Automatic filters for the detection of coherent structure in spatiotemporal systems

Cosma Rohilla Shalizi,[1,*] Robert Haslinger,[2,†] Jean-Baptiste Rouquier,[3,‡] Kristina Lisa Klinkner,[4,§] and Cristopher Moore[5,6,7,‖]
[1]*Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan 48109 USA*
[2]*Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts 02129 USA*
[3]*Laboratoire de l'Informatique du Parallélisme, École Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon, France*
[4]*Statistics Department, University of Michigan, Ann Arbor, Michigan 48109 USA*
[5]*Department of Computer Science, University of New Mexico, Albuquerque, New Mexico 87131 USA*
[6]*Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131 USA*
[7]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501 USA*
(Received 27 July 2005; published 2 March 2006)

Most current methods for identifying coherent structures in spatially extended systems rely on prior information about the form which those structures take. Here we present two approaches to *automatically* filter the changing configurations of spatial dynamical systems and extract coherent structures. One, *local sensitivity* filtering, is a modification of the local Lyapunov exponent approach suitable to cellular automata and other discrete spatial systems. The other, *local statistical complexity* filtering, calculates the amount of information needed for optimal prediction of the system's behavior in the vicinity of a given point. By examining the changing spatiotemporal distributions of these quantities, we can find the coherent structures in a variety of pattern-forming cellular automata, without needing to guess or postulate the form of that structure. We apply both filters to elementary and cyclical cellular automata (ECA and CCA) and find that they readily identify particles, domains, and other more complicated structures. We compare the results from ECA with earlier ones based upon the theory of formal languages and the results from CCA with a more traditional approach based on an order parameter and free energy. While sensitivity and statistical complexity are equally adept at uncovering structure, they are based on different system properties (dynamical and probabilistic, respectively) and provide complementary information.

DOI: 10.1103/PhysRevE.73.036104 PACS number(s): 89.75.Fb, 89.75.Kd, 05.65.+b, 02.50.Tt

I. INTRODUCTION

Coherent structures are ubiquitous in nature, the result of complex patterns of interaction between simple units [1,2]. Such structures are not simply epiphenomena of the microscopic physics, but instead often govern the system's macroscopic properties, providing powerful collective degrees of freedom—they make good "handles" on the system [3]. Condensed matter physics provides a host of equilibrium systems in which this is true [4]—for instance in antiferromagnets [5] and liquid crystals [6]—and the anomalous properties of the high-temperature superconductors may be due, in part, to the presence of ordered charge density waves (stripes) [7]. Coherent structures also govern the behavior of many nonequilibrium systems [2]. The obvious examples are biological systems which are rife with coherent structure and are most certainly far from equilibrium (until they die) [8,9]. Another example may be found in economics where the spatial pattern of economic activity has been postulated to be governed by the emergence of coherent structures in the form of cities and hierarchies of urban centers [10]. It has even been proposed that coherent structures can perform adaptive information processing tasks and are therefore the typical physical embodiment of emergent computation [11–15]. This conjecture is supported by both simulations of biological morphogenesis [16] and experiments on the adaptive responses of stomata in plants [17]. In view of the intrinsic scientific interest of complex, spatially extended systems and the potential power of coherent structures to describe them concisely, it is important to develop techniques which can uncover and define coherent structures in both the equilibrium and nonequilibrium cases. In this paper, we present two such filters, applicable to spatially extended discrete systems.

Most existing methods for identifying coherent structures rely on knowing some details about the structure one is looking for. *Matched filters* are designed so that signals of certain known character maximize the response [18]. Such filters may match either the structures themselves (e.g., [19]) or the uninteresting details of the background in which the structures are embedded (e.g., [20,21]). The latter technique has been particularly useful in equilibrium systems with broken symmetry. The identification of an appropriate order parameter and an associated free energy implicitly defines a filter which matches the background—i.e., the ordered state. Departures from the ordered state, such as domain walls, vortices, and other topological defects, constitute mesoscopic structure. While extremely successful, this method requires a lot of trial and error and some knowledge of the underlying

*Present address: Statistics Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Electronic address: cshalizi@cmu.edu
†Electronic address: robhh@nmr.mgh.harvard.edu
‡Electronic address: jean-baptiste.rouquier@ens-lyon.fr
§Present address: Statistics Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Electronic address: klinkner@cmu.edu.
‖Electronic address: moore@santafe.edu

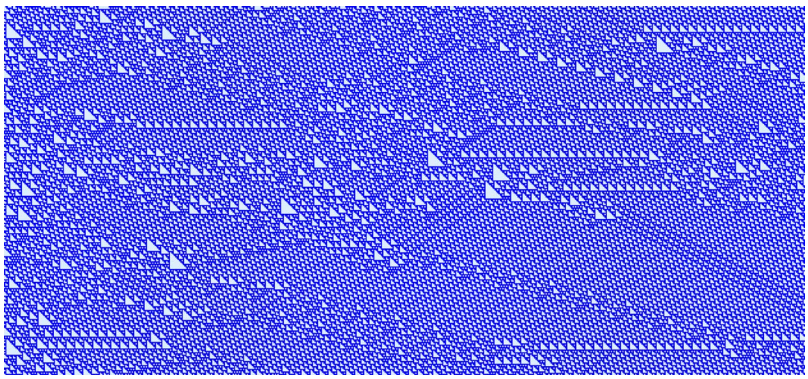

FIG. 1. (Color online) Typical space-time diagram of rule 110; time advances from left to right. Note the presence of the regular background domain and the particles which stand out in the background by contrast.

microdynamics. In far-from-equilibrium systems, much of the progress in detecting coherent structures has involved the application of regular language theory [22] to one-dimensional (1D) deterministic cellular automata [23–26]. Specific filters, adapted to particular cellular automata rules, have been designed to detect structures dynamically [27–31]. In these cases as well, however, one must know a lot about the system's dynamics to work out, by hand, a suitable filter. Furthermore, to apply such notions to cellular automata of higher dimension, one would have to extend regular language theory to such dimensions, which presents significant difficulties [32]. Extensions to stochastic systems would be even more problematic.

The two filters introduced in this paper address these issues.[1] Both are "automatic" filters, in that they require no prior information about the system's microscopic dynamics or the structures generated. They can be applied to both equilibrium and nonequilibrium systems of, in principle, any dimensionality. One of them is applicable to systems with stochastic dynamics. The two filters are, however, based upon very different system properties. Section II introduces the *local sensitivity*, an adaptation of Lyapunov exponents to cellular automata, measuring the degree to which small perturbations can alter the dynamics. Section III defines the *local statistical complexity*, which measures the spatiotemporal distribution of the amount of information needed to optimally predict the system's dynamics. It is therefore a probabilistic measure, completely compatible with any underlying stochastic dynamics. Both methods identify coherent structures by filtering the system's changing configuration with respect to sensitivity (or complexity) and tracking the changing spatial distributions of these fields. We apply each filter to the task of detecting coherent structures in cellular automata. Sections II B and III C look at "elementary" (one-dimensional, binary, range-1) cellular automata (ECA), where we find that both filters readily identify ECA particles, domains, and domain walls, although they emphasize these structures to differing degrees. In Sec. IV we use cyclic cellular automata (CCA), a self-organizing model of excitable media, to compare the local sensitivity and statistical complexity with a more traditional order parameter approach. (We identify the appropriate order parameter for this system in Sec. IV A.) Our conclusion gives a summary and discusses some open questions. Finally, the Appendixes discuss previous attempts to adapt the Lyapunov exponent to cellular automata (Appendix A) and provide details on the construction of the CCA order parameter and effective free energy (Appendix B).

[1]Open-source code, in the objective CAML language, is available from http://www.cscs.umich/afics/.

II. LOCAL SENSITIVITY

The first of our filters for coherent structures is based upon a measure of *local sensitivity*: the instability of the discrete cellular automata field under local perturbation. Filtering with respect to local sensitivity is motivated by a desire to distinguish autonomous objects from the rest of the CA's configuration field. By "autonomous objects" we mean those structures that have the greatest influence upon the future dynamics of the system. Perturbing an autonomous object will affect the future a great deal. In contrast, perturbations of the rest of the system, the dependent parts, should be quickly overridden and "healed" by the autonomous dynamics. The set of autonomous objects constitutes a high-level model of the system, and knowledge of them should allow us to infer most of the rest of the dynamics, as well as the objects' own futures.

A famous example of a system governed by autonomous objects is rule 110 of elementary cellular automata. This is a nonconservative, nearest-neighbor, binary rule, whose configuration field (Fig. 1) tends to exhibit background domains with long spatial and temporal periods and large-scale particles. Knowledge of the position of these particles carries considerable predictive power [25,33]. A second example may be found in the cyclic cellular automata (see below, Sec. IV). For certain parameter values the CCA configuration field evolves into competing spiral waves, the cores of which are autonomous (see Fig. 2).

For these two examples—rule 110 and spiral CCA—the presence of structure is obvious. However this is not always the case (see, e.g., Ref. [30] and the discussion of rule 146 in Sec. II B below). Even when it is, reliably identifying all the structure present in a CA's configuration field is difficult to do by eye and it is often not at all clear how autonomous such structures are. One might therefore hope for an automatic filter capable of not only distinguishing particles, domains, and extended objects like strings and domain walls in large spatiotemporal systems, but also inferring their degree of autonomy.

