

Solving the Quantum Many-Body Problem with Artificial Neural Networks

Giuseppe Carleo*

Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland

Matthias Troyer

Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland

Quantum Architectures and Computation Group,

Microsoft Research, Redmond, WA 98052, USA and

Station Q, Microsoft Research, Santa Barbara, CA 93106-6105, USA

The challenge posed by the many-body problem in quantum physics originates from the difficulty of describing the non-trivial correlations encoded in the exponential complexity of the many-body wave function. Here we demonstrate that systematic machine learning of the wave function can reduce this complexity to a tractable computational form, for some notable cases of physical interest. We introduce a variational representation of quantum states based on artificial neural networks with variable number of hidden neurons. A reinforcement-learning scheme is then demonstrated, capable of either finding the ground-state or describing the unitary time evolution of complex interacting quantum systems. We show that this approach achieves very high accuracy in the description of equilibrium and dynamical properties of prototypical interacting spins models in both one and two dimensions, thus offering a new powerful tool to solve the quantum many-body problem.

The wave function Ψ is the fundamental object in quantum physics and possibly the hardest to grasp in a classical world. Ψ is a monolithic mathematical quantity that contains all the information on a quantum state, be it a single particle or a complex molecule. In principle, an exponential amount of information is needed to fully encode a generic many-body quantum state. However, Nature often proves herself benevolent, and a wave function representing a *physical* many-body system can be typically characterized by an amount of information much smaller than the maximum capacity of the corresponding Hilbert space. A limited amount of quantum entanglement, as well as the typicality of a small number of physical states, are then the blocks on which modern approaches build upon to solve the many-body Schrödinger's equation with a limited amount of classical resources.

Numerical approaches directly relying on the wave function can either sample a finite number of physically relevant configurations or perform an efficient *compression* of the quantum state. Stochastic approaches, like quantum Monte Carlo (QMC) methods, belong to the first category and rely on probabilistic frameworks typically demanding a positive-semidefinite wave function. [1–3]. Compression approaches instead rely on efficient representations of the wave function, and most notably in terms of matrix product states (MPS) [4–6] or more general tensor networks [7, 8]. Examples of systems where existing approaches fail are however numerous, mostly due to the sign problem in QMC [9], and to the inefficiency of current compression approaches in high-dimensional systems. As a result, despite the striking success of these methods, a large number of unex-

explored regimes exist, including many interesting open problems. These encompass fundamental questions ranging from the dynamical properties of high-dimensional systems [10, 11] to the exact ground-state properties of strongly interacting fermions [12, 13]. At the heart of this lack of understanding lies the difficulty in finding a general strategy to reduce the exponential complexity of the full many-body wave function down to its most essential features [14].

In a much broader context, the problem resides in the realm of dimensional reduction and feature extraction. Among the most successful techniques to attack these problems, artificial neural networks play a prominent role [15]. They can perform exceedingly well in a variety of contexts ranging from image and speech recognition [16] to game playing [17]. Very recently, applications of neural network to the study of physical phenomena have been introduced [18–20]. These have so-far focused on the classification of complex phases of matter, when exact sampling of configurations from these phases is possible. The challenging goal of solving a many-body problem without prior knowledge of exact samples is nonetheless still unexplored and the potential benefits of Artificial Intelligences in this task are at present substantially unknown. It appears therefore of fundamental and practical interest to understand whether an artificial neural network can modify and adapt itself to describe and analyze a quantum system. This ability could then be used to solve the quantum many-body problem in those regimes so-far inaccessible by existing exact numerical approaches.

Here we introduce a representation of the wave function in terms of artificial neural networks specified by a set of internal parameters \mathcal{W} . We present a stochas-

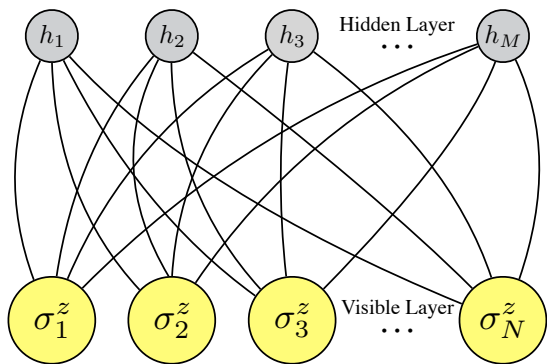


Figure 1. **Artificial Neural network encoding a many-body quantum state of N spins.** Shown is a restricted Boltzmann machine architecture which features a set of N visible artificial neurons (yellow dots) and a set of M hidden neurons (grey dots). For each value of the many-body spin configuration $\mathcal{S} = (\sigma_1^z, \sigma_2^z, \dots, \sigma_N^z)$, the artificial neural network computes the value of the wave function $\Psi(\mathcal{S})$.

tic framework for reinforcement learning of the parameters \mathcal{W} allowing for the best possible representation of both ground-state and time-dependent physical states of a given quantum Hamiltonian \mathcal{H} . The parameters of the neural network are then optimized (trained, in the language of neural networks) either by static variational Monte Carlo (VMC) sampling [21], or in time-dependent VMC [22, 23], when dynamical properties are of interest. We validate the accuracy of this approach studying the Ising and Heisenberg models in both one and two-dimensions. The power of the *neural-network quantum states* (NQS) is demonstrated obtaining state-of-the-art accuracy in both ground-state and out-of-equilibrium dynamics. In the latter case, our approach effectively solves the phase-problem traditionally affecting stochastic Quantum Monte Carlo approaches, since their introduction.

