

# Is there a physically universal cellular automaton or Hamiltonian?

Dominik Janzing

Max Planck Institute for Biological Cybernetics

Spemannstr. 38

72076 Tübingen, Germany

Email: [dominik.janzing@tuebingen.mpg.de](mailto:dominik.janzing@tuebingen.mpg.de)

September 09, 2010

## Abstract

It is known that both quantum and classical cellular automata (CA) exist that are computationally universal in the sense that they can simulate, after appropriate initialization, any quantum or classical computation, respectively. Here we introduce a different notion of universality: a CA is called physically universal if every transformation on any finite region can be (approximately) implemented by the autonomous time evolution of the system after the complement of the region has been initialized in an appropriate way. We pose the question of whether physically universal CAs exist.

Such CAs would provide a model of the world where the boundary between a physical system and its controller can be consistently shifted, in analogy to the Heisenberg cut for the quantum measurement problem. We propose to study the thermodynamic cost of computation and control within such a model because implementing a cyclic process on a microsystem may require a non-cyclic process for its controller, whereas implementing a cyclic process on system *and* controller may require the implementation of a non-cyclic process on a “meta”-controller, and so on. Physically universal CAs avoid this infinite hierarchy of controllers and the cost of implementing cycles on a subsystem can be described by mixing properties of the CA dynamics.

We define a physical prior on the CA configurations by applying the dynamics to an initial state where half of the CA is in the maximum entropy state and half of it is in the all-zero state (thus reflecting the fact that life requires non-equilibrium states like the boundary between a hot and a cold reservoir). As opposed to Solomonoff’s prior, our prior does not only account for the Kolmogorov complexity but also for the cost of isolating the system during the state preparation if the preparation process is not robust.

The main goal of this article is to formally state several open problems and sketch their relevance for the foundations of physics rather than providing results.

# 1 Towards a physical theory of control

In the abstract framework of both quantum theory and classical physics, the following concepts play a crucial role: (1) states (2) dynamical evolution (3) measurements (4) system composition and (5) restriction of the state of a composed system to one of its components. In quantum theory, states are given by density operators (e.g. positive operators with trace one) on the system Hilbert space  $\mathcal{H}$ , the dynamical evolution is described by a semi-group of completely positive trace-preserving maps, measurements are described by positive-operator-valued measures, and system composition is described by tensor products of Hilbert spaces [1, 2, 3]. Finally, partial traces define system restriction.

In classical physics, the states are probability distributions on a phase space, the dynamics is given by a semi-group of stochastic maps, system composition is given by the cartesian product of the phase spaces, and state restriction is given by marginalization of probability measures.

Having such a framework for the physical world raises the question to what extent the formalism also contains states, dynamical evolutions, and measurements that do not correspond to any physically possible situation or process. Restricting the attention to quantum theory, these questions thus read: (1) Is every density operator on  $\mathcal{H}$  a *physically* possible state, (2) is every completely positive trace-preserving operation a process that can be implemented in nature, (3) is there a measurement procedure for every POVM?

First we describe in what sense modern quantum computing (QC) research [3] has given an affirmative answer to all these questions and in what sense it has not. To this end, we first rephrase some terminology of QC. A quantum-bit (qubit) is a quantum system with Hilbert space  $\mathbb{C}^2$ , a quantum register is a collection of  $n$  qubits<sup>1</sup>. Researchers have described various physical systems having a quantum degree of freedom for which two states are universally controllable in the following sense: Any unitary operation on  $\mathbb{C}^2$  (“single qubit gate”) can be performed by appropriate operations on the system. Moreover, they have described how to implement controlled interactions between pairs of qubits, thus implementing a unitary on the Hilbert space  $\mathcal{H} := \mathbb{C}^2 \otimes \mathbb{C}^2$  that is not a product of single qubit operations (hence a proper “two-qubit gate”). It was then shown that sequences of one- and two-qubit gates are sufficient for implementing arbitrary unitary operations up to any desired precision [4]. Being able to prepare *one* pure state of the quantum register thus enables the preparation of *any* pure state. Moreover, measurements with respect to any measurement basis can be reduced to measurements with respect to a single reference basis by first transforming the state to the latter basis via a unitary transformation. Preparations of mixed states, implementation of general completely positive trace-preserving maps and measurements for general POVMs can be obtained by restriction of states to a subsystem. In this sense, questions (1)-(3) seem to be answered with “yes”. Then, the operational meaning of some multi-qubit states, dynamical evolutions, observables are only limited by the fact that the implementation time could even exceed the life-time of the universe<sup>2</sup>. There

---

<sup>1</sup>It should be noted that the restriction to two-dimensional systems is only a matter of convention.

<sup>2</sup>For a complexity theory of states and observables see e.g. [5, 6, 7, 8]

are, however, two other reasons why QC did not answer our questions in the sense intended here.

First, the quantum mechanical degrees of freedom defining the qubits in existing proposals for QC [3] are only a small part of the entire degrees of freedom of physical particles (e.g. the nuclear spin of a particle or it may be two levels in an internal degree of freedom of a trapped ion). So far, it has not been claimed that *all* the degrees of freedom of such a particle would be controllable simultaneously.

The second reason why we are not satisfied with the answer given by QC is that we would like to see a theoretical model of quantum control that treats the controller as the same type of physical system as the system to be controlled. Within such a unifying model – as proposed by the present article – we are able to explore the conditions under which one system acts as controller of the other, even though a physical interaction can send information in *both* directions.<sup>3</sup> Moreover, the question of how to control the controller then shifts the problem of how to control the system to the problem of how to control the controller by a “meta-controller”, leading to an infinite hierarchy of controllers.

Remarkably, the same shift between system and its interface is generally accepted for the quantum measurement problem: Once the quantum measurement process is described by an interaction between system and the measurement apparatus, the question occurs “who measures the measurement apparatus?”, which leads to the same chain of measurement instruments as we have stated for the controller problem above. For the measurement apparatus, it has been argued that the cut between system and measurement instrument is arbitrary, the description must remain consistent if the boundary is shifted. Likewise, we argue that quantum control has a consistent description if one can show that the cut between system and controller can be shifted. In [9], we have already described a toy model of quantum control with a fixed interaction between controller and system, where operations on the system are implemented by implementing transformations on the controller. In the present paper, we assume that we are only able to implement *state preparations* on the controller. We first state on an abstract level what we would consider a consistent model of physical control, before it will be made precise within the setting of cellular automata (CA):

**Definition 1 (model of physical control, abstract version )**

*Let  $(\alpha_t)$  with  $t \in \mathbb{R}$  or  $t \in \mathbb{Z}$  be a group describing the dynamical evolution on state space of the world  $W$ . Then every mathematically possible operation on the physical state space of some region  $R$  can be implemented by initializing the complement  $W \setminus R$  of the region to an appropriate state and waiting until  $\alpha_t$  implements the desired operation.*

To motivate Definition 1, we first consider an arbitrary experimental setup that is able to implement one particular control operation. The control operation may, for instance, be to change the quantum state of a few ions in an ion trap in some desired way. To this end, some sophisticated sequence of Laser pulses is applied to the

---

<sup>3</sup>Note that unidirectionality of causal influence not only occurs if the controller is significantly larger than the system to be controlled. Instead, it is also a matter of the *state* of the controller. For such toy models of quantum control see e.g., [9, 10]; Refs. [11, 12] discuss thermodynamic aspects of unidirectionality.

system. Assume that the pulses are controlled by a computer program so that there is no need for the experimentalist to intervene once the program runs. We can then consider computer, Laser and the ions in the trap as a big physical system on which the global dynamics of the world acts. Obviously, the computer software controlling this process is just a physical state of the computer. However, here we want to go further and also consider the presence or absence of the hardware of the experimental setup merely as different states of a larger system (which implicitly refers to a field-theoretic point of view). From such a perspective, there is no distinction between hardware and software in the experimental setup and the whole control operation on the system to be controlled (the ions) is implemented by changing the physical state of the system’s environment.

Following, for instance, [13, 14, 15] we will consider cellular automata (CA) as interesting models of the world and therefore study our problem in the context of CAs. Our main focus (Section 2) will be on classical CAs since the problem seems to be non-trivial even in the classical regime. Apart from describing possible definitions of physical universality (Subsection 2.1), we discuss some relations between physical universality to ergodic properties of CAs in Subsection 2.2. Subsections 2.3 and 2.4 argues why physically universal CAs are helpful for studying limits of control and thermodynamic laws from a new perspective. Subsection 2.5 proposes a prior distribution for physical states based on physically universal CAs. To this end, we consider an initial state of the CA where half of the cells are set to zero and the other half are in the maximum entropy state (thus modelling a hot and a cold part of the universe). Section 3 briefly discusses physical universality for *quantum* CAs (Subsection 3) and physically universal Hamiltonians as their continuous analog (Subsection 3.3), where controllability also implies the ability to control the preparation of quantum superpositions by a *classical* program. In the context of physically universal Hamiltonians, the terms “hot” and “cold” part of the universe can be taken more literally because they really refer to Gibbs states. This makes the physical interpretation of the prior more obvious.

