Quantum many-body dynamics in two dimensions with artificial neural networks

Markus Schmitt\(^1\) and Markus Heyl\(^2\)

\(^1\)Department of Physics, University of California at Berkeley, Berkeley, CA 94720, USA
\(^2\)Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Straße 38, 01187 Dresden, Germany

The efficient numerical simulation of nonequilibrium real-time evolution in isolated quantum matter constitutes a key challenge for current computational methods. This holds in particular in the regime of two spatial dimensions, whose experimental exploration is currently pursued with strong efforts in quantum simulators. In this work we present a versatile and efficient machine learning inspired approach based on a recently introduced artificial neural network encoding of quantum many-body wave functions. We identify and resolve some key challenges for the simulation of time evolution, which previously imposed significant limitations on the accurate description of large systems and long-time dynamics. As a concrete example, we study the dynamics of the paradigmatic two-dimensional transverse field Ising model, as recently also realized experimentally in systems of Rydberg atoms. Calculating the nonequilibrium real-time evolution across a broad range of parameters, we, for instance, observe collapse and revival oscillations of ferromagnetic order and demonstrate that the reached time scales are comparable to or exceed the capabilities of state-of-the-art tensor network methods.

**Introduction.** In the last two decades the field of nonequilibrium quantum many-body systems has seen a rapid development driven, in particular, by the remarkable progress in experiments \([11-14]\). Today, quantum simulators provide access to dynamics in quantum matter with an unprecedented control, which has led to the observation of genuinely nonequilibrium phenomena such as many-body localization \([5, 13, 16]\), discrete time crystals \([11, 17]\), dynamical quantum phase transitions \([7, 18-19]\), or quantum many-body scars \([8]\). A particular frontier pushed forward by experiments recently is towards the non-equilibrium dynamics in two-dimensional (2D) quantum many-body systems \([13, 14, 20, 21]\). The theoretical description of such unitary time evolution, yet, faces severe limitations. For instance, rapid entanglement growth or the exponential cost of contraction impose strong constraints on tensor network approaches. Nevertheless, considerable progress has been reported to capture transient dynamics \([22-29]\). Recently, it has been proposed that machine learning techniques might overcome these difficulties by encoding quantum many-body states in artificial neural networks (ANNs) \([30]\). Up to now it has, however, remained unclear whether this approach can be competitive with other techniques in the description of nonequilibrium quantum dynamics.

In this work we identify numerical instabilities as the key challenge of the real-time ANN approach and we provide prescriptions to mitigate and control their influence. On the one hand our proposed modifications target the structure of the ANN itself, for instance by utilizing deep architectures, i.e., convolutional neural networks. They naturally embody the fundamental physical principles of locality and causality, making the encoding much more efficient. On the other hand we significantly improve the numerical accuracy of solving the underlying time-dependent variational principle (TDVP), which, as we find, is essential to go beyond a transient temporal regime. We apply our approach to the paradigmatic transverse-field Ising model on a square lattice, whose nonequilibrium dynamics has recently been shown to be accessible in systems of Rydberg atoms \([14, 20, 21]\). With our resulting algorithm we obtain numerically exact results up to time scales comparable to or exceeding the capabilities of current tensor network algorithms by comparing to recent data from infinite Projected Entangled

![FIG. 1.](image-url) (a) Schematic illustration of the artificial neural network (ANN) encoding of many-body wave functions in 2D quantum spin systems. A given spin configuration \(s\), blue an red referring to the spin \(\uparrow\) and \(\downarrow\) state, respectively, functions as the input to an ANN whose output at the end is the corresponding wave function amplitude \(\psi_s\). (b) Collapse and revival of the ferromagnetic order in a quantum Ising model of \(8 \times 8\) spins on a square lattice after quenching the transverse field from \(h = 0\) to \(h = 2.63\hbar\). (c) Dynamics of the transverse magnetization \(\langle s^z(t) \rangle\). The quantum Fisher information density \(f_Q(t)\) in (d) reveals that genuine multipartite entanglement is generated by the unitary evolution.
Pair States (iPEPS) [26]. Computing the dynamics for a wide range of parameters, we observe, e.g., collapse and revival oscillations of the ferromagnetic order when strongly quenched by a transverse field, see Fig. 1. Importantly, we find that on the achieved time scales the expressivity of the ANN is not the limiting factor, but rather still some remaining instabilities.

Neural network wave functions. Considering a system of \(N\) spin-1/2 degrees of freedom, the quantum many-body wave function can be represented in the basis of spin configurations \(s=(s_1,s_2,\ldots,s_N)\), \(s_j=\uparrow,\downarrow\), as

\[
|\psi\rangle = \sum_s \psi(s)|s\rangle.
\]

Due to the exponentially large Hilbert space, wave function based numerical methods aiming at large systems need a strategy to avoid storing the individual amplitudes \(\psi(s)\) in memory. In this work we construct a general-purpose variational wave function \(\psi_\eta(s)\), parametrized by \(\eta=(\eta_1,\ldots,\eta_N)\), which constitutes an efficient representation of \(|\psi\rangle\) if \(M\) is much smaller than the Hilbert space size. Being able to provide a good approximation of the amplitudes on the fly \((\psi(s) \approx \psi_\eta(s))\), the variational wave function serves as a generative model, from which we can sample using conventional Monte-Carlo techniques.

