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 Tamblyn3,4

1Department of Physics and Astronomy, Ghent University, 9000 Ghent, Belgium
2Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA
3Department of Physics, University of Ottawa, K1N 6N5, Ontario, Canada
4Vector Institute for Artificial Intelligence, Toronto, MSG 1M1, Ontario, Canada

(Received 18 November 2020; revised 13 July 2021; accepted 28 July 2021; published 16 September 2021)

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.

DOI: 10.1103/PhysRevLett.127.120602

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 the use of 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 Letter consist of a lattice of $N$ binary spins $i = 1,...,N$, 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, $f_i = \sum_{j \in \text{NN}(i)} n_j$ 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

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

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.

0031-9007/21/127(12)/120602(6) 120602-1 © 2021 American Physical Society
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 \( \omega \) of length \( t \) is equal to \( K(\omega) \), the number of configuration changes within the trajectory. The probability distribution for activity adopts for long times the large-deviation form \( P(k) \approx e^{-\beta J(k)} \), 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 \( W' \), which we shall do using a variational method. The matrix elements of \( W' \) connecting microstates \( x \) and \( y \) are

\[
W'_{xy} = W_{xy} e^{-s(1 - \delta_{xy})} - R_s \delta_{xy}. \tag{2}
\]

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

\[
H' = \sum_{i} f_i [e^{-s} \sqrt{c(1-c)} \sigma_i^z - c(1-n_i) - (1-c)n_i], \tag{3}
\]

where \( \sigma^z \) is a Pauli matrix. The SCGF can therefore be obtained by solving the eigenproblem \( H'|\psi(s)\rangle = \theta(s) |\psi(s)\rangle \), where the eigenvectors \( |\psi(s)\rangle \) contain the configurational probabilities in the long-time limit for trajectories conditioned to have dynamical activity \( \langle k \rangle_t = -d\theta(s)/ds \). Because \( H' \) 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 \( x \equiv (x_1, \ldots, x_N) \) of an \( N \)-site lattice system to their corresponding probability amplitude \( \psi(x) \), which defines the state \( |\psi\rangle = \sum_x \psi(x)|x\rangle \). 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, and 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 highly efficient in the optimization of two-dimensional quantum systems. The probability amplitude of a configuration \( x \) with an RNN ansatz is defined as

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

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 nonlinear 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 \( x \), 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 \( \psi(x) \) of a configuration \( x \), 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 \( x \) distributed according to \( |\psi(x)|^2 \), 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 \( x \) and \( x' \) is independent, all operations can be performed in
hidden state vector by the choice of the RNN cell and by the dimension of its
The expressivity of this neural-network ansatz is determined
h
SCGF in this work, weights are optimized so that
according to the variational principle: to determine the
units. The weights of the neural network are updated
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. 1(a)], and traverse the lattice in a zigzag path [Fig. 1(b)].
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 
\(\langle \psi | H^2 | \psi \rangle \) is maximized. Additional details and schematics
describing this ansatz and its optimization are provided in the
Supplemental Material [73]. 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 the density matrix renormalization
group (DMRG) (see Supplemental Material [73]), 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 [74–76].
Though shown to be very accurate for two-dimensional
quantum systems, the computation of tensor network states
for two-dimensional systems is typically expensive, requir-
ing either a large number of variational parameters or
scaling unfavorably with the number of parameters.

To describe the dynamics of the two-dimensional FA
model, we first optimize neural-network states for an
\(8 \times 8\) lattice. The configuration with all sites in the down state is
disconnected from the rest of the configuration space due to the
kinetic constraints; we only consider dynamics without
this configuration during our optimization. In Fig. 2(a),
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, \(\langle k \rangle_s = -\theta^\prime(s)\)
[Fig. 2(b)]. 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 \(8 \times 8\) 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 \(8 \times 8\) system requires \(O(10^4)\)
optimization iterations, each successive optimization upon
increasing the system size typically converges after less
than \(O(10^5)\) 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 \(\chi(s) = \theta^\prime(s)\),
moves toward zero as the system size is increased. In the
inset of Fig. 2(a), we show the scaling of \(s_c(N)\) obtained for
three different values of \(c\). For each \(c\), the scaling is of the

FIG. 1. (a) For a two-dimensional system, an RNN cell with
learnable parameters \(W\) calculates a new hidden state \(h\), given the
hidden and visible states of previous lattice sites \(\alpha\) and \(\beta\), 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 zigzag 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 \(\theta(s)\).
form \( s_c \sim N^{-\alpha} \) where the exponent \( \alpha \geq 1 \) 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 Supplemental Material [73].

