

Cellular automata can classify data by inducing trajectory phase coexistence

Stephen Whitelam^{1,*} and Isaac Tamblyn^{2,3,†}

¹*Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA*

²*Department of Physics, University of Ottawa, Ottawa, ON, Canada K1N 6N5*

³*Vector Institute for Artificial Intelligence, Toronto, ON, Canada M5G 1M1*



(Received 15 March 2022; revised 13 April 2023; accepted 18 April 2023; published 24 July 2023)

We show that cellular automata can classify data by inducing a form of dynamical phase coexistence. We use Monte Carlo methods to search for general two-dimensional deterministic automata that classify images on the basis of activity, the number of state changes that occur in a trajectory initiated from the image. When the number of time steps of the automaton is a trainable parameter, the search scheme identifies automata that generate a population of dynamical trajectories displaying high or low activity, depending on initial conditions. Automata of this nature behave as nonlinear activation functions with an output that is effectively binary, resembling an emergent version of a spiking neuron.

DOI: [10.1103/PhysRevE.108.014126](https://doi.org/10.1103/PhysRevE.108.014126)

I. INTRODUCTION

Cellular automata are discrete dynamical systems. They are widely studied because they are simple to specify and display considerable complexity, including the ability to do universal computation [1–6]. Cellular automata have been used to do computations such as classification, either directly [7,8], or as a reservoir that enacts a nonlinear transformation prior to additional calculation [9–11], or in combination with a neural network [12]. Cellular automata do not currently perform as well at most machine-learning tasks as deep neural networks, but they continue to attract attention because discrete automata specified by integers consume less memory and power than do real-valued neural nets, a property useful for mobile devices [13].

In this paper, we search for general two-dimensional cellular automata able to classify images in MNIST [14,15], a standard data set. While image classification has been done by automata [11,12], our focus is how automata learn to do classification, and we describe a mechanism of decision making that resembles the phase transitions undergone by physical systems such as magnets [16,17] or glasses [18].

This connection is enabled by the introduction of *activity* as an order parameter for classification. We define activity as the total number of changes of state of all cells within a dynamical trajectory. Used in the study of glasses [18,19], it is a natural candidate for computations such as image classification because it is invariant to translations and is not biased toward particular spatial environments. If the number of time steps of the automaton (we call this the *depth* of the automaton) is a trainable parameter, we show that Monte Carlo search of the automaton rule table identifies rules and depths that perform classification by enacting a form of dynamical

phase coexistence. The action of these automata on the set of MNIST images produces a set of dynamical trajectories of two distinct types, possessing large or small activity. Automata of this nature resemble an emergent version of a spiking neuron [20,21], enacting a temporal calculation before committing to a state of large or small output. We show that collections or reservoirs of such automata can be used by a linear model to classify images with an accuracy comparable to that of fully connected neural networks that act on the pixels of the original image.

The automata described here perform classification by inducing a form of dynamical phase coexistence. Phase transitions and phase coexistence are seen in cellular automata in a variety of settings. Probabilistic automata such as the Ising model capture the essence of phase transitions in physical systems such as fluids or magnets [16,17]. Probabilistic and deterministic automata modeling traffic flow show the coexistence of spatial patterns [22–25], and they exhibit dynamical phase transitions as a function of density [25–27]. There is also evidence of a sharp transition in the space of rules [28,29] between different classes of deterministic automata [3]. The phase coexistence seen here occurs for deterministic automata of finite spatial extent, for sufficiently long trajectories. The trajectories in question contain a variety of spatial patterns but are defined by their dynamics, which falls into one of two classes and is accompanied by a bimodal distribution of a time-integrated order parameter. This bimodality is similar to that seen in stochastic models of growth [30] or the conditioned (atypical) trajectories of atomistic glasses [18]. In the present case, the coexistence emerges in response to an instruction to do classification. It occurs in a many-body system in the absence of explicit feedback, and in the typical (indeed deterministic) trajectory ensemble.

In Sec. II we introduce the model and simulation methods. In Sec. III we show that a cellular automaton can classify images by enacting a form of dynamical phase coexistence. In Sec. IV we discuss the accuracy with which linear models

*swhitelam@lbl.gov

†isaac.tamblyn@uottawa.ca

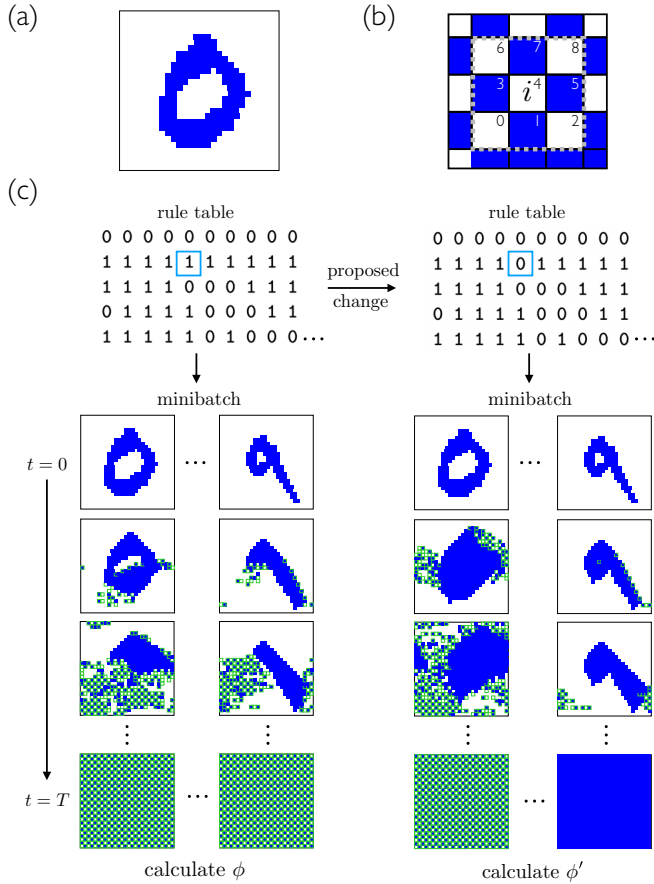


FIG. 1. (a) Example of a binarized MNIST digit on a 28×28 lattice. (b) Example of a 3×3 Moore neighborhood of site i , defining the 512-entry rule table of the cellular automaton. (c) Schematic of the Monte Carlo search procedure on the automaton rule table, designed to maximize ϕ , Eq. (4). Blue and white cells indicate states 1 and 0; green borders indicate cells whose states have changed from the previous time step.

acting on a collection of automata can classify images. We conclude in Sec. V.

