Characterisation of limit measures of higher-dimensional cellular automata

Martin Delacourt, Benjamin Hellouin de Menibus

September 30, 2015

Cellular automata are discrete dynamical systems defined by a local rule, introduced in the 40s by John von Neumann [15]. They model a large variety of discrete systems and are linked with various areas of mathematics or computer science, in particular computation theory, complex systems, ergodic theory and combinatorics.

One of the main catalysts of the study of cellular automata was their surprisingly complex and organised behaviours, even when iterated on configurations with no particular structure (e.g. chosen at random). To formalise these observations, many authors tried to describe their asymptotic behaviour by considering the limit set, which is the set of configurations that can be reached after arbitrarily many steps. These sets were shown to have potentially high computational complexity [13, 1], and any nontrivial property on them is undecidable [11]. These observations built a bridge between the variety of dynamical behaviours and the computational content of the model. Nevertheless, the problem of characterising which subshifts can be limit sets of CA remains open.

In 2000, Kůrka and Maass argued that limit sets did not provide a good description of empirical observations and introduced instead a measure-theoretical version of these sets [12]. The idea of μ-limit sets is to choose the initial configuration at random, according to some probability measure μ, to iterate the cellular automaton on this configuration and to consider all patterns whose probability to appear does not tend to 0. In the one-dimensional case, this approach yielded similar results of high complexity and undecidability [4, 3, 6, 2]. Althought these two families of results appear similar and both require sophisticated constructions inside cellular automata, they provide insight about different kinds of dynamics (topological vs. measure-theoretical) and computational power (deterministic vs. probabilistic).

In [5], H. and Sablik extended this approach to consider the limit probability measure(s). Still in the one-dimensional case, they provided a computational characterisation of the limit measures reachable by cellular automata, generalising the previous results.

This article is an extended version of [7]. In *op.cit*, we aimed at extending the previous results to the two-dimensional setting. More precisely, we characterised all subshifts that can be μ-limit sets of CA when μ is the uniform Bernoulli measure. The proof works by an explicit construction inspired by the one-dimensional constructions of [2, 5], althought the higher dimensional setting has many specific challenges. In the present article, this two-dimensional construction is generalised to d-dimensional space for any $d > 2$; furthermore, through a more careful analysis, we are able to characterise reachable limit measures, which is a more general result. As a corollary, we obtain an undecidability result on properties of limit measures, and cover as well Cesàro mean convergence and the case where the limit measure is not unique.

1 Definitions

1.1 Symbols, configurations and cellular automata

Let $\mathcal{A}$ be a finite set of symbols called *alphabet*. For $d > 0$, let $\mathcal{A}^{\mathbb{Z}^d}$ be the space of d-dimensional *configurations*.

On $\mathbb{Z}^d$, we define the basis vectors $e_i = (\delta_i(k))_{0 \leq k \leq d}$ (Kronecker deltas), that is, the vector worth 0 on all coordinates except the i-th where it is worth 1. Denote $\mathcal{U}nit(d) = \{\sum_{1 \leq j \leq d} \delta_j e_j \neq 0 : \forall j, \delta_j \in \{-1, 0, 1\}\}$ and $\mathcal{H}yp(d)$ the set of hyperplanes that have a normal vector in $\mathcal{U}nit(d)$; these hyperplanes have a basis of $d-1$ vectors in $\mathcal{U}nit(d)$.

We will use the ∞ and 1-distance between points of $\mathbb{Z}^d$:

$$\forall z_1, z_2 \in \mathbb{Z}^d, \qquad d_\infty(x, y) = \max_{1 \leq i \leq d} |x_i - y_i| \qquad \text{and} \qquad d_1(x, y) = \sum_{1 \leq i \leq d} |x_i - y_i|.$$

An ∞-path (respectively 1-path) between two points of $\mathbb{Z}^d$, is a sequence of points $z_1, \ldots, z_k$ such that $d_\infty(z_i, z_{i+1}) = 1$ for any i (d_1 respectively). An ∞-connected set (resp. 1-connected) is a subset of $\mathbb{Z}^d$ such that any pair of points are connected by an ∞-path (resp. 1-path).

If we endow $\mathcal{A}^{\mathbb{Z}^d}$ with the product topology of the discrete topology on $\mathcal{A}$, then $\mathcal{A}^{\mathbb{Z}^d}$ is compact, perfect and totally disconnected. This topology is also metrisable, for example using the *Cantor metric*:

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

For a subset $U \subset \mathbb{Z}^d$, denote $x_U \in \mathcal{A}^U$ the restriction of x to U. Denote $\mathcal{A}^* = \bigcup_{\substack{U \subset \mathbb{Z}^d \\ finite}} \mathcal{A}^U$ the set of finite *patterns*. For a pattern $w \in \mathcal{A}^U$, denote its *support* $\mathrm{supp}(w) = U$, and its *dimension* is the smallest d such that $\mathrm{supp}(w) \subset \mathbb{Z}^d$. We say a pattern is *cubic*, respectively *rectangular*, if its support is a $d-cube$, resp. a $d-box$ (Cartésian product of intervals).

The *cylinder* defined by a pattern $u \in \mathcal{A}^*$ and a position $i \in \mathbb{Z}^d$ is $[u]_i = \{x \in \mathcal{A}^{\mathbb{Z}^d} \ :\ x_{i+\mathrm{supp}(u)} = u\}$. For simplicity we sometimes write $[u]$ for $[u]_{(0,\ldots,0)}$.

Given two patterns $u \in \mathcal{A}^U$ and $v \in \mathcal{A}^V$, the *frequency of u in v* is defined as:

$$\mathrm{freq}(u, v) = \frac{\{i \in V \ :\ i + U \subset V, v_{i+U} = u\}}{\{i \in V \ :\ i + U \subset V\}} \qquad \text{if defined, 0 otherwise}$$

The *shift map*, or *shift*, is defined as:

$$\forall i \in \mathbb{Z}^d,\ \sigma_i(x) = (x_{i+j})_{j \in \mathbb{Z}^d}.$$

A *subshift* is a closed subset of $\mathcal{A}^{\mathbb{Z}^d}$ invariant under all shifts. Given a cubic pattern $u \in \mathcal{A}^{[0,n-1]^d}$, define the *periodic configuration* ${}^\infty u^\infty$ by ${}^\infty u^\infty_{[0,n-1]^d} = u$ and $\sigma^n_{e_k}({}^\infty u^\infty) = {}^\infty u^\infty$ for every $k \in [1, d]$.

A *cellular automaton* (or CA) is a continuous function $F : \mathcal{A}^{\mathbb{Z}^d} \to \mathcal{A}^{\mathbb{Z}^d}$ that commutes with all shifts ($F \circ \sigma_{e_k} = \sigma_{e_k} \circ F$ for every k). By the Curtis-Hedlund-Lyndon theorem [8], it can be defined equivalently as a function $F(x) = (f((x_j)_{j \in i + \mathbb{N}}))_{i \in \mathbb{Z}^d}$ where $\mathcal{N} \subset \mathbb{Z}^d$ is a finite set called *neighbourhood* and $f : \mathcal{A}^{\mathbb{N}} \to \mathcal{A}$ is called a *local rule*.

1.2 Probability measures on $\mathcal{A}^{\mathbb{Z}^d}$

Let $\mathfrak{B}$ be the Borel sigma-algebra of $\mathcal{A}^{\mathbb{Z}^d}$ and $\mathcal{M}(\mathcal{A}^{\mathbb{Z}^d})$ the set of probability measures on $\mathcal{A}^{\mathbb{Z}^d}$ defined on the sigma-algebra $\mathfrak{B}$. In this article, we focus on $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ the set of *σ-invariant probability measures* on $\mathcal{A}^{\mathbb{Z}^d}$, that is to say, the measures μ such that $\mu(\sigma_k^{-1}(B)) = \mu(B)$ for all $B \in \mathfrak{B}$ and $k \in \mathbb{Z}^d$. Cylinders corresponding to finite patterns form a base of the topology, and furthermore $\mu([u]_i) = \mu([u])$ for a measure

$\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$. Therefore μ is entirely characterised by $\{\mu([u]) \ : \ u \in \mathcal{A}^*\}$; actually, considering only cubic patterns is enough.

We endow $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ with the *weak** (or *weak convergence*) topology: a sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ converges to $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ if and only if, for all patterns $u \in \mathcal{A}^*$, one has $\lim_{n\to\infty} \mu_n([u]) = \mu([u])$. In the weak* topology, $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ is compact and metrisable. A metric is defined by

$$d_{\mathcal{M}}(\mu,\nu) = \sum_{n\in\mathbb{N}} \frac{1}{2^n} \max_{u\in\mathcal{A}^{[0,n]^d}} |\mu([u]) - \nu([u])|.$$

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

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

Let us define some examples that we use throughout the article.

The *Bernoulli measure* μ_λ associated to some vector $\lambda = (\lambda_a) \in [0;1]^{\mathcal{A}}$ satisfying $\sum_{a\in\mathcal{A}} \lambda_a = 1$ 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 $x \in \mathcal{A}^{\mathbb{Z}^d}$ is defined as $\delta_x(A) = \mathbf{1}_{x\in A}$. Generally δ_x is not σ-invariant. However, for any cubic pattern $w \in \mathcal{A}^{[0,n]^d}$, it is possible to define the *σ-invariant measure supported by* ${}^\infty w^\infty$ by taking the mean of the Dirac measures on the orbit under σ:

$$\widehat{\delta_w} = \frac{1}{|\operatorname{supp} w|} \sum_{i\in[0,n]^d} \delta_{\sigma_i({}^\infty w^\infty)}.$$

The set of measures $\left\{\widehat{\delta_w} : w \in \mathcal{A}^*\right\}$ is dense in $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ [14].

1.2.1 Action of a cellular automaton on $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ and limit measure

Let $F : \mathcal{A}^{\mathbb{Z}^d} \to \mathcal{A}^{\mathbb{Z}^d}$ be a cellular automaton and $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$. 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 $F_*(\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})) \subset \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$. This defines a continuous application $F_* : \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d}) \to \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$.

We consider in particular $F_*^t\mu$ the iterated image of μ by F_*. Since $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ is compact in the weak$*$ topology, the sequence $(F_*^t\mu)_{t\in\mathbb{N}}$ admits a set of limit points denoted $\mathcal{V}(F,\mu)$ and called the *μ-limit set of measures of* F. When $\mathcal{V}(F,\mu)$ is a singleton, i.e. when $F_*^t\mu \underset{n\to\infty}{\longrightarrow} \nu$, we say ν is the *limit measure* of F starting on μ.

2 Computability

We now introduce the computability notions that are needed to state our main results. This exposition is very similar to the one that can be found in [10], which was later expanded in [9], for the one-dimensional case. Indeed, most of the definitions and proofs only rely on the fact that the space is metric and separable, properties for which the increase in dimension is irrelevant. Therefore, we omit the proofs that can be obtained by a straightforward substitution ($\mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}^d}$) from the proofs found in *op.cit.*

2.1 Turing machines

Turing machines are a standard and robust tool to define the computability of mathematical operations. In the usual model, they have access to a one-dimensional, one- or two-sided infinite memory tape. In order

to simplify some constructions, we consider in this article that the tape is d-dimensional and infinite in all directions. This does not affect the computing power of the model.

