Dynamical Large Deviations of Two-Dimensional Kinetically Constrained Models Using a Neural-Network State Ansatz

Corneel Casert, Tom Vieijra, Stephen Whitelam, and Isaac Tamblyn

Phys. Rev. Lett. **127**, 120602 — Published 16 September 2021

DOI: 10.1103/PhysRevLett.127.120602

Dynamical large deviations of two-dimensional kinetically constrained models using a neural-network state ansatz

Corneel Casert,[1,*] Tom Vieijra,[1] Stephen Whitelam,[2] and Isaac Tamblyn[3,4,†]

[1] *Department of Physics and Astronomy, Ghent University, 9000 Ghent, Belgium*
[2] *Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA*
[3] *National Research Council Canada, Ottawa, Ontario, Canada*
[4] *Vector Institute for Artificial Intelligence, Toronto, Ontario, Canada*

We use a neural network ansatz originally designed for the variational optimization of quantum systems to study dynamical large deviations in classical ones. We use recurrent neural networks to describe the large deviations of the dynamical activity of model glasses, kinetically constrained models in two dimensions. We present the first finite size-scaling analysis of the large-deviation functions of the two-dimensional Fredrickson-Andersen model, and explore the spatial structure of the high-activity sector of the South-or-East model. These results provide a new route to the study of dynamical large-deviation functions, and highlight the broad applicability of the neural-network state ansatz across domains in physics.

Introduction— Dynamical systems, which include glassy [1–3], driven [4–8], and biochemical systems [9, 10], are defined by ensembles of stochastic trajectories, much as equilibrium systems are defined by ensembles of configurations. Trajectories can be characterized by time-extensive trajectory observables, such as dynamical activity [1, 2, 11], entropy production [12, 13], or other currents [14–16]. Fluctuations of these observables are often described by large-deviation functions—the scaled cumulant-generating function (SCGF) and the rate function—which play a role analogous to thermodynamic potentials for equilibrium systems [17, 18]. Calculating large-deviation functions is a challenging task, requiring the use of advanced methods based on e.g. cloning [19–21], or the use of guiding or auxiliary dynamics [22–24]. Recently, neural networks have been used to construct such auxiliary dynamics [25–27].

Here we demonstrate the ability of the neural-network state ansatz [28] to calculate the large-deviation functions of dynamical systems in both one and two dimensions. We use this ansatz to represent the long-time configurational probability distributions associated with rare trajectories, inspired by its recent success within the variational optimization of quantum systems [28]. The similarities between variational energy minimization in quantum systems and finding the SCGF as the largest eigenvalue of a tilted generator have inspired using variational techniques for studying large deviations in dynamical systems, in particular tensor network methods [29–33]. However, current variational approaches to calculating large-deviation functions are usually limited to one-dimensional systems, while the flexibility of the neural-network ansatz allows for straightforward generalization to higher spatial dimensions. We calculate the large-deviation functions for dynamical activity in prototypical models of slow dynamics, the Fredrickson-Andersen (FA) [34] and South-or-East models, in one and two dimensions, and present the first size-scaling analysis of the large-deviation functions for dynamical activity in two dimensions. We also explore and resolve the spatial structure of the high-activity sector of the South-or-East model. Although we focus on kinetically constrained models, our method for obtaining large-deviation functions is widely applicable. The ease of extension of this approach to two dimensions opens new avenues for the efficient study of dynamical large deviations, and demonstrates the broad applicability of the neural-network state ansatz to classical dynamical problems.

