It from Bit: Cellular Automata, Driven Ising Model, and Emergent Computation

Chris Miles[1], Elliot Nelson[2], Kyle Reing[3], Shing H. Zhan[4], and Mark Kirstein[5]

[1]Department of Physics, University of Michigan, Ann Arbor, MI, U.S.A.
[2]Perimeter Institute, Waterloo, ON, Canada
[3]Department of Computer Science, University of Southern California, Los Angeles, CA, U.S.A.
[4]Department of Zoology, University of British Columbia, Vancouver, BC, Canada
[5]Department of Economics, Technische Universität Dresden, Dresden, Germany

Abstract

We explored emergent computation within two systems — cellular automata (CA) and a driven Ising model. We attempted to generate emergent computation with a CA using a genetic algorithm, with the aim of understanding general conditions on information processing dynamics which increase computational capacity. Specifically, we aim to search the space of CAs to identify minimal computational units, which may in turn be used to execute specific computational tasks or optimize further search for collective computational behavior. In order to select for CAs with more computational capacity, we use an objective function defined in terms of dual total correlation (DTC), an information-theoretic quantity capturing the information in higher-order correlations between cells in a given spatial (and temporal) region of the time-dependent CA grid. The genetic algorithm evolves a population of CA update rules by testing average performance over a range of initial conditions, creating a new population favoring more successful rules, and iterating over a number of generations.

In addition, we have explored stochastic thermodynamics within a driven Ising model. This initial investigation has given us insight into thermodynamic relations present that will be beneficial for exploring this system's computational capacity in future work. In this preliminary study, the work distribution produced by many realization under the influence of a time-symmetric drive is consistent with known fluctuation relations in the field of stochastic thermodynamics. In future work, we hope to explore dissipative adaptation as a potential mechanism for promoting emergent computation.

1 Introduction

We are interested in exploring emergent computation. An emergent computation is when the emergent behavior of a system, arising from the nonlinear interactions, produces a computation [For90]. The long-term goal is to try to find an online algorithm that could learn a computational task.

2 ABMs and Physical Intelligence

Our group was formed based on joint interest in the Arbortron experiments performed by Alfred Hübler [JH05]. We were interested in the notion of alternate models of intelligence (such as the Physical Intelligence demonstrated in the experiments). Initially, our goal was to try to build an Agent Based Model (ABM) which reproduced the results from Hübler's experiments, with the hope of extending these models to perform some useful computation. After many attempts to implement Physical Intelligence in an ABM, we realized our interests lied more in emergent forms of computation. This lead us to explore three separate paths — swarm intelligence, Ising Model, and cellular automata (CA).

3 Swarm Intelligence and Optimization

For a short amount of time, we explored different concepts in swarm intelligence and optimization. The main objective of these methods is to use meta-heuristics to guide "agents" around an optimization landscape, with the hope that they will find a better global optimum. Examples of swarm intelligence models include ant foraging algorithms for combinatorial optimization problems (such as the Traveling Salesman Problem), and particle swarm / bee colony for objectives with continuous optimization space. These algorithms are similar

in theme to simulated annealing, which uses heuristics to help escape from local optima. While these methods are interesting in their own right, we found that they did not capture the notion of interaction and emergence producing computation. For this reason, we changed directions again to focus on emergent computation in CA.

4 Emergent Computation in Cellular Automata by Genetic Algorithms

We studied CA as a 'medium' for emergent computation. We first generated emergent computation via a genetic algorithm but we hope to explore other algorithms. Evolving CA by genetic algorithms [CM95] has been studied and loosely mimics the natural evolution of the emergent computational systems known in the biological world such as the human brain and ant colonies.

4.1 Setup

With 9 states for a Moore neighborhood (including the central cell), a single neighborhood configuration can be represented as a 9-bit string where each bit is the state of that neighbor. There are $2^9 = 512$ possible 9-bit strings. An update rule defines a state for each possible configuration, so we can represent the update rule as a 512-bit string. This 512-bit string is our 'chromosome' for our genetic algorithm. The total number of possible update rules is $2^{512} \approx 1.34 \times 10^{154}$.