The main contribution of this article is to raise the question of how to define the right framework for a physical control theory that also treats the controller as an object internal to the theory. In posing this question, the paper sketches the possible impact of such a framework, but it will not present any deep results on cellular automata.

## 2 Physically universal classical cellular automata

### 2.1 Possible options for defining physical universality

We first introduce some terminology and notation for classical cellular automata (CAs). Let  $L := \mathbb{Z}^d$  for some  $d \in \mathbb{N}$  be a  $d$ -dimensional lattice and  $A$  be an alphabet of states of a single cell (without loss of generality, let one of the symbols be “0”). The space of pure states of the CA is given by

$$S := A^L.$$

The space of mixed states is given by probability distributions on  $S$ . The maximally mixed state (maximum entropy state) is given by the uniform distribution over  $S$ , i.e.,

the infinite product of uniform distributions over  $A$ . For every state  $s \in S$  and every subset  $R \subset L$  the restriction of  $s$  to  $R$ , denoted by  $s|_R$  is defined by the substring  $s' \in A^R$  corresponding to  $R$ . A *region* will be a subset  $R \in L$ . Usually, our regions will be finite subsets unless we state the opposite.

By slightly abusing notation, we set

$$s|_x := s|_{\{x\}},$$

for any  $x \in L$ . A configuration of a region  $R$  is a string  $c \in A^R$ . It defines, in a straightforward way, the cylinder set

$$\{s \in A^L \mid s|_R = c\},$$

which will also be denoted by  $c$  whenever this causes no confusion. The entropy of  $R$  in the mixed state  $\nu$  is given by the Shannon entropy of the restriction of  $\nu$  to  $R$ , i.e.,

$$S(\nu|_R) := - \sum_{c \in A^R} \nu(c) \log \nu(c).$$

The time evolution  $(\alpha_t)_{t \in \mathbb{Z}}$  of a CA is a group (by assuming the group property we implicitly restrict the attention to reversible CAs) of translation covariant maps

$$\alpha_t : S \rightarrow S,$$

that is local in the sense that  $\alpha_{\pm 1}(s)|_x$  only depend on the state of the cells lying in some neighborhood of  $x$ . Here we consider the Moore neighborhood of radius one, i.e., all cells  $y$  with  $\|y - x\|_\infty \leq 1$  [16].

By slightly overloading notation, we also write  $\alpha_t(c)|_R$  if  $c \in A^{R'}$  is the configuration of any region  $R'$  that contains all cells relevant for determining the state of  $R$  at time  $t$  (which is, for instance, the case if  $R'$  contains the Moore neighborhood of  $R$  with radius  $t$ ). If a configuration  $c \in A^R$  is defined via  $c := (c_1, c_2)$  with  $c_1 \in A^{R_1}$  and  $c_2 \in A^{R_2}$  for  $R = R_1 \cup R_2$ , we write  $\alpha_t(c_1, c_2)|_R$  instead of  $\alpha_t((c_1, c_2))|_R$ .

The following definition formalizes the weakest form among all notions of physical universality that we define. It is the ability to change the state of a region  $R$  by initializing the complement of  $R$  in an appropriate way:

**Definition 2 (conditional state preparation)**

*A CA is said to allow for conditional state preparation if for every region  $R \subset L$  and every pair  $(c_i, c_f)$  of initial and final configurations of  $R$  there exists a configuration  $e \in A^{L \setminus R}$  and a time  $t \in \mathbb{N}_0$  such that*

$$\alpha_t(e, c_i)|_R = c_f.$$

*Less formally speaking, the dynamics prepares the final state  $c_f \in A^R$  after the time  $t$ , given that the environment started in the state  $e$  and the region in the state  $c_i$ .*

Note that the state  $e$  can be chosen differently for every initial state  $c_i$ . The following notion of state preparation is stronger since it demands the existence of a state  $e$  that works for *every* initial configuration  $c_i$ :

**Definition 3 (unconditional state preparation)**

A CA is said to allow for unconditional state preparation if for every finite subset  $R \subset L$  of cells and configurations  $c_f \in A^R$ , there exists a configuration in  $e \in A^{L \setminus R}$  and a time  $t \in \mathbb{N}$  such that

$$\alpha_t(e, c_i)|_R = c_f,$$

for every configuration  $c_i \in A^R$ .

Less formally speaking, the dynamics prepares the state  $c_f$  in the region  $R$  by initializing the complement of  $R$  to  $e$ , regardless of the initial state  $c_i$  of  $R$ .

It seems that Definition 2 already formalizes a sufficiently strong property because one could prepare the environment after having read out the initial state  $c_i$  of the region  $R$ . However, the entire process of readout and conditioning the initialization of the complement of  $R$  on the state  $c_i$  should also be implemented by the physical laws that govern the dynamics of the world. Therefore, we consider the latter definition as the better notion of universal state preparation. Nevertheless, the following example shows that Definition 3 is a rather weak notion of universality since it is already satisfied by a simple shift:

**Example 1 (shift)**

For  $x \in L$  let  $\alpha_1$  be given by shifting the state by the vector  $x$ , i.e.,

$$\alpha_1(s)|_i := s|_{i-x} \quad \forall s \in A^L.$$

For some finite region  $R \subset \mathbb{Z}$ , let  $c_f \in A^R$  be an arbitrary configuration. Then  $c_f$  can be prepared as follows. Choose some  $t_0$  such that

$$(R + xt_0) \cap R = \emptyset.$$

Initialize the region  $R' := R - xt_0$  to the translated copy of  $c_f$ . Then the region  $R$  is obviously in the configuration  $c_f$  at time  $t_0$ .

If  $d = 1$  and  $x = 1$ , the dynamics shifts the state of each site by one. Then the corresponding MDS is known as Bernoulli shift.

However, such a trivial model of dynamical evolution is unacceptable as a model for universal control. One reason is that it lacks computation power. We could ask for models that are computationally universal and allow for universal state preparation in the sense of Definitions 2 or 3. Rather than postulating computational power a priori, we prefer demanding that the model allows for non-trivial operations other than state preparation. The following condition includes conditional state preparation and is obviously not satisfied for the shift dynamics:

**Definition 4 (universal implementation of bijections)**

A CA is said to allow for universal implementation of bijections if for every finite region  $R \subset L$  and every bijective map

$$\pi : A^R \rightarrow A^R$$

there is a configuration  $e \in A^{L \setminus R}$  of the complement of  $R$  and a time  $t$  such that

$$\alpha_t(e, c)|_R = \pi(c) \quad \forall c \in A^R.$$

Note that the ability of implementing bijections implies the ability of implementing measurements in the following sense: apart from a region  $R$  whose state should be measured, define a region  $R_M$  which serves as a measurement apparatus. One can then implement a bijection  $\pi$  on  $R \cup R_M$  that changes the state of  $R_M$  depending on the state of  $R$ .

One of the main goal of this paper is to formulate the following open problem:

**Question 1 (existence of physically universal CA)**

*Is there a classical CA that is physically universal in the sense of Definition 4?*

It is easy to see that non-bijective maps  $\pi$  can be implemented by restricting bijections to smaller regions. For this reason, the bijectivity assumption in Definition 4 is irrelevant and it is a matter of taste whether one wants to keep it in the definition.

In case the answer to this question is negative, one should try to find a weaker sense of universal controllability. An affirmative answer, on the other hand, raises further questions since physically universal CAs are good candidates for studying thermodynamic cost of computation and (quantum) control from a new perspective. Some ideas on that will be presented in Subsection 2.4.

We will not formulate any conjecture regarding the solution of Question 1, but Subsection 2.3 will show that the controllability of the controller of a system imposes limitations on the controllability of the system itself.

## 2.2 Some relations between physical universality and ergodic properties

We want to discuss relations between physical universality and ergodicity of dynamical systems. To this end, we introduce the following terminology [17]:

**Definition 5 (measure-preserving dynamical systems (MDS))**

*Let  $(\Omega, \Sigma, \mu)$  be a measure space where  $\Omega$  is a set,  $\Sigma$  the  $\sigma$ -algebra of measurable subsets of  $\Omega$  and  $\mu$  a measure with  $\mu(\Omega) < \infty$ . Let  $\phi : \Omega \rightarrow \Omega$  be a measurable map with  $\mu(\phi^{-1}(B)) = \mu(B)$  for every measurable set  $B$ . Then  $(\Omega, \Sigma, \mu, \phi)$  is called a measure-preserving dynamical system (MDS).*

Then we have:

**Lemma 1 (CA is an MDS)**

*Every reversible CA as defined above is a measure-preserving dynamical system where  $\Omega := S$ ,  $\Sigma$  is generated by the set of cylinder sets,  $\mu$  is the product of uniform probability distributions on  $A$  and  $\phi := \alpha_1$ .*

Proof:  $\mu(\alpha_1^{-1}(B)) = \mu(B)$  can easily be checked for every cylinder set  $B$ . Since the latter ones generate the entire sigma algebra of measurable sets, conservation of measure follows.  $\square$

The following terminology will be useful [18, 17]:

**Definition 6 (ergodicity)**

An MDS is called ergodic if  $\phi^{-1}(B) \cong B$  implies  $B \cong \Omega$  or  $B \cong \emptyset$  for all  $B \in \Sigma$ , where  $\cong$  denotes equality up to sets of measure zero. Equivalently,  $\phi^{-1}(B) \cong B$  can also be replaced with  $\phi^{-1}(B) \subseteq B$  or  $\phi^{-1}(B) \supseteq B$  (up to sets of measure zero).