Model and observables— Kinetically constrained models discussed in this work consist of a lattice of N binary spins $i = 1, \ldots, N$, which take values $n_i = 1$ (up) or $n_i = 0$ (down). Spin i flips up (resp. down) with rate $f_i c$ (resp. $f_i(1-c)$), where c is a parameter (equal to the density of up-spins in equilibrium) and f_i is a model-dependent kinetic constraint that renders the dynamics of the model slow or glassy [34–37]. For the FA model, $f_i = \sum_{j \in \mathrm{nn}(i)} n_j$ is the number of nearest-neighbor up-spins. The East (1D) and South-or-East (2D) models have a directed kinetic constraint, with f_i equal to n_{i-1}, or to the number of nearest-neighbor up-spins to the left and above spin i, respectively. Their dynamics are described by the generator

$$W = \sum_i f_i \left[c(\sigma_i^+ + n_i - 1) + (1-c)(\sigma_i^- - n_i) \right], \quad (1)$$

where $\sigma_i^{\pm}$ flips site i up or down. We work with open boundary conditions by connecting each spin on the boundary of the lattice to an immobile site in the down state.

We will study the large deviation properties of the (intensive) dynamical activity $k = K/t$ of the FA and South-or-East models in two dimensions. The activity of trajectory ω of length t is equal to $K(\omega)$, the number of configuration changes within the trajectory. The

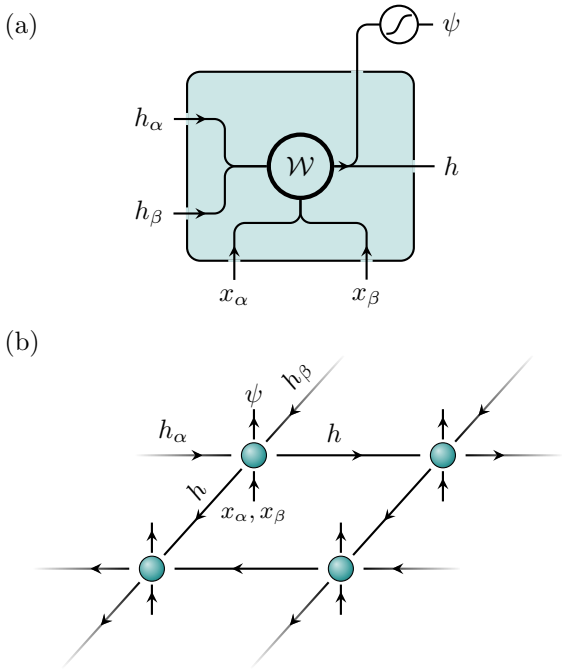

FIG. 1. (a) For a two-dimensional system, an RNN cell with learnable parameters $\mathcal{W}$ calculates a new hidden state h, given the hidden and visible states of previous lattice sites α and β, and passes this state to other lattice sites. The hidden state is further processed to calculate the normalized conditional probability amplitude $\psi(x_i|\{x_{j<i}\})$. (b) This RNN cell is applied to each site of a two-dimensional lattice, and the RNN traverses the lattice row by row in a zig-zag path to calculate the total probability amplitude of a configuration. Note that the hidden state is passed on in both the vertical and horizontal direction, respecting the geometry of the system under study. This probability amplitude is then used for the variational optimization of the scaled cumulant-generating function $\theta(s)$.

probability distribution for activity adopts for long times the large-deviation form $P(K) \approx e^{-tJ(k)}$, where the rate function $J(k)$ quantifies the likelihood of observing atypical values of activity [2, 17]. Information equivalent to that contained in $J(k)$ can be obtained from its Legendre transform, the SCGF: $\theta(s) = -\min_k (sk + J(k))$ [17].

The SCGF can be obtained as the largest eigenvalue of a modified or tilted generator W^s, which we shall do using a variational method. The matrix elements of W^s connecting microstates x and y are

$$W^s_{xy} = W_{xy}e^{-s}(1-\delta_{xy}) - R_x\delta_{xy}. \tag{2}$$

Here W_{xy} are the matrix elements of the original generator, in this case Eq. (1), and $R_x = \sum_{y\neq x} W_{xy}$ [17, 38, 39]. The dynamics described by the tilted generator W^s obeys detailed balance, so that a similarity transformation $P^{-1}W^sP = H^s$ can be performed. Here H^s is an Hermitian matrix with the same eigenvalue spectrum as W^s. It reads [2]

