

Can Neural Quantum States Learn Volume-Law Ground States?

Giacomo Passetti^{1,*}, Damian Hofmann², Pit Neitelei¹, Lukas Grunwald^{1,2}, Michael A. Sentef^{3,2} and Dante M. Kennes^{1,2,†}

¹Institut für Theorie der Statistischen Physik, RWTH Aachen University and

JARA-Fundamentals of Future Information Technology, 52056 Aachen, Germany

²Max Planck Institute for the Structure and Dynamics of Matter, Center for Free-Electron Laser Science (CFEL), Luruper Chaussee 149, 22761 Hamburg, Germany

³H H Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, United Kingdom



(Received 16 December 2022; accepted 27 June 2023; published 17 July 2023)

We study whether neural quantum states based on multilayer feed-forward networks can find ground states which exhibit volume-law entanglement entropy. As a testbed, we employ the paradigmatic Sachdev-Ye-Kitaev model. We find that both shallow and deep feed-forward networks require an exponential number of parameters in order to represent the ground state of this model. This demonstrates that sufficiently complicated quantum states, although being physical solutions to relevant models and not pathological cases, can still be difficult to learn to the point of intractability at larger system sizes. Hence, the variational neural network approach offers no benefits over exact diagonalization methods in this case. This highlights the importance of further investigations into the physical properties of quantum states amenable to an efficient neural representation.

DOI: 10.1103/PhysRevLett.131.036502

Introduction.—The exponential complexity of representing general quantum many-body states is a key challenge in computational quantum physics. To simulate systems beyond small sizes tractable by exact diagonalization methods, it is necessary to find an efficient representation of quantum states of interest. This is made possible by the fact that physically relevant states usually possess a high degree of structure, compared with an arbitrary Hilbert space vector. As a prominent example, ground states of local, gapped Hamiltonians exhibit an area law of the entanglement entropy, i.e., an entanglement entropy that scales like the boundary of the subregion instead of its volume. For systems with a low dimensionality, typically 1D, the area law allows for an efficient representation of the wave function as a matrix product state, which can be simulated by algorithms such as the density matrix renormalization group [1–5].

However, many quantum states of physical interest display a volume law scaling of the entanglement entropy [6], for which generally applicable efficient representations are not known to this date. One class of variational approximations that has been studied to overcome this challenge is neural quantum states (NQS) [7], which are based on an

artificial-neural-network representation of the wave function's probability amplitudes [8–10] and have shown promising results for the study of discrete lattice models even beyond one dimension [11–19]. Notably, it has been shown that a shallow NQS Ansatz is able to efficiently represent quantum states featuring volume-law entanglement [20,21], suggesting that this method could complement tensor network techniques for the purpose of uncovering the physics of highly entangled states. Nevertheless, while for matrix product states and more general tensor-network-based approaches it is known how the entanglement scaling limits the representation capabilities of the ansatz [3], there is so far no analogous physical property that directly relates to the ability of an NQS to learn a given quantum state. Universal approximation theorems, which have been proven for several broad classes of neural networks, guarantee that, in the limit of infinite network size, a neural network ansatz can theoretically represent any continuous function to arbitrary precision [22–25]. Still, these results do not provide bounds on the scaling of the required number of parameters with the system size. For practical applications of NQS, it is thus a central question to determine which classes of quantum many-body states can be efficiently represented that are impossible to tackle with other established variational Ansätze.