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 d-dimensional infinite memory tape is filled with #, except for a finite region (the input), and a computing head is located at coordinate $(0, \dots, 0)$;
- Q a finite set of states, with an initial state $q_0 \in Q$;
- $\delta : (Q \cup \#) \times \Gamma \to (Q \cup \#) \times \Gamma \times \{\pm e_i\}_{1 \leq i \leq d}$ the transition function. Given the current state and the letter it reads on the tape — which depends on its current position — the function returns the new state, the letter to be written on the tape at current position, and the vector by which the head moves.
- $Q_F \subset Q$ the set of final states — when a final state is reached, the computation stops and the output is the contents of the tape.

A function $f : \mathcal{A}^* \to \mathcal{A}^*$ is *computable* if there exists a Turing machine working on an alphabet $\Gamma \supset \mathcal{A}$ that, on any input $w \in \mathcal{A}^*$, eventually stops and outputs $f(w)$.

2.2 Computability of functions mapping countable sets

To generalise this definition to functions mapping arbitrary countable sets $X \to Y$, we need to define an *encoding*, that is, an alphabet $\mathcal{A}_X$ together with a bijection between X and some subset of $\mathcal{A}_X^*$, and similarly for Y. Then the computability of a function $X \to Y$ is defined up to some encoding. However, in practice, reasonable encodings yield the same notion of computability. To simplify notations, we fix some canonical encodings for the rest of the paper :

$\mathbb{Z}$ (or $\mathbb{N}$): Take $\mathcal{A}_\mathbb{Z} = \{0, 1\}$ and encode an element $k \in \mathbb{Z}$ as its binary expansion surrounded by blank symbols.

Product $X \times Y$: Take $\mathcal{A}_{X \times Y} = \mathcal{A}_X \times \mathcal{A}_Y$ and encode (x, y) as the product of encodings for x and y.

Using this last case, we define a canonical encoding for $\mathbb{Q}$ as the canonical encoding for $\mathbb{N} \times \mathbb{Z}$, up to the bijection $\frac{p}{q} \mapsto (p, q)$ (with p, q irreducible).

Furthermore, we define the computability of a set $K \subset X$ as the computability of the function $1_K : X \to \mathbb{N}$.

2.3 Computability of probability measures

As we mentioned above, a probability measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ is entirely described by the value of $\mu([u])$ for $u \in \mathcal{A}^*$. In other words, an element of $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ is described by a function $\mathcal{A}^* \to \mathbb{R}$. Since $\mathbb{R}$ is not countable, the standard ways to define notions of computability is to consider approximations by elements of $\mathbb{Q}$.

A measure $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ is *computable* if there exists a computable function $f : \mathcal{A}^* \times \mathbb{N} \to \mathbb{Q}$ such that

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

It is *limit-computable* if there exists a computable function $f : \mathcal{A}^* \times \mathbb{N} \to \mathbb{Q}$ such that

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

Additionally we define the notion of a *uniformly computable sequence*. Infomally, it means that a sequence of objects can be computed by a single algorithm which, given $n \in \mathbb{N}$ as input, returns a description of a n-th object of the sequence.

Formally, a sequence of measures $(\mu_i)_{i\in\mathbb{N}}$ is *uniformly computable* iff there exists $f : \mathcal{A}^* \times \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$ computable such that:

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

It is easy to see that the limit of a uniformly computable sequence of measures is limit-computable (it is not necessarily computable since the rate of convergence of μ_i to μ is not known).

Proposition 1 (Approximation by measures supported by periodic orbits)**.**
These notions can be defined in another equivalent way:

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

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

Notice the parallel with the definition of a computability of a real: in both cases, an object is computable if it is approximated by a uniformly computable sequence of elements taken from a dense subset ($\mathbb{Q}$ and the measures supported by periodic orbits, respectively) with a known rate of convergence.

2.4 Action of a cellular automaton on computable measures

Proposition 2 (First computability obstruction)**.** *Let $F : \mathcal{A}^{\mathbb{Z}^d} \to \mathcal{A}^{\mathbb{Z}^d}$ be a cellular automaton and $\mu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ be a computable measure. Then $(F_*^t\mu)_{t\in\mathbb{N}}$ is a uniformly computable sequence of measures. In particular, if $F_*^t\mu \underset{t\to\infty}{\longrightarrow} \nu$ then ν is limit-computable.*

In general, $F_*^t\mu$ does not have a single limit point, but a compact set of accumulation points. To obtain a similar obstruction, we extend our computability definitions to those objects.

2.5 Compact sets in computable analysis

Extending naively the definition for countable sets using the characteristic function does not work since the set of inputs would not be countable. Instead, we use a general definition for metric spaces that possess a countable dense subset, $(\widehat{\delta_w})_{w\in\mathcal{A}^*}$ in the case of $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$.

A closed set $\mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ is *computable* if the countable set $\left\{(w,r) \in \mathcal{A}^* \times \mathbb{Q} \;:\; \overline{\mathbf{B}(\widehat{\delta_w}, r)} \cap \mathcal{K} \neq \emptyset\right\}$ is computable, that is, if its characteristic function is.

However, the set of limit points of the sequence $(F_*^t\mu)_{t\in\mathbb{N}}$, where μ is computable, is not necessarily computable (or even limit-computable). We need to extend our definitions even further, obtaining an arithmetical hierarchy. We introduce these notions first on countable spaces, then on closed subsets of $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$.

Let X, Y be two countable sets, with Y being ordered. A sequence of functions $(f_i : X \to Y)_{i\in\mathbb{N}}$ is *uniformly computable* if $(i,x) \mapsto f_i(x)$ is computable.

A function $f : X \to Y$ is Π_2-*computable* (resp. Σ_2-*computable*) if $f = \inf_{i\in\mathbb{N}} \sup_{j\in\mathbb{N}} f_{i,j}$ (resp. $f = \sup_{i\in\mathbb{N}} \inf_{j\in\mathbb{N}} f_{i,j}$), where $(f_{i,j})_{(i,j)\in\mathbb{N}^2}$ is a uniformly computable sequence of functions.

A closed set $\mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ is Π_2-*computable* if the set

$$\left\{(w,r) \in \mathcal{A}^* \times \mathbb{Q} \;:\; \overline{\mathbf{B}(\widehat{\delta_w}, r)} \cap \mathcal{K} \neq \emptyset\right\}$$

is Π_2-computable, that is, its characteristic function is.

Remark. The symmetric notions of Π_2- and Σ_2-computability come from an analogy with the real arithmetic hierarchy [16, 17]. These definitions extend naturally to Π_n- and Σ_n-computability. Other equivalent definitions exists, see for example [10] for Π_2-computability or [9] for a more general result.

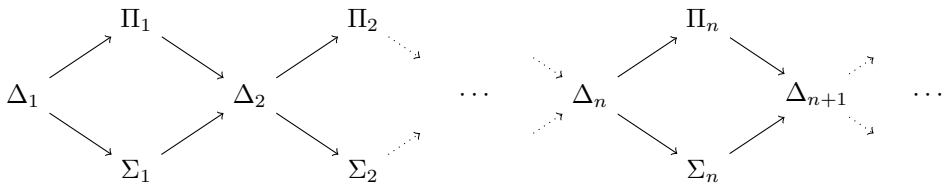

Figure 1: Representation of the arithmetical hierarchy. Arrows indicate strict inclusion relations.

Proposition 3 (Second computability obstruction)**.**
Let $F : \mathcal{A}^{\mathbb{Z}^d} \to \mathcal{A}^{\mathbb{Z}^d}$ be a cellular automaton and μ be a computable measure. Then $\mathcal{V}(F, \mu)$ is a nonempty Π_2-computable compact set.

Reciprocally, Π_2-computable compact sets can be all be described as the set of limit points of a uniformly computable sequence of measures $(w_n)_{n\in\mathbb{N}}$. However, our construction cannot do better that following such a sequence along a polygonal path, that it, along segments of the form $\left[\widehat{\delta_{w_i}}, \widehat{\delta_{w_{i+1}}}\right] = \left\{t\widehat{\delta_u} + (1-t)\widehat{\delta_v} : t \in [0,1]\right\}$. The following proposition shows that this corresponds to connected limit sets of measures.

Proposition 4 (Technical characterisation of Π_2-CCC sets)**.**
Let $\mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ be a non-empty Π_2-computable, compact, connected set (Π_2-CCC for short). Then there exists a uniformly computable sequence of cubic patterns $(w_n)_{n\in\mathbb{N}}$ such that $\mathcal{K}$ is the limit of the polygonal path defined by $(w_n)_{n\in\mathbb{N}}$, that is,

$$\mathcal{K} = \bigcap_{N>0} \overline{\bigcup_{n\geq N} \left[\widehat{\delta_{w_n}}, \widehat{\delta_{w_{n+1}}}\right]}.$$

3 Construction

In order to obtain the results announced in introduction, in conjunction to Proposition 4, we need the following result.

Theorem 1. *For any uniformly computable sequence $(w_n)_{n\in\mathbb{N}}$ of cubic patterns of $\mathcal{B}^*$ of dimension at most d, there exists a larger alphabet $\mathcal{A} \supset \mathcal{B}$ and a cellular automaton $F : \mathcal{A}^{\mathbb{Z}^d} \to \mathcal{A}^{\mathbb{Z}^d}$ such that:*

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

In this section, we present the construction of the alphabet $\mathcal{A}$ and cellular automaton F, given some alphabet $\mathcal{B}$ and some uniformly computable sequence $(w_n)_{n\in\mathbb{N}}$ of patterns of $\mathcal{B}^*$.

3.1 Sketch of the construction

We detail the construction of $\mathcal{A}$ and F by describing the tasks to be performed on the initial configuration. Each letter of $\mathcal{A}$ is a product of seven layers separated in three groups, each group representing some information needed to perform a given task. The alphabet of each layer contains a special *blank* symbol # to denote the absence of information.

- The first group is dedicated to the colonising of the configuration. Since we have no control over the contents of the initial configuration, we want to erase (almost) all letters present initially in favor of various processes that we can control and synchronize. To do this, the *birth layer* contains a *seed*

symbol $\boxed{*}$ that can only appear in the initial configuration. Each seed gives birth to a stationnary *heart* $\boxed{\heartsuit}$ on the same layer, and to a *membrane* on the *growth layer* which grows in every direction. As it grows, the membrane erases everything in its path, except for other membranes issued from a seed with which it merges.

- The second group is used to divide the colonised space into mostly independant areas called *organisms*, each organism having at its center a heart issued from a seed. The borders between organisms are redefined regularly by processes on the *organism layer*. Furthermore, organisms need to grow in size regularly, which is achieved by merging organisms whose hearts get too close using the *evolution layer*.
- The third group deals with the internal metabolism of the organisms. The goal is first to compute each w_n in succession, which is achieved by simulating a Turing machine in the *computing layer*; then, the main layer (defined below) of the whole body of the organism is filled with concatenated copies of the output by using a copying process taking place on the *copying layer*. The above is done synchronously, at some time t_n, in all the organisms.