$$\begin{aligned} H^s = \sum_i f_i[e^{-s}\sqrt{c(1-c)}\sigma_i^x \\ - c(1-n_i) - (1-c)n_i], \end{aligned} \tag{3}$$

where σ^x is a Pauli matrix. The SCGF can therefore be obtained by solving the eigenproblem $H^s|\psi(s)\rangle = \theta(s)|\psi(s)\rangle$, where the eigenvectors $|\psi(s)\rangle$ contain the configurational probabilities in the long-time limit for trajectories conditioned to have dynamical activity $\langle k\rangle_s = -\mathrm{d}\theta(s)/\mathrm{d}s$. Because H^s is Hermitian, the SCGF obtained using a variational method results in a lower bound on the exact SCGF; variational optimization can however still be applied for systems where detailed balance is violated, such as asymmetric simple exclusion processes [29, 31]. The large-deviation properties of the one-dimensional FA model are well-studied. In the limit of large system size there exists a singularity in the SCGF of the activity at a size-dependent value of s. Singularities in the SCGF are often associated with phase transitions — in this case a dynamical phase transition between an active and an inactive phase [2, 30, 40–42] — though this is not always the case [43]. In what follows, we show that a neural-network state ansatz can determine the scaling behavior of similar large-deviation singularities in a two-dimensional kinetically constrained model, and can describe the spatial correlations of trajectories displaying atypically large activity.

Recurrent neural-network states— Artificial neural networks can be used within a variational ansatz by mapping configurations $\boldsymbol{x} \equiv (x_1, \ldots, x_N)$ of an N-site lattice system to their corresponding probability amplitude $\psi(\boldsymbol{x})$, which defines the state $|\psi\rangle = \sum_{\boldsymbol{x}} \psi(\boldsymbol{x})|\boldsymbol{x}\rangle$. This ansatz has been shown recently to be capable of representing highly entangled quantum systems [28, 44–63], and has found use in quantum state tomography [64–67]. The expressivity of the neural-network ansatz depends on the architecture of the neural network, and typical choices include restricted Boltzmann machines, fully-connected and convolutional neural networks, and autoregressive neural networks. Here we use autoregressive neural networks, a popular architectural choice for complex machine learning tasks such as natural language processing, sequence generation, or handwriting recognition [68–72]. A state defined by such a network can be sampled in parallel without Markov chains, which is particularly useful for physical regimes in which Markov chains struggle to propose uncorrelated configurations (such as in glassy systems), and allows for the efficient use of state-of-the-art computing infrastructure such as massively parallel graphical processing units. Examples of autoregressive neural networks include PixelCNN [61] and recurrent neural networks (RNN) [62, 63]. We use the RNN ansatz of Refs. [62, 63], which was shown to be