In this Letter, we investigate the capabilities of NQS based on shallow and deep feed-forward neural networks (FFNNs) to represent ground states of the Sachdev-Ye-Kitaev (SYK) model [26–28], which is a paradigmatic model for quantum chaos and non-Fermi

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](#). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

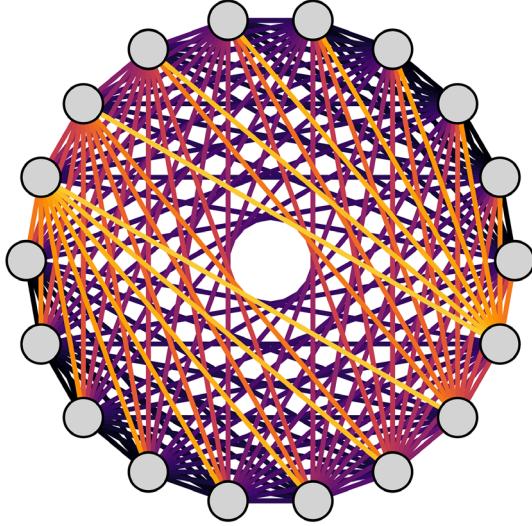


FIG. 1. Cartoon representation of the SYK model. Gray circles represent lattice sites, and every different color shown has two corresponding lines in total connecting four sites. Each color represents one element of the coupling matrix $J_{ij;kl}$ of the SYK model defined by Eq. (1).

liquid behavior [29] and which features a volume-law entanglement in the ground state [30]. Because of these nontrivial properties of the ground state, the SYK model is emerging as a canonical benchmark for variational quantum eigensolvers [31,32]. We present a systematic study of the representation accuracy achieved by the FFNN in dependence of the network hyperparameters. We find an exponential dependence on the system size for the number of network parameters required to learn the SYK ground state. This demonstrates limitations of fully general NQS to learn challenging volume-law ground states of physical interest, paving the way for future studies on scaling properties and boundaries of NQS approaches.

Model.—The SYK model, depicted in Fig. 1, describes strongly correlated fermions on L sites and is defined by the Hamiltonian [26–28]

$$\hat{H}_{\text{SYK}}(J) = \frac{1}{(2L)^{3/2}} \sum_{ijkl} J_{ij;kl} \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_k \hat{c}_l, \quad (1)$$

where $\hat{c}_i^{(\dagger)}$, $i \in \{1, \dots, L\}$, are fermionic ladder operators. The vertices $J_{ij;kl}$ have the symmetry $J_{ij;kl}^* = J_{lk;ji}$ and $J_{ij;kl} = -J_{ji;kl}$ and are random, uncorrelated, all-to-all couplings that are drawn from a Gaussian unitary ensemble (GUE) [33] with mean $\mathbb{E}[J_{ij;kl}] = 0$ and variance $\mathbb{E}[|J_{ij;kl}|^2] = 1$ [29]. Consequently, quantities of physical interest are expectation values over the ensemble of couplings J , which is evaluated after the quantum-expectation value. The ground state of the SYK model describes a strongly correlated non-Fermi liquid without quasiparticle excitations [29], that exhibits volume-law

entanglement entropy [34,35]. In the thermodynamic limit the model becomes self-averaging and exactly solvable, but despite this exact solvability, the ground state is not a Gaussian state, i.e., not a product of single particle wave functions [32]. At finite sizes, particularly studied in the context of quantum chaos [36–38] and experimental realizations [39], no exact solutions are known. Different variational Ansätze to represent the ground state have been proposed recently [32,40]. Here the model can be analyzed by employing approximations, or numerically, by drawing a set of couplings $\{J^{(n)}\}_{n=1}^N$ from the GUE, constructing the corresponding Hamiltonians $\hat{H}_{\text{SYK}}(J^{(n)})$, and solving for the ground states $|\Psi_{\text{GS}}(J^{(n)})\rangle$. Finally, the properties of interest, such as expectation values, are averaged over this ground state ensemble. Because of the self-averaging property of the SYK model, it suffices to evaluate expectation values for a single realization of J in the thermodynamic limit [29]. The self-averaging behavior of the finite sizes simulated together with a comparison with the thermodynamic limit for the ground state energy are discussed in detail in Secs. IV and V of the Supplemental Material [41].