Concretely, the expectation value of any observable \(O\) can be obtained as:

\[
\langle \psi_\eta | O | \psi_\eta \rangle = \sum_s |\psi_\eta(s)|^2 O_\eta(s),
\]

with \(O_\eta(s) = \sum_{s'} \langle s|\hat{O}|s'\rangle \psi_\eta(s')/\psi_\eta(s)\). Since \(\langle s|\hat{O}|s'\rangle\) is sparse for few-body observables, the expectation value can be computed efficiently via sampling the probability \(p_\eta(s) = |\psi_\eta(s)|^2\) with Monte Carlo; importantly, there is no sign problem associated with this procedure.

Clearly, it might appear difficult to construct a general-purpose generative machine. However, simple but powerful versions have already been constructed recently for tailored problems [31, 32]. Aiming for a more versatile approach we now follow the proposal to employ artificial neural networks (ANNs) [30]. ANNs have the crucial advantage that they are universal function approximators [33, 34]. As a consequence, any quantum many-body wave function can, in principle, be represented by ANNs provided the network is sufficiently large. Consequently, the network size acts as a control parameter for our simulations that can be used to check convergence of the results. Moreover, the celebrated gradient backpropagation algorithm [30, 38] enables the efficient numerical treatment of this class of variational wave functions.

As one of the key improvements we propose two modifications of the ANN structure compared to previous works. First, we explore deep architectures by means of convolutional neural networks, which naturally respect the fundamental principles of locality and causality. While we provide a detailed description of the CNN wave function in the supplemental material [39], let us point out that CNNs include the Restricted Boltzmann Machines (RBMs), which have been used in previous works for quantum dynamics [40, 41], as the special case of a fully connected single layer CNN with a fixed activation function. By contrast, CNNs are typically constructed as deep networks with sparse connectivity and arbitrary activation functions. For ground-state searches, CNN architectures have already been explored previously [43] with a polynomially enhanced efficiency in encoding entanglement as compared to the RBM [44]. The hidden unit density \(\alpha\), which specifies the size of an RBM [39], corresponds to the number of channels in terms of a CNN architecture. Accordingly, we will denote the size of a CNN with \(L\) layers by a tuple \(\alpha=(\alpha_1,\ldots,\alpha_L)\) with \(\alpha_k\) the number of channels in the \(k\)-th layer [39].

Second, we find that it is crucial for the description of the unitary dynamics to use analytic activation functions for the complex-valued ANNs. In contrast to ground state searches, which are resilient to the encountering of poles and branch cuts of typical activation functions such as \(\ln(cosh(z))\) due to the projective nature of imaginary time evolution, real-time evolution relies on the differentiability of the wave function at any point of the variational manifold in the full complex plane. In our simulations we use the Taylor expansion to \(6\)th order of \(\ln(cosh(z))\) around \(z=0\) as activation function [39].

Training and the TDVP. Training, i.e., optimizing \(\psi_\eta(s)\) to represent the dynamical quantum many-body wave function, is performed using a TDVP [30, 45], which yields a first order differential equation for the variational parameters \(\eta_k(t) \in \mathbb{C}\)

\[
S_{k,k'}\eta_{k'} = iF_k,
\]
where \(S_{k,k'} = \langle \langle O_k O_{k'} \rangle \rangle_c\) and \(F_k = \{\langle O_k^2 E_{\text{loc}} \rangle \rangle_c \) with \(k,k'=1,\ldots,M\) and \(\langle AB \rangle_c = \{\langle AB \rangle \rangle - \langle A \rangle \langle B \rangle \rangle\) a connected correlation function. Here, we introduced the variational derivatives \(O_k(s) = \partial \ln \psi_\eta(s) / \partial \eta_k\) and the local energy \(E_{\text{loc}}(s) = \sum_{s'} \langle s|H|s'\rangle \psi_\eta(s')/\psi_\eta(s)\). The brackets \(\langle \cdot \rangle\) denote expectation values with respect to the normalized probability distribution obtained from \(|\psi_\eta(s)\rangle^2\).

For holomorphic \(\psi_\eta(s)\) Eq. 3 equivalently follows from an action principle or the minimization of the distance between the time-evolved state \(e^{-itH} |\psi_\eta(t)\rangle\) and \(|\psi_{\eta(t)+\delta}\rangle\) as measured by the Fubini-Study metric \(D\) [30, 40, 48]. Considering the Fubini-Study metric provides in addition a practical figure of merit for the accuracy of the resulting time evolution: A small error rate

\[
r^2(t) = \frac{D(|\psi_{\eta(t)}+\delta\rangle, e^{-itH}|\psi_{\eta(t)}\rangle)^2}{D(|\psi_{\eta(t)}\rangle, e^{-itH}|\psi_{\eta(t)}\rangle)^2}
\]
indicates for a small time step \(\delta\) that the variational evolution closely follows the exact dynamics [30]. In the
FIG. 2. Time evolution after quenching a transverse field Ising model of size $N = 8 \times 8$ from the paramagnetically polarized initial state $|\psi_0\rangle = |\uparrow\rangle$ a) into the paramagnetic phase at $h = 2h_c$, b) to the critical point, and c) into the ferromagnetic phase at $h = h_c/10$. For direct comparison the top row includes data obtained with iPEPS from Ref. [26]. The agreement is very good in all cases for the largest networks. The deep CNN for $h_z = 2h_c$ is of size $\alpha = 4,3,2$, in the other cases it is of size $\alpha = 5,4,3$. The second row shows correlation functions along the lattice axis and the bottom row shows the integrated TDVP error $R^2(t)$.

