A sequence of functions $(f_n : \mathcal{M} \times \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z}) \longrightarrow \mathbb{R})_{n\in\mathbb{N}}$ is **a computable sequence of functions with oracle in** $\mathcal{M}$ if

- there exists $a : \mathcal{M}\times\mathbb{N}\times\mathbb{N}\times\mathcal{A}^* \longrightarrow \mathbb{Q}$ computable with oracle in $\mathcal{M}$ such that $\left| f_n(\mu, \widehat{\delta_w}) - a(\mu, n, m, w)\right| \leq \frac{1}{m}$ for all $\mu \in \mathcal{M}$, $w \in \mathcal{A}^*$ and $n, m \in \mathbb{N}$;
- there exists $b : \mathcal{M} \times \mathbb{N} \longrightarrow \mathbb{Q}$ computable with oracle in $\mathcal{M}$ such that $d_\mathcal{M}(\nu, \nu') < b(\mu, m)$ implies $|f_n(\mu, \nu) - f_n(\mu, \nu')| \leq \frac{1}{m}$ for all $\mu \in \mathcal{M}$ and $n, m \in \mathbb{N}$.

Let $\mathfrak{K}$ be a set of compact subsets of $\mathcal{M}_\sigma(\mathcal{B}^\mathbb{Z})$. A function $\Psi : \mathcal{M} \longrightarrow \mathfrak{K}$ is **Σ_2-computable** if there exists a computable sequence of functions $(f_n : \mathcal{M} \times \mathcal{M}_\sigma(\mathcal{B}^\mathbb{Z}) \longrightarrow \mathbb{R})_{n\in\mathbb{N}}$ with oracle in $\mathcal{M}$ such that $f_\mu(\nu) = \lim_{n\to\infty} f_n(\mu, \nu)$ for all $\mu \in \mathcal{M}$ and $\nu \in \mathcal{M}_\sigma(\mathcal{B}^\mathbb{Z})$ and $\Psi(\mu) = f_\mu^{-1}(\{0\})$.

The proofs of Section 2 can be easily adapted in this framework. For any cellular automaton F on $\mathcal{A}^\mathbb{Z}$, one has:

- following Proposition 2, the function $\mu \longmapsto F_*\mu$ is computable with oracle in $\mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$;
- following Proposition 4, $\mu \longmapsto \mathcal{V}(F, \mu)$ and $\mu \longmapsto \mathcal{V}'(F, \mu)$ are Σ_2-computable with oracle in $\mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$;
- following Proposition 5, if $\Psi : \mathcal{M} \longrightarrow \mathfrak{K}$ is a Σ_2-computable function with oracle in $\mathcal{M}$ and if every element of $\mathfrak{K}$ is connected, then there exists a computable function $f : \mathcal{M} \times \mathbb{N} \longrightarrow \mathcal{A}^*$ with oracle in $\mathcal{M}$ such that $\Psi(\mu) = \mathcal{V}((f(\mu, n))_{n\in\mathbb{N}})$ (closure of the limit points of the polygonal path).

4.4.2. **Towards a reciprocal**

In this section, we give a partial reciprocal to the last fact. To use the initial measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^\mathbb{Z})$ as an oracle, we need to keep some information from the initial configuration. We adapt the original construction in the following way:

Each segment keeps a sample of the initial configuration, using the frequency of patterns inside this sample as an oracle in the computation. We need to ensure that the frequency of a pattern $u \in \mathcal{A}^k$ in this sample is close to $\mu([u])$ with a high probability. For $\mu \in \mathcal{M}^{\text{full}}_{\sigma-\text{mix}}(\mathcal{A}^{\mathbb{Z}})$ we have an exponential rate of convergence for every length (Theorem III.1.7 of [Shi96]). More precisely:

$$\mu\left(\left\{x \in \mathcal{A}^{\mathbb{Z}} : \max_{u \in \mathcal{A}^k}\{|\mu([u]) - \text{Freq}(u, x_{[0,n]})| \geq \varepsilon\}\right\}\right) \leq (k+m)\psi(m)^{\frac{n}{k}}\left(\frac{n}{k}+1\right)^{\text{Card}(A)^k} 2^{-\frac{nc\varepsilon^2}{4k}},$$

where $m \in \mathbb{N}$, $c > 0$.

However, in our case, not all the information in the initial configuration can be kept since sweeping destroys information in the segment. In all the following, we will only keep information about the density of $\boxed{\text{I}}$ symbols. It would actually be possible to adapt the construction and keep information on longer words, only considering the positions of $\boxed{\text{I}}$ symbols.

