# Adiabatic Transport of Neural Network Quantum States

Matija Medvidović,[1,*] Alev Orfi,[2,3] Juan Carrasquilla,[1] and Dries Sels[2,4]
[1]*Institute for Theoretical Physics, ETH Zürich, CH-8093 Zürich, Switzerland*
[2]*Center for Computational Quantum Physics, Flatiron Institute, 162 5th Avenue, New York, New York 10010, USA*
[3]*Center for Quantum Phenomena, Department of Physics, New York University, 726 Broadway, New York, New York 10003, USA*
[4]*Department of Physics, Boston University, 590 Commonwealth Ave., Boston, Massachusetts 02215, USA*

(Received 21 October 2025; revised 21 May 2026; accepted 29 July 2026; published 15 September 2026)

Variational methods have offered controllable and powerful tools for capturing many-body quantum physics for decades. The recent introduction of expressive neural network quantum states has enabled the accurate representation of a broad class of complex wavefunctions for many Hamiltonians of interest. We introduce a first-principles method for building neural network representations of many-body excited states by adiabatically continuing eigenstates of simple Hamiltonians into the strongly correlated regime. With controlled access to the full many-body gap, we obtain accurate estimates of critical exponents. Successive eigenstate estimates can be run entirely in parallel, enabling precise targeting of excited-state properties without reference to the rest of the spectrum, opening the door to large-scale numerical investigations of universal properties of entire phases of matter.

DOI: 10.1103/khy4-5v68

## I. INTRODUCTION

Neural quantum states (NQSs) have emerged as efficient and accurate representations of quantum states in many-body systems in recent years. Coinciding with the rapid progress of artificial intelligence for classical learning tasks, the NQS community has adopted and expanded state-of-the-art neural network models and pipelines [1–4]. Variational Monte Carlo (VMC) calculations with NQSs inherit a key advantage from the broader quantum Monte Carlo family of methods—They are *first principles* and do not require an independent dataset [5] to produce quantitative predictions of correlated many-body states.

Initially used as trial states for variational ground-state optimization for spin systems [6,7], they have quickly expanded to fermionic problems on lattices [8–12] and directly in real space [13–21]. Due to the expressive power of neural networks, NQSs have shown particular promise in capturing difficult volume-law states [22], where tensor network approaches [23,24] may struggle [25]. As a consequence, they are increasingly being used as a method of choice to capture nonequilibrium phenomena with a nontrivial sign structure, in calculations at finite temperature [26,27] and for real-time dynamics [28–36].

Despite the rapid progress of NQS methods, access to many-body excited states is often obstructed by prohibitive computational scaling. Excited states contain key information for a range of downstream physical observables such as gaps, critical exponents, and quasiparticle excitation spectra. These states are also a requirement for efficient and accurate impurity solvers for Green's function algorithms like dynamical mean-field theory [37,38] or GW [39–41] at scale. Targeting the $n$th excited state, traditional approaches rely on projecting out contributions from lower states, adding $\sim n^2$ penalty terms to the calculation, and sacrificing accuracy and computational efficiency. Consequently, these methods require the first $n-1$ states before studying the $n$th excited state. More recently, a generalized variational principle was proposed to capture the subspace spanned by the chosen number of variational states [42–44]. This approach allows simultaneous access to all $n$ states, but requires optimizing them collectively, leading to substantially higher computational overhead.

We introduce a method for accessing ground and excited states, based on adiabatic transport [45,46] projected onto a high-dimensional NQS parameter manifold. Starting from exact solutions of simple classical Hamiltonians, states are systematically *dressed* with interactions. Unlike fine-tuning strategies that adjust preoptimized networks for nearby parameters [47,48], our approach directly updates the variational parameters under Hamiltonian changes using efficient Monte Carlo estimators, allowing excited states to be targeted. Crucially, this approach is *parallel* in excited states, allowing for the calculation of higher states without reference to the rest of the spectrum. The resulting finite-size scaling of the gap recovers the expected critical behavior at the transition, yielding precise estimates of the critical exponents and universal gap ratios in the underlying conformal field theory (CFT). By confirming the conclusions of Ref. [49] via direct computation, we demonstrate that adiabatic transport captures the nontrivial operator content of three-dimensional (3D) Ising and Ising* CFTs near the thermodynamic limit. Because the method provides NQS representations at neighboring