Network architecture.—We use a fully connected FFNN [Figs. 2(a) and 3(a)],

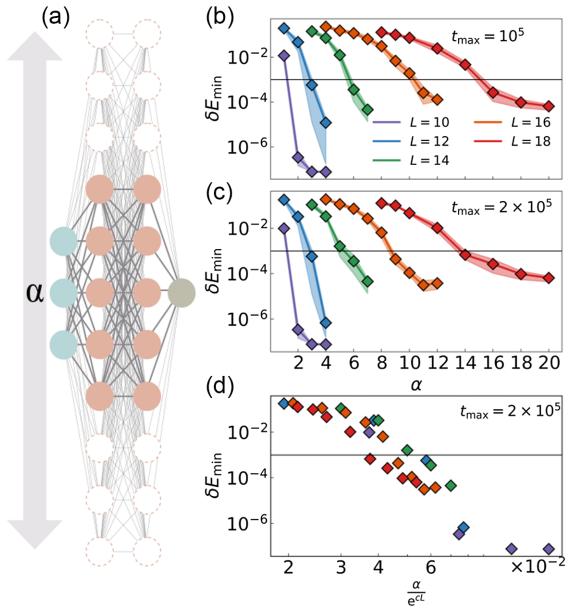


FIG. 2. (a) Shallow fully connected feed-forward neural network, where α denotes the hidden unit density of each layer and thus parametrizes the network width. (b),(c) Relative ground state energy error δE as function of α for several system sizes and random initializations after (b) $t = 10^5$ and (c) $t = 2 \times 10^5$ simulation steps, respectively. The color of each set of data points corresponds to the average over four independent realizations of the network initial weights, for the system size L as indicated in the legend. Black bars indicate $\delta E_{\text{threshold}} = 10^{-3}$. (d) Same data as in panel (c) as a function of rescaled α / e^{cL} . The constant $c \approx 0.33$ was optimized to achieve an approximate collapse of the displayed curves for different system sizes.

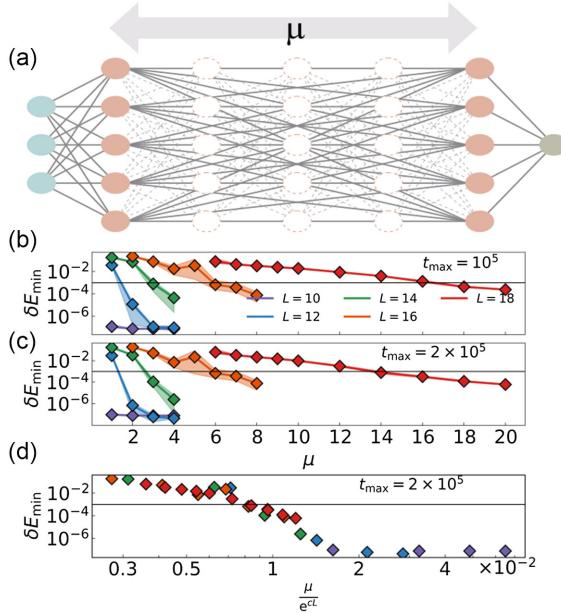


FIG. 3. (a) Deep fully connected feed-forward neural network, where μ denotes the number of layers and thus parametrizes the network depth. (b),(c) Relative energy error δE as a function of μ for several system sizes and random initializations after (b) $t = 10^5$ and (c) $t = 2 \times 10^5$ simulation steps, respectively. The color of each set of data points corresponds to the average over four independent realizations of the network's initial weights, for the system size L as indicated in the legend. Black bars indicate $\delta E_{\text{threshold}} = 10^{-3}$. (d) Same data as in panel (c) as a function of rescaled μ / e^{cL} . The constant $c \approx 0.41$ was optimized to achieve an approximate collapse of the displayed curves for different system sizes.

$$\begin{aligned} F(x) &= f^{(\mu)} \circ \dots \circ f^{(1)}(x), \\ f^{(l)}(y) &= \phi(W^{(l)}y + b^{(l)}), \end{aligned} \quad (2)$$

which is a composition of μ layers $f^{(l)}$, each applying an affine transformation and a scaled exponential linear unit activation function ϕ [43] as pointwise nonlinearity. Each layer has αL neurons, where α is the fixed hidden unit density. The output of the final layer is reduced to a (scalar) log-probability amplitude with respect to the computational basis $\{|x\rangle\}$ by an exponential sum,

$$\log \langle x | \psi_\theta \rangle = \log \sum_{i=1}^{\alpha L} \exp[F_i(x)]. \quad (3)$$