We will use an $N \times N$, 2-dimensional grid with periodic boundary conditions. The initial conditions on the array will be generated by (i) randomly selecting a probability p (with uniform distribution from 0 to 1) that a cell is on, and (ii) randomly selecting the on/off state of each of the N^2 cells independently, with probability p.

4.2 Information Theory and Computation

Any computation will require information to be stored, transmitted, and processed between different parts of the system. In a CA, this will likely necessitate the communication of information through correlations between a number of cells, rather than bits of information in the computation being carried and processed by only one or two cells at a time. For example, we expect the state of a cell at time t to depend on various correlations between the 9 cells at time $t-1$ on which its state depends.

We would like to use an information theoretic metric which captures this idea. Dual total correlation (DTC) is an information theoretic measure which quantifies information in the correlations between a large fraction of N variables (in our case, bits). In the context of the CA, a high DTC for the joint distribution over $n \times n$ sub-grid configurations indicates a large amount of synergistic information between the n^2 cells.

The DTC is a strict criterion in the sense that it places the most weight on the highest-order correlations, for example, between all cells in a 3×3 grid. We may wish to relax this criterion by preferring, for example, correlations between a minimum fraction of cells in a given region, or greater redundancy of information encoded in lower-order correlations. A less restrictive requirement for synergy might select for greater robustness to noise in the dynamics, in addition to allowing for more computational routines.

Ideally, we would like to know if there exists an optimal ratio of synergy to redundancy in the objective function which selects most effectively for computational CAs. This could be studied, for instance, by interpolating the objective function between DTC and total correlation (TC), which quantifies redundant information, and identifying resulting variation in the resulting computation.

4.3 Performance metrics

The objective function for a given CA simulation will either use measures mentioned in the previous section to capture the computational capacity of the CA or a measure that captures the error on a computational task such as the majority classification task or executing a logic gate function.

4.4 Genetic algorithm

We start with an initial population of P chromosomes, and evolve each of them on the $N \times N$ grid over T timesteps, resulting in a space-time array of N^2T bits. We repeat this process for N_{IC} randomly chosen initial conditions. We evaluate the fitness function of each chromosome by averaging the objective function defined in 4.3 over all N_{IC} runs.

(We can adjust this as needed:) A second generation of chromosomes is generated from the first in the following way: The fraction f of chromosomes which score the highest fitness function will be replicated in the new generation unchanged. The remaining $(1-f)P$ chromosomes are determined by (i) randomly selecting two members of the top fP chromosomes and randomly pairing half of one 512-bit string with half of the other

string to form a new chromosome. Point mutations are introduced randomly into all chromosomes at rate γ per bit. (For example, $\gamma = 1/512$ results in roughly one bit flip per chromosome.)

4.5 Propagation of information

Because the CA dynamics are local, any CA computation must take place in a finite subset of the grid. Locality also imposes a maximum propagation speed of $\sqrt{2}$ cells per timestep in diagonal directions, and 1 cell per timestep in the horizontal or vertical directions.[1] A computation distributed spatially over k^2 cells will therefore take at least $\sim k$ timesteps to execute, since the relevant information cannot be shared across the region in fewer timesteps.

Therefore, if we would like to select for CAs with emergent computation at some minimum scale k, we evaluate the performance of each CA by studying the information structure in cubic 3-dimensional sub-arrays with k^3 cells, spaced evenly to tile the entire array of N^2T cells.

We may also want to select for CAs with a collective computation distributed over larger scales. In this case, we will also study the information structure in larger cubic sub-arrays. In particular, for a hierarchically structured computation, the entire computation must be confined to some scale K, while sub-computations could take place at all scales down to the minimum k. If we want to select for larger computations like this, we will account for the information structure in cubic arrays of various sizes between k^3 and K^3.