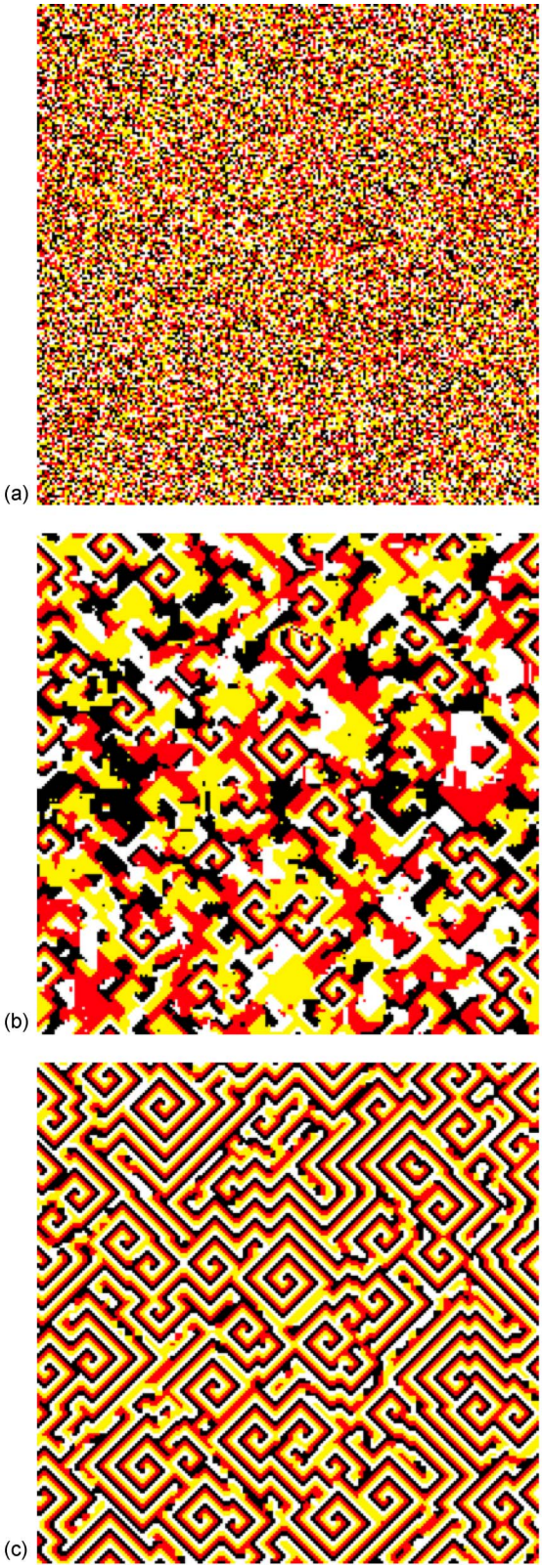

FIG. 2. (Color online) Phenomenology of the cyclic CA (defined in Sec. IV) in one of its spiral-forming phases ($\kappa=4$ colors, excitation threshold $T=2$, rule radius $r=1$). Time progresses from top to bottom, the panels depicting (a) the random initial conditions, (b) the formation of spiral wave cores, and (c) their eventual domination of the entire lattice.

A. Defining local sensitivity

Since we want to filter cellular automata with respect to autonomy, a natural point of departure is some mathematical formalization of the idea that a system's future may be very sensitive to details of its present state. Devaney's definition of chaos [34] and the notions of topological and metric (Kolmogorov-Sinai) entropy rates [35] are two ways in which this intuition can be formalized. Here, however, we will begin with the idea of a dynamical system's Lyapunov exponents, since these provide a more refined and quantitative measure of sensitivity dependence than either the mere fact of chaos or than the metric entropy rate (which is generally the sum of the positive Lyapunov exponents). We briefly review the definition of Lyapunov exponents before developing our filter.

A one-dimensional map f, for which the system's state is given by a single real variable x, has a single Lyapunov exponent, given by the following limit:

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log|Df^n(x_0)|,$$

where D is the derivative operator. (The limit can be shown to exist and be identical for almost all initial points x_0.) The interpretation is that, if one begins with two points x_0 and $x_0+\epsilon$, the distance between their future images, after n time steps, is approximately $\epsilon e^{n\lambda}$; this relation becomes exact as $\epsilon \to 0$. In the case of a dynamical system described by m state variables, the Lyapunov exponents are defined as the m eigenvalues, $\lambda_1, \ldots, \lambda_m$, of the matrix

$$\Lambda = \lim_{n\to\infty} [D^\dagger f^n(x_0) Df^n(x_0)]^{1/2n},$$

where D is again the derivative operator and $A^\dagger$ denotes the adjoint of A. These give the exponential growth rates of small perturbations applied along different directions (which are not, however, the eigenvectors of Λ, but those of Df). The spectrum of Lyapunov exponents thus gives a fairly detailed picture of the kinds and degrees of sensitivity displayed by the system dynamics.

These definitions take no account of the system's spatial structure; this should be rectified in an application to cellular automata. In addition, we must define the exponents in a way which is spatially *local*, so that their variation over space can be studied and structures identified. In continuous, spatially extended systems the traditional approach has been to consider the global configurations of the system as points in a Hilbert space ([35], p. 53) and then expand the above definition to include infinite-dimensional spaces. Although this produces a meaningful set of global exponents, these exponents contain no information about whether the system is more sensitive at *particular* points in space. Such information is crucial for identifying coherent structure. Also crucial is an ability to identify structures at varying spatial scales. For example the elementary CA rule 110 has particles that can be over 20 cells wide—and all parts of the particle need not be equally sensitive. Some more recent work has attempted to define Lyapunov exponents in a spatially local manner, for precisely these reasons [2].

A second challenge is to define Lyapunov exponents in a

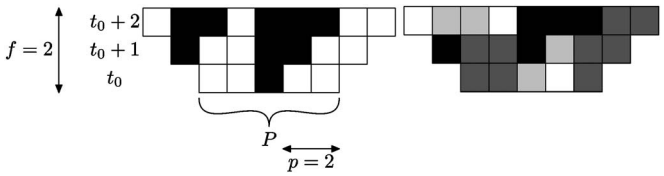

FIG. 3. Illustration of the calculation of local sensitivity in rule 110. Left: unperturbed evolution of a configuration for rule 110. Here 0's are shown as white cells and 1's as black cells. Time progresses upwards. Right: results of apply a perturbation of width $p=2$, in this case the perturbed initial configuration s is 11001. Cells which switched from 0 to 1 under the perturbation are colored dark gray; cells which switched from 1 to 0 under the perturbation are colored light gray. The evolution of the perturbation is tracked forward for f time steps. For this perturbation, the area of the difference plumes (gray cells) is $(1/f)\left(\frac{5}{7}+\frac{4}{9}\right)$. This is then averaged over all radius-p perturbations to get the sensitivity ξ.

way which is applicable to systems which are *discretized*, in both state and space. The classical definition of the Lyapunov exponent invokes limits taken as two initial conditions approach arbitrarily close to one another in continuous-state spaces. (This is implicit in the use of derivatives.) In cellular automata, in which the state is discretized, the closest distinct configurations possible are two configurations differing only on one cell.[2] Space is also discretized: there are no distances intermediate between one and two cells. Several attempts have been made to frame definitions of Lyapunov exponents which accommodate these facts; we discuss these attempts and why we feel they are not suitable for our present purposes in Appendix A. Here we take the straightforward approach of perturbing small patches of contiguous cells, evolving the system forward in time and measuring the distance between the perturbed and unperturbed configurations. This eliminates the need to take derivatives, and by varying the size of the perturbed region, one can examine the structures present at different length scales.

Formally, we define the local sensitivity at point $\vec{r},t_0$ as the result of the following calculation. Let P be the set of all cells at distance at most p from $\vec{r}$. The parameter p is called the *perturbation range*. Let S be the set of all possible configurations restricted to P (i.e., the set of words of length $|P|$ taking the states as alphabet). For one $s \in S$, we set the cells in P to their state in s and keep the original configuration for all other cells. Then we let the automaton evolve for f steps. We name the parameter f the *future depth*. Finally, we measure the area of the difference plumes, the total number of cells which differ from the original configuration, at each time step and compute the mean, weighting the area at time t by $1/|w|$, where w is the set of points at time t that depend on $(\vec{r},t_0)$. (Figure 3 illustrates this calculation for rule 110.) We average this over all $s \in S$ to get the sensitivity at that point, $\xi(\vec{r},t_0)$, which, intuitively, gauges how unstable the system is with respect to perturbations there. Note that ξ is normalized so that it always lies between 0 and 1, facilitating the comparison of distinct rules.

[2]There are metrics, such as the Cantor metric, in which CA configurations can approach arbitrarily close, but they are inappropriate for local studies, such as ours.

Two parameters enter into our calculation of ξ, though we suppress them in the notation: the perturbation range p and the future depth f. Both must be chosen with some care. If they are too small, the calculated ξ is excessively sensitive to fluctuations inside the background domain or within particles, but overly large parameter choices tend to make things blurry and are quite time consuming. Choosing the right p is like adjusting a microscope to the size of the object one wants to observe. If the magnification factor is too high, one might miss the bigger structures or even encounter diffraction artifacts (counterpart of being too close to the discretization scale). It may seem self-defeating that we should need to know the size of the relevant structures to successfully use an automatic filter. However, there may be structures at different scales and varying p allows one to filter these preferentially. Similarly, the coherent structures of many systems can be arranged in a causal hierarchy, where higher-level objects drive lower-level ones, and this can be detected by increasing the future depth f, which tends to pick out the most autonomous features. (See the example of ECA rule 146 below, especially Fig. 7, obtained with $f=30$.) Note that the levels of the spatial and causal hierarchies need not be aligned with one another; i.e., the largest structures need not be the most autonomous.

B. Results from elementary cellular automata

We now present the results of applying the dynamical sensitivity filter to several (1+1)D elementary CA. In Fig. 4 we show the configuration field of rule 110 (a) and the field filtered with respect to the sensitivity (b). Figure 5 does the same for rule 54, another nonconservative elementary rule which tends to produce a background domain of alternating triangles. The filtering automatically distinguishes the autonomous features of the CA, in particular the particles' evolution through time. The autonomy of the particles is reflected by their dark tone. The domains, in contrast, are much less autonomous and appear lighter. This makes intuitive sense. Domains are large, ordered regions of the CA, and their perturbation has little effect upon the dynamics, because any defects are either quickly healed or remain confined. Particles in contrast are relatively complex and localized objects which travel through the domains. Perturbing a particle generally strongly changes the dynamics of the CA, either through its destruction or via a significant alteration of its attributes—location, internal phase, type of particle, etc. [25].