II. MODEL AND SIMULATION DETAILS

We consider the MNIST data set, a collection of 28×28 -pixel grayscale images of 70 000 handwritten digits of class $\alpha = 0, 1, \dots, 9$ [14,15]. We binarized images by setting to 0 or 1 any pixel whose value is less than or greater than $1/4$ of the maximum possible intensity, respectively. One of these images is shown in Fig. 1(a). These binary images serve as the initial conditions for simulations, which take place on a square lattice of size $N = 28 \times 28$ with periodic boundary conditions. On each site $i \in [1, N]$ of the lattice lives a binary cell $S_i = 0, 1$, indicated white (0) or blue (1) in images. We evolve the lattice in time using a discrete-time deterministic cellular automaton. The automaton is a table of rules $S_i(t+1) = \mathcal{R}(s_i(t))$ that specifies the state of cell i at time $t+1$ given the state $s_i(t)$ of its environment $\{S_i(t)\}$ at time t . The table $\mathcal{R}(s)$, where $s \in [0, 511]$, consists of 512 integers 0 or 1, and is defined as follows.

The environment of cell i is the nine-neighbor-square (Moore) construction [31] shown in Fig. 1(b), denoted $\{S_i\} = (S_i^{(0)}, S_i^{(1)}, \dots, S_i^{(8)})$. Cells in this environment are indicated $S_i^{(m)}$, the m^{th} neighbor of the central site i , where $m = 0, 1, 2, \dots, 8$. As indicated in the figure, we define $m = 0$ as the bottom-left site, and increment m by moving left to right, row by row (note that $S_i^{(4)} = S_i$). The state of the environment of cell i at time t is indicated by the integer

$$s_i(t) = \sum_{m=0}^8 2^m S_i^{(m)}(t), \quad (1)$$

where $S_i^{(m)}(t) = 0, 1$ is the spin state of neighbor m at time t . For example, the state of the environment with all cells zero (white) is $s_i = 0$, and the state of the environment with all cells unity (blue) is $s_i = 2^9 - 1 = 511$. The state of the environment shown in Fig. 1(b) is $s_i = 2^1 + 2^3 + 2^5 + 2^7 = 170$.

The table $\mathcal{R}(s)$ thus has 512 entries, indicating the state $S_i(t+1)$ of cell i at time $t+1$ given the state $s_i(t)$ of its environment at time t . Each entry in the table is 0 or 1, and so there are $2^{512} \sim 10^{154}$ possible rules [31]. We start with the identity rule table, for which $S_i(t+1) = S_i(t)$. One application of the table to each cell on the lattice constitutes one step of the automaton. The total number of steps T we call the depth of the automaton.

To search for automata able to classify images, we selected a subset (a minibatch) of $M = 5000$ images from the MNIST training set, and we did a zero-temperature Metropolis Monte Carlo search [32] on the rule table. This procedure, shown schematically in Fig. 1(c), proceeds as follows [33]. Starting from the identity rule table and an automaton of depth T , we calculate an order parameter ϕ (described shortly) for the minibatch. We then propose a change of the 512-entry rule table \mathcal{R} in k positions. We alternate, every Monte Carlo step, between the choices $k = 1$ and k a uniform random number on $[1, 10]$; in Fig. 1(c), we show a change to a single entry of the rule table. Using this new rule table, we calculate the new value of ϕ , called ϕ' . If $\phi' \geq \phi$, then the change to \mathcal{R} is accepted: the proposed rule table becomes the current rule table, and ϕ is set equal to ϕ' . Otherwise, the original rule table and value of ϕ are retained.

We trained automata in two modes: fixed-depth mode, with T fixed to 10, and variable-depth mode. In the latter, we set $T = 2$ initially, and we propose a change $T \rightarrow T \pm 1$ (with equal likelihood) every 10 Monte Carlo steps. When proposing a change of T , no change is made to the rule table. In variable-depth mode, both the automaton rule table \mathcal{R} and its depth T are trainable parameters.

The fixed-depth search scheme provides a comparison for the variable-depth search scheme. The results of the latter show the emergence of trajectory phase coexistence when the depth T becomes sufficiently large. We did not anticipate the existence of this phase coexistence, nor do we know in advance, for a given image class and given realization of the Monte Carlo search scheme, how large T must be to allow this behavior. Knowing the existence of this behavior, we can attempt to reproduce it by training at various large fixed depths, but the variable-depth scheme removes the need

for such guesswork. It also emphasizes that the action of an automaton is a function of both its rule table and its depth.

To classify images, we need a suitable order parameter ϕ . One possibility is to consider the state of a given cell [34] or to count types of local environment, but these break translational invariance and are biased toward certain image types, respectively. An alternative, borrowed from studies of glasses, is *activity*, the total number of changes of state of all cells in the course of a simulation. Activity is a simple measure that does not affect translational invariance and is not biased toward any particular spatial pattern. If $S_i(t)$ is the state of cell i at time t , then we define activity

$$A(T) \equiv \sum_{t=0}^{T-1} \sum_{i=1}^N (1 - \delta_{S_i(t), S_i(t+1)}) \quad (2)$$

as the total number of changes of state of all cells under the action of the automaton. Here $\delta_{x,y}$ is the Kronecker delta, equal to 1 if $x = y$ and 0 otherwise. We further define $A_j(T)$ as the value of (2) upon starting from image j , i.e., upon setting $S_i(0) = B_{i,j}$ for all i , where $B_{i,j}$ is the binary pixel pattern of image j [shown for one image in Fig. 1(a)]. Let

$$a_j(T) \equiv \frac{A_j(T)}{NT} \in [0, 1] \quad (3)$$

be the scaled counterpart of $A_j(T)$. We then define the order parameter

$$\phi \equiv \frac{1}{M_{\bar{\alpha}}} \sum_{C(j) \neq \alpha} a_j(T) - \frac{1}{M_{\alpha}} \sum_{C(j) = \alpha} a_j(T), \quad (4)$$

where the first sum runs over all $M_{\bar{\alpha}} \equiv M - M_{\alpha}$ images not of class α in the minibatch, and the second sum runs all M_{α} images of class α [$C(j)$ returns the class of image j]. Note that ϕ is a property of the minibatch, not of individual automaton trajectories, and it has a range $[-1, 1]$ in principle and $[0, 1]$ in practice, given the initial conditions and acceptance criterion of the search scheme. The instruction to maximize (4) is the instruction to find an automaton whose dynamics is inactive if initiated from an image of class α , and as active as possible otherwise.

We note that the model and simulation algorithm are translationally invariant, but not rotationally invariant (by design, because the class of a number is not invariant to rotation). Rotational invariance (and other symmetries) could be imposed during the search scheme, by limiting the search to automaton rules that respect those symmetries. Here we carry out a general search, with no restrictions.