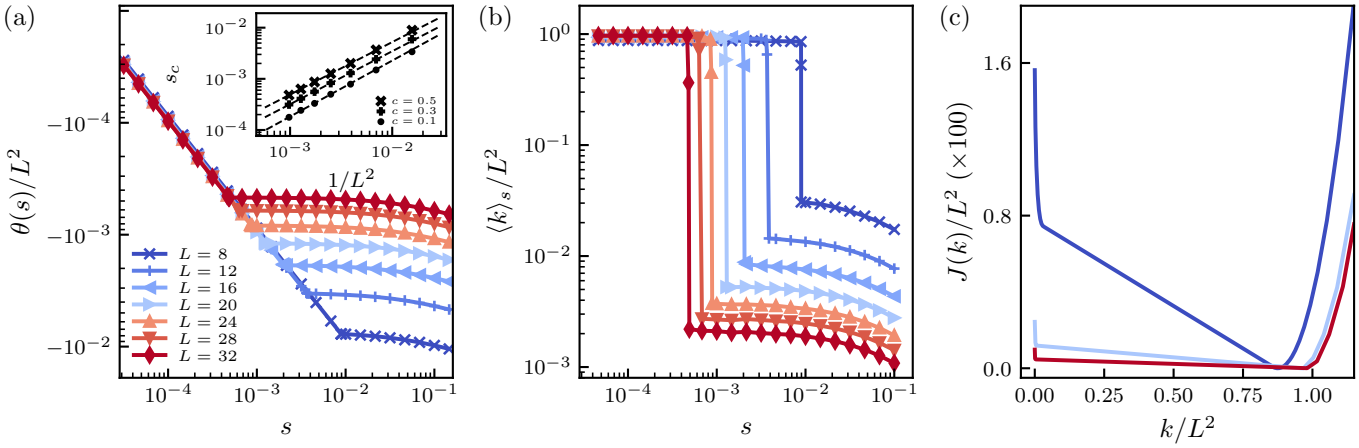

FIG. 2. (a) Scaled cumulant-generating function $\theta(s)$ of the two-dimensional Fredrickson-Andersen model at $c = 0.5$ with two-dimensional recurrent neural-network states, on $L \times L$ square lattices with length between $L = 8$ and $L = 32$. Inset: Scaling of the location s_c of the singularity in the SCGF with the number of lattice sites for three values of c. (b) The dynamical activity $\langle k \rangle_s = -\theta'(s)$ per lattice site. (c) The rate function $J(k)$ for $L = 8, 16, 32$.

highly efficient in the optimization of two-dimensional quantum systems. The probability amplitude of a configuration $\boldsymbol{x}$ with an RNN ansatz is defined as

$$\psi(\boldsymbol{x}) = \prod_{i=1}^{N} \psi(x_i | x_{i-1}, \ldots, x_1), \tag{4}$$

where $\psi(x_i|x_{i-1}, \ldots, x_1)$ is a conditional probability amplitude depending entirely on $\{x_{j<i}\}$ encountered earlier on the lattice. An RNN is defined by its elementary building block, the RNN cell, which is a parametrized non-linear function that sweeps over the lattice site-by-site and is used to calculate $\psi(x_i|\{x_{j<i}\})$ for each. For a one-dimensional configuration $\boldsymbol{x}$, the RNN cell receives at each lattice site i the "visible" state x_{i-1} from the previous site, as well as the "hidden" state vector h_{i-1}, which contains information from the previously encountered degrees of freedom $\{x_{j<i}\}$ and serves as a form of memory. From this, the RNN cell calculates the hidden state of the current lattice site, h_i. This hidden state is processed further to obtain $\psi(x_i|\{x_{j<i}\})$, and is also passed to the next site. In order to calculate the probability amplitude $\psi(\boldsymbol{x})$ of a configuration $\boldsymbol{x}$, we start from an initial visible and hidden state and traverse the lattice site by site with the RNN cell to calculate $\psi(x_i|\{x_{j<i}\})$; finally, we multiply these conditional probability amplitudes per Eq. (4). To draw a new configuration $\boldsymbol{x}$ distributed according to $|\psi(\boldsymbol{x})|^2$, again starting from an initial visible and hidden state, we sample at each site a new visible state x_i from the distribution $P(x_i|\{x_{j<i}\}) = |\psi(x_i|\{x_{j<i}\})|^2$. Together with the new hidden state, this quantity is used as input for the next site. We repeat this process N times. Because the sampling of new configurations $\boldsymbol{x}$ and $\boldsymbol{x}'$ is independent, all operations can be performed in parallel. The RNN ansatz can be naturally extended to higher dimensions; for a two-dimensional system, we provide the RNN cell with a hidden and visible state from two directions (Fig. 1a), and traverse the lattice in a zig-zag path (Fig. 1b). The expressivity of this neural-network ansatz is determined by the choice of the RNN cell and by the dimension of its hidden state vector d_h, also known as the number of hidden units. The weights of the neural network are updated according to the variational principle: to determine the SCGF in this work, weights are optimized so that $\langle \psi | H^s | \psi \rangle$ is maximized. Additional details and schematics describing this ansatz and its optimization are provided in the SM. Because the RNN cell itself is not explicitly dependent on the number of lattice sites of the system, it serves as an optimized starting point for further study of large systems: an RNN cell is first optimized on small lattices, which is computationally relatively cheap, after which it can be optimized for a larger system, often requiring only a few hundred iterations until convergence [63]. Hence, the more costly parts of the optimization procedure, such as determining the optimal hyperparameters and avoiding local minima, are only performed for a small lattice, and obtaining results on very large lattices becomes computationally efficient.

