Quantifying Hidden Order out of Equilibrium

Stefano Martiniani,[1,*] Paul M. Chaikin,[1,†] and Dov Levine[2,‡]
[1]Center for Soft Matter Research, Department of Physics, New York University, New York 10003, USA
[2]Department of Physics, Technion-IIT, 32000 Haifa, Israel

(Received 21 June 2018; revised manuscript received 27 November 2018; published 14 February 2019)

While the equilibrium properties, states, and phase transitions of interacting systems are well described by statistical mechanics, the lack of suitable state parameters has hindered the understanding of nonequilibrium phenomena in diverse settings, from glasses to driven systems to biology. The length of a losslessly compressed data file is a direct measure of its information content: The more ordered the data file is, the lower its information content and the shorter the length of its encoding can be made. Here, we describe how data compression enables the quantification of order in nonequilibrium and equilibrium many-body systems, both discrete and continuous, even when the underlying form of order is unknown. We consider absorbing state models on and off lattice, as well as a system of active Brownian particles undergoing motility-induced phase separation. The technique reliably identifies nonequilibrium phase transitions, determines their character, quantitatively predicts certain critical exponents without prior knowledge of the order parameters, and reveals previously unknown ordering phenomena. This technique should provide a quantitative measure of organization in condensed matter and other systems exhibiting collective phase transitions in and out of equilibrium.

DOI: 10.1103/PhysRevX.9.011031

Subject Areas: Computational Physics, Soft Matter, Statistical Physics

I. INTRODUCTION

Intuitively, the more ordered a system is, the shorter the description required to specify a typical microstate. If the probability distribution of the ensemble of microstates is known, then the Shannon entropy [1] provides a quantitative measure of the information content and order. For a random variable X the Shannon entropy is defined as

$$H(X) = -\sum_{\{x\}} p(x) \log p(x), \tag{1}$$

which may be thought of as the average uncertainty in X. Here, $p(x)$ is the probability that a given signal x is generated by a given source; in statistical physics terms, x would be a microstate, and the source may be thought of as defining an ensemble. If we take x to specify microstates of an equilibrium thermodynamic ensemble, and $p(x)$ to be the probabilities of their occurrence, then Eq. (1) reproduces the thermodynamic entropy appropriate to this ensemble. It is important to understand that the framework of equilibrium statistical thermodynamics provides *a priori* probabilities, but this is not the case for systems out of equilibrium, making the explicit computation of H, in general, impossible [2].

Knowledge of the probability distribution is not required for the algorithmic approach to information content pioneered by Kolmogorov and Chaitin [4,5]. This approach culminated in the definition of the *Kolmogorov complexity* K [6], as (loosely speaking) the length of the shortest computer program able to generate a given data sequence (which, for us, is a microstate of the system). When an ensemble of microstates exists, H and K are closely related, and, under fairly general conditions, become equal in the large system limit [6]. However, although elegant, the Kolmogorov complexity is not typically computable, and so cannot be used for physical systems.

A. Computable information density

In this paper we study an easily accessible proxy for these measures, which we will refer to as *computable information density* (CID), which is proportional to the length of a losslessly compressed data string [7]. Concretely we define

$$\mathrm{CID} \equiv \frac{\mathcal{L}(x)}{L}, \tag{2}$$

where $\mathcal{L}(x)$ is the total binary code length of the compressed sequence, and L is the length of the original sequence x

*smartiniani@nyu.edu
†chaikin@nyu.edu
‡dovlevine19@gmail.com

(the number of sites in the system); see Appendix A for further discussion. We note that although we are using the term "sequence," we are not restricted to 1D strings—microstates in any dimension may be compressed by appropriate procedures [9,10].

Despite the wealth of emergent phenomena discovered in classical nonequilibrium many-body physics, for example, in active matter [11,12], driven colloidal systems [13], and granular systems [14], there is still a dearth of viable techniques capable of characterizing their collective behavior, with numerous studies relying instead on the identification of *ad hoc* order parameters and phenomenological descriptions.

One may ask whether the information content as measured by the CID of microstates of a nonequilibrium system can provide, even in principle, useful information about the system. Although in thermodynamic equilibrium we know that the entropy is a central quantity, this is not known for systems far from equilibrium. Moreover, even conceding this point in theory, there is no guarantee that CID is sensitive enough to be fruitful. In this paper, we answer both of these questions in the affirmative, and demonstrate that CID is a practical and rather general approach for characterizing the phase-behavior of many-body nonequilibrium systems, in particular those where the nature of ordering is unclear.

The central idea of lossless data compression algorithms is simple: we wish to encode the data in a shorter binary representation, which may then be decoded to recover the original data exactly. It is natural to ask whether there is a minimal size for the encoding, such that any shorter encoding will lead to loss of information. The answer to this question is provided by Shannon's source coding theorem [1], which states that (in the large system limit [15]) the size of the shortest encoding that can be achieved without loss of information is H. Among lossless compression algorithms, those that do not require *a priori* knowledge of the underlying ensemble are known as *universal* (or adaptive) [6]. The Lempel-Ziv 77 coding algorithm (LZ77) [16] is one such method that is also asymptotically *optimal*, meaning that it produces encodings whose size converges to H in the large system (thermodynamic) limit. Thus, optimal universal data compression algorithms may be used to approximate H for a broad class of data [15], saturating the bound asymptotically.

For our studies, we have computed the CID using the LZ77 compression algorithm (though other choices of universal codes are available), with extrapolation to the thermodynamic limit performed according to Eq. (C5); see Appendixes A–C for a description of the algorithm and a discussion of the extrapolation.

Data compression was first applied to the two-dimensional Ising model by Sheinwald, Lempel, and Ziv [10] as a benchmark for image compression. More recently, application to statistical physics has been mostly through the analysis of the time dependence of single-site variables. For equilibrium systems, a time series of the spin or the Edwards-Anderson autocorrelation parameter at a given site, obtained by Monte Carlo simulation, was used to locate the critical points of the 3D Edwards-Anderson spin glass [17] and the 2D and 3D Ising models [18,19], and to approximate the entropy of the 2D Ising model [19]. Data compression has also proven to be a useful tool in the definition and characterization of complexity of (mostly one-dimensional) dynamical models, such as cellular automata and dynamical systems [20–26], as well as for turbulence [27]. Methods based on data compression have also been used to estimate the entropy production of a nonequilibrium stationary state [28,29] and to detect the onset of chaos in biological systems [30–32].

In this paper, we consider several different interacting nonequilibrium systems, both on and off lattice, in one and two dimensions [33]. We show that CID provides an easily applied and quantitatively accurate measure of information content which can serve as a simple and sensitive way to quantify order, its evolution in time, and its dependence on control parameters [34]. In particular, we show that nonequilibrium analogs of both discontinuous and continuous phase transitions are well characterized by singularities in CID, that certain critical exponents can be extracted without *a priori* knowledge of the order parameter, and that previously unknown ordering phenomena can be discovered.

We note that the order reflected in the CID is distinct from that obtained from dynamical order parameters. In equilibrium systems an order parameter often reflects the onset of a broken symmetry accompanying a phase transition. When this is the case, the singular part of the entropy (and free energy) is expressed as an integral over a power-series expansion in the order parameter [36]. Conversely, the order parameter can be obtained from derivatives of the entropy functional. Thus, the order revealed by entropy and order parameter are intimately related.

In the nonequilibrium systems studied here, it is the dynamics of the system which is singular at the phase transition. The transition is characterized by an order parameter quantifying the change in the dynamics, e.g., the fraction of "active" sites or particles. As of now, we know of no way to relate the dynamics to an entropylike quantity, e.g., the information content of the configurations of such systems. Yet we still expect that a sudden change in the amount of information reflects a change in "order." In particular, for the systems we study here, the order parameters measure whether the system is active (viz., continues to evolve with time), while the CID quantifies the spatial correlations.