Here, θ denotes the vector of all variational parameters, which contains all entries of the weight matrices $W^{(l)}$ and bias vectors $b^{(l)}$. The variational parameters and therefore network outputs are complex numbers, with the activation function being applied separately to real and imaginary parts. The total number of network parameters scales as $N_{\text{par}} = \mathcal{O}(\mu \alpha^2 L^2)$. We choose the occupation number

basis, as has been done in previous NQS studies of fermionic molecular Hamiltonians [51–53] and target the ground state at half filling ($L/2$ fermions), which corresponds to the global SYK ground state in the $N \rightarrow \infty$ limit. Therefore, the input to the neural network [Eq. (2)] is a vector of occupation numbers $x \in \{0, 1\}^L$ such that $\sum_i x_i = L/2$.

We have verified our results for several variations of this network architecture. In particular, we have evaluated using \tanh as a nonlinear activation function as well as the addition of skip connections, which can be used to counteract the increased training complexity of networks beyond a certain depth [47,48]. These variations did not achieve better results compared with those presented in the main text. Details can be found in Sec. III of the Supplemental Material [41].

Optimization.—The ground state of the network is obtained by numerically minimizing the overlap difference

$$\delta O(\theta, J) = 1 - \left| \frac{\langle \psi_\theta | \psi_{\text{GS}}(J) \rangle}{\langle \psi_\theta | \psi_\theta \rangle} \right| \quad (4)$$

between the variational state $|\psi_\theta\rangle$ and the ground state $|\psi_{\text{GS}}(J)\rangle$ with respect to the variational parameters θ using Adam [46]. We work with system sizes up to $L = 18$ sites, which are accessible via exact diagonalization (ED) and thus enable training using a supervised learning protocol targeting the overlap with the ED ground state $|\psi_{\text{GS}}(J)\rangle$ [45]. The system size allows us to evaluate the loss function [Eq. (4)] by summation over the full Hilbert space (preventing any potential errors arising from Monte Carlo sampling) and to assess the quality of our results using the relative energy error

$$\delta E(\theta; J) = \frac{E(\theta; J) - E_{\text{GS}}(J)}{E_{\text{GS}}(J)} \quad (5)$$

compared with the target ground state energy $E_{\text{GS}}(J) = \langle \psi_{\text{GS}}(J) | \hat{H}_{\text{SYK}}(J) | \psi_{\text{GS}}(J) \rangle$. Details on the optimization scheme are reported in Sec. II of the Supplemental Material [41].

Results.—To start, we discuss the minimum energy error $\delta E_{\min} = \min_{t \in [0, t_{\max}]} \delta E(\theta, J)$ reached within a maximum number of iterations t_{\max} of the optimization protocol. Figures 2(b) and 2(c) show the dependence of δE_{\min} on the network width α for a network with a fixed number of $\mu = 2$ layers, while Figs. 3(b) and 3(c) show the results as a function of network depth μ for deep networks with constant width $\alpha = 4$. We select $\delta E_{\text{threshold}} = 10^{-3}$ as a threshold error to assess successful convergence to the desired ground state. With this threshold, one can see in Figs. 2(b) and 3(b) that at any fixed number of training iterations t_{\max} there is a systematic improvement of the accuracy with respect to increasing both α and μ , as one