FA model— Having first verified the efficacy of the RNN states in computing large-deviation functions for the one-dimensional FA model and comparing its accuracy to previous results using DMRG (see SM), we turn to the previously unstudied large-deviation behavior of the FA model in two dimensions. To this end we use the two-dimensional RNN shown in Fig. 1. Obtaining large-deviation functions in two dimensions with tensor networks has so far been limited to exclusion processes, using either DMRG [31] or projected entangled pair

states (PEPS) [32]; similar two-dimensional models have also been studied exactly or with macroscopic fluctuation theory [73–75]. Though shown to be very accurate for two-dimensional quantum systems, the computation of tensor network states for two-dimensional systems is typically expensive, requiring either a large number of variational parameters or scaling unfavorably with the number of parameters. Autoregressive neural-network states were recently used to study two-dimensional quantum systems, and have been shown to outperform DMRG [62] and PEPS [61] for several prototypical models while using far fewer parameters.

To describe the dynamics of the two-dimensional FA model, we first optimize neural-network states for an 8×8 lattice. The configuration with all sites in the down state is disconnected from the rest of configuration space due to the kinetic constraints; we only consider dynamics without this configuration during our optimization. In Fig. 2a, we show the resulting SCGF at $c = 0.5$ and for a range of s-values. The dynamical activity can be calculated as a numerical derivative of the SCGF; $\langle k \rangle_s = -\theta'(s)$ (Fig. 2b). Studying the large-deviation behavior of the dynamical activity by varying s reveals a singularity in the SCGF at s_c which separates an active and an inactive sector, similar to observations in one dimension. To further characterize this singularity, we calculate how s_c varies with the number of lattice sites N. Using the RNN states obtained for the 8×8 system as a starting point, we further optimize neural-network states for progressively larger system sizes, repeatedly increasing the linear system size by four sites at a time in order to obtain the SCGF for system sizes up to $N = 1024$. While the training of the initial RNN state for the 8×8 system requires $\mathcal{O}(10^4)$ optimization iterations, each successive optimization upon increasing system size typically converges after less than $\mathcal{O}(10^2)$ iterations. The result of this procedure, shown in Fig. 2, reveals that the value of $s_c(N)$, obtained from the location of the peak of the susceptibility $\chi(s) = \theta''(s)$, moves toward zero as the system size is increased. In the inset of Fig. 2a, we show the scaling of $s_c(N)$ obtained for three different values of c. For each c, the scaling is of the form $s_c \sim N^{-\alpha}$ where the exponent $\alpha \gtrsim 1$ increases slightly for smaller values of c. A similar value for these exponents was recently found for the one-dimensional FA model [30]. We discuss a collapse of the SCGF in the SM.

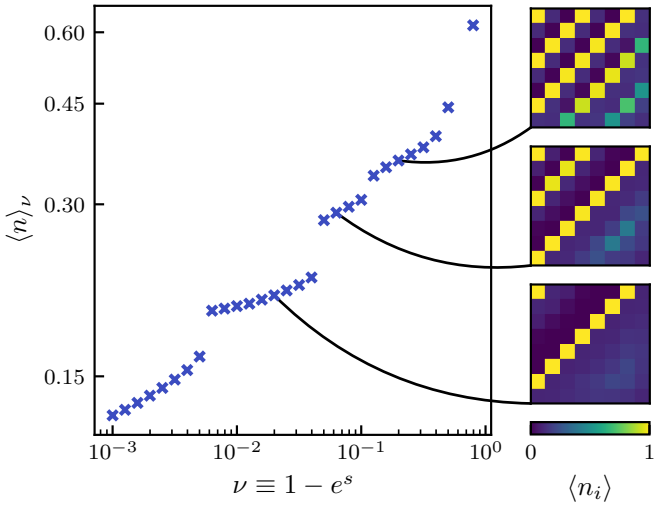

FIG. 3. Average density in the active sector ($s < 0$) of the South-or-East model, as a function of $\nu \equiv 1 - e^s$, for an 8×8 system at $c = 0.1$. On the right, we show a density profile for each of the levels in $\langle n \rangle_\nu$.