The utility of the local sensitivity filter can be seen by comparing the raw configurations produced by rules 22 and 146 to their filtered fields (Figs. 6 and 7, respectively). The configuration fields of the two rules look very similar, both appearing highly disordered. Once filtered, however, it is clear that they are quite different. Rule 22 is chaotic, and all points have roughly equal, and strong, influence on the future of the system. All of this, along with the absence of autonomous objects, is plain from the sensitivity filter [Fig. 6(b)]. In contrast, rule 146 (Fig. 7) is composed of domains separated by boundaries which wander over time. The domain walls appear as light traces, indicating that they are less autono-

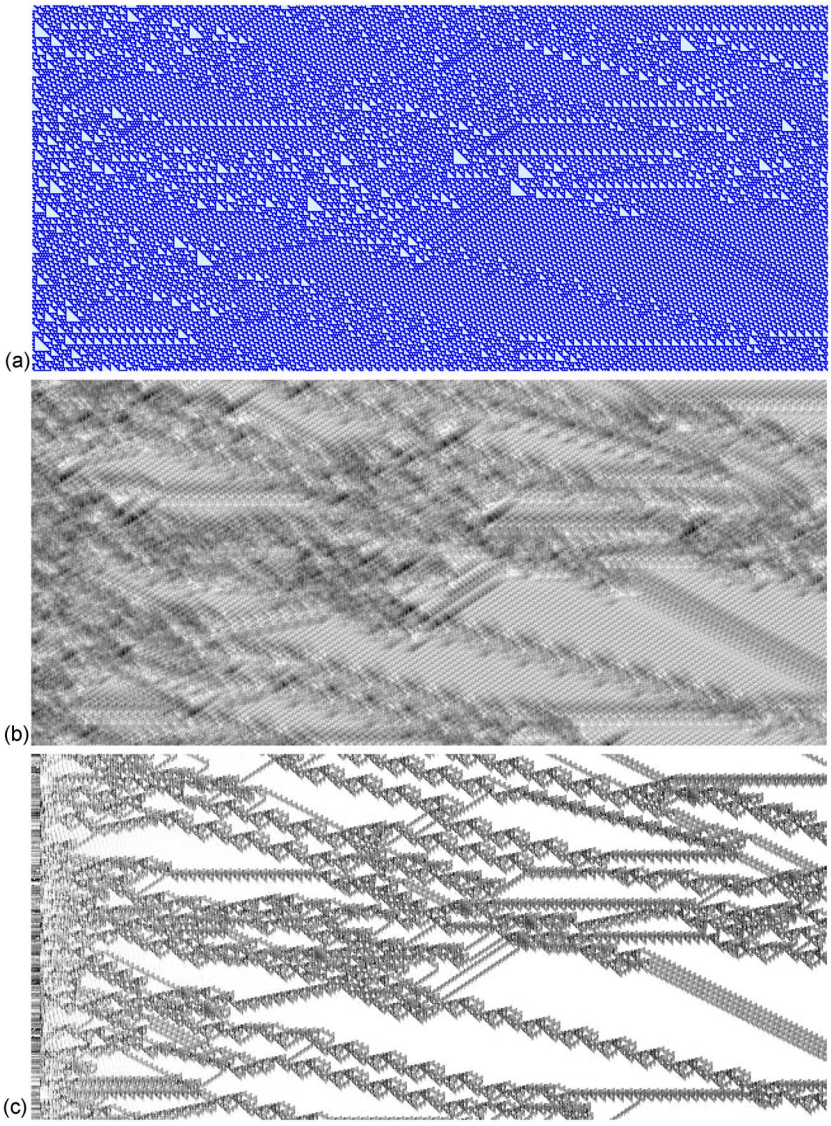

FIG. 4. (Color online) Evolution of rule 110 from a random initial condition (the same as that shown in Fig. 1, repeated for convenience): (a) configuration field, (b) sensitivity field (calculated with $p=1$, $f=10$), and (c) complexity field, using light cones of depth 3. In (b) and (c), higher values of the field are denoted by darker points. In (c), the vertical grey band on the left is due to the fact that we need a past to compute the causal states. There are a lot of particles in the beginning, which makes the background domain rather unusual and complex, but it gradually bleaches.

mous than the domains they separate. The domain walls of rule 146 are dependent objects because their motion is determined by the chaotic dynamics of the domains to either side of them. This is a clear instance of the fact that while the presence of an autonomous object implies the existence of structures, the presence of structures does not necessarily imply autonomy.

The local sensitivity is spatially uniform in linear CA rules [those which are a sum (mod 2) on a subset of the neighborhood]. This is because the Hamming distance between the original and perturbed configurations is independent of the perturbed cell, which in turn is because the difference plume always has the same shape. As expected from this argument, direct calculations of the sensitivity field on additive rules such as ECA 60, 85, 90, 105, 150, and 170 produce perfectly uniform results. (We omit plots for reasons of visual monotony.)

III. LOCAL STATISTICAL COMPLEXITY

Although the sensitivity is very effective at uncovering structures, as well as telling us how stable those structures are, it has a significant drawback. The CA configuration field must be actively perturbed many times and the results compared to the unperturbed case. Such a procedure is computationally intensive and often impossible to apply effectively to experimental data. We now discuss the *local statistical complexity* $C(\vec{r},t)$ [36–38], the calculation of which only requires observations, rather than active and repeated perturbations. We shall see that the statistical complexity has other advantages as well.

A. Local causal states and their complexity

Let $x(\vec{r},t)$ be an $(n+1)$D field, possibly stochastic, in which interactions between different space-time points propagate at speed c. As in [39], define the *past light cone* of the space-time point $(\vec{r},t)$ as all points which could influence $x(\vec{r},t)$—i.e., all points $(\vec{q},u)$ where $u<t$ and $||\vec{q}-\vec{r}|| \leqslant c(t-u)$. The *future light cone* of $(\vec{r},t)$ is the set of all points which could be influenced by what happens at $(\vec{r},t)$. $l^-(\vec{r},t)$ is the configuration of the field in the past light cone and $l^+(\vec{r},t)$ the field in the future light cone. The distribution of future

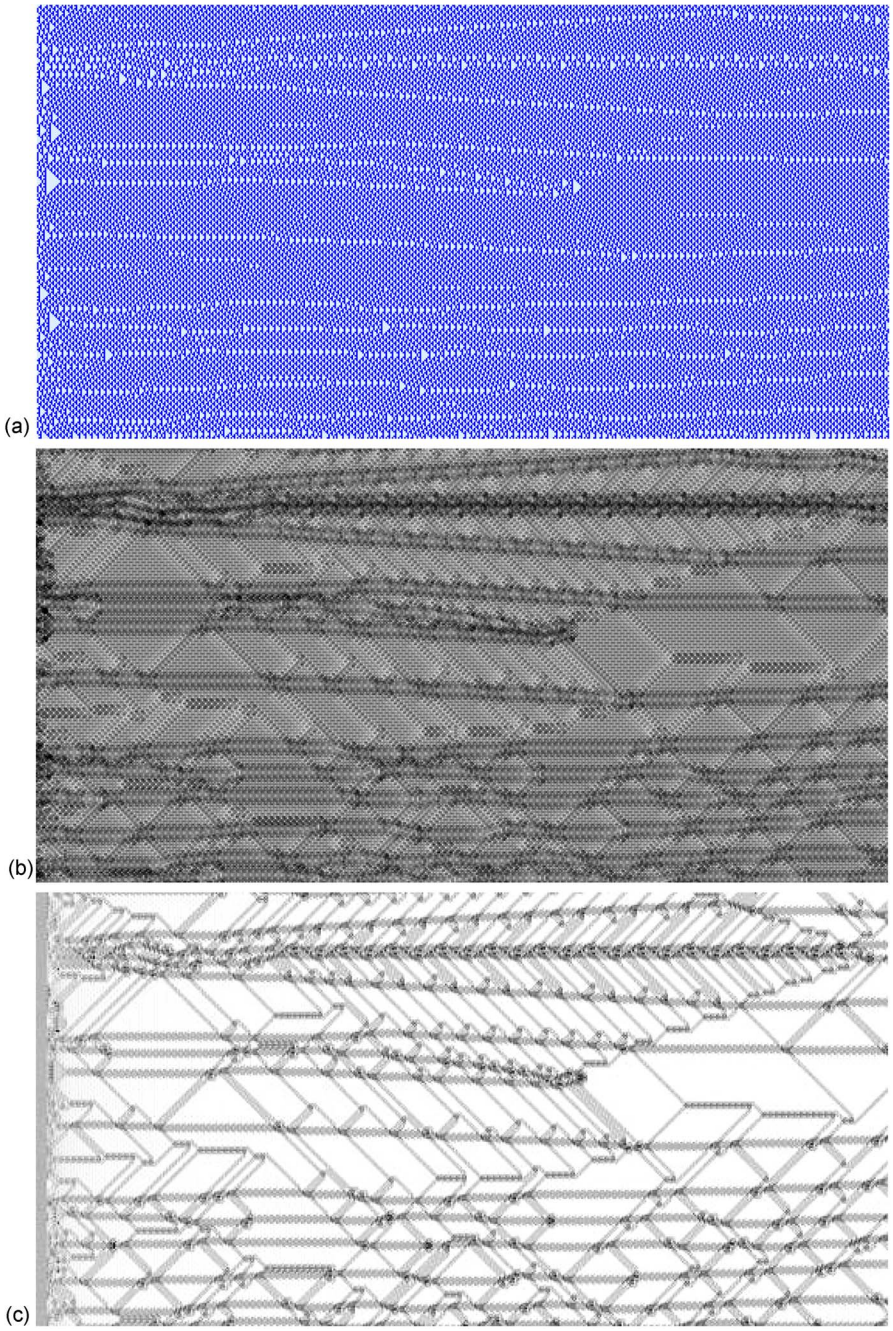

FIG. 5. (Color online) Evolution of rule 54 from a random initial condition: (a) configuration field, (b) local sensitivity ($p=1$, $f=10$), and (c) statistical complexity for rule 54 (light-cone depth of 3). In (c), note that the background domain is light (low complexity), the particles are grey, and most of the collisions are darker (highest complexity). The tiny particles that are merely phase shifts in the periodic background are made clear here, while they are hard to identify by eye in the configuration field (though misalignment between that field and your printer matrix can help).

light cone configurations, given the configuration in the past, is $\mathrm{P}(l^+|l^-)$.

Any function η of l^- defines a *local statistic*. It summarizes the influence of all the space-time points which could affect what happens at $(\vec{r},t)$. Such local statistics should tell us something about "what comes next," which is l^+. Information theory lets us quantify how informative different statistics are and so guides our choice among them.

The information about variable x in variable y is

$$I[x;y] \equiv \left\langle \log_2 \frac{\mathrm{P}(x,y)}{\mathrm{P}(x)\mathrm{P}(y)} \right\rangle, \tag{1}$$

where $\mathrm{P}(x,y)$ is the joint probability, $\mathrm{P}(x)$ is the marginal probability, and $\langle\cdot\rangle$ is the expectation [40]. The information a statistic η conveys about the future is $I[l^+;\eta(l^-)]$. A statistic is *sufficient* if it is as informative as possible [40], here if and only if $I[l^+;\eta(l^-)]=I[l^+;l^-]$. This is the same [40] as requiring that $\mathrm{P}(l^+|\eta(l^-))=\mathrm{P}(l^+|l^-)$. A sufficient statistic retains all the predictive information in the data. Since we want *optimal* prediction, we confine ourselves to sufficient statistics.