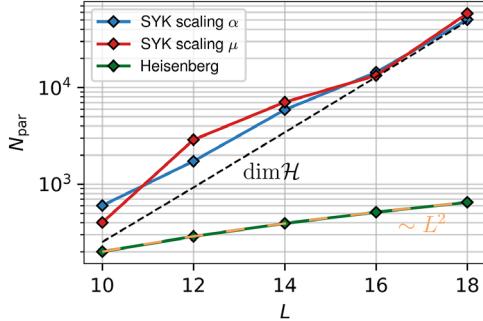


FIG. 4. Minimum number of parameters N_{par} required for the FFNN to learn the ground state of the SYK model as a function of the system size L . Results are shown for the scaling with network width in a shallow ($\mu = 2$) network (blue lines) and for the scaling with network depth for fixed $\alpha = 4$ (red line). In both cases, an exponential scaling in the system size is observed, which matches the scaling of the full Hilbert space dimension $\dim \mathcal{H}$ (dashed line). The N_{par} scaling for the ground state of the Heisenberg model (blue) and the associated quadratic polynomial law are reported for comparison.

would expect given the increased representation capabilities of the network at larger sizes.

See Sec. II A of the Supplemental Material [41] for additional details on this data. Figures 2(d) and 3(d) show the same data for $t = 2 \times 10^5$ as a function of an exponentially rescaled system size. This approximate curve collapse hints at an exponential scaling of the required network size to represent the ground state.

Next, we determine the minimum number of variational parameters at which the network is able to learn the ground state with the desired energy of $\delta E_{\text{threshold}}$. Especially for the smallest system sizes, there is a clear transition between regimes where the network is able or unable to learn the state (in particular as a function of α in the shallow network). For larger system sizes, it is somewhat more difficult to assess convergence. While both very small and very large networks converge to energies above or below the desired threshold within a reasonable optimization time, there is an intermediate regime where the energy gets close to the threshold but only converges at very long timescales. In order to systematically identify a value of α or μ at that boundary, we have developed a criterion used to truncate optimization runs after a reasonable optimization time when those runs are predicted to ultimately converge to a $\delta E(\theta, J)$ higher than $\delta E_{\text{threshold}}$. See Sec. II B of the Supplemental Material [41] for details. In Fig. 4 we show the number of network parameters at the critical α_{\min} or μ_{\min} at which the network is able to reach the target energy accuracy threshold. This allows for a comparison of network expressiveness for both varying width and depth on equal footing. We find that for both the shallow and deep network, an exponentially growing number of parameters is needed to achieve the target energy error. A comparison with the Hilbert space dimension reveals that the network

only reaches this threshold once the number of variational parameters exceeds the number of probability amplitudes contained in the respective state vector. Hence, we find that our deep feed-forward NQS Ansatz as trained here does not learn a more efficient representation of the SYK ground state than the full state vector representation. It is conceivable, in particular given the fully connected nature of our Ansatz, that there is some redundancy in the learned variational parameters, which could be used to achieve a degree of compression after training. In order to investigate this possibility, we have performed a low-rank approximation based on singular value decomposition of the weight matrices [50], the details of which are reported in Sec. VI of the Supplemental Material [41]. This analysis, however, has not revealed such a redundancy.

Our scaling results cannot be interpreted as an immediate consequence of the entanglement scaling of the SYK model, as NQS are known to be able to efficiently represent some volume-law quantum states [20], while they seem to fail for others (as shown here). While a particular realization of the SYK Hamiltonian is of significantly higher complexity than a low-dimensional local lattice Hamiltonian (both because of its fully connected structure and the $\propto L^4$ randomly drawn interaction matrix elements), its ground state still exhibits more structure than a random Hilbert space vector. Since it is well known that deep (and, in fact, already two-layer) networks are able to memorize even completely random data once the number of network parameters exceeds the number of data points [54], these results provide evidence that our FFNN Ansatz does not learn to utilize any of this structure but only manages to learn it as unstructured random data. This is in stark contrast to more structured lattice Hamiltonians, where it is clear from previous works that neural quantum states can approximate ground state energies with subexponential scaling and thus do manage to make use of structure present in the quantum ground state [55,56], although exponential scaling results as a function of real time have been previously found for time-evolved states in a one-dimensional lattice spin model [57]. We have found comparable subexponential behavior when evaluating our training procedure on the ground state of the Heisenberg spin model $\hat{H}_{\text{Heisb}} = \sum_{i=1}^N \sum_{q=1}^3 \hat{\sigma}_i^{(q)} \hat{\sigma}_{i+1}^{(q)}$ on a one-dimensional chain with periodic boundary conditions diagonalized in the same zero-magnetization subspace used for the SYK computations. The scaling of the required number of parameters to reach $\delta E_{\text{threshold}}$ in this model is also reported in Fig. 4. In this case, a relatively small and fixed $\alpha = 1$ and $\mu = 2$ independent of the system size are sufficient to reach this threshold, implying a polynomial scaling of the required number of parameters $N_{\text{par}} = \mathcal{O}(L^2)$. This corresponds to an effective compression of the information contained in the exact state vector and allows one to study sizes beyond those tractable by full

