Large deviations of one-dimensional kinetically constrained models with recurrent neural networks

Corneel Casert
Department of Physics and Astronomy
Ghent University
9000 Ghent, Belgium
corneel.casert@ugent.be

Tom Vieijra
Department of Physics and Astronomy
Ghent University
9000 Ghent, Belgium
tom.vieijra@ugent.be

Stephen Whitelam
Molecular Foundry
Lawrence Berkeley National Laboratory
1 Cyclotron Road, Berkeley
CA 94720, USA
swhitelam@lbl.gov

Isaac Tamblyn
National Research Council of Canada
Ottawa, ON K1N 5A2, Canada
Vector Institute for Artificial Intelligence
Toronto, Ontario, Canada
isaac.tamblyn@nrc.ca

Abstract

We show how neural-network states can be used to calculate the dynamical large-deviation functions of classical systems, which play the role of thermodynamic potentials for trajectories. We use a recurrent neural network to obtain the scaled cumulant-generating function for the dynamical activity of the Fredrickson-Andersen model, a prototypical kinetically constrained model for glass formers, in one dimension with high accuracy. These results offer a novel method for the determination of large-deviation functions, and highlight the broad applicability of the neural-network state ansatz.

1 Introduction

Whereas equilibrium systems are studied with ensembles of configurations, dynamical systems—including glassy [1, 2, 3], driven [4, 5, 6, 7, 8], and biochemical systems [9, 10]—are studied with ensembles of stochastic trajectories. Time-extensive trajectory observables, examples of which include dynamical activity [1, 2, 11], entropy production [12, 13], or other currents [14, 15, 16], characterize these trajectories. Large-deviation functions—the scaled cumulant-generating function (SCGF) and rate function—describe the fluctuations of these observables, and play the role of thermodynamic potentials for trajectories [17, 18]. Here, we demonstrate the use of the neural-network state ansatz [19], which shall represent the long-time configurational probability distributions associated with rare trajectories, to calculate these large-deviation functions. In particular, we will study the large deviations of a time-extensive observable $\mathcal{A}$, which is defined along a trajectory ω of length t with $K(\omega)$ configurational changes as

$$\mathcal{A} = \sum_{k=0}^{K(\omega)-1} \alpha_{x_k x_{k+1}}, \tag{1}$$

with α_{xy} the change in $\mathcal{A}$ when jumping between the discrete configurations x and y. Such an observable is said to follow a large-deviation principle if in the limit of $t \to \infty$ it follows the distribution

$$P(\mathcal{A}) \approx e^{-tJ(a)}, \tag{2}$$

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

where $a = \mathcal{A}/t$, and $J(a)$ is the rate function which quantifies the exponential decrease of observing atypical values of $\mathcal{A}$ [17]. The cumulants of $\mathcal{A}$ can in this limit be obtained as derivatives of the scaled cumulant-generating function $\theta(s)$

$$\theta(s) = \lim_{t\to\infty} \frac{1}{t} \ln \left\langle e^{-s\mathcal{A}} \right\rangle . \tag{3}$$

The SCGF $\theta(s)$ and rate function $J(a)$ are connected through a Legendre transform, $\theta(s) = -\min_a\{sa + J(a)\}$.

Obtaining these rate functions is challenging for many physical systems, requiring the use of advanced Monte Carlo methods based on *e.g.* cloning [20, 21, 22] or auxiliary dynamics [23, 24, 25]. Recently, neural networks have been used to construct such auxiliary dynamics [26, 27, 28]. Alternatively, the SCGF $\theta(s)$ can be obtained as the largest eigenvalue of a modified or "tilted" generator W^s whose matrix elements connecting microstates x and y are

$$W^s_{xy} = W_{xy} e^{-s\alpha_{xy}} (1 - \delta_{xy}) - R_x \delta_{xy}. \tag{4}$$

Here W_{xy} are the matrix elements of the original generator describing the dynamical process and $R_x = \sum_{y\neq x} W_{xy}$ [29, 17, 30]. The SCGF can hence be obtained by solving the eigenproblem $W^s|P^s\rangle = \theta(s)|P^s\rangle$, where the right eigenvectors $|P^s\rangle$ contain the configurational probabilities in the long-time limit for trajectories conditioned on $\langle a\rangle = -\mathrm{d}\theta(s)/\mathrm{d}s$. The similarities between the variational energy minimization in quantum systems and finding the SCGF as largest eigenvalue of a tilted generator have led to the use of variational techniques for studying large deviations in dynamical systems, in particular the density matrix renormalization group (DMRG), [31, 32, 33]. Inspired by their recent applications in the variational optimization of quantum systems, we here demonstrate the use of the neural-network state ansatz [19] for determining the SCGF of a one-dimensional dynamical system.

