Reversible Cellular Automata in Presence of Noise Rapidly Forget Everything

Siamak Taati ✉ 🏠 🆔
Department of Mathematics, American University of Beirut, Lebanon

Abstract

We consider reversible and surjective cellular automata perturbed with noise. We show that, in the presence of positive additive noise, the cellular automaton forgets all the information regarding its initial configuration exponentially fast. In particular, the state of a finite collection of cells with diameter n becomes indistinguishable from pure noise after $O(\log n)$ time steps. This highlights the seemingly unavoidable need for irreversibility in order to perform scalable reliable computation in the presence of noise.

1 Introduction

1.1 Background

A major challenge regarding the physical implementation of computation is the inevitability of transient errors due to noise. The difficulty is that, even if each component of the system is built to be highly accurate and has a very small probability of error, in a lengthy computation, occasional errors are bound to occur. Such errors may then propagate and entirely corrupt the computation. The problem of how to perform computation reliably in the presence of noise (via suitable error-correcting mechanisms) goes back to von Neumann and has been studied since [36, 12].

At the nanoscopic scale, the issue of thermal noise becomes even more pressing. Not only are the nano-scale components more sensitive to any sort of fluctuations, but also the heat generated by the computational process has little time and space to escape the system, and thus leads to an increase in thermal noise. Landauer argued that the heat generated by a computational process is associated to its logical irreversibility, and identified a theoretical lower bound for the amount of heat dissipated in the process of erasing a bit of information [18]. Bennett, Fredkin and Toffoli showed that, at least in theory, any computation can be efficiently simulated by a logically reversible one, hence requiring virtually no dissipation [2, 3, 9, 4]. The output of the reversible computation will consist of the intended output as well as some extra information that allows one to trace the computation backwards. If need be, this extra information can be erased away from the computer core, hence avoiding the accumulation of heat.

By the virtue of their "physics-like" features, cellular automata (CA) have been a popular mathematical model for studying the physical aspects of computation. The notion of reversibility has a natural formulation in the setting of CA, and reversible CA have been widely studied, not only as models of reversible computers, but also as models of physical

processes and from other points of view [30, 35, 22, 31, 26, 32, 16, 7, 25, 29, 28, 17]. The reliability of computation in the presence of noise is also studied in the setting of cellular automata [33, 10, 13, 5, 6, 11, 23, 14, 20].

Although logical reversibility solves the issue of heat generation in a computational process, it leads to another difficulty. Even if it does not dissipate heat itself, the process is still exposed to external noise. This external noise can potentially be reduced with proper insulation but can never be eliminated altogether. The logical reversibility of the process entails that the noise entering the system is not dissipated and hence accumulates inside the system. This means that, unless one finds a clever workaround, the state of the system will eventually be overcome by noise [3].

For a model of noisy reversible circuits, Aharonov, Ben-Or, Impagliazzo and Nisan [1] identified the limitation imposed by the accumulation of noise. They considered reversible circuits in which errors occur on wires at each "time unit", and proved that the output of such a circuit is indistinguishable from pure noise unless the size of the circuit is exponential in its depth. Conversely, they showed that every classical Boolean circuit with size s and depth d can be simulated by a noisy reversible circuit with size $O\big(s \times 2^{O(d)}\big)$ and depth $O(d)$. In particular, polynomial-size noisy reversible circuits have the power of the complexity class $\mathbf{NC}^1$. They also proved a similar (though not as sharp) result concerning noisy quantum circuits.

The result we present here formulates a similar limitation imposed by the accumulation of noise in the setting of cellular automata. We show that a reversible CA subject to positive additive noise forgets its initial configuration exponentially fast, in the sense that the state of any finite collection of its cells with diameter n becomes indistinguishable from pure noise after $O(\log n)$ number of time steps. It remains open whether any meaningful computation can be done reliably with such a limitation.

Mathematically, the forgetfulness of a reversible CA subject to noise corresponds to the exponential ergodicity of the resulting probabilistic CA. This means that the distribution of the process converges exponentially fast to a unique invariant measure, which in this case is the uniform Bernoulli measure (i.e., the distribution of a configuration chosen by independent coin flips). This result improves upon an earlier partial ergodicity result [20], which was limited to shift-invariant initial measures, and in which the rate of convergence was not identified. The exponential ergodicity of noisy reversible CA is a special case of a more general ergodicity result concerning positive-rate probabilistic CA with Bernoulli invariant measures and their asynchronous counterparts [21].

The structure of the paper is as follows. In Section 1.2, we introduce the setting, and in Section 1.3 we state the main result. The proof of the main result, which is based on entropy, appears in Section 2. Section 3 is dedicated to the interpretation of the result regarding the rapid information loss of reversible computers in the presence of noise. The paper is concluded with some discussions in Section 4.

1.2   Setting

General notation

We will use the notation $\mathbb{N} \triangleq \{0, 1, 2, \ldots\}$ and $\mathbb{Z}^+ \triangleq \{1, 2, 3, \ldots\}$. We will write $[\![n, m]\!]$ to denote the integer interval $\{n, n+1, \ldots, m\}$. We use the notation x_A for the restriction of a function x to a subset A of its domain. We write $Z \sim q$ to indicate that Z is a random variable with distribution q. The total variation distance between two probability distributions p and

q on a finite set A is

$$\|q - p\|_{\text{TV}} \triangleq \sup_{E \subseteq A} |q(E) - p(E)| = \frac{1}{2} \sum_{a \in A} |q(a) - p(a)| \ .$$

Throughout this article, $\log(\cdot)$ stands for the natural logarithm.

Cellular automata

Cellular automata (CA) are abstract models of massively parallel computation. A *configuration* of the model is an assignment of symbols from a finite alphabet Σ to every site of the lattice $\mathbb{Z}^d$ (for $d = 1, 2, \ldots$). The sites of the lattice are called *cells* and the symbol on each cell is referred to as its *state*. At each step of the computation, the states of all cells are simultaneously updated according to a *local rule*. The local rule takes into account the current state of the cell to be updated as well as its neighbours. More specifically, the local rule is a function $f : \Sigma^{\mathcal{N}} \to \Sigma$, where $\mathcal{N} \subseteq \mathbb{Z}^d$ is a finite set indicating the relative positions of the *neighbours* of each cell, possibly including the cell itself. A configuration $x \in \Sigma^{\mathbb{Z}^d}$ is updated to a configuration $Fx \in \Sigma^{\mathbb{Z}^d}$, where $(Fx)_i \triangleq f\big((x_{i+a})_{a \in \mathcal{N}}\big)$ for each cell $i \in \mathbb{Z}^d$. We refer to F as the *global map* of the CA. The computation thus consists in iterating the global map F on an initial configuration. We identify a CA with its global map F and speak of the CA F.