Neural-Network Quantum States — Consider a quantum system with N discrete-valued degrees of freedom $\mathcal{S} = (\mathcal{S}_1, \mathcal{S}_2 \dots \mathcal{S}_N)$, which may be spins, bosonic occupation numbers, or similar. The many-body wave function is a mapping of the N -dimensional set \mathcal{S} to (exponentially many) complex numbers which fully specify the amplitude and the phase of the quantum state. The point of view we take here is to interpret the wave function as a computational black box which, given an input many-body configuration \mathcal{S} , returns a phase and an amplitude according to $\Psi(\mathcal{S})$. Our goal is to approximate this computational black box with a neural network, trained to best represent $\Psi(\mathcal{S})$. Different possible choices for the artificial neural-network architectures have been proposed to solve specific tasks, and the best architecture to describe a many-body quantum system may vary from one case to another. For the sake of concreteness, in the following we specialize our discussion to restricted Boltz-

mann machines (RBM) architectures, and apply them to describe spin 1/2 quantum systems. In this case, RBM artificial networks are constituted by one visible layer of N nodes, corresponding to the physical spin variables in a chosen basis (say for example $\mathcal{S} = \sigma_1^z, \dots, \sigma_N^z$), and a single hidden layer of M auxiliary spin variables ($h_1 \dots h_M$) (see Fig. 1). This description corresponds to a variational expression for the quantum states which reads:

$$\Psi_M(\mathcal{S}; \mathcal{W}) = \sum_{\{h_i\}} e^{\sum_j a_j \sigma_j^z + \sum_i b_i h_i + \sum_{ij} W_{ij} h_i \sigma_j^z},$$

where $h_i = \{-1, 1\}$ is a set of M hidden spin variables, and the weights $\mathcal{W} = \{a_i, b_j, W_{ij}\}$ fully specify the response of the network to a given input state \mathcal{S} . Since this architecture features no intra-layer interactions, the hidden variables can be explicitly traced out, and the wave function reads $\Psi(\mathcal{S}; \mathcal{W}) = e^{\sum_i a_i \sigma_i^z} \times \prod_{i=1}^M F_i(\mathcal{S})$, where $F_i(\mathcal{S}) = 2 \cosh \left[b_i + \sum_j W_{ij} \sigma_j^z \right]$. The network weights are, in general, to be taken complex-valued in order to provide a complete description of both the amplitude and the wave-function's phase.

The mathematical foundations for the ability of NQS to describe intricate many-body wave functions are the numerous established representability theorems [24–26], which guarantee the existence of network approximates of high-dimensional functions, provided a sufficient level of smoothness and regularity is met in the function to be approximated. Since in most physically relevant situations the many-body wave function reasonably satisfies these requirements, we can expect the NQS form to be of broad applicability. One of the practical advantages of this representation is that its quality can, in principle, be systematically improved upon increasing the number of hidden variables. The number M (or equivalently the density $\alpha = M/N$) then plays a role analogous to the bond dimension for the MPS. Notice however that the correlations induced by the hidden units are intrinsically non local in space and are therefore well suited to describe quantum systems in arbitrary dimension. Another convenient point of the NQS representation is that it can be formulated in a symmetry-conserving fashion. For example, lattice translation symmetry can be used to reduce the number of variational parameters of the NQS ansatz, in the same spirit of shift-invariant RBM's [27, 28]. Specifically, for integer hidden variable density $\alpha = 1, 2, \dots$, the weight matrix takes the form of feature filters $W_j^{(f)}$, for $f \in [1, \alpha]$. These filters have a total of αN variational elements *in lieu* of the αN^2 elements of the asymmetric case (see Supp. Mat. for further details).

Given a general expression for the quantum many-body state, we are now left with the task of solving the many-body problem upon machine learning of the network parameters \mathcal{W} . In the most interesting applications the exact many-body state is unknown, and it is typically found upon solution either of the static Schrödinger

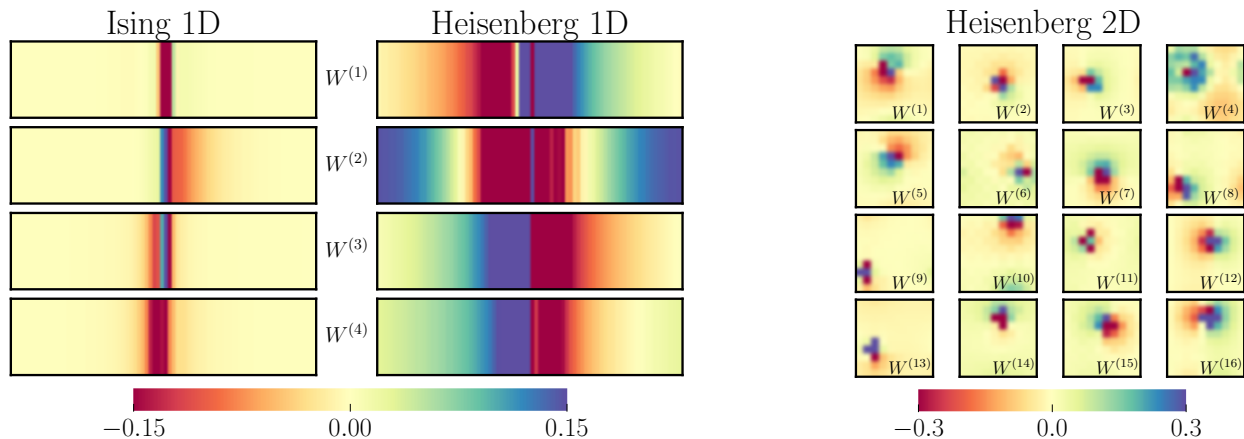


Figure 2. **Neural Network representation of the many-body ground states of prototypical spin models in one and two dimensions.** In the left group of panels we show the feature maps for the one-dimensional TFI model at the critical point $h = 1$, as well as for the AFH model. In both cases the hidden-unit density is $\alpha = 4$ and the lattices comprise 80 sites. Each horizontal colormap shows the values that the f -th feature map $W_j^{(f)}$ takes on the j -th lattice site (horizontal axis, broadened along the vertical direction for clarity). In the right group of panels we show the feature maps for the two-dimensional Heisenberg model on a square lattice, for $\alpha = 16$. In this case the the horizontal (vertical) axis of the colormaps correspond to the $x(y)$ coordinates on a 10×10 square lattice. Each of the feature maps act as effective filters on the spin configurations, capturing the most important quantum correlations.

equation $\mathcal{H}|\Psi\rangle = E|\Psi\rangle$, either of the time-dependent one $i\mathcal{H}|\Psi(t)\rangle = \frac{d}{dt}|\Psi(t)\rangle$, for a given Hamiltonian \mathcal{H} . In the absence of samples drawn according to the exact wave function, supervised learning of Ψ is therefore not a viable option. Instead, in the following we derive a consistent reinforcement learning approach, in which either the ground-state wave function or the time-dependent one are learned on the basis of feedback from variational principles.