III. CLASSIFICATION BY DYNAMICAL PHASE COEXISTENCE

In Fig. 2 we show results of two instances of the Monte Carlo search scheme for image class $\alpha = 0$; these are typical of results for other image classes. The fixed-depth automaton learns to increase the value of ϕ , shown in panel (a), with Monte Carlo step number n . The associated histograms of activity for 60 000 binarized MNIST digits not seen during training are shown in panel (b), top. Symbols indicate nonzero values in windows of width $1/50$, while lines are a guide to the eye. The sum of symbol values for each histogram

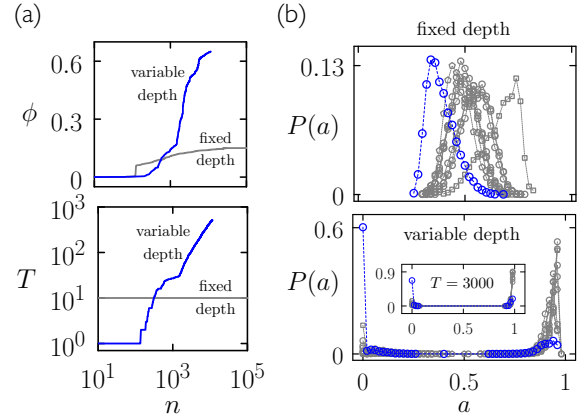


FIG. 2. (a) Monte Carlo search of the automaton rule table to maximize ϕ , Eq. (4), for digit class $\alpha = 0$. The two panels show ϕ and automaton depth T as a function of Monte Carlo step number n , for simulations carried out in fixed-depth mode ($T = 10$; gray) and variable-depth mode (blue). (b) Histograms of activity, Eq. (3), for 60 000 MNIST digits, unseen during training, of class 0 (blue) and otherwise (gray). The automaton produced by training in variable-depth mode induces a bimodal distribution of dynamical activities for each digit class.

is unity. The distribution of activity for each digit class is unimodal, and class 0 (blue lines and symbols) is on average less active than the others (gray lines and symbols; image class $\alpha \neq 0$ is denoted by gray polygons with $\alpha + 3$ sides). There is considerable overlap between classes.

Automata trained by the variable-depth search scheme behave differently. About one-third of these adopted a depth of 1, but in a majority of cases we observed behavior of the kind shown in the figure. The order parameter soon exceeds that of the fixed-depth scheme, and the depth of the automaton grows steadily, exceeding 500 time steps within the period allotted to training [Fig. 2(a)]. The activity distributions also differ from those of the fixed-depth scheme, being bimodal for each class [Fig. 2(b)]. Zeros are considerably more likely to give rise to inactive trajectories than active ones, while the opposite is true of the other classes, although exceptions are numerous (we will return to the effectiveness of these rules as classifiers). The inset shows histograms generated using the same automaton run for 3000 time steps, revealing that long trajectories begun from the MNIST digits display either high-activity or low-activity dynamics.

In Fig. 3(a) we show 100 trajectories of the automaton discovered by variable-depth search when presented with 100 images. 50 images are of class 0 (blue lines) and 50 are of other classes (gray lines). The top panel shows the time-integrated scaled activity of each trajectory, Eq. (3), as a function of the number of automaton time steps. Two distinct populations of trajectories are apparent, with the majority of zeros belonging to the low-activity population. The bottom panel shows, for the same trajectories, the instantaneous activity

$$\Delta A_j(t) \equiv A_j(t) - A_j(t-1) \quad (5)$$

accrued at time step t . This format shows that trajectories tend, after about 250 time steps, to persistently inactive or

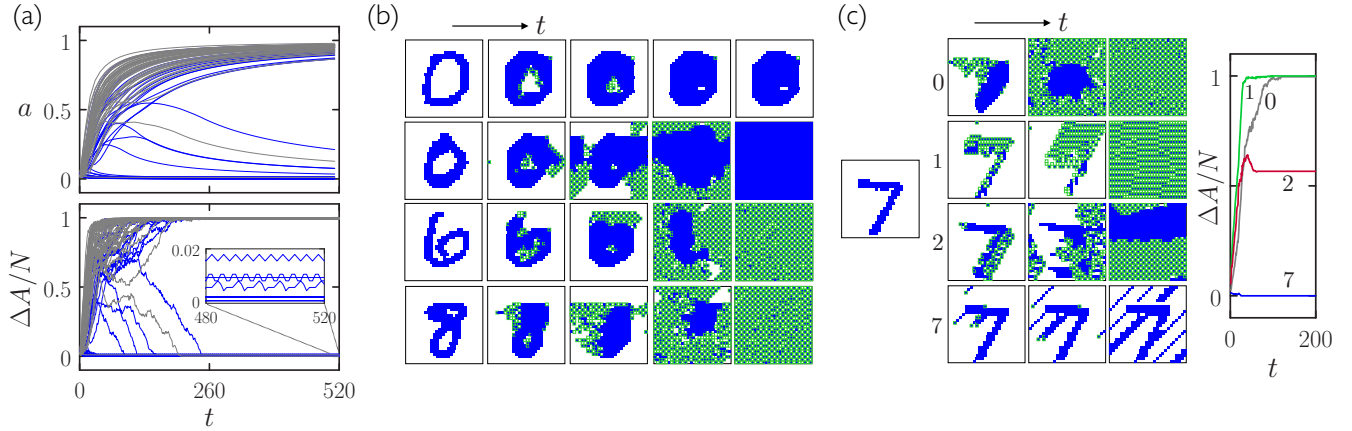


FIG. 3. (a) Time-integrated activity (top), Eq. (3), and its instantaneous counterpart (bottom), Eq. (5), as a function of time, for the automaton discovered by the variable-depth search of Fig. 2. Shown are 100 of the 60 000 trajectories comprising the histogram of the bottom panel of Fig. 2(b). Trajectories begun from digits of class 0 are blue. (b) Time-ordered snapshots generated by the automaton when presented with four different images. Cells with green boundaries have changed state from the preceding time step. (c) Time-ordered snapshots (left) and instantaneous activity as a function of time (right) for four automata. Each was trained in variable-depth mode to propagate an inactive trajectory when presented with a digit of the indicated class, and an active trajectory otherwise. Each is presented with the same digit, a 7 (shown left).

active solutions. In the inset to the bottom figure, we show some of the low-activity trajectories on a larger scale. Some of the solutions adopted are absorbing states, having zero activity, some have persistent small values of activity, and some solutions are periodic, moving between configurations that generate slightly different values of activity. Trajectories thus show a range of behavior, and they do not all correspond to the same values of activity, but nonetheless divide naturally into two populations or “phases” whose order-parameter values are tightly distributed and well separated from each other. The time-integrated order parameter shown in the top panel of Fig. 3(a) will, in the limit of infinite time, adopt the value consistent with the limiting form of the instantaneous order parameter ΔA shown in the bottom panel.