2 Recurrent neural-network states

Artificial neural networks can be used as 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 recently been shown to be capable of faithfully representing highly entangled quantum systems [19, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53], and has found use in quantum state tomography [54, 55, 56, 57]. The expressivity of the neural-network state 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 which are popular architectural choices for complex machine learning tasks such as natural language processing, sequence generation or handwriting recognition [58, 59, 60, 61, 62]. A state defined by such a network can be sampled in parallel without Markov chains, which is particularly useful for physical regimes where Markov chains struggle to propose uncorrelated configurations. Examples of autoregressive neural networks include PixelCNN [51] and recurrent neural networks (RNN) [52, 53]. Here, we use the RNN ansatz of [52, 53], which was shown to be highly efficient in the optimization of quantum systems. 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. For a given configuration $\boldsymbol{x}$, at each lattice site i the RNN cell receives the "visible" state x_{i-1} from the previous site, as well as the "hidden" state h_{i-1}, which contains information from the previously encountered degrees of freedom and serves as a form of memory. From this, the RNN cell calculates the hidden state for the current lattice site, h_i, and further processes this hidden state to obtain a conditional probability amplitude $\psi(x_i|x_{i-1}, \ldots, x_1)$ for a new visible state—depending entirely on $x_{j<i}$ encountered earlier on the lattice. A new visible state is obtained by sampling from the distribution $P(x_i|x_{i-1}, \ldots, x_1) = |\psi(x_i|x_{i-1}, \ldots, x_1)|^2$ which, together with the new hidden state, can be used as input for the next site. Starting from an initial visible and hidden state, the whole lattice can be traversed in this way, which allows for highly parallel sampling and calculation of probability amplitudes. The probability amplitude of the entire configuration $\boldsymbol{x}$ with an autoregressive neural-network state is defined as

$$\psi(\boldsymbol{x}) = \prod_{i=1}^{N} \psi(x_i|x_{i-1}, \ldots, x_1). \tag{5}$$

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. The expectation value of the tilted generator can be written as

$$\langle\psi|W^s|\psi\rangle = \sum_x |\psi(x)|^2 \sum_{x'} W^s_{xx'} \frac{\psi(x')}{\psi(x)} \tag{6}$$

$$\equiv \sum_x |\psi(x)|^2 \, \theta_{\text{loc}}(x). \tag{7}$$

The goal of our training routine is the maximization of this expectation value, as to obtain the SCGF $\theta(s)$. This is done using variational Monte Carlo: given a parameterized state, $\psi_{\mathcal{W}}$, we sample the current value of $\langle\psi_{\mathcal{W}}|W^s|\psi_{\mathcal{W}}\rangle \equiv \tilde{\theta}(s)$ using N_S samples $\{x_S\}$ drawn according to $|\psi_{\mathcal{W}}|^2$:

$$\tilde{\theta}(s) \approx \frac{1}{N_S} \sum_{x\in\{x_S\}} \theta_{\text{loc}}(x). \tag{8}$$

The gradients of $\tilde{\theta}(s)$ w.r.t. to the variational parameters $\mathcal{W}$ are then calculated as

$$\partial_{\mathcal{W}}\tilde{\theta}(s) = \frac{2}{N_S} \sum_{x\in\{x_S\}} \partial_{\mathcal{W}} \log\psi_{\mathcal{W}}(x) \left(\theta_{\text{loc}}(x) - \tilde{\theta}(s)\right), \tag{9}$$

and the parameters $\mathcal{W}$ are updated as to maximize $\tilde{\theta}(s)$. This is repeated until convergence is achieved, at which point we obtain our best estimate for the SCGF $\theta(s)$.