Ground State — To demonstrate the accuracy of the NQS in the description of complex many-body quantum states, we first focus on the goal of finding the best neural-network representation of the unknown ground state of a given Hamiltonian \mathcal{H} . In this context, reinforcement learning is realized through minimization of the expectation value of the energy $E(\mathcal{W}) = \langle \Psi_M | \mathcal{H} | \Psi_M \rangle / \langle \Psi_M | \Psi_M \rangle$ with respect to the network weights \mathcal{W} . In the stochastic setting, this is achieved with an iterative scheme. At each iteration k , a Monte Carlo sampling of $|\Psi_M(S; \mathcal{W}_k)|^2$ is realized, for a given set of parameters \mathcal{W}_k . At the same time, stochastic estimates of the energy gradient are obtained. These are then used to propose a next set of weights \mathcal{W}_{k+1} with an improved gradient-descent optimization [29]. The overall computational cost of this approach is comparable to that of standard ground-state Quantum Monte Carlo simulations (see Supp. Material).

To validate our scheme, we consider the problem of finding the ground state of two prototypical spin models, the transverse-field Ising (TFI) model and the anti-ferromagnetic Heisenberg (AFH) model. Their Hamilto-

nians are

$$\mathcal{H}_{\text{TFI}} = -h \sum_i \sigma_i^x - \sum_{\langle i,j \rangle} \sigma_i^z \sigma_j^z \quad (1)$$

and

$$\mathcal{H}_{\text{AFH}} = \sum_{\langle i,j \rangle} \sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y + \sigma_i^z \sigma_j^z, \quad (2)$$

respectively, where $\sigma^x, \sigma^y, \sigma^z$ are Pauli matrices.

In the following, we consider the case of both one and two dimensional lattices with periodic boundary conditions (PBC). In Fig. 2 we show the optimal network structure of the ground states of the two spin models for a hidden variables density $\alpha = 4$ and with imposed translational symmetries. We find that each filter $f = [1, \dots, \alpha]$ learns specific correlation features emerging in the ground state wave function. For example, in the 2D case it can be seen (Fig. 2, rightmost panels) how the neural network learns patterns corresponding to anti-ferromagnetic correlations. The general behavior of the NQS is completely analogous to what observed in convolutional neural networks, where different layers learn specific structures of the input data.

In Fig. 3 we show the accuracy of the NQS states, quantified by the relative error on the ground-state energy $\epsilon_{\text{rel}} = (E_{\text{NQS}}(\alpha) - E_{\text{exact}}) / |E_{\text{exact}}|$, for several values of α and model parameters. In the left panel, we compare the variational NQS energies with the exact result obtained by fermionization of the TFI model, on a one-dimensional chain with PBC. The most striking result is that NQS achieve a controllable and arbitrary

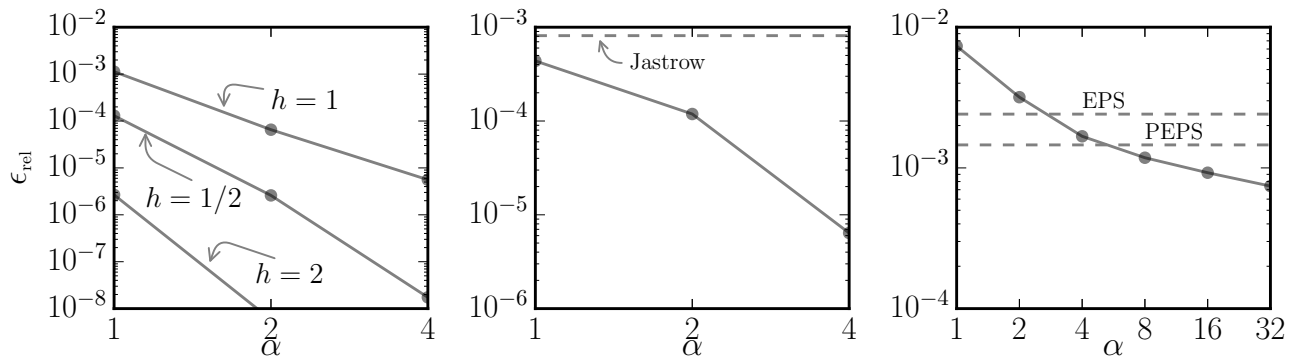


Figure 3. **Finding the many-body ground-state energy with neural-network quantum states.** Shown is the error of the NQS ground-state energy relative to the exact value, for several test cases. Arbitrary precision on the ground-state energy can be obtained upon increasing the hidden units density, α . (Left panel) Accuracy for the one-dimensional TFI model, at a few values of the field strength h , and for a 80 spins chain with PBC. Points below 10^{-8} are not shown to easy readability. (Central panel) Accuracy for the one-dimensional AFH model, for a 80 spins chain with PBC, compared to the Jastrow ansatz (horizontal dashed line). (Right panel) Accuracy for the AFH model on a 10×10 square lattice with PBC, compared to the precision obtained by EPS (upper dashed line) and PEPS (lower dashed line). For all cases considered here the NQS description reaches MPS-grade accuracies in 1D, while it systematically improves the best known variational states for 2D lattice systems.