II. DISCRETE SYSTEMS

A. Conserved lattice gas in 1D

To illustrate the use of CID, we consider a particularly simple model with a nonequilibrium phase transition, the

conserved lattice gas (CLG) in 1D. Initially, N particles are distributed randomly on $L \geq N$ sites with no multiple occupancy. An occupied site is considered active if one of its neighbors is also occupied. The dynamics consist of moving particles randomly from active sites to unoccupied neighboring sites, as illustrated in Fig. 1(a) (in practice we implement random sequential updates, so we displace one particle at a time). The statistical state of the system is characterized by the order parameter f_a, the fraction of sites that are active. An "absorbing state" is attained when $f_a = 0$, at which point the dynamics ends. No absorbing states are possible for densities $\rho \equiv N/L$ higher than the geometrical limit $\rho_G = 0.5$. For absorbing state models in general [37], it is well known that there exists a critical

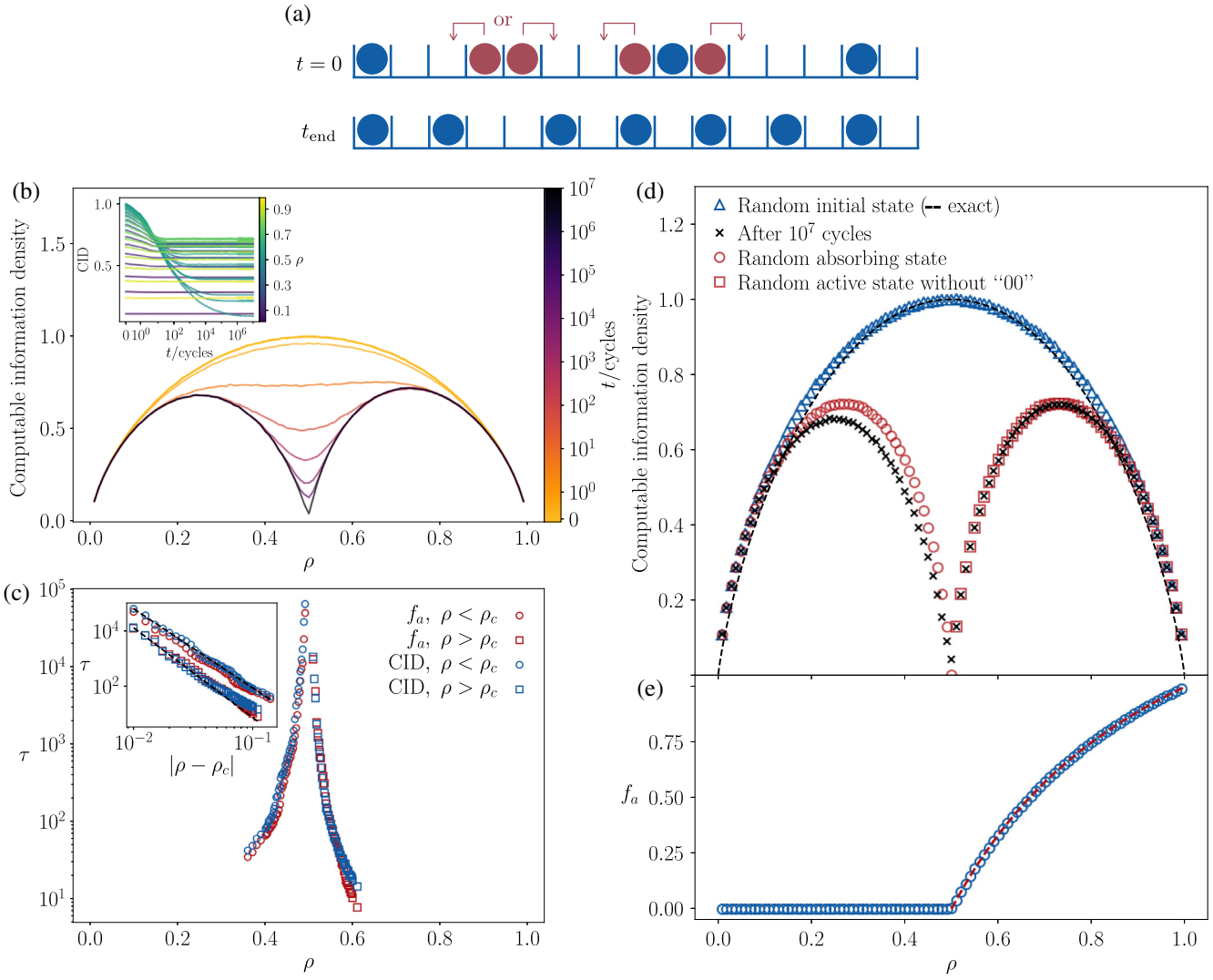

FIG. 1. The 1D conserved lattice gas model of size $L = 10^5$. (a) At time $t = 0$ the system is in an active randomly sampled state (active sites in red) and the possible moves prescribed by the dynamics are indicated by the arrows. When the particle density is below the critical density ρ_c, the system relaxes to an absorbing state, such that the fraction of active sites $f_a = 0$. (b) Time dependence of the CID as a function of particle density ρ, a cycle corresponds to L randomly attempted moves. The system orders as a function of time, developing a cusp minimum at the critical density $\rho_c = 0.5$. The inset shows the CID time-evolution profile for several densities. (c) Characteristic time τ as a function of density ρ, as measured by the decay of f_a and CID, exhibiting a divergence near ρ_c, in line with the cusp minimum in panel (b). Data are from an independent set of calculations in the neighborhood of ρ_c, averaged over ten independent initial conditions. The inset shows τ as a function of $|\rho - \rho_c|$, exhibiting identical power-law divergence and critical exponent $\nu_{\|} = 3 \pm 0.3$ from both measures. Lines of best fit (dashed black lines) were obtained by bootstrapped minimum mean-square error fits using a robust covariance estimator [38,39]. (d) Comparison of the CID for random initial states (blue triangles), states found by the dynamics after 10^7 cycles (black crosses), random (uniformly sampled) absorbing states below ρ_c generated by Monte Carlo sampling (red circles), and active states without "00" pairs above ρ_c (red squares) also generated by Monte Carlo sampling. (e) Fraction of active sites as a function of ρ, red dashed line is the exact solution from Ref. [40].

density ρ_c, such that for $\rho > \rho_c$ the system evolves to an active, fluctuating steady state with a well-defined $f_a > 0$, while for $\rho < \rho_c \leq \rho_G$, the system evolves to an absorbing state. The 1D CLG is atypical in the sense that $\rho_c = \rho_G$ [37], but will be seen to have nontrivial correlations in the absorbing phase. For the 1D CLG the total number of possible absorbing state configurations is $\binom{(1-\rho)L}{\rho L}$ when $\rho \leq \rho_G$ and 0 otherwise; see Appendix D for further discussion. An example of one such state is shown in Fig. 1(a) for $t = t_{\text{end}}$. States of the 1D CLG may be represented simply as a binary string of 0's and 1's signifying the occupation of the sites. These strings can easily be compressed by a large variety of universal codes, we do so by the "unrestricted" Lempel Ziv string-matching code, also known as LZ77 algorithm [16], described in detail in Appendix A.

We analyze a 1D CLG model of size $L = 10^5$ with periodic boundary conditions, for 99 densities in the range $0.01 \leq \rho \leq 0.99$. Starting from random (Bernoulli distributed) initial configurations, we let the system evolve for 10^7 full cycles (sweeps) by random sequential updates. At regular time intervals we measure the CID by LZ77.

In Fig. 1(b) we show the CID as a function of ρ for different times, with the inset indicating the CID time-evolution profiles. At time $t = 0$ the CID matches the Shannon entropy of a Bernoulli sequence $H = -\rho \log \rho - (1-\rho)\log(1-\rho)$, and at low and high densities ρ the CID remains unchanged in time. For densities near the critical point $\rho_c = 0.5$ the system organizes in time and the CID tends to 0 at the critical point, where only a single absorbing state is allowed. We fit the time-dependent CID in the inset of Fig. 1(b) and the fraction of active sites $f_a(t)$ (not shown) with the functional form $y(t) = (y_0 - y_\infty)e^{-t/\tau}(t/t_0)^{-\delta} + y_\infty$, where y_0 and y_∞ are the zero and infinite time limits, respectively, and δ and t_0 are fitted parameters that are roughly constant for all densities. In Fig. 1(c) we show the characteristic time τ as a function of ρ for CID and f_a, both exhibiting a divergence precisely near ρ_c, in line with the cusp in Fig. 1(a). Hence, an analysis of the CID reveals a divergence of the correlation time (critical slowing down) in quantitative agreement with measurements performed on the time decay of f_a, which is the standard order parameter for the analysis of this model. The transition is thus continuous in nature and a fit of τ shows a power-law divergence of the form $\tau \sim |\rho - \rho_c|^{-\nu_\parallel}$, with $\nu_\parallel = 3 \pm 0.3$, see inset in Fig. 1(b).

Analysis of the CID immediately shows the extent that the dynamics orders the states. In Fig. 1(d) we show the CID of the final (absorbing or stationary active) states as obtained by the dynamics (black crosses). On the active side, $\rho > \rho_c$, we compare dynamically obtained states (black crosses) with uniformly sampled unrestricted active states (blue triangles—these are all allowed configurations, the same as the initial random states), as well as with random (uniformly sampled) active states with no 00 pairs (red squares), since they are disallowed by the dynamics [41]. The active configurations with no 00 pairs were generated by Monte Carlo sampling, and their total number is $\binom{\rho L}{(1-\rho)L}$; see Appendix D for further discussion. The perfect match in CID between the random active states with no 00 and those obtained from the dynamics indicates that this is precisely the ensemble sampled by the dynamics above ρ_c. That the CID of the unrestricted active states is much higher than these clearly highlights the degree to which the dynamically accessible states are more ordered than the unrestricted active states.