Theorem 2. *Let $\Psi : \mathcal{M}^{full}_{\sigma-mix}(\{0,1\}^{\mathbb{Z}}) \to \mathfrak{K}$ be a Σ_2-computable function where $\mathfrak{K}$ is a set of compact connected subsets of $\mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}})$. Assume that $\Psi(\mu) = \Psi(\mu')$ if $\mu([\boxed{I}]) = \mu'([\boxed{I}])$ for $\mu, \mu' \in \mathcal{M}^{full}_{\sigma-mix}(\{0,1\}^{\mathbb{Z}})$.*

There exists a cellular automaton $(\mathcal{A}^{\mathbb{Z}}, F)$ such that $\mathcal{V}(F,\mu) = \Psi(\pi\mu)$ for all $\mu \in \mathcal{M}_{\sigma-mix}(\mathcal{A}^{\mathbb{Z}})$, where π is a 1-block map defined by $\pi(x)_i = 1$ when $x_i = \boxed{I}$, and $\pi(x)_i = 0$ otherwise.

Notice that since only one density is considered, it would be equivalent in this case to consider a Σ_2-computable function $\mathbb{R} \to \mathfrak{K}$.

Proof. Let $f : \mathcal{M}^{\text{full}}_{\sigma-\text{mix}}(\{0,1\}^{\mathbb{Z}}) \times \mathbb{N} \longrightarrow \mathcal{A}^*$ be a computable function with oracle in $\mathcal{M}^{\text{full}}_{\sigma-\text{mix}}(\{0,1\}^{\mathbb{Z}})$ such that $\Psi(\mu) = \mathcal{V}((f(\mu,n))_{n\in\mathbb{N}})$ and consider the associated Turing machine with oracle.

Let F be the cellular automaton defined in Theorem 1. We add a new layer $\mathcal{A}_{\texttt{oracle}}$ in which each segment at time t stores the frequency of the state $\boxed{\text{I}}$ in this segment at time 0. To do that, we modify the construction in the following way:

- We subdivide the layer $\mathcal{A}_{\texttt{oracle}}$ in two parts, on which each wall $\boxed{\text{W}}$ keeps on its left:
  - the first counter for the number of $\boxed{\text{I}}$ symbols that have been destroyed in its left segment;
  - the second counter for its length, worth 0 if the segment is not swept.
- Another counter accompanies each sweeping counter, measuring the length of the segment as it progresses.
- The second counter is initialized as 0. When the time counter attached to this wall makes a comparison with an initialized sweeping counter (the comparison returns the result "="), the second counter stores the length of the segment. It may take the value 0 again after merging with a non-swept segment (see below).
- When a wall is destroyed by a merging process, it sends to its right a signal at speed 1 containing all the stored information. Such a signal should not cross a sweeping counter, so it is slowed down if necessary.
- When a wall has stored (c_1, c_2) as oracle and receives the signal (c'_1, c'_2) from its left, there are three cases:
  - If $c_2 = 0$, the left segment was not swept, the signal cannot come from an initialized wall and can be safely ignored. The oracle remains (c_1, c_2).
  - If $c_2 \neq 0$, the information comes from an initialized wall. Put $c''_1 = c_1 + c'_1 + 1$ to take the merging into account. If $c'_2 = 0$, the segment just merged with a non swept segment and $c''_2 = 0$; otherwise $c''_2 = c_2 + c'_2$. The new oracle is (c''_1, c''_2).

  See Figure 14. We remark that if the length of the segment is k, the information can be coded in space $\log(k)$, and it is possible to actualize the values before another signal can come from the left.