*Contact author: mmedvidovic@ethz.ch

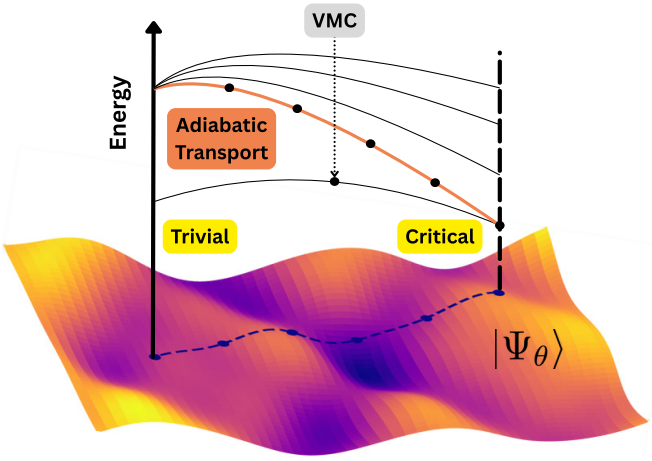


FIG. 1. An overview of the adiabatic transport of an NQS ground state with traditional VMC ground-state optimization. Successive eigenstate approximations trace a curve in the space of parameters $\theta$ of the trial state $\Psi_\theta$.

Hamiltonian parameters, geometric probes can be easily evaluated, offering generic signatures of critical behavior.

## II. METHODS

Consider a many-body quantum system with a Hilbert space $\mathcal{H}$ spanned by an arbitrary computational basis $\{|\mathbf{x}\rangle\}$ and described by a Hamiltonian $H_\lambda$ depending on a real parameter $\lambda$. For small changes in the Hamiltonian parameter, $\lambda \mapsto \lambda + \delta\lambda$, the instantaneous eigenstates and corresponding energies vary smoothly, unless a phase transition is crossed in the thermodynamic limit. While recent work [46] developed adiabatic transport methods for matrix product states, here we introduce a procedure for transporting NQS eigenstates as $\lambda$ is varied, enabling adiabatic transport into critical regions of two-dimensional (2D) systems. Figure 1 illustrates the difference between adiabatic transport of eigenstates and traditional variational ground-state search.

We choose $H_0$ such that its relevant eigenstates can be determined exactly. The $n$th exact eigenstate $|n_0\rangle$ of $H_0$ serves as the initial condition of the adiabatic transport into the critical region, defined by $H_\lambda|n_\lambda\rangle = E_{n,\lambda}|n_\lambda\rangle$. The eigenstate $|n_\lambda\rangle$ for $\lambda > 0$ is approximated by an NQS variational state $|\Psi_\theta\rangle$ by assuming that the network parameters $\theta$ themselves depend on $\lambda$,

$$|n_\lambda\rangle \approx |\Psi_{\theta(\lambda)}\rangle \propto \sum_{\mathbf{x}} \psi_{\theta(\lambda)}(\mathbf{x})|\mathbf{x}\rangle, \tag{1}$$

for a fixed functional form of the unnormalized trial wavefunction $\psi_\theta$.

We introduce the *inverse power iteration* (IPI) method as a solver for updating general variational representations of quantum states. Recently, a similar method has been used as an improved optimizer for ground-state problems in Ref. [50], showing faster convergence than natural gradient descent [51,52] in certain cases. The IPI method uses an approximate target eigenvalue $\omega \approx E$ of a Hamiltonian $H$ to recover the corresponding eigenstate $|\Psi\rangle$ such that $H|\Psi\rangle = E|\Psi\rangle$. Sequential eigenstate estimates are then refined by the well-known shift-and-invert procedure,

$$|\Psi'\rangle \propto (H - \omega)^{-1}|\Psi\rangle, \tag{2}$$

ignoring the normalization factor. Iterating the update in Eq. (2) recovers the eigenstate with the energy $E$ closest to the estimate $\omega$.