The bimodal dynamical behavior shown in Fig. 3(a) is similar to that seen in systems with explicit feedback or memory [30,35], and it resembles an emergent version of a spiking neuron [20,21]. Spiking neurons are time-dependent models of biological neurons, often modeled by differential equations, whose outputs can vary sharply with time in response to a stimulus. Here the spiking results from the many-body dynamics of the automaton: it performs a time-dependent computation from an initial condition specified by an image, and (with some error) becomes asymptotically quiescent if it detects a specified image type and spikes otherwise. Cellular automata resemble deep neural networks in the sense that they enact a nonlinear transformation by propagating a dynamics or a computer program as a function of time or depth [36–38]. However, neural-net weights usually vary with depth, making a single automaton rule closer in a sense to a recurrent neural network, or to a single activation function that performs a time-dependent computation.

In Fig. 3(b) we show time-ordered snapshots of the automaton dynamics starting from four different images. Blue and white pixels denote the two automaton cell states. Cells with a green boundary are active, meaning that their state in the previous time step was different, while cells not

labeled in this way are inactive, having the same state as in the previous time step. The top panel shows the automaton identifying a zero by propagating a low-activity trajectory. It performs what is in effect a local computation, causing the image to become increasingly compact until it adopts a state in which only one cell on the periphery of the image remains persistently active, changing state every time step. The second row shows the automaton identifying a different zero by propagating an asymptotically inactive trajectory, but in this case the computation is carried out across the whole lattice, and involves a long transient. One part of the image gives rise to an active pattern and another to an inactive pattern, and these coexist for some time before the inactive pattern consumes the lattice. The bottom two rows show the automaton identifying images that are not zeros, by propagating active trajectories. In both cases, parts of the images serve as nucleation sites for active patterns, which, after some time, consume the inactive portions of the lattice. A similar dynamics is seen when the automaton misclassifies an image: in those cases, the competition between active and inactive portions of the lattice is won by the “wrong” phase.

In Fig. 3(c) we show time-ordered snapshots of four different automata [39]. Each automaton was trained, in variable-depth mode, to generate an inactive trajectory when presented with an image of the indicated class, and an active trajectory otherwise. In each case the depths identified by the training process were several hundred time steps. All are presented with the same digit, a 7. The plot at right shows the activity accrued at each time step, Eq. (5). The 0- and 1-automata generate active trajectories, while the 7-automaton generates an inactive one. The 2-automaton illustrates a minority behavior seen in some of the automata, which display a bimodal activity distribution peaked at high and low values but with some small fraction of trajectories having intermediate values. This example falls in the latter category. The automaton produces an ambiguous answer in which, at long times, high- and low-activity patterns coexist in a spatial sense,

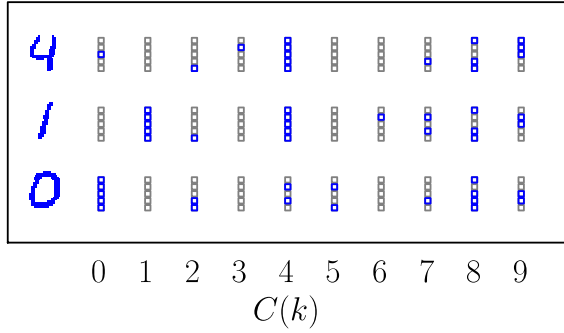


FIG. 4. Action on three images (left) of 50 automata k (squares) trained to recognize image class $C(k)$. The presence of a blue square indicates that the automaton recognizes the image by propagating a trajectory belonging to its low-activity dynamical phase.

similar to behavior seen in automata such as models of traffic [22–25] or the Ising model [16,17].

In Appendix we provide the rule tables for three automata trained to recognize different digit classes.

IV. RESERVOIR COMPUTING WITH AUTOMATA

A single automaton has a modest ability to classify images: for instance, the automaton obtained by variable-depth training in Fig. 2 classifies zeros with about 2/3 accuracy. In addition, distinct automata trained using the same procedure can identify specific digits differently, as shown in Fig. 4. There we show the action of 50 trained automata, 5 per class, on 3 MNIST digits. A blue square indicates that the automaton propagates an inactive trajectory [defined as $a_j(T) < 1/4$] when presented with the digit j shown at left, and so it recognizes it as a member of the class it is trained to identify. A gray square indicates the opposite. In these examples, all the automata that should recognize the digit as a member of their own classes do so, but false positives are common. For example, the 4-automata all identify the 1 as a 4, and the 0 is incorrectly recognized as a member of their own class by automata of 6 other classes.

Acting in concert, automata are considerably more accurate. To illustrate this fact, we adopt the approach of reservoir computing, in which a nonlinear system or reservoir—which can be a neural network, a cellular automaton, or even a physical system such as a bucket of water—enacts a transformation on a set of data, before a final layer, usually a linear model, is trained in order to do computation [9–11,40]. In reservoir computing, the reservoir is not usually trained, obviating the need for doing gradient descent on a complex nonlinear system. Here we use as a reservoir a set of cellular automata, each trained by Monte Carlo search, and we pass their output to a final layer that is subsequently trained.

We proceed as follows. We construct N_f cellular automata using the variable-depth search algorithm described in the previous section. We considered the original order parameter ϕ , Eq. (4), and five others, described below. Automata were trained on 5000 digits of the MNIST training set, and training was terminated after 5 CPU hours. Any automaton whose depth was unity after 500 Monte Carlo steps was reset (to

the identity rule table and depth 2). The automata produced in this way are heterogeneous, some exhibiting the phase coexistence-like behavior discussed previously, and some remaining relatively shallow and exhibiting unimodal activity distributions.

Once trained, we pass all 70,000 MNIST digits through all N_f automata. For each digit, we consider the N_f values of activity that result to comprise an N_f -dimensional feature vector. We train a final layer, using these feature vectors, on the 60 000 digits of the MNIST training set. With this trained layer we calculate A , its classification accuracy on the 10,000 digits of the MNIST test set.

We considered two versions of the final layer. The first is a logistic (generalized linear) model, conventional in reservoir computing. This has

$$N_{lc}(N_f) = (N_f + 1)C \quad (6)$$

real-valued parameters, where $C = 10$ is the number of digit classes. For a given digit, its output is a C -dimensional vector whose entries correspond to its prediction for the class of the digit. We trained this model using multinomial regression. The second was a nonlinear model, a single-layer neural network with 128 hidden neurons, each with tanh nonlinearities. This model has

$$N_{net}(N_f) = (128 + C)N_f + 128 + C \quad (7)$$

real-valued parameters. For a given digit its output is also a C -dimensional vector whose entries correspond to its prediction for the class of the digit. We trained this model using the Adam optimizer [41] in PyTorch [42] (using minibatches of size 1000 we trained for 200 epochs with a learning rate of 10^{-3}).