Ordering due to the dynamics is even more dramatic in the absorbing phase, when $\rho < \rho_c$. Comparison of the CID between random (uniformly sampled) absorbing states generated by Monte Carlo sampling (red circles) and those arrived at by the dynamics (black crosses) shows that the dynamical states are more ordered than the random absorbing states, with the relative gap between the two growing as $\rho \to \rho_c$. This shows that the dynamics sample only a small subset of ordered states out of all the possible

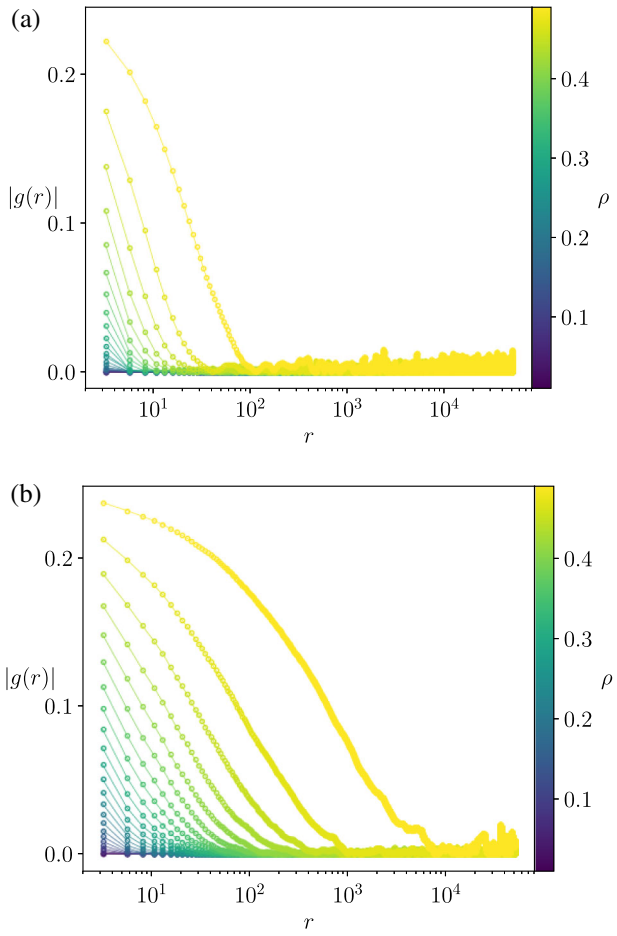

FIG. 2. Absolute value of the autocorrelation function $g(r) \equiv \langle x_i x_{i+r}\rangle - \langle x\rangle^2$ for the 1D conserved lattice gas model for (a) uniformly sampled absorbing states and (b) absorbing states arrived at by the dynamics. The 1D CLG dynamical states are correlated over a much larger scale.

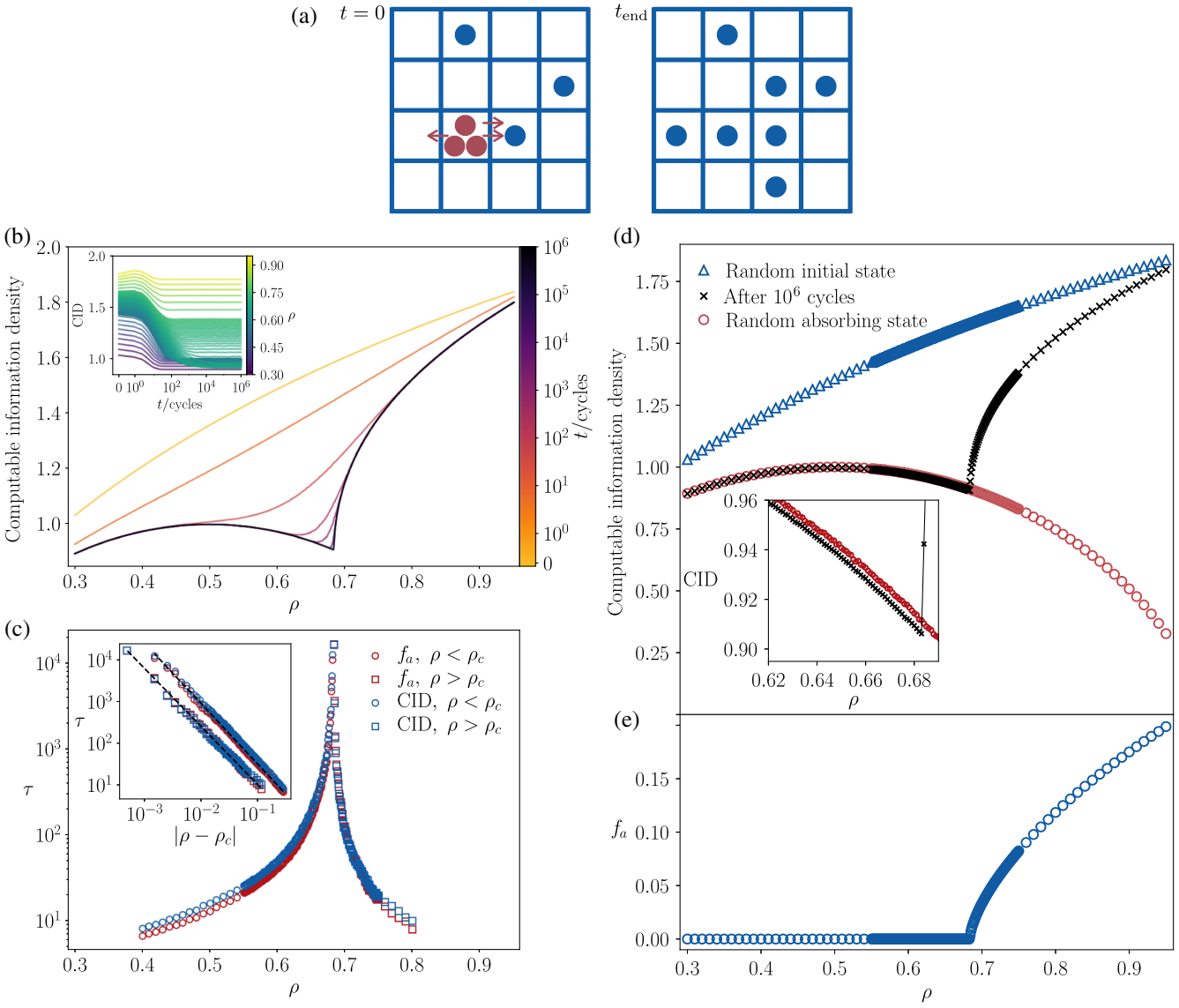

FIG. 3. 2D Manna model with size $2^{10} \times 2^{10}$ and $z_{\max} = 1$. (a) At time $t = 0$ the system is in a randomly sampled state and sites occupied by more than $z_{\max}$ particles are active (active sites in red). At each time step, a randomly sampled active site is emptied by redistributing all particles to its nearest neighbors, a possible move is illustrated by the red arrows. When the particle density is below the critical density ρ_c, the system relaxes to an absorbing state, such that the fraction of active sites $f_a = 0$. (b) CID as a function of particle density ρ for different times; a cycle corresponds to L randomly attempted moves. The system orders as a function of time, developing a cusp minimum at the critical density $\rho_c \approx 0.683$. The inset shows the CID time-evolution profile for several densities. (c) Characteristic time τ as a function of density ρ, as measured by the decay of f_a and CID, exhibiting a divergence near ρ_c, in line with the cusp minimum in panel (b). The inset shows τ as a function of $|\rho - \rho_c|$, exhibiting identical power-law divergence and critical exponent $\nu_{\parallel} = 1.3 \pm 0.2$ from both measures. Lines of best fit (dashed black lines) were obtained by bootstrapped minimum mean-square error fits using a robust covariance estimator [38,39]. (d) Comparison of the random initial states (blue triangles), states found by the dynamics after 10^6 cycles (black crosses) and random (uniformly sampled) absorbing states (red circles); the inset shows that near the critical point the absorbing states arrived at by the dynamics have lower CID, and are more ordered than the uniformly sampled ones. (e) Fraction of active sites as a function of ρ.

absorbing states. To understand the nature of the ordering we compute the autocorrelation function $g(r) \equiv \langle x_i x_{i+r} \rangle - \langle x \rangle^2$ [42] for the random absorbing states [Fig. 2(a)] and the dynamically sampled absorbing states [Fig. 2(b)]. In both cases, the values of $g(r)$ alternate between positive and negative values due to the effective nearest neighbor repulsion, but as $\rho \to \rho_c$ it is apparent that the correlations are much longer ranged for the dynamically sampled absorbing states than for the uniformly sampled absorbing states. These longer ranged correlations indicate that the

dynamics spreads out the particles in a very uniform way as the critical point is approached, a point which was not appreciated in this model before it was revealed by CID.