3 Dynamical activity of the Fredrickson-Andersen model

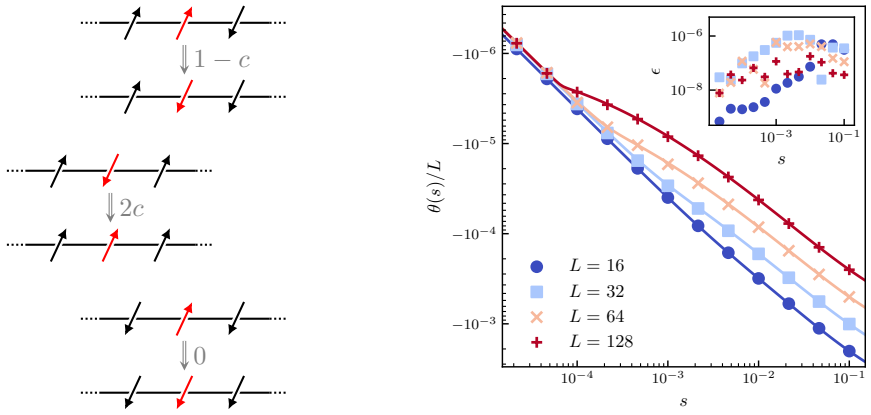
The Fredrickson-Andersen (FA) model describes glassy dynamics by placing kinetic constraints on the transitions between configurations of the system, which gives rise to slow relaxation [63, 64, 65]. On each lattice site, a binary spin can only flip up (resp. down) with rate c (resp. $1-c$) if one of its neighbouring spins is up, and this rate increases for each of its neighbours in the up state, see Fig. 1a. These dynamics are described by the generator W, given by

$$W = \sum_i (n_{i-1} + n_{i+1}) \left[c\sigma_i^+ + (1-c)\sigma_i^- - c(1-n_i) - (1-c)n_i\right], \tag{10}$$

where n_i is 1 if site i is in the up state and zero otherwise, and $\sigma_i^\pm$ flips site i up or down. We work with open boundary conditions by connecting each spin on the boundary of the lattice to a site in the down state. 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 state. In the following, we will study the large-deviation properties of the dynamical activity of the FA model in one dimension. The dynamical activity measures the number of spin flips during a trajectory of time t, so that $\alpha_{xy} = 1$ in Eq. (1). The tilted generator W^s (Eq. (4)) for the dynamical activity of the FA model is then obtained by multiplying the off-diagonal elements of W with the factor e^{-s}, and the corresponding SCGF can be calculated as the largest eigenvalue of W^s. A dynamical phase transition separates an active and inactive dynamical phase at a size-dependent critical value of the tilting parameter $s_c \geq 0$ [2, 32, 66, 67, 68], which is marked by a singularity in the SCGF. The dynamics described by W^s obey detailed balance, so that a similarity transformation $P^{-1}W^sP = H^s$ —which leaves the eigenvalue spectrum of W^s unchanged—can be performed, which results in a Hermitian matrix given by [2]

$$H^s = \sum_i (n_{i-1} + n_{i+1}) \left[e^{-s}\sqrt{c(1-c)}\sigma_i^x - c(1-n_i) - (1-c)n_i\right], \tag{11}$$

where σ_x is a Pauli matrix. As H^s is a Hermitian operator, the SCGF obtained through the variational method results in a lower bound on the exact SCGF—whereas if it were obtained through W^s, the result can be above or below it—allowing for a straightforward comparison of the accuracy of the RNN ansatz with other methods.



(a) (b)

Figure 1: (a) Example transition rates of the one-dimensional Fredrickson-Andersen model studied in this work. (b) Scaled cumulant-generating function $\theta(s)$ per lattice site of the one-dimensional Fredrickson-Andersen model with a recurrent neural network ansatz (marks) and DMRG (lines) for various values of s, and system sizes between $L = 16$ and $L = 128$. Inset: Difference $\epsilon \equiv \theta(s)_{\text{DMRG}}/L - \theta(s)_{\text{RNN}}/L$ between DMRG calculations and the RNN ansatz.