We will also need another equivalent formulations of ergodicity [18]:

**Lemma 2 (different characterization of ergodicity)**

An MDS is ergodic if and only if for every  $B, D \in \Sigma$  there is an  $t \in \mathbb{N}$  such that

$$\phi^{-t}(B) \cap D \neq \emptyset.$$

Then we have:

**Theorem 1 (state preparation in ergodic CAs)**

If a CA is an ergodic MDS, it allows for conditional state preparation in the sense of Definition 2.

Proof: Let  $B_i \subset S$  and  $B_f \subset S$  be the cylinder sets corresponding to the initial and the final configuration  $c_i$  and  $c_f$  of  $R$ , respectively. Then there is a  $t$  such that

$$\alpha_t^{-1}(B_f) \cap B_i \neq \emptyset.$$

Choose  $c \in \alpha_t^{-1}(B_f) \cap B_i$ . Since  $c$  is an element of  $B_i$ , it is of the form  $c = (e, c_i)$ . On the other hand,  $\alpha_t(e, c_i)|_R = c_f$  because  $c \in \alpha_t^{-1}(B_f)$ .  $\square$

Ergodicity of CAs has already been studied in the literature<sup>4</sup> [20, 21], but the fact that the Bernoulli shift (Example 1) is ergodic [17] shows that even ergodicity does not imply physical universality in the sense of Definition 4.

## 2.3 Limits of controllability

Being able to prepare a certain state, one may also wish to keep it at least for some time. In the context of quantum information processing, for instance, it is considered as an important problem to prevent a quantum state from decaying too quickly (where decay can be understood in the sense of both decoherence or relaxation). To ensure this, one tries to isolate the system as much as possible from influences of the environment. On the other hand, implementing control operations requires interactions with the environment. We expect that this conflict between protecting the state by isolating the system and nevertheless still being able to access it, can be nicely explored in the setting of physically universal CAs. Then, isolating the system only means to prepare the environment into a state that effectively turns off the interaction. The question of whether this conflict implies serious restrictions to physical universality will mainly be unanswered, but we mention some small observations that may suggest a future direction for research. The following statement, for instance, is almost obvious, but we phrase it as a theorem because it shows that too strong controllability assumptions are self-contradictory:

---

<sup>4</sup>Note that [19] studies ergodic quantum CAs, but not in the sense of MDS. Instead, ergodicity is meant in the sense of a topological dynamics having a unique invariant state.

**Theorem 2 (some configurations are unstable)**

Let  $R$  be a region that includes at least the Moore neighborhood of one cell  $x$ . Let  $\alpha_t$  be physically universal in the sense of Definition 2, then there is a configuration  $c \in A^R$  such that

$$\alpha_1(e, c)|_R \neq c \quad \forall e \in A^{L \setminus R}.$$

Proof: If the dynamics of the CA is non-trivial (which is certainly the case for physically universal CAs) there must be a configuration  $c \in A^R$  such that

$$\alpha_1(c)|_x \neq c|_x.$$

Hence,

$$\alpha_1(e, c)|_x \neq c_x$$

for all  $e \in A^{L \setminus R}$ .  $\square$

The following result is only slightly less straightforward, but it already illustrates how controllability of the controller of a region  $R$  restricts the controllability of  $R$ :

**Theorem 3 (no configuration lasts forever)**

Given a CA that is physically universal in the sense of Definition 4, then it is impossible that there exists initial and final configurations  $c_i, c_f \in A^R$ , a finite “program” region  $R_p$  with initialization  $c_p$ , and a time  $t_0 \in \mathbb{N}$  such that

$$\alpha_t(c_i, c_p)|_R = c_f \quad \forall t \geq t_0. \tag{1}$$

Proof: Assume that (1) is satisfied. Set  $R' := R \cup R_p$  and choose a vector  $x \in L$  such that  $(R' + x) \cap R' = \emptyset$  and that  $\|x\|_\infty \geq t_0$ . Let  $\beta$  be the transformation on  $R' \cup (R' + x)$  that swaps the state between  $R'$  and  $R' + x$ . By physical universality in the sense of Definition 4, there is a configuration of the complement of  $R' \cup (R' + x)$  such that  $\alpha_{t_1}$  implements  $\beta$  for some  $t_1$ . After the implementation of  $\beta$ , the region  $R$  is only in the state  $c_f$  if the initial state of the region  $R + x$  has been the shifted copy of  $c_f$ . Hence,  $t_1$  must be smaller than  $t_0$  since (1) states that the state of  $R$  is  $c_f$  *regardless* of the state of  $R' + x$  (note that  $R' + x$  is part of the complement of  $R_p$  by assumption and its state is thus irrelevant for (1)). On the other hand, the implementation of the swap  $\beta$  requires at least the time  $t_0$  since the information can propagate one cell per time step only, which leads to a contradiction.  $\square$

Theorem 3 shows that initializing a finite region can never prepare a state that lasts forever. If possible at all, it requires an *infinite* region. To show more powerful results about control tasks that are self-contradictory has to be left to the future (in this context it may also be worth mentioning Ref. [22] which describes some impossibility results for *inference* tasks instead of control tasks within a computation model of the world and relate them to the Halting problem).

## 2.4 Space and energy requirements of computations and control operations

In this section we want to mention some potential implications for the resource requirements of computation processes, given that physically universal CAs define a reasonable model of the world. Even though we have proved only a few results on this, the following high-level arguments motivate why physically universal CAs shed a different light on thermodynamics.

1. *The thermodynamic cost of isolating systems:* the difficulty of isolating physical objects from its environment is one of the main obstacles in controlling microphysics. In usual quantum control, this appears more or less as a practical problem and the question is how to turn off the disturbing interactions. Physical universality, however, implies that the system is never isolated and that only appropriate states of the environment ensure that the system behaves for some time period as if it would be isolated. The fact that, in turn, also the environment of the system is permanently coupled to its environment (by physical universality) implies that this “isolating state” is perturbed after a while. Preparing the environment into a state that effectively isolates the system for a long time, probably requires a lot of thermodynamic resources. To discuss these costs, one probably needs a model where all interactions are permanently present and cannot be turned on and off by the experimentalist. Within the framework of physically universal CAs it is not only possible to address the requirements of extracting heat from a system [23] but also of preventing the heat from reentering the system.
2. *Thermodynamic reversibility:* It is commonly assumed that the implementation of a bijective transformation of the states of a microscopic system is thermodynamically reversible. The fact that the experimental setup controlling the implementation generates a lot of heat is usually considered as a problem of current technology rather than being a fundamental law of physics. Physically universal CAs provide a model that makes it possible to explore how the controller (i.e., the region  $R_p$  around the region  $R$  to be controlled) changes its state during this implementation. From the point of view of traditional thermodynamics, this state transition is again reversible if it is a bijection of the state space of a microsystem. However, inverting this bijection will then change the state of the environment around  $R_p$ . Then, the question of thermodynamic reversibility leads, again, to our infinite sequence of meta-controllers. We will not present any solution to this deep problem. We only emphasize that the existence of thermodynamic reversible processes is challenged by the ideas above.
3. *Space and energy requirements of computation:* In complexity theory, the space requirements of a computation is defined as the size of the memory band of a Turing machine that is written on during the computation process. The complexity class PSPACE, for instance, is defined as the class of problems whose space requirements increase only polynomial in the size of the input string [24]. It is known [25] that appropriate CAs can simulate a universal Turing machine efficiently with respect to both space and time resources.

In our context, we want to redefine the space requirements of a computation in a way that is motivated by ideas from thermodynamics: we do not only count those cells of the CA that are *actively* involved in the computation in the sense that their state changes during the process. Instead, we count all cells whose state matters. In the simplest case, it may be necessary to set a large set of cells to some fixed symbol (e.g. to zero) to avoid that these cells disturb the computation by influencing the cells involved in the computation. From a purely computer scientific point of view, it is natural to study the resources of computation within a setting where all the sites are set to zero except for those involved in the computation. In our physical model, however, this would correspond to cooling all cells down to zero temperature, which requires infinite thermodynamic resources. We assume that we can only extract the entropy of a *finite* region and use this free memory space for the computation. In a physically universal CA, we then get the problem that this region can never *remain* free of entropy because the interaction that guarantees universality necessarily transfers entropy into the free memory space.