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 \( P(K) \approx e^{-J(k)} \). In Fig. 2(c), 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 Supplemental Material [73] 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 \( \langle n \rangle_s = \langle 1/N \rangle \sum_{i=1}^{N} \langle n_i \rangle_s \); here \( \langle n_i \rangle_s = \langle \psi_i | n_i | \psi_i \rangle \) where \( | \psi_i \rangle \) 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 \( \nu \) 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 apart from boundary effects (the spatial structure of the 2D FA model is discussed in the Supplemental Material [73]).

**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-Anderson 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. V. is supported as an “FWO-aspirant” under Contract No. 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
tisaac.tamblyn@uottawa.ca

[1] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. V. Duijvendijk, and F. Van Wijland, Phys. Rev. Lett. 98, 195702 (2007).
[2] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. V. Duijvendijk, and F. V. Wijland, J. Phys. A 42, 075007 (2009).
[3] J. P. Garrahan, Physica (Amsterdam) 504A, 130 (2018).
[4] S. Vaikuntanathan, T. R. Gingrich, and P. L. Geissler, Phys. Rev. E 89, 062108 (2014).
[5] P. Visco, A. Puglisi, A. Barrat, E. Trizac, and F. Van Wijland, J. Stat. Phys. 125, 533 (2006).
[6] G. Bunin, Y. Kafri, and D. Podolsky, Europhys. Lett. 99, 20002 (2012).
[7] J. Mehl, T. Speck, and U. Seifert, Phys. Rev. E 78, 011123 (2008).
[8] B. Derrida, J. L. Lebowitz, and E. R. Speer, in J. Stat. Phys. 110, 775 (2003).
[9] U. Seifert, Rep. Prog. Phys. 75, 126001 (2012).
[10] T. McGrath, N. S. Jones, P. R. Ten Wolde, and T. E. Ouldridge, Phys. Rev. Lett. 118, 028101 (2017).
[11] R. L. Jack and P. Sollich, J. Phys. A 47, 015003 (2014).
[12] U. Seifert, Phys. Rev. Lett. 95, 040602 (2005).
[13] C. Maes, Phys. Rev. Lett. 119, 160601 (2017).
[14] T. Bodineau and B. Derrida, C.R. Phys. 8, 540 (2007).
[15] V. Lecomte, A. Imparato, and F. Van Wijland, in Progress of Theoretical Physics (Oxford Academic, New York, 2010), Vol. 184, pp. 276–289.
[16] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Phys. Rev. Lett. 116, 120601 (2016).
[17] H. Touchette, Phys. Rep. 478, 1 (2009).
[18] H. Touchette and R. J. Harris, in Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond, edited by R. Klages, W. Just, and C. Jarzynski (Wiley-VCH, Weinheim, 2013), p. 335.
[19] C. Giardinà, J. Kurchan, and L. Peliti, Phys. Rev. Lett. 96, 120603 (2006).
[20] V. Lecomte and J. Tailleur, J. Stat. Mech. (2007) P03004.
[21] T. Nemoto, F. Bouchet, R. L. Jack, and V. Lecomte, Phys. Rev. E 93, 062123 (2016).
[22] U. Ray, G. K. L. Chan, and D. T. Limmer, Phys. Rev. Lett. 120, 210602 (2018).
[23] D. Jacobson and S. Whitelam, Phys. Rev. E 100, 052139 (2019).
[24] U. Ray and G. Kin-Lic Chan, J. Chem. Phys. 152, 104107 (2020).
[25] S. Whitelam, D. Jacobson, and I. Tamblyn, J. Chem. Phys. 153, 044113 (2020).
[26] T. H. E. Oakes, A. Moss, and J. P. Garrahan, Mach. Learn. Sci. Technol. 1, 035004 (2020).
[27] D. C. Rose, J. F. Mair, and J. P. Garrahan, New J. Phys. 23, 013013 (2021).
[28] G. Carleo and M. Troyer, Science 355, 602 (2017).
[29] M. Gorissen, J. Hooyberghs, and C. Vanderzande, Phys. Rev. E 79, 020101(R) (2009).
[30] M. C. Bañuls and J. P. Garrahan, Phys. Rev. Lett. 123, 200601 (2019).
[31] P. Helms, U. Ray, and G. K.-L. Chan, Phys. Rev. E 100, 022101 (2019).
[32] P. Helms and G. K.-L. Chan, Phys. Rev. Lett. 125, 140601 (2020).
[33] L. Causer, M. C. Bañuls, and J. P. Garrahan, Phys. Rev. E 103, 062144 (2021).
[34] G. H. Fredrickson and H. C. Andersen, Phys. Rev. Lett. 53, 1244 (1984).
[35] S. Butler and P. Harrowell, J. Chem. Phys. 95, 4454 (1991).
[36] F. Ritort and P. Sollich, Adv. Phys. 52, 219 (2003).