state simulation [7,55]. However, the same approach fails to be useful in the more complex SYK model case.

Discussion.—We have tackled the prototypical SYK model using an NQS variational Ansatz, presenting a systematic study of the ability of deep FFNNs to learn the volume-law entangled ground states of this model. Focusing on the scaling of the required number of parameters to describe the ground state to a desired and fixed accuracy, we find that the size of the FFNN Ansatz needs to grow exponentially in the system size. For practical studies, the neural-network Ansatz is hence outperformed by exact diagonalization with Krylov space methods, which have been applied to systems with up to 30 lattice sites in the context of the SYK model [38,58,59]. With this we show explicitly that the neural network Ansatz is unable to efficiently represent SYK ground states in larger systems in spite of general results raising such hopes. We have performed this analysis using a variety of training techniques (as detailed in the Supplemental Material [41]), showing that the observed scaling is robust to such implementation choices. Investigating whether these training protocols are able to exploit the particle number symmetry, without the need to restrict to a fixed number of particles Hilbert subspace, represents an interesting avenue for future studies.

While the proven capability of random restricted Boltzmann machines to represent volume-law quantum states [20,21] indicates that NQS methods have the potential to tackle problems out of the reach of established tensor-network-based methods, our results demonstrate that the entanglement entropy is not the property that determines whether or not a physical quantum state can be efficiently represented by an NQS. There is still an intriguing open question regarding which additional properties of a physical quantum state determine the efficient applicability of NQS-based methods. NQS Ansätze more specifically tailored to fermionic systems could potentially achieve better scaling [53,60]. Studies in this direction would help elucidate to what extent the nonlocal parity structure inherent to fermionic models [61] affects the learnability of the SYK ground state. Separating this influence from other sources of complexity, such as the lack of spatial structure and the disorder induced by random couplings, and thereby exploring the intermediate region between states that can be learned with compression (such as in the Heisenberg and similar spin models) and states that cannot (such as the SYK results presented here) can provide an improved understanding of the complexity of physical quantum states.

We acknowledge helpful discussions with Giuseppe Carleo, Sebastian Goldt, Javed Lindner, Claudia Merger, Alexandre René, and Attila Szabó. NQS calculations have been performed using NETKET3 [44,62] with JAX[42]. Computations were performed on the HPC system Ada at the Max Planck Computing and Data Facility (MPCDF).

The authors also gratefully acknowledge computing time granted by the JARA Vergabegremium and provided on the JARA partition part of the supercomputer JURECA at Forschungszentrum Jülich [63] under the project ID enhancerg. We acknowledge support by the Max Planck-New York City Center for Nonequilibrium Quantum Phenomena. We acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under RTG 1995 and under Germany's Excellence Strategy—Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1—390534769.

