Life Won't Wait!
(On the Slowdown of Asynchronous Automata Networks)

Thomas P. Hayes* Michael Janes* Christopher Moore *

Abstract

Nakamura [1974], and later independently, Toffoli [1978] and Nehaniv [2002] proved that any synchronous cellular automaton can be simulated by an asynchronous cellular automaton. Their constructions effectively discard many of the updates in order to create a limited artificial synchrony between the cells. We consider the overhead cost of this procedure, assuming that cells update in random order. In particular, we prove explicit upper bounds on this overhead that depend only on the maximum degree of the network.

1 Introduction

When *you* think of cellular automata, what example comes to mind first? For many, the answer would be John Conway's Game of Life, in which the cells of a square lattice change states from alive to dead and back according to simple rules, which nevertheless lead to surprisingly complex behaviors over time. This CA is *synchronous* in the sense that, depending on our exact implementation, the cells must either all update simultaneously, or at least in a more or less deterministic order; therefore, some sort of global clock is required to ensure that none of the cells race ahead or lag behind, in terms of how many times they have updated.

Since one of the nice features of cellular automata is the extent to which all computation is done locally, we would prefer to avoid the need for globally coordinated update

*University of New Mexico, Department of Computer Science. email: `hayes,mjanes,moore@cs.unm.edu`

timing. Indeed, we would like our computation to be almost totally independent of the order in which cells are updated.

Fortunately, this is always possible! As first shown in 1974 by Nakamura [1], and later independently rediscovered by Toffoli [3] and Nehaniv [2], there is a simple construction that allows any synchronous CA to be simulated by a completely asynchronous CA, as long as each cell eventually updates infinitely often. The elegant idea is to augment each cell's state space slightly, to maintain a mod-3 counter of the number of times it has successfully updated, and to ensure that each cell only changes state when its neighbors have all caught up to it or are one step ahead.

This construction ensures that each cell will go through the same sequence of updates as the original synchronous CA, although each cell does this on its own schedule. For two cells that are far apart in the network, the coupling between their update schedules may be very loose indeed.

In this paper, we examine the overhead incurred by the above construction, in terms of the fraction of attempted updates which must be wasted because the counters of neighboring cells have fallen behind. For simplicity, we will assume that, at any given time, the next cell to update is chosen uniformly at random. More specifically, we will think of each cell as updating at each ring of its own independent Poisson clock, that rings with an average rate of 1.

Our main result is the following, which partially answers open problems (1) and (2) from [2]. For a fixed graph G, define the *asymptotic slowdown*, S, as the asymptotic ratio of the elapsed time to the number of successful updates of a given cell.

Theorem 1. *Let G be a graph with maximum degree Δ. Then, assuming i.i.d. Poisson updates, the asymptotic slowdown, S, for asynchronous simulation of any synchronous CA network on G, satisfies*

$$S \exp(1 - S) \geq \frac{1}{1 + \Delta},$$

or, equivalently,

$$S \leq W_{-1}\left(\frac{-1}{\mathrm{e}(1 + \Delta)}\right),$$

where W_{-1} denotes the lower branch of the Lambert W function.

Note that in the worst case when G is the complete graph on n vertices, the slowdown equals the expected time for the standard Coupon Collector problem, i.e., $\ln(n)$. The bound from Theorem 1 is approximately $\ln(n) + \ln(\ln(n))$ in this case.

Lattice	Δ	$S \leq$
1-D (or path or cycle)	2	3.289
2-D square lattice	4	3.994
3-D cubic lattice	6	4.436
2-D square lattice with diagonals (Game of Life)	8	4.757

Figure 1: Bounds to the slowdown of some lattices of interest

2 Preliminaries

We will assume throughout the paper, that the automata network is given by the graph G on n vertices with maximum degree Δ. We will be concerned with the following random process, which models the number of successful updates at each vertex.

Each vertex v has its own i.i.d. Poisson clock, with parameter 1. At time t, we denote by $c(v,t)$ the count of successful updates at v. Initially, $c(v,0) = 0$. Subsequently, each time v's Poisson clock rings at time t, we set $c(v,t) = c(v,t^-) + 1$ if, for all $w \in N(v)$, $c(v,t^-) \leq c(w,t)$, and otherwise, $c(v,t) = c(v,t^-)$. Here, the shorthand t^- refers to any time before t but after the preceding ring of v's Poisson clock.