Following we will regard the integrated residual, $R^2(t) = \int_0^t dt' r^2(t')$, as measure of the accuracy of our simulations. Importantly, we demonstrate that the error systematically converges by increasing the neural network size, rendering the approach well controlled. For completeness, we include a derivation of Eq. (3) and the explicit form of the residual [1] in [39].

**Technical aspects.** Besides the proposed modifications of the ANN structure, a further central finding of our work is that the solution of the TDVP inherits instabilities, which can be accounted for and mitigated by two strategies. First of all, due to the non-linearity of the problem, the accurate numerical integration of Eq. (3) can require strongly varying integration step sizes over the course of time ranging over orders of magnitude. Thus, an integrator with adaptive step size is essential. Secondly, the inversion of Eq. (3) is often ill-conditioned and solutions are prone to overfitting Monte Carlo fluctuations. We employed Tikhonov regularization with generalized cross validation to determine the regularization parameter [19]. An alternative approach to be explored in future work would be a regularization scheme based on the L-curve criterion [40].

Finally, the Monte Carlo sampling from the CNN wave function amplitudes is computationally very intense. Hence, an efficient parallel implementation of the sampling is inevitable. The Monte Carlo sampling is straightforwardly parallelizable over many processors of a distributed memory machine using a message passing scheme. The network evaluation allows for a shared memory parallelization using the individual cores of a processor or GPUs. Thereby, this machine learning approach allows us to make full use of the computational resources of cutting edge supercomputers for the simulation of quantum many-body dynamics. Further details are contained in the supplemental material [39].

**Transverse-field Ising model.** As a paradigmatic example of a quantum many-body system we consider the transverse-field Ising model on a 2D square lattice, defined by the Hamiltonian

$$H = -J \sum_{\langle i,j \rangle} \sigma^x_i \sigma^x_j - h \sum_j \sigma^z_j. \quad (5)$$

Here, the $\sigma^{x/z}_i$ denote the Pauli $x$ and $z$ matrices and $\langle i,j \rangle$ is the set of all neighboring sites in the lattice. The model exhibits a quantum phase transition at the critical transverse field $h_c/J = 3.04438(2)$ [71] separating a ferromagnetic phase at $h < h_c$ from a paramagnetic phase. This model has recently developed a particular practical relevance, as it is now naturally realized in Rydberg atom quantum simulators [14,20,21]. Different aspects of its dynamics in 2D have been addressed previously in Refs. [25,26,29,81,52,53].

In the following we demonstrate that the far from equilibrium dynamics induced by quantum quenches can be efficiently simulated using neural network wave functions, independent of the considered parameter regimes. We choose typical initial conditions of quantum simulators, namely uncorrelated product states $|\psi_0\rangle$. After preparation the dynamics generated by the Hamiltonian $H$ yields
the formal solution $|\psi(t)\rangle = e^{-iHt}|\psi_0\rangle$.

**Collapse and revival oscillations.** We start with a quench from a ferromagnetically polarized state $|\psi_0\rangle = |\uparrow\rangle = \prod_{l} |\uparrow_{l}\rangle$ into the paramagnetic phase at $h = 2.63h_c$. The resulting dynamics is shown in Fig. 1(b-d). The order parameter $\langle \sigma_z^i \sigma_z^j \rangle$ initially decays up to a point where it vanishes completely consistent with the expectation that $H$ doesn’t support long-range order. At later times, however, the signal revives with an inverted magnetization. This sequence of decay and inversion continues in a regular pattern. Over time the amplitude of these collapse and revival oscillations decays, which is a consequence of relaxation due to interactions. As one can see in Fig. 1(c) this is accompanied by the oscillatory buildup of a transverse magnetization. Notably, significant entanglement is also generated, see Fig. 1(d), where we show the quantum Fisher information density $f_Q(t) = \frac{1}{\pi} \sum_{i,j} \langle \sigma_z^i \sigma_z^j \rangle c$. After two oscillations of the order parameter, $f_Q(t) > 8$ implying that genuine multipartite entanglement has been developed of at least 9 spins [54, 55]. We checked the accuracy upon increasing the network size and found that a single layer fully connected CNN with $\alpha = 5$ is sufficient for convergence [39].

**Quench from a paramagnetic initial condition.** Next, we consider quenches starting from a paramagnetic initial state $|\psi_0\rangle = |\rangle$. In this case we can compare our results to data obtained recently with an iPEPS algorithm [26].

In Fig. 2 we show results for quenches to weak and strong fields as well as to the quantum critical point, which has previously been identified to constitute a particularly challenging regime for the neural network approach [30, 40]. For large fields $h_c = 2h_c$, we can observe relaxation of the transverse magnetization $\langle \sigma_z^i \rangle$ to a steady state value with remaining temporal fluctuations due to the finite system size. In this regime quantum correlations only develop dominantly for nearest-neighboring spins. For the critical transverse field $h_c = h_c$ the magnetization decays to a much smaller value and significant quantum correlations are generated also at larger distances indicating a strongly correlated state. At weak transverse-fields, a case that has been studied using classical networks recently [31], the dynamics appears more local than in the case of strong fields with quantum correlations emerging mainly between nearest neighbors on the accessible time scales.