The discussion below tries to support the vague statements above by formal arguments. We will not always distinguish between computation processes and other control processes.<sup>5</sup> The following theorem is actually a simple observation, but we phrase it as a theorem because it confirms the last sentence of item 3 above:

**Theorem 4 (lower bound on entropy influx)**

Let  $R$  be an arbitrary region and  $\nu$  be a probability distribution on  $S$  whose restriction to  $L \setminus R$  is the uniform distribution. Let the CA be universal in the sense of Definition 4 and  $x$  be some vector such that  $R \cap (R + x) = \emptyset$ . If  $R_p$  denotes a region such that for some  $c_p \in A^{R_p}$  the state  $c$  of  $R$  is transferred to  $R + x$ , i.e.,

$$\alpha_t(c_p, c)|_{R+x} = c \quad \forall c \in A^R,$$

for some appropriate  $t$ , then the entropy of  $R$  after the time  $t$  is at least

$$S((\nu \circ \alpha_t)|_R) \geq \frac{|R|}{|A|^{|R_p|}} \log |A|.$$

Proof: For  $\nu_t := \nu \circ \alpha_t$  we consider the conditional distribution given  $\alpha_t^{-1}(c_p)$ . Its restriction to  $R$  is the uniform distribution because the initial state  $c_p$  triggers the implementation of the swap between  $R + x$  and  $R$ . The entropy of the uniform distribution on  $R$  reads  $|R| \log |A|$ . Since  $R_p$  is initially also in the maximum entropy mixture, the probability for being in the state  $c_p$  is  $|A|^{-|R_p|}$ . Weighting the entropy  $|R| \log |A|$  with this factor yields the desired bound.  $\square$

The theorem shows a trade-off between being able to implement bijections and being able to isolate a region: if  $\beta$  can be *easily* implemented on  $R \cup (R+x)$  (i.e., by initializing a *small* region  $R_p$ ) then  $R \cup (R + x)$  is badly isolated because we get large entropy

---

<sup>5</sup>On the elementary level of nature, thermodynamic and computation processes are closely related, anyway [26]

influx. Note that no such statement holds for *computationally* universal CAs since they could have a “death state” that remains forever and turns off all interactions with the surrounding cells. A boundary with dead cells could then prevent the memory space from getting entropy from its environment. In a physically universal CA, the environment is always able to “revitalize” the “dead cells”. It is possible that in physically universal CAs, the region that needs to be initialized to enable a computation process grows proportionally with the computation time. Loosely speaking, the size of the region that needs to be initialized is related to the amount of free energy that must be available in order to run the computation properly. This is because Landauer’s principle [27, 28, 23] states that it requires the energy  $E = kT \ln 2$  to initialize one bit. From a more accurate point of view, however, we have to account for the fact that the region that we must initialize not necessarily needs to be prepared to *one* specific configuration. Instead, it could be that there is a whole set of configurations that ensure that the desired computation process works properly. This corresponds to a smaller amount of free energy. The following definition formalizes the free energy content of configurations:

**Definition 7 (free energy of a set of configurations)**

Let  $B \subset A^L$  be a set of configurations and  $\mu$  be the uniform distribution on  $S$  (which is defined via the product of uniform distributions on each  $A$ ). Then

$$F(B) = -\log_2 \mu(B)$$

is the free energy required to ensure that the world is in a state  $s \in B$ .

The definition is motivated by the following interpretation of probability distributions. The mixed state  $\mu$ , which is the uniform distribution over all configuration, is thought to be the thermodynamic equilibrium of the world, i.e., the analog of the Gibbs state. We define its free energy to be zero. In physics, the free energy of a mixed state is, up to the factor  $kT$ , given by its relative entropy distance from thermal equilibrium [29]. Here, mixed states are probability distributions on  $A^L$  and the free energy is thus (up to constants that we ignore for sake of convenience) given by the relative entropy distance from  $\mu$ , i.e.,

$$F(\tilde{\mu}) := D(\tilde{\mu}||\mu).$$

If  $\tilde{\mu}$  is any distribution with support  $B$ , the relative entropy distance to  $\mu$  is minimal if  $\tilde{\mu}$  is the uniform distribution on  $B$ . One checks easily that

$$D(\tilde{\mu}||\mu) = -\log_2 \mu(B).$$

Within this setting, we can easily define the free energy needed for a preparation process:

**Definition 8 (free energy required for a preparation process)**

Assume a region is in the state  $c_i \in A^R$  and we want it to be in the state  $c_f$  at time  $t$ . Interpreting  $c_i$  and  $c_f$  as cylinder sets, the state of the lattice  $s \in A^{L \setminus R}$  must be chosen such that

$$s \in c_i \cap \alpha_t^{-1}(c_f),$$

where the right hand side interprets  $c_i$  and  $c_f$  as sets (as defined previously).

Then,

$$F(c_i \mapsto c_f) := -\log \mu(c_i \cap \alpha_t^{-1}(c_f))$$

is the free energy needed to implement the preparation process  $c_i \mapsto c_f$  after the time  $t$ .

Note that this definition includes the free energy content of  $c_i$  which is given by  $|R| \log_2 |A|$ , since  $c_i$  is one configuration in a set of  $|A|^{|R|}$  possible ones.

We also define the free energy required for a computation process:

**Definition 9 (energy requirements for computation)**

Assume that the physical universal CA is only able to perform a desired computation process  $C$  if the state  $s$  of the world lies in the set  $B \subset A^L$ . Then

$$F(B) := -\log \mu(B)$$

is the free energy required for  $C$ .

We will not elaborate on this any further, but consider the thermodynamic costs of implementing *sequences of state transitions* on some region  $R$  since this task is easier to address than computation tasks. Consider the following sequence of state transitions

$$c_0 \xrightarrow{t_1} c_1 \xrightarrow{t_2} c_1 \cdots \xrightarrow{t_n} c_n,$$

and define the corresponding free energy resource requirements by

$$-\log \mu(c_0 \cap \alpha_{t_1}^{-1}(c_1) \cap \alpha_{t_1+t_2}^{-1}(c_2) \cap \cdots \cap \alpha_{t_1+\dots+t_n}^{-1}(c_n)).$$

An interesting special instance is to implement  $k$  cycles

$$\underbrace{c_1 \mapsto c_2 \mapsto \cdots \mapsto c_n}_{\text{1th cycle}} \mapsto \underbrace{c_1 \mapsto c_2 \cdots c_n}_{\text{2nd cycle}} \cdots, \quad (2)$$

where the transition from  $c_i$  and  $c_{i+1}$  and from  $c_n$  to  $c_1$  is implemented by one time step of the CA. We do not know whether physically universal CAs also allow for the implementation of arbitrarily many cycles of this form, but given that they do, we have the following statement for ergodic CAs:

**Theorem 5 (cost of implementing repeated cycle processes)**

Let  $c_1, \dots, c_n$  be configurations of a region  $R$  such that

$$\bigcup_{j=1}^n c_j \neq A^R. \quad (3)$$

Then, in an ergodic CA, the cost of implementing  $k$  cycles of the form (2) converges to infinity for  $k \rightarrow \infty$ .

Proof: define  $B := \bigcup_{j=1}^n c_j$  and

$$D := \bigcap_{j=1}^k \alpha_k^{-1}(B).$$

Clearly,

$$\alpha_1(D) \subset D. \quad (4)$$

Due to eq. (3), we have  $\mu(D) \neq 1$ . Since  $\alpha_1$  is ergodic, all sets satisfying the invariance condition (4) have measure zero or one, hence  $\mu(D) = 0$ . Due to

$$\lim_{k \rightarrow \infty} \mu \left( \bigcap_{j=1}^k \alpha_j(B) \right) = \mu(D) = 0,$$

the statement follows.  $\square$ .

A weaker task than implementing a cycle is to periodically restore the same configuration  $c$  again and again after  $\tau$  time steps, without specifying what happens between the  $\tau$  steps:

$$c \xrightarrow{\tau} c \xrightarrow{\tau} c \xrightarrow{\tau} \dots$$

According to Definition 7, the free energy requirements are given by

$$-\log \mu \left( \bigcap_{j=0}^n \alpha_{j\tau}^{-1}(c) \right). \quad (5)$$

To derive statements on the resources needed, we first recall the following mixing property (see [18], page 38), which is known to imply ergodicity [18, 17]:

**Definition 10 (weakly mixing MDS)**

*An MDS is called weakly mixing if*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(\phi^{-j}(B) \cap D) = \mu(B)\mu(D),$$

for all measurable sets  $B, D$ .