*Corresponding author.

passetti@physik.rwth-aachen.de

†Corresponding author.

dante.kennes@mpsd.mpg.de

- [1] F. Verstraete and J. I. Cirac, *Phys. Rev. B* **73**, 094423 (2006).
- [2] F. Verstraete, V. Murg, and J. Cirac, *Adv. Phys.* **57**, 143 (2008).
- [3] J. Eisert, M. Cramer, and M. B. Plenio, *Rev. Mod. Phys.* **82**, 277 (2010).
- [4] U. Schollwöck, *Ann. Phys. (Amsterdam)* **326**, 96 (2011).
- [5] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, *Rev. Mod. Phys.* **93**, 045003 (2021).
- [6] E. Bianchi, L. Hackl, M. Kieburg, M. Rigol, and L. Vidmar, *PRX Quantum* **3**, 030201 (2022).
- [7] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [8] J. Schmidhuber, *Neural Netw.* **61**, 85 (2015).
- [9] Y. LeCun, Y. Bengio, and G. Hinton, *Nature (London)* **521**, 436 (2015).
- [10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016), <http://www.deeplearningbook.org>.
- [11] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, *Phys. Rev. X* **8**, 011006 (2018).
- [12] S. R. Clark, *J. Phys. A* **51**, 135301 (2018).
- [13] R. Kaubruegger, L. Pastori, and J. C. Budich, *Phys. Rev. B* **97**, 195136 (2018).
- [14] K. Choo, T. Neupert, and G. Carleo, *Phys. Rev. B* **100**, 125124 (2019).
- [15] G. Fabiani and J. H. Mentink, *SciPost Phys.* **7**, 004 (2019).
- [16] M. Schmitt and M. Heyl, *Phys. Rev. Lett.* **125**, 100503 (2020).
- [17] G. Fabiani, M. Bouman, and J. Mentink, *Phys. Rev. Lett.* **127**, 097202 (2021).
- [18] N. Astrakhantsev, T. Westerhout, A. Tiwari, K. Choo, A. Chen, M. H. Fischer, G. Carleo, and T. Neupert, *Phys. Rev. X* **11**, 041021 (2021).
- [19] C. Roth, A. Szabó, and A. MacDonald, High-accuracy variational Monte Carlo for frustrated magnets with deep neural networks, [arXiv:2211.07749](https://arxiv.org/abs/2211.07749).
- [20] D.-L. Deng, X. Li, and S. Das Sarma, *Phys. Rev. X* **7**, 021021 (2017).
- [21] X.-Q. Sun, T. Nebabu, X. Han, M. O. Flynn, and X.-L. Qi, *Phys. Rev. B* **106**, 115138 (2022).
- [22] G. Cybenko, *Math. Control Signals Syst.* **2**, 303 (1989).
- [23] K. Hornik, *Neural Netw.* **4**, 251 (1991).