Importantly, we find excellent agreement with the dynamics computed using iPEPS for all cases up to the maximally reached times in iPEPS, which are included in Fig. 2 as dashed lines for comparison. While iPEPS directly operates in the thermodynamic limit, the utilized machine learning approach enables us to reach significantly larger times for system sizes up to $N = 8 \times 8$. The direct comparison shows that the system size we reach is sufficient to exclude finite-size effects in local observables up to the time scales reached with iPEPS.

To independently assess the accuracy, we perform our simulations with varying network sizes and architectures. While fully-connected single-layer CNNs are sufficient to reach convergence on time scales similar to or exceeding iPEPS for quenches into the paramagnetic phase or to the critical point, going to a deep CNN with sparse connectivity yields a substantial improvement over the single-layer network for $h = h_c/10$, indicated also by a significant reduction of the TDVP error $R^2(t)$. Consequently, the dynamics remains more local, which can be exploited by using CNNs as we discuss in the following.

**Exploiting locality and causality with the CNN.** Figure 3 visualizes how the build-up of quantum correlations under time evolution is reflected in the network parameters for an RBM and a CNN. In the interest of clarity we consider the dynamics of a one-dimensional (1D) transverse-field Ising model quenched from the uncorrelated paramagnetic state to the critical point. The hierarchical build-up of correlations is displayed in Fig. 3. The color plots in Fig. 3 show snapshots of the magnitudes of the individual couplings of the two networks at three different times. In the RBM the build-up of strong correlations at short distances is reflected in peaks of the coupling magnitudes of the two channels. The couplings at long distances can be regarded as superfluous, they could have been discarded from the beginning. The sparsely connected deep CNN, by contrast,
is constructed such that long distance correlations are mediated through the deep layers. Thereby, locality and causality of the dynamics are exploited and the available variational parameters can be used more efficiently.

**Discussion.** We have shown that variational time evolution of artificial neural network states constitutes a controlled and accurate approach to simulate dynamics in 2D quantum matter, which can be made competitive with current state-of-the-art tensor network algorithms. An alternative tensor network approach besides iPEPS is based on matrix product states and the approximation of 2D systems using cylindrical geometries [25, 29]. For our purpose, however, we chose iPEPS as a reference, because it reflects the full $C_4$ symmetry of the square lattice and we avoid ambiguities caused by finite-size effects in the reference data.

The availability of a versatile numerical method for time evolution paves the way to study the non-equilibrium quantum many-body dynamics in 2D and for new benchmarks of quantum simulators against classical simulations. The main obstacles that restrict the maximal times reached are still numerical issues; importantly, the limitation at this point is not related to the network expressivity. We expect that further refinement similar to other machine learning applications will substantially expand the capabilities of the approach [50–58].

Note: During the preparation of this manuscript we became aware of related work by I. Lopez Gutierrez and C. Mendl, which will appear on the arXiv at the same day.

We thank P. Czarnik, J. Dziarmaga, and P. Corboz for providing the iPEPS data from their work [26]. Moreover, we acknowledge fruitful discussions with J. Budich, M. Bukov, H. Burau, G. Carleo, M. Dupont, and P. Karpov. MS was supported through the Leopoldina Fellowship Programme of the German National Academy of Sciences Leopoldina (LPDS 2018-07) with additional support from the Simons Foundation. MH acknowledges support by the Deutsche Forschungsgemeinschaft via the Gottfried Wilhelm Leibniz Prize program. The numerical simulations were performed at the Max Planck Computing and Data Facility in Garching.