Finally, the *main layer* is where copies of w_n are written, which implies that the corresponding alphabet is $\mathcal{B} \cup \{\#\}$. To sum up, the global alphabet is $\mathcal{A} = \mathcal{A}_{\text{birth}} \times \mathcal{A}_{\text{growth}} \times \mathcal{A}_{\text{orga}} \times \mathcal{A}_{\text{evol}} \times \mathcal{A}_{\text{comp}} \times \mathcal{A}_{\text{copy}} \times (\mathcal{B} \cup \{\#\})$. We check that $\mathcal{B} \subset \mathcal{A}$ up to the bijection $b \mapsto (\#, \#, \#, \#, \#, \#, b)$. Denote p_{birth}, p_{growth}, p_{orga}, p_{evol}, p_{comp}, p_{copy}, p_{main} the projections on each coordinate.

During the description of F, we will treat each layer successively. The layers were introduced in order of dependency, in the sense that the time evolution of symbols in a given layer only depends on the contents of layers in the same group and the one immediatly preceding it. Furthermore, the main layer is only affected by the writing layer.

3.2 Space colonisation: Seeds and membranes

In this section, we describe the cleaning of the configuration through the seeds and the birth, growth and fusion of membranes. We deal only with the birth layer and alphabet $\mathcal{A}_{\text{birth}}$ for the moment. Section 3.2.1 gives the general ideas of the process, while Section 3.2.2 focuses on technical difficulties of the cellular implementation.

3.2.1 Creation myth: a sketch

Birth and growth and membranes This alphabet $\mathcal{A}_{\text{birth}}$ contains the seed symbol $\boxed{*}$, which can only appear in the initial configuration since it cannot be produced by the local rule of F. At the first step, each seed spawns a number of processes and turns into a heart $\boxed{\heartsuit} \in \mathcal{A}_{\text{birth}}$. This heart and those processes (and those spawned from them) are called *initialised*, which means that their behaviour is well controlled and synchronised (since they are all born at time 1). All other symbols are *uninitialised*.

If two seeds are too close from each other (d_∞ less than 5), the largest (in lexicographic order) is erased to give enough space to the other seed to spawn its processes. A seed that is not destroyed at time 1 this way is called *viable*. By abuse of notation we write $p_{\text{birth}}(c_x) = \boxed{*}^V$ to mean that the configuration c has a viable seed at coordinate x.

Each occurence of $\boxed{*}$ triggers the birth at time 1 of a living *membrane*. The membrane consists in membrane symbols $\boxed{\square}$ that form initially the surface of an hypercube of edge length 5 centred on the seed. The membrane is oriented, being able to distinguish inside from outside through arrow labels. Furthermore, to each membrane symbol is associated an *age counter*, which is a binary counter initialised at 0 and increasing at each step, whose aim is to keep track of the age of the membrane. Notice that in an initialised membrane all age counters are equal.

From there on, the membrane grows slowly towards the outside, erasing the content of other layers as it progresses with the exception of other membranes. This is governed by the *respiration process*: each time the age stored in its counter is the square of an integer, the membrane grows to the outside, making one step in every direction. Again, this supposes that storage, incrementation and decrementation of counters are

instantaneous, we will remedy this problem in the next section. Figure 2 represents some part of a membrane with arrows and counters.

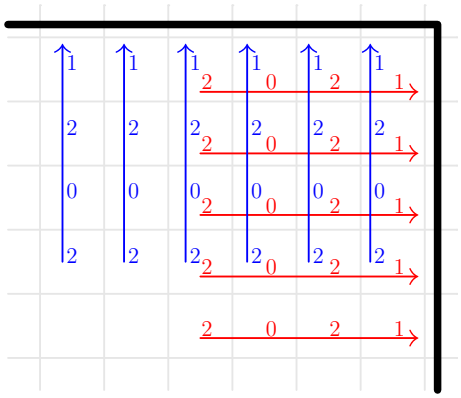

Figure 2: A corner of a membrane in dimension 2. The arrows give the orientation and the counters store the age of the membrane.

Fight for survival When the growing membranes meet other membrane symbols, they try to determine locally whether they are part of an initialised membrane (in which case the two should merge), or some uninitialised symbols which should be erased. We call *dead* an uninitialised group of membrane symbols that present some locally detectable malformation, such as non-connexity, the absence of or inconsistence between age counters/respiration processes, inconsistence between inner and outer orientation for neighbors, etc. In this case, the malformation generates a death signal [☠] that spreads through the whole membrane erasing it. However, such a group can also form a *zombie membrane*, that is apparently well-formed though uninitialised. Initialised (living) and zombie membranes are distinguished through age counters.

Fact 1. *At time t, all age counters associated with a well-formed membrane have value at least $t-1$, the minimum being reached only for initialised membranes.*

Indeed, age counters of initialised membranes are initialised at 0 at time 1, while age counters of zombie membranes were already present (with a positive value) at time 0, and both are incremented by 1 at each step.

3.2.2 Membranes and age counters

This section is dedicated to the details of the cellular implementation of the ideas exposed in the previous section. First we show how to implement age counters with binary counters using logarithmic space.

As we saw, $\mathcal{A}_{\text{birth}}$ contains a blank state # and symbols for seeds [*], hearts [♥], and membranes. Each membrane symbol [‖] contains an outward orientation label consisting of a vector of $\mathcal{U}nit(d)$.

Definition. A membrane m at time t is a maximal 1-connected set of coordinates containing membrane symbols [‖] at time t with consistent outward orientation; i.e., orientation of neighbouring membrane symbols differ in at most one coordinate, and at most by 1.

When a membrane forms a closed curve (which is the case for initialised membranes), we denote $\mathbf{Supp}(m)$ its support of m. In this case, $\mathbf{Supp}(m)$ partitions $\mathbb{Z}^d$ into a finite set $\overline{\mathbf{Int}}(m)$ and an infinite set $\mathbf{Ext}(m)$ which are ∞-connected. We also denote $\mathbf{Int}(m) = \overline{\mathbf{Int}}(m) \setminus \mathbf{Supp}(m)$. By "outward" in the previous definition, we mean that the orientation vectors of m are directed towards $\mathbf{Ext}(m)$. Notice this cannot be checked locally, but any locally detectable malformation spawns a death process.

To implement all counters, we use a redundant binary basis:

t	*phase*	A	B	*age*	*breath*	$5+2\lfloor\sqrt{t}\rfloor$
1	$+$	1	0	0		5
2	$-$	0	1	1		5
3	$-$	1	0	2		5
4	$+$	2	0	11	$\times$	7
5	$+$	$1\overline{1}$	1	12		7
6	$-$	0	2	21		7
7	$-$	1	$1\overline{1}$	102		7
8	$-$	2	0	111		7
9	$+$	11	0	112	$\times$	9
10	$+$	10	1	121		9

Figure 3: Values of the three counters for $t \leq 10$.

Definition (Redundant binary basis)**.** Let $c = \underline{c_{n-1}} \dots c_0 \in \{0, 1, 2, \overline{1}\}^n$ be a counter whose most significant bit is marked. The **value** of c is $\sum_{i=0}^{n-1} c_i 2^i$ (reverse order) where $\overline{1}$ has value -1. Since $2 = 10$, 2 can be seen as a 0 with a carry, and $\overline{1}$ as a 0 with a "negative" carry.

At each time step, carries are propagated along the counter, which can be done in a local manner ($02 \to 10, 12 \to 20, \#2 \to 10$). If the counter is incremented by one, which is the case for age counters, the rule is adapted at the least significant bit at the end of the counter ($0 \to 1, 1 \to 2, 2 \to 1$). Decrementation through negative carries is done in a symmetric manner ($1\overline{1} \to 01, 0\overline{1} \to \overline{1}1$) except that we erase additional zeroes at the beginning of the counter ($\underline{1\overline{1}} \to \#\underline{1}$).

The age counters are implemented in this way. The least significant bit of each counter is next to its corresponding membrane symbol, and the following bits lie on a line directed towards the inside of the membrane. To each possible direction (corresponding to some $\pm e_j$) corresponds a different sublayer, which allows counters to cross near the corners. Thus the age counters use $2d$ sublayers, each sublayer containing symbols $\{\overline{1}, 0, 1, 2\}$.

Recall that any inconsistency such that the absence of an age counter for some membrane symbols or parallel counters containing different symbols spawns a death counter, which spreads in the whole membrane and erases all layers of these cells.

3.2.3 The respiration process

The goal of the respiration process is to govern a slow growth of the membrane. Along with the age counter, on two other sublayers of $\mathcal{A}_{\text{birth}}$, to counters A and B are initialised at time 0 with values 1 and 0, respectively. From there on two phases alternate, the current phase being labelled on the membrane symbol:

Phase $+$ A is decremented while B is incremented, until A reaches 0. The phase passes to $-$;

Phase $-$ A is incremented while B is decremented, until B reaches 0. On the next step A is incremented once more while a *breath* is triggered (explained below), then the phase passes to $+$.

$A + B$ is constant on a cycle and a cycle takes a total time $2(A + B) + 1$, after which the sum $A + B$ is incremented by 1. Therefore a breath occurs at each time t^2 for $t > 1$ (the starting time being 1). Each breath makes the membrane progress by one cell on every direction, from which we can see that the membrane forms an hypercube of edge length $5+2\lfloor\sqrt{t}\rfloor$ at time t. Furthermore, the counters of an initialised membranes are initialised to 0 at time 1, hence the maximum size of the counters is $\lceil \log_2(t-1) \rceil$ at time t.

In Figure 3 we represent the update operation of all three counters, that is, incrementing the age and updating A and B according to the phase.

Lemma 1. *The counter update can be performed locally with radius* 2.

Proof. The incrementations and decrementations described above can be achieved with radius 1. The least significant bit can be distinguished by being next to the membrane symbol which contains the information on the current phase and the most significant bit is next to a blank symbol (on its layer).

We show that the fact that a counter is worth 0 is locally detectable (to see this is nontrivial, consider the update $\underline{1}\overline{1}\underline{1}\ldots\overline{1} \to \#00\ldots0$). During a decrementation the least significant bit alterns between 0 and $\overline{1}$. Since they progress at "speed" one, two negative carries can never be next to each other. Therefore the only possible representations of 1 are $\underline{1}$ and $\underline{1}\overline{1}$, and both yield $\underline{0}$ at the next step. Therefore detecting when the counter is worth 0 requires radius two, in order to "see" the 0 and the # state indicating this is the most significant bit. □

In Lemma 2, we quantify the maximal number of extensions of a membrane in a given time.

Lemma 2. *The number of breaths triggered for any membrane symbol between times t and $t+k$ is at most $\lfloor\sqrt{t+k}\rfloor - \lfloor\sqrt{t}\rfloor$.*

Proof. Notice this lemma is not restricted to initialised membranes. Apart from time 0 (when a breath symbol could be present), a breath is only triggered when the B counter of a membrane symbol without local malformations reaches 0. This symbol must be issued from a seed or from a membrane symbol already present at time 0. In the first case, since a breath is triggered at each step when the time t is the square of an integer (except for 1), the number of breaths before time t is $\lfloor\sqrt{t}\rfloor - 1$. The lemma follows.