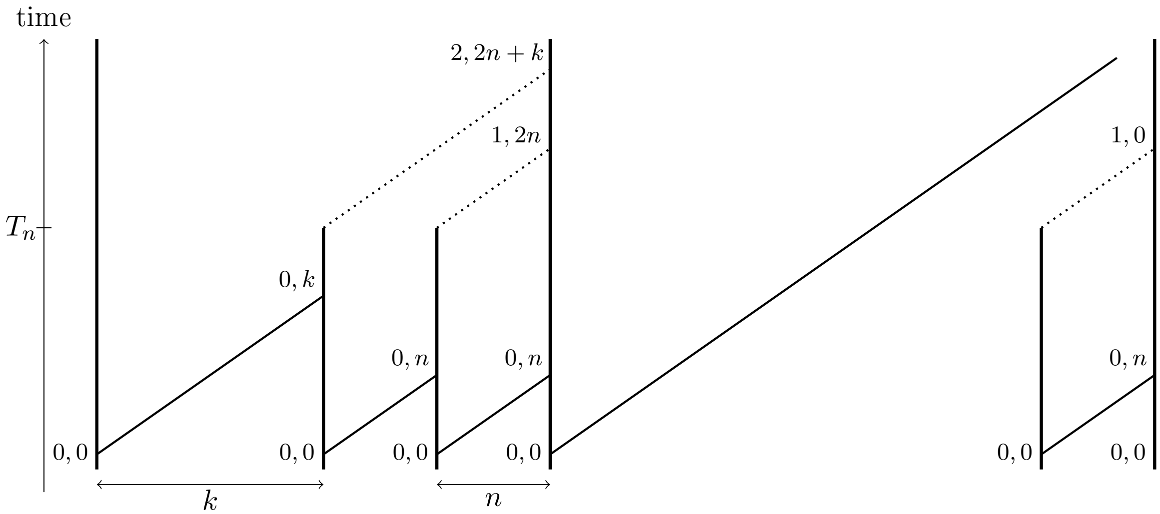

FIGURE 14. Each wall has its counter displayed when its value changes. Slanted thick lines are sweeping counters, dotted lines are signals transmitting information.

- If two symbols $\boxed{\text{I}}$ are too close in the initial configuration, they are destroyed by the bootstrapping process (see Section 3.2.1). If a $\boxed{\text{I}}$ is in a group of $\boxed{\text{I}}$ separated by two cells or less, the rightmost $\boxed{\text{I}}$ sends a sweeping counter and the leftmost one starts a time counter. Thus a group of $\boxed{\text{I}}$ separated by two cells or less behave as a single symbol for initialization purposes. All the $\boxed{\text{I}}$ except the leftmost one are transformed immediately into oracle signals (supposing the basis of the counter is larger than 3 then they occupy only one cell) and the other cells present initially are erased.
- The Turing machine simulation described in Section 3.3.2 can be adapted to simulate a Turing machine with oracle. When there is an oracle query for the value of $\mu([\boxed{\text{I}}])$ with the precision 2^{-i} at time $t \in [T_n, T_{n+1}]$, there are two possibilities:
  - if $n^{-\frac{1}{6}} \leq 2^{-i}$, the Turing machine uses the information stored in the oracle layer to return the frequency of $\boxed{\text{I}}$ on the segment at time 0, and this corresponds to an approximation of $\mu([\boxed{\text{I}}])$ with sufficient precision;
  - if $n^{-\frac{1}{6}} > 2^{-i}$, the computation stops, and the last word successfully computed is output. The same thing happens until a time when enough information is available.

Let us check that $\mathcal{V}(F,\mu) = \Psi(\pi\mu)$ for $\mu \in \mathcal{M}^{\text{full}}_{\sigma-\text{mix}}(\mathcal{A}^{\mathbb{Z}})$. It is clear that the density of auxiliary states tends to 0, so if the sample approximates correctly $\mu([\boxed{\text{I}}])$, the sequence of words $(w_n)_{n\in\mathbb{N}}$ produced by the cellular automaton correspond to $(f(\mu,n))_{n\in\mathbb{N}}$ up to some repetition. Thus we only need to prove that the probability that a cell belongs to a segment which sample correspond to a "bad" approximation tends to 0 when t tends to ∞. Recall that $\Gamma^{T_n}_{[i,j]} = \{x \in \mathcal{A}^{\mathbb{Z}} \mid [i,j]$ is a segment at time $T_n\}$.

$$\begin{aligned}
B_n &= \mu\left(\left\{x \in \mathcal{A}^{\mathbb{Z}} : x_0 \text{ belongs in a segment with a ``bad'' sample at time } T_n\right\}\right)\\
&= \sum_{i<0,j>0} \mu\left(\left\{x \in \Gamma^{T_n}_{[i,j]} : |\mu([u]) - Freq(u, x_{[i,j]})| > n^{-\frac{1}{6}}\right\}\right)\\
&= \sum_{k>0} k\cdot\mu\left(\left\{x \in \Gamma^{T_n}_{[0,k]} : |\mu([u]) - Freq(u, x_{[0,k]})| > n^{-\frac{1}{6}}\right\}\right),
\end{aligned}$$

by σ-invariance. By restricting ourselves to $n \leq k \leq K_n$:

$$\begin{array}{rcl} B_n & \leq & \mu\left(\Gamma^{T_n}_{0,\geq K_n}\right) + \displaystyle\sum_{k=n}^{K_n} k \cdot \mu\left(\left\{x \in \mathcal{A}^{\mathbb{Z}} : |\mu([u]) - Freq(u, x_{[0,k]})| > n^{-\frac{1}{6}}\right\}\right) \\ & \leq & \mu\left(\Gamma^{T_n}_{0,\geq K_n}\right) + K_n^2(1+m)\psi(m)^n\,(n+1)^{\mathrm{Card}(A)}\,2^{-\frac{c}{4}n^{\frac{2}{3}}} \\ & \underset{n\to\infty}{\longrightarrow} & 0. \end{array}$$

The result follows.