---

[1] M. Greiner, O. Mandel, T. W. Hansch, and I. Bloch, Nature **419**, 51 (2002).
[2] T. Kinoshita, T. Wenger, and D. S. Weiss, Nature **440**, 900 (2006).
[3] I. M. Georgescu, S. Ashhab, and F. Nori, Rev. Mod. Phys. **86**, 153 (2014).
[4] E. A. Martinez, C. A. Muschik, P. Schindler, D. Nigg, A. Erhard, M. Heyl, P. Hauke, M. Dalmonte, T. Monz, P. Zoller, and R. Blatt, Nature **534**, 516 (2016).
[5] J.-y. Choi, S. Hild, J. Zeiher, P. Schauß, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross, Science **352**, 1547 (2016).
[6] C. Gross and I. Bloch, Science **357**, 995 (2017).
[7] P. Jurcevic, H. Shen, P. Hauke, C. Maier, T. Brydges, C. Hempel, B. P. Lanyon, M. Heyl, R. Blatt, and C. F. Roos, Phys. Rev. Lett. **119**, 080501 (2017).
[8] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletič, and M. D. Lukin, Nature **551**, 579 (2017).
[9] J. Zhang, G. Pagano, P. W. Hess, A. Kyriianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, Nature **551**, 601 (2017).
[10] M. Gärtnner, J. G. Böhnert, A. Safavi-Naini, M. L. Wall, J. J. Bollinger, and A. M. Rey, Nature Physics **13**, 781 (2017).
[11] S. Choi, J. Choi, R. Landig, G. Kucsko, H. Zhou, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, V. Khemani, C. von Keyserlingk, N. Y. Yao, E. Demler, and M. D. Lukin, Nature **543**, 221 (2017).
[12] H. Levine, A. Keesing, A. Omran, H. Bernien, S. Schwartz, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, Phys. Rev. Lett. **121**, 123603 (2018).
[13] S. Hild, T. Fukuhara, P. Schauß, J. Zeiher, M. Knap, E. Demler, I. Bloch, and C. Gross, Phys. Rev. Lett. **113**, 147205 (2014).
[14] D. Barredo, V. Lienhard, S. de Léséleuc, T. Lahaye, and A. Browaeys, Nature **561**, 79 (2018).
[15] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch, Science **349**, 842 (2015), arXiv:1501.05661v1 [cond-mat.quant-gas].
[16] J. Smith, A. Lee, P. Richerme, B. Neyenhuis, P. W. Hess, P. Hauke, M. Heyl, D. A. Huse, and C. Monroe, Nature Phys. **12**, 907 (2016).
[17] J. Zhang, P. W. Hess, A. Kyriianidis, P. Becker, A. Lee, J. Smith, G. Pagano, I. D. Potirniche, A. C. Potter, A. Vishwanath, N. Y. Yao, and C. Monroe, Nature **543**, 221 (2017).
[18] J. Zhang, G. Pagano, P. W. Hess, A. Kyriianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z. X. Gong, and C. Monroe, Nature (London) **551**, 601 (2017), arXiv:1708.01044 [quant-ph].
[19] N. Flaschner, D. Vogel, M. Tarnowski, B. S. Rem, D. S. Lühmann, M. Heyl, J. C. Budich, L. Mathey, K. Senfstock, and C. Weitenberg, Nature Physics **14**, 265 (2018).
[20] H. Labuhn, D. Barredo, S. Ravets, S. de Léséleuc, T. Macrì, T. Lahaye, and A. Browaeys, Nature (London) **534**, 667 (2016), arXiv:1509.04543 [cond-mat.quant-gas].
[21] E.Guardado-Sánchez, P. T. Brown, D. Mitra, T. Devakul, D. A. Huse, P. Schauß, and W. S. Bakr, Physical Review X **8**, 021069 (2018).
[22] V. Murg, F. Verstraete, and J. I. Cirac, Phys. Rev. A **75**, 033605 (2007).
[23] M. P. Zaletel, R. S. K. Mong, C. Karrasch, J. E. Moore, and F. Pollmann, Phys. Rev. B **91**, 165112 (2015).
[24] J. Hauschild, F. Pollmann, and F. Heidrich-Meisner, Phys. Rev. A **92**, 053629 (2015).
[25] T. Hashizume, I. P. McCulloch, and J. C. Halimeh, "Dynamical phase transitions in the two-dimensional transverse-field Ising model," (2018), arXiv:1811.09275 [cond-mat.str-el].
[26] P. Czarnik, J. Dziarmaga, and P. Corboz, Phys. Rev. B **106**, 014307 (2017).
[27] C. Hubig and J. I. Cirac, SciPost Phys. 6, 31 (2019).
[28] C. Hubig, A. Bohrdt, M. Knap, F. Grusdt, and J. I. Cirac, “Evaluation of time-dependent correlators after a local quench in ipeps: hole motion in the t-j model,” (2019), arXiv:1911.01159 [cond-mat.str-el].
[29] T. Hashizume, J. C. Halimeh, and I. P. McCulloch, “Hybrid infinite time-evolving block decimation algorithm for long-range multi-dimensional quantum many-body systems,” (2019), arXiv:1910.10726 [cond-mat.str-el].
[30] G. De Tomasi, F. Pollmann, and M. Heyl, Phys. Rev. B 99, 241114 (2019).
[31] M. Schmitt and M. Heyl, SciPost Phys. 4, 013 (2018).
[32] G. Carleo, F. Becca, M. Schirò, and M. Fabrizio, Scientific Reports 2, 243 (2012).
[33] J. Broeckhove, L. Lathouwers, E. Kesteloot, and P. V. Leuven, Chemical Physics Letters 149, 547 (1988).
[34] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pižorn, H. Verschelde, and F. Verstraete, Phys. Rev. Lett. 107, 070601 (2011).
[35] G. Carleo, Spectral And Dynamical PropertiesOf Strongly Correlated Systems, Ph.D. thesis, International School for Advanced Studies (SISSA) (2011).
[36] G. H. Golub, M. Heath, and G. Wahba, Technometrics 21, 215 (1979).
[37] P. Hansen, “The l-curve and its use in the numerical treatment of inverse problems,” (2001) pp. 119–142.
[38] H. W. J. Blöte and Y. Deng, Phys. Rev. E 66, 066110 (2002).
[39] B. Blaß and H. Rieger, Scientific Reports 6, 38185 (2016).
[40] G. H. Golub, M. Heath, and G. Wahba, Technometrics 21, 215 (1979).
Supplemental material to
Quantum many-body dynamics in two dimensions with artificial neural networks
Markus Schmitt and Markus Heyl

CONVOLUTIONAL NEURAL NETWORK WAVE FUNCTION

The convolutional neural network (CNN) is defined as a nested function of alternating non-linear and affine maps, which can be visualized in a layered structure as in Fig. 3 of the main text. In this picture, the input layer is filled with the basis configuration \( s = (s_1, \ldots, s_N) \) in order to obtain the corresponding coefficient \( \psi_\eta(s) \) from the output layer. The term “convolutional” owes to the fact that in the CNN architecture the linear part of the affine map actually resembles a convolution of the previous layer with a filter along the orbit generated by the lattice translations. In particular, the values on the vertices in the \( l \)-th layer are obtained as

\[
v_{c,j}^{(l)}(s) = \sigma \left( \sum_{c'=1}^{\alpha_l} \sum_{k=1}^{N_F} F_{c,c',k} v_{c',k}^{(l-1)}(s) + b_c^{(l)} \right).
\]