If we use a sufficient statistic η for prediction, we must describe or encode it. Since $\eta(l^-)$ is a function of l^-, this encoding takes $I[\eta(l^-);l^-]$ bits. If knowing η_1 lets us compute η_2, which is also sufficient, then η_2 is a more concise summary and $I[\eta_1(l^-);l^-] \geq I[\eta_2(l^-);l^-]$. A *minimal sufficient statistic* [40] can be computed from any other sufficient statistic. In the present context, the minimal sufficient statistic is

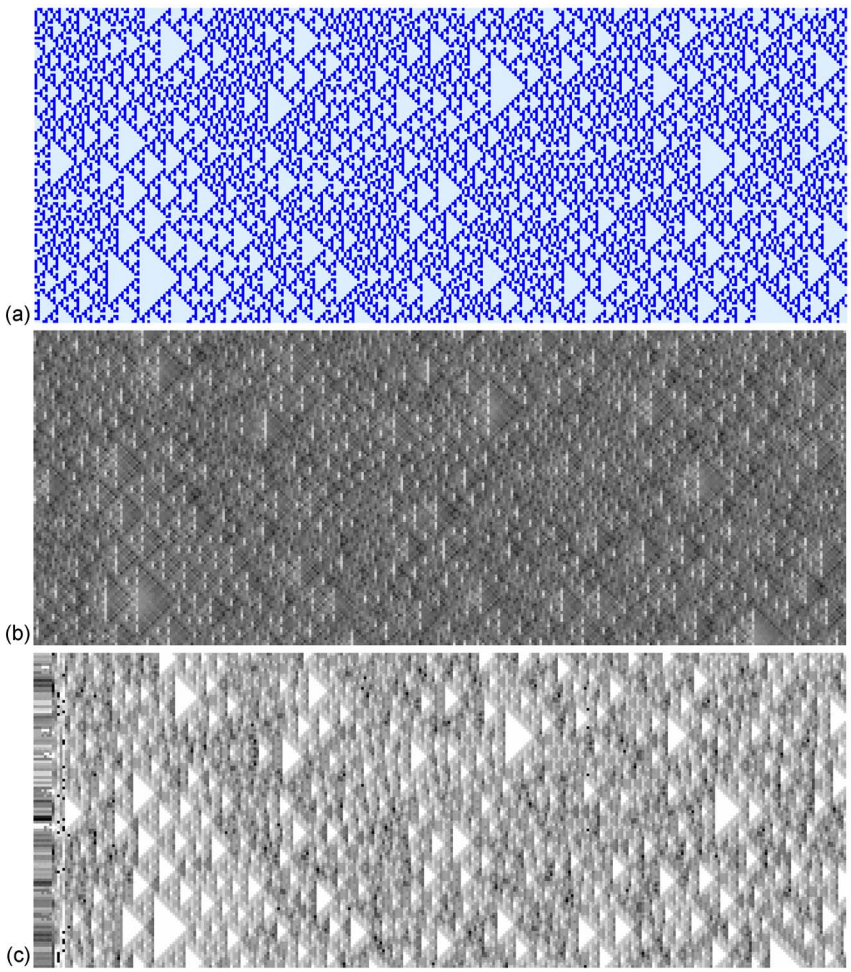

FIG. 6. (Color online) The evolution of rule 22 from a random initial condition: (a) configuration field, (b) local sensitivity (calculated with $p=1$, $f=10$), and (c) statistical complexity. The nearly uniform sensitivity field in (b) reflects the chaotic nature of the rule.

essentially unique (as discussed below) and can be reached through the following construction.

Take two past light-cone configurations l_1^- and l_2^-. Each has some conditional distribution over future light-cone configurations: $\mathrm{P}(l^+|l_1^-)$ and $\mathrm{P}(l^+|l_2^-)$, respectively. The two past configurations are equivalent, $l_1^- \sim l_2^-$, if those conditional distributions are equal. The set of configurations equivalent to l^- is $[l^-]$. Our statistic is the function which maps past configurations to their equivalence classes:

$$\epsilon(l^-) \equiv [l^-] = \{\lambda : \mathrm{P}(l^+|\lambda) = \mathrm{P}(l^+|l^-)\}. \tag{2}$$

Clearly, $\mathrm{P}(l^+|\epsilon(l^-)) = \mathrm{P}(l^+|l^-)$ and so $I[l^+;\epsilon(l^-)] = I[l^+;l^-]$, making ϵ a sufficient statistic. The equivalence classes, the values ϵ can take, are the *causal states* [37,41–43]. Each causal state is a set of specific past light cones, and all the cones it contains are equivalent, predicting the same possible futures with the same probabilities. Thus there is no advantage to subdividing the causal states, which are the coarsest set of predictively sufficient states.

For any sufficient statistic η, $\mathrm{P}(l^+|l^-) = \mathrm{P}(l^+|\eta(l^-))$. So if $\eta(l_1^-) = \eta(l_2^-)$, then $\mathrm{P}(l^+|l_1^-) = \mathrm{P}(l^+|l_2^-)$ and the two pasts belong to the same causal state. Since we can get the causal state from $\eta(l^-)$, we can use the latter to compute $\epsilon(l^-)$. Thus, ϵ is minimal. Moreover, ϵ is the *unique* minimal sufficient statistic ([37], Corollary 3): any other just relabels the same states.

Because ϵ is minimal, $I[\epsilon(l^-);l^-] \leq I[\eta(l^-);l^-]$, for any other sufficient statistic η. Thus we can speak objectively about the minimal amount of information needed to predict the system, which is how much information about the past of the system is relevant to predicting its own dynamics. This quantity, $I[\epsilon(l^-);l^-]$, is a characteristic of the system and not of any particular model. We define the *local statistical complexity* as

$$C(\vec{r},t) \equiv I[\epsilon(l^-(\vec{r},t));l^-(\vec{r},t)]. \tag{3}$$

For a discrete field, C is also equal to $H[\epsilon(l^-)]$, the Shannon entropy of the local causal state.[3] C is the amount of information required to describe the behavior at that point and equals the logarithm of the effective number of causal states—i.e., of different distributions for the future. Complexity lies between disorder and order [41,44,45], and $C=0$ both when the field is completely disordered (all values of x are independent) and completely ordered (x is constant).

[3]The proof is as follows. $I[\epsilon(l^-);l^-] = H[\epsilon(l^-)] - H[\epsilon(l^-)|l^-]$, the amount by which the uncertainty in $\epsilon(l^-)$ is reduced by knowing l^-. But for any discrete-valued function f, $H[f(x)|x]=0$, because a function is certain, given its argument. Hence $I[\epsilon(l^-);l^-] = H[\epsilon(l^-)]$.

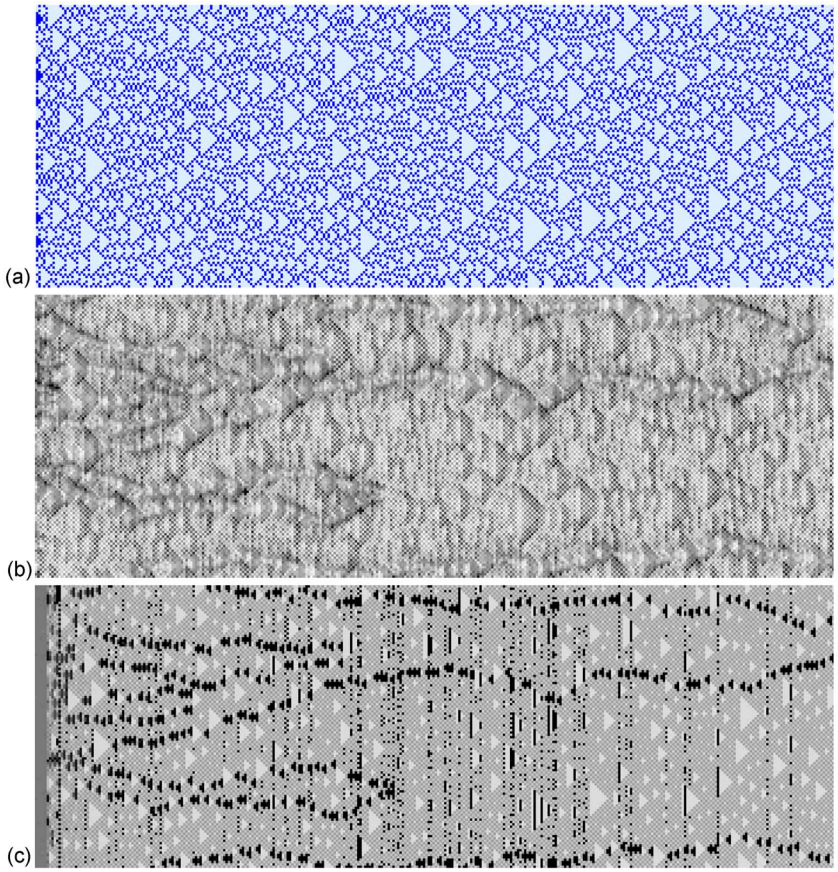

FIG. 7. (Color online) The evolution of rule 146 from a random initial condition: (a) the configuration field, (b) the sensitivity field (calculated using $p=1$ and $f=30$), and (c) the complexity field. Although the configuration field looks very similar to that of rule 22 (Fig. 6), this field is not chaotic and has a domain structure. Local sensitivity filtering (b) reveals the domain walls as light (low sensitivity) traces bounded by darker zones. The domain walls are not autonomous; their behavior is instead determined by what happens inside the domains, hence their low sensitivity. Note also the increase in sensitivity when two walls are near: a small perturbation can lead them to merge (or prevent them from doing so). Statistical complexity filtering (c) reveals the domain walls as dark (high complexity) traces composed of localized triangular regions. Because the domain walls require more predictive information than the domains themselves, the statistical complexity is higher for the walls. The dark vertical lines are finite-size effects and should be ignored

C grows when the field's dynamics become more flexible and intricate, and more information is needed to describe the behavior.