Having access to the SCGF allows us to determine, via a Legendre transform, the rate function $J(k)$. The latter defines the distribution $P(K)$ of the activity in the long-time limit, via $P(K) \approx e^{-tJ(k)}$. In Fig. 2c we show the rate function for three different system sizes. These rate functions demonstrate the strongly non-Gaussian distribution of the dynamical activity. In the SM we verify that this is also the case for other values of c.

South-or-East model—The South-or-East model is a two-dimensional generalization of the East model, and has a directed kinetic constraint equal to the number of nearest-neighbor up-spins to the left and above each spin. We consider here only configurations with the spin in the top left corner in the up-state, which allows access to the largest ergodic component of configuration space. Studying the SCGF of the South-or-East and 2D FA models reveals that both models exhibit qualitatively similar large-deviation behavior. However, the spatial structure of trajectories with atypically high activity ($s < 0$) as revealed by time-integrated density profiles shows markedly different behavior. The average density of up-spins can be measured as $\langle n \rangle_s = \frac{1}{L^2} \sum_{i=1}^{N} \langle n_i \rangle_s$; here $\langle n_i \rangle_s = \langle \psi_s | n_i | \psi_s \rangle$ where $|\psi_s\rangle$ is the eigenstate of H^s (Eq. (3)) with eigenvalue $\theta(s)$. While the average density in the active sector of both models is similar at large values of c, more interesting behavior emerges at small values of $c \lesssim 0.1$. For the one-dimensional East model, it was proven [11] and later numerically verified [30] that for $s < 0$ the average density as a function of $\nu \equiv 1 - e^s$ shows distinct plateaus as ν increases for very small values of c. In Fig. 3, we demonstrate that two-dimensional RNN states now allow us to uncover that similar plateau-like features are also present for the South-or-East model at $c = 0.1$. The corresponding density profiles exhibit large anticorrelations in the form of diagonal bands of up-spins surrounded by vacant bands. The number of such bands is different between the density levels. This behavior contrasts with that of the 2D FA model, where density plateaus are absent even for very small c and the local density profiles

are homogeneous up to boundary effects (the spatial structure of the 2D FA model is discussed in the SM).

Outlook— We have presented a study of the large-deviation behavior of two two-dimensional kinetically constrained models. In particular, we have characterized the scaling behavior of the dynamical activity of the two-dimensional Fredrickson-Andersen model, and have described the spatial structure of trajectories with atypically high activity for the South-or-East model. This was made possible by introducing artificial neural-network states as a variational ansatz for obtaining large-deviation functions of classical dynamical systems, drawing from its success in the variational optimization of quantum ones. Our results highlight how the neural-network state ansatz can be employed to efficiently and accurately study large-deviation functions. Although we have focused our study on prototypical models, this ansatz is broadly applicable. Given the rapid improvements being made to the neural-network state ansatz, we expect it to play an important role in the study of dynamical large deviations for higher-dimensional systems.

Computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), and the Flemish Government – department EWI. T. Vieijra is supported as an 'FWO-aspirant' under contract number FWO18/ASP/279. S.W. was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02–05CH11231. I.T. acknowledges NSERC.