Surjectivity, injectivity, and reversibility

A CA is said to be *surjective* (resp., *injective*, *bijective*) if its global map is surjective (resp., injective, bijective). It is well-known that every injective CA is also surjective, and hence bijective. In fact, the Garden-of-Eden theorem states that surjectivity is equivalent to pre-injectivity [24, 27]. A CA F is said to be *pre-injective* if whenever two distinct configurations x and y agree on all but finitely many cells, their images Fx and Fy are distinct.

A CA F is said to be *reversible* if F is an invertible map and F^{-1} is itself a CA. It follows from a topological argument that every bijective CA is automatically reversible [15]. Thus, injectivity, bijectivity and reversibility are equivalent conditions.

Noise

In this paper, we are concerned with the effect of transient noise on the computation carried out by a CA. In the presence of noise, random errors might occur during the updates of the cells.

We restrict ourselves to a specific model of noise, namely *additive noise*, although the results of the current paper remain true with the somewhat more general model of *permutation noise* [20]. We (arbitrarily) identify Σ with an Abelian group $(\Sigma, +)$. Subject to an additive noise with *noise distribution* q, a symbol a is replaced with $a + Z$, where Z is a random variable with distribution q. We will assume that the noise distribution q is strictly positive. In particular, $\mathbb{P}(a + Z = b) = q(b - a) > 0$ for every $a, b \in \Sigma$.

At each time step of the computation, the state of each cell is first updated according to the local rule of the CA and is then perturbed with positive additive noise. The noise variables at different cells and different time steps are all assumed to be independent.

More specifically, the noise is described by a family $(Z_i^t)_{i \in \mathbb{Z}^d, t \in \mathbb{Z}^+}$ of independent random variables with distribution q. The trajectory of the noisy computation starting from a configuration x is given by a sequence of random configurations $(X^t)_{t \in \mathbb{N}}$, where $X^0 \triangleq x$, and

$$X_i^t \triangleq f\big((X_{i+a}^{t-1})_{a \in \mathcal{N}}\big) + Z_i^t$$

for every time step $t > 0$ and every cell $i \in \mathbb{Z}^d$.

Probabilistic CA

The noisy computation can be described by a probabilistic CA. In a *probabilistic* CA (PCA), the local rule is probabilistic and the updates at different cells and different time steps are performed independently. More specifically, the local rule is given by a stochastic matrix $\varphi : \Sigma^{\mathcal{N}} \times \Sigma \to [0,1]$ (hence, $\sum_{b \in \Sigma} \varphi(u, b) = 1$ for each $u \in \Sigma^{\mathcal{N}}$). A configuration x is updated to a random configuration Y with distribution

$$\mathbb{P}(Y_A = y_A) = \prod_{i \in A} \varphi\big((x_{i+a})_{a \in \mathcal{N}}, y_i\big)$$

for every finite set $A \subseteq \mathbb{Z}^d$. The role of the global map in the deterministic case is played by a (global) transition kernel Φ where $\Phi(x, \cdot)$ indicates the distribution of Y. (See [34] or [20] for more details.) The trajectory of a PCA Φ is a Markov process with transition kernel Φ, that is, a sequence $(X^t)_{t \in \mathbb{N}}$ of random configurations such that

(i) Given X^t, the configuration X^{t+1} is independent of the configurations $X^0, X^1, \ldots, X^{t-1}$.

(ii) Given X^t, the distribution of X^{t+1} is given by $\Phi(X^t, \cdot)$.

In the case of a CA F with additive noise, the local rule of the resulting PCA is given by

$$\varphi(u, b) \triangleq q\big(b - f(u)\big) \,,$$

where f is the local rule of F and q is the noise distribution.

Ergodicity

We say that a probability measure λ on $\Sigma^{\mathbb{Z}^d}$ is *invariant* under a PCA Φ if $X^{t+1} \sim \lambda$ whenever $X^t \sim \lambda$. We say that Φ is *ergodic* if it has a unique invariant measure λ and furthermore, for any (possibly random) starting configuration X^0, the distribution of X^t converges weakly to λ. This means that for every finite set $A \subseteq \mathbb{Z}^d$ and every $u \in \Sigma^A$,

$$\mathbb{P}(X_A^t = u) \to \lambda\big(\{\hat{x} : \hat{x}_A = u\}\big) \qquad \text{as } t \to \infty.$$

We can interpret the ergodicity of a PCA Φ as Φ "forgetting" its initial configuration. However, the convergence (and hence the process of forgetting the initial configuration) can potentially be slow.

Among the PCA that are ergodic, it is quite common that the convergence towards the unique invariant measure is exponentially fast, in the sense that, for every finite set $A \subseteq \mathbb{Z}^d$,

$$\big\|\mathbb{P}(X_A^t \in \cdot) - \lambda\big(\{\hat{x} : \hat{x}_A \in \cdot\}\big)\big\|_{\mathrm{TV}} \leq \alpha_A \mathrm{e}^{-\beta t} \,,$$

where $\beta > 0$ is a constant (independent of A) and α_A depends on the set A but not on t or the initial configuration $X^0 = x$. (Here, $\mathbb{P}(X_A^t \in \cdot)$ stands for the distribution of X_A^t and $\lambda\big(\{\hat{x} : \hat{x}_A \in \cdot\}\big)$ for the marginal of λ on A.) It is the dependence of α_A on the set A that has more relevant information on the speed of convergence. We will discuss this further in Section 3.

In the current paper, the unique invariant measure of the PCA we study will be the *uniform Bernoulli measure*, that is, the distribution of a random configuration in which the states of different cells are chosen uniformly at random from Σ and independently from one another.

1.3 Statement of the theorem

While the primary interest here is in reversible CA in the presence of noise, our main result remains true for surjective CA. By the *diameter* of a finite set $A \subseteq \mathbb{Z}^d$ we mean the smallest $n \in \mathbb{N}$ such that $A \subseteq u + [\![0, n-1]\!]^d$ for some $u \in \mathbb{Z}^d$.