In the second case, the membrane symbol had at time 0 counters A and B with some positive values a_0 and b_0 and some phase ε_0, values which correspond to those of an initialised set of counters at some time $t_0 > 0$. From there on the time evolution of the membrane symbol is similar to the evolution of an initialised membrane symbol of age $t_0 + t$, which means that the number of breaths between times t and $t+k$ is $\lfloor\sqrt{t+t_0+k}\rfloor - \lfloor\sqrt{t+t_0}\rfloor \leq \lfloor\sqrt{t+k}\rfloor - \lfloor\sqrt{t}\rfloor$. □

Breathing process Now we describe the effect of a breath symbol on the membrane. If such a symbol is not produced synchronously by the whole membrane, then ☠ signals spawn and spread to erase the membrane.

Recall that each membrane symbol at coordinates x is labelled with an outward growth direction, which is a vector $v = \sum \varepsilon_j e_j \in \mathcal{U}nit$. The membrane symbol is the border between $\mathbf{Int}(m)$ and $\mathbf{Ext}(m)$, with the orientation vector indicating which part $\mathbf{Ext}(m)$ is. If a breath occurs for some symbols but not all, a death process is triggered. Otherwise, the membrane symbols are removed and new symbols are created on all cells of $\mathbf{Ext}(m)$ that were ∞-adjacent to any membrane symbol of m. The new orientation vectors are found by remaining coherent with the old orientation.

The remaining task is to reproduce the counters for the new symbols.

First consider the case of a face symbol x and orientation e_j. Right after a breath, when a new symbol is created at coordinate $x + e_j$, the symbol at x is replaced by a placeholder symbol s_{lim}. This symbol progressively shifts the counters of x by one cell in direction e_j, marking at each step the limit between the part which is to be shifted and the part already shifted. The counters keep updating by ignoring this symbol, which increase the radius to 3. Figure 4 illustrates the shift.

For an initial configuration c, define $M_t(c)$ be the set of initialised membranes at time t. Then the *colonised space* at time t is:

$$\mathbf{Col}_t(c) = \bigcup_{m \in M_t(c)} \overline{\mathbf{Int}}(m).$$

When a membrane grows, it erases the content of every other layer of the cells it encounters, except when the birth layer contains the outer border of a membrane. In this case, the comparison process starts, which is the topic of the next section.

Hence alphabet $\mathcal{A}_{\text{birth}}$ contains the seed, the blank state and the states used for membranes. As we will see in the next section, we actually need to allow 2^d different membranes to share the same cell.

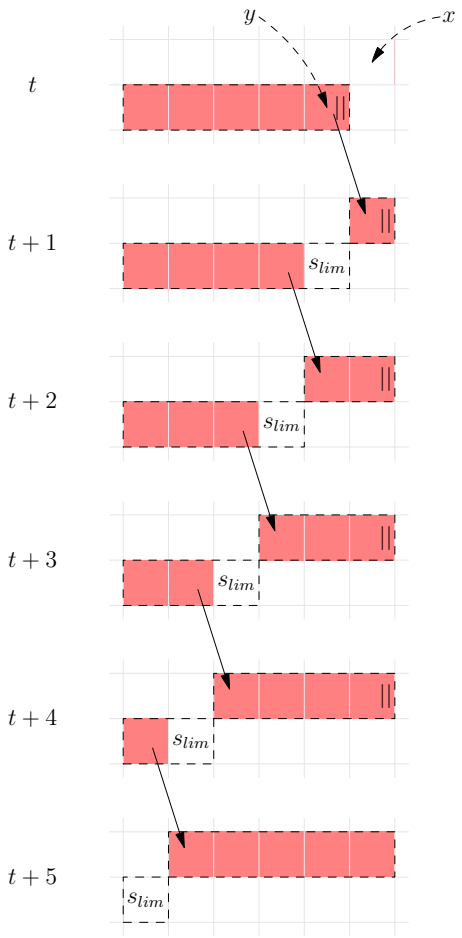

Figure 4: After a breath, a membrane symbol at cell y is erased and new membrane symbols appear at x and x'. At each step between $t+1$ and $t+4$, the red cells represent superposition of the age, A and B counters. They are copied to the new membrane symbols, but the incrementation does not stop.

3.2.4 Forming colonies

As membranes grow and tend to cover the whole space, different membranes eventually meet. The result of the encounter should depend on the nature of the membranes: two initialised membranes should merge while an initialised membrane should erase an uninitialised membrane (what happens between uninitialised membrane is irrelevant). In this section, we devise a comparison process to distinguish initialised from uninitialised membranes. We now consider the growth layer and its alphabet $\mathcal{A}_{\text{growth}}$.

The process aims at comparing the value of the age counters to let the younger membrane survive, with merging occuring in case of equality (see Lemma 1.

When we say that two membranes m and m' meet at time t in cells x and x', we mean that there exists $1 \leq j \leq d$ such that:

- either $x \in \mathbf{Supp}(m) \cap \mathbf{Supp}(m')$, $x + e_j \in \mathbf{Ext}(m)$ and $x - e_j \in \mathbf{Ext}(m')$, in which case take $x' = x$;
- or $x \in \mathbf{Supp}(m) \cap \mathbf{Ext}(m')$ and $x + e_j \in \mathbf{Supp}(m') \cap \mathbf{Ext}(m)$, in which case take $x' = x + e_j$.

The two possible situations are illustrated in Figure 5.

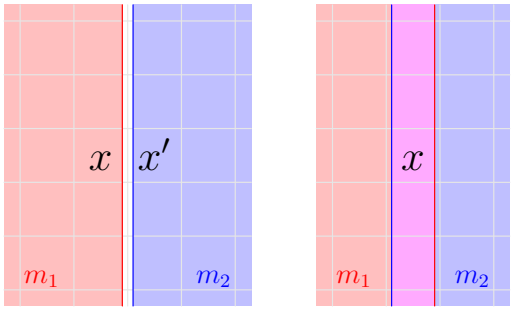

Figure 5: Depending on the parity of the distance between membranes, they will meet either when they share some border cells ($x' = x$), or when the borders are adjacent($x' = x + e_j$).

In particular, the membranes arriving from opposite directions, they have (at least) an age counter in opposite directions, say e_j and $-e_j$. The idea is now to copy these age counters on the growth layer and compare them. At positions x and x', two symbols C_j^1 and C_-^1 are written on the growth layer to trigger the process (if $x = x'$, a symbol $C_\pm^1$ represents the superposition of those symbols).

At the next step, both symbols begin to progress at speed one in the corresponding direction, copying at each step one bit from the age counter. Carries 2 are copied as 0: indeed, the copy is performed at the same speed as the carry progresses, so the carry will be taken into account one it meets a 0 to turn into a 1 (otherwise it would be copied at each step). More generally, only carries that appeared before the beginning of the copy can influence the copied bits. Thus we see that the copied counter (that is not incremented) have the same value as the age counter at the beginning of the copy. When it reaches the end of its counter, each copy symbol turns into a comparison symbol C_j^2 (resp. C_-^2), which triggers the comparison phase.

In the comparison phase, the comparison symbols return towards the meeting point, "pushing" in front of them the copied bits in a catterpillar-like movement, starting from the most significant bit. The returning bits use a third sub-layer. The process is represented in Figure 6.

As the returning bits reach the meeting point, one of the following situations occur:

- the most significant bit from one side arrives earlier than the most significant bit from the other side. In this case the age counter of the corresponding side is shorter, which means that the membrane of this side is younger;
- both most significant bits arrive simultaneously at the meeting points x and x'. Then bits are compared as they arrive. The first bit that is smaller than its counterpart corresponds to the side of the younger membrane;

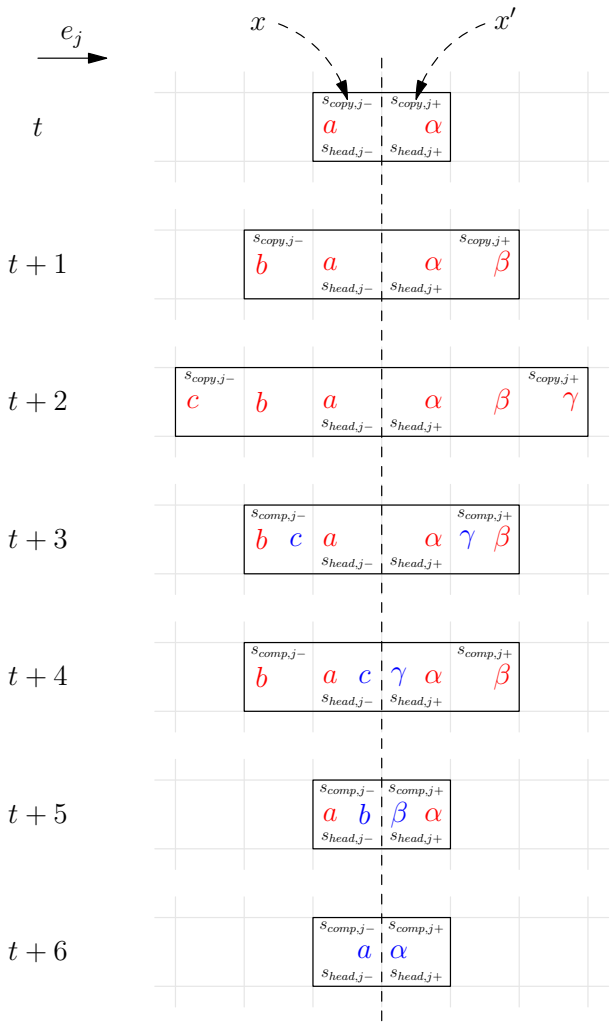

Figure 6: In this example, two membranes meet in cells x and x' at time t. Their age counters are respectively abc and $\alpha\beta\gamma$ of same length 3. Only the growth layer is represented. At the end $(t+6)$, the decision can be made in both x and x'. In this particular case, the symbols s_{head} have not been moved, which means neither of the membranes did extend during the comparison.

- in the previous case, if all bits are equal until the end, both membranes are exactly as old.

Those three possibilities are tested locally at the meeting point and the result is written at the meeting point under the form of a symbol (on its own sublayer) marking the direction of the younger membrane, with $=$ in case of a tie. If for some reason a ☠ process reaches the symbol of one of the sides, the comparison stops and the surviving membrane is marked as younger "by default".

If a membrane is declared younger, all auxiliary symbols used for comparison are erased and a death process triggers in the older membrane. The younger membrane will resume its growth naturally at the next breath. If the result is a tie, both membrane symbols are erased along with all associated auxiliary states: the membrane are merged.

Remark.

- By Lemma 1, initialised membranes are always youngest, and only tie with other initialised membranes.
- Two membranes may meet with more than one meeting point. In that case, comparisons are performed simultaneously at every point and in every concerned direction. In the case of a tie, all symbols participating in the meeting should be erased simultaneously; any local discrepancy results in the spawn of a death process.
- In the worst case 2^d different membranes can meet in the same cell x (corners of hypercube arriving from all possible directions). To solve this problem, we duplicate each sublayer (in $\mathcal{A}_{\text{birth}}$ and $\mathcal{A}_{\text{growth}}$) used in the comparison into 2^d copies, each copy being able to perform a comparison independently of the others. If the membrane is older than at least another membrane, a death process is spawned; similarly for the tie case.
- Let ℓ be the length of the shortest age counter. The previous process needs ℓ steps to copy this age counter on the growth layer, and 2ℓ steps to send them one by one to the meeting point. Regardless of the length of the other counter, the comparison reaches a result after the last bit of the shortest counter arrives. Therefore the whole process takes at most 2ℓ steps, with $\ell = \lceil \log t - 1 \rceil$ if one of the membranes is initialised (where t is the time at the beginning of the process).