By the *asymptotic efficiency*, we mean the limit

$$\frac{1}{S} := \lim_{t \to \infty} \frac{c(v,t)}{t},$$

of the number of successful updates to the expected number of total updates. We refer to the inverse efficiency, S, as the *asymptotic slowdown*.

Our assumption of continuous time is really just a convenience. By standard techniques, all our results immediately translate into analogous results in discrete time, assuming each vertex is sampled independently uniformly at random.

3 A Comparison Argument for the 1-D Case

Note that, although the successful update counts $c_t(v) = c(v,t)$ tend to infinity, if we mod out by additive constants, there are only finitely many possibilities, for a fixed graph, G. Indeed, in this case, the function $c_t - \min c_t$ is a finite Markov chain with respect to the continuous time parameter t. It follows from general principles that there is a unique stationary distribution that this chain approaches in the limit as $t \to \infty$.

The asymptotic slowdown is therefore a function of this stationary distribution. Indeed, we make the following observation.

Observation 2. *If, for a vertex v, the stationary probability is p that, for all $w \in N(v)$, $c(v) \leq c(w)$, then $S = 1/p$.*

Note that this probability cannot depend on the vertex v, since no two vertices can ever have their counts differ by more than n (we assume G is connected).

Unfortunately, the stationary distribution for c_t (mod translation) is not so easy to describe. In this section, we consider the special case when G is a $2n$-cycle, for which a slight variant of the Markov chain c_t converges to a uniform distribution. Thereby, through an easy comparison argument, we can prove a clean upper bound on the slowdown for this graph.

Definition 3. Define the variant count functions $\widetilde{c}(v,t)$ by the same update rule as $c(v,t)$, but with the initial condition

$$\widetilde{c}(v,0) = \begin{cases} 0 & \text{if } v \text{ is even} \\ 1/2 & \text{if } v \text{ is odd} \end{cases}$$

Now, $\widetilde{c}$ differs from c in that, for adjacent vertices v, w, we can never have a tie, $\widetilde{c}_t(w) = \widetilde{c}_t(v)$. Surprisingly, this leads to the following consequence. For convenience, we will assume that $\widetilde{c}$ is only defined up to translation by additive constants.

Lemma 4. *The stationary distribution for $2\widetilde{c}$ is uniform over all functions $2\widetilde{c} : [2n] \to \mathbb{Z}$ that satisfy $|\widetilde{c}(i+1) - \widetilde{c}(i)| = 1$, for all i, where $i+1$ is computed mod $2n$.*

Proof. It is clear by induction that every $\widetilde{c}$ is of the form described. A greedy construction shows that every function of the form described is achievable starting from $\widetilde{c}_0$. All that remains is to verify that the uniform distribution is stationary.

To see this, fix a function $\widetilde{c}_t$. Observe that the following are all equal:

- the number of distinct successor states from $\widetilde{c}_t$
- the number of distinct predecessor states to $\widetilde{c}_t$
- 1 plus the number of local maxima of $\widetilde{c}_t$
- 1 plus the number of local minima of $\widetilde{c}_t$.

This is clear, because each successful update converts a (preceding) local minimum into a (subsequent) local maximum. And the number of maxima equals the number of minima because ties are not allowed, so minima and maxima must alternate. Thus, under the uniform distribution, the detailed balance condition is satisfied. $\square$

Lemma 5. *The $\widetilde{c}_t$ process has asymptotic slowdown ≤ 4.*

Proof. Consider a sequence of $2n$ coin flips used to generate $\widetilde{c}_t$ (heads correspond to a change of $+1/2$ from the previous value, tails to a change of $-1/2$). Since a local minimum corresponds to a tails followed by a heads, we expect a $1/4$ fraction of these. This is in fact the asymptotically correct value for the slowdown as $n \to \infty$, by an application of the central limit theorem. However, for small n, the expected number of local minima is larger, because the cycle topology creates some negative dependence between successive differences. □

Lemma 6. *The c_t process on the cycle has asymptotic slowdown ≤ 4.*

Proof. For any fixed order of vertex updates, the process $\widetilde{c}_t - \widetilde{c}_0$ will always be $\leq c_t$. Whenever they are equal, the $\widetilde{c}_t$ process declines to update the odd cells unless they are strict local minima, whereas the c_t process will allow updates to non-strict local minima. Since successful updates always increase the counts by 1, there is no way for $\widetilde{c}$ to "come from behind" to pass up c. The result follows from Lemma 5 and the definition of slowdown. □

Before going on, we observe that the uniformity of the stationary distribution of $\widetilde{c}$ allows us to prove all sorts of other nice properties of it, such as that, with high probability, $\max_v \widetilde{c}(v,t) - \min_w \widetilde{c}(w,t) = O(\sqrt{n})$. Unfortunately, it seems that the actual distribution of $c(v,t)$ is quite far from uniform. Furthermore, we do not know how to generalize the techniques of this section to higher dimensions.