[37] J. P. Garrahan, P. Sollich, and C. Toninelli, in Dynamical Heterogeneities in Glasses, Colloids, and Granular Media, edited by L. Berthier, G. Biroli, J.-P. Bouchaud, L. Cipelletti, and W. van Saarloos, International Series of Monographs on Physics, (Oxford University Press, Oxford, England, 2011).
[38] J. L. Lebowitz and H. Spohn, J. Stat. Phys. 95, 333 (1999).
[39] B. Derrida and T. Sadhu, J. Stat. Phys. 176, 773 (2019).
[40] T. Nemoto, R. L. Jack, and V. Lecomte, Phys. Rev. Lett. 118, 115702 (2017).
[41] T. Bodineau and C. Toninelli, Commun. Math. Phys. 311, 357 (2012).
[42] T. Bodineau, V. Lecomte, and C. Toninelli, J. Stat. Phys. 147, 1 (2012).
[43] S. Whitelam and D. Jacobson, Phys. Rev. E 103, 032152 (2021).
[44] A. Nagy and V. Savona, Phys. Rev. Lett. 122, 250501 (2019).
[45] M. J. Hartmann and G. Carleo, Phys. Rev. Lett. 122, 250502 (2019).
[46] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, Phys. Rev. Lett. 122, 250503 (2019).
[47] N. Yoshioka and R. Hamazaki, Phys. Rev. B 99, 214306 (2019).
[48] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Phys. Rev. Lett. 121, 167204 (2018).
[49] K. Choo, T. Neupert, and G. Carleo, Phys. Rev. B 100, 125124 (2019).
[50] T. Vieijra, C. Casert, J. Nys, W. De Neve, J. Haegeman, J. Ryckebusch, and F. Verstraete, Phys. Rev. Lett. 124, 097201 (2020).
[51] R. G. Melko, G. Carleo, J. Carraquilla, and J. I. Cirac, Nat. Phys. 15, 887 (2019).
[52] S. Pilati, E. M. Inack, and P. Pieri, Phys. Rev. E 100, 043301 (2019).
[53] F. Ferrari, F. Becca, and J. Carraquilla, Phys. Rev. B 100, 125131 (2019).
[54] D. Sehayek, A. Golubeva, M. S. Albergo, B. Kulchytskyy, G. Torlai, and R. G. Melko, Phys. Rev. B 100, 195125 (2019).
[55] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson, and A. A. Bagrov, Nat. Commun. 11, 1593 (2020).
[56] A. Szabó and C. Castelnovo, Phys. Rev. Research 2, 033075 (2020).
[57] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Phys. Rev. B 96, 205152 (2017).
[58] D. L. Deng, X. Li, and S. Das Sarma, Phys. Rev. B 96, 195145 (2017).
[59] D. L. Deng, X. Li, and S. Das Sarma, Phys. Rev. X 7, 021021 (2017).
[60] G. Carleo, Y. Nomura, and M. Imada, Nat. Commun. 9, 5322 (2018).
[61] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, Phys. Rev. Lett. 124, 020503 (2020).
[62] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carraquilla, Phys. Rev. Research 2, 023358 (2020).
[63] C. Roth, arXiv:2003.06228.
[64] G. Torlai and R. G. Melko, Phys. Rev. Lett. 120, 240503 (2018).
[65] G. Torlai, G. Mazzola, J. Carraquilla, M. Troyer, R. Melko, and G. Carleo, Nat. Phys. 14, 447 (2018).
[66] G. Torlai, B. Timar, E. P. L. Van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, Phys. Rev. Lett. 123, 230504 (2019).
[67] J. Carraquilla, G. Torlai, R. G. Melko, and L. Aolita, Nat. Mach. Intell. 1, 155 (2019).
[68] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, arXiv:1412.3555.
[69] A. Graves, in Supervised Sequence Labelling with Recurrent Neural Networks (Springer, Berlin, Heidelberg, 2012), pp. 5–13.
[70] A. Graves, arXiv:1308.0850.
[71] A. Graves, A. r. Mohamed, and G. Hinton, in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE, Vancouver, 2013), pp. 6645–6649.
[72] A. Van Den Oord, N. Kalchbrenner, and K. Kavukcuoglu, in Proceedings of the 33rd International Conference on International Conference on Machine Learning (JMLR, New York, 2016), Vol. 48, pp. 1747–1756.
[73] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.127.120602 for details on the neural network architecture and training, and additional results on scaling, spatial structure and rate functions.
[74] R. Villavicencio-Sanchez, R. J. Harris, and H. Touchette, Europhys. Lett. 105, 30009 (2014).
[75] N. Tizón-Escamilla, C. Pérez-Espigares, P. L. Garrido, and P. I. Hurtado, Phys. Rev. Lett. 119, 090602 (2017).
[76] C. Pérez-Espigares, P. L. Garrido, and P. I. Hurtado, Phys. Rev. E 93, 040103(R) (2016).