The following result of ergodic theory (Corollary 14.15 in [17]) will be helpful:

**Lemma 3 (mixing of all orders)**

*Every weakly mixing MDS is weakly mixing of all orders in the sense that*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{n=0}^{n-1} \mu \left( B_0 \cap \phi^{-n}(B_1) \cap \phi^{-2n}(B_2) \cap \dots \cap \phi^{-(k-1)n}(B_{k-1}) \right) = \mu(B_0) \cdots \mu(B_{k-1}), \quad (6)$$

for all  $k \in \mathbb{N}$  and every  $B_0, \dots, B_{k-1} \in \Sigma$ .

We apply this result to our setting and obtain:

**Theorem 6 (cost of restoring states in weakly mixing CAs)**

Let the CA be weakly mixing and assume that there is a configuration  $c \in A^{\mathbb{R}}$  for which it is possible to implement the following  $k$ -fold recurrence

$$\underbrace{c \xrightarrow{\tau} c \xrightarrow{\tau} c \xrightarrow{\tau} \cdots \xrightarrow{\tau} c}_k,$$

for all  $\tau \geq \tau_0$  for some  $\tau_0 \in \mathbb{N}$ . Let  $F_k(\tau)$  be the free energy required for implementing this process. Define the average free energy requirements over all  $\tau \geq \tau_0$  by

$$\bar{F}_k := \liminf_{\tau_1 \rightarrow \infty} \frac{1}{\tau_1 - \tau_0 + 1} \sum_{\tau=\tau_0}^{\tau_1} F_k(\tau). \quad (7)$$

Then it satisfies the lower bound

$$\bar{F}_k \geq -k \log_2 \mu(c).$$

Proof: According to Definition 7,  $F_k(\tau)$  reads

$$F_k(\tau) := -\log_2 \mu \left( \bigcap_{j=0}^{k-1} \alpha_{j\tau}^{-1}(c) \right).$$

Hence,

$$\bar{F}_k := -\liminf_{\tau_1 \rightarrow \infty} \frac{1}{\tau_1 - \tau_0 + 1} \sum_{\tau=\tau_0}^{\tau_1} \log_2 \mu \left( \bigcap_{j=0}^{k-1} \alpha_{j\tau}^{-1}(c) \right).$$

The convexity of the logarithm implies

$$\bar{F}_k \geq \log_2 \lim_{\tau_1 \rightarrow \infty} \frac{1}{\tau_1 - \tau_0 + 1} \sum_{\tau=\tau_0}^{\tau_1} \mu \left( \bigcap_{j=0}^{k-1} \alpha_{j\tau}^{-1}(c) \right) = -\log_2 \mu(c)^k = -k \log_2 \mu(c),$$

where the second last equality uses eq. (6).  $\square$

Theorem 6 states that the cost of repeatedly restoring the same state  $k$  times (after  $\tau$  time steps) grows linearly in  $k$  when averaged over all  $\tau$ . The physical relevance of this statement is speculative for two reasons. First, we do not know whether the appropriate mixing properties follow from physical universality. Second, it is unclear whether the assumption that the sequence of state transitions can be implemented for all  $\tau \geq \tau_0$  is reasonable. We will therefore formulate another open problem:

**Question 2 (thermodynamic cost of cycles)**

Given any desired configuration  $c$ , how does the free energy (5) of restoring it again and again grow with the number  $k$  of cycles?

In case the energy grows at least linearly in  $k$  for physically realistic models, this would suggest that implementing cycles on microscopic systems involves an experimental setup whose energy content grows linearly in the number of cycles. On the

one hand, the energy content does not seem to be *used up*, since it just needs to be *available*. On the other hand, this amount of energy cannot be used to implement the next cycles because, if reusing the energy was possible, the amount of energy that needs to be present would not grow linearly in  $k$ . Note that the above ergodic theory based framework avoids exploring the thermodynamic cost of an infinite sequence of controllers and meta-controllers as sketched in item 2 at the beginning of this subsection because the universal CA describes the whole hierarchy of controllers simultaneously.

## 2.5 Towards a physical analog of Kolmogorov complexity and Solomonoff’s prior

Several authors have already pointed out the physical relevance of algorithmic information (“Kolmogorov complexity”), e.g., [30, 31]. For any binary string  $\{0, 1\}^*$ , the algorithmic information  $K(s)$  is defined by the length of the shortest program on a universal prefix Turing machine that outputs  $s$  and halts then [32, 33, 34].

The thermodynamic relevance of Kolmogorov complexity has, for instance, been emphasized in [35, 30, 36], its importance for statistical inference has already been described by Solomonoff [33], and also the foundation of modern machine learning methodology often refer to Kolmogorov complexity, e.g. [37, 38]. Recently, [39, 40, 41] postulated *causal* inference rules that also use algorithmic information. A crucial concept for algorithmic information based inference is Solomonoff’s prior:

### Definition 11 (Solomonoff’s prior)

*Given a universal Turing machine  $T$  with prefix coding. Then, for any binary string  $s \in \{0, 1\}^*$ , one defines  $m(s)$  by the probability that  $T$  produces the output  $s$  and stops after every bit of the infinite input tape has been randomly set to 0 or 1 with probability  $1/2$  each.*

Note that these random programs do not contain any *additional* symbol that indicates the end of the program code. Since the Turing machine uses prefix coding, no valid program is the prefix of another one. For this reason, the uniform distribution over *all* binary words (defined by the infinite product of uniform distributions on  $\{0, 1\}$ ) automatically defines a distribution on the set of valid programs.

Even though Solomonoff’s prior has shown to be a powerful concept for the foundation of inference, the following modifications may be appropriate for a prior on the states of the physical world:

1. *Symmetries*: What prior probability should, for instance, be assigned to the event that a next lightning hits the earth at a longitude of  $0^\circ$  (up to an error of  $\epsilon$ )? There is no reason why it should be larger than the probability of hitting the earth at  $24.35219^\circ$ , because nature does not care about whether the numerical value of the location can be computed by a short program. The physical laws that govern lightning fulfill some symmetries that should be respected by our prior. To construct a prior that accounts for these symmetries and still captures the aspect of description length, we propose to use a computation model that is inherently symmetric with respect to some transformations.

2. *Complexity of isolating systems:* According to Solomonoff’s prior, any state having a short program as description is likely to occur in nature, no matter whether the running time is large or not and no matter how robust the output is with respect to perturbing the state of the Turing machine during the computation. Physical prior probability should also account for the *robustness* of the computation process since no system is perfectly isolated from its environment. Physically universal CAs are good models to take this into account because the coupling between system and its environment is always present by definition.

We now define a prior via a physically universal CA. A naive analog of randomizing the input of the Turing machine would be to initialize the CA to the uniform distribution over all pure states and then applying the dynamics  $\alpha_t$ , yields a trivial prior for every  $t$  since our bijective dynamics preserves the uniform distribution. We want to define a prior that gives higher probability to *simple* patterns like  $0^R$  (all cells in  $R$  are in the state 0). It will therefore be based on the following initial state:

**Definition 12 (initial state of the universe)**

Let  $L = L_+ \cup L_-$  be a partition of the lattice into two infinite subsets ( $L_+$  could, for instance, be all cells with  $x_1 > 0$ ). Define a probability measure  $Q$  by setting all sites in  $L_-$  to zero and choosing the uniform distribution on  $A^{L_+}$  (i.e. for every site in  $L_+$ , a symbol is chosen independently with probability  $1/|A|$  each).

We consider  $L_+$  and  $L_-$  as hot and cold parts of the world, respectively. Then, interesting structure can only start growing at the boundary between hot and cold regions. This accounts for the fact that life requires thermal non-equilibrium, which is most naturally provided by temperature gradients.

Such a state ensures the availability of an infinite amount of free memory space. – A similar convention would also be required for Solomonoff’s prior if it was defined with respect to a *reversible* Turing machine [42]. Then one would also need to provide free memory space for free in order to ensure that the string  $0^k$  obtains a higher prior probability than a typical  $k$ -bit string. We now define:

**Definition 13 (physical prior)**

For every time  $t \in \mathbb{N}$ , let  $P_t$  be the probability distribution on  $S$  that is obtained by applying  $\alpha_t$  to the initial mixed state  $Q$ , as given by Definition 12.

Let us discuss some properties of  $P_t$ . As opposed to Solomonoff’s prior, it depends on  $t$ . This is because the Turing machine stops for appropriate inputs whereas the dynamics of our CA does not. It is not clear whether one should consider this as a feature rather than as a drawback of our definition – one may argue that in the early stage of the universe other states were more likely than today and others were less likely. Note, however, that

$$P(c) \propto \sum_{t \in \mathbb{N}_0} P_t(c) 2^{-K(t)}$$

would be an option to define a time-independent prior. To elaborate on this goes beyond the scope of this paper, but the additional term  $K(t)$  will also appear in our definition of physical complexity below.