The local complexity field $C(\vec{r},t)$ is simply $-\log \Pr(s(\vec{r},t))$ where $s(\vec{r},t)$ is the local causal state at space-time point $(\vec{r},t)$. The local complexity is the number of bits which would be required to encode the causal state at $\vec{r},t$, if we used the optimal (Shannon) coding scheme. Equivalently, it is the number of bits of information about $l^-(\vec{r},t)$ which are used to determine the causal state.

It is appropriate, at this point, to take a step back and consider what we are doing. Why should we use the light-cone construction, as opposed to any other kind of localized predictor? Indeed, why use localized statistics at all, rather than global methods? Let us answer these in reverse order. The use of local predictors is partly a matter of interest—in studying coherent structures, we care essentially about spatial organization, and so global approaches, which would treat the system's sequence of configurations as one giant time series, simply do not tell us what we want to know. In part, too, the local approach makes a virtue of necessity, because global prediction quickly becomes impractical for systems of any real size. The number of modes required by methods attempting global prediction, like Karhunen-Loeve decomposition, grows extensively with system volume [46,47]. Global methods thus provide no *advantage* in terms of compression or accuracy.

The use of light cones for the local predictors, first suggested by [39],[4] rather than some other shape, is motivated partly by physical considerations and partly the nice formal features which follow from the shape, of which we will mention three [37].

(i) The light-cone causal states, while local statistics, do not lose any global predictive power. To be precise, if we specify the causal state at each point in a spatial region, that array of states is itself a sufficient statistic for the future configuration of the region, even if the region is the entire lattice.

(ii) The light-cone states can be found by a recursive filter. To illustrate what this means, consider two space-time points $(\vec{r},t)$ and $(\vec{q},u)$, $u \geqslant t$. The state at each point is determined by the configuration in its past light cone: $s(\vec{r},t) = \epsilon(l^-(\vec{r},t))$, $s(\vec{q},u) = \epsilon(l^-(\vec{q},u))$. The recursive-filtration property means that we can construct a function which will give us $s(\vec{q},u)$ as a function of $s(\vec{r},t)$, plus the part of the past light cone of $(\vec{q},u)$ that is not visible from $(\vec{r},t)$. Not only does this greatly simplify state estimation, it opens up powerful connections to the theory of two-dimensional automata [32].

(iii) The local causal states form a Markov random field, once again allowing very powerful analytical techniques to

[4]Kolmogorov seems to have been the first to use light cones to analyze spatial stochastic processes, calling them "causal sets" in a model of crystallization [48]—see [49].

```
U ← list of all pasts in random order
Move the first past in U to a new state
for each past in U
    noMatch ← TRUE
    state ← first state on the list of states
    while (noMatch and more states to check)
          noMatch ← (Significant difference between
                     past and state?)
          if (noMatch)
            state ← next state on the list
          else
              Move past from U to state
              noMatch ← FALSE
    if (noMatch)
      make a new state and move past into it from U
```

FIG. 8. Algorithm for grouping past light cones into estimated states.

be employed which would not otherwise be available [50].

In general, if we used some other shape than the light cones, we would not retain any of these properties.

B. A Reconstruction algorithm

We now sketch an algorithm to recover the causal states from data and so estimate C. (See Fig. 8 for the pseudocode, Ref. [37] for details, and cf. [39].) At each time t, list the observed past and future light-cone configurations and put the observed past configurations in some arbitrary order, $\{l_i^-\}$. (In practice, we must limit how far light cones extend into the past or future.) For each past configuration l_i^-, estimate $\mathrm{P}_t(l^+|l_i^-)$. We want to estimate the states, which ideally are groups of past cones with the same conditional distribution over future cone configurations. Not knowing the conditional distributions *a priori*, we must estimate them from data, and with finitely many samples, such estimates always have some error. Thus, we approximate the true causal states by clusters of past light cones with similar distributions over future light cones; the conditional distribution for a cluster is the weighted mean of those of its constituent past cones. Start by assigning the first past l_1^- to the first cluster. Thereafter, for each l_i^-, go down the list of existing clusters and check whether $\mathrm{P}_t(l^+|l_i^-)$ differs significantly from each cluster's distribution, as determined by a fixed-size χ^2 test. (We used $\alpha=0.05$ in all our calculations.) If the discrepancy is insignificant, add l_i^- to the first matching cluster, updating the latter's distribution. Make a new cluster if l_i^- does not match any existing cluster. Continue until every l_i^- is assigned to some cluster. The clusters are then the estimated causal states at time t. Finally, obtain the probabilities of the different causal states from the empirical probabilities of their constituent past configurations and calculate $C(\vec{r}t)$. This procedure converges on the correct causal states as it gets more data, independent of the order of presentation of the past light cones, the ordering of the clusters, or the size α of the significance test [37]. For finite data, the order of presentation matters, but we finesse this by randomizing the order.

C. Results on elementary cellular automata

In Sec. II B, we demonstrated the ability of local sensitivity filtering to find coherent structures in elementary cellular automata. Here we apply local statistical complexity filtering to the same cellular automata and compare the results of the two procedures.

Figures 4(c) and 5(c) show the local complexity fields $C(\vec{r},t)$ of two of the classic rules 110 and 54, respectively. Particles stand out clearly and distinctly on clean backgrounds. Note that we achieve this result by applying the same filter to both systems, even though their background domains and particles are completely different. Were one to use, say, the conventional regular-language filter constructed in [31] for rule 54 on rule 110, it would produce nonsense. (Reference [25] discusses the domains, particles, and conventional filters of both these rules.) Note that the filter in [31] is handcrafted, based on a detailed understanding of the CA's dynamics, whereas our filter, as we have said, is completely automatic and requires no human intervention. One might expect this generality to be paid for in a loss of resolving power or missing system-specific features, but this does not appear to be the case.[5] For instance, the regular-language analysis of rule 54 [31] identifies a subtle kind of particle, which consists of a phase shift in the spatially periodic background domain. These phase shifts are hard to identify by eye, but show up very cleanly as particles in Fig. 5(c).

For completeness and further comparison with the local sensitivity, we include rules 22 and 146 filtered with respect with statistical complexity in Figs. 6(c) and 7(c). Statistical complexity is also able to distinguish between the two CA configuration fields, even though the difference is not readily apparent to the naked eye.

Experimentally, we find that the depth of the past light cone is more relevant to proper filtering than the depth of the future light cone. (This may be related to the recursive estimation property of the causal states.) On elementary cellular automata, depth 2 is often sufficient, in the sense that further extensions of the cones do not change the states identified, but the results presented here use depth 3 for both future and past light cones. (Rule 41, not shown for reasons of space, required a past depth of 5 in order to reach convergence.)

While both our techniques make it easy to identify objects like particles, even when they were previously hard to detect, they are very different filters, not just in their definition but also in their results. This can immediately be seen from the correlation coefficients[6] of the sensitivity and complexity fields—for rule 110, for instance, the correlation is a negligible 0.014.

More abstractly, statistical complexity is a local quantity that is calculated using a global object: namely, the probability distribution over causal states. Its accurate estimation thus needs a quite large number of cells, since, lacking an analyti-

[5]This statement must be qualified by a recognition that we must supply the filter with enough data for it to find the right states. The learning rate of the state reconstruction algorithm is an important topic, beyond the scope of this paper.

[6]The correlation coefficient of statistics, $\rho_{xy}=(\langle XY\rangle-\langle X\rangle\langle Y\rangle)/\sigma_X\sigma_Y$, is a dimensionless counterpart to the correlation function of statistical physics, $C_{XY}=\langle XY\rangle-\langle X\rangle\langle Y\rangle$, normalized so that its value lies between -1 and $+1$ and is zero when X and Y are linearly unrelated.

cal form for that distribution, the latter must itself be estimated. A further consequence is that the separation between particles and other structures and the background tends to become cleaner over time, as the domains grow and the density of particles decays, suppressing the complexity of the former and raising that of the later. [This may be clearly seen in Fig. 4(c).] Identification of structures from the complexity field is thus best undertaken after a (hopefully short) transient regime has passed, during which the local sensitivity filter may be more useful. It is not, of course, necessary for the system to have reached a stationary regime in order to use the local complexity filter.

IV. RESULTS ON CYCLIC CELLULAR AUTOMATA

In this section we compare and contrast local sensitivity and statistical complexity with a more traditional order parameter approach in a spiral-forming CA model of an excitable medium. By applying all three analyses to the same (2+1)D CA, we will clarify the specific structural details emphasized by each method and bring out the advantages of automatic filtering over approaches where the details must be put in by hand.

Our model system in this section is the cyclic cellular automata on a square lattice, which have been studied in considerable detail in the literature on spatial stochastic processes [51,52]. In the general case, there are κ colors, numbered from 0 to $\kappa-1$. A cell of color k changes its color only if at least T of its neighbors are of color $k+1$ mod κ, in which case it assumes that color. In this paper, we confine ourselves to the special case of range 1 neighborhoods, $\kappa=4$, and $T=2$. The behavior of the system, started from uniform random initial conditions, is illustrated in Fig. 2. In [38], we demonstrated the ability of statistical complexity to quantify the extent to which the CA self-organizes and the effect of changing parameter values on the degree of self-organization. Here, however, we are more interested in the patterns formed than in whether significant pattern formation is taking place, so we deal only with the most strongly self-organizing variant.

A. Order parameter and effective free energy

While, as remarked, cyclic CA have been extensively studied as models of excitable media, they have not (to the best of our knowledge) hitherto been examined with the tools of the order parameter approach. To better highlight the distinctive characteristics of the automatic filtering methods we propose, we first identify an order parameter and effective free energy for the present version of CCA. Our free energy will be a functional of the configuration field alone and is most appropriate for the long-run stationary distribution, rather than the initial transient stages, when the system is far from statistical equilibrium. Accordingly, we suppress time as an argument to the configuration field in what follows.