▶ **Theorem 1** (Surjective CA with additive noise). *Let $F : \Sigma^{\mathbb{Z}^d} \to \Sigma^{\mathbb{Z}^d}$ be a surjective CA. Let Φ be a PCA describing the perturbation of F with positive additive noise. Then, Φ is exponentially ergodic with the uniform Bernoulli measure λ as the unique invariant measure.*

More specifically, there exist constants $\alpha, \beta, a, b > 0$ such that if $(X^t)_{t \in \mathbb{N}}$ is a trajectory of Φ with arbitrary initial configuration, then

$$\left\| \mathbb{P}(X_A^t \in \cdot) - \lambda\big(\{\hat{x} : \hat{x}_A \in \cdot\}\big) \right\|_{\mathrm{TV}} \leq \alpha \mathrm{e}^{-\beta t} n^{(d-1)/2} ,$$

for every finite set $A \subseteq \mathbb{Z}^d$ with diameter n and every $t \geq a \log n + b$.

The above theorem completes an earlier partial result in which the uniqueness and convergence (without rate of convergence) was established only among shift-invariant measures [20]. Theorem 1 is a special case (relevant to reversible computing) of a more general result: every probabilistic CA with strictly positive transition probabilities which has a Bernoulli invariant measure is exponentially ergodic [21].

2 Accumulation of entropy

Like the result of Aharonov et al. [1], the proof of Theorem 1 is based on entropy. Recall that the (Shannon) *entropy* of a discrete random variable X (measured in *nats*) is defined as

$$H(X) \triangleq -\sum_x \mathbb{P}(X = x) \log \mathbb{P}(X = x) .$$

The entropy of X measures the average information content of X. If X takes its values within a finite set Γ, then $H(X) \leq \log|\Gamma|$, with equality if and only if X is uniformly distributed over Γ. We refer to [8] for information on entropy and its properties.

We follow the approach of the earlier proof of ergodicity modulo shift [20]. Namely, we use the fact that positive additive noise increases the entropy of every finite collection of cells, while a surjective CA does not erase the entropy and only "diffuses" it. In order to achieve complete ergodicity with sharp rate of convergence, we use two new ingredients:

(a) An explicit bound on the amount of entropy increase due to noise,

(b) A "bootstrap argument" showing that if the rate of entropy increase is high compared to the rate of entropy leakage, then the entropy of each finite set will inevitably accumulate and rise up to its maximum capacity.

That the proof is based on entropy is natural. A reversible CA in the presence of noise can be thought of as a (microscopically reversible) physical system in contact with a heat bath. The entropy increase is therefore a manifestation of the second law of thermodynamics for such systems.

2.1 Effect of additive noise on entropy

We set $\overline{h} \triangleq \log|\Sigma|$, so that $\overline{h}$ is the highest possible entropy of a Σ-valued random variable. Throughout this section, we also let $q : \Sigma \to (0, 1)$ be a strictly positive probability distribution, and set $\kappa \triangleq |\Sigma| \min_{a \in \Sigma} q(a)$. Note that $0 < \kappa \leq 1$. To avoid trivial situations, we assume that Σ has at least two elements and that q is not uniform. Hence, $\overline{h} > 0$ and $\kappa < 1$.

▶ **Lemma 2** (Effect of additive noise on entropy). *If A and N are independent Σ-valued random variables with $N \sim q$, then*

$$H(A + N) \geq \kappa \overline{h} + (1 - \kappa) H(A) .$$

Proof. The distribution q can be decomposed as

$$q = \kappa u + (1 - \kappa) \tilde{q}$$

where u is the uniform distribution on Σ and $\tilde{q}$ is another distribution on Σ given by $\tilde{q}(a) \triangleq \frac{q(a) - \kappa / |\Sigma|}{1 - \kappa}$ for $a \in \Sigma$. Thus, without loss of generality (by defining a new probability space if necessary), we can assume that $N = BU + (1 - B)\tilde{N}$ where $B \sim \mathsf{Bernoulli}(\kappa)$, $U \sim \mathsf{Uniform}(\Sigma)$ and $\tilde{N} \sim \tilde{q}$, and the variables A, B, U and $\tilde{N}$ are independent.

Using this representation, we have

$$
\begin{aligned}
H(A &+ N) \\
&= H(A + BU + (1 - B)\tilde{N}) \\
&\geq H(A + BU + (1 - B)\tilde{N} \mid B) \\
&= \kappa H(A + \underbrace{BU + (1 - B)\tilde{N}}_{= \, U \text{ when } B = 1} \mid B = 1) + (1 - \kappa) H(A + \underbrace{BU + (1 - B)\tilde{N}}_{= \, \tilde{N} \text{ when } B = 0} \mid B = 0) \\
&= \kappa \overline{h} + (1 - \kappa) H(A + \tilde{N}) .
\end{aligned}
$$

The claim follows once we recall that $H(A + \tilde{N}) \geq H(A)$ whenever A and $\tilde{N}$ are independent Σ-valued random variables. Namely, we have

$$
\begin{aligned}
H(A + \tilde{N}, \tilde{N}) &= H(\tilde{N}) + \overbrace{H(A + \tilde{N} \mid \tilde{N})}^{= H(A)} \\
H(A + \tilde{N}, \tilde{N}) &= H(A + \tilde{N}) + H(\tilde{N} \mid A + \tilde{N})
\end{aligned}
$$

from which we get $H(A + \tilde{N}) = H(A) + I(A + \tilde{N}; \tilde{N})$ where $I(A + \tilde{N}; \tilde{N}) \geq 0$ is the mutual information between $A + \tilde{N}$ and $\tilde{N}$. ◀

The conditional version of the above lemma can be proven similarly, or by reducing it to the unconditional version.

▶ **Lemma 3** (Effect of additive noise on conditional entropy). *If A and N are Σ-valued random variables and C is another random variable conditioned on which A and N are independent with $N \sim q$, then*

$$H(A + N \mid C) \geq \kappa \overline{h} + (1 - \kappa) H(A \mid C) .$$

For a collection of random symbols subjected to independent noise, we have the following lemma as a corollary.