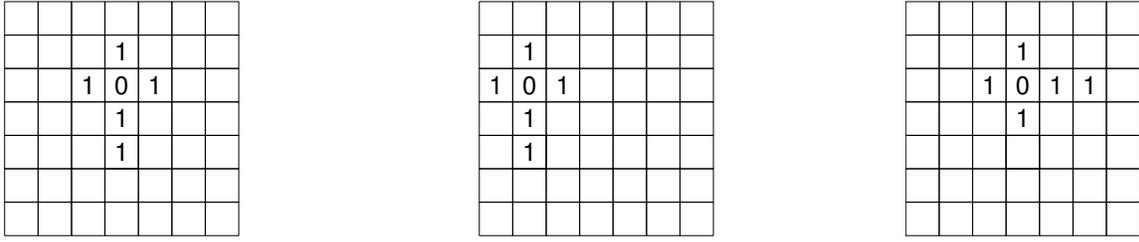


Figure 1: The prior probability of the configuration on the left is the same as for its shifted copy (middle) and its rotated copy (right) provided that these transformations preserve the partition into  $L_+$  and  $L_-$ . For instance, if  $L_+$  is the set with  $x_1 > 0$ , then it is allowed to shift along the  $x_1 = 0$  axis. In a lattice with  $d \geq 3$ , there are also rotations that preserve the axis.

A second feature of  $P_t$  is that the prior probability of a configuration (and also the physical complexity that we define below) depends on its location on the lattice: creating a cold region in the middle of the hot region involves much more sophisticated initialization than creating it close to the boundary to the cold region. In the former case, the entropy of the hot region needs to be transported over a long way to the cold region.

Recalling our motivation for defining a prior different from Solomonoff's, we note that  $P_t$  indeed respects some of the symmetries of physical laws. Consider, for some time  $t$ , the probability  $P_t$  of the pattern in Fig. 1, left, consisting of symbols 1 and 0. The empty squares indicate cells whose value is unspecified. Fig. 1, middle, and right, show shifted and rotated copies of the same pattern, respectively. If the shift and the rotation are chosen such that they leave  $L_{\pm}$  invariant, then  $P_t$  is obviously the same for these copies.

To discuss item 2 in the above list of desired modifications, we assume that the generation of some  $c$  requires only a short program on a Turing machine but one needs to initialize a large region  $R_p$  to generate it on a physically universal CA. One reason could be that it involves a long and fragile computation process which only outputs the correct result if a large environment is correctly initialized. Then  $P_t(c)$  would be small for all  $t$ .

In the spirit of Solomonoff's prior, we would like to ensure that every  $c \in A^R$  (for an arbitrary finite region  $R$ ) gets non-zero probability for some  $t \in \mathbb{N}_0$ . Note that the maximally mixed state on  $L_+$  can be interpreted as a mixture over "random programs", and it is not clear whether programs on  $L_+$  are sufficient for preparing any desired configuration (also on  $L_-$ ). It could be that this defines an even stronger kind of physical universality. This problem will also be left open.

To define a physical analog of Kolmogorov complexity we first discuss why the following straightforward definition is inappropriate for our purposes: For any  $c \in A^R$  one could define the complexity of  $c$  as the size of the region  $R_p$  for which there is a

state  $c_p \in A^{R_p}$  such that

$$\exists t \in \mathbb{N}_0 \quad \alpha_t(c_p, e)|_R = c \quad \forall e \in A^R.$$

Then the complexity of  $c$  is at least  $|R|$  because  $\alpha_t$  is a bijection. We want to define complexity of a state in such a way that simple patterns like  $0^R$  have low complexity. Unfortunately, we are not able to show that this will be the case for the complexity measure below, but there is at least no obvious argument why it cannot be (as opposed to the above measure). The decisive assumption that we make is that  $R_p$  must be contained in  $L_+$ . This is in agreement with the fact that our “random programs” that define  $P_t$  are contained in  $L_+$  while  $L_-$  only contains free memory space.

Even though we must leave it open, whether *every* configuration can be prepared by programs in  $L_+$  (see also the remarks above regarding the physical prior), we now define the “program size complexity”, but we phrase it more general and define the complexity of processes other than state preparation:

**Definition 14 (physical complexity)**

Let  $R$  be some region and

$$M : A^R \rightarrow A^R$$

be an arbitrary map. The physical complexity of  $M$  is defined by the minimum

$$C(M) := \min\{|R_M| \log_2 |A| + K(t)\},$$

where the minimum is taken over all  $t \in \mathbb{N}$  and all regions  $R_M$  and initializations  $c_i \in S^{R_M}$  for which

$$\alpha_t(c_M, c_i)|_{R_M} = M(c_i).$$

The physical complexity  $C(c)$  of a configuration  $c$  is defined by the complexity of the map  $M$  with  $M(c_i) = c$  for all  $c_i \in A^R$ .

The additional term  $K(t)$  will later be needed to ensure that our complexity measure satisfies Kraft’s inequality. A more intuitive justification may be that the time parameter must be provided as *external* information. Since there is probably no finite initialization that prepares a state and keeps it forever, we must be told when the desired state is present or the desired transformation is performed. The following relation between physical complexity and the physical prior is almost obvious:

**Lemma 4 (lower bound on physical complexity)**

$$C(c) \geq \min_{t \in \mathbb{N}_0} \{-\log_2 P_t(c) + K(t)\}. \tag{8}$$

Proof: By definition of the physical prior,

$$P_t(c) \geq |A|^{-|R_p|}, \tag{9}$$

for all  $c_p$  that prepare  $c$  after the time  $t$ . By definition of physical complexity,

$$C(c) = \min_{t \in \mathbb{N}_0} \min_{c_p} \{|R_p| \log_2 |A| + K(t)\},$$

where the minimum is taken over all  $c_p$  that prepare  $c$  after the time  $t$ . Using (9) we obtain

$$C(c) \geq \min_{t \in \mathbb{N}_0} \{-\log_2 P_t(c) + K(t)\}. \quad (10)$$

□

Rather than having inequality (8) only one may wish to show a tighter link between the physical prior and physical complexity – in analogy to the tight connection between Solomonoff's prior and Kolmogorov complexity [43]:

**Theorem 7 (Coding Theorem of Levin)**

$$-\log m(s) = K(s) + O(1),$$

where  $O(1)$  means that the error can be bounded by a constant that depends on the Turing machine, but does not depend on  $s$ .

Hence,  $m(s) \approx 2^{-K(s)}$  up to a multiplicative term that is bounded by some constant. For this reason, tighter connections between the physical prior and physical complexity are desirable.

The following theorem describes a mathematical property of physical complexity that it shares with Kolmogorov complexity:

**Theorem 8 (Kraft's inequality)**

Let  $U$  be a set of mutually exclusive configurations of arbitrary size. Then, physical complexity satisfies

$$\sum_{c \in U} 2^{-C(c)} \leq 1.$$

Proof: Let  $t_c$  be the time that minimizes the right hand side of (10), hence

$$C(c) \geq -\log_2 P_{t_c}(c) + K(t_c).$$

We conclude

$$\begin{aligned} \sum_{c \in U} 2^{-C(c)} &\leq \sum_{c \in U} P_{t_c}(c) 2^{-K(t_c)} \\ &\leq \sum_{t \in \mathbb{N}_0} \sum_{c \in U} P_t(c) 2^{-K(t)} \\ &\leq \sum_{t \in \mathbb{N}_0} 2^{-K(t)} \leq 1, \end{aligned}$$

where the second last inequality holds because the configurations are mutually exclusive and the last step uses the usual Kraft inequality for Kolmogorov complexity. □

The fact that Kolmogorov complexity satisfies Kraft's inequality (which was not the case in Kolmogorov's version since he did not use prefix codes) made it possible to renormalize it to a probability distribution on strings, yielding Solomonoff's prior.

Although a better understanding of our notion of physical complexity has to be left to the future, it is, by construction, clear that it takes into account whether running a

process requires to adjust a large part of the environment – even though the process may be simple from the point of view of algorithmic information. Such a strong disagreement between Kolmogorov complexity and physical complexity occurs e.g. if  $R_p$  is large but  $c_p$  mainly consists of zeros, or some other algorithmically simple pattern. If a physical process requires, for instance, cooling a large region (e.g. setting many cells to zero) around the system this could formally appear as large physical complexity.

### 3 Physical universality in the quantum world

#### 3.1 Informal description of some differences to the classical case

The main question that arises when we translate the notion of universal state preparation into the quantum world is whether the configuration of the environment is supposed to be a *basis* state. In other words, we ask whether the preparation of general quantum superposition should be reducible to the preparation of basis states in the environment.

On the one hand, it seems to be artificial to select a certain subset of states as being more fundamental than others. On the other hand, the following model suggests that basis states should be sufficient: we could think of the basis states as states in the register of a classical processor that controls a quantum preparation machine. Then the register is the region that we act on by changing its *classical* state only.