□

This result may seem surprising since the same cellular automaton has very different asymptotical behaviors depending on the initial measure.

Open question 4. *Is it possible to improve Theorem 2 and characterize functions $\Psi : \mathcal{M}^{full}_{\psi-mix}(\{0,1\}^{\mathbb{Z}}) \to \mathfrak{K}$, where $\mathfrak{K}$ is a set of compact subsets of $\mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}})$, that are realisable as the action of a cellular automaton F in the sense that for all μ, $\mathcal{V}(F,\mu) = \Psi(\mu)$?*

5. Removing the auxiliary states

In this section, our aim is to carry the previous results to the case where the cellular automaton does not use auxiliary states. A straightforward extension is impossible: for example, consider ν a semi-computable measure with full support and $F : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ a cellular automaton such that $F_*\mu \to \nu$ for any "simple" measure μ. Since ν has full support, F is a surjective automaton, and hence the uniform Bernoulli measure is invariant under F_*. Thus ν must be the uniform Bernoulli measure.

However, if the limit measure does not have full support, the previous results can be extended by using a word not charged by the measure to encode the auxiliary states in some sense.

Theorem 3. *Let $(w_n)_{n\in\mathbb{N}}$ be a computable sequence of words of $\mathcal{B}^*$, where $\mathcal{B}$ is a finite alphabet, and assume there exists a word u that does not appear as factor in any of the w_n. Then there is a cellular automaton $F : \mathcal{B}^{\mathbb{Z}} \to \mathcal{B}^{\mathbb{Z}}$ such that for any measure $\mu \in \mathcal{M}^{full}_{\sigma-mix}(\mathcal{B}^{\mathbb{Z}})$, $\mathcal{V}(F,\mu) = \mathcal{V}((w_n)_{n\in\mathbb{N}})$.*

Proof. Let $\mathcal{A}$ be the alphabet and F be the CA associated to the sequence $(w_n)_{n\in\mathbb{N}}$ by Theorem 1. Our aim is to provide an encoding of any configuration of $\mathcal{A}^{\mathbb{Z}}$ in $\mathcal{B}^{\mathbb{Z}}$ and a cellular automaton F' that behaves similarly to F after encoding.

Denote $U_d \subset \mathcal{B}^d$ be the set of words of length d beginning with u, that do not contain u as factor (except at the first letter), and that do not end with a prefix of u. $\#(U_d) \underset{d\to\infty}{\longrightarrow} \infty$, so for d large enough, we can find an injection $\varphi : \mathcal{A}\backslash\mathcal{B} \to U_d$ (encoding the auxiliary states), and we extend it by putting $\varphi = Id$ on $\mathcal{B}$. For a finite word, we define $\varphi(u_1 \dots u_n) = \varphi(u_1)\dots\varphi(u_n)$, and this can be naturally extended further to configurations $\Phi : \mathcal{A}^{\mathbb{Z}} \mapsto \mathcal{B}^{\mathbb{Z}}$ by considering that $\varphi(a_0)$ starts on the column zero.

Let $\mathbf{T} \subset \mathcal{A}^{\mathbb{Z}}$ be the set of configurations such that the word u does not appear on the main layer ($\mathbf{T}$ is a subshift of finite type). Since u marks unambiguously the beginning of a word of $\varphi(\mathcal{A}\backslash\mathcal{B})$, the restriction $\Phi : \mathbf{T} \to \mathcal{B}^{\mathbb{Z}}$ is injective.

Each configuration from $\mathcal{B}^{\mathbb{Z}}$ can thus be divided uniquely into words from $\varphi(\mathcal{A})$, that we will call **clusters** from now on. **Output cells** are elements of $\mathcal{B} = \varphi(\mathcal{B})$ that occupy only 1 cell (corresponding to $(b, \#, \#, \#, \#, \#)$ for $b \in \mathcal{B}$ in the previous construction) and **auxiliary clusters** are elements of $\varphi(A \backslash B)$ that occupy d cells while containing one letter of output. Thus we can define a decoding $\Psi : \mathcal{B}^{\mathbb{Z}} \to \mathbf{T}$ such that $\Psi \circ \Phi = \mathrm{Id}$.