▶ **Lemma 4** (Effect of additive noise on joint entropy). *If $\underline{A} \triangleq (A_1, A_2, \ldots, A_n)$ and $\underline{N} \triangleq (N_1, N_2, \ldots, N_n)$ are two independent collections of Σ-valued random variables and $N_1, N_2, \ldots, N_n$ are i.i.d. with distribution q, then*

$$H(\underline{A} + \underline{N}) \geq n\kappa \overline{h} + (1 - \kappa) H(\underline{A}) .$$

Proof. Using the chain rule and the fact that N_k's are independent of one another and independent A_k's, we have

$$H(\underline{A}) = \sum_{k=1}^{n} H\big(A_k \mid (A_i)_{i<k}\big) \tag{1}$$

and

$$\begin{aligned}
H(\underline{A} + \underline{N}) &= \sum_{k=1}^{n} H\big(A_k + N_k \mid (A_i + N_i)_{i<k}\big) \\
&\geq \sum_{k=1}^{n} H\big(A_k + N_k \mid (A_i)_{i<k}, (N_i)_{i<k}\big) \\
&= \sum_{k=1}^{n} H\big(A_k + N_k \mid (A_i)_{i<k}\big) .
\end{aligned} \tag{2}$$

Applying Lemma 3 to the corresponding terms in (1) and (2) yields the result. ◀

2.2 Effect of a surjective CA on entropy

The effect of surjective CA on entropy was clarified in the earlier proof of ergodicity modulo shift [20]. For completeness, we recall the proof. Given a set $J \subseteq \mathbb{Z}^d$ and an integer $r \geq 0$, we denote by $\mathcal{M}^r(J) \triangleq J + [\![-r, r]\!]^d$ the set of all cells that are within distance r from J. We also define $\partial\mathcal{M}^r(J) \triangleq \mathcal{M}^r(J) \setminus J$.

▶ **Lemma 5** (Effect of a surjective CA on entropy). *Let $F : \Sigma^{\mathbb{Z}^d} \to \Sigma^{\mathbb{Z}^d}$ be a surjective CA with neighbourhood $\mathcal{N} \subseteq [\![-r, r]\!]$. Then, for every random configuration X and every finite set $J \subseteq \mathbb{Z}^d$ we have*

$$H\big((FX)_J\big) \geq H(X_J) - c(J)$$

where $c(J) \triangleq \big(\big|\partial\mathcal{M}^{2r}(J)\big| + \big|\partial\mathcal{M}^r(J)\big|\big)\overline{h}$.

Proof. By the Garden-of-Eden theorem, F is pre-injective [24, 27]. From the pre-injectivity of F it follows that for every configuration x, the pattern x_J is uniquely determined from the patterns $x_{\partial\mathcal{M}^{2r}(J)}$ and $(Fx)_{\mathcal{M}^r(J)}$. Indeed, suppose that x and x' are two configurations such that $x'_{\partial\mathcal{M}^{2r}(J)} = x_{\partial\mathcal{M}^{2r}(J)}$ and $(Fx')_{\mathcal{M}^r(J)} = (Fx)_{\mathcal{M}^r(J)}$. Let x'' be another configuration that agrees with x' on $\mathcal{M}^{2r}(J)$ and with x outside J. Then, x and x'' agree everywhere except possibly on J. On the other hand, $Fx = Fx''$ because

- Fx'' and Fx agree outside $\mathcal{M}^r(J)$ because x'' and x agree outside J,
- Fx'' and Fx' agree on $\mathcal{M}^r(J)$ because x'' and x' agree on $\mathcal{M}^{2r}(J)$, and Fx' and Fx agree on $\mathcal{M}^r(J)$ by assumption.

The pre-injectivity of F now implies $x'' = x$. It follows that x and x' agree on J.

Now, consider a random configuration X. Since X_J is a function of $X_{\partial\mathcal{M}^{2r}(J)}$ and $(FX)_{\mathcal{M}^r(J)}$, we have

$$\begin{aligned}
H(X_J) &\leq H\big(X_{\partial\mathcal{M}^{2r}(J)}, (FX)_{\mathcal{M}^r(J)}\big) \\
&= H\big((FX)_J\big) + H\big(X_{\partial\mathcal{M}^{2r}(J)}, (FX)_{\partial\mathcal{M}^r(J)} \mid (FX)_J\big) \\
&\leq H\big((FX)_J\big) + \big(\big|\partial\mathcal{M}^{2r}(J)\big| + \big|\partial\mathcal{M}^r(J)\big|\big)\overline{h} ,
\end{aligned}$$

proving the claim. ◀

2.3   Evolution of entropy

Combining the above lemmas we obtain the following proposition.

▶ **Proposition 6** (Evolution of entropy). *Let Φ be a PCA describing the perturbation of a surjective CA $F : \Sigma^{\mathbb{Z}^d} \to \Sigma^{\mathbb{Z}^d}$ with positive additive noise, and let $q : \Sigma \to (0,1)$ denote the noise distribution. Let $X^0, X^1, \ldots$ be a trajectory of Φ starting from an arbitrary configuration X^0. Then, for every finite $J \subseteq \mathbb{Z}^d$ and each $t \geq 0$ we have*

$$H\big(X_J^t\big) \geq [1 - (1 - \kappa)^t]|J|\overline{h} - \tilde{c}(J)$$

where $\kappa \triangleq |\Sigma| \min_{a \in \Sigma} q(a)$ and $\tilde{c}(J) \triangleq \left(\frac{1-\kappa}{\kappa}\right)\big(\big|\partial\mathcal{M}^{2r}(J)\big| + \big|\partial\mathcal{M}^r(J)\big|\big)\overline{h}$.

Proof. According to Lemma 5, for every $s > 0$ we have

$$H\big((FX^{s-1})_J\big) \geq H(X_J^{s-1}) - c(J) \ .$$

Lemma 4 on the other hand gives

$$H(X_J^s) \geq \kappa|J|\overline{h} + (1 - \kappa)H\big((FX^{s-1})_J\big) \ .$$

Combining the two, we find that for $s > 0$,

$$H(X_J^s) \geq (1 - \kappa)H(X_J^{s-1}) + \kappa|J|\overline{h} - (1 - \kappa)c(J) \ .$$

Multiplying by $(1 - \kappa)^{t-s}$, we obtain

$$(1 - \kappa)^{t-s}H(X_J^s) \geq (1 - \kappa)^{t-s+1}H(X_J^{s-1}) + \kappa(1 - \kappa)^{t-s}|J|\overline{h} - (1 - \kappa)^{t-s+1}c(J) \ .$$

Summing over s from 1 to t, we get

$$\sum_{s=1}^t (1 - \kappa)^{t-s}H(X_J^s) \geq \sum_{s=1}^t (1 - \kappa)^{t-s+1}H(X_J^{s-1})$$
$$+ \big[1 - (1 - \kappa)^t\big]|J|\overline{h} - \overbrace{\big[1 - (1 - \kappa)^t\big]}^{<1}\frac{1 - \kappa}{\kappa}c(J)$$

which after cancellation of the common terms gives

$$H(X_J^t) \geq (1 - \kappa)^t H(X_J^0) + \big[1 - (1 - \kappa)^t\big]|J|\overline{h} - \tilde{c}(J)$$
$$\geq \big[1 - (1 - \kappa)^t\big]|J|\overline{h} - \tilde{c}(J) \ . \qquad\qquad\blacktriangleleft$$

As a corollary, we get the following proposition.