A possibility we did not take into account is that one of the membranes breaths (grows) during the comparison. For each meeting of a pair of membranes, call *instigating membrane* the one whose breath has triggered the meeting (possibly both if they moved simultaneously; this is the case for initialised membranes).

Lemma 3. *Let m be a living membrane meeting another membrane at time t. During the comparison process, m may breath at most one time if it is not instigating, and cannot breath at all if m is instigating.*

Proof. Using the above remark, we know that the comparison process takes at most $k = 3\lceil \log t - 1 \rceil$ steps. Using Lemma 2, the number of breaths of m between times $t-1$ and $t+k$ is at most:

$$\begin{aligned}
\left\lfloor \sqrt{t+k} \right\rfloor - \lfloor \sqrt{t-1} \rfloor &\leq \left\lceil \sqrt{t+k} - \sqrt{t-1} \right\rceil \\
&\leq \left\lceil \frac{k}{2\sqrt{t-1}} \right\rceil \quad \text{(since the derivative of } \sqrt{t-1} \text{ is decreasing)} \\
&\leq \left\lceil \frac{3\log t - 1}{2\sqrt{t-1}} \right\rceil \leq 1.
\end{aligned}$$

If m is instigating, then by definition m breathed at time $t-1$ and cannot breath again before time $t+k$. Otherwise, m breathes at most one time. □

Therefore, if during the comparison, one or both membranes move, due to the respiration process presented earlier, it is enough to give the information of the extension to the head $s_{head,j+}$ written on the second layer. Indeed, as we juste proved in Lemma 3, there cannot be more than one extension during a comparison involving a living membrane. Therefore, if a membrane extends more than twice before the end of the comparison, the ☠ state is written in this membrane. As the radius of F is 2, the border of the membrane has always immediate access to the result of the comparison.

Lemma 4. *Take any $t > 0$ and initial configuration $c \in \mathcal{A}^{\mathbb{Z}^d}$. Then:*

$$\mathbf{Col}_t(c) = \{x \in \mathbb{Z}^d : \exists y \in \mathbb{Z}^d, d_\infty(x, y) \leq 1 + \sqrt{t}, p_{\text{birth}}(c_y) = \boxed{*}^V\}.$$

In other words, the colonised space at time t is exactly the set of cells that, at time a, are at distance less than $1 + \sqrt{t}$ from a viable seed.

Proof. We prove this result by structural induction. If $t = 1$, then the colonised space is the set of all initialised membranes which are hypercubes of length 5 around each viable seed, and the result is proved.

Now suppose that the hypothesis holds at time t. Notice than $\mathbf{Col}_t(c) \subset \mathbf{Col}_{t+1}(c)$, and that merging does not add any cell to the colonised space: an initialised membrane cannot be erased, and the colony obtained after merging two membranes is the union of the colonies defined by the two merged membranes. Only the breathing process may add new cells to the colonised space.

Consequently, the induction step is empty if $(t + 1)$ is not a square since $\mathbf{Col}_t(c) = \mathbf{Col}_{t+1}(c)$ and $d_\infty(x, y) \leq 1 + \sqrt{t} \Leftrightarrow d_\infty(x, y) \leq 1 + \sqrt{t+1}$ (distances are integers). If $(t+1)$ is a square, then all membrane symbols in initialised membranes take a breath and extend by one cell towards the outside. Now, take a cell y at distance $1 + \sqrt{t+1}$ from the nearest viable seed. By the induction hypothesis, $y \notin \mathbf{Col}_t(c)$, but y has a neighbour $y + v$ with $v \in \mathcal{U}nit$ at distance $1 + \sqrt{t}$ for that seed, so that $y + v \in \mathbf{Col}_t(c)$. Therefore $y + v$ must be a membrane symbol in the support of an initialised membrane that breathes at time $t + 1$, and therefore $y \in \mathbf{Col}_{t+1}(c)$. Conversely, if a cell z is at distance greater than $1 + \sqrt{t+1}$ from the nearest viable seed, it cannot have a membrane symbol belonging to an initialised membrane as a neighbour, so that $z \notin \mathbf{Col}_{t+1}(c)$. □

3.3 Colonies: evolution of the population

Thanks to Lemma 4, we only have to consider initialised membranes and colonies. In this section, we describe the interaction of organisms inside colonies. In all the following, we assume we are inside a colony, and the support of the surrounding membrane acts as an impassable wall for any symbols in the second group layers: the organism and evolution layers.

3.3.1 Hearts and organisms

This section describes the organism layer, and every state presented here belongs to $\mathcal{A}_{\text{orga}}$.

As we saw before, seeds $\boxed{*}$ at time 1 spawn a membrane and turn into hearts $\boxed{\heartsuit}$. Each heart will be the center of an *organism* which is itself a subset of the colony. At first each colony have only one heart, but as initialised membranes merge together, the colonies become multi-hearted, and the colony space should be partitioned into organisms (except possibly a negligible part). For various reasons, the size of the organisms should grow in a controlled way, which requires some hearts to be progressively removed.

In the present section, we present the cycle of division of colony space and life of the organisms.

The life of an organism consists in a succession of generations. We introduce a sequence of times $(t_n)_{n\geq 1}$ (to be fixed later), marking the limit between the $n-1$-th and nth generation. Time is tracked by the heart through a binary *time counter*, initialised at 1 at time 1 (along with the heart) and remaining stationnary next to the heart. Details on the implementation and the way to determine when $t = t_n$ will be given in Section 3.4.1.

At time t_n, *organism-building signals* spread from every heart, progressing as membrane symbols but with speed 1 (although they do not have any counters). While progressing, they erase the old contents of the second and third group layers (but not the main layer). When they meet a membrane or another organism-building signal, they vanish leaving behind a *neutral border symbol* $\boxed{\$}$. For reasons of parity, if two signals emitted by hearts in x and x' arrive simultaneously in two neighbor cells y and y', they receive a *pseudo border symbol* $\boxed{\$'}$ that contains an orientation towards the interior of the corresponding organism, that is the opposite of the direction of the organism-building signal. Just as membrane symbols, $3^d - 1$ different organism-building symbols and pseudo border symbols are required (one for each orientation).

The *territory* of a heart is the largest set of 1-connected cells containing the heart and no neutral border symbol $\boxed{\$}$ nor pseudo border symbol pointing towards another organism, that is the set of cells reached first by organism-building signals emitted by this heart. At time $t_n + k$ (assuming $t_n + k < t_{n+1}$), the only cells of the colony that are not part of some organism are either at distance more than k from the nearest heart, or were outside the membrane at time t_n (and a breath had included them since).

Fact 2. *Let x be a cell containing a heart at time t with $t_n \leq t \leq t_{n+1}$, and let y be a cell in its territory. Then the organism-building signal emitted by x reached y at time $t_n + d_\infty(x, y)$ and no other organism-building signal reached a neighbor of y before that time.*

The following lemma gives insight about the shape of the territories, namely, that they are a (discrete) star domain, whether you include the borders or not.

Lemma 5. *If a cell y belongs to the territory of a heart in cell x, then each cell y' such that $d_\infty(x, y') + d_\infty(y', y) = d_\infty(x, y)$ is also in this territory. Furthermore, y' can be a border only if y is a border.*

Proof. For the first part of the lemma, suppose such a y' is not in the territory of x. We can build an ∞-path between y' and y consisting of cells $(y^{(i)})_{0 \leq i \leq K}$ such that $d_\infty(x, z) + d_\infty(z, y) = d_\infty(x, y)$. Take $y^{(j)}$ the first $y^{(i)}$ that belongs to the territory of x.

Denote $T = t_n + d_\infty(x, y^{(j-1)})$ the time when the organism-building signal emitted by x should have reached $y^{(j-1)}$ in the absence of any other heart. Since $y^{(j-1)}$ is not in the territory of x, there must exist another heart x' that emitted an organism-building signal that arrived in $y^{(j-1)}$ before time T (recall that the pseudo-borders are considered as parts of organisms).

Since $y^{(j)}$ is adjacent to $y^{(j-1)}$, $y^{(j)}$ is reached by some signal before time $T + 1$. But the signal from x cannot reach $y^{(j)}$ before time $t_n + d_\infty(x, y^{(j)}) = T + 1$. Therefore $y^{(j)}$ is not in the territory of x, a contradiction.

For the second case, notice that y' is a border if and only if both signals reached this cell simultaneously. Then the same reasoning along the path $(y^{(i)})_{0 \leq i \leq K}$ shows that these cells cannot be inside the territory of x. □

3.3.2 Natural selection

In this section we consider the evolution layer and the alphabet $\mathcal{A}_{\text{evol}}$.

To have enough computation space and ensure that the auxiliary symbols are in negligible density, the minimal size of the organisms should grow regularly. More precisely, we require that the territory of any organism during the n-th generation contains at least a hypercube of side $2n + 1$ centred at its heart. If two hearts are at distance less than $2n + 1$, they are said to be in *conflict*. In this section, we devise a selection process to detect this fact and to erase one of them.

In order to detect conflicting organisms, we draw a *body*, that is a hypercube of side $2n + 1$, around each heart. Two hearts are in conflict exactly when their bodies intersect. To draw the body of an organism, we use the current value of n that is kept under the form of a binary counter on the computation layer (see next section). At time t_n, (this is detected by the computation layer), the heart builds an hypercube of side length 3 consisting of *body-building* symbols. Then the heart sends $n - 1$ successive impulses (hypercube-shaped signals) that progress at speed 1 in every direction and push the body-building signals outward. The count is kept through a decrementing binary counter, as usual.

Whenever two hearts conflict, a death process triggers for the one in the smallest cell (in lexicographic order). In this case, a death symbol spreads through the entire body cells (not the inside), erasing the selection layer. When this process reaches a body *corner symbol* (defined as in Section 3.2.2, but for body symbols), the corner sends a *heartbreak signal* at speed 1 in the direction opposite to its orientation (this is a unique feature of corner symbols). The heart, after receiving d heartbreak symbols (the current number being kept track of on a dedicated fixed-length binary counter), self-destroys. Not being able to send border-building signals, its territory will be occupied by other organisms at the next phase.

Thanks to this process we can bound the *radius* of an organism, which is the largest distance from a cell of its territory to its heart.

Lemma 6. *For any constant $K > 1$, denoting $K_n = K^{n^{d-\frac{1}{2}}}$,*

$$\max_{t_n \le t \le t_{n+1}} \mu\left(\left\{c \in \mathcal{A}^{\mathbb{Z}^d} : \exists x \in \mathbb{Z}^d, d_1(x,0) \le K_n, p_{\text{orga}}(F^t(c)_x) = \boxed{\heartsuit}\right\}\right) \to_{n\to\infty} 1.$$

Proof. We can assume 0 is in colonised space. Since the shape of organisms only change at times t_n, it is enough to show that the sets:

$$\Gamma_n^D = \left\{c \in \mathcal{A}^{\mathbb{Z}^d} : \exists x \in \mathbb{Z}^d, d_1(x,0) \le D, p_{\text{orga}}(F^{t_n}(c)_x) = \boxed{\heartsuit}\right\}$$

are such that $\mu(\Gamma_n^{K_n})$ tend to 1 as n tends to infinity.