* corneel.casert@ugent.be
† isaac.tamblyn@nrc.ca

[1] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. Van Duijvendijk, and F. Van Wijland, Physical Review Letters **98**, 195702 (2007).
[2] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. v. Duijvendijk, and F. v. Wijland, Journal of Physics A: Mathematical and Theoretical **42**, 075007 (2009).
[3] J. P. Garrahan, Physica A: Statistical Mechanics and its Applications **504**, 130 (2018).
[4] S. Vaikuntanathan, T. R. Gingrich, and P. L. Geissler, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **89**, 062108 (2014).
[5] P. Visco, A. Puglisi, A. Barrat, E. Trizac, and F. Van Wijland, Journal of Statistical Physics **125**, 529 (2006).
[6] G. Bunin, Y. Kafri, and D. Podolsky, EPL (Europhysics Letters) **99**, 20002 (2012).
[7] J. Mehl, T. Speck, and U. Seifert, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **78**, 011123 (2008).
[8] B. Derrida, J. L. Lebowitz, and E. R. Speer, in *Journal of Statistical Physics*, Vol. 110 (Springer, 2003) pp. 775–810.
[9] U. Seifert, Reports on Progress in Physics **75**, 126001 (2012).
[10] T. McGrath, N. S. Jones, P. R. Ten Wolde, and T. E. Ouldridge, Physical Review Letters **118**, 028101 (2017).
[11] R. L. Jack and P. Sollich, Journal of Physics A: Mathematical and Theoretical **47**, 015003 (2014).
[12] U. Seifert, Physical Review Letters **95**, 040602 (2005).
[13] C. Maes, Physical Review Letters **119**, 160601 (2017).
[14] T. Bodineau and B. Derrida, Comptes Rendus Physique **8**, 540 (2007).
[15] V. Lecomte, A. Imparato, and F. Van Wijland, in *Progress of Theoretical Physics*, Vol. 184 (Oxford Academic, 2010) pp. 276–289.
[16] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Physical Review Letters **116**, 120601 (2016).
[17] H. Touchette, Physics Reports **478**, 1 (2009).
[18] H. Touchette and R. J. Harris, Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond , 335 (2011).
[19] C. Giardinà, J. Kurchan, and L. Peliti, Physical Review Letters **96**, 120603 (2006).
[20] V. Lecomte and J. Tailleur, Journal of Statistical Mechanics: Theory and Experiment **2007**, P03004 (2007).
[21] T. Nemoto, F. Bouchet, R. L. Jack, and V. Lecomte, Physical Review E **93**, 062123 (2016).
[22] U. Ray, G. K. L. Chan, and D. T. Limmer, Physical Review Letters **120**, 210602 (2018).
[23] D. Jacobson and S. Whitelam, Physical Review E **100** (2019), 10.1103/PhysRevE.100.052139.
[24] U. Ray and G. Kin-Lic Chan, Journal of Chemical Physics **152**, 104107 (2020).
[25] S. Whitelam, D. Jacobson, and I. Tamblyn, arXiv:1909.00835 (2019).
[26] T. H. E. Oakes, A. Moss, and J. P. Garrahan, Machine Learning: Science and Technology (2020), 10.1088/2632-2153/AB95A1.
[27] D. C. Rose, J. F. Mair, and J. P. Garrahan, arXiv:2005.12890 (2020).
[28] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[29] M. Gorissen, J. Hooyberghs, and C. Vanderzande, Physical Review E **79**, 020101 (2009).
[30] M. C. Bañuls and J. P. Garrahan, Physical Review Letters **123**, 200601 (2019).
[31] P. Helms, U. Ray, and G. K. L. Chan, Physical Review E **100**, 022101 (2019).
[32] P. Helms and G. K.-L. Chan, arXiv:2003.03050 (2020).
[33] L. Causer, M. C. Bañuls, and J. P. Garrahan, Physical Review E **103**, 062144 (2021).
[34] G. H. Fredrickson and H. C. Andersen, Physical Review Letters **53**, 1244 (1984).
[35] S. Butler and P. Harrowell, The Journal of Chemical Physics **95**, 4454 (1991).
[36] F. Ritort and P. Sollich, Advances in Physics **52**, 219 (2003).
[37] J. P. Garrahan, P. Sollich, and C. Toninelli, Dynamical Heterogeneities in Glasses, Colloids, and Granular Media **9780199691470** (2010).
[38] J. L. Lebowitz and H. Spohn, Journal of Statistical Physics **95**, 333 (1999).
[39] B. Derrida and T. Sadhu, Journal of Statistical Physics **176**, 773 (2019).
[40] T. Nemoto, R. L. Jack, and V. Lecomte, Physical Review Letters **118**, 115702 (2017).