#### 3.2 Defining the problem

To formally define quantum CAs, we assume that every site  $x \in L$  contains a quantum system with Hilbert space  $\mathcal{H} := \mathbb{C}^a$ , where  $a := |A|$  and the basis vectors  $|j\rangle$  are labelled by symbols  $j \in A$ . The Hilbert space of a region  $R$  is then given by the tensor product of copies of  $\mathcal{H}$ , but to avoid problems with infinite tensor products we follow [44] and use an operator algebraic framework [45, 46]: Let every site  $x$  be described by a copy of the same matrix algebra  $\mathcal{A}_x$  of  $a \times a$  matrices. The self-adjoint part of  $\mathcal{A}_x$  is interpreted as the observables corresponding to site  $x$ . For every finite set  $\Lambda \in L$ , let  $\mathcal{A}_\Lambda$  be the tensor product

$$\mathcal{A}_\Lambda := \bigotimes_{x \in \Lambda} \mathcal{A}_x.$$

For  $\Lambda \subset \Lambda'$ ,  $\mathcal{A}_\Lambda$  is considered as subalgebra of  $\mathcal{A}_{\Lambda'}$  in a canonical way by adding the tensor product of an appropriate number of  $a \times a$  identity matrices. For every infinite set  $\Lambda$ , we define  $\mathcal{A}_\Lambda$  as the  $C^*$ -completion over the union of algebras of finite regions  $\mathcal{A}_{\Lambda_f}$ . This defines the  $C^*$ -algebra  $\mathcal{A}_L$  which contains all local algebras<sup>6</sup>.

The set  $S(\mathcal{A}_L)$  of states is the set of positive linear functionals  $\phi : \mathcal{A}_L \rightarrow \mathbb{C}$  with  $\phi(\mathbf{1}) = 1$ . The state space  $S(\mathcal{A}_L)$  is a convex set whose extreme points are called pure states, this definition generalizes density operators of rank one to the infinite system.

---

<sup>6</sup>the “quasi-local” algebra [45]

A pure state  $\phi$  is said to be a basis state on a region  $R$  if it is given by

$$\phi(a) = \text{tr}(\rho a) \quad \forall a \in \mathcal{A}_R,$$

where  $\rho$  is a *diagonal* matrix with diagonal  $(0, \dots, 0, 1, 0, \dots, 0)$ . A pure state is said to be a (global) basis state if its restriction to every finite region is a basis state.

It is convenient to describe the dynamics in the Heisenberg picture, it is then given by a group  $(\alpha_t)$  of  $C^*$ -automorphisms of  $\mathcal{A}_L$  satisfying the following locality condition:

$$\phi(\mathcal{A}_x) \in \mathcal{A}_R,$$

for every region  $R$  that contains the Moore neighborhood of  $x$  with radius one. The dynamics transfers the state  $\phi$  into  $\phi \circ \alpha_t$ . For any observable  $a \in \mathcal{A}_R$  for which  $\alpha_t(a) \in \mathcal{A}_{R'}$  for some region  $R'$ , the value  $(\phi \circ \alpha_t)(a)$  is already determined by the restriction of  $\phi$  to  $R'$ . Therefore,  $(\rho \circ \alpha_t)(a)$  is also a well-defined expression if  $\rho$  is a state on  $\mathcal{A}_{R'}$ .

The following notion of physical universality can be seen as a quantum analog of Definition 2 to the quantum world. As opposed to the set of classical configurations of a finite region, the set of pure states is (uncountably) infinite. On the other hand, the set of basis states of a region  $R_p$  is finite and the set of all basis states of the whole lattice still is countable, we cannot prepare all states on  $R$  exactly but at most up to any desired accuracy:

**Definition 15 (conditional quantum state preparation)**

A quantum CA is said to allow for conditional state preparation if for every pair of states  $(\rho_i, \rho_f) \in S(\mathcal{A}_R) \times S(\mathcal{A}_R)$  of a region  $R$  and every  $\epsilon > 0$  there is a basis state  $\gamma \in S(\mathcal{A}_{L \setminus R})$  of the complement and a time  $t$  such that

$$|(\gamma \otimes \rho_i) \circ \alpha_t(a) - \rho_f(a)| \leq \epsilon \|a\| \quad \forall a \in \mathcal{A}_R,$$

where  $\|\cdot\|$  denotes the operator norm.

It is important to note that the program state  $\gamma$  is a basis state, i.e., the program is *classical* software. As opposed to the classical case, this notion of universality is not satisfied by the “trivial” CA that only shifts the state. Instead, it includes problems like how to prepare sophisticated multi-particle entanglement using a given interaction via preparing the environment to *basis* states. We thus formulate the following open problem:

**Question 3 (physically universal quantum CA)**

*Is there a quantum CA that is physically universal in the sense of Definition 15?*

We will not translate Definitions 3 and 4 to the quantum setting since even our “weak” form of universality is not obvious to exist for quantum CAs.

### 3.3 Physically universal Hamiltonians

To account for the fact that time evolutions are actually continuous, we may want to switch from CAs to Hamiltonians. In the literature there exists a large number of translation invariant finite range Hamiltonians on lattices that are universal for quantum computing, e.g., [47, 48, 49, 50], but physical universality has not been considered. A characteristic feature of many constructions for computational universal Hamiltonians is the separation between a “program region” and a “data region” where the former controls the operations performed on the latter. Physical universality would imply that we are also able to operate on the program region, which could require an infinite hierarchy of program regions. To formally define physical universality, we can straightforwardly adapt Definition 15 by replacing the group  $(\alpha_t)_{t \in \mathbb{Z}}$  with the continuous version  $(\alpha_t)_{t \in \mathbb{R}}$ . To properly state what it means that a dynamics of an infinite lattice is given by a finite range translation invariant Hamiltonian we consider an operator  $h \in \mathcal{A}_R$  for some region  $R$  and define for every vector  $x \in \mathbb{Z}^d$ , the shifted copy of  $h$  by  $\tau_x(h)$ . Then it is known that the differential equation

$$\frac{d}{dt} \alpha_t(a) = i \left[ \sum_{x \in \mathbb{Z}^d} \tau_x(h), a \right] \quad (11)$$

defines uniquely a group of  $C^*$ -automorphisms [46]. Definition 15 and, correspondingly, Question 3 then straightforwardly translate to the group  $\alpha_t$  defined by (11).

The considerations on the thermodynamic costs change more significantly because we replace the maximum entropy state by the state of minimum free energy, i.e., the Gibbs state (for defining thermal equilibrium states for infinite lattices see [46]), which ensures that we are getting closer to real physics. We may then even allow for lattices having an infinite dimensional algebra at each site. We also want to translate the physical prior and the physical complexity in Subsection 2.5. Now, the notion of hot and cold parts is taken more literally than above since the definition of Hamiltonians allows us to defined thermal states for temperatures other than  $T = 0$  and  $T = \infty$ . Thermal equilibrium states on infinite quantum lattice systems can be defined via limits of Gibbs states for finite regions [46] (we do not care about the potential non-uniqueness of limit points here). We restrict these states of the infinite lattice to  $L_+$  and  $L_-$ , respectively and “glue” them together to define our initial state:

**Definition 16 (initial state of the universe)**

Let  $\phi_T : \mathcal{A}_L \rightarrow \mathbb{C}$  be Gibbs states for temperature  $T$  on the entire lattice. For some  $T_2 > T_1 > 0$ , let  $\phi_+$  be the restriction of  $\phi_{T_2}$  to  $\mathcal{A}_{L_+}$  and  $\phi_-$  the restriction of  $\phi_{T_1}$  to  $\mathcal{A}_{L_-}$ . Then we define the “initial state of the universe” by

$$\phi := \phi_+ \otimes \phi_- .$$

**Definition 17 (physical prior for Hamiltonian systems)**

For every  $t$  we define the mixed state

$$\phi_t := \phi \circ \alpha_t .$$

Let  $|\psi\rangle$  be the state vector of some pure state on  $\mathcal{A}_R$ . Then

$$\phi_t(|\psi\rangle\langle\psi|)$$

is the probability for obtaining the state  $|\psi\rangle\langle\psi|$  after the time  $t$  when measuring a non-degenerate self-adjoint operator that contains  $|\psi\rangle$  as one of its eigenvectors.

In the spirit of Solomonoff’s prior, we would like to give higher prior to states that are simple in an intuitive sense than to complex ones. For instance, we would consider the basis state  $|0\rangle\langle 0|^R$  (i.e., all cells in the region  $R$  are in the state  $|0\rangle\langle 0|$ ) as simple. It is possible that a small program makes the Hamiltonian dynamics generating free memory space via using the temperature gradient. This is at least not forbidden by any obvious thermodynamic laws. Thermodynamics also allows for processes that use the existing temperature gradient to either lower the temperature of some region in  $L_-$  (refrigerator driven by a heat engine, see also [23]) or increase the temperature of  $L_+$  even further. The size of the program required to make  $\alpha_t$  implementing such a process would then be the physical complexity of the process. This is only meant to be one of many examples how physically universal CAs define the complexity of physical processes, no matter whether they are computation processes or not.