accuracy which is compatible with a power-law behavior in α . The hardest to learn ground-state is at the quantum critical point $h = 1$, where nonetheless a remarkable accuracy of one part per million can be easily achieved with a relatively modest density of hidden units. The same remarkable accuracy is obtained for the more complex one-dimensional AFH model (central panel). In this case we observe as well a systematic drop in the ground-state energy error, which for a small $\alpha = 4$ attains the same very high precision obtained for the TFI model at the critical point. Our results are compared with the accuracy obtained with the spin-Jastrow ansatz (dashed line in the central panel), which we improve by several orders of magnitude. It is also interesting to compare the value of α with the MPS bond dimension M , needed to reach the same level of accuracy. For example, on the AFH model with PBC, we find that with a standard DMRG implementation [30] we need $M \sim 160$ to reach the accuracy we have at $\alpha = 4$. This points towards a more compact representation of the many-body state in the NQS case, which features about 3 orders of magnitude less variational parameters than the corresponding MPS ansatz.

We next study the AFH model on a two-dimensional square lattice, comparing in the right panel of Fig. 3 to QMC results [31]. As expected from entanglement considerations, the 2D case proves harder for the NQS. Nonetheless, we always find a systematic improvement of the variational energy upon increasing α , qualitatively similar to the 1D case. The increased difficulty of the problem is reflected in a slower convergence. We still obtain results at the level of existing state-of-the-art methods or better. In particular, with a relatively small hid-

den unit density ($\alpha \sim 4$) we already obtain results at the same level than the best known variational ansatz to-date for finite clusters (the EPS of Ref. [32] and the PEPS states of Ref. [33]). Further increasing α then leads to a sizable improvement and consequently yields the best variational results so-far-reported for this 2D model on finite lattices.

Unitary Dynamics — NQS are not limited to ground-state problems but can be extended to the time-dependent Schrödinger equation. For this purpose we define complex-valued and time-dependent network weights $\mathcal{W}(t)$ which at each time t are trained to best reproduce the quantum dynamics, in the sense of the Dirac-Frenkel time-dependent variational principle [34, 35]. In this context, the variational residuals

$$R(t; \dot{\mathcal{W}}(t)) = \text{dist}(\partial_t \Psi(\mathcal{W}(t)), -i\mathcal{H}\Psi) \quad (3)$$

are the objective functions to be minimized as a function of the time derivatives of the weights $\dot{\mathcal{W}}(t)$ (see Supp. Mat.) In the stochastic framework, this is achieved by a time-dependent VMC method [22, 23], which samples $|\Psi_M(S; \mathcal{W}(t))|^2$ at each time and provides the best stochastic estimate of the $\dot{\mathcal{W}}(t)$ that minimize $R^2(t)$, with a computational cost $\mathcal{O}(\alpha N^2)$. Once the time derivatives determined, these can be conveniently used to obtain the full time evolution after time-integration.

To demonstrate the effectiveness of the NQS in the dynamical context, we consider the unitary dynamics induced by quantum quenches in the coupling constants of our spin models. In the TFI model we induce a non-trivial quantum dynamics by means of an instantaneous change in the transverse field: the system is initially prepared in the ground-state of the TFI model for some

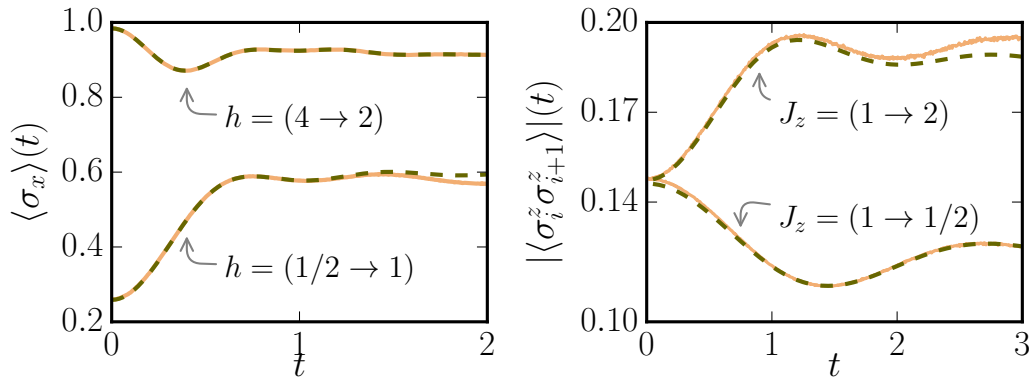


Figure 4. **Describing the many-body unitary time evolution with neural-network quantum states.** Shown are results for the time evolution induced by a quantum quench in the microscopical parameters of the models we study (the transverse field h , for the TFI model and the coupling constant J_z in the AFH model). (Left Panel) NQS results (solid lines) are compared to exact results for the transverse spin polarization in the one-dimensional TFI model (dashed lines). (Right Panel) In the AFH model, the time-dependent nearest-neighbors spin correlations are compared to exact numerical results obtained with t-DMRG for an open one-dimensional chain representative of the thermodynamic limit (dashed lines).

transverse field, h_i , and then let evolve under the action of the TFI Hamiltonian with a transverse field $h_f \neq h_i$. We compare our results with the analytical solution obtained from fermionization of the TFI model for a one-dimensional chain with PBC. In the left panel of Fig. 4 the exact results for the time-dependent transverse spin polarization are compared to NQS with $\alpha = 4$. In the AFH model, we study instead quantum quenches in the longitudinal coupling J_z and monitor the time evolution of the nearest-neighbors correlations. Our results for the time evolution (and with $\alpha = 4$) are compared with the numerically-exact MPS dynamics [36–38] for a system with open boundaries (see Fig. 4, right panel).