Take any $n > 0$. At time t_n, only hearts that are at distance $2n$ or $2n+1$ of another heart are potentially destroyed. For $c \in \mathcal{A}^{\mathbb{Z}^d}$ and $z \in \mathbb{Z}^d$, we define the *n-heart chain* in c starting from z inductively:

- $z_0 = z$ assuming that $F^{t_{n-1}}(c)_z = \boxed{\heartsuit}$ (otherwise the chain is empty);
- Assuming z_n is defined for some $n > 0$,
  - if there exist coordinates $z' > z_n$ such that $d_\infty(z_n, z') \in \{2n, 2n+1\}$ and $F^{t_{n-1}}(c)_{z'} = \boxed{\heartsuit}$, define z_{n+1} as the maximal such z';
  - otherwise, z_{n+1} is undefined and the chain stops.

Lemma 7. *For k and $L < D$ positive integers such that $D - L > 2k+1$,*

$$\mu\left(c \notin \Gamma_k^D \mid c \in \Gamma_{k-1}^{D-L}\right) \le \left[1 - \left(1 - \mu\left([\boxed{*}^V]\right)\right)^{d(2k+1)^{d-1}}\right]^{\frac{L}{2k+1}}.$$

Proof of Lemma 7. If $c \in \Gamma_k^{D-L}$, there exists $z \in \mathbb{Z}^d$ such that $d(0,z) < D - L$ and $F^{t_{k-1}}(c)_z = \boxed{\heartsuit}$; denote $m_{k-1}(c)$ one such z minimising the distance $d(0,z)$ and $(z_i^k(c))$ the k-heart chain in c starting from $m_{k-1}(c)$. Notice that 0 cannot be part of this chain.

For any i, the heart at $z_i^k(c)$ (if defined) is destroyed at time t_k if, and only if, $z_{i+1}^k(c)$ is defined. Furthermore by straightforward induction we have $d_\infty(z_i^k(c), 0) \le D + (2k+1)i$. Consequently, to get $c \notin \Gamma_k^D$, the chain must contain at least $\frac{L}{2k+1}$ elements.

For any i,

$$\begin{aligned}
\mu\left(z_{i+1}^k(c) \text{ is defined } \mid z_i^k(c) \text{ is defined }\right) &= \mu\left(\exists z' > z_i^k(c), d_\infty(z_i^k(c), z') \in \{2n, 2n+1\} \wedge F^{t_{k-1}}(c)_{z'} = \boxed{\heartsuit}\right) \\
&\le \mu\left(\exists z' > z_i^k(c), d_\infty(z_i^k(c), z') \in \{2n, 2n+1\} \wedge c_{z'} = \boxed{*}^V\right) \\
&\le 1 - (1 - \mu([\boxed{*}^V]))^{d(2k+1)^{d-1}}
\end{aligned}$$

In the second step, we used the fact that a heart can only exist if a viable seed was present at this coordinate at time 0. The third step comes from the fact that the measure is Bernoulli and that the considered coordinates form the (two-cell thick) surface area of a half-hypercube of edge length $2k+1$.

To conclude, we apply this computation inductively, using the fact that at each step of the chain, the coordinates considered in the computation were never considered in a previous step. Together with the fact that μ is a Bernoulli measure, we obtain that the events $\{z_{i+1}^k(c) \text{ is defined } \mid z_i^k(c) \text{ is defined }\}$ are independant. □

We come back to the proof of Lemma 6. Take some $c \notin \Gamma_n^{K_n}$. By the pidgeonhole principle, one of the following is true:

- $c \notin \Gamma_0^{K_n/n}$, or
- $\exists i < n,\ c \in \Gamma_i^{iK_n/n}$ and $c \notin \Gamma_{i+1}^{(i+1)K_n/n}$.

By Birkhoff's theorem, since $\mu([\boxed{*}]^V) > 0$, $\mu(c \notin \Gamma_0^{K_n/n}) \underset{n\to\infty}{\longrightarrow} 0$. Using Lemma 7, we obtain:

$$\begin{aligned}\mu\left(c \notin \Gamma_n^{K_n}\right) &\leq \sum_{i<n} \left[1 - \left(1 - \mu\left([\boxed{*}^V]\right)\right)^{d(2i+1)^{d-1}}\right]^{\frac{K_n}{n(2i+1)}} + o(1)\\ &\leq n \cdot \left[1 - (1 - \mu([\boxed{*}^V]))^{d(2n+1)^{d-1}}\right]^{\frac{K_n}{n(2n+1)}} + o(1)\end{aligned}$$

and since $K_n \sim K^{n^{d-\frac{1}{2}}}$, we can see that $\log \mu\left(c \notin \Gamma_n^{K_n}\right) \leq \log n + \frac{1}{n^2} K^{n^{d-\frac{1}{2}}} (1 - \mu([\boxed{*}^V]))^{d(2n+1)^{d-1}} \to -\infty$, from which we deduce that $\mu\left(c \notin \Gamma_n^{K_n}\right) \to 0$. □

The same method can be used to show that, around any heart and with asymptotic probability 1, other hearts can be found at distance at most $K^{n^{d-\frac{1}{2}}}$ in each quadrant (sets of the form $\{x \in \mathbb{Z}^d : \forall i,\ \varepsilon_i x_i > 0\}$ for some $\varepsilon_i = \pm 1$). This is enough to show that, with a probability tending to 1, the central cell belongs to an organism whose radius is less than K_n.

Definition. An organism is *healthy* is its radius is less than $K_n = K^{n^{d-\frac{1}{2}}}$.

3.4 Individual organisms: internal metabolism

The last group of layers is used to govern the internal metabolism of the organisms. In this section, we consider some organism and describe how it behaves during a generation.

3.4.1 Computing

In this section, we use symbols of the alphabet $\mathcal{A}_{\text{comp}}$ in the computational layer. Let $(w_n)_n$ be the uniformly computable sequence of patterns given as an hypothesis of the theorem. Our goal is to delimit a small computation space around the heart where each w_n will be computed in succession.

We use standard techniques to embed the time evolution of any Turing machine $TM = (Q, \Gamma, \#, q_0, \delta, Q_F)$ inside our cellular automaton. We use for this the alphabet $(\Gamma \cup \#) \times (Q \cup \#)$: the left part contains the tape symbol, and the right part contains the current state for the cell where the head is located, and # everywhere else. Then each step of the Turing machine moves the head and modifies the tape around the head according to local information, which can be done through the local rule of a CA.

In our case, the alphabet $\mathcal{A}_{\text{comp}}$ is divided into 3 sublayers, each one simulating in parallel a Turing machine with a d-dimensional tape. The TM of each sublayer may access to the contents of the tape of another sublayer when indicated. Let us assume we are at time $t = t_n$, that the first TM has the current value of t as a binary *time counter* on its tape, and the second one the current value of n (*generation counter*). We describe the behaviour of the each machine:

1. The first machine increment t at each step to keep the time value updated. This is similar to the binary age counters, but we will see that the time counter is folded in a square, which makes this task less trivial.

2. The second machine computes the value of t_{n+1} (not modifying the generation counter), then keep watch on the time counter on the first tape. When the time counter reaches t_{n+1}, the machine increments the generation counter by one, which triggers many other processes.

3. The third machine reads the value of n on the second tape, then computes the cubic pattern w_n along with its side length k.

We want to ensure that these computations can be performed between times t_n and t_{n+1} and without leaving the hypercube centred on the heart of side length $2\left\lceil n^{\frac{d-\frac{1}{4}}{d}}+1\right\rceil$, that is, that they can be performed in time $t_{n+1}-t_n = K^{n^{d-\frac{1}{4}}}$ and space $(2n)^{d-\frac{1}{4}}$.

By hypothesis, the patterns w_n are assumed cubic. Without loss of generality, we also assume that $w_n \in \mathcal{A}^{[0,k]^d}$ for some $\frac{1}{2}n^{\frac{d-\frac{1}{2}}{d}} < k \leq n^{\frac{d-\frac{1}{2}}{d}}$ (by concatenating copies of w_n if necessary), and that they are computable in time $O(K^{n^{d-\frac{1}{4}}})$ and space $O(n^{d-\frac{1}{4}})$. This is done by repeating each pattern w_n in the sequence until the next pattern satisfies those constraints. Furthermore the time counter occupies a space $\lceil \log t \rceil \leq \lceil \log t_{n+1} \rceil \sim \log(K^{n^{d-\frac{1}{4}}}) = O(n^{d-\frac{1}{4}})$, and computing the value of $t_{n+1} = \sum_{k<n} K^{k^{d-\frac{1}{4}}}$ takes the same space and time $O(K^{n^{d-\frac{1}{4}}})$.

To get rid of the multiplicative constant contained in the O notation, we use the standard techniques of linear speedup and tape compression for Turing machines. For any fixed constant C, by grouping cubes C^d tapes cells together in a single letter and performing C computation steps at once, we can divide required time and space by C. Of course the tape alphabet of the Turing machines increases exponentially (ic C).

Therefore the described computations are doable within these time and space constraints. w_n is computed before time t_{n+1}, and at time t_{n+1} the second machine enters a special states that triggers various processes: organism-building signals, body-building, and the object of the next section, a copying process that will write concatenated copies of the pattern w_n all over the territory of the organism.

The alphabet $\mathcal{A}_{\text{comp}}$ is thus $(\Gamma_1 \cup \#) \times (Q_1 \cup \#) \times (\Gamma_2 \cup \#) \times (Q_2 \cup \#) \times (\Gamma_3 \cup \#) \times (Q_3 \cup \#)$, where Q_i, Γ_i are the state space and the tape alphabet of (the compressed version of) the i-th Turing machine described above.

3.4.2 Copying

The second task of an organism is to write concatenated copies of the previously computed pattern on the whole territory of the organism. In this section, auxiliary symbols belong in the copy layer $\mathcal{A}_{\text{copy}}$ but the pattern is written in the main layer with alphabet $\mathcal{B}$.

Remind that we assume $w_n \in \mathcal{A}^{[0,k]^d}$ for some $\frac{1}{2}n^{\frac{d-\frac{1}{2}}{d}} < k \leq n^{\frac{d-\frac{1}{2}}{d}}$. The global copying process relies on a cubic grid of side length k that covers the whole territory of the organism. Starting from the cells centred on the heart, the pattern w_n is copied in each cell of this grid passing from neighbour to neighbour, through a translation of vector ke_j or $-ke_j$ for each $1 \leq j \leq d$.

First denote x the coordinates of the central heart, and assume that the borders of the computed pattern w_n are marked with a special symbol $\boxed{G}$.

For a set of coordinates $i \in \mathbb{Z}^d$, define the corresponding grid element $\Sigma_i = \{\sum_{1\leq j\leq d} \alpha_j e_j : \forall 1 \leq j \leq d, ki_j \leq \alpha_j \leq ki_j + k\}$, and $\overline{\Sigma_x}$ its border (extremal cells). Notice that the computed pattern is supported by $\Sigma_{0,\dots,0}$, and $\overline{\Sigma_{0,\dots,0}}$.