B. Manna model in 2D

We next consider a two-dimensional system, a discrete lattice sandpile model known as the Manna model. In this model, the sites of a $M \times M$ square lattice (with periodic boundary conditions) are considered active if they are occupied by more than $z_{\max}$ particles. The model allows for an unlimited number of particles at each site and the initial configuration is generated by depositing $N = \rho L$ particles at random on the lattice sites, where $L = M^2$ is the total number of sites. At each time step, one active site is selected and all of its particles redistributed to the neighboring sites at random. This procedure is performed repeatedly, until either there are no active sites or the system arrives to a stationary (steady) state with a characteristic average fraction of active sites f_a. Here, we take $z_{\max} = 1$, such that $\rho_G = 1$ and $\rho_c \approx 0.683$. An example of an initial random state and a final absorbing state is given in Fig. 3(a).

In order to compute the CID of a two- or higher-dimensional system we flatten the grid. In 2D, we use a Peano-Hilbert space filling curve [43]; this is also known as a "Hilbert scan" and it requires $M = 2^m$. This scan, which covers the lattice in a self-similar fashion and preserves locality [44], has been shown to give optimal (distortion-free) compression as $L \to \infty$ [45].

In Fig. 3 we show results for the 2D Manna model of size $2^{10} \times 2^{10}$ over 246 densities in the range $0.3 \leq \rho \leq 0.95$, evolved for approximately 10^6 full cycles. In the CLG the alphabet (possible site occupancies) is $\{0, 1\}$, whereas in the Manna model the alphabet may contain any positive integer and its size may change as the system evolves. A reduction in alphabet size during the evolution contributes to a decrease in the CID. In Fig. 3(b) we show the time evolution of the CID as a function of ρ. The inset shows the CID time-evolution profiles for some of the densities; the curves are averaged over six independently sampled random initial conditions. Note that CID > 1 for the initial random configurations because the alphabet size is greater than 2.

At long times, the system develops a sharp cusplike minimum around $\rho_c \approx 0.683$ indicating a continuous phase transition, and the critical slowing-down is characterized in Fig. 3(c). The correlation times measured from the CID and the fraction of active sites f_a are in quantitative agreement with each other and we obtain critical exponent $\nu_\parallel = 1.3 \pm 0.2$, in agreement with the known value [37]. In Fig. 3(c) we compare the states after 10^6 iterations (black crosses) with randomly generated absorbing states (red circles), i.e., random binary sequences where 1's and 0's occur with frequency ρ and $1 - \rho$, respectively; these have degeneracy $\binom{L}{\rho L}$. For the Manna model $\rho_c < \rho_G$, hence the CID for random (uniformly sampled) absorbing states (red circles) is a smooth function around ρ_c. The inset shows how the absorbing states found by the dynamics have smaller CID and thus are more ordered than the uniformly sampled ones. Recent studies of the 2D Manna and related models [46] indicate that they are hyperuniform [47] at the critical point, meaning that in this limit large-scale density fluctuations are anomalously suppressed.

III. CONTINUUM MODELS

A. Random organization in 2D

The utility of the CID measure rests on the possibility to analyze experimental data, which do not, typically, lie on a lattice. We therefore investigate a 2D continuum system, the "random organization" (RandOrg) model [48], which was developed to explain the reversible-irreversible transition observed in experiments on sheared colloidal suspensions [13]. In RandOrg, the state of the system is given by positions of particles in real space, and in order to calculate the CID, the space must be discretized (quantized). We choose a grid size such that there is at most one particle center in each bin. Note that the resulting configuration is a coarse-grained representation of the original system and therefore the CID estimate may be subject to systematic deviations; we briefly discuss this issue in the Supplemental Material [49].

In the simplest variant of RandOrg, identical disks are initially distributed randomly in space, with disks being considered active if they overlap. Over each cycle all active disks are given in parallel an independent random displacement $\vec{\epsilon}$, whose size ϵ is typically a fraction of a particle diameter. The control parameter for this model is the area fraction, $\phi = Na_0/A$, where A is the area of the system and a_0 is the area of a particle. In this system $\phi_G = \phi_{\text{close-packed}} \approx 0.91$, while $\phi_c \approx 0.43$ in the limit $\epsilon \to 0$. RandOrg can be considered a continuous version of the Manna model and has been shown to belong to the same universality class [50]. An example of an initial random state and a final absorbing state is given in Fig. 4(a).

We study RandOrg using a box of fixed area A, and generate initial configurations by randomly depositing $26\,569 \times \phi$ monodisperse disks with diameter d for 256 area fractions in the range $0.05 \leq \phi \leq 0.8$. We then let the system evolve for approximately 1.5×10^6 full cycles with $\epsilon = d/3$ and periodic boundary conditions. For this system $\phi_c \approx 0.364$. We quantize the coordinates of the system using a square grid fine enough that the centers of two nonoverlapping disks cannot occupy the same grid site and also require that the total number of bins be $2^m \times 2^m$, which is required by the Hilbert scan; in practice this results in a bin-size of approximately $d/\sqrt{3}$.

In Fig. 4(b) we show the time evolution of the CID as a function of ϕ and in the inset we show the CID time-evolution profiles for some of the area fractions; the curves