The high accuracy obtained also for the unitary dynamics further confirms that neural network-based approaches can be fruitfully used to solve the quantum many-body problem not only for ground-state properties but also to model the evolution induced by a complex set of excited quantum states. It is all in all remarkable that a purely stochastic approach can solve with arbitrary degree of accuracy a class of problems which have been traditionally inaccessible to QMC methods for the past 50 years. The flexibility of the NQS representation indeed allows for an effective solution of the infamous phase problem plaguing the totality of existing exact stochastic schemes based on Feynman’s path integrals.

Outlook — Variational quantum states based on artificial neural networks can be used to efficiently capture the complexity of entangled many-body systems both in one a two dimensions. Despite the simplicity of the restricted Boltzmann machines used here, very accurate results for both ground-state and dynamical properties of prototypical spin models can be readily obtained. Potentially many novel research lines can be envisaged in the near future. For example, the inclusion of the most recent

advances in machine learning, like deep network architectures, might be further beneficial to increase the expressive power of the NQS. Furthermore, the extension of our approach to treat quantum systems other than interacting spins is, in principle, straightforward. In this respect, applications to answer the most challenging questions concerning interacting fermions in two-dimensions can already be anticipated. Finally, at variance with Tensor Network States, the NQS feature intrinsically non-local correlations which can lead to substantially more compact representations of many-body quantum states. A formal analysis of the NQS entanglement properties might therefore bring about substantially new concepts in quantum information theory.

We acknowledge discussions with F. Becca, J.F. Carrasquilla, M. Dolfi, J. Osorio, D. Pataniçœ, and S. Sorella. The time-dependent MPS results have been obtained with the open-source ALPS implementation [30, 39]. This work was supported by the European Research Council through ERC Advanced Grant SIMCOFE by the Swiss National Science Foundation through NCCR QSIT, and by Microsoft Research. This paper is based upon work supported in part by ODNI, IARPA via MIT Lincoln Laboratory Air Force Contract No. FA8721-05-C-0002. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright annotation thereon.

* gcarleo@ethz.ch

- [1] Ceperley, D. & Alder, B. Quantum Monte Carlo. *Science* **231**, 555–560 (1986).
- [2] Foulkes, W. M. C., Mitas, L., Needs, R. J. & Rajagopal, G. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.* **73**, 33–83 (2001).
- [3] Carlson, J. *et al.* Quantum Monte Carlo methods for nuclear physics. *Rev. Mod. Phys.* **87**, 1067–1118 (2015).
- [4] White, S. R. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.* **69**, 2863–2866 (1992).
- [5] Rommer, S. & Ostlund, S. Class of ansatz wave functions for one-dimensional spin systems and their relation to the density matrix renormalization group. *Phys. Rev. B* **55**, 2164–2181 (1997).
- [6] Schollwöck, U. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics* **326**, 96–192 (2011).
- [7] Orús, R. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics* **349**, 117–158 (2014).
- [8] Verstraete, F., Murg, V. & Cirac, J. I. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in Physics* **57**, 143–224 (2008).
- [9] Troyer, M. & Wiese, U.-J. Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations. *Physical Review Letters* **94** (2005).
- [10] Polkovnikov, A., Sengupta, K., Silva, A. & Vengalattore, M. Colloquium: Nonequilibrium dynamics of closed interacting quantum systems. *Reviews of Modern Physics* **83**, 863–883 (2011).
- [11] J. Eisert, M. Friesdorf & C. Gogolin. Quantum many-body systems out of equilibrium. *Nat Phys* **11**, 124–130 (2015).
- [12] Montorsi, A. *The Hubbard Model: A Collection of Reprints* (World Scientific, 1992).
- [13] Thouless, D. J. *The Quantum Mechanics of Many-Body Systems: Second Edition* (New York, 1972), reprint of the academic press edn.
- [14] Freericks, J. K., Nikolić, B. K. & Frieder, O. The nonequilibrium quantum many-body problem as a paradigm for extreme data science. *Int. J. Mod. Phys. B* **28**, 1430021 (2014).
- [15] Hinton, G. E. & Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. *Science* **313**, 504–507 (2006).
- [16] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- [17] Silver, D. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
- [18] Schoenholz, S. S., Cubuk, E. D., Sussman, D. M., Kaxiras, E. & Liu, A. J. A structural approach to relaxation in glassy liquids. *Nat Phys* **12**, 469–471 (2016).
- [19] Carrasquilla, J. & Melko, R. G. Machine learning phases of matter. *arXiv:1605.01735 [cond-mat]* (2016). ArXiv: 1605.01735.
- [20] Wang, L. Discovering Phase Transitions with Unsupervised Learning. *arXiv:1606.00318 [cond-mat, stat]* (2016). ArXiv: 1606.00318.
- [21] McMillan, W. L. Ground State of Liquid He4. *Phys. Rev.* **138**, A442–A451 (1965).
- [22] Carleo, G., Becca, F., Schiro, M. & Fabrizio, M. Localization and Glassy Dynamics Of Many-Body Quantum Systems. *Scientific Reports* **2**, 243 (2012).
- [23] Carleo, G., Becca, F., Sanchez-Palencia, L., Sorella, S. & Fabrizio, M. Light-cone effect and supersonic correlations in one- and two-dimensional bosonic superfluids. *Phys. Rev. A* **89**, 031602 (2014).
- [24] Kolmogorov, A. N. On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. *Doklady Akademii Nauk SSSR* **108**, 179–182 (1961).
- [25] Hornik, K. Approximation capabilities of multilayer feed-forward networks. *Neural Networks* **4**, 251–257 (1991).
- [26] Le Roux, N. & Bengio, Y. Representational Power of Restricted Boltzmann Machines and Deep Belief Networks. *Neural Computation* **20**, 1631–1649 (2008).
- [27] Sohn, K. & Lee, H. Learning Invariant Representations with Local Transformations. 1311–1318 (2012).
- [28] Norouzi, M., Ranjbar, M. & Mori, G. Stacks of convolutional Restricted Boltzmann Machines for shift-invariant feature learning. In *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*, 2735–2742 (2009).
- [29] Sorella, S., Casula, M. & Rocca, D. Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods. *The Journal of Chemical Physics* **127**, 014105 (2007).
- [30] Dolfi, M. *et al.* Matrix product state applications for the ALPS project. *Computer Physics Communications* **185**, 3430–3440 (2014).
- [31] Sandvik, A. W. Finite-size scaling of the ground-state parameters of the two-dimensional Heisenberg model. *Physical Review B* **56**, 11678–11690 (1997).
- [32] Mezzacapo, F., Schuch, N., Boninsegni, M. & Cirac, J. I. Ground-state properties of quantum many-body systems: entangled-plaquette states and variational Monte Carlo. *New J. Phys.* **11**, 083026 (2009).
- [33] Lubasch, M., Cirac, J. I. & Bañuls, M.-C. Algorithms for finite projected entangled pair states. *Phys. Rev. B* **90**, 064425 (2014).
- [34] Dirac, P. a. M. Note on Exchange Phenomena in the Thomas Atom. *Mathematical Proceedings of the Cambridge Philosophical Society* **26**, 376–385 (1930).
- [35] Frenkel, I. *Wave Mechanics: Advanced General Theory*. No. v. 2 in The International series of monographs on nuclear energy: Reactor design physics (The Clarendon Press, 1934).
- [36] White, S. R. & Feiguin, A. E. Real-Time Evolution Using the Density Matrix Renormalization Group. *Phys. Rev. Lett.* **93**, 076401 (2004).
- [37] Vidal, G. Efficient Simulation of One-Dimensional Quantum Many-Body Systems. *Phys. Rev. Lett.* **93**, 040502 (2004).
- [38] Daley, A. J., Kollath, C., Schollwöck, U. & Vidal, G. Time-dependent density-matrix renormalization-group using adaptive effective Hilbert spaces. *Journal of Statistical Mechanics-Theory and Experiment* P04005 (2004).
- [39] Bauer, B. *et al.* The ALPS project release 2.0: open source software for strongly correlated systems. *J. Stat. Mech.* **2011**, P05001 (2011).