We carried out this exercise using six different reservoirs, each trained by Monte Carlo search to maximize one of six order parameters, in order to understand how important is the choice of order parameter to the ability of the final layer to do classification. To introduce these order parameters, we define the average

$$\langle \cdot \rangle_\alpha \equiv \frac{1}{M_\alpha} \sum_{C(j)=\alpha} (\cdot), \quad (8)$$

the sum running over all M_α images of class α in the minibatch, and

$$\langle \cdot \rangle_{\bar{\alpha}} \equiv \frac{1}{M_{\bar{\alpha}}} \sum_{C(j) \neq \alpha} (\cdot), \quad (9)$$

the sum running over all $M_{\bar{\alpha}} \equiv M - M_\alpha$ images not of class α in the minibatch. The first order parameter considered is the original, Eq. (4). The second is its sign-reversed counterpart,

$$\phi_2 = \langle a_j(T) \rangle_\alpha - \langle a_j(T) \rangle_{\bar{\alpha}}, \quad (10)$$

which is the instruction to maximize the activity of image class α and minimize the activity of all other classes.

The third is

$$\phi_3 = |\langle a_j(T) \rangle_\alpha - \langle a_j(T) \rangle_{\bar{\alpha}}|, \quad (11)$$

which encourages separation of the mean activity of class α from the mean of the other nine image classes, but it does not

TABLE I. Reservoir computing with trained automata. MNIST is binarized and passed through N_f automata, each trained using the order parameter ϕ_i . The resulting activities constitute a set of N_f -dimensional feature vectors on which we train a logistic model (log) or a fully connected neural network (net). A is the resulting MNIST test-set accuracy percentage, with parameter memory cost in kilobytes in brackets.

Scheme	N_f /Features	A_{log} (Mem)	A_{net} (Mem)
MNIST	784	92.6 (31)	98.0 (433)
binarized MNIST	784	92.1 (31)	97.7 (433)
ϕ	500	95.3 (52)	96.3 (309)
ϕ_2	500	95.4 (52)	96.2 (309)
ϕ_3	500	95.3 (52)	96.3 (309)
ϕ_4	500	96.8 (52)	97.2 (309)
	870	97.3 (91)	98.1 (537)
ϕ_5	500	94.6 (52)	95.3 (309)
ϕ_6	80	91.9 (8)	95.2 (50)
	255	95.7 (27)	96.9 (158)
	500	96.7 (52)	97.3 (309)
all	3370	97.7 (350)	98.1 (2076)
untrained	500	88.0 (52)	93.0 (309)

specify whether the activity of class α should be greater than or less than those of the other classes.

The fourth is the preceding order parameter with an additional constraint on the standard deviation $\sigma_\alpha \equiv \sqrt{\langle a_j(T)^2 \rangle_\alpha - \langle a_j(T) \rangle_\alpha^2}$ of the distributions of the chosen class and its complement:

$$\phi_4 = \phi_3 - \frac{1}{2}(\sigma_\alpha + \sigma_{\bar{\alpha}}). \quad (12)$$

For each of the order parameters ϕ to ϕ_4 we constructed reservoirs in which automata instructed to identify each class were equally numerous.

The fifth order parameter is a version of ϕ_4 that rewards separation of all $C = 10$ classes and the minimization of the variance of their distributions:

$$\phi_5 = \frac{1}{P} \sum_{\alpha \neq \beta} |\langle a_j(T) \rangle_\alpha - \langle a_j(T) \rangle_\beta| - \frac{1}{C} \sum_{\alpha=1}^C \sigma_\alpha, \quad (13)$$

the first sum running over all $P = C(C-1)/2 = 45$ pairs of classes and the second running over all $C = 10$ classes.

The final order parameter is the analog of ϕ_5 for two class types, α and β ,

$$\phi_6 = |\langle a_j(T) \rangle_\alpha - \langle a_j(T) \rangle_\beta| - \frac{1}{2}(\sigma_\alpha + \sigma_\beta). \quad (14)$$

When constructing reservoirs using ϕ_6 as an order parameter, we assign each automaton at random to one of the 45 class-pairs (α, β) . Each used only digits of those two classes from the 5000-digit minibatch, approximately 1000 digits in all cases.

Results are shown in Table I. For reference, the first two rows show classification accuracy A of the two versions of the final layer using the original MNIST feature vectors and

MNIST binarized using the threshold (1/4 of the maximum pixel intensity) used by the automata. Some small loss of accuracy is incurred by binarizing MNIST.

We shall focus first on results using the logistic model as a final layer, shown in the third column of the table. The order parameters ϕ to ϕ_3 , which encourage automata to make the dynamics of one digit class more or less active than the others, perform similarly. Reservoirs of 500 automata each provide about 3% improvement in logistic-model classification accuracy over the original MNIST data, indicating that the nonlinear transformations enacted by the automata are beneficial, separating classes in phase space in a way that a logistic model finds easier to process. The three order parameters provide essentially the same numerical benefit, emphasizing that activity is an order parameter agnostic to digit type: it does not matter if automata are instructed to make certain classes more or less active than others.

Moving from the order parameter ϕ_3 to ϕ_4 , we see that some additional numerical improvement is afforded by the extra terms in ϕ_4 that encourage automata to produce narrow distributions of activity. Moreover, more reservoir automata generally equates to better performance, and it is possible to construct reservoirs possessing more members than there are features in MNIST: with 870 features, the logistic-model accuracy exceeds by about 5% the corresponding number using the original data set.

The order parameter ϕ_5 , which encourages each automaton to separate activity distributions for all classes, performs less well than the other order parameters. ϕ_5 is the most difficult order parameter for an individual automaton to maximize, and this difficulty results in a reservoir slightly less expressive than those made using “easier” order parameters.

The order parameter ϕ_6 , which encourages separation of two classes and ignores the rest, performs similarly to ϕ_4 . We show also that classification accuracy increases with increasing reservoir size. Indeed, the accuracy of the logistic model increases further if presented with all of the previous reservoirs combined, shown in the penultimate row of the table.

As a point of comparison with the trained reservoirs, we show in the final row of the table the performance of a reservoir of 500 automata whose rules were chosen randomly, as is conventional in reservoir computing. For each, we drew a random number λ uniformly on $(0,1]$, and populated each entry of its rule table with a 1 with probability λ and a 0 otherwise (λ is then the mean fraction of 1’s in the rule table [44]). We trained the depth of each automaton by starting at $T = 1$ and increasing T by 1 until the order parameter ϕ_5 began to decrease. This reservoir of randomly chosen rules performs less well with the linear classifier than does the original data set, indicating that information has been lost by the reservoir and highlighting the utility of training the automata.

The performance of the nonlinear final layer, shown in the fourth column of Table I, is different from that of the logistic model. With the neural network, the best reservoirs and the original data set yield similar results. Comparison of the performance of the two versions of the final layer indicates that the nonlinear transformations enacted by the reservoir are unnecessary for the neural network, which is expressive

enough that it can extract similar information using the original data set.