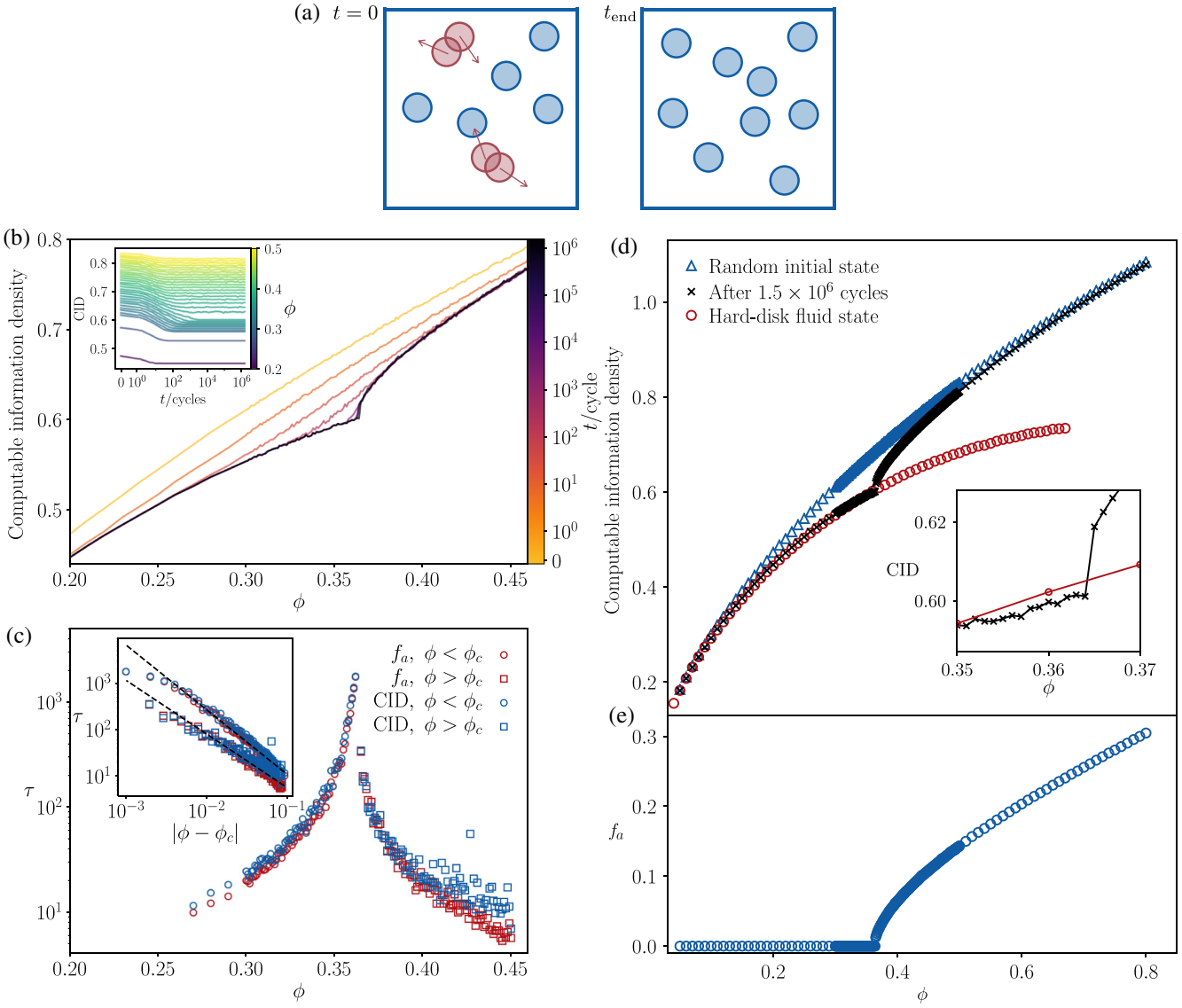

FIG. 4. 2D random organization model with $26\,569 \times \phi$ disks of diameter d. Coordinates are quantized (digitized) using a square grid with bin size approximately $d/\sqrt{3}$. (a) At time $t = 0$ the system is in a randomly sampled ideal gas configuration and disks are considered active when overlapping (active disks in red). Over each cycle the active disks are given a random displacement of size $\epsilon = d/3$, possible moves are illustrated by the red arrows. When the area fraction is below ϕ_c, the system relaxes to an absorbing state, such that the fraction of active particles $f_a = 0$. (b) CID as a function of area fraction ϕ for different times, a cycle corresponds to N attempted parallel moves (one per particle). The system orders as a function of time, developing a cusp at the critical area fraction $\phi_c \approx 0.364$. the inset shows the CID time-evolution profile for several densities. (c) Characteristic time τ as a function of density ρ, as measured by the decay of f_a and CID, exhibiting a divergence near ρ_c, in line with the cusp minimum in panel (b). The inset shows τ as a function of $|\rho - \rho_c|$, exhibiting identical power-law divergence and critical exponent $\nu_\parallel = 1.2 \pm 0.2$ for both measures. Lines of best fit (dashed black lines) were obtained by bootstrapped minimum mean-square error fits using a robust covariance estimator [38,39]. (d) Comparison of the random initial states, states found by the dynamics after 1.5×10^6 cycles and hard-disk fluid state corresponding to uniformly sampled absorbing states; the inset shows that near the critical point the absorbing states arrived at by the dynamics have lower CID, and are more ordered than the uniformly sampled ones. (e) Fraction of active sites as a function of ϕ.

are averaged over seven independent random initial conditions. As we would have expected by analogy with the Manna model, at long times the system develops a CID cusp around $\phi_c \approx 0.364$ and the critical slowing down is characterized in Fig. 4(c). Note the remarkable agreement between the correlation times measured from the fraction of active particles f_a and from the CID; we find $\nu_\parallel = 1.2 \pm 0.2$ on either side of the transition. In Fig. 4(d) we compare the CID of RandOrg final stationary states (black crosses) with those of an equilibrium hard-disk fluid (red circles),

corresponding to uniformly sampled absorbing states. The CID of the hard-disk fluid is smooth around ϕ_c because the hard-disk fluid does not crystallize until the melting density $\phi_{\text{melting}} \approx 0.71$. The inset shows how the absorbing states found by the Manna dynamics are more ordered than those of the hard-disk fluid; just as for the Manna model, the absorbing states of RandOrg at the critical point are hyperuniform [46].

B. Active Brownian particles

With our last example, we show that CID analysis is not limited to absorbing state models. Here, we consider a system of active Brownian particles exhibiting a motility-induced phase separation at a characteristic area fraction ϕ_c; such behavior has been seen in experiments [51] and studied in theory [52–54]. The model by Fily and Marchetti [52] consists of soft disks interacting via a short-ranged repulsive harmonic force $\mathbf{F}_{ij} = k(d - |\mathbf{r}_{ij}|)\Theta(d - |\mathbf{r}_{ij}|)\mathbf{r}_{ij}/|\mathbf{r}_{ij}|$ where k is the spring constant, d is the particle diameter, Θ is the Heavyside step function and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. The particles self-propel at fixed speed v_0 with orientation $\hat{\mathbf{n}}_i = \binom{\cos\theta_i}{\sin\theta_i}$. The dynamics are overdamped, with mobility μ and zero-mean Gaussian rotational white noise $\eta_i(t)$ with rotational diffusion rate ν_r, and are governed by the equations

$$\partial_t \mathbf{r}_i = v_0 \hat{\mathbf{n}}_i + \mu \sum_{j \neq i} \mathbf{F}_{ij},$$
$$\partial_t \theta_i = \eta_i(t). \tag{3}$$

We prepare the system by depositing $N = 16\,384 \times \phi$ monodisperse disks in a fixed area for 95 area fractions in the range $0.01 \leq \phi \leq 0.95$. We minimize the energy by steepest descent [55], and then let the system evolve under periodic boundary conditions with velocity $v_0 = 0.1$, mobility $\mu = 1$, rotational diffusion rate $\nu_r = 5 \times 10^{-4}$, and spring constant $k = 1$. We evolve the system according to Eq. (3) for time $t_{\max} = 3 \times 10^5$ and time step $\Delta t = 10^{-2}$. We quantize the coordinates analogously to the protocol we followed for the RandOrg model, with a bin size of approximately $d/\sqrt{5}$. In Fig. 5 we show the CID as a function of area fraction at different times, and in the inset we show the CID time-evolution profiles; curves are averaged over six independent random initial configurations.

At the lowest area fractions, the system is in an homogeneous gaslike state, and although the system is changing constantly with time, the CID remains unchanged from that of the initial nonoverlapping random configurations. As seen in the inset, at $\phi_c \approx 0.37$ (the third curve from the bottom in the inset), the CID remains essentially constant until about $t \approx 10^4$ iterations, when it drops, indicating the formation of a more ordered state.

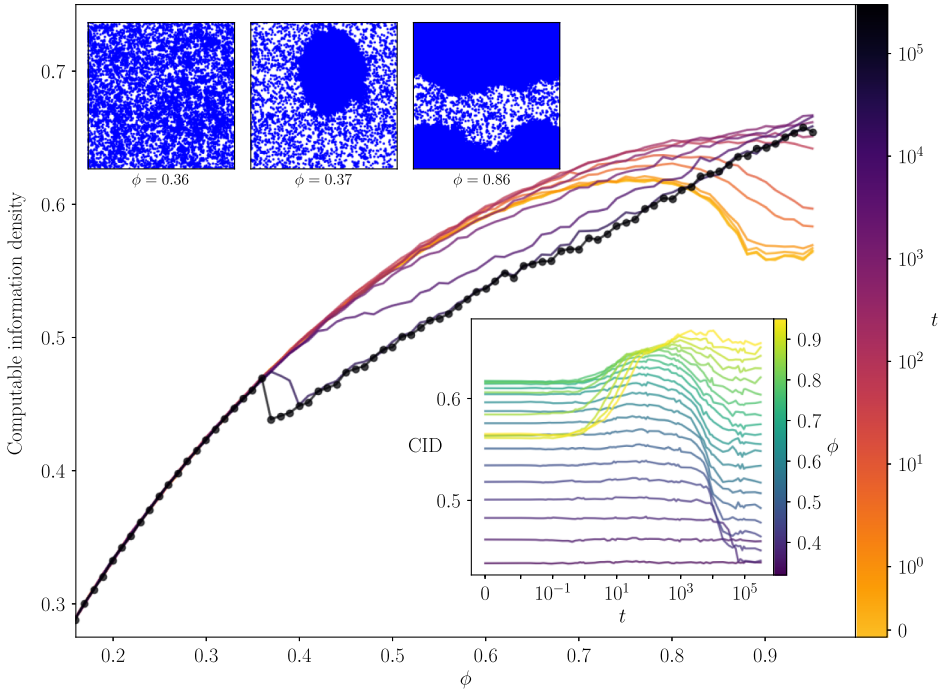

FIG. 5. Active Brownian particles: Systems of $16\,384 \times \rho$ disks with short-range repulsion, self-propelling at speed $v_0 = 0.1$. Coordinates are quantized (digitized) using a square grid with bin-size approximately $d/\sqrt{5}$. At $\phi \approx 0.37$ the CID drops precipitously, indicating ordering associated with clustering and motility-induced phase separation [56]. Representative configurations are shown for $\phi = 0.37$, 0.39, 0.86. For the initial quenched configurations (yellow curve) the flat region for $\phi \gtrapprox 0.88$ corresponds to samples consisting of small grain crystals. The inset shows the time dependence of the CID for different densities.

Inspection of the configurations shown at the top of Fig. 5 shows this to be the result of a phase separation into dense liquidlike and less dense gaslike regions. The steplike discontinuity in the CID between the initial time and the long-time steady state indicate a first order phase transition. These results confirm the density and velocity dependent phase transition previously reported [53], but present a clearer indication of the transition and clearly identify it as first order, in agreement with existing theoretical results [56,57]. In the Supplemental Material [49], we show results for a different velocity, which shifts the critical point. At still higher densities, the CID is not monotonic in time, initially increasing before dropping. In this case the initial configurations, after relaxing by steepest decent to a minimum energy configuration, are highly structured, and almost crystalline. When the particle activity is turned on, the order initially becomes disturbed. At later stages the phase separation sets in and a different order sets in, reducing the CID.