Each layer consists of a number of channels labeled by \( c = 1 \ldots \alpha_l \) as indicated in Fig. 3 of the main text by the shaded rectangles in different colors. Each channel has \( N_T \) vertices, where \( N_T \) is the number of elements \( \tau_j \) in the orbit \( T \) generated by all lattice translations. The filters of the convolution are denoted by \( F_{c,c',k}^{(l)} \) involving \( N_F \leq N_T \) vertices of each channel \( c' \) of the previous layer. The affine map can include a bias \( b_c^{(l)} \) for each channel. The filters and biases constitute the set of variational parameters in this ansatz, \( \eta \equiv (F,b) \), and they have to be complex numbers in order to encode the complex-valued wave function. Finally, a non-linear activation function \( \sigma(\cdot) \) is applied to obtain the values of the vertices in the next layer. The recursive relation starts with the initial layer that consists simply of the basis configuration \( v_{1,j}^{(0)}(s) = s_j = \pm 1 \).

To obtain a wave function \( \psi_\eta(s) \) that is invariant under both translations and point symmetries \( \pi \in P \) of a lattice, we consider

\[
\ln \psi_\eta(s) = \sum_{\pi \in P} \sum_{c,j} v_{c,j}^{(L)}(\pi(s)).
\]

Here, we let the CNN encode the logarithm of the wave function coefficients, because in this way the variational derivatives \( O_k(s) \) are directly obtained through the backpropagation algorithm.

In all cases presented in this work single layer CNNs are fully connected, whereas for CNNs with multiple layers we utilized filters that connect to patches of \( 4 \times 4 \) neurons in each channel of the previous layer. As activation function we use the first three non-vanishing terms of the series expansion of \( \ln \cosh(z) \) around \( z = 0 \), i.e.,

\[
\sigma(z) = \frac{z^2}{2} - \frac{z^4}{12} + \frac{z^6}{45}.
\]

When testing lower order polynomials we observed that the expressivity of the network is significantly reduced.

TIME-DEPENDENT VARIATIONAL PRINCIPLE

For the formulation of the time-dependent variational principle (TDVP) we follow Ref. [30] of the main text and consider the distance measured by the Fubini-Study metric between the updated wave function \( |\psi_{\eta+\dot{\eta}\Delta}\rangle \) and the one that is obtained by unitary evolution for a short time \( \Delta \), \( e^{-iH\Delta} |\psi_\eta\rangle \),

\[
\mathcal{D}(|\psi_{\eta+\dot{\eta}\Delta}\rangle, e^{-iH\Delta} |\psi_\eta\rangle)^2 = \arccos \left( \frac{\langle \psi_{\eta+\dot{\eta}\Delta} | e^{-iH\Delta} |\psi_\eta\rangle \langle \psi_\eta | e^{iH\Delta} |\psi_{\eta+\dot{\eta}\Delta}\rangle}{\langle \psi_{\eta+\dot{\eta}\Delta} | \psi_{\eta+\dot{\eta}\Delta}\rangle \langle \psi_\eta | \psi_\eta\rangle} \right) = \arccos \left( \frac{\langle \varphi | \phi \rangle \langle \phi | \varphi \rangle}{\langle \varphi | \varphi \rangle \langle \phi | \phi \rangle} \right)^2
\]

This quantity becomes tractable in the limit of small \( \Delta \), where we can expand

\[
\varphi(s) \equiv (1 - i\Delta E_{\text{loc}}(s))\psi_\eta(s)
\]
and

$$\phi(s) = (1 + \Delta \dot{\eta}_k O_k(s)) \psi_\eta(s).$$  \hfill (11)

Notice, that although we will consider a second order consistent expansion of the Fubini-Study metric below, it is sufficient to keep first order terms in the expression above, because the second order terms will cancel each other. Here, $O_k(s) = \frac{d \log \phi(s)}{d \eta_k}$ denotes the variational derivative.

To abbreviate the notation we write

$$\varphi(s) = (1 - i \Delta E_{\text{loc}}(s)) \psi_0(s) \equiv (1 + \mathcal{E}) \psi_\eta(s)$$  \hfill (12)

and

$$\phi(s) = (1 + \Delta \dot{\eta}_k O_k(s)) \psi_0(s) \equiv (1 + Q) \psi_\eta(s).$$  \hfill (13)

In the following $\bar{\cdot}$ denotes expectation values with respect to $|\psi_\eta(s)|^2$.