We turn finally to the memory consumption of each method. We represent real-valued parameters with 32-bit precision, and binary variables, which specify the automata rule tables, cost 1 bit. Thus for N_f features, the linear classifier costs $32 \text{ bits} \times N_{lc}(N_f)$, given by Eq. (6), and the net costs $32 \text{ bits} \times N_{net}(N_f)$, given by Eq. (7), while each automaton costs $1 \text{ bit} \times 512$. The resulting memory cost of the methods (to the nearest kilobyte; 8 bits = 1 byte) is shown in brackets in the table. For example, using binarized MNIST, the net achieves an accuracy of 97.7% at a memory cost of 433 kb; all the reservoirs connected to a linear classifier achieve 97.7%, with a memory cost of 350 kb. 870 automata trained using the order parameter ϕ_4 achieve the slightly lower accuracy of 97.3%, but at a reduced memory cost of 91 kb. This is equivalent to about 22,750 real-valued parameters, and we can assess the accuracy of similarly sized fully connected networks acting directly on MNIST or binarized MNIST. 22 750 parameters equates to a single-layer fully connected net of about 28 hidden nodes, which scores 96.4% (MNIST) or 95.7% (binarized MNIST); or to a two-layer net with about 27 nodes per layer (96.2% or 95.9%); or to a three-layer net with about 26 nodes per layer (96.1% or 95.6%); or to a four-layer net with about 25 nodes per layer (96.2% or 95.1%); or to a ten-layer net with about 22 nodes per layer (94.6% or 93.5%). Thus a linear classifier acting on a set of automata can achieve comparable accuracy to fully connected neural networks acting directly on MNIST, with a smaller memory cost.

Convolutional neural networks, whose inputs involve local kernels that sweep over an entire image [14,15,37,38], are closer in construction to a cellular automaton than is a fully connected neural network. To compare the performance of fully connected and convolutional neural networks, we calculated the accuracy on real-valued (not binarized) MNIST using convolutional neural networks having a kernel size of 3×3 pixels. We considered 15 different network architectures, divided into three groups according to the number of parameters they contain. Networks in these groups have a parameter memory cost of 29.7, 62.6, and 142.1 kilobytes, respectively, and yield a mean classification accuracy of 96.8%, 98.1%, and 98.4%. Those numbers (the memory cost and accuracy) can be compared with the top-right entry of Table I, which shows the fully connected net to be less accurate per unit memory cost.

V. CONCLUSIONS

We have introduced dynamical activity, an order parameter with no bias toward any particular environment or spatial position, as a metric for image classification by cellular automata. Zero-temperature Metropolis Monte Carlo search of rules and depths for general 2D automata has identified algorithms that classify MNIST images by enacting a form of dynamical phase coexistence, propagating trajectories belonging to high- or low-activity dynamical phases. Individual automata achieve this classification with modest accuracy. Collections or reservoirs of such automata can be used by a linear model to classify images with comparable accuracy (and smaller

memory cost) than fully connected neural networks that act on the pixels of the original image.

Conceptually, the automata described here evoke concepts or display features seen in other studies, but in detail they are distinct from automata described previously. For example, several authors have discussed the computational potential of automaton rules close to phase transitions in rule space [28,29,43,44]. We can view the transition between the behaviors shown in the top and bottom panels of Fig. 2(b) as a phase transition in rule space, analogous to the transition shown by the Ising model as its temperature is reduced below the critical value. However, the defining character of the automata described here is that they do computation by enacting a form of dynamical phase coexistence, and so the computation is enabled by a phase transition taking place in trajectory space, not rule space.

In addition, automata of this nature display elements of the four Wolfram classes of cellular automata [3], but they are not completely described by these rules. For example, Fig. 3 shows that a single automaton, starting from different initial conditions, can generate a trajectory similar to that of a Class 1 automaton (ending in a homogeneous state, i.e., a limit point), or a Class 2 automaton (generating structures periodic in time, i.e., limit cycles), and it can generate complex patterns with long transients. In these respects, it resembles a Class 4 automaton, which can generate both repetitive and complex patterns, but the defining characteristic of the automaton is that its dynamical ensemble contains trajectories of two distinct types, defined by their activity, a concept separate from those considered in the Wolfram scheme. Indeed, the dynamical phase coexistence seen in these automata resembles that seen in some stochastic systems [16–18,30], a connection of the kind anticipated in Question 10 (“What is the correspondence between cellular automata and stochastic systems?”) of Ref. [45].

ACKNOWLEDGMENTS

This work was performed as part of a user project at the Molecular Foundry, Lawrence Berkeley National Laboratory, supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02–05CH11231. I.T. acknowledges Natural Sciences and Engineering Research Council of Canada (NSERC).

APPENDIX: EXAMPLE AUTOMATA

In this Appendix, we specify three automata that learned, under variable-depth search, to recognize three different digit classes via the dynamical phase coexistence mechanism described in the text. Entries in the rule table are numbered sequentially from 0 to 511, according to Eq. (1) and the associated description.

(i) Class 0: depth $T = 512$, rule table $\mathcal{R} =$

```

0 0 0 0 0 0 0 1 1 0 1 0 0 1 1 1 0 1 1 1 1 0 0
0 1 0 1 0 1 0 0 1 1 0 1 1 0 0 0 1 0 1 1 1 0 0
0 1 0 0 1 1 0 0 0 1 0 0 0 1 0 1 0 1 0 0 0 1 0
0 1 1 0 0 1 1 1 0 1 0 1 0 1 1 0 0 0 0 0 1 0 1
0 1 0 1 1 1 1 1 1 0 1 0 1 0 1 1 1 1 1 1 1 0 0

```

```

0 0 0 0 0 0 1 0 1 1 1 0 1 0 0 0 0 0 0 1 0 1 1
1 1 1 1 1 1 1 0 1 1 0 0 1 0 0 0 0 0 0 1 1 1 1
1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 0 1 0 0 1 0
1 0 0 0 0 1 0 1 0 0 1 1 0 0 0 0 0 0 1 1 1 1 0
1 0 0 0 0 1 0 1 1 0 1 0 1 1 1 1 1 1 1 1 1 0 1
1 1 1 1 0 1 0 1 0 1 1 1 1 1 0 0 1 1 1 0 0 1 1 0
0 1 1 0 1 0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 1 1
1 0 0 0 0 1 0 0 0 0 1 1 1 1 0 0 1 0 0 0 1 0 0 0
0 0 0 0 1 0 0 0 1 1 0 0 1 0 1 0 0 1 0 0 0 0 0
0 1 0 0 0 0 1 0 1 1 1 1 0 0 1 0 1 0 1 0 1 0 1
0 0 1 0 0 0 0 1 0 1 0 0 0 0 0 1 1 1 1 1 0 0 0
1 0 0 1 1 0 1 1 0 1 1 0 0 0 0 0 0 1 1 1 1 0 0
0 1 0 0 1 0 0 0 1 1 0 1 1 1 0 0 1 1 0 0 0 1 1
1 1 1 0 0 0 1 0 1 0 1 0 0 1 1 1 1 1 1 0 1 0 1
0 1 1 1 0 1 1 0 1 0 0 0 1 1 1 1 1 1 0 1 0 0 0
1 0 0 0 0 1 0 0 1 0 0 1 1 1 0 1 1 1 1 1 1 1 1
0 1 1 1 1 1
    
```