IV. CONCLUSIONS

In what sense, then, does CID uncover "hidden" order? We may think of hidden order as that which is not revealed by an order parameter, but which is uncovered by the information content of the system. For example, in the absorbing state models we consider, we have seen that the CID for dynamically obtained absorbing states is lower than that for the set of all possible absorbing states, suggesting that the dynamically obtained states are more ordered. In the study of the conserved lattice gas we showed evidence of the enhanced order by comparing the pair-correlation functions for all absorbing states and dynamically found absorbing states; to our knowledge this enhanced spatial order was previously unknown. In the random organization and Manna models, the lower CID of the dynamically obtained absorbing states also suggests spatial order, but the nature of this order remains hidden. One may speculate that this order is related to the hyperuniformity associated with the critical points of these systems.

In this work we make a particular choice of universal code for data compression (LZ77), but other approaches are worth exploring, such as Kieffer-Yang grammar-based codes [58–60], prediction by partial matching [61], context tree weighting [62], block sorting [63], and deep neural networks [64,65]. Image-compression techniques based on machine learning approaches have recently resurged due to improved methods for training deep networks [66–68] and may inspire the development of better CID estimators.

The advent and use of powerful lossless data compression algorithms is a half-century old. During this period, its application to many problems as well as its limitations have been extensively explored. Lossless compression not only provides a bound for entropy, but it is a surprisingly good one. The aim of this paper is to illustrate that the CID provides a useful and readily implemented measure for systems out of equilibrium, accurately predicting critical points of phase transitions, their first or second order nature, and even yielding some critical exponents. It allows a quantitative comparison of different states of a system and their time evolution, and it enables the discovery of new phases whose order can subsequently be characterized and studied. These features give us reason to think that CID may find wide use in many areas of statistical many-body physics, especially in the study of disordered and glassy systems, and make an important contribution to our understanding of correlation and organization.

ACKNOWLEDGMENTS

We would like to thank Mark Adler, Daniel Hexner, Yariv Kafri, Johannes Klicpera, Yuval Lemberg, Neri Merhav, Emre Telatar, and Adi Wyner for interesting and useful discussions. We thank Ron Alfia for preliminary numerical results. We are grateful to Ram Avinery, Roy Beck, and Micha Kornreich for useful discussions and for providing their preprint on the application of data compression to study protein folding [69]. This work was primarily supported by the National Science Foundation Physics of Living Systems Grant No. 1504867. D. L. thanks the U.S.-Israel Binational Science Foundation (Grant No. 2014713), the Israel Science Foundation (Grant No. 1866/16), and the Initiative for the Theoretical Sciences at the Graduate Center of CUNY. P. M. C. was supported partially by the Materials Research Science and Engineering Center (MRSEC) Program of the National Science Foundation under Grant No. DMR-1420073.

APPENDIX A: LEMPEL-ZIV STRING-MATCHING CODE (LZ77)

A lossless encoding of a data file may be thought of as a string which may be decoded to yield the original data exactly, with no changes or errors. There are many different algorithms which effect lossless data compression; we employ the unrestricted Lempel Ziv string-matching algorithm, also known as LZ77 [16,70]. LZ77 is a "universal" compression algorithm in that it will compress any given data set, with no *a priori* knowledge about it. To do this, it searches for repeating patterns in the data, and uses this information to construct an efficient encoding.

LZ77 works as follows, beginning from an ordered sequence of data: Looking to the right of a cursor located at some position in the data sequence (initially, the cursor is placed before the first character in the sequence; it will move as the process evolves as will be detailed presently), we search for the "longest previous factor" (LPF), which is the longest subsequence (beginning at the cursor) which has already occurred in the past. We represent this LPF by a tuple (i, ℓ), where i is a pointer to the position of the

matching subsequence (or the character itself when it is observed for the first time) and ℓ is the length of the matching subsequence. The cursor is then moved to the position just following the LPF, and the process is repeated.

The algorithm is best illustrated with an example. Consider the sequence

$$x = abcbabababab$$

of length $L = 12$ and alphabet $\alpha = \{a, b, c\}$ of size $|\alpha| = 3$. Taking the first character of the string to be at position 1, we get, for this example,

$$\mathrm{LZ77}(x) = \{(a,0),(b,0),(c,0),(2,1),(1,2),(5,6)\}.$$

To understand this, we note that, at the outset, no factors have been identified, so that the first factor is the first character, $x_1 = a$ and the length of the matching subsequence is obviously 0, hence the LPF is $(a, 0)$. Moving one position to the right, we encounter $x_2 = b$, which has not yet been seen, and likewise for the next character, $x_3 = c$, thus we have LPF $(b, 0)$ and $(c, 0)$, respectively. Moving to the next position 4, we see that of the subsequences starting at this position, $(b, ba, bab, \ldots)$, only the single character subsequence b has been encountered (at position 2); and the LPF is (2,1) instructing the decoder to copy 1 character starting at position 2. Starting at the following position 5, we note that the words a and ab have already occurred (starting at position 1), but not $aba, abab, \ldots$, thus the LPF is (1,2) instructing the decoder to copy two characters beginning at position 1. Moving to position 7 we note that the entire remaining string, $ababab$, corresponds to the previous subsequence $x_5x_6 = ab$ copied cyclically for six characters, thus the LPF is (5,6). This gives us a list of $C = 6$ tuples with which the entire original string may be reconstructed; it is the LZ77 encoding of the sequence.

The total binary code length $\mathcal{L}(x)$ of the LZ77 encoding can be computed from the number C of longest previous factors: It takes $\log(|\alpha| + L)$ bits to specify a position in the sequence x or a location in the dictionary α, and it takes $\log \ell_j + O(\log\log \ell_j)$ bits to specify ℓ_j, the length of the matching subsequence for the jth factor [70]. Hence, the total binary code length is bounded as

$$\begin{aligned}\mathcal{L}_{\mathrm{LZ77}}(x) &\le C\log(L+|\alpha|) + \sum_{j=1}^{C}\log \ell_j \\ &\quad + O\left(\sum_{j=1}^{C}\log\log \ell_j\right) \\ &\le C\log C + 2C\log\frac{L}{C} + O\left(C\log\log\frac{L}{C}\right), \end{aligned} \tag{A1}$$

where the final bound was obtained by concavity of the log (Jensen's inequality), we assumed $L \gg |\alpha|$, and all log are base 2 throughout. We define the CID to be the ratio

$$\mathrm{CID} \equiv \frac{\mathcal{L}(x)}{L}, \tag{A2}$$

where, for the calculations in this paper, we approximate $\mathcal{L}$ by

$$\mathcal{L} \approx C\log C + 2C\log\frac{L}{C}. \tag{A3}$$

For all the systems we have studied in this paper, we have found that the variance in $\mathcal{L}(x)$, evaluated for different microstates x, is negligible. Note that, although closely related, the CID is not the same as the compression factor [8]

$$\varrho = \frac{\mathrm{CID}}{\log|\alpha|}, \tag{A4}$$

corresponding to the amount of information per character of the *binary* representation of the uncompressed sequence x, although they are equivalent for binary sequences (when $|\alpha| = 2$). Thus, we have that $0 \le \varrho \le 1$ while the CID ≥ 0 is not bounded from above and it is indeed an information "density."

APPENDIX B: RATE OF CONVERGENCE

How well a code compresses a sequence is measured in terms of the "*redundancy*"

$$R \equiv \mathbb{E}\left(\frac{\mathcal{L}(x)}{L}\right) - H, \tag{B1}$$

that is the amount by which the average CID $\equiv \mathcal{L}(x)/L$ exceeds the entropy (per character) of the source. Shannon demonstrated that the redundancy cannot be negative and there exist optimal codes for which the redundancy is zero [1]. It is known that when the sequence x is sampled from a stationary and ergodic process, LZ codes achieve optimal compression in the thermodynamic limit, that is, $R \to 0$ as $L \to \infty$. For individual deterministic sequences, LZ77 codes do at least as well as the empirical (block) entropy, and often better [70].

The rate at which optimality can be attained is rather slow, converging as $\sim \log\log L/\log L$ [71,72]. In the following section we discuss extrapolation to the thermodynamic limit, and in the Supplemental Material [49] we analyze two examples for which the exact value of the entropy is known analytically: a Bernoulli sequence and the two-dimensional Ising model. This gives some insight into the rate of convergence of LZ77 on the basis of numerical results and established theoretical results.