However, Φ and Ψ are not σ-invariant, so $\Phi \circ F \circ \Psi$ is not a cellular automaton. We must build manually a cellular automaton on $\mathcal{B}^{\mathbb{Z}}$ that behaves in roughly the same way as $\Phi \circ F \circ \Psi$. Provided the neighborhood is larger than $[-4d, 4d]$, each cell can “read" the cluster in which it belongs, and the three clusters to its right and left.

If a word u appears outside of an auxiliary cluster, it is replaced by some output cells and can never be created again. To avoid creating an auxiliary cluster by mistake, we fix to this purpose a letter $b \in \mathcal{B}$ such that $b^d \notin U_d$. Similarly, auxiliary clusters that are destroyed for any reason leave behind them output b cells.

Remark. For clarity, in all diagrams of this section, we suppose that $\mathcal{B} = \{0, 1\}$, $d = 3$ (it would be much larger in real implementations) and we represent auxiliary clusters as blocks with layers, instead of words from $\mathcal{B}^d$. Also we fix $b = 0$ in the definition above.

The different parts of the construction are modified in the following way.

- $\boxed{\text{I}}$ and $\boxed{\text{W}}$ clusters, time counters, and Turing machines have the same behavior as in the previous construction. However, since the counters take more space, it is necessary to erase $3d$ cells to the left and right of each $\boxed{\text{I}}$ cluster at time 0.
- The tail of copying process progresses to the left at speed one, and behaves normally as long as it does not meet another auxiliary state (see Figure 15).

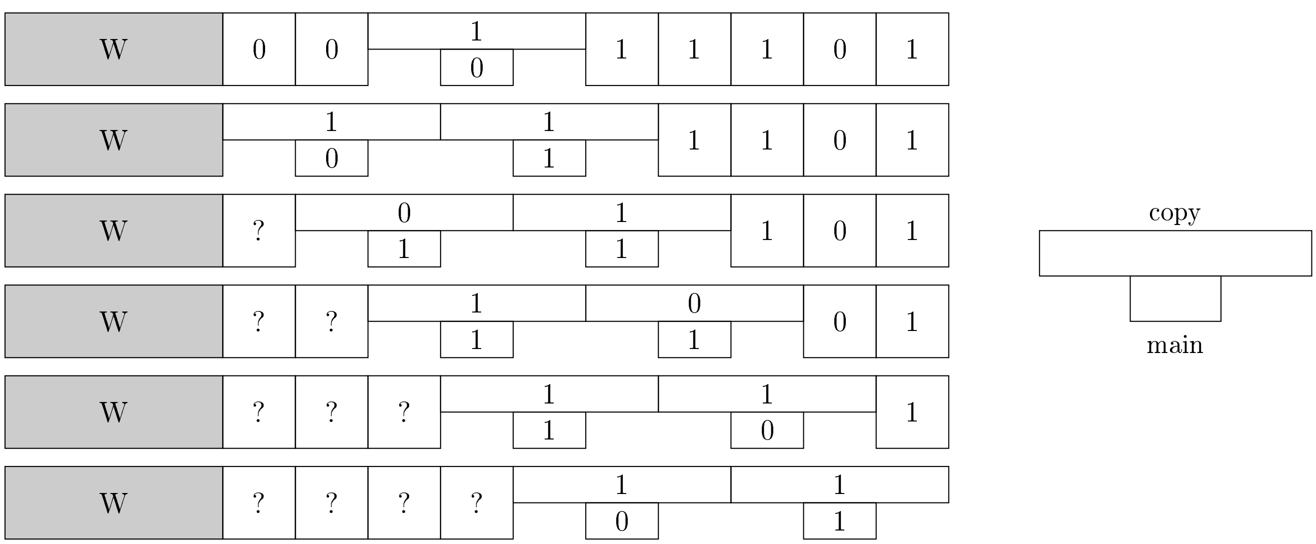

FIGURE 15. End of the copying process described in Figure 7, copying 1101.