(ii) Class 1: depth $T = 698$, rule table $\mathcal{R} =$

```

0 0 0 1 0 1 0 1 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0
0 0 1 0 1 0 0 0 0 0 1 1 1 0 1 0 1 0 0 0 1 0 1
1 1 0 0 1 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 0 0 0
1 0 1 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0 1 0 1 0 1
0 0 1 1 0 1 0 1 0 0 1 1 0 0 1 1 0 1 0 1 0 0 0
0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 1 1 0 1 1 1 0 0
1 0 0 0 1 0 0 0 1 1 1 1 1 1 0 1 0 1 0 1 0 0 0
1 1 1 0 1 1 1 1 0 1 0 0 0 0 1 1 0 1 0 1 0 1 1
0 0 1 0 1 1 0 1 0 0 1 1 1 1 1 1 1 0 0 1 0 0 0
1 1 0 0 1 0 1 0 0 0 1 1 1 0 0 1 0 1 1 1 1 1 1
1 0 0 0 1 0 0 1 0 0 0 0 0 1 1 1 0 1 0 0 0 1 0
0 1 1 1 1 0 1 0 0 0 1 0 1 1 1 0 1 1 1 0 0 1 1
0 0 1 0 0 0 0 1 0 0 0 0 0 1 1 1 0 0 0 1 1 0 0
1 0 1 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 1
    
```

```

0 1 1 1 1 1 0 0 1 0 0 0 1 1 1 0 0 0 0 1 0 0 0
1 1 1 0 1 0 0 1 1 1 1 0 1 0 1 0 0 1 0 0 0 1 0
1 0 0 0 0 1 0 0 1 1 0 1 0 1 1 0 1 1 1 1 0 1 1
1 1 0 1 1 1 0 1 1 0 0 1 1 0 0 1 1 0 1 0 1 1 0
0 0 1 1 1 1 0 1 1 1 1 0 1 0 1 1 0 1 1 0 1 1 0
0 1 1 0 1 0 0 0 0 0 0 1 1 0 1 1 1 1 1 1 1 0 1 0
1 0 1 1 0 0 0 0 1 1 1 0 0 1 0 1 0 1 0 1 1 1 0
1 1 1 0 1 0 1 0 1 1 1 1 1 0 0 0 1 1 0 1 0 0 0
0 1 1 1 0 0
    
```

(iii) Class 3: depth $T = 699$, rule table $\mathcal{R} =$

```

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 1 1 1 0 1
1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 0 1 0 1 0 0 0 1
1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 1 0 0 0 0
0 1 1 1 1 0 0 0 1 1 1 1 1 1 0 1 1 0 1 0 0 1 0 1
1 1 1 1 1 0 1 0 0 0 1 1 1 1 0 1 1 1 0 1 0 0 0 1
1 1 0 0 0 1 1 1 0 1 0 0 1 0 0 0 0 0 0 0 1 0 1 1
0 0 0 1 1 1 1 1 0 0 1 0 0 0 1 1 0 0 1 0 1 1 0
0 0 1 0 1 0 1 0 1 1 0 0 1 1 1 1 1 1 0 0 1 1 0 1
1 0 1 0 0 1 0 1 0 0 1 0 0 0 1 0 1 1 0 0 1 0 1
1 0 1 0 0 1 0 0 1 1 0 0 0 0 0 0 1 0 0 0 1 1 1
1 0 1 0 0 1 0 1 1 0 0 1 0 0 1 1 0 1 1 1 1 1 1
1 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 1 0 1
1 0 0 0 1 1 1 0 1 1 0 1 1 1 1 1 1 1 0 1 1 1 0
1 0 1 1 0 1 1 1 0 0 1 1 1 1 1 0 0 1 0 1 1 1 0
1 1 0 1 0 1 0 1 0 0 1 1 1 1 1 0 0 1 1 1 1 1 1
0 0 0 1 0 1 1 1 0 1 1 0 1 1 1 0 1 1 1 1 1 1 1
1 0 1 0 1 1 1 1 1 0 0 0 1 1 0 1 0 0 1 0 1 1 0
1 0 0 0 0 0 1 1 1 1 0 0 1 1 0 0 1 0 1 0 0 1 0
1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 0
0 1 1 1 1 0 0 1 1 1 1 0 1 1 0 0 0 1 1 0 1 0 0
1 1 1 1 1 1 0 0 1 0 0 0 1 0 1 0 1 1 0 1 1 1 0
1 0 1 1 0 1 1 0 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1
1 1 1 1 1 1
    
```

[1] J. Von Neumann, The general and logical theory of automata, in *Systems Research for Behavioral Science Systems Research* (Routledge, London, UK, 1968).

[2] J. Conway *et al.*, The game of life, *Sci. Am.* **223**, 4 (1970).

[3] S. Wolfram, Universality and complexity in cellular automata, *Physica D* **10**, 1 (1984).

[4] S. Wolfram, Statistical mechanics of cellular automata, *Rev. Mod. Phys.* **55**, 601 (1983).

[5] B. Chopard and M. Droz, *Cellular Automata* (Springer, Berlin, Germany, 1998), Vol. 1.

[6] I. Santé, A. M. García, D. Miranda, and R. Crecente, Cellular automata models for the simulation of real-world urban processes: A review and analysis, *Landscape Urban Plann.* **96**, 108 (2010).

[7] N. Ganguly, P. Maji, S. Dhar, B. K. Sikdar, and P. P. Chaudhuri, Evolving cellular automata as pattern classifier, in *International Conference on Cellular Automata* (Springer, Berlin, Germany, 2002), pp. 56–68.

[8] P. Maji, C. Shaw, N. Ganguly, B. K. Sikdar, and P. P. Chaudhuri, Theory and application of cellular automata for pattern classification, *Fundam. Inform.* **58**, 321 (2003).

[9] O. Yilmaz, Machine learning using cellular automata based feature expansion and reservoir computing. *J. Cell. Autom.* **10** (2015).

[10] S. Nichele and M. S. Gundersen, Reservoir computing using non-uniform binary cellular automata, [arXiv:1702.03812](https://arxiv.org/abs/1702.03812).

[11] A. Morán, C. F. Frasser, and J. L. Rosselló, Reservoir computing hardware with cellular automata, [arXiv:1806.04932](https://arxiv.org/abs/1806.04932).