4 Results

We demonstrate the success of the neural-network state ansatz in computing large deviation functions by obtaining the SCGF for the dynamical activity of the FA model on a one-dimensional lattice. We optimize RNN states to find the largest eigenvalue of H^s (Eq. (11)), at $c = 0.1$ and for values of s corresponding to both the active and inactive dynamical phases. The derivaties of the SCGF provide us with e.g. the average value of the dynamical activity or its susceptibility, while the probability distribution of the dynamical activity can be obtained by taking its Legendre transform, allowing us to quantify the rarity of certain values of the activity. Similar results were recently obtained with very high precision using DMRG [32]. The cell type of the RNN used here is a gated recurrent unit [69], and we set $d_h = 128$. The RNN is first optimized on an $L = 16$ chain (Fig. 1b), using $N_S = 5000$ samples and updating the weights using the Adam optimizer [70] with a learning rate $\alpha = 10^{-4}$, which is lowered to $\alpha = 10^{-5}$ until convergence is acquired. Additionally, we enforce parity symmetry [52]. The optimized RNN cell obtained in this way is not explicitly dependent on the lattice size, and can be used as a starting point for training of larger systems belonging to the same dynamical phase [53]. Hence, we can first extensively train the RNN cell on a small lattice, which is computationally cheap for actions such as determining the optimal training hyperparameters and avoiding local minima. These cells are then used as a starting point to obtain the SCGF for large lattices in an iterative fashion; $L = 16 \to L = 32 \to L = 64 \to L = 128$ (Fig. 1b). Each of these steps typically requires only a few hundred training iterations until convergence, so that even large lattices can be studied in a computationally efficient way. In the inset of Fig. 1b, we compare the SCGF obtained with the RNN states to those calculated with DMRG. The error per lattice site made with the RNN ansatz $\epsilon \equiv \theta(s)_{\text{DMRG}}/L - \theta(s)_{\text{RNN}}/L$ is typically very small, $\epsilon \lesssim \mathcal{O}(10^{-6})$—with the largest errors often occurring near the transition point—which demonstrates the capability of the RNN states in representing the steady-state configurational probability distribution for both dynamical phases.

5 Discussion

We have demonstrated the use of neural-network states for obtaining large-deviation functions of dynamical systems. In particular, we have shown that a recurrent neural-network state ansatz is capable of accurately determining the scaled cumulant-generating function of the Fredrickson-Andersen model on large one-dimensional lattices. Similar ansätze have previously been shown to outperform tensor network techniques for two-dimensional quantum systems [51, 52]. Hence, this ansatz could prove powerful in studying large deviation functions of two-dimensional systems which

are relatively unexplored; we are currently investigating this. As improvements are still being made rapidly to this ansatz, we can expect it to play an important role for the study of large deviations of dynamical systems.

Broader Impact

The authors do not foresee their work having any negative societal consequences or causing ethical issues.

Acknowledgments and Disclosure of Funding

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.

References

[1] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. Van Duijvendijk, and F. Van Wijland. Dynamical first-order phase transition in kinetically constrained models of glasses. *Physical Review Letters*, 98(19):195702, 5 2007.

[2] Juan P Garrahan, Robert L Jack, Vivien Lecomte, Estelle Pitard, Kristina van Duijvendijk, and Frédéric van Wijland. First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories. *Journal of Physics A: Mathematical and Theoretical*, 42(7):075007, 1 2009.

[3] Juan P. Garrahan. Aspects of non-equilibrium in classical and quantum systems: Slow relaxation and glasses, dynamical large deviations, quantum non-ergodicity, and open quantum dynamics. *Physica A: Statistical Mechanics and its Applications*, 504:130–154, 8 2018.

[4] Suriyanarayanan Vaikuntanathan, Todd R. Gingrich, and Phillip L. Geissler. Dynamic phase transitions in simple driven kinetic networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 89(6):062108, 6 2014.

[5] Paolo Visco, Andrea Puglisi, Alain Barrat, Emmanuel Trizac, and Frédéric Van Wijland. Fluctuations of power injection in randomly driven granular gases. *Journal of Statistical Physics*, 125(3):529–564, 11 2006.