For $u \in \mathcal{U}nit(d)$ and $i \in \mathbb{Z}^d$, we define the *copying operation* $\mathcal{C}_i(u)$ that copies the contents of the main layer from Σ_i to Σ_{i+u}. It consists of simulating a Turing machine (see previous section) accomplishing the following steps:

Reproducing the borders: The first step is to write $\boxed{G}$ in every cell of $\overline{\Sigma_{i+u}}$. The Turing machine determines the value of k by counting the side length of $\overline{\Sigma_i}$. It then travels to the coordinate $k(i+u)$ and builds an hypercube of symbols $\boxed{G}$ of side length k (Σ_{i+u}). This takes $O(k^2)$ time steps. If the new hypercube is not entirely included in the territory of the organism, the process stops there.

Reproducing the pattern: The second step is to copy the pattern letter by letter. The machine copies each letter in lexicographic order, marking with a symbol letters already copied. Each letter needs at most $O(k)$ steps to be copied, so the whole process takes $O(k^3)$ steps.

Cleaning the auxiliary states: The third step is to remove all the auxiliary states that remain on the tape in the original grid hypercube Σ_i (including $\boxed{G}$). This is done by going through all k^2 cells of Σ_i, taking $O(k^2)$ steps.

Selectionning heirs: The hypercube Σ_{i+u} will in turn spawn new copy processes. The rule to carry on is the following:

$$\mathcal{C}_i(u) \to \begin{cases} \{C_{i+u}(v) : v = e_j + \sum_{k\neq j} \lambda_k e_k, \lambda_k \in \{-1,0,+1\}\} & \text{if } u = e_j \\ C_{i+u}(u) & \text{otherwise} \end{cases}$$

Those new processes are performed in parallel by duplicating the copy layer 2^d times.

At the initial step, it is enough to trigger a copy process in all directions $u \in \mathcal{U}nit(d)$. The copying operations then progressively fill the whole organism, as can be seen in Figure 7. To do: Il faudrait changer w'_n en w_n. Aussi, peut-etre montrer comment chaque flèche se dédouble ?

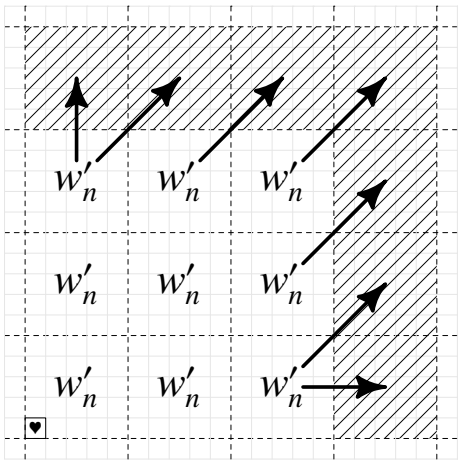

Figure 7: In this 2-dimensional example, the pattern w_n is copied from the heart of the organism towards its boundaries in successive steps.

Each copying operation takes $O(k^3)$ steps, and the active copying operations expand outward from the heart as a (thick) hypercube. Therefore, if the radius of the organism is r, the total time needed to finish the copying process is $\frac{r}{k} \cdot O(k^3)$. We can take $r \leq K^{n^{d-\frac{1}{2}}}$ by Lemma 6 and $k \leq n^{\frac{d-\frac{1}{2}}{d}}$, which gives a total time of $O(n^2 K^{n^{d-\frac{1}{2}}}) = O(K^{n^{d-\frac{1}{4}}})$. Lowering if needed the multiplicative constant by the linear speedup theorem, we see that the process ends before time t_{n+1}.

3.5 Proof of the main theorem

We first prove that the density of auxiliary states tend to 0 as time tends to infinity, which ensures they are not charged by any limit measure.

Lemma 8. *Borders have negligible density asymptotically, i.e.,*

$$F^t\mu\left(\left\{c \in \mathcal{A}^{\mathbb{Z}^d} : p_{\text{orga}}(c)_0 = \boxed{\$}\right\}\right) \to_t 0$$

Proof. By Lemma 4, we can consider only the cells in the colonised space, i.e. inside a living membrane. Given a configuration c and some time $t_n \leq t < t_{n+1}$ during the n-th generation, denote $S_{\boxed{\$}}(t) = \{x \in \mathbb{Z}^d :$

$p_{\text{orga}}(F^t(c)_x) = \boxed{\$}\} \cap \mathbf{Col}_t(c)$ the set of colonised cells containing a border, and $\overline{S}_{\boxed{\$}}(t)$ the complement of the previous set. We show that there exists a constant λ such that $\frac{|S_{\boxed{\$}}(t)|}{|\overline{S}_{\boxed{\$}}(t)|} \leq \frac{\lambda}{n}$. This property being true for every initial configuration c, the lemma follows using Birkhoff's theorem.

The idea is to partition the borders between organisms into (subsets of) hyperplanes. To each such hyperplane subset bordering two given organisms, we associate some volume inside the territory of one of the organisms that is large enough (linear in n). Practically, we prove that any planar surface included in the border is at distance at least n of one of the hearts.

Lemma 9. *The common border of two hearts is included in a (finite) union of hyperplanes of $\mathcal{H}yp(d)$.*

Proof. Let x and x' be two cells containing a heart each. Assuming no other heart existed in the space, the border between these two hearts would be included in the union of $H^{\varepsilon}_{k,n}$ sets, where $H^{\varepsilon}_{k,n}$ are defined as:

$$y \in H^{\varepsilon}_{k,n} \iff \begin{cases} |2y_n - x_n - x'_n| \leq 1 & \text{if } k = n \text{ and } x_n = x'_n \\ |y_k - x_k - \varepsilon(y_n - x'_n)| \leq 1 & \text{otherwise} \end{cases}$$

All these sets are unions of one or two hyperplanes of $\mathcal{H}yp(d)$ (depending on the parity of $x_k - \varepsilon x'_n$). In the presence of other hearts, the border between x and x' is a subset of this "ideal border", which proves the Lemma. □

Given two hearts at cells x_0 and x_1, denote $B(x_0, x_1)$ the set of cells corresponding to their common border. Partition this set into a finite collection $\{H_1, \ldots, H_k\}$ of disjoint subset of hyperplanes according to the previous fact. For each such H_i, as $d_1(x_0, x_1) \geq 2n$, we have either $d_1(x_0, H_i) \geq n$ or $d_1(x_1, H_i) \geq n$. Denote $A(s)$ the area of s and V_0 and V_1 the volumes of the d-polytopes limited by the surface s and respectively the points H_0 and H_1. And $V_\iota = \frac{1}{d} d_1 H_\iota, PA$ for each $\iota \in \{0, 1\}$. Denote $V(s) = V_0 + V_1$, then $\frac{A(s)}{V_s} \leq \frac{d}{n}$.

We now do this operation for every organism, that is split $S_{\boxed{\$}}(t)$ into a collection $\mathcal{S}$ of disjoint hyperplanar surfaces that belong to the common border of two organisms. For every two such different surfaces, the corresponding volumes inside organisms are also disjoints, hence

$$\frac{|S_{\boxed{\$}}(t)|}{|\overline{S}_{\boxed{\$}}(t)|} \leq \frac{\sum_{\mathcal{S}} A(s)}{\sum_{\mathcal{S}} V(s)} \tag{1}$$

$$\leq \frac{d}{n} \tag{2}$$

□

Lemma 10. $\forall z \in \mathbb{Z}^d$,

$$\mu\left(F^t(c)_z \in \mathcal{A} \setminus \mathcal{B}\right) \underset{t \to \infty}{\longrightarrow} 0.$$

Proof. We handle each layer separately.

Uncolonised space and membranes First, by Lemma 4, we can see that c_0 can belong to the uncolonised space at time t only if the nearest viable seed at time 0 is at distance more than $\sqrt{t}$. For the same reason, c_0 can be part of a living membrane or a related process (age counter, respiration process, comparison process) only if the nearest viable seed is at distance more than $\sqrt{t} - \log t$. Since a viable seed appear with a nonzero probability, the probability of this event tends to 0 as t tends to infinity.

It remains to handle symbols appearing inside the colonised space on the layers bla. By the ergodic theorem, it is equivalent to prove that the density of auxiliary states in a configuration tends to 0 almost surely when time tends to infinity.

Hearts, computing symbols In the colonised space hearts $\boxed{\heartsuit}$ must be issued from a seed, and as explained in Section bla they each have at time T_n a body, which are non-overlapping hypercubes of side $2n+1$ centered on the heart (more precisely, they can overlap shortly but are destroyed before the next T_n). Thus the density of hearts $\boxed{\heartsuit}$ in c between times T_n and T_{n+1} is almost surely less than $\frac{1}{(2n-1)^d}$. Since the computing process taking place around the heart is contained in an hypercube of side $\sqrt{n}$, the density of cells with nonempty computing layer is almost surely less than $\frac{1}{(2n-1)^{d/2}}$ in this period.

Bodies and bodybuilding signals As for the symbols for the body of the heart, they form the surface of an hypercube of side $2n+1$ (when it is fully built) or less (during the construction), and therefore there are less than $2d(2n+1)^{d-1}$ such symbols for each heart. The impulses used to grow the body being sent one at a time, they occupy at most as much space as the body itself at any given time. Therefore all those symbols have density less than $2d(2n+1)^{d-1} \cdot \frac{1}{(2n-1)^d} = O\left(\frac{1}{n}\right)$.

Borders and border-building signals Borders $\boxed{\$}$ were handled in Lemma 8. We use a similar argument to show that the density of symbols in signals used to build borders is asymptotically negligible. The signal is born around the heart and progresses at speed one. Therefore, m steps after its birth, the set of cells in the organism containing the signal is an hypercube of side $2m+1$ centred on the heart (intersected with the inside of the organism). In particular, in an organism of healthy size, the signal sent at time t_n has disappears before time $t_n + K^{n^d} \leq t_{n+1}$, so at most one signal appear in a give organism at the same time.

If $m \leq n$, since the organism contains at least n^d cells, the signal appear with frequency less than $\frac{2d(2m+1)^{d-1}}{n^d} = O(\frac{1}{n})$. If $m > n$, notice that for each cell z of the organism satisfying $d_\infty(\boxed{\heartsuit}, z) = m$, by Lemma 5 the entire line between $\boxed{\heartsuit}$ and z is inside the organism and does not contain the signal (since its distance to the heart is less than m). Therefore if a certain proportion of the surface area of the hypercube is inside the organism (and therefore contains symbols), the same proportion of the volume of the hypercube is inside the organism as well (without symbols), from which we deduce that the symbol density is at most $\frac{2d(2m+1)^{d+1}}{m^d} = O\left(\frac{1}{n}\right)$ (since $m > n$).