- [40] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21**, 1087–1092 (1953).
- [41] Choi, S.-C. T. & Saunders, M. A. Algorithm 937: MINRES-QLP for Symmetric and Hermitian Linear Equations and Least-Squares Problems. *ACM Trans Math Softw* **40** (2014).

Appendix A: Stochastic Optimization For The Ground State

In the first part of our Paper we have considered the goal of finding the best representation of the ground state of a given quantum Hamiltonian \mathcal{H} . The expectation value over our variational states $E(\mathcal{W}) = \langle \Psi_M | \mathcal{H} | \Psi_M \rangle / \langle \Psi_M | \Psi_M \rangle$ is a functional of the network weights \mathcal{W} . In order to obtain an optimal solution for which $\nabla E(\mathcal{W}^*) = 0$, several optimization approaches can be used. Here, we have found convenient to adopt the Stochastic Reconfiguration (SR) method of Sorella et al. [29], which can be interpreted as an effective imaginary-time evolution in the variational subspace. Introducing the variational derivatives with respect to the k -th network parameter,

$$\mathcal{O}_k(\mathcal{S}) = \frac{1}{\Psi_M(\mathcal{S})} \partial_{\mathcal{W}_k} \Psi_M(\mathcal{S}), \quad (\text{A1})$$

as well as the so-called local energy

$$E_{\text{loc}}(\mathcal{S}) = \frac{\langle \mathcal{S} | \mathcal{H} | \Psi_M \rangle}{\Psi_M(\mathcal{S})}, \quad (\text{A2})$$

the SR updates at the p -th iteration are of the form

$$\mathcal{W}(p+1) = \mathcal{W}(p) - \gamma S^{-1}(p) F(p), \quad (\text{A3})$$

where we have introduced the (positive-definite) covariance matrix

$$S_{kk'}(p) = \langle \mathcal{O}_k^* \mathcal{O}_{k'} \rangle - \langle \mathcal{O}_k^* \rangle \langle \mathcal{O}_{k'} \rangle, \quad (\text{A4})$$

the *forces*

$$F_k(p) = \langle E_{\text{loc}} \mathcal{O}_k^* \rangle - \langle E_{\text{loc}} \rangle \langle \mathcal{O}_k^* \rangle, \quad (\text{A5})$$

and a scaling parameter $\gamma(p)$. Since the covariance matrix can be non-invertible, S^{-1} denotes its Moore-Penrose pseudo-inverse. Alternatively, an explicit regularization can be applied, of the form $S_{k,k'}^{\text{reg}} = S_{k,k'} + \lambda(p) \delta_{k,k'} S_{k,k}$. In our work we have preferred the latter regularization, with a decaying parameter $\lambda(p) = \max(\lambda_0 b^p, \lambda_{\text{min}})$ and typically take $\lambda_0 = 100$, $b = 0.9$ and $\lambda_{\text{min}} = 10^{-4}$.

Initially the network weights \mathcal{W} are set to some small random numbers and then optimized with the procedure outlined above. In Fig. 5 we show the typical behavior of the optimization algorithm, which systematically approaches the exact energy upon increasing the hidden units density α .

Appendix B: Time-Dependent Variational Monte Carlo

In the second part of our Paper we have considered the problem of solving the many-body Schrödinger equation with a variational ansatz of the NQS form. This task

can be efficiently accomplished by means of the Time-Dependent Variational Monte Carlo (t-VMC) method of Carleo et al.