Because a genuine computation would involve information being shared and processed between all cells in a given region, and would fail for a reversal of any cell state, we are interested in using a performance metric that quantifies the information shared between *all* cells in a certain region.

5 Dissipative adaptation and stochastic thermodynamics

Dissipative adaptation [Eng15, PME16, KOE17] may be a potential mechanism to generate emergent computation. Dissipative adaptation is the tendency for physical systems to develop structures that are more capable of absorbing work from their environment. By clever encoding information into the driving source of work, we may be able to use this mechanism to produce a computation. If there is time-correlated information in the driving source of work, the system may have to perform a computation to anticipate the outcome at later times in order to absorb more energy.

We will consider the driven Ising model for exploring dissipative adaptation. We will first explore thermodynamic relations in this setting to try to understand the dynamics of the driven Ising model. Then, we would like to explore extensions of the Ising model with the goal of applying this system to computational tasks.

5.1 Ising model

Consider a grid of spins on a $N \times N$ lattice. $\mathbf{x}$ is a vector of individial spins x_i representing the system configuration. The energy $E[\mathbf{x}(t), \lambda(t)]$ of the system at time t is given by

$$E[\mathbf{x}(t), \lambda(t)] = -\sum_{\langle i,j \rangle} J x_i(t) x_j(t) - \sum_i \lambda(t) x_i(t)$$

where J is a coupling constant,$\langle i, j \rangle$ means that the summation occurs over all neighboring pairs of sites i and j, and $\lambda(t)$ a time-dependent external field. When the system is in equilibrium and $\lambda(t) = \lambda_0$ is fixed in time, the probability of being in a state is given by the Boltzmann distribution:

$$P[\mathbf{x}] = \frac{\exp\left(-\frac{E[\mathbf{x}]}{k_B T}\right)}{Z}$$

where k_B is the Boltzmann constant, T is the temperature, and $Z = \sum_{\mathbf{x}} \exp\left(-\frac{E[\mathbf{x}]}{k_B T}\right)$ is the partition function.

We simulate the dynamics of the ising model using the Metropolis-Hastings algorithm. At each time step, a spin i is chosen at random. This spin is flipped with probability

$$p = \min\left(\exp\left(-\frac{(e'[-x_i] - e[x_i])}{k_B T}\right), 1\right)$$

where $e[x_i] = -J \sum_{\text{neighbors } j \text{ of } i} x_i x_j - \lambda x_i$.

[1] Consequently, any square region of the grid will fully determine the dynamics in a future "light-pyramid" and will be fully determined by the corresponding past last-pyramid.

5.2 Stochastic thermodynamics of the Ising model

Following Crooks [Cro98], we define thermodynamic quantities heat ΔQ and work ΔW. Heat generated from time $t-1$ and t is defined as

$$\Delta Q = E[\mathbf{x}(t), \lambda(t-1)] - E[\mathbf{x}(t-1), \lambda(t-1)].$$

The work done on the system is given by

$$\Delta W = E[\mathbf{x}(t), \lambda(t)] - E[\mathbf{x}(t), \lambda(t-1)].$$

The energy change is

$$\Delta E = E[\mathbf{x}(t), \lambda(t)] - E[\mathbf{x}(t-1), \lambda(t-1)] = \Delta W + \Delta Q.$$

Therefore these definitions are consistent with the 1st law. The above construction has split the single time step into two substeps to indentify work and heat production during this interval. Note that work is only done on the system by a time-varying λ. Therefore, if $\lambda(t)$ is fixed in time, the system will simply exchange energy with a temperature bath in the form of heat.

For an initial investigation, we will consider driving the system by a time-symmetric $\lambda(t)$ and assume that the system relaxes to the equilibrium state before applying work to the system ($\lambda(t) = 0$ up until this point). After the application of work, we let the system relax back to equilibrium with $\lambda = 0$. In this setting, we hope to explore fluctuation relations to understand the driven Ising model before exploring model extensions and applications to computation.