An interesting modification of the above would be given by replacing the lattice with a field-theoretic model, where nets of subalgebras  $\mathcal{A}_\Lambda$  are assigned to regions in  $\mathbb{R}^d$  [51] and define physical universality for a field theory. As opposed to the discrete model, this would allow for the definition of an even “more physical” prior that is invariant under the full Lorentz group.

## 4 Conclusions

The main contribution of this paper is to introduce and motivate the concept of physically universal CAs and Hamiltonians. Their non-existence would probably have interesting consequences for the limits of controlling microscopic systems. But also their existence poses questions that are equally fundamental, because such CAs are nice models for studying the thermodynamic cost of computation and control.

We also use physically universal CAs to define the complexity of states and a corresponding prior probability that is considered as a physically motivated analog of Solomonoff’s prior. An interesting feature of this prior is that it is invariant under some physical symmetries. Moreover, it tries to capture the amount of adjustments that is needed in the environment to run a preparation process, which includes also the cost of removing disturbing heat and the cost of keeping it away from the system during the implementation of the process.

The author would like to thank Bastian Steudel and David Balduzzi for helpful comments on an earlier draft and Aram Harrow and Armen Allahverdyan for interesting discussions. This work has partially been supported by the VW-project “Quantum Thermodynamics: energy and information flow at nanoscale”.

## References

- [1] J. Jauch. *Foundations of quantum mechanics*. Addison-Wesley, Reading, Mass., 1968.
- [2] E. Davies. *Quantum theory of open systems*. Academic Press, London, 1976.
- [3] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [4] D. DiVincenzo. Two-qubit gates are universal for quantum computation. *Phys. Rev A*, 51:1015–1022, 1995.
- [5] D. Janzing and Th. Beth. Remark on multi-particle states and observables with constant complexity. <http://arxiv.org/abs/quant-ph/0003117>.
- [6] A. Soklakov and R. Schack. Efficient state preparation for a register of quantum bits. <http://arxiv.org/abs/quant-ph/0408045>.
- [7] P. Wocjan, D. Janzing, and Th. Decker. Measuring 4-local n-qubit observables could probabilistically solve PSPACE. *Quantum Information and Computation*, 4(8 & 9):741–755, 2008.
- [8] D. Janzing. *Computer science Approach to Quantum control*. Habilitationsschrift, UniVerlag Karlsruhe. 2006.
- [9] D. Janzing, F. Armknecht, R. Zeier, and Th. Beth. Quantum control without access to the controlling interaction. *Phys. Rev. A*, 65:022104, 2002.
- [10] D. Janzing and Decker T. How much is a quantum controller controlled by the controlled system? *Applicable Algebra in Engineering, Communication and Computing*, 19(3), 2008.
- [11] A. Allahverdyan, D. Janzing, and G. Mahler. Thermodynamic efficiency of information and heat flow. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(09):P09011 (35pp), 2009.
- [12] A. Allahverdyan and D. Janzing. Relating the thermodynamic arrow of time to the causal arrow. *J. Stat. Mech.*, page P04001, 2008.
- [13] K. Zuse. Calculating space. *MIT Technical Translation*, German Original: Rechner Raum, Friedrich Vieweg & Sohn, Braunschweig, 1969, 1970. Cambridge, Ma.
- [14] E. Fredkin. An informational process based on reversible universal cellular automata. *Physica D*, 45(1–3):254–270, 1990.
- [15] D. Cheung. Cellular automata as a model of physical systems. *Journal of Cellular Automata*, 5(6):469–480, 2010.
- [16] J. Kari. Theory of cellular automata: a survey. *Theoretical Computer Science*, 344:3–33, 2005.
- [17] B. Farkas, T. Eisner, M. Haase, and R. Nagel. *Ergodic Theory – An Operator-theoretic Approach*. 12th International Internet Seminar, 2009. <http://isem.mathematik.tu-darmstadt.de/isem/InternetSeminar?action=AttachFile&do=get&target=isem08-announcement.pdf>.

- [18] P.R. Halmos. *Lectures on Ergodic Theory*. AMS Chelsea Publishing, New York, 1956.
- [19] S. Richter and Werner. R. Ergodicity of quantum cellular automata. *Journal of Statistical Physics*, 82(3–4):963–998, 1996.
- [20] M. Shirvani and T. Rogers. On ergodic one-dimensional cellular automata. *Communications in Mathematical Physics*, 136:599–605, 1991.
- [21] S. Willson. On the ergodic theory of cellular automata. *Theory of Computing Systems*, 9(2):132–141, 1975.
- [22] D. Wolpert. Physical limits of inference. *Physica D*, 237:1257–1281, 2008.
- [23] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and Th. Beth. Thermodynamic cost of reliability and low temperatures : Tightening Landauer’s principle and the Second Law. *Int. Jour. Theor. Phys.*, 39(12):2217–2753, 2000.
- [24] Ch. Papadimitriou. *Computational Complexity*. Addison Wesley, Reading, Massachusetts, 1994.
- [25] E. Berlekamp, J. Conway, and R. Guy. *Winning ways for your mathematical plays*. Academic Press, New York.
- [26] D. Janzing. On the computational power of molecular heat engines. *J. Stat. Phys.*, 122(3):531–556, 2006.
- [27] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Develop.*, 5:183–191, 1961.
- [28] C. Bennett. The thermodynamics of computation – a review. *Int. J. Theor. Phys.*, 21:905–940, 1982.
- [29] M. Ohya and D. Petz. *Quantum entropy and its use*. Springer Verlag, 1993.
- [30] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer, New York, 1997 (3rd edition: 2008).
- [31] W. Zurek, editor. *Complexity, entropy and the physics of information*. Addison-Wesley, 1990.
- [32] A. Kolmogorov. Three approaches to the quantitative definition of information. *Problems Inform. Transmission*, 1(1):1–7, 1965.
- [33] R. Solomonoff. A formal theory of inductive inference. *Information and Control, Part II*, 7(2):224–254, 1964.
- [34] G. Chaitin. On the length of programs for computing finite binary sequences. *J. Assoc. Comput. Mach.*, 13:547–569, 1966.
- [35] W. Zurek. Algorithmic randomness and physical entropy. *Phys Rev A*, 40(8):4731–4751, 1989.
- [36] C. Mora, B. Kraus, and H. Briegel. Quantum Kolmogorov complexity and its applications. *International Journal of Quantum Information*, 5:729–750, 2007.
- [37] M. Hutter. On universal prediction and Bayesian confirmation. *Theoretical Computer Science*, 384(1):33–48, 2007.

- [38] P. Grünwald. *The minimum description length principle*. MIT Press, Cambridge, MA, 2007.
- [39] J. Lemeire and E. Dirks. Causal models as minimal descriptions of multivariate systems. <http://parallel.vub.ac.be/~jan/>, 2006.
- [40] J. Lemeire and K. Steenhaut. Inference of graphical causal models: Representing the meaningful information of probability distributions. *Journal of Machine Learning Research, Workshop and Conference Proceedings*, 6:107–120, 2010.
- [41] D. Janzing and B. Schölkopf. Causal inference using the algorithmic Markov condition. to appear in *IEEE Transactions on Information Theory*. See also <http://arxiv.org/abs/0804.3678>.
- [42] C. H. Bennett. Time/space trade-offs for reversible computation. *SIAM J. Computing*, 18(4):766–776, 1989.
- [43] L. Levin. Laws of information conservation (non-growth) and aspects of the foundation of probability theory. *Problems Information Transmission*, 10(3):206–210, 1974.
- [44] D. Gross, V. Nesme, H. Vogts, and R. Werner. Index theory of one dimensional quantum walks and cellular automata. <http://arxiv.org/abs/0910.3675>.
- [45] O. Bratteli and D. Robinson. *Operator algebras and quantum statistical mechanics*, volume 1. Springer, New York, 1987.
- [46] O. Bratteli and D. Robinson. *Operator algebras and quantum statistical mechanics*, volume 2. Springer, New York, 1987.
- [47] D. Janzing and P. Wocjan. Ergodic quantum computing. *Quant. Inf. Process.*, 4(2):129–158, 2005.
- [48] D. Janzing. Spin-1/2 particles moving on a 2D lattice with nearest-neighbor interactions can realize an autonomous quantum computer. *Phys. Rev.*, A:012307, 2007.
- [49] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe. The power of quantum systems on a line. *Comm. Math. Physics*, 287(1):41–65, 2009.
- [50] K. Vollbrecht and I. Cirac. Quantum simulators, continuous-time automata, and translationally invariant systems. *Phys Rev Lett*, 100:010501, 2008.
- [51] Rudolf Haag. *Local quantum physics: fields, particles, algebras*. Texts and monographs in physics. Springer, Berlin; Heidelberg, 1992.