[6] Guy Bunin, Yariv Kafri, and Daniel Podolsky. Large deviations in boundary-driven systems: Numerical evaluation and effective large-scale behavior. *EPL (Europhysics Letters)*, 99(2):20002, 7 2012.

[7] Jakob Mehl, Thomas Speck, and Udo Seifert. Large deviation function for entropy production in driven one-dimensional systems. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 78(1):011123, 7 2008.

[8] B. Derrida, J. L. Lebowitz, and E. R. Speer. Exact Large Deviation Functional of a Stationary Open Driven Diffusive System: The Asymmetric Exclusion Process. In *Journal of Statistical Physics*, volume 110, pages 775–810. Springer, 3 2003.

[9] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75(12):126001, 11 2012.

[10] Thomas McGrath, Nick S. Jones, Pieter Rein Ten Wolde, and Thomas E. Ouldridge. Biochemical Machines for the Interconversion of Mutual Information and Work. *Physical Review Letters*, 118(2):028101, 1 2017.

[11] Robert L Jack and Peter Sollich. Large deviations of the dynamical activity in the East model: analysing structure in biased trajectories. *Journal of Physics A: Mathematical and Theoretical*, 47(1):015003, 12 2014.

[12] Udo Seifert. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Physical Review Letters*, 95(4):040602, 7 2005.

[13] Christian Maes. Frenetic Bounds on the Entropy Production. *Physical Review Letters*, 119(16):160601, 10 2017.

[14] Thierry Bodineau and Bernard Derrida. Cumulants and large deviations of the current through non-equilibrium steady states. *Comptes Rendus Physique*, 8(5-6):540–555, 6 2007.

[15] Vivien Lecomte, Alberto Imparato, and Frédéric Van Wijland. Current fluctuations in systems with diffusive dynamics, in and out of equilibrium. In *Progress of Theoretical Physics*, volume 184, pages 276–289. Oxford Academic, 3 2010.

[16] Todd R. Gingrich, Jordan M. Horowitz, Nikolay Perunov, and Jeremy L. England. Dissipation Bounds All Steady-State Current Fluctuations. *Physical Review Letters*, 116(12):120601, 3 2016.

[17] Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, 7 2009.

[18] Hugo Touchette and Rosemary J. Harris. Large deviation approach to nonequilibrium systems. *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond*, pages 335–360, 10 2011.

[19] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2 2017.

[20] Cristian Giardinà, Jorge Kurchan, and Luca Peliti. Direct evaluation of large-deviation functions. *Physical Review Letters*, 96(12):120603, 3 2006.

[21] Vivien Lecomte and Julien Tailleur. A numerical approach to large deviations in continuous time. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(03):P03004, 3 2007.

[22] Takahiro Nemoto, Freddy Bouchet, Robert L. Jack, and Vivien Lecomte. Population-dynamics method with a multicanonical feedback control. *Physical Review E*, 93(6):062123, 6 2016.

[23] Ushnish Ray, Garnet Kin Lic Chan, and David T. Limmer. Exact Fluctuations of Nonequilibrium Steady States from Approximate Auxiliary Dynamics. *Physical Review Letters*, 120(21):210602, 5 2018.

[24] Daniel Jacobson and Stephen Whitelam. Direct evaluation of dynamical large-deviation rate functions using a variational ansatz. *Physical Review E*, 100, 2019.

[25] Ushnish Ray and Garnet Kin-Lic Chan. Constructing auxiliary dynamics for nonequilibrium stationary states by variance minimization. *Journal of Chemical Physics*, 152(10):104107, 3 2020.

[26] Stephen Whitelam, Daniel Jacobson, and Isaac Tamblyn. Evolutionary reinforcement learning of dynamical large deviations. *arXiv:1909.00835*, 9 2019.

[27] Tom H E Oakes, Adam Moss, and Juan P Garrahan. A deep learning functional estimator of optimal dynamics for sampling large deviations. *Machine Learning: Science and Technology*, 5 2020.

[28] Dominic C. Rose, Jamie F. Mair, and Juan P. Garrahan. A reinforcement learning approach to rare trajectory sampling. *arXiv:2005.12890*, 5 2020.