- Sweeping counters progress to the right at speed d. This is too fast to keep the output information, so the counter leaves behind output cells b defined above. Any moving signal it meets (e.g. copying process or length signal) is destroyed. When entering the time counter, if it cannot progress by d cells exactly, it is offset by less than d cells (see Figure 16). Thus sweeping clusters separated by small offsets are still considered to be the same counter.
- Merging signals which determine length of segments also progress at speed d. To avoid being offset by copying processes (which would modify the “measured length”), the determination of length starts only after the copy is finished. Thus the signal is only offset once, when

0	0	0	0	0	0	0	0	0	?	time sweeping	time sweeping
0	0	0	0	0	0	sweeping			?	time sweeping	time
0	0	0	sweeping			sweeping			?	time	time
sweeping			sweeping			?	?	?	?	time	time

FIGURE 16. A sweeping counter gets offset when entering the time counter area. Notice the auxiliary clusters being replaced by output cells containing zeroes.

entering the time counter area. After bouncing off the right wall, it returns to the left wall where its offset can be measured. If it takes t_0 steps to return with an offset of α, then the segment has length $\frac{t_0}{2} \cdot d + \alpha$ (see Figure 17). On the left side of the wall, a Turing machine computes the measured length and compares it with n, and a $\boxed{\mathrm{M}}$ symbol is created if needed.

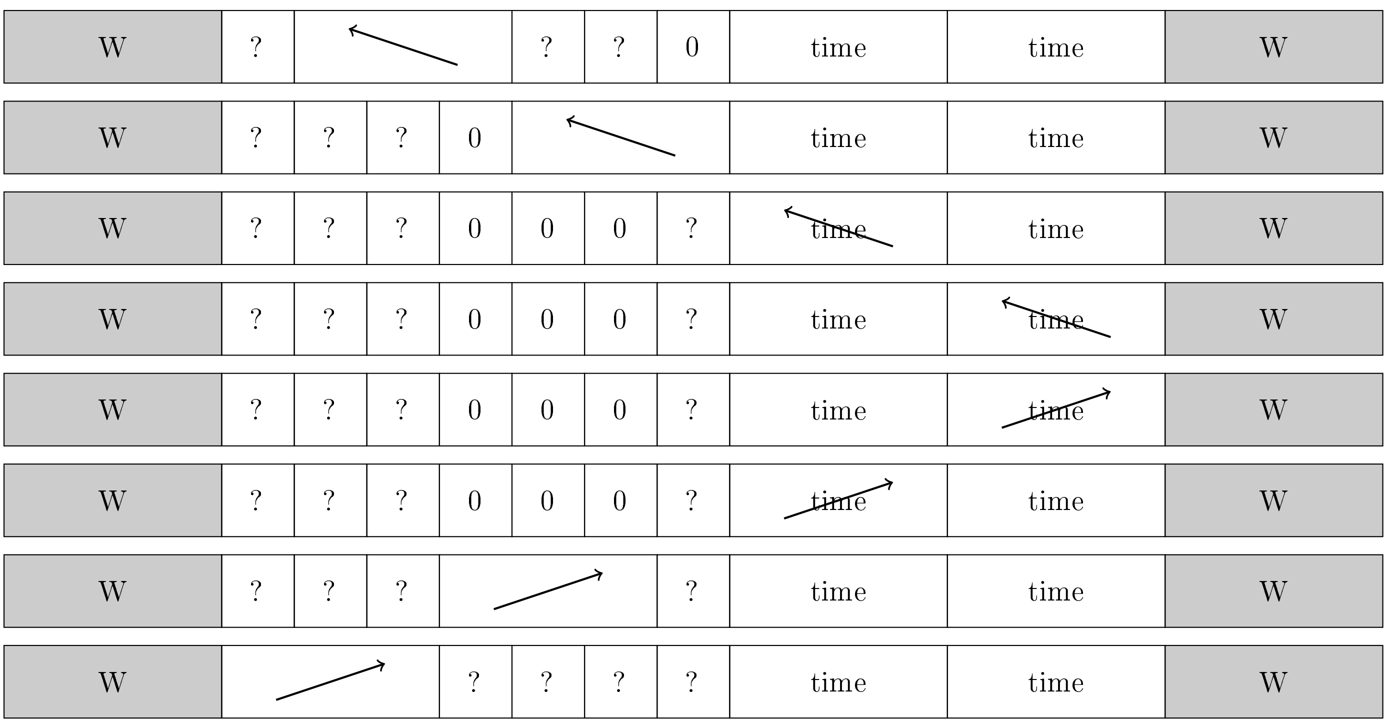

FIGURE 17. Determination of length. Here $d = 3$, $t_0 = 8$ and $o = 1$, for a measured length of 13.