▶ **Proposition 7** (Evolution of entropy). *Let Φ be a PCA describing the perturbation of a surjective CA $F : \Sigma^{\mathbb{Z}^d} \to \Sigma^{\mathbb{Z}^d}$ with positive additive noise, and let $q : \Sigma \to (0,1)$ denote the noise distribution. There are two constants $a_0, b_0 > 0$ with the following property. If $X^0, X^1, \ldots$ is a trajectory of Φ starting from an arbitrary configuration X^0, then for every finite set $J \subseteq \mathbb{Z}^d$, we have*

$$H(X_J^t) \geq |J|\overline{h} - 2\tilde{c}(J) \qquad \text{for all } t \geq a_0 \log \frac{|J|}{\tilde{c}(J)} + b_0$$

where $\tilde{c}(J) \triangleq \left(\frac{1-\kappa}{\kappa}\right)\big(\big|\partial\mathcal{M}^{2r}(J)\big| + \big|\partial\mathcal{M}^r(J)\big|\big)\overline{h}$.

Proof. From Proposition 6, it follows that in order to have $H(X_J^t) \geq |J|\overline{h} - 2\tilde{c}(J)$, it is sufficient that $(1-\kappa)^t |J|\overline{h} \leq \tilde{c}(J)$. We have,

$$(1-\kappa)^t |J|\overline{h} \leq \tilde{c}(J) \Longleftrightarrow t\log(1-\kappa) \leq \log\tilde{c}(J) - \log|J| - \log\overline{h}$$

$$\Longleftrightarrow t \geq \frac{\log|J| - \log\tilde{c}(J) + \log\overline{h}}{-\log(1-\kappa)}$$

$$\Longleftrightarrow t \geq a_0 \log\frac{|J|}{\tilde{c}(J)} + b_0$$

where $a_0 \triangleq -1/\log(1-\kappa)$ and $b_0 \triangleq -\log\overline{h}/\log(1-\kappa)$. ◀

The latter proposition can be interpreted as follows. For $n \geq 0$, consider a hypercube $S_n \triangleq [\![0, n-1]\!]^d$ of size n^d in the lattice. Then, $\tilde{c}(S_n) = \Theta(n^{d-1})$ as $n \to \infty$. Thus, according to Proposition 7, irrespective of the distribution of the initial configuration X^0, we have $H(X_{S_n}^t) \geq n^d\overline{h} - \Theta(n^{d-1})$ (i.e., S_n lacks no more than $\Theta(n^{d-1})$ nats of entropy at time t) as soon as $t \geq \Theta(\log n)$.

2.4 A bootstrap lemma

The next step is a "bootstrap argument". The intuitive idea is as follows. The effect of noise on a hypercube $S_n \triangleq [\![0, n-1]\!]^d$ is to accumulate entropy as long as the entropy of S_n is less than its maximum capacity $|S_n|\overline{h}$. A surjective CA on the other hand keeps the entropy of S_n almost preserved except for a leakage of size $O(|\partial S_n|)$ per iteration through the boundary of S_n. The "equilibrium" is reached at time $t_n = O(\log n)$, when the rate of accumulation and the maximum rate of leakage roughly match. Now consider a much larger hypercube S_m which contains many disjoint copies of S_n. For S_m, a similar "equilibrium" is reached at time $t_m = O(\log m)$. However, the entropy leaking from the copies of S_n will not have enough time to reach and escape through the boundary of S_m before time t_m, and will hence have to accumulate inside S_m. This implies that the entropy missing from each copy of S_n at time t_m must in fact be much less than $\Theta(|\partial S_n|)$.

Let us make this argument precise. Given a random configuration X and finite set $A \subseteq \mathbb{Z}^d$, let us write $\Xi(X_A) \triangleq |A|\overline{h} - H(X_A)$ for the difference between the entropy of X_A and the maximum entropy capacity of A. Note that $\Xi(X_A) \geq 0$ with equality if and only if X_A is uniformly distributed over Σ^A.

▶ **Lemma 8** (Bootstrap lemma). *Let Φ be a PCA on $\Sigma^{\mathbb{Z}^d}$ with neighbourhood $\mathcal{N} \subseteq [\![-r, r]\!]^d$. Let $\tau, \delta : \mathbb{Z}^+ \to [0, \infty)$ be two functions satisfying the following property:*

- *for every trajectory $X^0, X^1, \ldots$ of Φ and each $n \in \mathbb{Z}^+$, we have $\Xi(X_{S_n}^t) \leq \delta(n)$ for all $t \geq \tau(n)$, irrespective of the distribution of X^0.*

Let $k, m, n, t \in \mathbb{Z}^+$ be such that $m \geq k(n + 2rt)$ and $t \geq \tau(m)$. Then, for every trajectory $X^0, X^1, \ldots$ of Φ, we have $\Xi(X_{S_n}^t) \leq \delta(m)/k^d$.

Proof. Observe that we can pack k^d disjoint copies of $\mathcal{M}^{rt}(S_n)$ in S_m. Namely, for $w \in S_k$, let $Q_w \triangleq (n + 2rt)w + [\![rt, rt + n - 1]\!]^d$. Then, the sets $\mathcal{M}^{rt}(Q_w)$ (for $w \in S_k$) are disjoint and are all included in S_m. Construct a random configuration Y by choosing the patterns $Y_{\mathcal{M}^{rt}(Q_w)}$ (for $w \in S_k$) independently according to the distribution of $X_{\mathcal{M}^{rt}(S_n)}^0$, and assigning arbitrary values to the remaining cells. Consider a trajectory $Y^0, Y^1, \ldots$ of Φ with initial configuration $Y^0 \triangleq Y$. Clearly, the patterns $Y_{Q_w}^t$ (for $w \in S_k$) are independent and have the

same distribution as $X_{S_n}^t$. Therefore, using the chain rule, we have

$$H(Y_{S_m}^t) = \sum_{w \in S_k} H(Y_{Q_w}^t) + H\left(Y_{S_m \setminus \bigcup_{w \in S_k} Q_w}^t \,\middle|\, Y_{\bigcup_{w \in S_k} Q_w}^t\right)$$

$$\leq k^d H(X_{S_n}^t) + \left|S_m \setminus \bigcup_{w \in S_k} Q_w\right| \overline{h} \,,$$

which implies

$$\Xi(Y_{S_m}^t) \geq k^d \Xi(X_{S_n}^t) \,.$$

Now, since $Y^0, Y^1, \ldots$ is a trajectory of Φ and $t \geq \tau(m)$, we have $\Xi(Y_{S_m}^t) \leq \delta(m)$. It follows that $\Xi(X_{S_n}^t) \leq \delta(m)/k^d$, as claimed. ◀

2.5   Proof of the theorem

Proof of Theorem 1. Let $X^0, X^1, \ldots$ be a trajectory of Φ. For every finite set $A \subseteq \mathbb{Z}^d$, we show that $\Xi(X_A^t) = |A|\overline{h} - H(X_A^t) \to 0$ exponentially fast as $t \to \infty$. This would imply that the distribution of X^t converges weakly to the uniform Bernoulli measure λ. We then translate the bound on $\Xi(X_A^t)$ to a bound on total variation distance.