4 Proof of Theorem 1

To prove a bound on the slowdown for general graphs, we will consider the ways a missing update at some other vertex could interfere with the counter at vertex v reaching a certain value. Fortunately, the locality of updates in a CA network means that the count at each cell can only be influenced by events taking place within a certain "light cone" in space-time (Figure 2). In our context, space is measured through the shortest-path distance in G, while local time is measured in rings of the Poisson clock at a node.

Proof of Theorem 1. Construct a "space-time" graph $G' = (V', E')$, where $V' = V \times \{0, 1, 2, 3, \dots\}$, and E' consists of all directed edges $((v,t),(w,t+1))$, where either $v = w$ or $\{v,w\} \in E$. See Figure 3.

In this graph, the node (v,t) corresponds to the event that the count $c(v,t')$ is at least t. The edges into (v,t) correspond to the requirements that v and all of its neighbors must previously have reached at least $t-1$.

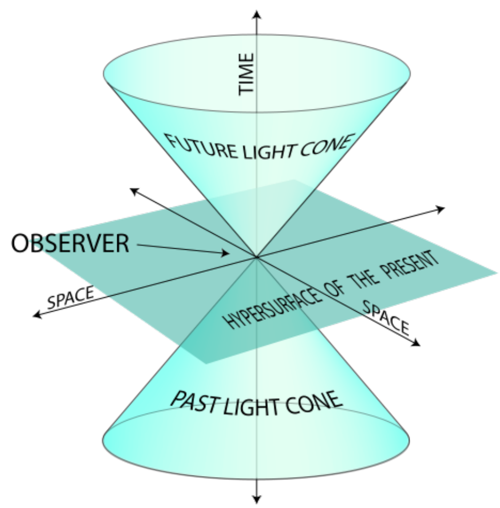
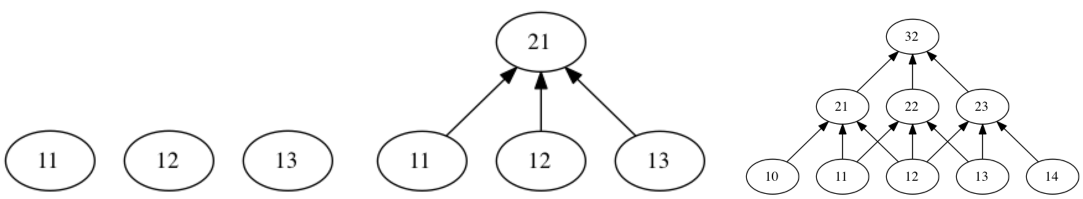


Figure 2: A Light Cone.
[Source: Wikimedia Commons Image:World_line.png]



(a) First layer of G' (b) First two layers of G' (c) First three layers of G'

Figure 3: A "past" light cone when G is a path or cycle.

Say that node (v, t) in G' is *occupied* at time t' if $c(v, t') \geq t$. Then, starting from $c_0 = 0$, each time the Poisson clock at a node u rings, we find the maximum t such that all in-neighbors of (u, t) are occupied, and make (u, t) occupied, if it was not already.

The time it takes for a node $v \in V$ to achieve ℓ updates equals to the time it takes for node (v, ℓ) to be occupied. By an easy induction, this in turn equals the sum of times between consecutive Poisson clock rings along the slowest path from the base layer to v_ℓ.