As noted, with these parameter settings, the CA configuration field forms rotating spiral waves, which grow to engulf the entire lattice, with disordered domain walls at the boundaries between competing spiral cores. (See Fig. 9.) The

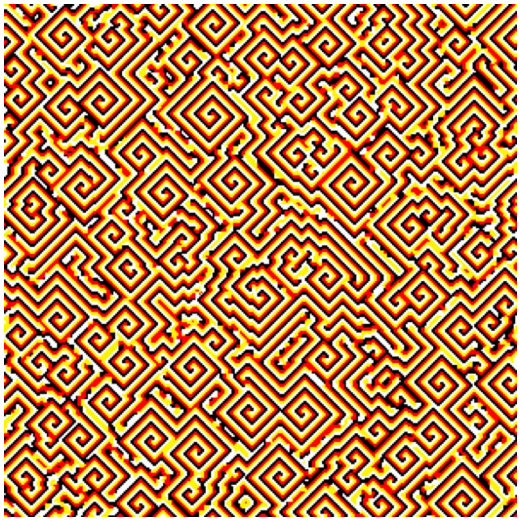

FIG. 9. (Color online) A typical configuration of the spiral-forming CCA cyclic cellular automaton in the asymptotic regime ($t=200$), illustrating the presence of coherent structures (rotating spiral waves) dominating the lattice.

mechanisms by which spirals form and grow are fairly well understood, and topological arguments [52] pick out the key role of both the conservation of winding number and of the spiral cores in this process. Accordingly we consider CCA as a kind of discretized XY model, similar to a clock model ([4], Sec. 3.6.3),[7] and, as in such models generally, the appropriate order parameter has two components [4]. Here the crucial observation is that the ground state consists of plane waves, with stripes of cells of constant color extending perpendicular to the direction of propagation of the wave. The order parameter is the local normalized wave vector

$$\boldsymbol{\Psi}(x,y) = \bar{\Psi}_x(x,y)\hat{x} + \bar{\Psi}_y(x,y)\hat{y},$$

defined such that it takes one of four values $\boldsymbol{\Psi}=(1/\sqrt{2})(\pm 1,\pm 1)$ in each of the four domains surrounding a spiral core. The exact definition of the wave vector in terms of the states of the CCA configuration field is given in Appendix B. This wave vector can be used to define a phase for the spiral wave at each lattice site:

$$\Phi(x,y) = \tan^{-1}\frac{\bar{\Psi}_y(x,y)}{\bar{\Psi}_x(x,y)}.$$

The local free energy is then calculated as the discretized version of

$$F(x,y) = [\nabla\Phi(x,y)]^2. \tag{4}$$

We show the free energy of the CA configuration field in Fig. 10. Note the increased free energy caused by topological defects such as the domain walls and spiral cores.

[7] A class of cellular automata very similar to CCA are treated as antiferromagnetic Potts models in [53]. We experimented with such an order parameter, but the results were poor.

FIG. 10. Free energy per site of the CCA configuration field given in Fig. 9. Darker cells have higher energy values. Note the increased free energy at the topological defects—e.g., the spiral cores and domain walls.

B. Local sensitivity and statistical complexity for CCA

Filtering by sensitivity and complexity, which we demonstrated above for elementary cellular automata, works without modification on cyclic CA. As might be expected, the two filters reveal different, but compatible, aspects of the system.

Figure 11 shows the local sensitivity field for the same configuration as in Fig. 9, with a one-cell perturbation range ($p=0$). Figure 12, on a smaller field, shows that with $p=1$ we get qualitatively similar results. The spiral cores are easily spotted in both figures, demonstrating that they are autonomous objects. There is no clear difference between spiral angles (the horizontal and vertical lines) and spiral domains. Spiral boundaries are white (lowest sensitivity), because any perturbation here will be quickly erased under the pressure of

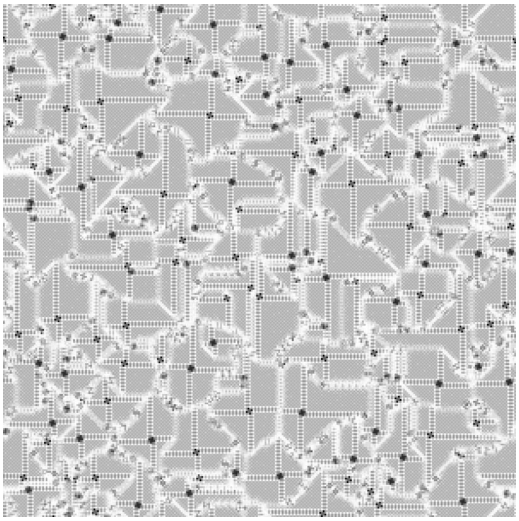

FIG. 11. Local sensitivity ($p=0$, $f=5$) for the cyclic cellular automaton, using the CCA configuration field of Fig. 9

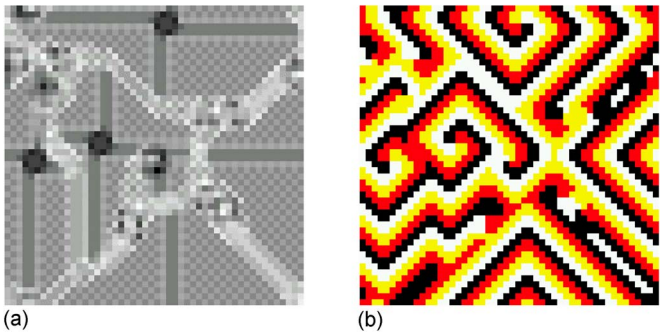

FIG. 12. (Color online) Local sensitivity (a) for the cyclic cellular automaton, calculated with $p=1$, $f=1$, and corresponding configuration (b).

the radiating spiral cores (cf. [54]). Calculating local sensitivity for (2+1)D automata is very slow: each cell has 8 neighbors and there are 4 states, which makes 4^9-1 possible perturbations. While sampling a random subset of all possible perturbations would be faster, the resulting approximation error would have to be carefully determined.

We show the CCA field filtered with respect to statistical complexity in Fig. 13. Comparison with Fig. 10 shows that causal states and order parameter analysis yield the same information. The spiral cores are still among the most complex areas. Here spiral boundaries are also complex: under statistical complexity filtering, this is because some prediction is possible, but requires much information; under free energy analysis, this is due to high phase differences.

C. Comparison of free energy, sensitivity, and complexity

The three methods described here—the free energy, the local sensitivity, and the statistical complexity—all uncover

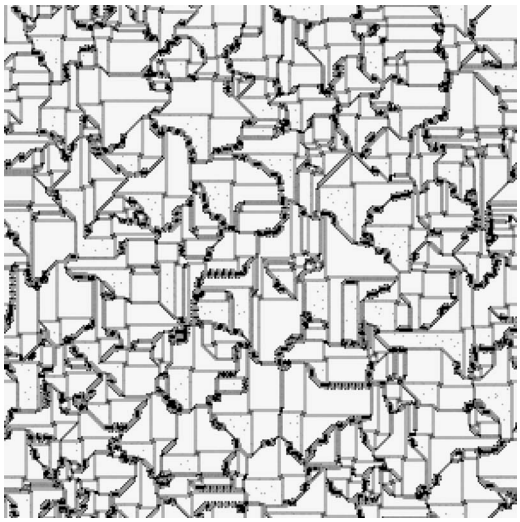

FIG. 13. Local statistical complexity of the CCA configuration field in Fig. 9, calculated with depth-1 light cones. Darker cells have higher values for the field. This figure should be compared to both Figs. 10 and 11. Note that free energy and statistical complexity are strongly correlated, while local sensitivity and statistical complexity are not because they emphasize different aspects of the coherent structure.

significant structures in the spatiotemporal dynamics of the CA. However, the methods are not equivalent and emphasize different aspects of those structures, as one might expect of different filters.

Qualitatively, we can see the difference in the order in which different features are ranked. For sensitivity, the ordering is particles (spiral cores), domains, and then domain walls last. For both complexity and free energy, the order is by contrast domain walls, particles, and domains. A quantitative calculation of the correlation coefficients between the different filtered fields confirms this qualitative impression. The most strongly correlated—as one might expect by comparing Fig. 10 with Fig. 13—are the statistical complexity and the free energy, $\rho=0.690$; we discuss the roots of this strong correlation below. The complexity and the sensitivity, however, are almost completely uncorrelated, just as we observed in the case of rule 110—the correlation coefficient calculated from sample data, $\rho=-0.006$, is not materially different from 0. Unsurprisingly, there is also no important correlation between the sensitivity and free energy ($\rho=-0.008$). Despite the fact that sensitivity and complexity are nearly orthogonal, plotting complexity as a function of sensitivity (not shown) reveals an interesting relationship: as sensitivity increases, the *minimum* value of complexity rises, though the converse is not true (minimally sensitive points are found at all values of complexity).

Sensitivity looks for regions that are unstable to perturbation; because of this instability, prediction requires very precise discriminations among histories. Moreover, localized unstable regions are presumably rare and the corresponding states uncommon. (Recall that the complexity is both $H[S]$ and $I[S;X^-]$.) Thus the spiral cores are rare, sensitive, and complex. On the other hand, uncommon, complex structures need not be unstable. Domain walls, for instance, are comparatively rare and require considerable historical information for their prediction, essentially because they are regions where the evolution of the phase is hard to determine; this in turns means that their causal states need more bits to specify. They are, however, very stable, because they have considerable spatial extent, and to destroy or move a wall implies changing the domains on either side. The domains themselves, while only minimally complex, are more sensitive than the domain walls, because perturbations there can create *localized* defects.

The relationship between free energy and complexity, while strong and at first sight clear, is actually a bit tricky. The easy, but flawed, argument for why effective free energy and complexity should be correlated runs as follows. The effective free energy relies on the assumption that the microscopic dynamics tend to create ordered regions, where "ordered" is understood in the sense given by the order parameter. All departures from the ordered state are unlikely, and large departures are exponentially unlikely. (This is not tautologically true, but when it is, we can define an effective free energy.) These departures take the form of topological defects, like spiral cores and domain walls, which pay an energetic penalty. The exact size of this penalty and the resulting rarity of these features will depend on the details of the microscopic dynamics. But precisely because they are rare, the causal states corresponding to them should be rare too and so have high local complexity.