APPENDIX C: EXTRAPOLATION TO THE THERMODYNAMIC LIMIT

It is known that the convergence rate of LZ77 for many positive entropy sources ($H > 0$) scales as $\sim \log\log L/\log L$ [71,72]. Such an exponentially slow rate of convergence calls for careful extrapolation to the thermodynamic limit $L \to \infty$. In particular, we expect that

$$R_{\mathrm{LZ77}} = AH\frac{\log\log L}{\log L}, \tag{C1}$$

where R_{LZ77} is the redundancy for the LZ77 algorithm as defined in Eq. (B1), H is the true Shannon entropy, and the constant A depends on the specifics of the source (for example, in physics terms, A would be dependent on a control parameter such as temperature). Let us denote the average CID for a system of size L by $\hat{H}_L$, so that $\hat{H}_L \equiv \mathbb{E}[\mathrm{CID}(x)]$; this is obtained operationally by measuring the CID for many microstates of the same ensemble and averaging. By rearranging Eq. (C1), we can write an estimator for the true Shannon entropy H:

$$H = \frac{\hat{H}_L \log L}{\log L + A^\star \log\log L} \quad \text{as } L \to \infty, \tag{C2}$$

where we have replaced A by $A^\star$, which will be extracted from finite-size scaling analysis. This is done by plotting $\hat{H}_L$ vs $\log\log L/\log L$, which assuming Eq. (C2), should go as

$$\hat{H}_L = \hat{H}_\infty + A^\star \hat{H}_\infty \frac{\log\log L}{\log L}. \tag{C3}$$

This is a straight line with intercept $\hat{H}_\infty$, which is our estimate for H and slope $A^\star \hat{H}_\infty$, both of which will depend on the system being studied.

The finite-size scaling analysis above is quite compute intensive, and we have found that a good alternative estimator is obtained by using the value of $A^\star$ for a Bernoulli sequence. Rearranging Eq. (C2), and using the fact that $H = 1$ for Bernoulli sequences, gives

$$A^{\star(\mathrm{Bern})} = (\hat{H}_L^{(\mathrm{Bern})} - 1)\frac{\log L}{\log\log L}, \tag{C4}$$

where $\hat{H}_L^{(\mathrm{Bern})}$ is the average CID for a Bernoulli sequence of length L. Note that typically for binary sequences $A^\star \geq \hat{A}^{\star(\mathrm{Bern})}$, a fact that we verify in Figs. S2d and S3d [49]. This choice yields the estimator

$$\hat{H}_\infty = \frac{\hat{H}_L}{\hat{H}_L^{(\mathrm{Bern})}}. \tag{C5}$$

In other words, this extrapolation corresponds to pinning the CID to one for random binary sequences (so, for example, the CID for the high T Ising model tends to one). This is the estimator we have used in this paper. In the Supplemental Material [49], we show explicit finite-size scaling analysis for two analytically tractable systems, the 1D Bernoulli sequences and the 2D Ising model.

APPENDIX D: NUMBER OF STATES OF CONSERVED LATTICE GAS IN 1D

The absorbing states of the conserved lattice gas in 1D are defined such that no "11" can be observed [40]. Likewise, the dynamically accessible active states are such that no "00" can be observed. The implication of this fact is that the number of states is symmetrical above and below the absorbing transition at $\rho_c = 0.5$. For absorbing states at density ρ, replacing all the 0's with 1's will yield a valid active state with density $1-\rho$, and vice versa.

We calculate the number of states for random absorbing states and, equivalently, active states. This can be achieved easily by considering a system of length L and noting that the combination of elements $\mathbb{D}\mathbb{C}$ and $\mathbb{D}1\mathbb{C}$, where $\mathbb{C}\mathbb{D} \equiv 0$ such that $\mathbb{C}\mathbb{D}1\mathbb{C}\mathbb{D}1\mathbb{C}\mathbb{D}\mathbb{C}\mathbb{D} = 010100$, yields valid absorbing states with density ρ satisfying these simple relations

$$\rho = N_{\mathbb{D}1\mathbb{C}}/L \tag{D1}$$

$$L = 2N_{\mathbb{D}1\mathbb{C}} + N_{\mathbb{D}\mathbb{C}} \tag{D2}$$

Then the number of accessible states Ω will be equivalent to the number of permutations of the two sets:

$$\Omega = \frac{(N_{\mathbb{D}1\mathbb{C}} + N_{\mathbb{D}\mathbb{C}})!}{N_{\mathbb{D}1\mathbb{C}}!N_{\mathbb{D}\mathbb{C}}!} = \frac{[L(1-\rho)]!}{(\rho L)![L(1-2\rho)]!} \tag{D3}$$

Note that the expression for the active side can be obtained from the expression for the absorbing side by the simple substitution $\rho \to 1-\rho$, and vice versa.

[1] C. E. Shannon, *A Mathematical Theory of Communication*, Bell Syst. Tech. J. **27**, 379 (1948).

[2] H may be estimated from block entropies (based on the occurrence frequencies of blocks of finite size). Sampling issues attendant to this method are discussed in Ref. [3].

[3] P. Grassberger, *Entropy Estimates from Insufficient Samplings*, arXiv:physics/0307138.

[4] A. N. Kolmogorov, *Three Approaches to the Quantitative Definition of Information*, Int. J. Comput. Math. **2**, 157 (1968).

[5] G. J. Chaitin, *On the Length of Programs for Computing Finite Binary Sequences*, J. Assoc. Comput. Mach. **13**, 547 (1966).

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, Hoboken, 2012).

[7] Note that the CID is not the same as the compression ratio (or compressibility) ϱ of the sequence, in fact $\mathrm{CID} = \varrho \log_2 |\alpha|$, where $|\alpha|$ is the dictionary size of the sequence [8]; see Appendix A for further discussion.

[8] J. Ziv and A. Lempel, *Compression of Individual Sequences via Variable-Rate Coding*, IEEE Trans. Inf. Theory **24**, 530 (1978).

[9] Typically, but not necessarily, higher dimensional microstates are "unwound" to give linear sequences before applying a compression algorithm.

[10] D. Sheinwald, A. Lempel, and J. Ziv, *Two-Dimensional Encoding by Finite-State Encoders*, IEEE Trans. Commun. **38**, 341 (1990).

[11] A. Zöttl and H. Stark, *Emergent Behavior in Active Colloids*, J. Phys. Condens. Matter **28**, 253001 (2016).

[12] S. Ramaswamy, *Active Matter*, J. Stat. Mech. (2017) 054002.

[13] D. J. Pine, J. P. Gollub, J. F. Brady, and A. M. Leshansky, *Chaos and Threshold for Irreversibility in Sheared Suspensions*, Nature (London) **438**, 997 (2005).

[14] G. Ristow, *Pattern Formation in Granular Materials*, Springer Tracts in Modern Physics (Springer-Verlag, Berlin, Heidelberg, 2000).

[15] In particular, for stationary and ergodic processes. By stationary we mean that the probability of an event is invariant with time and by ergodic we mean that the strong law of large numbers holds, viz., the sample average tends to the expected value $\lim_{n\to\infty} 1/n \sum_{i=1}^{n} X_i = \mu$.

[16] J. Ziv and A. Lempel, *A Universal Algorithm for Sequential Data Compression*, IEEE Trans. Inf. Theory **23**, 337 (1977).

[17] V. Cortez, G. Saravia, and E. E. Vogel, *Phase Diagram and Reentrance for the 3D Edwards-Anderson Model Using Information Theory*, J. Magn. Magn. Mater. **372**, 173 (2014).

[18] E. E. Vogel, G. Saravia, and L. V. Cortez, *Data Compressor Designed to Improve Recognition of Magnetic Phases*, Physica(Amsterdam) **391A**, 1591 (2012).

[19] O. Melchert and A. K. Hartmann, *Analysis of the Phase Transition in the Two-Dimensional Ising Ferromagnet Using a Lempel-Ziv String-Parsing Scheme and Black-Box Data-Compression Utilities*, Phys. Rev. E **91**, 023306 (2015).

[20] F. Kaspar and H. G. Schuster, *Easily Calculable Measure for the Complexity of Spatiotemporal Patterns*, Phys. Rev. A **36**, 842 (1987).

[21] W.-H. Steeb and R. Stoop, *Exact Complexity of the Logistic Map*, Int. J. Theor. Phys. **36**, 949 (1997).

[22] V. Benci, C. Bonanno, S. Galatolo, G. Menconi, and M. Virgilio, *Dynamical Systems and Computable Information*, arXiv:cond-mat/0210654.

[23] A. Baronchelli, E. Caglioti, and V. Loreto, *Measuring Complexity with Zippers*, Eur. J. Phys. **26**, S69 (2005).

[24] P. Grassberger, *Randomness, Information, and Complexity*, arXiv preprint arXiv:1208.3459 (2012).