In the case of adiabatic transport, we use the IPI solver to propagate an approximate energy $E_\lambda$ and the corresponding eigenstate $|\Psi_{\theta(\lambda)}\rangle$ from some $\lambda$ to $\lambda + \delta\lambda$. We set the target energy $\omega = \omega_\lambda$ to the first-order perturbative estimate

$$\omega_\lambda = E_\lambda + \delta\lambda \, \langle\Psi_{\theta(\lambda)}|\frac{\mathrm{d}H_\lambda}{\mathrm{d}\lambda}|\Psi_{\theta(\lambda)}\rangle, \tag{3}$$

and iterate the IPI scheme until convergence. After sufficiently many iterations, the resulting state $|\Psi'\rangle$ in Eq. (2) is identified with $|\Psi'\rangle \to |\Psi_{\theta+\delta\theta}\rangle$.

By directly substituting $|\Psi'\rangle \to |\Psi_{\theta+\delta\theta}\rangle$ into Eq. (2), the update $\delta\theta$ can be determined, corresponding to a single IPI step on the variational manifold, as derived in Appendix A. This procedure leads to a linear system, $\mathbf{G}\,\delta\theta = -f$, where

$$\begin{aligned} \mathbf{G}_{\mu\nu} &= 2\mathrm{Re}\,\langle\partial_\mu\Psi_\theta|(H_\lambda - \omega_\lambda)|\partial_\nu\Psi_\theta\rangle, \\ f_\mu &= 2\mathrm{Re}\,\langle\partial_\mu\Psi_\theta|H_\lambda|\Psi_\theta\rangle. \end{aligned} \tag{4}$$

Here, the parameters $\theta$ are indexed with Greek indices and $\partial_\mu = \partial/\partial\theta^\mu$. In practice, the variational parameters are updated as $\theta' = \theta + \eta\,\delta\theta$ with an empirically tuned mixing parameter $\eta$ to stabilize the iterations and enforce the condition that each $\delta\theta$ is small.

## III. RESULTS

The adiabatic transport method is validated on the prototypical transverse-field Ising model (TFIM) with periodic boundary conditions on a one-dimensional (1D) chain and a two-dimensional square lattice. The model Hamiltonian is

$$H_\lambda = -\lambda \sum_{\langle i,j\rangle} \sigma_i^z\sigma_j^z - \sum_i \sigma_i^x, \tag{5}$$

where $\sigma_i$ denotes Pauli operators acting on site $i$, and the first sum is taken over all nearest-neighbor pairs $i$ and $j$.

Defining the single-spin state $|+\rangle$ through $\sigma^x|+\rangle = |+\rangle$, the ground state at $\lambda = 0$ takes the simple product form $|0\rangle = \bigotimes_i |+\rangle_i$ over all lattice sites. The single-particle excitation sector is spanned by momentum eigenstates $|\mathbf{k}\rangle = \frac{1}{\sqrt{N}}\sum_i e^{i\mathbf{k}\cdot\mathbf{r}_i}|i\rangle$, where $\mathbf{r}_i$ is the position of site $i$ and $|i\rangle = |-\rangle_i \bigotimes_{j\neq i} |+\rangle_j$ denotes the configuration with a single spin flipped into the $|-\rangle$ state. These states provide exact low-lying excited states at $\lambda = 0$ and are used as initial states for adiabatic transport. Transport of the ground and excited states is performed from a perturbatively small $\lambda_0$ to beyond the known critical points, $\lambda_c = 1$ in one dimension and $\lambda_c \approx 0.329$ in two dimensions [53].