The bootstraping and sweeping processes work essentially in the same way as previously, except that a sweeping counter erases any copy process and merging signal it meets, along with output information. Hence Propositions 9 and 8 can be extended. Furthermore, at time t, with $T_n \leq t < T_{n+1}$, the copy process followed by the process of determination of length for segments of size $n+1$

still take less than $T_{n+1} - T_n$ steps. Hence the proof in section 3.5.3 can be extended, and the theorem follows. □

However, because of the destructive nature of the sweeping counter, the proof in Section 3.5.4 cannot be adapted and we cannot weaken the hypothesis to $\mu \in \mathcal{M}^{\text{full}}_{\sigma-\text{erg}}(\mathcal{B}^{\mathbb{Z}})$ when $\mathcal{K}$ is a singleton. Since this result is a counterpart to the second point of Theorem 1 that does not use auxiliary states, it is natural to give similar counterparts to corollaries 2 to 7.

Definition 14. A word $u \in \mathcal{A}^*$ is said to be **not charged** by a set $\mathcal{K} \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}})$ if for all $\nu \in \mathcal{K}$, $\nu([u]) = 0$.

Corollary 8. *Let $\mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}})$ be a non-empty Σ_2-CCC subset of $\mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}})$ that does not charge a word $u \in \mathcal{B}^*$. Then there is a cellular automaton $F : \mathcal{B}^{\mathbb{Z}} \to \mathcal{B}^{\mathbb{Z}}$ such that for any measure $\mu \in \mathcal{M}^{full}_{\sigma-mix}(\mathcal{B}^{\mathbb{Z}})$, $\mathcal{V}(F, \mu) = \mathcal{K}$. In particular, any semi-computable measure which does not have full support can be obtained this way.*

Proof. Since $\mathcal{K}$ does not charge u, we can assume without loss of generality that no word in the computable sequence $(w_n)_{n\in\mathbb{N}}$ associated to $\mathcal{K}$ by Proposition 1 contains u as factor. Thus Theorem 3 applies. □

The proofs of the following corollaries are adaptations of the proofs of their counterparts using Theorem 3. Corollary 1 does not have a counterpart since its proof uses the first point of Theorem 1.

Corollary 9. *Let $\mathcal{K}' \subset \mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}})$ two non-empty Σ_2-CCC sets that both do not charge the same word $u \in \mathcal{B}^*$. Then there exists a cellular automaton $F : \mathcal{B} \to \mathcal{B}$ such that for any $\mu \in \mathcal{M}^{full}_{\sigma-mix}(\mathcal{A}^{\mathbb{Z}})$,*

- *$\mathcal{V}(F, \mu) = \mathcal{K}$;*
- *$\mathcal{V}'(F, \mu) = \mathcal{K}'$.*

Corollary 10 (Rice theorem on μ-limit measures sets)**.** *Let $\mathcal{B}$ be an alphabet, $\mu \in \mathcal{M}^{full}_{\sigma-mix}(\mathcal{B}^{\mathbb{Z}})$, $u \in \mathcal{B}^*$, and P be a nontrivial property on non-empty Σ_2-CCC sets that do not charge u. Then it is undecidable, given a CA $F : \mathcal{B}^{\mathbb{Z}} \to \mathcal{B}^{\mathbb{Z}}$, whether $\mathcal{V}(F, \mu)$ satisfies P.*

This result extends to single measures and Cesàro mean μ-limit measures set, in a similar way as Corollaries 6 and 7.

We leave open in particular the case of limit measures with full support. For corollaries 8 and 9, solving this case would imply to characterize the possible asymptotic behaviors of surjective automata, for which a similar construction seems difficult. As for Corollary 10, if we fix μ the uniform Bernoulli measure, the problem of whether $\mathcal{V}(F, \mu)$ contains only the uniform Bernoulli measure is equivalent to the surjectivity of F, which is decidable [AP72]. Hence the question of which nontrivial properties on limit measures and μ-limit measures sets with full support are decidable remains open.

Open question 5. *Which sets of measures are reachable by surjective cellular automata?*

References

[AP72] S. Amoroso and Y.N. Patt. Decision procedures for surjectivity and injectivity of parallel maps for tessellation structures. *Journal of Computer and System Sciences*, 6(5):448–464, 1972.

[AS11] Nathalie Aubrun and Mathieu Sablik. Simulation of effective subshifts by two-dimensional subshift of finite type. *Acta Applicandae Mathematicae, Volume 128, issue 1, p.35-63, 2013.*

[BDS10] Laurent Boyer, Martin Delacourt, and Mathieu Sablik. Construction of μ-limit sets. In *JAC*, pages 76–87, 2010.