Now, for any fixed path of length ℓ, the probability that t time units do not suffice for the ℓ Poisson clocks to ring in the correct order, equals

$$\mathrm{e}^{-t} \sum_{j<\ell} \frac{t^j}{j!}$$

Thus, by a union bound over the $(1+\Delta)^\ell$ paths into (v,ℓ), the probability that (v,ℓ) is unoccupied at time t is at most

$$(1+\Delta)^\ell \mathrm{e}^{-t} \sum_{j\le\ell} \frac{t^j}{j!}$$

Setting $t = S\ell$ and approximating the sum by its dominant term, corresponding to $j=\ell$, we obtain

$$\frac{(1+\Delta)^\ell}{\mathrm{e}^{S\ell}} \frac{(S\ell)^\ell}{\ell!} \le \left(\frac{\mathrm{e}S(1+\Delta)}{\mathrm{e}^S}\right)^\ell$$

which tends asymptotically to zero exactly below the threshold from the Theorem statement. Since the convergence is exponential in ℓ, the approximation of the ℓ-term sum by its maximum cannot affect the answer. Therefore, the proof is complete. $\square$

It is worth noting that despite our inability to describe the stationary distribution, and the huge union bound in its proof, Theorem 1 still gives a noticeably better result in the 1-D case than the comparison argument from Section 3 (namely 3.289 instead of 4).

5 Tugs-of-War on Distributions

Brain-teasers of the style "What is the value of $\sqrt{1+\sqrt{1+\sqrt{1+\ldots}}}$" are commonplace. In this case, the answer is well-defined, namely the golden ratio, $1.618\ldots$. We can view this as a "tug-of-war" between alternating operations: "add one" and "square root." The first tends to drive us upward towards its fixed point at $+\infty$, while the second tends to drive us down towards its fixed point at 1. In general, the question of which of two operations will "win" in the long run can be quite interesting.

Consider the following pair of operations A_2, M_2, on distributions. For a distribution X, to sample from A_2X, take two independent samples from X, and average them. To sample from M_2X, take two independent samples from X, and return the larger one. More generally, we can consider operations A_k, M_k that sample k times independently and then average or max respectively. The M operation increases its argument. The A operation concentrates its argument, while keeping its mean fixed. Both operations have the same infinity of fixed points, namely all distributions supported on a single real value. What can we say in general about which operation wins in a tug-of-war?

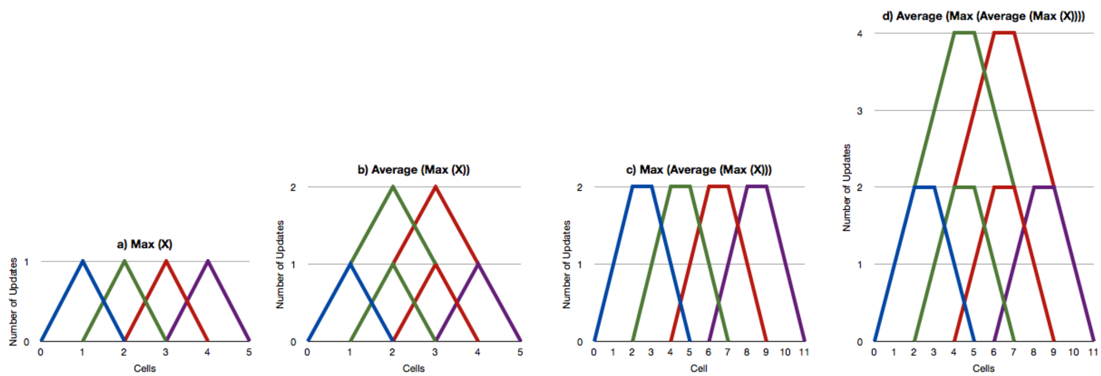


Figure 4: A visual representation of repeatedly maxing and averaging on a space-time graph. Here the underlying topology is one-dimensional, and the alternating operations are M_4 and A_2.

This question is motivated by an alternative approach to Theorem 1 that proved to be less general. Basically, the idea is that in our time-space graph G', the delay to occupy vertex $(v, 2)$ along a particular path equals the average of two independent Poisson clocks. The overall delay to occupy vertex $(v, 2)$ is the maximum of these delays, taken over the 3 different children of $(v, 2)$. More generally, the recursive structure of the light cone suggests a long alternating sequence of averaging and max operations. However, there are two caveats: firstly, the max operations are not taken over independent samples (this turns out to not really matter in our analysis), and secondly, in order to get the operations to take the same number of arguments at each level, we take advantage of a high degree of self-similarity for lattices, that is not present in more general graphs. See Figure 4 Although the application to slowdowns yields inferior results to Theorem 1, we feel that this approach may be of some interest on its own.

If p is the probability that $X \leq A$, where X is a random variable and A is constant, then clearly the probability that the sum of ℓ independent copies of X is at most ℓA is at least p^ℓ, since this is the likelihood that each X_i are at most A. The next result shows that, for exponential moment bounds, something akin to a converse holds.