The initial condition $\psi_0(\mathbf{x})$ is built into the wavefunction as $\psi_{\theta(\lambda)}(\mathbf{x}) = (\psi_0(\mathbf{x}))^a \times (\phi_{\theta(\lambda)}(\mathbf{x}))^b$, where $\phi_{\theta(\lambda)}(\mathbf{x})$ is an NQS parametrized by $\theta(\lambda)$. Variational parameters $a$ and $b$ are optimized together with $\theta$, and $\psi_0(\mathbf{x})$ is the exact amplitude at small $\lambda$. The NQS amplitude $\phi_{\theta(\lambda)}(\mathbf{x})$ is constructed as a custom architecture, combining convolutional [54] and residual [55] layers with a top-level restricted Boltzmann machine

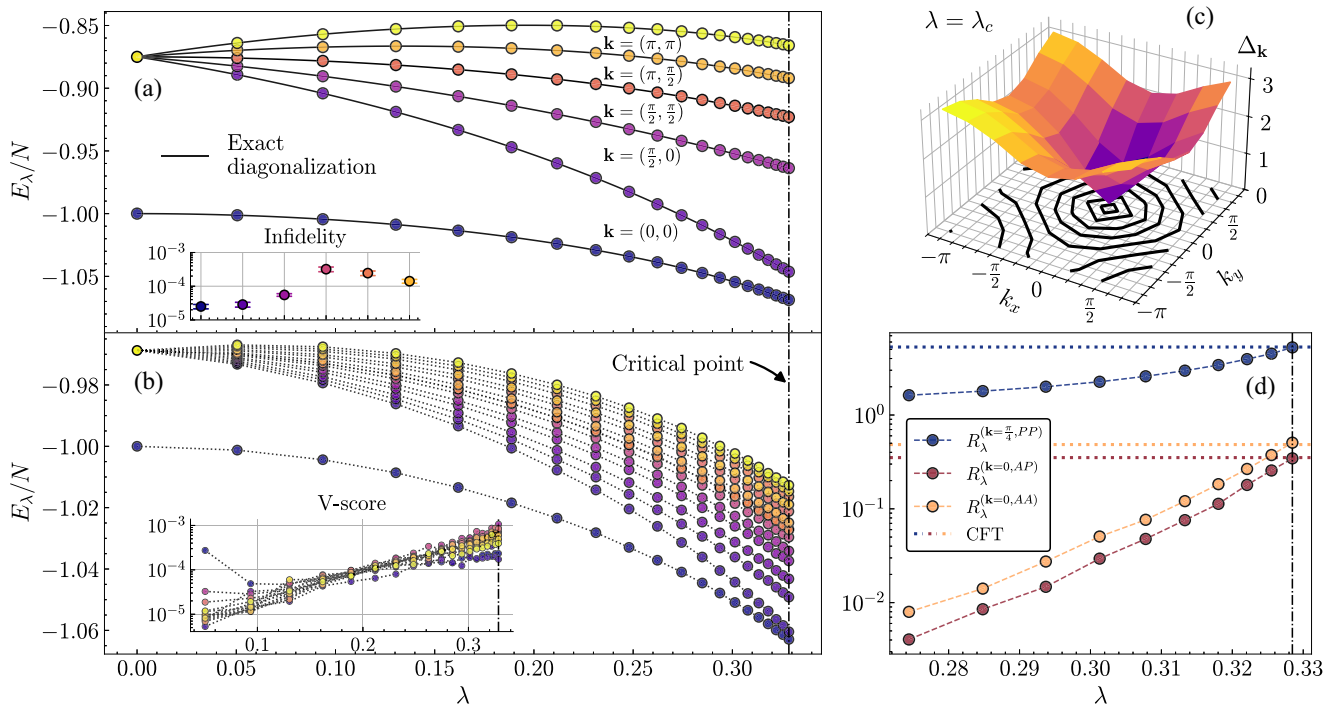


FIG. 2. Adiabatic transport of ground and excited states of the 2D TFIM to the critical point. Panel (a) shows the $4 \times 4$ lattice, with energies compared to exact diagonalization (solid lines) and an inset displaying the average infidelity for one representative eigenstate per degenerate manifold. Panel (b) shows the $8 \times 8$ lattice, with V-scores reported in the inset, and dispersion relation at the critical point shown in panel (c). Panel (d) shows the universal gap ratios $R_\lambda^X$ defined in Eq. (10) for the $8 \times 8$ 2D TFIM, in agreement with the CFT estimates of Ref. [49] (horizontal lines) at the critical point $\lambda_c$.