Let $r \in \mathbb{N}$ be such that the neighbourhood of F (and hence also of Φ) is included in $[\![-r, r]\!]^d$. Let $\tilde{c}(J)$ and $a_0, b_0 > 0$ be as in Proposition 7, and note that $\tilde{c}(S_n) = \Theta(n^{d-1})$ as $n \to \infty$. Choose constants $a_1, b_1, c_1 > 0$ such that $\tau(n) \triangleq a_1 \log n + b_1 \geq a_0 \log \frac{|S_n|}{\tilde{c}(S_n)} + b_0$ and $\delta(n) \triangleq c_1 n^{d-1} \geq 2\tilde{c}(S_n)$ for every $n \in \mathbb{Z}^+$. Then, by Proposition 7, for each $n \in \mathbb{Z}^+$ we have $\Xi(X_{S_n}^t) \leq \delta(n)$ whenever $t \geq \tau(n)$, hence the hypothesis of Lemma 8 is satisfied.

Suppose that $A \subseteq \mathbb{Z}^d$ is a finite set of cells with diameter n. This means that $A \subseteq u + S_n$ for some $u \in \mathbb{Z}^d$. For $t \geq 0$, define $m_t \triangleq k_t(n + 2rt)$, where $k_t \in \mathbb{Z}^+$ is to be determined. Then, according to Lemma 8,

$$\Xi(X_A^t) \leq \frac{\delta(m_t)}{k_t^d} = \frac{c_1 k_t^{d-1}(n + 2rt)^{d-1}}{k_t^d} = c_1 k_t^{-1}(n + 2rt)^{d-1} \,,$$

provided that

$$t \geq \tau(m_t) = a_1 \log k_t + a_1 \log(n + 2rt) + b_1 \,. \tag{3}$$

Now, pick β_1 such that $0 < \beta_1 < 1/a_1$, and set $k_t \triangleq \lfloor e^{\beta_1 t} \rfloor$. Observe that with this choice, condition (3) is satisfied for all sufficiently large t. In particular, we can find constants $a, b > 0$ such that (3) holds whenever $t \geq a \log n + b$. It follows that, for a suitable constant $c_2 > 0$, $\Xi(X_A^t) \leq c_2 e^{-\beta_1 t}(n + 2rt)^{d-1}$ for all $t \geq a \log n + b$. (Here, we need a new constant c_2 instead of c_1 in order to compensate for replacing $\lfloor e^{\beta_1 t} \rfloor$ with $e^{\beta_1 t}$.) In particular, $\Xi(X_A^t) \to 0$ exponentially fast as $t \to \infty$.

Next, let μ^t denote the distribution of X^t, and denote by μ_A^t and λ_A the marginals of μ^t and λ on A. Observe that $\Xi(X_A^t)$ is the same as $D(\mu_A^t \| \lambda_A)$, the Kullback–Leibler divergence of μ_A^t relative to λ_A. According to Pinsker's inequality [8, Lemma 11.6.1], we have

$$\|\mu_A^t - \lambda_A\|_{\text{TV}} \leq \sqrt{\frac{1}{2} D(\mu_A^t \| \lambda_A)} = \sqrt{\frac{1}{2} \Xi(X_A^t)} \,.$$

We conclude that

$$\|\mu_A^t - \lambda_A\|_{\text{TV}} \leq \sqrt{c_2/2} \, e^{-(\beta_1/2)t}(n + 2rt)^{(d-1)/2}$$

for every finite $A \subseteq \mathbb{Z}^d$ with diameter n and all $t \geq a \log n + b$. The claim follows by choosing $\beta > 0$ slightly smaller than $\beta_1/2$ and $\alpha > 0$ sufficiently large. ◀

3 Rapid information loss

3.1 Reversible CA on an infinite lattice

Theorem 1 shows that for a surjective CA subject to positive additive noise, mixing occurs quite fast, in the sense that it takes only $O(\log n)$ steps before the marginal on each hypercube of size n^d is within ε-distance from its stationary value.

To make this precise, let us use the notation

$$\|\mu - \nu\|_A \triangleq \|\mu_A - \nu_A\|_{\mathrm{TV}} = \frac{1}{2} \sum_{u \in \Sigma^A} |\mu([u]) - \nu([u])|$$

for the total variation distance between the marginals of two measures μ and ν on $A \Subset \mathbb{Z}^d$. Let

$$d_A(t) \triangleq \sup_{x \in \Sigma^{\mathbb{Z}^d}} \|\Phi^t(x, \cdot) - \lambda\|_A$$

be the maximum distance from stationarity of the marginal on A at time t. Note that $d_A(t)$ is non-increasing with t. Given a finite set $A \subseteq \mathbb{Z}^d$ and $\varepsilon > 0$, we let

$$t_{\mathsf{mix}}(A, \varepsilon) \triangleq \inf\{t : d_A(t) \leq \varepsilon\} .$$

We call $t_{\mathsf{mix}}(A, \varepsilon)$ the *mixing time* of set A at accuracy level ε (cf. [19]).

As before, we let $S_n \triangleq [\![0, n-1]\!]^d$ be a hypercube of size n^d in the lattice.

▶ **Corollary 9** (Mixing time of surjective CA with additive noise). *Let Φ be a PCA describing the perturbation of a surjective CA $F : \Sigma^{\mathbb{Z}^d} \to \Sigma^{\mathbb{Z}^d}$ with positive additive noise. For every $\varepsilon > 0$, we have $t_{\mathsf{mix}}(S_n, \varepsilon) = O(\log n)$ as $n \to \infty$.*

Proof. According to Theorem 1,

$$d_{S_n}(t) \leq \alpha \mathrm{e}^{-\beta t} n^{(d-1)/2}$$

for every $n \in \mathbb{Z}^+$ and $t \geq a \log n + b$. Therefore, $d_{S_n}(t) \leq \varepsilon$ as soon as

$$t \geq \max\left\{a \log n + b, \frac{d-1}{2\beta} \log n + \frac{1}{\beta}(\log \alpha - \log \varepsilon)\right\} ,$$

which means $t_{\mathsf{mix}}(S_n, \varepsilon) = O(\log n)$. ◀

In other words, for every $n > 0$ and any accuracy $\varepsilon > 0$, the distribution of the pattern on S_n becomes ε-indistinguishable from the uniform distribution after $O(\log n)$ time steps.