In particular, the *residuals*

$$R(t; \dot{\mathcal{W}}(t)) = \text{dist}(\partial_t \Psi(\mathcal{W}(t)), -i\mathcal{H}\Psi) \quad (\text{B1})$$

are a functional of the variational parameters derivatives, $\dot{\mathcal{W}}(t)$, and can be interpreted as the quantum distance between the exactly-evolved state and the variationally evolved one. Since in general we work with unnormalized quantum states, the correct Hilbert-space distance is given by the Fubini-Study metrics, given by

$$\text{dist}_{\text{FS}}(\Phi, \Phi') = \arccos \sqrt{\frac{\langle \Phi' | \Phi \rangle \langle \Phi | \Phi' \rangle}{\langle \Phi' | \Phi' \rangle \langle \Phi | \Phi \rangle}}. \quad (\text{B2})$$

The explicit form of the residuals is then obtained considering $\Phi = \Psi + \delta \partial_t \Psi(\mathcal{W}(t))$ and $\Phi' = \Psi - i\delta \mathcal{H}\Psi(\mathcal{W}(t))$. Taking the lowest order in the time-step δ and explicitly minimizing $\text{dist}_{\text{FS}}(\Phi, \Phi')^2$, yields the equations of motion

$$\dot{\mathcal{W}}(t) = -iS^{-1}(t)F(t), \quad (\text{B3})$$

where the correlation matrix and the forces are defined analogously to the previous section. In this case the diagonal regularization, in general, cannot be applied, and $S^{-1}(t)$ strictly denotes the Moore-Penrose pseudo-inverse.

The outlined procedure is globally stable as also already proven for other wave functions in past works using the t-VMC approach. In Fig. 6 we show the typical behavior of the time-evolved physical properties of interest, which systematically approach the exact results when increasing α .

Appendix C: Efficient Stochastic Sampling

We complete the supplementary information giving an explicit expression for the variational derivatives previously introduced and of the overall computational cost of the stochastic sampling. We start rewriting the NQS in the form

$$\Psi_M(\mathcal{S}) = e^{\sum_i a_i \sigma_i^z} \times \prod_{j=1}^M 2 \cosh \theta_j(\mathcal{S}), \quad (\text{C1})$$

with the effective *angles*

$$\theta_j(\mathcal{S}) = b_j + \sum_i W_{ij} \sigma_i^z. \quad (\text{C2})$$

The derivatives then read

$$\frac{1}{\Psi_M(\mathcal{S})} \partial_{a_i} \Psi_M(\mathcal{S}) = \sigma_i^z, \quad (\text{C3})$$

$$\frac{1}{\Psi_M(\mathcal{S})} \partial_{b_j} \Psi_M(\mathcal{S}) = \tanh[\theta_j(\mathcal{S})], \quad (\text{C4})$$

$$\frac{1}{\Psi_M(\mathcal{S})} \partial_{W_{ij}} \Psi_M(\mathcal{S}) = \sigma_i^z \tanh[\theta_j(\mathcal{S})]. \quad (\text{C5})$$

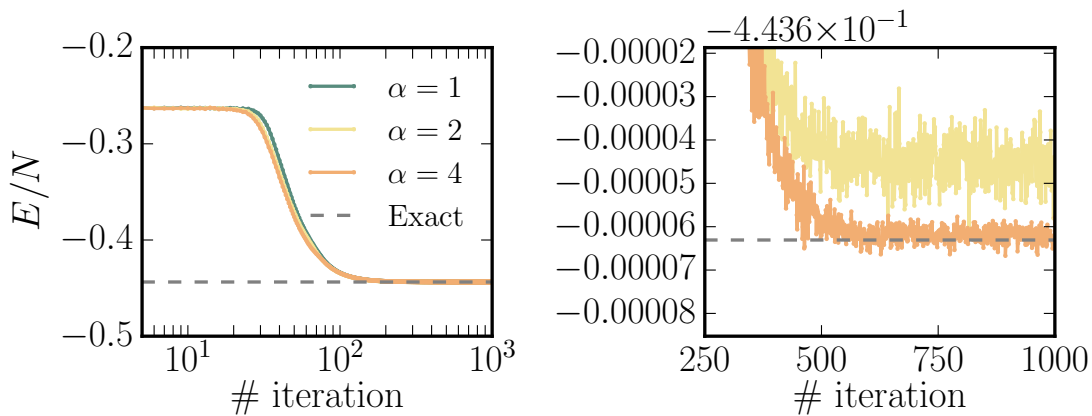


Figure 5. **Convergence properties of the stochastic optimization.** Variational energy for the 1D Heisenberg model as a function of the Stochastic Reconfiguration iterates, and for different values of the hidden units density α . The system has PBC over a chain of $N = 40$ spins. The energy converges smoothly to the exact energy (dashed horizontal line) upon increasing α . In the Left panel we show a complete view of the optimization procedure and on the Right panel a zoom in the neighborhood of the exact energy.