(RBM) [7,56,57], maintaining the translational invariance of the wavefunction input $\mathbf{x}$ introduced by periodic boundary conditions. We use a residual encoder $f_\alpha$ parametrized by $\alpha$ to generate hidden spins $\mathbf{h} = f_\alpha(\mathbf{x})$ that are used to augment the existing input data as

$$\ln \phi_\theta(\mathbf{x}) = \mathtt{RBM}_\beta([\mathbf{x}, f_\alpha(\mathbf{x})]), \tag{6}$$

where $[\cdots]$ is the concatenation operation and $\theta = \{\alpha, \beta\}$. Further details of the network architecture and transport procedure are detailed in Appendix B and the Supplemental Material [58].

Figure 2(a) shows the adiabatic transport of the ground state and several excited states of the $4 \times 4$ TFIM to the critical point. The corresponding energies faithfully reproduce the results from exact diagonalization (ED) of the Hamiltonian using the Lanczos algorithm. Average infidelities over all $\lambda$ are shown in the inset, demonstrating the quality of the variational representation. For each degenerate manifold, one representative eigenstate is tracked, and its infidelity is evaluated with respect to the exact manifold. The excellent agreement with ED confirms that both ground- and excited-state wavefunctions are reliably captured across all values of $\lambda$.

We observe that energy crossings with states belonging to different discrete symmetry sectors do not destroy the transported state. In fact, energy levels shown in Fig. 2 undergo several crossings with odd-parity excited states that are not shown but are present in the full Hamiltonian spectrum. We also successfully transport states past the critical region in finite systems, indicating that with small enough steps, the method can resolve small gaps. If exact degeneracies appear during transport, parameter update equations become singular. However, we expect that using degenerate perturbation theory at the crossing should be sufficient to recover the transported state.

Far beyond the reach of ED, Fig. 2(b) shows the transport of eigenstates of the $8 \times 8$ lattice. In the absence of ED reference data, we employ the V-score introduced in Ref. [25] as a figure of merit. The V-score is a rescaled, dimensionless energy variance defined as V-score $= N\mathrm{Var}H/(E - E_\infty)^2$, with $N$ being the total number of spins and $E_\infty = 0$ for spin systems such as the TFIM. We obtain reliable energies and wavefunctions for 15 eigenstates computed in parallel, with V-scores remaining below 0.001 even for the highest excited state at criticality. Despite being designed as a figure of merit for ground-state calculations, these V-score values suggest excellent agreement with the target state. Obtained values for higher excited states are comparable with state-of-the-art results for ground states of frustrated magnetic systems or fermionic lattice models.

Using the accurate ground and first excited states, we extract the critical exponents $z$ and $\nu$, which govern the scaling of the correlation length $\xi$ and the energy gap $\Delta$ near the critical point:

$$\xi \sim |\lambda - \lambda_c|^{-\nu} \quad \text{and} \quad \Delta \sim \xi^{-z}. \tag{7}$$

Despite diverging at criticality in the thermodynamic limit, the correlation length is bounded by the system's linear size $L$ for