The flaw is that the causal state at a point is a function of the configuration in its *past* light cone, while the order parameter and the free energy are functions of the *immediate* neighborhood configuration. Since the two filters use different inputs, their strong correlation is not trivial. However, under fairly general conditions, the local causal states form a Markov random field [37] and the Gibbs-Markov theorem tells us that a random field is Markovian if and only if there is an additive effective free energy which gives the probability of configurations [55]. Conversely, it can be argued that a complete set of thermodynamic macrovariables should be equivalent to the causal state of the system and should be Markovian in time, so that they can be calculated entirely from the present configuration [43]. Thus, there should, in fact, be a relationship between the complexity and the free energy *if* we have identified the proper order parameter; the strong correlation we find between the two fields suggests that we have done so. The fact that the correlation, while strong, is not perfect could be either due to our definition of the order parameter being slightly off or due to finite-sample errors in the identification of the causal states.[8]

V. CONCLUSIONS

We have introduced two complementary filtering methods for spatially extended dynamical systems. Local sensitivity calculates the degree to which local perturbations alter the system and picks out autonomous features. Local statistical complexity calculates the amount of historical information required for optimal prediction and identifies the most highly organized features. We emphasize that "organized" is not equivalent to ordered, because the latter corresponds to low entropy which is not the same as high complexity or organization [38]. A regular lattice is highly ordered, but not very organized. On the other hand, high complexity is not equivalent to high entropy. A random field has high entropy but low complexity. Complexity, as we said earlier, lies between order and randomness and is a reflection of the probabilistic properties of the system. In contrast, sensitivity is a measure of the system's dynamical properties. The two are linked through ergodic theory, but they remain distinct. Both sensitivity and complexity pick out spatiotemporal coherent structures, and the structures they identify match those known from previous, more *ad hoc* approaches, whether based on regular languages (as in ECA) or order parameters and topological considerations (as in CCA). In no case, however, is prior knowledge about the system or its coherent structures used in constructing our filters; at most, we have tuned calculational parameters in a way akin to adjusting a microscope until the image comes into focus. Ideally, one would use both sensitivity and complexity filtering, because they provide distinct kinds of information about the system, but

[8] Smith [64] has pointed out that this argument suggests a relationship between the time depth needed to identify the causal states and the spatial range needed to calculate the free energy. However, pinning down that relationship would require careful treatment of many mathematical issues regarding sofic shifts, long-range correlations in Markovian fields, etc., beyond the scope of this paper.

we suspect complexity will be much easier to calculate from empirical data since it only requires observation and not perturbation.

Many theoretical and mathematical questions present themselves about both filtering methods. We have mentioned some in passing; here, we wish to highlight just a few. On the local sensitivity method: What is the exact relationship between the perturbation range p and the spatial scale of identified structures? How much error would be introduced by averaging over a random subset of perturbations, rather than an exhaustive enumeration? Can one identify a typical lifespan for a perturbation, after which it is erased by its surroundings, and if so is this lifespan related to f? (This last is presumably related to dynamical mixing properties.) On the local complexity method: What quantitative factors relate the volume of data available to the error in our estimates of the causal states and so of the complexity? Can we use causal states to give algebraic and automata-theoretic definitions of "domain" and "particle," like those in the 1D case, without entangling ourselves in the difficulties of higher-dimensional languages? (Cf. [36], Sec. 10.5.) When we do have partial knowledge of the correct pattern basis, can we use this to hasten the identification of the local causal states? Perhaps most ambitiously, could one reverse engineer a good order parameter from the local causal state field and its transition structure?

As very large, high-dimensional data sets become increasingly common and important in science, human perception will become increasingly inadequate to the task of identifying appropriate patterns [56,57]. It is desirable, therefore, to move towards more automatic filtering methods and automatic ways of detecting coherent objects. Because our filtering methods do not presuppose any prior knowledge or require human insight about what the right structures are, they should work generically, across systems of highly varying nature and dynamics. This is a hypothesis rather than a theorem, but it can be tested simply by applying our filtering methods to a wide range of systems with known coherent structures—and to others where the appropriate structures are *not* known. In the latter cases, the test of our methods will be whether the structures they identify can be used to frame interesting and insightful higher-level models about the dynamics and functions of the systems involved (cf. [43,58]).

ACKNOWLEDGMENTS

Thanks to Michel Morvan for valuable discussions and for arranging J. B. R.'s visit to Ann Arbor; to Vincenzo Capasso for Refs. [48,49]; to Scott Page and Anna Amirdjanova for their generous construals of "diversity" and "filtering," respectively; and to Dave Albers, Sandra Chapman, Tracy Conrad, Chris Genovese, David Griffeath, J. Hines, Alfred Hübler, Jürgen Jost, Kara Kedi, Eric Mjolsness, Eric Smith, Padhraic Smyth, Peter Stadler, Naftali Tishby, Dowman Varn, and Nicholas Watkins for comments, discussion, and advice. C.R.S. is supported by a grant from the James S. McDonnell Foundation. R.H. is supported by NIH Grant Nos. R01 DA015644, R01 MH59733 and the Jenny Fund at the Dept. of Anesthesia and Critical Care, Massachusetts General Hospital. J. -B. R. is supported by the ENS-Lyon training program. K. L. K. is supported by a grant from the NSA. C.M is supported by NSF Grant No. PHY-0200909.

APPENDIX A: ATTEMPTS TO ADAPT LYAPUNOV EXPONENTS TO CELLULAR AUTOMATA

There have been at least three attempts to define Lyapunov exponents for cellular automata. We review them in order of priority, concluding that none of them is altogether suited to the aims of this paper. All are based on measuring the rate at which small perturbations cause a divergence from the original trajectory.

Shereshevsky [59] defines a Lyapunov exponent for one-dimensional cellular automata as the velocity of propagation of the edge of the difference plume. More exactly, he defines two Lyapunov exponents for the velocities of the left and right edges. By assuming that an invariant measure over configurations exists and has certain properties, he is able to show that these velocities have reasonable long-time limits. (Tisseur [60] shows that the limit exists with somewhat weaker conditions on the invariant measure.) To show that the limit is uniform over the lattice, he invokes a further spatial ergodicity property.

We, of course, would like things *not* to be uniform, so the fact that we do not have measures which are so nicely ergodic and stationary should not trouble us. But this definition allows a cell which only makes a difference to one other cell in the future to count as highly influential, provided that said cell moves very rapidly. (Consider the shift rule 170 "copy your neighbor to your left.") Looking at the area of the difference plume (i.e., the number of differing cells between the original configuration and the perturbed configuration) seems far more reasonable.

Bagnoli *et al.* [61] come closest to the way we define local sensitivity in Sec. II by looking at the total number of "defects," here meaning the number of cells which are different between the original and perturbed configurations. They proceed as follows. They perturb a single site of the lattice and iterate forward one time step, so that the the difference plume now embraces m sites. They then create m copies of the lattice, each identical to the time-evolved unperturbed configuration, except at one of the m sites in the difference plume. They then repeat this procedure, accumulating more and more copies of the lattice, each of which differs from the unperturbed trajectory only at a single site. Suppose the initial perturbation was applied at cell $\vec{x}_0$ and time t_0. Then for each site $\vec{x}$ and time $t \geqslant t_0$, the number of copies of the system which have a defect at $\vec{x}$ at time t is $N_{\vec{x}_0,t_0}(\vec{x},t)$ and $N_{\vec{x}_0,t_0}(t)=\Sigma_x N(\vec{x},t)$. $N_{\vec{x}_0,t_0}(t)$ can grow exponentially, and Bagnoli *et al.* define the local Lyapunov exponent to be that exponential growth rate.

This is far from anything that could be called a Lyapunov exponent in the strict sense of the term. Furthermore, to avoid the extremely involved (exponential) calculation which their definition implies, they make use of a kind of derivative, as in the continuous-system definition of the Lyapunov exponent. But these derivatives are only defined for Boolean-valued rules, which makes them useless for, e.g., cyclic cellular automata. The direct calculation could be

somewhat simplified for deterministic cellular automata by means of dynamic programming, but it would still be a hard calculation to obtain a number whose physical significance is unclear.

Finally, Urías *et al.* [62] define a quantity which is a Lyapunov exponent (and does not require binary values), but only under a very specific metric, with no clear physical interpretation; it is not clear how, if at all, their metric could be extended to higher-dimensional cellular automata. The ultimate result they get is that the Lyapunov exponent is just the maximum velocity at which the envelope of the difference region spreads—taking the maximum over all possible semi-infinite (not single-point) perturbations.

In this paper we wished not to measure a single Lyapunov exponent for the entire system, but rather the local (in both space and time) effects of perturbations. This allowed us to determine the degree to which different local structures of the CA are autonomous.

APPENDIX B: ORDER PARAMETER AND FREE ENERGY FOR SPIRAL-FORMING CYCLIC CELLULAR AUTOMATA

A classic method of describing the equilibrium ordered phases of a system is to define an appropriate order parameter for each phase and an associated free energy determined from symmetry considerations [3,4]. In this Appendix we will perform this analysis for the spiral cellular automaton in its equilibrium state. By "equilibrium" we mean the cellular automaton's long-time behavior after the domain walls and spiral cores have stabilized. The transient period of the cellular automaton corresponds to the phase transition which is not described by the order parameter formalism. Examination of the cellular automaton field (e.g., Fig. 9) reveals that, while it certain possesses order, it is also highly frustrated and exhibits many topological defects, most notably the vortices at the spiral centers and the domain walls between adjacent spirals. However the order parameter is defined, we expect the free energy to increase at these defects.

We begin by defining an appropriate order parameter on the discrete lattice. We take our cue from the fact that for a continuous field, the spirals would consist of concentric rings possessing two-dimensional rotational symmetry. We therefore construct a discretized *XY* model order parameter and free energy. In the *XY* model the Ginsburg-Landau free energy density at a given lattice site is

$$F(x,y) = \alpha[\nabla\Phi(x,y)]^2, \tag{B1}$$

where α is a positive constant which we arbitrarily set to 1 and henceforth ignore. Φ is a localized phase defined at each lattice site. In practice we calculate the Laplacian using each lattice site's nearest neighbors:

$$\begin{aligned} F(x,y) = \frac{1}{4}\{&[\Psi(x+1,y) - \Psi(x,y)]^2 \\ &+ [\Psi(x,y) - \Psi(x-1,y)]^2 \\ &+ [\Psi(x,y+1) - \Psi(x,y)]^2 \\ &+ [\Psi(x,y) - \Psi(x,y-1)]^2\}. \end{aligned} \tag{B2}$$

Φ is defined in terms of the states of the configuration field $\sigma \in \{0,1,2,3\}$. Noting that the spiral cellular automaton field consists mainly of plane waves cycling through the four colors and traveling mainly along the diagonals, although at times along the x and y axes, it seems reasonable to define the spiral cellular automaton phase Φ as a function of the *direction* in which the wave is traveling. We define a normalized two-component wave vector at each lattice site (x,y),

$$\mathbf{\Psi}(x,y) = \frac{1}{\sqrt{\Psi_x^2 + \Psi_y^2}}[\Psi_x(x,y)\hat{x} + \Psi_y(x,y)\hat{y}] = (\bar{\Psi}_x, \bar{\Psi}_y), \tag{B3}$$

where $\bar{\Psi}_x$ and $\bar{\Psi}_y$ are the components of the normalized vector and the unnormalized components Ψ_x and Ψ_y are defined as follows:

$$\begin{aligned} \Psi_x(x,y) = &\sum_{i=\pm 1} g(\sigma(x+i,y), \sigma(x,y)) \\ &+ \frac{1}{2}\sum_{i=\pm 1, j=\pm 1} g(\sigma(x+i,y+j), \sigma(x,y)), \end{aligned} \tag{B4}$$

$$\begin{aligned} \Psi_y(x,y) = &\sum_{i=\pm 1} g(\sigma(x,y+i), \sigma(x,y)) \\ &+ \frac{1}{2}\sum_{i=\pm 1, j=\pm 1} g(\sigma(x+i,y+j), \sigma(x,y)). \end{aligned} \tag{B5}$$

Note that the first sum is over nearest neighbors in the x (y) direction and the second sum is over all next nearest neighbors. The factor $1/2$ comes from the corresponding $1/2$ in the definition of the order parameter. The function g is defined as

$$\begin{aligned} g(\sigma_1, \sigma_2) &= +1 \text{ if } \sigma_1 = \sigma_2 + 1 \\ &= -1 \text{ if } \sigma_1 = \sigma_2 - 1 \\ &= 0 \text{ otherwise}. \end{aligned} \tag{B6}$$

It is to be understood that the above definition of g involves addition and subtraction modulo 4.