Then

$$\begin{align*}
\langle \varphi|\dot{\varphi} \rangle = \frac{(1 + \mathcal{E}^*)(1 + Q) (1 + Q^*)(1 + \mathcal{E})}{(1 + \mathcal{E}^*)(1 + \mathcal{E}^*) (1 + Q)(1 + Q^*)} \\
= \frac{1 + \mathcal{E}^* + \mathcal{E}^* Q + \mathcal{E}^* Q^* + \mathcal{E}^* Q^* + (\mathcal{E} + \mathcal{E}^*) (Q + Q^*)}{(Q + Q^*)^2}
\end{align*}$$  \hfill (14)

Now we expand $1/(1 + x) = -x + x^2 + \mathcal{O}(x^3)$ and keep only terms of $\mathcal{O}(\Delta^2)$ ($\mathcal{E}$ and $Q$ are of $\mathcal{O}(\Delta)$):

$$\begin{align*}
\ldots &= \left(1 + \mathcal{E}^* + Q + \mathcal{E}^* Q + \mathcal{E}^* Q^* + \mathcal{E}^* Q^* + (\mathcal{E} + \mathcal{E}^*) (Q + Q^*) \right) \\
&\quad \times \left(1 - \mathcal{E} - \mathcal{E}^* - \mathcal{E}^* Q - \mathcal{E}^* Q^* - \mathcal{E} - \mathcal{E}^* Q - \mathcal{E}^* Q^* + (\mathcal{E} + \mathcal{E}^* + Q + Q^*)^2 \right) + \mathcal{O}(\Delta^3) \\
&= 1 - (Q^* Q - Q^* Q^*) + (\mathcal{E}^* Q^* - \mathcal{E}^* Q^*) + (\mathcal{E}^* Q - \mathcal{E}^* Q) + (\mathcal{E}^* \mathcal{E} - \mathcal{E}^* \mathcal{E}) + \mathcal{O}(\Delta^3)
\end{align*}$$  \hfill (15)

Notice that at this point second order terms that we dropped already in Eqs. (10) and (11) would have cancelled the same way as $\mathcal{E}$ and $Q$ did in the expression above.

Then the expansion $\arccos(\sqrt{1 + x})^2 = -x + \mathcal{O}(x^2)$, yields

$$\mathcal{D}(\varphi, \phi)^2 = \Delta^2 \left( \dot{\eta}_k^* S_{k,k'} \dot{\eta}_{k'} - F_k \dot{\eta}_k - F_k^* \dot{\eta}_k^* + \text{Var}_{|\psi_\eta|}(\hat{H}) \right) + \mathcal{O}(\Delta^3)$$  \hfill (16)

with

$$S_{k,k'} = \bar{O}_{k} O_{k'} - \bar{O}_{k} \bar{O}_{k'}$$  \hfill (17)

$$F_k = i \left( \bar{E}_{\text{loc}} O_{k} - \bar{E}_{\text{loc}} O_{k'\ast} \right)$$  \hfill (18)

$$\text{Var}_{|\psi_\eta|}(\hat{H}) = \bar{E}_{\text{loc}} E_{\text{loc}}^\ast - \bar{E}_{\text{loc}}^\ast E_{\text{loc}}$$  \hfill (19)

Finally, requiring stationarity of $\mathcal{D}(\varphi, \phi)^2$ with respect to $\dot{\eta}_k$ up to $\mathcal{O}(\Delta^3)$ yields the linear TDVP equation

$$S_{k,k'} \dot{\eta}_{k'} = F_k^*$$  \hfill (20)

Again following Ref. [30] of the main text, we estimate the TDVP error as

$$\gamma^2(t) = \frac{\mathcal{D}(\{|\psi_\eta + \dot{\eta}_\Delta\rangle, e^{-iH\Delta} |\psi_\eta\rangle\})^2}{\mathcal{D}(\{|\psi_\eta\rangle, e^{-iH\Delta} |\psi_\eta\rangle\})^2} = 1 + \frac{\dot{\eta}_k^* R_k - F_k^* \dot{\eta}_k}{\text{Var}_{|\psi_\eta|}(\hat{H})}$$  \hfill (21)

with the residual vector $R_k = S_{k,k'} \dot{\eta}_{k'} - F_k^*$. This expression is again obtained via a second order consistent expansion in powers of $\Delta$. 
DETAILS OF THE NUMERICAL PROCEDURE

1. Computational basis: We always choose the quantization axis along the polarization of the initial state. This facilitates the initialization in the desired state.

2. Initialization: The numerical scheme is very sensitive to a suited initialization of the variational parameters, depending also on the depth of the network. We draw random initial weights, where real and imaginary parts are uniformly distributed on an interval $[-w, w]$. For the single layer networks $w = 10^{-3}$ and for the deep networks $w = 10^{-1}$. Upon random initialization we perform imaginary time evolution with $H = -\sum_j \sigma^x_j$ or $H = -\sum_{\langle i,j \rangle} \sigma^x_i \sigma^x_j - \frac{1}{2} \sum_j \sigma^z_j$ to obtain the paramagnetically or ferromagnetically polarized state, respectively.

3. Adaptive time step: We use a second order consistent numerical integration scheme and estimate errors based on varying step sizes. This reduces the required number of Monte Carlo samplings compared to higher order schemes like the Dormand-Prince method. Using the maximum norm to quantify the error, we find that a tolerance $\epsilon = ||\eta||_\infty \times 10^{-5}$ yields suited integration step sizes. Initially, step sizes as small as $\Delta = 10^{-5}/J$ can be required, because the matrix $S_{k,k'}$ almost vanishes, resulting in large updates of the variational parameters.

4. Monte Carlo sampling: We perform simple Markov Chain Monte Carlo to sample $|\psi(s)|^2$ and find that $N_{MC} = 8 \times 10^5$ samples are sufficient to obtain stable dynamics.