3.2 Reversible parallel computers with finite space

In the proof of Theorem 1, the bootstrap lemma was needed to handle the potential diffusion of entropy on the infinite lattice. For a reversible parallel computer with finite space, a simpler entropy argument can be used to show that, in the presence of positive additive noise, the state of the system becomes indistinguishable from pure noise in logarithmic number of steps. This is a reformulation of Theorem 2 of Aharonov et al. [1] in a slightly more general setup.

We consider a model of parallel reversible computation with finite space in which every piece of data is (reversibly) processed at each time step, and is hence exposed to noise. More specifically, let A be a finite set and Σ a finite alphabet. Let $\mathscr{F}$ be a family of bijective maps

$F : \Sigma^A \to \Sigma^A$. Consider a computational process whose data is an element of Σ^A, and in which, at every step, an arbitrary map from $\mathscr{F}$ is applied to the data, and the result is then subjected to positive additive noise. For instance, a reversible logic circuit with t layers, n nodes per layer, and wires between consecutive layers only, in which every node is subject to positive noise fits in this setting.

The above process can be described by a time-inhomogeneous finite-state Markov chain which mixes rapidly. To make this precise, let us review some terminology and notation regarding finite-state Markov chains [19]. Let Φ be an ergodic (possibly time-inhomogeneous) Markov chain with finite state space $\mathcal{X}$ and unique stationary distribution π. We write $\Phi^{s \to t}$ for the transition matrix from time s to time t. Let

$$d_\Phi(t) \triangleq \sup_{x \in \mathcal{X}} \|\Phi^{0 \to t}(x, \cdot) - \pi\|_{\mathrm{TV}}$$

denote the maximum distance from stationarity of the distribution at time t. Since Φ is ergodic, $d(t) \to 0$ monotonically as $t \to \infty$. The *mixing time* at accuracy $\varepsilon > 0$ is defined as

$$t_{\mathsf{mix}}(\Phi, \varepsilon) \triangleq \inf\{t : d_\Phi(t) \leq \varepsilon\} \ .$$

▶ **Theorem 10** (Rapid mixing of reversible computer with noisy components). *Let A be a finite set and Σ a finite alphabet. Let $\mathscr{F}$ be a family of bijective maps $F : \Sigma^A \to \Sigma^A$. Let Φ be a time-inhomogeneous Markov chain on Σ^A in which, at each time step, first an (arbitrary) element of $\mathscr{F}$ is applied to the current state, and then the state is subjected to positive additive noise with fixed noise distribution. Then, Φ is ergodic with the uniform distribution on Σ^A as the stationary distribution. Furthermore, for every $\varepsilon > 0$, $t_{\mathsf{mix}}(\Phi, \varepsilon) = O(\log|A|)$ as $|A| \to \infty$.*

Proof. Let X^t denote the state of the Markov chain at time t. Let λ denote the uniform distribution on Σ^A. Let $F_t \in \mathscr{F}$ be the map applied at time t, and let q denote the noise distribution. Thus, X^t is obtained from X^{t-1} by applying additive noise with distribution q to $F_t(X^{t-1})$. Let $\kappa \triangleq |\Sigma| \min_{a \in \Sigma} q(a)$.

The bijective maps F_t do not change the entropy, hence according to Lemma 4,

$$H(X^t) \geq |A| \kappa \overline{h} + (1 - \kappa) H\big(F_t(X^{t-1})\big)$$
$$= |A| \kappa \overline{h} + (1 - \kappa) H(X^{t-1})$$

for each $t \geq 1$. Therefore, setting $\Xi(X^t) \triangleq |A| \overline{h} - H(X^t)$, we have

$$\Xi(X^t) \leq (1 - \kappa)^t \Xi(X^0) \leq (1 - \kappa)^t |A| \overline{h}$$

for every $t \geq 0$. This shows that the Markov chain is ergodic with λ as the unique stationary distribution.

Now, recall that $\Xi(X^t)$ is the same as the Kullback–Leibler divergence $D(\mu^t \,\|\, \lambda)$, where μ^t is the distribution of the Markov chain at time t. Therefore, Pinsker's inequality [8, Lemma 11.6.1] gives

$$\|\mu^t - \lambda\|_{\mathrm{TV}} \leq \sqrt{\frac{1}{2} D(\mu^t \,\|\, \lambda)} = \sqrt{\frac{1}{2} \Xi(X_A^t)} \leq \sqrt{\overline{h}/2} |A|^{1/2} (1 - \kappa)^{t/2} \ .$$

Since X^0 is arbitrary, we get

$$d_\Phi(t) \leq \sqrt{\overline{h}/2} |A|^{1/2} (1 - \kappa)^{t/2} \ ,$$

from which it follows that $t_{\mathsf{mix}}(\Phi, \varepsilon) = O(\log|A|)$ as $|A| \to \infty$.   ◀

4 Discussion

Is there anything we can do with a noisy reversible CA?

Since an input of size n is lost in $O(\log n)$ steps, no cell on the lattice will have time to "sense" all the input. Thus, it seems unlikely that one can do any meaningful computation with a noisy reversible CA in a scalable fashion.

Theorem 1 of Aharoni et al. [1] states that, with exponential redundancy, any logic circuit can be simulated by a reliable reversible circuit. In order to simulate a computation with d steps (i.e., a circuit of depth d), every input bit is provided in 3^d separate copies. The i-th step of the computation is performed 3^{d-i+1} times in parallel on separate copies, and majority vote on groups of 3 is used to pass the results of the i-th step to the next step. Such an error-correcting mechanism cannot be implemented in the setting of cellular automata on $\mathbb{Z}^d$, because the dependence graph of each cell grows at most polynomially.

Reversible serial computers

Theorem 10 describes the loss of information in a parallel model of a reversible computer in which every bit of data is updated at every step of the computation and is therefore subjected to noise. In a serial computer (such as a Turing machine) on the other hand, only a bounded portion of data is operated on at each step, and only that portion is affected by significant noise. We can there wonder about the possibility of having a reversible serial computer that is capable of performing a significantly long computation reliably in the presence of noise.