[41] T. Bodineau and C. Toninelli, Communications in Mathematical Physics **311**, 357 (2011).

[42] T. Bodineau, V. Lecomte, and C. Toninelli, Journal of Statistical Physics **147**, 1 (2012).

[43] S. Whitelam and D. Jacobson, Physical Review E **103**, 032152 (2021).

[44] A. Nagy and V. Savona, Physical Review Letters **122**, 250501 (2019).

[45] M. J. Hartmann and G. Carleo, Physical Review Letters **122**, 250502 (2019).

[46] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, Physical Review Letters **122**, 250503 (2019).

[47] N. Yoshioka and R. Hamazaki, Physical Review B **99**, 214306 (2019).

[48] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Physical Review Letters **121**, 167204 (2018).

[49] K. Choo, T. Neupert, and G. Carleo, Physical Review B **100**, 125124 (2019).

[50] T. Vieijra, C. Casert, J. Nys, W. De Neve, J. Haegeman, J. Ryckebusch, and F. Verstraete, Physical Review Letters **124**, 097201 (2020).

[51] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirac, Nature Physics **15**, 887 (2019).

[52] S. Pilati, E. M. Inack, and P. Pieri, Physical Review E **100**, 043301 (2019).

[53] F. Ferrari, F. Becca, and J. Carrasquilla, Physical Review B **100**, 125131 (2019).

[54] D. Sehayek, A. Golubeva, M. S. Albergo, B. Kulchytskyy, G. Torlai, and R. G. Melko, Physical Review B **100**, 195125 (2019).

[55] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson, and A. A. Bagrov, Nature Communications **11**, 1 (2020).

[56] A. Szabó and C. Castelnovo, arXiv:2002.04613 (2020).

[57] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Physical Review B **96**, 205152 (2017).

[58] D. L. Deng, X. Li, and S. Das Sarma, Physical Review B **96**, 195145 (2017).

[59] D. L. Deng, X. Li, and S. Das Sarma, Physical Review X **7**, 021021 (2017).

[60] G. Carleo, Y. Nomura, and M. Imada, Nature Communications **9**, 1 (2018).

[61] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, Physical Review Letters **124** (2019), 10.1103/PhysRevLett.124.020503.

[62] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Physical Review Research **2**, 023358 (2020).

[63] C. Roth, arXiv:2003.06228 (2020).

[64] G. Torlai and R. G. Melko, Physical Review Letters **120**, 240503 (2018).

[65] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nature Physics **14**, 447 (2018).

[66] G. Torlai, B. Timar, E. P. Van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, Physical Review Letters **123**, 230504 (2019).

[67] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Nature Machine Intelligence **1**, 155 (2019).

[68] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, arXiv:1412.3555 (2014).

[69] A. Graves, in *Supervised Sequence Labelling with Recurrent Neural Networks* (Springer, Berlin, Heidelberg, 2012) pp. 5–13.

[70] A. Graves, arXiv:1308.0850 (2013).

[71] A. Graves, A. R. Mohamed, and G. Hinton, in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (2013) pp. 6645–6649.

[72] A. Van Den Oord, N. Kalchbrenner, and K. Kavukcuoglu, in *33rd International Conference on Machine Learning, ICML 2016*, Vol. 4 (International Machine Learning Society (IMLS), 2016) pp. 2611–2620.

[73] R. Villavicencio-Sanchez, R. J. Harris, and H. Touchette, EPL **105**, 30009 (2014).

[74] N. Tizón-Escamilla, C. Pérez-Espigares, P. L. Garrido, and P. I. Hurtado, Physical Review Letters **119**, 090602 (2017).

[75] C. Pérez-Espigares, P. L. Garrido, and P. I. Hurtado, Physical Review E **93**, 040103 (2016).