CONVERGENCE CHECKS

A key feature of the neural network approach is the ability to systematically test the accuracy of simulations by comparing results with different network sizes. Here we provide additional data for those cases, for which convergence is not evident from the figures shown in the main text.

1. Collapse and revival of ferromagnetic order (Fig. 1 of the main text): In Fig. 4 we show results for the time evolution of the order parameter $\langle \sigma^z_i \rangle$, Fisher information density $f_\Omega(t)$ and the integrated TDVP error $R^2(t)$ obtained with fully connected single layer CNNs of the size $\alpha = 1$, $\alpha = 3$, and $\alpha = 5$. The results for the order parameter fully coincide for $\alpha = 3$ and $\alpha = 5$, whereas there are small deviations in $f_\Omega(t)$. Going from $\alpha = 1$ to $\alpha = 3$ substantially reduces the TDVP error, which is also reflected in both observables. Remarkably, the inaccurate result obtained with the smallest network exhibits a much larger Fisher information density, which means that entanglement is by far overestimated. This behavior is in stark contrast to tensor network simulations, which systematically underestimate entanglement when the expressivity is insufficient.

2. Quench into the ferromagnetic phase (Fig. 2, right column, of the main text): In Fig. 5 we include results obtained with smaller CNNs to demonstrate convergence up to the maximal time shown. The TDVP error in Fig. 5a) is systematically decreased with increasing network size and, accordingly, the result for the transverse magnetization in Fig. 5a) converges.

FIG. 4. Additional data for the quench from the ferromagnetic product state to $h = 2.633 h_c$ with different network sizes.
FIG. 5. Additional data for the quench from the paramagnetic product state to $h = 0.1h_c$ with CNNs of different size.

FINITE SIZE EFFECT FOR QUENCH TO CRITICAL POINT

For the quench to the critical point our simulations reveal that the finite system size affects the dynamics of local observables already on timescales that are exceeded by the iPEPS simulation if the system is smaller than $N = 8 \times 8$. In Fig. 6 we show results for the dynamics of transverse magnetization after quenching from the paramagnetic product state to the critical point, $h = h_c$, for system sizes $N = 6 \times 6$, $N = 7 \times 7$, and $N = 8 \times 8$. The time interval of agreement with the iPEPS data, which corresponds to dynamics in an infinite system, is extended systematically with increasing system size.

FIG. 6. Dynamics of transverse magnetization after quenching from the paramagnetic product state to the critical point, $h = h_c$, for different system sizes.

PARALLEL COMPUTE PERFORMANCE

Variational time evolution with neural network states can benefit from massively parallel compute resources. The algorithm allows for a hybrid parallelization that exploits distributed memory parallelism of multiple compute nodes as well as shared memory resources available on individual processors. For the example of computing the energy expectation value the hierarchy of parallelism is schematically depicted in Fig. 7. On the top level the Markov Chain Monte Carlo sampling can be distributed over independent processors of a distributed memory machine, exploiting the independence that is inherent to Monte Carlo sampling. For each sample $s_j$ the off-diagonal matrix elements $\langle s_j | \hat{H} | s_j^{(k)} \rangle$ and the connected basis configurations $s_j^{(k)}$ have to be determined, which constitutes a set of independent operations that can be carried out in parallel on individual cores of a processor or on a GPU. Finally, the wave function amplitudes have to be determined for all configurations $s_j$ and $s_j^{(k)}$. This corresponds to a large number of independent network evaluations, where, again, the operations of each individual evaluation are well suited for parallelization on a shared memory unit, especially a GPU. Global communication is only required once in order to perform the sum over all matrix elements.

We realized the parallelization on the level of distributed memory using an implementation of Message Passing Interface (MPI) and found perfect speedup with up to 256 processes when taking $N = 8 \times 10^5$ samples. On the shared memory level we used OpenMP for a CPU implementation and CUDA to alternatively utilize GPU accelerators. In Fig. 8 we show the speedup of both implementations obtained over the serial performance for the network evaluation
and for gradient backpropagation using the test case of a square lattice with 49 spins and a CNN of size \( \alpha = 5, 4, 3 \). Our OpenMP implementation clearly falls behind perfect scaling when using 20 threads. By contrast, both operations are hugely accelerated when using the GPU. In particular, the GPU implementation is 30 times faster than our OpenMP implementation on 20 cores and still 15 times faster than an ideal parallelization on 20 CPU cores.

\[
E = \frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{K_j} \langle s_j | \hat{H} | s_j^{(k)} \rangle \frac{\psi(s_j^{(k)})}{\psi(s_j)}
\]

FIG. 7. Schematic visualization of the parallel implementation of the time evolution algorithm. The top level parallelism of independent Monte Carlo chains can be exploited on a distributed memory machine using a message passing scheme. On the lower level computing matrix elements and evaluating the network for large numbers of input configurations can be parallelized with shared memory using multiple cores of a CPU or GPU accelerators.

FIG. 8. Acceleration of the network evaluation and gradient backpropagation using a GPU in comparison with the OpenMP parallelized version on a system of size \( N = 7 \times 7 \) and with \( \alpha = 5, 4, 3 \). Two typical numbers of configurations were considered, namely 40 for MC sampling and 1000 for the evaluation of expectation values. The CPU timings were obtained on an Intel Xeon E5-2680 and the GPU was an NVIDIA V100.