[BP11] Mike Boyle and Karl Petersen. Hidden Markov processes in the context of symbolic dynamics. In *Entropy of hidden Markov processes and connections to dynamical systems*, volume 385 of *London Math. Soc. Lecture Note Ser.*, pages 5–71. Cambridge Univ. Press, Cambridge, 2011.

[BPT06] Laurent Boyer, Victor Poupet, and Guillaume Theyssier. On the complexity of limit sets of cellular automata associated with probability measures. In *MFCS*, pages 190–201, 2006.

[Del11] Martin Delacourt. Rice's Theorem for μ-Limit Sets of Cellular Automata. In *ICALP (2)*, pages 89–100, 2011.

[dMS11] Benjamin Hellouin de Menibus and Mathieu Sablik. Self-organization in Cellular Automata: A Particle-Based Approach. In *Developments in Language Theory*, pages 251–263, 2011.

[DPST11] Martin Delacourt, Victor Poupet, Mathieu Sablik, and Guillaume Theyssier. Directional dynamics along arbitrary curves in cellular automata. *Theor. Comput. Sci.*, 412(30):3800–3821, 2011.

[Fis90] Robert Fisch. The one-dimensional cyclic cellular automaton: a system with deterministic dynamics that emulates an interacting particle system with stochastic dynamics. *J. Theoret. Probab.*, 3(2):311–338, 1990.

[FMMN00] Pablo A. Ferrari, Alejandro Maass, Servet Martínez, and Peter Ney. Cesàro mean distribution of group automata starting from measures with summable decay. *Ergodic Theory Dynam. Systems*, 20(6):1657–1670, 2000.

[Gác01] Peter Gács. Reliable cellular automata with self-organization. *J. Statist. Phys.*, 103(1-2):45–267, 2001.

[HM10] Michael Hochman and Tom Meyerovitch. A characterization of the entropies of multidimensional shifts of finite type. *Ann. of Math. (2)*, 171(3):2011–2038, 2010.

[Hoc09] Michael Hochman. On the dynamics and recursive properties of multidimensional symbolic systems. *Invent. Math.*, 176(1):131–167, 2009.

[KM00] Petr Kůrka and Alejandro Maass. Limit sets of cellular automata associated to probability measures. *J. Statist. Phys.*, 100(5-6):1031–1047, 2000.

[Ků05] Petr Kůrka. On the measure attractor of a cellular automaton. *Discrete Contin. Dyn. Syst.*, (suppl.):524–535, 2005.

[Maz96] Jacques Mazoyer. On optimal solutions to the firing squad synchronization problem. *Theoret. Comput. Sci.*, 168(2):367–404, 1996. Universal machines and computations (Paris, 1995).

[Mey11] Tom Meyerovitch. Growth-type invariants for $\mathbb{Z}^d$ subshifts of finite type and arithmetical classes of real numbers. *Inventiones Mathematicae*, 184:567–589, 2011.

[MM98] Alejandro Maass and Servet Martínez. On Cesàro limit distribution of a class of permutative cellular automata. *J. Statist. Phys.*, 90(1-2):435–452, 1998.

[Pet83] Karl Petersen. *Ergodic theory*, volume 2 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1983.

[PY02] Marcus Pivato and Reem Yassawi. Limit measures for affine cellular automata. *Ergodic Theory Dynam. Systems*, 22(4):1269–1287, 2002.

[Shi96] Paul C. Shields. *The ergodic theory of discrete sample paths*, volume 13 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.

[Wei00] Klaus Weihrauch. *Computable analysis*. Texts in Theoretical Computer Science. An EATCS Series. Springer-Verlag, Berlin, 2000. An introduction.

[You02] Lai-Sang Young. What are SRB measures, and which dynamical systems have them? *J. Statist. Phys.*, 108(5-6):733–754, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

[Zie05] Martin Ziegler. Computability and Continuity on the Real Arithmetic Hierarchy and the Power of Type-2 Nondeterminism. In *CiE*, pages 562–571, 2005.

[ZW01] Xizhong Zheng and Klaus Weihrauch. The arithmetical hierarchy of real numbers. *MLQ Math. Log. Q.*, 47(1):51–65, 2001.

LATP, Aix-Marseille Université, 39, rue F. Joliot Curie,13453 Marseille Cedex 13, France
E-mail address: benjamin.hellouin@gmail.com
URL: http://www.latp.univ-mrs.fr/∼hellouin/

LATP, Aix-Marseille Université, 39, rue F. Joliot Curie,13453 Marseille Cedex 13, France
E-mail address: sablik@latp.univ-mrs.fr
URL: http://www.latp.univ-mrs.fr/∼sablik/