[29] Joel L. Lebowitz and Herbert Spohn. A gallavotti-cohen-type symmetry in the large deviation functional for stochastic dynamics. *Journal of Statistical Physics*, 95(1-2):333–365, 4 1999.

[30] Bernard Derrida and Tridib Sadhu. Large deviations conditioned on large deviations I: Markov chain and Langevin equation. *Journal of Statistical Physics*, 176(4):773–805, 8 2019.

[31] Mieke Gorissen, Jef Hooyberghs, and Carlo Vanderzande. Density-matrix renormalization-group study of current and activity fluctuations near nonequilibrium phase transitions. *Physical Review E*, 79(2):020101, 2 2009.

[32] Mari Carmen Bañuls and Juan P. Garrahan. Using Matrix Product States to Study the Dynamical Large Deviations of Kinetically Constrained Models. *Physical Review Letters*, 123(20):200601, 11 2019.

[33] Phillip Helms, Ushnish Ray, and Garnet Kin Lic Chan. Dynamical phase behavior of the single-A nd multi-lane asymmetric simple exclusion process via matrix product states. *Physical Review E*, 100(2):022101, 8 2019.

[34] Alexandra Nagy and Vincenzo Savona. Variational Quantum Monte Carlo Method with a Neural-Network Ansatz for Open Quantum Systems. *Physical Review Letters*, 122(25):250501, 6 2019.

[35] Michael J. Hartmann and Giuseppe Carleo. Neural-Network Approach to Dissipative Quantum Many-Body Dynamics. *Physical Review Letters*, 122(25):250502, 6 2019.

[36] Filippo Vicentini, Alberto Biella, Nicolas Regnault, and Cristiano Ciuti. Variational Neural-Network Ansatz for Steady States in Open Quantum Systems. *Physical Review Letters*, 122(25):250503, 6 2019.

[37] Nobuyuki Yoshioka and Ryusuke Hamazaki. Constructing neural stationary states for open quantum many-body systems. *Physical Review B*, 99(21):214306, 6 2019.

[38] Kenny Choo, Giuseppe Carleo, Nicolas Regnault, and Titus Neupert. Symmetries and Many-Body Excitations with Neural-Network Quantum States. *Physical Review Letters*, 121(16):167204, 10 2018.

[39] Kenny Choo, Titus Neupert, and Giuseppe Carleo. Two-dimensional frustrated J1-J2 model studied with neural network quantum states. *Physical Review B*, 100(12):125124, 9 2019.

[40] Tom Vieijra, Corneel Casert, Jannes Nys, Wesley De Neve, Jutho Haegeman, Jan Ryckebusch, and Frank Verstraete. Restricted Boltzmann Machines for Quantum States with Non-Abelian or Anyonic Symmetries. *Physical Review Letters*, 124(9):097201, 3 2020.

[41] Roger G. Melko, Giuseppe Carleo, Juan Carrasquilla, and J. Ignacio Cirac. Restricted Boltzmann machines in quantum physics. *Nature Physics*, 15(9):887–892, 9 2019.

[42] S. Pilati, E. M. Inack, and P. Pieri. Self-learning projective quantum Monte Carlo simulations guided by restricted Boltzmann machines. *Physical Review E*, 100(4):043301, 10 2019.

[43] Francesco Ferrari, Federico Becca, and Juan Carrasquilla. Neural Gutzwiller-projected variational wave functions. *Physical Review B*, 100(12):125131, 9 2019.

[44] Dan Sehayek, Anna Golubeva, Michael S. Albergo, Bohdan Kulchytskyy, Giacomo Torlai, and Roger G. Melko. Learnability scaling of quantum states: Restricted Boltzmann machines. *Physical Review B*, 100(19):195125, 11 2019.

[45] Tom Westerhout, Nikita Astrakhantsev, Konstantin S. Tikhonov, Mikhail I. Katsnelson, and Andrey A. Bagrov. Generalization properties of neural network approximations to frustrated magnet ground states. *Nature Communications*, 11(1):1–8, 12 2020.

[46] Attila Szabó and Claudio Castelnovo. Neural network wave functions and the sign problem. *arXiv:2002.04613*, 2 2020.