[25] S. Aaronson, S. M. Carroll, and L. Ouellette, *Quantifying the Rise and Fall of Complexity in Closed Systems: The Coffee Automaton*, arXiv preprint arXiv:1405.6903 (2014).

[26] E. Estevez-Rams, R. Lora-Serrano, C. A. J. Nunes, and B. Aragón-Fernández, *Lempel-Ziv Complexity Analysis of One Dimensional Cellular Automata*, Chaos **25**, 123106 (2015).

[27] R. T. Cerbus and W. I. Goldburg, *Information Content of Turbulence*, Phys. Rev. E **88**, 053012 (2013).

[28] É. Roldán and J. M. R. Parrondo, *Estimating Dissipation from Single Stationary Trajectories*, Phys. Rev. Lett. **105**, 150607 (2010).

[29] É. Roldán and J. M. R. Parrondo, *Entropy Production and Kullback-Leibler Divergence between Stationary Trajectories of Discrete Systems*, Phys. Rev. E **85**, 031129 (2012).

[30] M. Nykter, N. D. Price, M. Aldana, S. A. Ramsey, S. A. Kauffman, L. E. Hood, O. Yli-Harja, and I. Shmulevich, *Gene Expression Dynamics in the Macrophage Exhibit Criticality*, Proc. Natl. Acad. Sci. U.S.A. **105**, 1897 (2008).

[31] D. J. Galas, M. Nykter, G. W. Carter, N. D. Price, and I. Shmulevich, *Biological Information as Set-Based Complexity*, IEEE Trans. Inf. Theory **56**, 667 (2010).

[32] N. S. Flann, H. Mohamadlou, and G. J. Podgorski, *Kolmogorov Complexity of Epithelial Pattern Formation: The Role of Regulatory Network Configuration*, BioSystems **112**, 131 (2013).

[33] We have studied systems in three and four dimensions, with similar results.

[34] Different interesting measures, often termed "complexity," which are minimal in perfectly ordered and perfectly random cases, have also been proposed; see, e.g., Ref. [35].

[35] P. Grassberger, *Toward a Quantitative Theory of Self-Generated Complexity*, Int. J. Theor. Phys. **25**, 907 (1986).

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

[37] M. Henkel, H. Hinrichsen, and S. Lübeck, *Absorbing Phase Transitions*, Non-Equilibrium Phase Transitions Vol. 1 (Springer, Dordrecht, 2008).

[38] M. Hubert and M. Debruyne, *Minimum Covariance Determinant*, Wiley Interdiscip. Rev. Comput. Stat. **2**, 36 (2010).

[39] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (CRC Press, Dordrecht, 1994).

[40] M. J. de Oliveira, *Conserved Lattice Gas Model with Infinitely Many Absorbing States in One Dimension*, Phys. Rev. E **71**, 016112 (2005).

[41] The 1D CLG dynamics does not allow for the creation of "00" pairs where none existed, a fact already noted by de Olivera [40].

[42] We computed the autocorrelation function by Wiener-Khinchin $\langle x_i x_{i+r} \rangle = \mathcal{F}^{-1}(|\mathcal{F}(x)|^2)$, that is the inverse Fourier transform of the power spectral density, both computed by discrete FFT.

[43] D. Hilbert, *Ueber die stetige abbildung einer line auf ein flächenstück*, Math. Ann. **38**, 459 (1891).

[44] In the sense that any two points close along the curve are also close in real space, though points nearby in space are not necessarily close on the curve.

[45] A. Lempel and J. Ziv, *Compression of Two-Dimensional Data*, IEEE Trans. Inf. Theory **32**, 2 (1986).

[46] D. Hexner and D. Levine, *Hyperuniformity of Critical Absorbing States*, Phys. Rev. Lett. **114**, 110602 (2015).

[47] S. Torquato and F. H. Stillinger, *Local Density Fluctuations, Hyperuniformity, and Order Metrics*, Phys. Rev. E **68**, 041113 (2003).

[48] L. Corté, P. M. Chaikin, J. P. Gollub, and D. J. Pine, *Random Organization in Periodically Driven Systems*, Nat. Phys. **4**, 420 (2008).

[49] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.9.011031 for further discussion on the rate of convergence of the CID for deterministic sequences; an explicit finite-size scaling analysis for two analytically tractable systems, the 1D Bernoulli sequences and the 2D Ising model; further discussion of continuum systems, and supplementary data on the Brownian active particles system discussed in Sec. III.B.

[50] G. I. Menon and S. Ramaswamy, *Universality Class of the Reversible-Irreversible Transition in Sheared Suspensions*, Phys. Rev. E **79**, 061108 (2009).

[51] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin, *Living Crystals of Light-Activated Colloidal Surfers*, Science **339**, 936 (2013).

[52] Y. Fily and M. C. Marchetti, *Athermal Phase Separation of Self-Propelled Particles with No Alignment*, Phys. Rev. Lett. **108**, 235702 (2012).

[53] Y. Fily, S. Henkes, and M. C. Marchetti, *Freezing and Phase Separation of Self-Propelled Disks*, Soft Matter **10**, 2132 (2014).

[54] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur, *Generalized Thermodynamics of Phase Equilibria in Scalar Active Matter*, Phys. Rev. E **97**, 020602 (2018).

[55] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses* (Cambridge University Press, Cambridge, 2003).

[56] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur, *Generalized Thermodynamics of Motility-Induced Phase Separation: Phase Equilibria, Laplace Pressure, and Change of Ensembles*, New J. Phys. **20**, 075001 (2018).

[57] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur, *Generalized Thermodynamics of Phase Equilibria in Scalar Active Matter*, Phys. Rev. E **97**, 020602 (2018).

[58] J. C. Kieffer and E.-H. Yang, *Grammar-Based Codes: A New Class of Universal Lossless Source Codes*, IEEE Trans. Inf. Theory **46**, 737 (2000).

[59] E.-H. Yang and J. C. Kieffer, *Efficient Universal Lossless Data Compression Algorithms Based on a Greedy Sequential Grammar Transform. I. Without Context Models*, IEEE Trans. Inf. Theory **46**, 755 (2000).

[60] J. C. Kieffer, E.-H. Yang, G. J. Nelson, and P. Cosman, *Universal Lossless Compression via Multilevel Pattern Matching*, IEEE Trans. Inf. Theory **46**, 1227 (2000).

[61] J. Cleary and I. Witten, *Data Compression Using Adaptive Coding and Partial String Matching*, IEEE Trans. Commun. **32**, 396 (1984).

[62] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens, *The Context-Tree Weighting Method: Basic Properties*, IEEE Trans. Inf. Theory **41**, 653 (1995).

[63] H. Cai, S. R. Kulkarni, and S. Verdú, *Universal Entropy Estimation via Block Sorting*, IEEE Trans. Inf. Theory **50**, 1551 (2004).

[64] J. Jiang, *Image Compression with Neural Networks—A Survey*, Signal Processing: Image Communication; Polymer-Plastics Technology and Engineering **14**, 737 (1999).

[65] A. J. Hussain, A. Al-Fayadh, and N. Radi, *Image Compression Techniques: A Survey in Lossless and Lossy Algorithms*, Neurocomputing;Variable Star Bulletin **300**, 44 (2018).

[66] K. Gregor, F. Besse, D. J. Rezende, I. Danihelka, and D. Wierstra, *Towards Conceptual Compression*, in *Proceedings of Advances in Neural Information Processing Systems 29* (NIPS 2016) (Neural Information Processing Systems Foundation, Barcelona, 2016), pp. 3549–3557.

[67] G. Toderici, D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor, and M. Covell, *Full Resolution Image Compression with Recurrent Neural Networks*, in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, New York, 2017), pp. 5435–5443.

[68] D. Minnen, G. Toderici, M. Covell, T. Chinen, N. Johnston, J. Shor, S. J. Hwang, D. Vincent, and S. Singh, *Spatially Adaptive Image Compression Using a Tiled Deep Network*, in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)* (IEEE, New York, 2017), pp. 2796–2800.

[69] R. Avinery, M. Kornreich, and R. Beck, *Universal and Efficient Entropy Estimation Using a Compression Algorithm*, arXiv:1709.10164.

[70] P. C. Shields, *Performance of LZ Algorithms on Individual Sequences*, IEEE Trans. Inf. Theory **45**, 1283 (1999).

[71] A. J. Wyner, *The Redundancy and Distribution of the Phrase Lengths of the Fixed-Database Lempel-Ziv algorithm*, IEEE Trans. Inf. Theory **43**, 1452 (1997).

[72] S. A. Savari, *Redundancy of the Lempel-Ziv String Matching Code*, IEEE Trans. Inf. Theory **44**, 787 (1998).