- [24] A. Pinkus, *Acta Numer.* **8**, 143 (1999).
- [25] P. Kidger and T. Lyons, in Proceedings of Thirty Third Conference on Learning Theory, *Proceedings of Machine Learning Research* Vol. 125, edited by J. Abernethy and S. Agarwal (PMLR, 2020), pp. 2306–2327, <https://proceedings.mlr.press/v125/kidger20a.html>.
- [26] S. Sachdev and J. Ye, *Phys. Rev. Lett.* **70**, 3339 (1993).
- [27] A. Kitaev, A simple model of quantum holography (2015).
- [28] J. Maldacena and D. Stanford, *Phys. Rev. D* **94**, 106002 (2016).
- [29] D. Chowdhury, A. Georges, O. Parcollet, and S. Sachdev, *Rev. Mod. Phys.* **94**, 035004 (2022).
- [30] C. Liu, X. Chen, and L. Balents, *Phys. Rev. B* **97**, 245126 (2018).
- [31] M. B. Hastings and R. O’Donnell, arXiv:2110.10701.
- [32] A. Haldar, O. Tavakol, and T. Scaffidi, *Phys. Rev. Res.* **3**, 023020 (2021).
- [33] *The Oxford Handbook of Random Matrix Theory*, edited by G. Akemann, J. Baik, and P. D. Francesco (Oxford University Press, New York, 2015), 10.1093/oxfordhb/9780198744191.001.0001.
- [34] W. Fu and S. Sachdev, *Phys. Rev. B* **94**, 035135 (2016).
- [35] P. Zhang, arXiv:2203.01513.
- [36] A. Altland and D. Bagrets, *Nucl. Phys.* **B930**, 45 (2018).
- [37] A. M. García-García and J. J. M. Verbaarschot, *Phys. Rev. D* **94**, 126010 (2016).
- [38] B. Kobrin, Z. Yang, G. D. Kahanamoku-Meyer, C. T. Olund, J. E. Moore, D. Stanford, and N. Y. Yao, *Phys. Rev. Lett.* **126**, 030602 (2021).
- [39] M. Brzezinska, Y. Guan, O. V. Yazyev, S. Sachdev, and A. Kruchkov, arXiv:2208.01032 [Phys. Rev. Lett. (to be published)].
- [40] J. Kim, J. Kim, and D. Rosa, *Phys. Rev. Res.* **3**, 023203 (2021).
- [41] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.131.036502> for details on the network architecture, training protocols, and comparisons with different optimization schemes, which includes Refs. [42–50].
- [42] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of Python + NumPy programs (2018), <http://github.com/google/jax>.
- [43] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, arXiv:1706.02515.
- [44] F. Vicentini, D. Hofmann, A. Szabó, D. Wu, C. Roth, C. Giuliani, G. Pescia, J. Nys, V. Vargas-Calderón, N. Astrakhantsev, and G. Carleo, *SciPost Phys. Codebases* **7** (2022).
- [45] B. Jónsson, B. Bauer, and G. Carleo, Neural-network states for the classical simulation of quantum computing, arXiv: 1808.05232.
- [46] D. P. Kingma and J. Ba, arXiv:1412.6980.
- [47] K. He, X. Zhang, S. Ren, and J. Sun, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778, 10.1109/CVPR.2016.90.
- [48] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, in *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 2018, Montréal, Canada*, edited by S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (2018), pp. 6391–6401.
- [49] D. N. Page, *Phys. Rev. Lett.* **71**, 1291 (1993).
- [50] J. Xue, J. Li, and Y. Gong, in *Interspeech 2013* (ISCA, Lyon, 2013).
- [51] K. Choo, A. Mezzacapo, and G. Carleo, *Nat. Commun.* **11**, 2368 (2020).
- [52] P.-J. Yang, M. Sugiyama, K. Tsuda, and T. Yanai, *J. Chem. Theory Comput.* **16**, 3513 (2020).
- [53] J. Hermann, J. Spencer, K. Choo, A. Mezzacapo, W. M. C. Foulkes, D. Pfau, G. Carleo, and F. Noé, Ab-initio quantum chemistry with neural-network wavefunctions, arXiv:2208.12590.
- [54] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, *Commun. ACM* **64**, 107 (2021).
- [55] D. Sehayek, A. Golubeva, M. S. Albergo, B. Kulchytskyy, G. Torlai, and R. G. Melko, *Phys. Rev. B* **100**, 195125 (2019).
- [56] L. L. Viteritti, F. Ferrari, and F. Becca, *SciPost Phys.* **12**, 166 (2022).
- [57] S.-H. Lin and F. Pollmann, *Phys. Status Solidi (b)* **259**, 2100172 (2022).
- [58] G. Gur-Ari, R. Mahajan, and A. Vaezi, *J. High Energy Phys.* **11** (2018) 070.
- [59] B. Bhattacharjee, P. Nandy, and T. Pathak, arXiv:2210.02474.
- [60] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2122059119 (2022).
- [61] F. Verstraete and J. I. Cirac, *J. Stat. Mech.* (2005) P09012.
- [62] G. Carleo, K. Choo, D. Hofmann, J. E. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Mendl, E. van Nieuwenburg, O. O’Reilly, H. Théveniaut, G. Torlai, F. Vicentini, and A. Wietek, *SoftwareX* **10**, 100311 (2019).
- [63] P. Thörnig, *J. Large-Scale Res. Facil.* **7** (2021).