[47] Yusuke Nomura, Andrew S. Darmawan, Youhei Yamaji, and Masatoshi Imada. Restricted Boltzmann machine learning for solving strongly correlated quantum systems. *Physical Review B*, 96(20):205152, 11 2017.

[48] Dong Ling Deng, Xiaopeng Li, and S. Das Sarma. Machine learning topological states. *Physical Review B*, 96(19):195145, 11 2017.

[49] Dong Ling Deng, Xiaopeng Li, and S. Das Sarma. Quantum entanglement in neural network states. *Physical Review X*, 7(2):021021, 5 2017.

[50] Giuseppe Carleo, Yusuke Nomura, and Masatoshi Imada. Constructing exact representations of quantum many-body systems with deep neural networks. *Nature Communications*, 9(1):1–11, 12 2018.

[51] Or Sharir, Yoav Levine, Noam Wies, Giuseppe Carleo, and Amnon Shashua. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *Physical Review Letters*, 124(2), 2 2019.

[52] Mohamed Hibat-Allah, Martin Ganahl, Lauren E. Hayward, Roger G. Melko, and Juan Carrasquilla. Recurrent neural network wave functions. *Physical Review Research*, 2(2):023358, 6 2020.

[53] Christopher Roth. Iterative Retraining of Quantum Spin Models Using Recurrent Neural Networks. *arXiv:2003.06228*, 3 2020.

[54] Giacomo Torlai and Roger G. Melko. Latent Space Purification via Neural Density Operators. *Physical Review Letters*, 120(24):240503, 6 2018.

[55] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, 14(5):447–450, 5 2018.

[56] Giacomo Torlai, Brian Timar, Evert P.L. Van Nieuwenburg, Harry Levine, Ahmed Omran, Alexander Keesling, Hannes Bernien, Markus Greiner, Vladan Vuletić, Mikhail D. Lukin, Roger G. Melko, and Manuel Endres. Integrating Neural Networks with a Quantum Simulator for State Reconstruction. *Physical Review Letters*, 123(23):230504, 12 2019.

[57] Juan Carrasquilla, Giacomo Torlai, Roger G. Melko, and Leandro Aolita. Reconstructing quantum states with generative models. *Nature Machine Intelligence*, 1(3):155–161, 3 2019.

[58] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv:1412.3555*, 12 2014.

[59] Alex Graves. Supervised Sequence Labelling. In *Supervised Sequence Labelling with Recurrent Neural Networks*, pages 5–13. Springer, Berlin, Heidelberg, 2012.

[60] Alex Graves. Generating Sequences With Recurrent Neural Networks. *arXiv:1308.0850*, 8 2013.

[61] Alex Graves, Abdel Rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 6645–6649, 10 2013.

[62] Aäron Van Den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. In *33rd International Conference on Machine Learning, ICML 2016*, volume 4, pages 2611–2620. International Machine Learning Society (IMLS), 1 2016.

[63] Glenn H. Fredrickson and Hans C. Andersen. Kinetic Ising model of the glass transition. *Physical Review Letters*, 53(13):1244–1247, 9 1984.

[64] F. Ritort and P. Sollich. Glassy dynamics of kinetically constrained models. *Advances in Physics*, 52(4):219–342, 6 2003.

[65] Juan P. Garrahan, Peter Sollich, and Cristina Toninelli. Kinetically Constrained Models. *Dynamical Heterogeneities in Glasses, Colloids, and Granular Media*, 9780199691470, 9 2010.

[66] Takahiro Nemoto, Robert L. Jack, and Vivien Lecomte. Finite-Size Scaling of a First-Order Dynamical Phase Transition: Adaptive Population Dynamics and an Effective Model. *Physical Review Letters*, 118(11):115702, 3 2017.

[67] Thierry Bodineau and Cristina Toninelli. Activity phase transition for constrained dynamics. *Communications in Mathematical Physics*, 311(2):357–396, 1 2011.

[68] Thierry Bodineau, Vivien Lecomte, and Cristina Toninelli. Finite Size Scaling of the Dynamical Free-Energy in a Kinetically Constrained Model. *Journal of Statistical Physics*, 147(1):1–17, 4 2012.

[69] Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8)*, pages 103–111, 9 2014.

[70] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, 12 2015.