The phase Φ at a given lattice site is defined simply as

$$\Phi = \tan^{-1}\frac{\bar{\Psi}_y}{\bar{\Psi}_x}. \tag{B7}$$

Thus we see that in the upper right quadrant of a spiral, $\Psi = (1/\sqrt{2}, 1/\sqrt{2})$ and $\Phi = \pi/4$, whereas in the upper left quadrant $\Psi = (1/\sqrt{2}, -1/\sqrt{2})$ and $\Phi = 3\pi/4$. Thus there is a phase gradient along the boundary between these two quadrants of the spiral and an increase of free energy here. This can be seen in Fig. 10. (The localization of the free energy increase to the boundary between quadrants is, of course, an artifact of the discretized spatial lattice. In the continuum limit, the gradient energy would be distributed evenly around the central vortex.) The other locations where one expects to see increased free energy are in the vortex cores and at the boundaries between spirals. Our free energy captures both of these effects (Fig. 10).

[1] P. Ball, *The Self-Made Tapestry: Pattern Formation in Nature* (Oxford University Press, Oxford, 1999).

[2] M. C. Cross and P. Hohenberg, Rev. Mod. Phys. **65**, 851 (1993).

[3] M. H. Krieger, *Doing Physics: How Physicists Take Hold of the World* (Indiana University Press, Bloomington, 1992).

[4] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, England, 1995).

[5] C. Moore, M. G. Nordahl, N. Minar, and C. R. Shalizi, Phys. Rev. E **60**, 5344 (1999).

[6] P.-G. de Gennes and J. Prost, *The Physics of Liquid Crystals*, 2nd ed. (Clarendon Press, Oxford, 1993).

[7] S. A. Kivelson, I. P. Bindloss, E. Fradkin, V. Oganesyan, J. M. Tranquada, A. Kapitulnik, and C. Howald, Rev. Mod. Phys. **75**, 1201 (2003).

[8] A. T. Winfree, *The Geometry of Biological Time* (Springer-Verlag, Berlin, 1980).

[9] S. Camazine, J.-L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau, *Self-Organization in Biological Systems* (Princeton University Press, Princeton, NJ, 2001).

[10] M. Fujita, P. Krugman, and A. J. Venables, *The Spatial Economy: Cities, Regions, and International Trade* (MIT Press, Cambridge, MA, 1999).

[11] K. Steiglitz, I. Kamal, and A. Watson, IEEE Trans. Comput. **37**, 138 (1988).

[12] *Emergent Computation: Self-Organizing, Collective, and Cooperative Phenomena in Natural and Artificial Computing Networks*, edited by S. Forrest (North-Holland, Amsterdam, 1990).

[13] R. K. Squier and K. Steiglitz, Complex Syst. **8**, 311 (1994).

[14] *Collision-Based Computing*, edited by A. Adamatzky (Springer-Verlag, Berlin, 2002).

[15] J. P. Crutchfield and M. Mitchell, Proc. Natl. Acad. Sci. U.S.A. **92**, 10742 (1995).

[16] T. Rohlf and S. Bornholdt, Morphogenesis by coupled regulatory networks, e-print q-bio.MN/0401024.

[17] D. Peak, J. D. West, S. M. Messinger, and K. A. Mott, Proc. Natl. Acad. Sci. U.S.A. **101**, 918 (2004).

[18] N. Gershenfeld, *The Nature of Mathematical Modeling* (Cambridge University Press, Cambridge, England, 1999).

[19] J. E. Ruppert-Felsot, O. Praud, E. Sharon, and H. L. Swinney, Phys. Rev. E **72**, 016311 (2005).

[20] G. H. Gunaratne, D. K. Hoffman, and D. J. Kouri, Phys. Rev. E **57**, 5146 (1998).

[21] G. Nathan and G. Gunaratne, Phys. Rev. E **71**, 035101(R) (2005).

[22] H. R. Lewis and C. H. Papadimitriou, *Elements of the Theory of Computation*, 2nd ed. (Prentice-Hall, Upper Saddle River, NJ, 1998).

[23] S. Wolfram, Commun. Math. Phys. **96**, 15 (1984), reprinted in [63],

[24] J. E. Hanson and J. P. Crutchfield, J. Stat. Phys. **66**, 1415 (1992).

[25] W. Hordijk, C. R. Shalizi, and J. P. Crutchfield, Physica D **154**, 240 (2001).

[26] M. Pivato, e-print math.DS/0506417.

[27] P. Grassberger, Phys. Rev. A **28**, 3666 (1983).

[28] P. Grassberger, Physica D **10**, 52 (1984).

[29] N. Boccara, J. Nasser, and M. Roger, Phys. Rev. A **44**, 866 (1991).

[30] J. P. Crutchfield and J. E. Hanson, Physica D **69**, 279 (1993).

[31] J. E. Hanson and J. P. Crutchfield, Physica D **103**, 169 (1997).

[32] K. Lindgren, C. Moore, and M. Nordahl, J. Stat. Phys. **91**, 909 (1998).

[33] M. Cook, Complex Syst. **15**, 1 (2004).

[34] J. Banks, J. Brooks, G. Cairns, G. Davis, and P. Stacy, Am. Math. Monthly **99**, 332 (1992).

[35] D. Ruelle, *Chaotic Evolution and Strange Attractors: The Statistical Analysis of Time Series for Deterministic Nonlinear Systems* (Cambridge University Press, Cambridge, England, 1989), notes prepared by Stefano Isola.

[36] C. R. Shalizi, Ph.D. thesis, University of Wisconsin-Madison, 2001.

[37] C. R. Shalizi, Discrete Math. Theor. Comput. Sci. AB(DMCS), 11 (2003).

[38] C. R. Shalizi, K. L. Shalizi, and R. Haslinger, Phys. Rev. Lett. **93**, 118701 (2004).

[39] U. Parlitz and C. Merkwirth, Phys. Rev. Lett. **84**, 1890 (2000).

[40] S. Kullback, *Information Theory and Statistics*, 2nd ed. (Dover, New York, 1968).

[41] J. P. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1989).

[42] C. R. Shalizi and J. P. Crutchfield, J. Stat. Phys. **104**, 817 (2001).

[43] C. R. Shalizi and C. Moore, e-print cond-mat/0303625.

[44] R. Badii and A. Politi, *Complexity: Hierarchical Structures and Scaling in Physics* (Cambridge University Press, Cambridge, England, 1997).

[45] P. Grassberger, Int. J. Theor. Phys. **25**, 907 (1986).

[46] S. M. Zoldi and H. S. Greenside, Phys. Rev. Lett. **78**, 1687 (1997).

[47] S. M. Zoldi, J. Liu, K. M. S. Bejaj, H. S. Greenside, and G. Ahlers, Phys. Rev. E **58**, R6903 (1998).

[48] A. N. Kolmogorov, Bull. Acad. Sci. USSR, Phys. Ser. **1**, 355 (1937).

[49] V. Capasso and A. Micheletti, in *Topics in Spatial Stochastic Processes*, edited by E. Merzbach (Springer-Verlag, Berlin, 2002), pp. 1-39.

[50] R. Kindermann and J. L. Snell, *Markov Random Fields and their Applications* (American Mathematical Society, Providence, RI, 1980).

[51] R. Fisch, J. Gravner, and D. Griffeath, Stat. Comput. **1**, 23 (1991).

[52] R. Fisch, J. Gravner, and D. Griffeath, in *Spatial Stochastic Processes: A Festschrift in Honor of Ted Harris on His Seventieth Birthday*, edited by K. Alexander and J. Watkins (Birkhäuser, Boston, 1991), pp. 171-188.

[53] A. Szolnoki, G. Szabo, and M. Ravasz, Phys. Rev. E **71**, 027102 (2005).

[54] A. T. Winfree, *When Time Breaks Down: The Three-Dimensional Dynamics of Electrochemical Waves and Cardiac Arrhythmias* (Princeton University Press, Princeton, 1987).

[55] D. Griffeath, in *Denumerable Markov Chains*, 2nd ed., edited by J. G. Kemeny, J. L. Snell, and A. W. Knapp (Springer-Verlag, Berlin, 1976), pp. 425-457.

[56] D. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining* (MIT Press, Cambridge, MA, 2001).

[57] K. Young, Y. Chen, J. Kornak, G. B. Matson, and N. Schuff, Phys. Rev. Lett. **94**, 098701 (2005).

[58] N. Israeli and N. Goldenfeld, Phys. Rev. Lett. **92**, 074105 (2004).
[59] M. A. Shereshevsky, J. Nonlinear Sci. **2**, 1 (1992).
[60] P. Tisseur, Nonlinearity **13**, 1547 (2000).
[61] F. Bagnoli, R. Recthman, and S. Ruffo, Phys. Lett. A **172**, 34 (1992).
[62] J. Urías, R. Rechtman, and A. Enciso, Chaos **7**, 688 (1997).
[63] S. Wolfram, *Cellular Automata and Complexity: Collected Papers* (Addison-Wesley, Reading, MA, 1994).
[64] E. Smith (private communication).