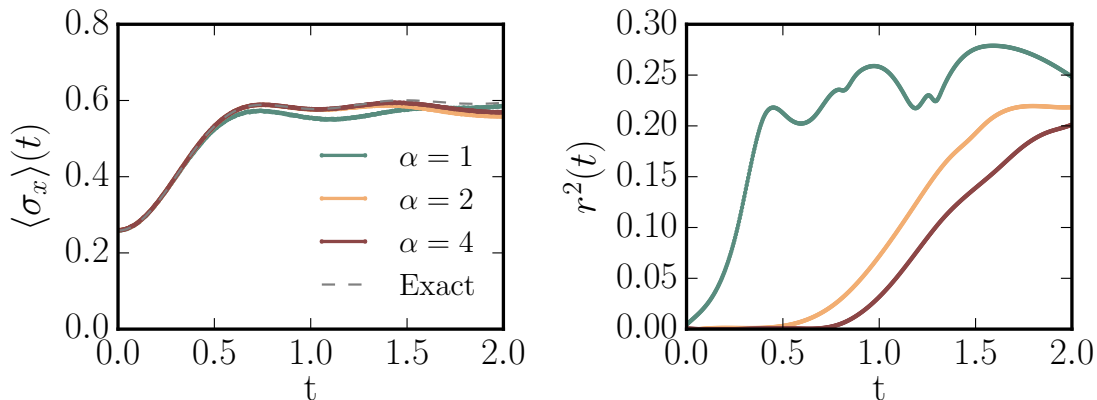


Figure 6. **Convergence properties of the stochastic unitary evolution.** Time-dependent expectation value of the transverse polarization along the x direction in the TFI model, for a quantum quench from $h_i = 1/2$ to the critical interaction $h_f = 1$. t -VMC results are shown for different values of the hidden units density α . The system has periodic boundary conditions over a chain of $N = 40$ spins. (Left panel) The variational curves for the expectation value of the transverse polarization converge smoothly to the exact solution (dashed line) upon increasing α . (Right panel) The relative residual error $r^2(t) = R^2(t)/D_0^2(t)$, where $D_0^2(t) = \text{dist}_{\text{FS}}(\Phi, \Phi - i\delta\mathcal{H})^2$ is shown for different values of the hidden unit density, and it is systematically reduced increasing α .

In our stochastic procedure, we generate a Markov chain of many-body configurations $\mathcal{S}^{(1)} \rightarrow \mathcal{S}^{(2)} \rightarrow \dots \mathcal{S}^{(P)}$ sampling the square modulus of the wave function $|\Psi_M(\mathcal{S})|^2$ for a given set of variational parameters. This task can be achieved through a simple Metropolis-Hastings algorithm [40], in which at each step of the Markov chain a random spin s is flipped and the new configuration accepted according to the probability

$$A(\mathcal{S}^{(k)} \rightarrow \mathcal{S}^{(k+1)}) = \min \left(1, \left| \frac{\Psi_M(\mathcal{S}^{(k+1)})}{\Psi_M(\mathcal{S}^{(k)})} \right|^2 \right). \quad (\text{C6})$$

In order to efficiently compute these acceptances, as well as the variational derivatives, it is useful to keep in mem-

ory look-up tables for the effective angles $\theta_j(\mathcal{S}^{(k)})$ and update them when a new configuration is accepted. These are updated according to

$$\theta_j(\mathcal{S}^{(k+1)}) = \theta_j(\mathcal{S}^{(k)}) - 2W_{kj}\sigma_s^z, \quad (\text{C7})$$

when the spin s has been flipped. The overall cost of a Monte Carlo sweep (i.e. of $\mathcal{O}(N)$ single-spin flip moves) is therefore $\mathcal{O}(N \times M) = \mathcal{O}(\alpha N^2)$. Notice that the computation of the variational derivatives comes at the same computational cost as well as the computation of the local energies after a Monte Carlo sweep.

Appendix D: Iterative Solver

The most time-consuming part of both the SR optimization and of the t-VMC method is the solution of the linear systems (A3 and B3) in the presence of a large number of variational parameters N_{var} . Explicitly forming the correlation matrix S , via stochastic sampling, has a dominant quadratic cost in the number of variational parameters, $\mathcal{O}(N_{\text{var}}^2 \times N_{\text{MC}})$, where N_{MC} denotes the number of Monte Carlo sweeps. However, this cost can be significantly reduced by means of iterative solvers which never form the covariance matrix explicitly. In particular, we adopt the MINRES-QLP method of Choi and Saunders [41], which implements a modified conjugate-gradient iteration based on Lanczos tridiagonalization. This method iteratively computes the pseudo-inverse S^{-1} within numerical precision. The backbone of iterative solvers is, in general, the application of the matrix to be inverted to a given (test) vector. This can be efficiently implemented due to the product structure of the covariance matrix, and determines a dominant complexity of $\mathcal{O}(N_{\text{var}} \times N_{\text{MC}})$ operations for the sparse solver. For example, in the most challenging case when translational symmetry is absent, we have $N_{\text{var}} = \alpha N^2$, and the dominant computational cost for solving (A3 and B3) is in line with the complexity of the previously described Monte Carlo sampling.

Appendix E: Implementing Symmetries

Very often, physical Hamiltonians exhibit intrinsic symmetries which must be satisfied also by their ground- and dynamically-evolved quantum states. These symmetries can be conveniently used to reduce the number of

variational parameters in the NQS.

Let us consider a symmetry group defined by a set of linear transformations T_s , with $s = 1, \dots, S$, such that the spin configurations transform according to $T_s \sigma^z = \tilde{\sigma}^z(s)$. We can enforce the NQS representation to be invariant under the action of T defining

$$\Psi_\alpha(\mathcal{S}; \mathcal{W}) = \sum_{\{h_{i,s}\}} \exp \left[\sum_f a^{(s)} \sum_s \sum_j \tilde{\sigma}_j^z(s) + \sum_f b^{(s)} \sum_s h_{f,s} + \sum_f \sum_s h_{f,s} \sum_j W_j^{(f)} \tilde{\sigma}_j^z(s) \right], \quad (\text{E1})$$

where the network weights have now a different dimension with respect to the standard NQS. In particular, $a^{(f)}$ and $b^{(f)}$ are vectors in the *feature* space with $f = 1, \dots, \alpha_s$ and the connectivity matrix $W_j^{(f)}$ contains $\alpha_s \times N$ elements. Notice that this expression corresponds effectively to a standard NQS with $M = S \times \alpha_s$ hidden variables. Tracing out explicitly the hidden variables, we obtain

$$\Psi_\alpha(\mathcal{S}; \mathcal{W}) = e^{\sum_{f,s,j} a^{(f)} \tilde{\sigma}_j^z(s)} \times \prod_f \prod_s 2 \cosh \left[b^{(f)} + \sum_j W_j^{(f)} \tilde{\sigma}_j^z(s) \right]. \quad (\text{E2})$$

In the specific case of site translation invariance, we have that the symmetry group has an orbit of $S = N$ elements. For a given feature f , the matrix $W_j^{(f)}$ can be seen as a filter acting on the N translated copies of a given spin configuration. In other words, each feature has a pool of N associated hidden variables that act with the same filter on the symmetry-transformed images of the spins.