[12] E. Randazzo, A. Mordvintsev, E. Niklasson, M. Levin, and S. Greydanus, Self-classifying mnist digits, *Distill* **5**, e00027 (2020).

[13] A. Morán, C. F. Frasser, M. Roca, and J. L. Rosselló, Energy-efficient pattern recognition hardware with elementary cellular automata, *IEEE Trans. Comput.* **69**, 392 (2019).

[14] Y. LeCun, L. D. Jackel, L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. A. Muller, E. Sackinger, P. Simard *et al.*, Learning algorithms for classification: A comparison on handwritten digit recognition, *Neural Netw.: Stat. Mech. Persp.* **261**, 2 (1995).

[15] <http://yann.lecun.com/exdb/mnist/>.

[16] W. Kinzel, Phase transitions of cellular automata, *Z. Phys. B* **58**, 229 (1985).

- [17] J. J. Binney, N. J. Dowrick, A. J. Fisher, and M. Newman, *The Theory of Critical Phenomena: An Introduction to the Renormalization Group* (Oxford University Press, Oxford, UK, 1992).
- [18] L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, Dynamic order-disorder in atomistic models of structural glass formers, *Science* **323**, 1309 (2009).
- [19] S. V. Meshkov, Low-frequency dynamics of lennard-jones glasses, *Phys. Rev. B* **55**, 12113 (1997).
- [20] H. Paugam-Moisy and S. M. Bohte, Computing with spiking neuron networks, *Handbook Nat. Comput.* **1**, 1 (2012).
- [21] W. Gerstner and W. M. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity* (Cambridge University Press, Cambridge, UK, 2002).
- [22] R. Jiang and Q.-S. Wu, Cellular automata models for synchronized traffic flow, *J. Phys. A: Math. Gen.* **36**, 381 (2003).
- [23] D. E. Wolf, Cellular automata for traffic simulations, *Physica A* **263**, 438 (1999).
- [24] J. P. L. Neto, M. L. Lyra, and C. R. Da Silva, Phase coexistence induced by a defensive reaction in a cellular automaton traffic flow model, *Physica A* **390**, 3558 (2011).
- [25] R. M. DôSouza, Coexisting phases and lattice dependence of a cellular automaton model for traffic flow, *Phys. Rev. E* **71**, 066112 (2005).
- [26] A. Schadschneider, D. Chowdhury, E. Brockfeld, K. Klauck, L. Santen, and J. Zittartz, A new cellular automata model for city traffic, [arXiv:cond-mat/9911312](https://arxiv.org/abs/cond-mat/9911312).
- [27] S. Yukawa, M. Kikuchi, and S.-I. Tadaki, Dynamical phase transition in one dimensional traffic flow model with blockage, *J. Phys. Soc. Jpn.* **63**, 3609 (1994).
- [28] W. Li, N. H. Packard, and C. G. Langton, Transition phenomena in cellular automata rule space, *Physica D* **45**, 77 (1990).
- [29] W. K. Wootters and C. G. Langton, Is there a sharp phase transition for deterministic cellular automata? *Physica D* **45**, 95 (1990).
- [30] K. Klymko, J. P. Garrahan, and S. Whitlam, Similarity of ensembles of trajectories of reversible and irreversible growth processes, *Phys. Rev. E* **96**, 042126 (2017).
- [31] N. H. Packard and S. Wolfram, Two-dimensional cellular automata, *J. Stat. Phys.* **38**, 901 (1985).
- [32] Also known as random-mutation hill climbing [46,47].
- [33] This approach is similar to those that use genetic algorithms to search for automaton rules that perform particular computations [48–51], except that the present search scheme applies mutations and the Metropolis acceptance criterion to a single individual, rather than applying genetic operations to a population of individuals. For continuous variables and small mutations, the present procedure constitutes noisy clipped gradient descent on the loss surface [52].
- [34] F. Gers, H. de Garis, and M. Korkin, Codi-1bit: A simplified cellular automata based neuron model, in *European Conference on Artificial Evolution* (Springer, Berlin, Germany, 1997), pp. 315–333.
- [35] S. Whitlam and D. Jacobson, Varied phenomenology of models displaying dynamical large-deviation singularities, *Phys. Rev. E* **103**, 032152 (2021).
- [36] J. Schmidhuber, Discovering neural nets with low kolmogorov complexity and high generalization capability, *Neural Netw.* **10**, 857 (1997).
- [37] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Netw.* **61**, 85 (2015).
- [38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).
- [39] Times of snapshots for each row are as follows. 0: 10,38,200; 1: 3,9,200; 2: 3,13,200; 7: 5,10,200.
- [40] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, Recent advances in physical reservoir computing: A review, *Neural Netw.* **115**, 100 (2019).
- [41] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [42] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, Pytorch: An imperative style, high-performance deep learning library, *Advances in Neural Information Processing Systems 32* (2019).
- [43] M. Mitchell, J. P. Crutchfield, and P. T. Hraber, Dynamics, computation, and the edge of chaos: A re-examination, in *Santa Fe Institute Studies in the Sciences of Complexity* (Addison-Wesley, Reading, MA, 1994), Vol. 19, pp. 497–497.
- [44] C. G. Langton, Computation at the edge of chaos: Phase transitions and emergent computation, *Physica D* **42**, 12 (1990).
- [45] S. Wolfram, Twenty problems in the theory of cellular automata, *Phys. Scr.* **1985**, 170 (1985).
- [46] M. Mitchell, J. Holland, and S. Forrest, When will a genetic algorithm outperform hill climbing, *Advances in Neural Information Processing Systems 6* (1993).
- [47] M. Mitchell, *An Introduction to Genetic Algorithms* (MIT Press, Cambridge, MA, 1998).
- [48] T. Suzudo, Searching for pattern-forming asynchronous cellular automata—an evolutionary approach, in *International Conference on Cellular Automata* (Springer, Berlin, Germany, 2004), pp. 151–160.
- [49] M. Mitchell, J. P. Crutchfield, R. Das *et al.*, Evolving cellular automata with genetic algorithms: A review of recent work, in *Proceedings of the First International Conference on Evolutionary Computation and Its Applications (EvCA96)* (Moscow, 1996), Vol. 8.
- [50] G. M. B. Oliveira, L. G. A. Martins, L. B. de Carvalho, and E. Fynn, Some investigations about synchronization and density classification tasks in one-dimensional and two-dimensional cellular automata rule spaces, *Electron. Notes Theor. Comput. Sci.* **252**, 121 (2009).
- [51] A. Chavoya and Y. Duthen, Using a genetic algorithm to evolve cellular automata for 2d/3d computational development, in *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation* (2006), pp. 231–232.
- [52] S. Whitlam, V. Selin, S.-W. Park, and I. Tamblyn, Correspondence between neuroevolution and gradient descent, *Nat. Commun.* **12**, 6317 (2021).