How much irreversibility is needed to do reliable computation with a CA?

Theorem 1 suggests that, in order to perform reliable computation (with a CA-like computer), some degree of irreversibility is unavoidable. Can we quantify the degree of irreversibility needed to perform reliable computation at a given noise level?

References

1. D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan. Limitations of noisy reversible computation, 1996. `arXiv:quant-ph/9611028`.

2. C. H. Bennett. Logical reversibility of computation. *IBM Journal of Research and Development*, 17(6):525–532, 1973. `doi:10.1147/rd.176.0525`.

3. C. H. Bennett. The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982. `doi:10.1007/BF02084158`.

4. C. H. Bennett. Time/space trade-offs for reversible computation. *SIAM Journal on Computing*, 18(4):766–776, 1989. `doi:10.1137/0218053`.

5. P. Berman and J. Simon. Investigations of fault-tolerant networks of computers. In *STOC '88: Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 66–77. ACM, 1988. `doi:10.1145/62212.62219`.

6. M. Bramson and C. Neuhauser. Survival of one-dimensional cellular automata under random perturbations. *The Annals of Probability*, 22(1):244–263, 1994. `doi:10.1214/aop/1176988858`.

7. B. Chopard and M. Droz. *Cellular Automata Modeling of Physical Systems*. Cambridge University Press, 1998. `doi:10.1017/CBO9780511549755`.

8. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, second edition, 2006. `doi:10.1002/047174882X`.

9. E. Fredkin and T. Toffoli. Conservative logic. *International Journal of Theoretical Physics*, 21(3/4):219–253, 1982. `doi:10.1007/BF01857727`.

10 P. Gács. Reliable computation with cellular automata. *Journal of Computer and System Sciences*, 32(1):15–78, 1986. `doi:10.1016/0022-0000(86)90002-4`.

11 P. Gács. Reliable cellular automata with self-organization. *Journal of Statistical Physics*, 103(1–2):45–267, 2001. `doi:10.1023/A:1004823720305`.

12 P. Gács. Reliable computation, 2005. URL: `https://www.cs.bu.edu/fac/gacs/`.

13 P. Gács and J. Reif. A simple three-dimensional real-time reliable cellular array. *Journal of Computer and System Sciences*, 36(2):125–147, 1988. `doi:10.1016/0022-0000(88)90024-4`.

14 P. Gács and I. Törmä. Stable multi-level monotonic eroders, 2018. `arXiv:1809.09503`.

15 G. Hedlund. Endomorphisms and automorphisms of shift dynamical systems. *Mathematical Systems Theory*, 3:320–375, 1969. `doi:10.1007/BF01691062`.

16 J. Kari. Reversibility and surjectivity problems of cellular automata. *Journal of Computer and System Sciences*, 48(1):149–182, 1994. `doi:10.1016/S0022-0000(05)80025-X`.

17 J. Kari. Reversible cellular automata: From fundamental classical results to recent developments. *New Generation Computing*, 36:145–172, 2018. `doi:10.1007/s00354-018-0034-6`.

18 R. Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961. `doi:10.1147/rd.53.0183`.

19 D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2009.

20 I. Marcovici, M. Sablik, and S. Taati. Ergodicity of some classes of cellular automata subject to noise. *Electronic Journal of Probability*, 24(41), 2019. `doi:10.1214/19-EJP297`.

21 I. Marcovici and S. Taati, In preparation.

22 N. Margolus. Physics-like models of computation. *Physica D: Nonlinear Phenomena*, 10(1–2):81–95, 1984. `doi:10.1016/0167-2789(84)90252-5`.

23 M. McCann and N. Pippenger. Fault tolerance in cellular automata at high fault rates. *Journal of Computer and System Sciences*, 74(5):910–918, 2008. `doi:10.1016/j.jcss.2008.02.003`.

24 E. F. Moore. Machine models of self-reproduction. In *Mathematical Problems in the Biological Sciences*, volume 14 of *Proceedings of Symposia in Applied Mathematics*, pages 17–33. American Mathematical Society, 1962.

25 K. Morita. Reversible computing and cellular automata—a survey. *Theoretical Computer Science*, 395(1):101–131, 2008. `doi:10.1016/j.tcs.2008.01.041`.

26 K. Morita and M. Harao. Computation universality of one-dimensional reversible (injective) cellular automata. *The Transactions of the IEICE*, E72(6):758–762, 1989.

27 J. Myhill. The converse of Moore's Garden-of-Eden theorem. *Proceedings of the American Mathematical Society*, 14(4):685–686, 1963. `doi:10.2307/2034301`.

28 V. Salo and I. Törmä. A one-dimensional physically universal cellular automaton. In J. Kari, F. Manea, and I. Petre, editors, *CiE 2017: Unveiling Dynamics and Complexity*, volume 10307 of *LNCS*, pages 375–386. Springer, 2017. `doi:10.1007/978-3-319-58741-7_35`.

29 L. Schaeffer. A physically universal cellular automaton. In *ITCS '15: Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pages 237–246. ACM, 2015. `doi:10.1145/2688073.2688107`.

30 T. Toffoli. Computation and construction universality of reversible cellular automata. *Journal of Computer and System Sciences*, 15(2):213–231, 1977. `doi:10.1016/S0022-0000(77)80007-X`.

31 T. Toffoli and N. Margolus. *Cellular Automata Machines*. MIT Press, 1987.

32 T. Toffoli and N. Margolus. Invertible cellular automata: A review. *Physica D: Nonlinear Phenomena*, 45(1–3):229–253, 1990. `doi:10.1016/0167-2789(90)90185-R`.

33 A. Toom. Stable and attractive trajectories in multicomponent systems. In R. L. Dobrushin and Ya. G. Sinai, editors, *Multicomponent Random Systems*, pages 549–575. Marcel Dekker, 1980.

34 A. L. Toom, N. B. Vasilyev, O. N. Stavskaya, L. G. Mityushin, G. L. Kuryumov, and S. A. Pirogov. Discrete local Markov systems. In R. L. Dobrushin, V. I. Kryukov, and A. L. Toom, editors, *Stochastic cellular systems: ergodicity, memory, morphogenesis*. Manchester University Press, 1990.

35 G. Y. Vichniac. Simulating physics with cellular automata. *Physica D: Nonlinear Phenomena*, 10(1–2):96–116, 1984. `doi:10.1016/0167-2789(84)90253-7`.

36 J. von Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 43–98. Princeton University Press, 1956. `doi:10.1515/9781400882618`.