Copying processes The copying grid $\boxed{G}$ is simply a grid of side length $\sqrt{n}$, and therefore the density of symbols $\boxed{G}$ is less than $\frac{2d(2n+1)^{d-1}}{n^d} = O\left(\frac{1}{n}\right)$. Each copying operation contains symbols in at most two squares at any given time : one from which it copies and one to which it copies. Furthermore, because all copying operations take the same amount of time $C(n)$ to copy one square, the whole copying process of an organism in the time interval $[t_n + kC(n), t_n + (k+1)C(n)]$ is contained in the squares located at "distance" k and $k+1$ from the heart, i.e. the cells whose distance from the heart is between $k\sqrt{n}$ and $(k+2)\sqrt{n}$. Furthermore, by Lemma 6, the probability that an organism contains only one copying process tends to 1. The previous argument (used for the border-building signals) shows that copying symbols have density at most $O\left(\frac{2\sqrt{n}}{n}\right) = O\left(\frac{1}{\sqrt{n}}\right)$. □

From this lemma, we see that no limit measure can assign a nonzero probability to any pattern with a nonempty auxiliary layer.

Lemma 11.

$$d_{\mathcal{M}}(F^{t_n}\mu, \widehat{\delta_{w_n}}) \underset{n\to\infty}{\longrightarrow} 0 \qquad \textit{and} \qquad \max_{t_n \leq t \leq t_{n+1}} d_{\mathcal{M}}\left(F^t\mu, [\widehat{\delta_{w_n}}, \widehat{\delta_{w_{n+1}}}]\right) \underset{n\to\infty}{\longrightarrow} 0.$$

Proof. Take any finite square pattern $u \in \mathcal{A}^{[0,\ell]^d}$. From Lemma 10 and by σ-invariance, we can see that if c is drawn according to μ then the probability that $F^t(c)_{[0,\ell]^d}$ has any part outside the colonised space or with a nonempty auxiliary layer is $O\left(\frac{1}{\sqrt{n}}\right)$. Inside any organism at time $t = t_n$, the main layer contains concatenated copies of w_{n-1} in all directions except for those cells at distance more than $\frac{t_n - t_{n-1}}{C(n)}\sqrt{n}$ from

the heart (see the last paragraph of the previous proof), which form an asymptotically negligible set by Lemma 6. By σ-invariance, we obtain that:

$$|F^{t_n}\mu([u]) - \widehat{\delta_{w_n}}([u])| \underset{n\to\infty}{\longrightarrow} 0.$$

Since this is true for any square pattern, we get the first part of the result.

At time t_n, the copying process for w_n is triggered. As explained in the last paragraph of the previous proof, between times $t_n + kC(n)$ and $t_n + (k+1)C(n)$ the copying process is contained in cells at distance $k\sqrt{n}$ to $(k+2)\sqrt{n}$ from the nearest heart. In particular, the main layer of cells at distance less than $k\sqrt{n}$ from the nearest heart contain concatenated copies of w_n while those at distance more than $(k+2)\sqrt{n}$ still contain concatenated copies of w_{n-1}.

Therefore, denoting by $h(c)$ the minimum distance between 0 and an heart in c, we have for any $t_n \leq t \leq t_{n+1}$:

$$F^t\mu([u]) = \mu\left(h(c) \leq \frac{t-t_n}{C(n)}\sqrt{n}\right) \cdot \widehat{\delta_{w_n}}([u]) + \mu\left(h(c) > \frac{t-t_n}{C(n)}\sqrt{n}\right) \cdot \widehat{\delta_{w_{n-1}}}([u]) + \underset{n\to\infty}{o}(1).$$

The second term contains $\mu\left(h(c) > \frac{t-t_n}{C(n)}\sqrt{n}\right)$ instead of the expected $\mu\left(h(c) > \left(\frac{t-t_n}{C(n)} + 2\right)\sqrt{n}\right)$ to get an actual barycenter, the difference between asymptotically negligible in n. This equation holding for any square pattern u, we obtain:

$$d_{\mathcal{M}}\left(F^t\mu\ ,\ \mu\left(h(c) \leq \frac{t-t_n}{C(n)}\sqrt{n}\right) \cdot \widehat{\delta_{w_n}} + \mu\left(h(c) > \frac{t-t_n}{C(n)}\sqrt{n}\right) \cdot \widehat{\delta_{w_{n-1}}}\right) \underset{n\to\infty}{\longrightarrow} 0.$$

The right-hand measure belonging to the segment $[\widehat{\delta_{w_{n-1}}}, \widehat{\delta_{w_n}}]$, and this being true for any $t_n \leq t \leq t_{n+1}$, we obtain the desired result. □

Proof (of Theorem 1). The main theorem follows from Lemma 11. The right-hand part proves that $\mathcal{V}(F,\mu)$ is included in the closure of the polygonal path delineated by the sequence $(\widehat{\delta_{w_n}})_{n\in\mathbb{N}}$. To get the other inclusion, notice that $d_{\mathcal{M}}\left(F^t\mu, F^{t+1}\mu\right) \to 0$. Indeed, using the second part of the previous proof, we have $d_{\mathcal{M}}\left(F^t\mu, F^{t+1}\mu\right) \leq 2\mu\left(\frac{t-t_n}{C(n)} \leq h(c) \leq \frac{t+1-t_n}{C(n)}\right) + \underset{n\to\infty}{o}(1) = \underset{n\to\infty}{o}(1)$. Therefore any point of the segment $\left[\widehat{\delta_{w_n}}, \widehat{\delta_{w_{n+1}}}\right]$ has a measure of $(F^t\mu)_{t_n\leq t\leq t_{n+1}}$ at distance $\underset{n\to\infty}{o}(1)$, which gives the other inclusion. □

4 Statement of the results

From Theorem 1 we deduce a number of results which are our main contributions.

Corollary 1. *The measures $\nu \in \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ for which there exist:*

- *an alphabet $\mathcal{B} \supset \mathcal{A}$,*
- *a cellular automaton $F : \mathcal{B}^{\mathbb{Z}^d} \to \mathcal{B}^{\mathbb{Z}^d}$, and*
- *a nondegenerate Bernoulli measure $\mu \in \mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}^d})$*

such that $F^t\mu \to \nu$, are exactly the limit-computable measures.

Corollary 2. *The connected sets of measures $\mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ for which there exist:*

- *an alphabet $\mathcal{B} \supset \mathcal{A}$,*
- *a cellular automaton $F : \mathcal{B}^{\mathbb{Z}^d} \to \mathcal{B}^{\mathbb{Z}^d}$, and*

- *a nondegenerate Bernoulli measure* $\mu \in \mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}^d})$

such that $\mathcal{V}(F,\mu) = \mathcal{K}$*, are exactly the* Π_2*-computable, connected, compact sets of measures.*

Furthermore, both corollaries hold if one requires the convergence to hold for all nondegenerate Bernoulli measures.

Proof. Apply Theorem 1 to Proposition 4. To get Corollary 1, use the fact that ν is a limit-computable measure if and only if the singleton $\{\nu\}$ is a Π_2-computable set of measures (and of course connected). □

Following [5], we obtain a similar characterisation using convergence in Cesàro mean (Corollary 5 in op.cit) and a Rice-style theorem on μ-limit measures set (Corollary 7 in op.cit). Since the proofs of op.cit. only involve finding an appropriate uniformly computable sequence (w_n) without modifying the cellular automaton, they can be carried straightforwardly to the d-dimensional case by replacing $\mathcal{A}^{\mathbb{Z}}$ by $\mathcal{A}^{\mathbb{Z}^d}$ and we do not repeat them here.

Corollary 3. *The sets of measures* $\mathcal{K}' \subset \mathcal{K} \subset \mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$ *for which there exist:*

- *an alphabet* $\mathcal{B} \supset \mathcal{A}$*,*
- *a cellular automaton* $F : \mathcal{B}^{\mathbb{Z}^d} \to \mathcal{B}^{\mathbb{Z}^d}$*, and*
- *a nondegenerate Bernoulli measure* $\mu \in \mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}^d})$

such that $\mathcal{V}(F,\mu) = \mathcal{K}$ *and* $\mathcal{V}'(F,\mu) = \mathcal{K}'$*, are exactly the* Π_2*-computable, connected, compact sets of measures.*

In particular we characterise all sets of measures reachable at the limit in convergence in Cesàro mean from a Bernoulli measure, since those sets are necessarily connected (Section 1.2.3 in op.cit.). Here again, the result holds if one requires the convergence to hold for all nondegenerate Bernoulli measures.

Corollary 4. *Let* P *be a nontrivial property (i.e. not always or never true) on non-empty* Π_2*-computable, compact, connected sets of* $\mathcal{M}_\sigma(\mathcal{A}^{\mathbb{Z}^d})$*. There is no algorithm that can decide, given an alphabet* $\mathcal{B}$*, a cellular automaton* $F : \mathcal{B}^{\mathbb{Z}^d} \to \mathcal{B}^{\mathbb{Z}^d}$ *and a Bernoulli measure* $\mu \in \mathcal{M}_\sigma(\mathcal{B}^{\mathbb{Z}^d})$*, whether* $\mathcal{V}(F,\mu)$ *satisfies* P*.*

Here it is assumed that the Bernoulli measure is finitely described by a list of (rational) parameters. A similar statement follows on nontrivial properties of limit-computable measures. This corollary would also hold if the property was required to hold for all, or for some, nondegenerate Bernoulli measure(s).

References

[1] Alexis Ballier, Pierre Guillon, and Jarkko Kari. Limit sets of stable and unstable cellular automata. *Fundam. Inf.*, 110(1-4):45–57, January 2011.

[2] Laurent Boyer, Martin Delacourt, Victor Poupet, Mathieu Sablik, and Guillaume Theyssier. μ-limit sets of cellular automata from a computational complexity perspective. *CoRR*, abs/1309.6730, 2014.

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

[4] 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.

[5] Benjamin Hellouin de Menibus and Mathieu Sablik. Characterisation of sets of limit measures after iteration of a cellular automaton on an initial measure. *CoRR*, abs/1301.1998, 2013.

[6] Martin Delacourt. Rice's theorem for -limit sets of cellular automata. In *ICALP (2)*, pages 89–100, 2011.

[7] Martin Delacourt and Benjamin Hellouin de Menibus. Construction of mu-Limit Sets of Two-dimensional Cellular Automata. In Ernst W. Mayr and Nicolas Ollinger, editors, *32nd International Symposium on Theoretical Aspects of Computer Science (STACS 2015)*, volume 30 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 262–274, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[8] Gustav A. Hedlund. Endomorphisms and automorphisms of the shift dynamical system. *Mathematical Systems Theory*, 3(4):320–375, 1969.

[9] Benjamin Hellouin de Menibus. *Asymptotic Behaviour of Cellular Automata: Computation and Randomness*. PhD thesis, Aix-Marseille University, 2010.

[10] Benjamin Hellouin de Menibus and Mathieu Sablik. Characterisation of sets of limit measures after iteration of a cellular automaton on an initial measure. arXiv:1301.1998, accepted to *Ergodic Theory and Dynamical Systems*, 2013.

[11] J. Kari. Rice's theorem for the limit sets of cellular automata. *Theoretical Computer Science*, 127:229–254, 1994.

[12] P. Kůrka and A. Maass. Limit Sets of Cellular Automata Associated to Probability Measures. *Journal of Statistical Physics*, 100(5-6):1031–1047, 2000.

[13] Alejandro Maass. On the sofic limit sets of cellular automata. *Ergodic Theory and Dynamical Systems*, 15:663–684, 7 1995.

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

[15] John von Neumann. *Theory of Self-Reproducing Automata*. University of Illinois Press, Champaign, IL, USA, 1966.

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

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