The Metropolis algorithm used for simulating the Ising model satisfies the microscopic reversibility condition with fixed λ:

$$\frac{P(\mathbf{x} \to \mathbf{x}')}{P(\mathbf{x}' \to \mathbf{x})} = e^{-\frac{E(\mathbf{x}', \lambda) - E(\mathbf{x}, \lambda)}{k_B T}}. \tag{1}$$

Given this condition, the results from Crooks (1998) predict the following fluctuation relation for a time-symmetric drive $\lambda(t)$:

$$\frac{P(\beta W = \beta w)}{P(\beta W = -\beta w)} = e^{\beta w} \tag{2}$$

where W is the total work and $\beta = \frac{1}{k_B T}$. Furthermore, Crooks [Cro99] derived the Jarzynski equality [Jar97] from a similar relation as (2). For the drive under consideration, we find a special case of the Jarzynski equality,

$$\langle e^{-\beta w} \rangle = 1 \tag{3}$$

where $\langle \cdot \rangle$ is the average over all realizations. By Jensen's inequality, we find that

$$\langle \beta w \rangle \geq 0 \tag{4}$$

which is consistent with the classical thermodynamic result $W \geq \Delta F = 0$ where ΔF is the free energy difference between initial and final states. $\Delta F = 0$ since we assume that we start and end in the same equilibrium state.

To verify these relations, we drove the system with a time-symmetric λ as shown in figure 1. We also allowed for plenty of time for the system to relax to equilibrium prior to the application of work. We ran the model for 100 runs and generated the work distribution shown in 2. Note the Gaussian shape this is consistent with the finding of Crooks [Cro99] for similar Markovian systems. Also, the mean is clearly greater than zero and is consistent with (4).

5.3 Applications to computation

Dissipative adaptation [Eng15] is a theorized mechanism present in driven non-equilibrium systems. Dissipative adaptation causes a system to self-organize into structures that are adapted at absorbing work. Given the theoretical setup presented in the previous section, we are now in a good position to explore dissipative adaptation further within the driven Ising model. In future work, we hope to explore the following question: "Is there a way to encode information in the initial state $\mathbf{x}$ or in the signal $\lambda(t)$, so that the system can process this information or rather complete a computation?"

If we can demonstrate the dissipative adaptation is present within our driven Ising model, we would like to make the work-absorbing states also the high-performing computational states. If this can be accomplished, then it may be possible for a system to self-organized into a high-performing computational system.

In view of applying this to computation, we may find that the dynamics encoded via the Hamiltonian are too simplistic. We may alter the Hamiltonian so that the system has more rich dynamics that could potentially be more useful for computation.

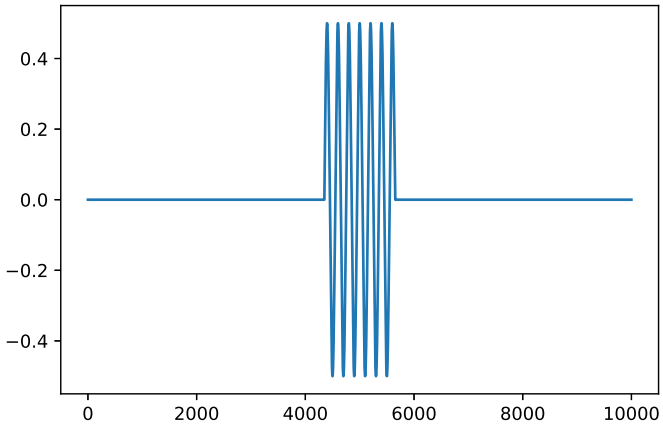

Figure 1: The driving magnetic field. The associate work distribution is shown in figure 2