Lemma 7. *If p is the best possible exponential moment bound for $\mathbf{Pr}\left(X \leq A\right)$ then p^ℓ is the best possible exponential moment bound for $\mathbf{Pr}\left(\sum^\ell X \leq \ell A\right)$.*

Proof. By Markov's inequality we have the following exponential moment bound, for all

$t > 0$,

$$\mathbf{Pr}\left(X \geq A\right) = \mathbf{Pr}\left(e^{tX} \geq e^{tA}\right) \leq \frac{\mathbf{E}\left(e^{tX}\right)}{e^{tA}}$$

A second application of Markov's inequality gives the following exponential bound

$$\begin{aligned}\mathbf{Pr}\left(\sum^{\ell} X \geq \ell A\right) &= \mathbf{Pr}\left(e^{t' \sum^{\ell} X} \geq e^{t'\ell A}\right) \\ &\leq \frac{\mathbf{E}\left(e^{t' \sum^{\ell} X}\right)}{e^{t'\ell A}} = \frac{\mathbf{E}\left(\Pi^{\ell} e^{t'X}\right)}{e^{t'\ell A}} = \left(\frac{\mathbf{E}\left(e^{t'X}\right)}{e^{t'A}}\right)^{\ell}\end{aligned}$$

Setting $t' = t$, where t was chosen optimally, we conclude

$$\mathbf{Pr}\left(\sum^{\ell} X \geq \ell A\right) \leq p^{\ell}$$

□

Now we are ready to prove that, as long as our starting distribution doesn't decay too slowly, each averaging operator beats every max operator.

Lemma 8. *Suppose X has a finite exponential moment, and let $k, \ell \geq 2$ be fixed integers. Let A_k denote the "average k i.i.d. copies" operator, and M_ℓ denote any operator that takes the max of ℓ copies, arbitrarily correlated. Then the alternating sequence $(A_k M_\ell)^i X$ converges to a finite real number. Indeed, let $p = (1/\ell)^{1-1/k}$. Then if a is such that p is an exponential moment upper bound on $\mathbf{Pr}\left(X \geq a\right)$, then $(A_k M_\ell)^i X$ converges to a real number between $\mathbf{E}\left(X\right)$ and a.*

Proof. It is easy to verify that, when p is an exponential moment bound on $\mathbf{Pr}\left(X \geq a\right)$, that ℓp is an exponential moment bound on $\mathbf{Pr}\left(M_\ell X \geq a\right)$. By Lemma 7, we know that $(\ell p)^k$ is an exponential moment bound on $\mathbf{Pr}\left(A_k M_\ell X \geq a\right)$. But, by choice of p, $(\ell p)^k \leq p$. By an induction argument, we can deduce that the probabilities $\mathbf{Pr}\left((A_k M_\ell)^i X \geq a\right)$ converge exponentially to zero, and hence the distributions $(A_k M_\ell)^i X$ converge in measure to a distribution supported on reals $\leq a$. Since the operator M_ℓ increases the expectation of any non-constant distribution, it follows that $(A_k M_\ell)^i$ converges in probability to a real constant between $\mathbf{E}\left(X\right)$ and a. □

6 Conclusion

We have seen that, under a random vertex order, the overhead for implementing the asynchronous simulation of synchronous CA's is not only bounded as a function of the vertex degrees, but is a rather small constant for most networks one might seriously consider building. This can be viewed as further evidence of the practicality of such simulations, and of the advantage of planning synchronously, but implementing asynchronously.

We conjecture that there are interesting results that remain to be proven about the stationary distribution of the Markov chain c_t. For instance, it seems plausible that the "lag," $\max_{v,w} c(v,t) - c(w,t)$ is $O(\sqrt{n})$ with high probability, which would tell us that, viewed globally, the asynchronous local counters look much closer to being a distributed global clock than we currently know.

References

[1] K. Nakamura. Asynchronous cellular automata and their computational ability. *Systems, Computers, Controls* **5**:5 (1974) 58–66.

[2] C. L. Nehaniv. Asynchronous Automata Networks Can Emulate Any Synchronous Automata Network *International Journal of Algebra & Computation*, **14**:5-6 (2004), 719–739.

[3] T. Toffoli Integration of Phase-Difference Relations in Asynchronous Sequential Networks. In: G. Ausiello and C. Bohm, eds., *Automata, Languages, and Programming (Fifth Colloquium, Udine, July 1978)*, Lecture Notes in Computer Science 62, Springer Verlag, (1978) 457–463.