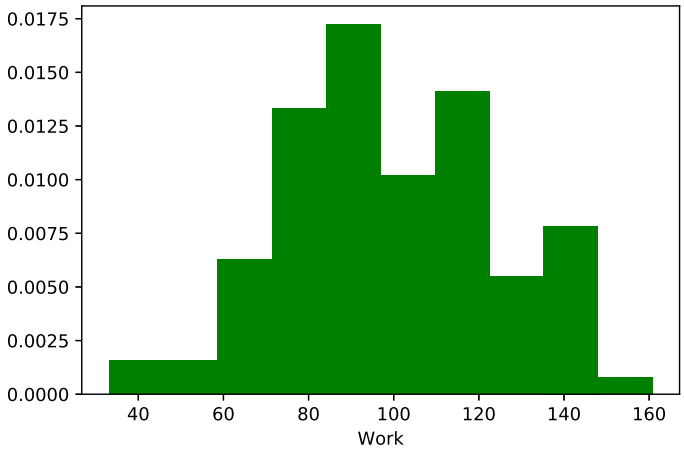

Figure 2: This is the work distribution for 100 runs. The driving magnetic field is shown in figure 1.

6 Conclusion

We implemented a genetic algorithm on a 1D and 2D cellular automata by using the DEAP python library. We have attempted to reproduce results from Crutchfield and Mitchell [CM95]. In particular, we are trying to get the CA to accomplish the majority classification task. For the 1D case with 7-cell-neighborhood (including the central cell) update functions, We ran the genetic algorithm with 200 generations, 100 individuals per generation, and mutation rate of 0.01. The resulting best update function returned all black in the final state — representing a classification of majority black in the initial condition. This update function thus generates the correct answer 50% of the time but does not actually process information. We will try to run the genetic algorithm for more generations and attempt to increase selective pressure.

We have also explored stochastic dynamics and thermodynamic relations in the context of the driven Ising model. The purpose of this preliminary study was to gain an understanding of the thermodynamics of the system before applying and modifying the system for computational means. We showed consistent results with known fluctuation relations from the field of stochastic thermodynamics. We hope to explore different driving signals that could potentially have more rich structure than the simple driving signal shown in figure 1.

7 Future Directions

We are interested in exploring the following items:

- the information structure of hierarchical computation on multiple scales and of robustness to noise or errors
- the utility of our evolved CAs for specific computational tasks
- emergent computation via dissipative adaptation to an external driving force
- ways to implement online reinforcement learning through the CA's internal dynamics.
- What information theoretic criteria correlate with the capacity for replication?

References

[CM95] James P Crutchfield and Melanie Mitchell. The evolution of emergent computation. *Proceedings of the National Academy of Sciences*, 92(23):10742–10746, 1995.

[Cro98] Gavin E Crooks. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *Journal of Statistical Physics*, 90(5/6):1481–1487, 1998.

[Cro99] Gavin Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60(3):2721–2726, 1999.

[Eng15] Jeremy L. England. Dissipative adaptation in driven self-assembly. *Nature Nanotechnology*, 10(11):919–923, 2015.

[For90] Stephanie Forrest. Emergent computation: Self-organizing, collective, and cooperative phenomena in natural and artificial computing networks. Introduction to the proceedings of the ninth annual CNLS conference. *Physica D: Nonlinear Phenomena*, 42(1-3):1–11, 1990.

[Jar97] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters*, 78(14):2690–2693, 1997.

[JH05] J. K. Jun and A. H. Hubler. Formation and structure of ramified charge transportation networks in an electromechanical system. *Proceedings of the National Academy of Sciences*, 102(3):536–540, 2005.

[KOE17] Tal Kachman, Jeremy A Owen, and Jeremy L England. Self-organized resonance during search of a diverse chemical space. pages 1–5, 2017.

[PME16] Nikolay Perunov, Robert A. Marsland, and Jeremy L. England. Statistical physics of adaptation. *Physical Review X*, 6(2):1–12, 2016.