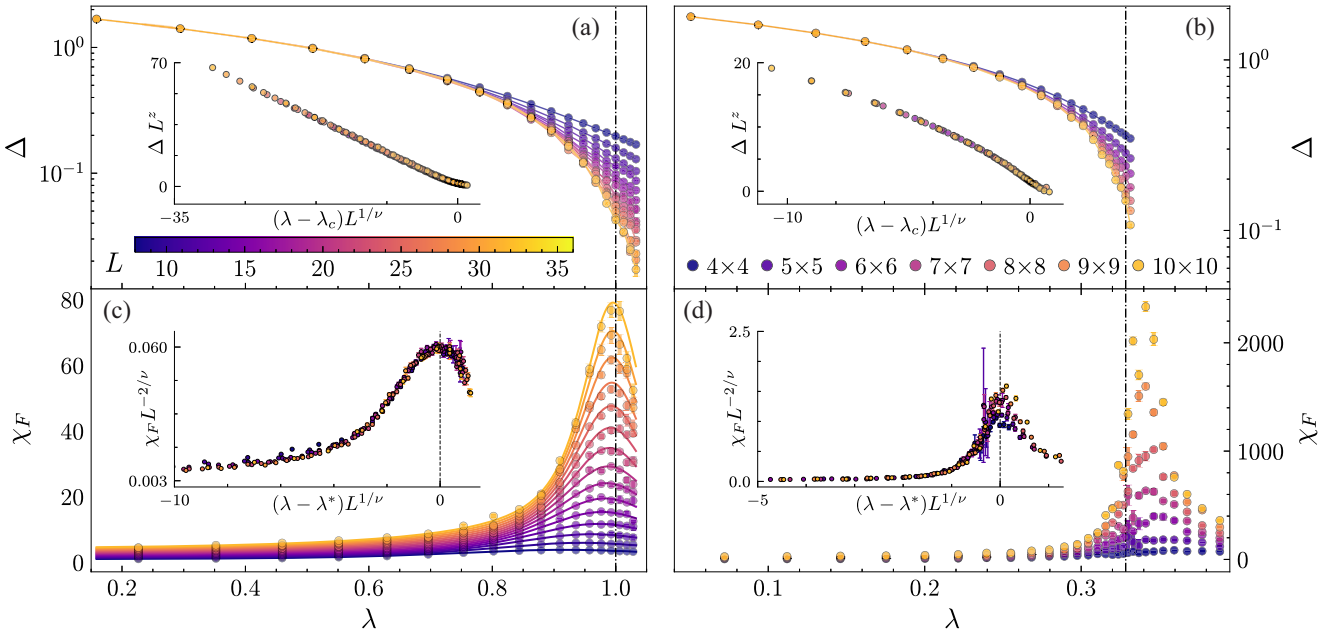


FIG. 3. Energy gap Δ vs λ for 1D [panel (a)] and 2D [panel (b)] systems. The scaling $\Delta \sim L^{-z}$ at the critical point is used to extract the dynamical exponent $z$, while the insets display the finite-size data collapse used to determine the correlation-length exponent $\nu$. The extracted critical exponents are listed in Table I. Panels (c) and (d) show the corresponding ground-state fidelity susceptibility for 1D and 2D systems, with the scaling collapse illustrated in the insets. In one dimension [panel (c)], the exact solution obtained via Jordan-Wigner transformation is shown as solid lines, demonstrating excellent agreement with the NQS results.

finite systems. This leads to the finite-size scaling relation

$$\Delta = L^{-z} F\big((\lambda-\lambda_c)L^{1/\nu}\big), \tag{8}$$

with a universal function $F$. The energy gap as a function of λ is shown in Figs. 3(a) and 3(b) for one and two dimensions, respectively. At the critical point, the scaling $\Delta \sim L^{-z}$ allows $z$ to be estimated via linear regression. Once $z$ is found, the exponent $\nu$ is obtained by collapsing the gap data across different system sizes according to Eq. (8), with the insets displaying the resulting high-quality collapse. Details of this procedure can be found in the Supplemental Material [58]. Numerical values of the critical exponents are given in Table I, showing good agreement with the known values.

The availability of NQS wavefunctions at different λ enables direct estimation of the ground-state fidelity susceptibility, a general probe of quantum phase transitions that does not rely on prior knowledge of the order parameter [62]. It is defined as

$$\chi_{\mathcal{F}}(\lambda) = -\lim_{\epsilon\to 0}\frac{\partial^2}{\partial\epsilon^2}\ln \mathcal{F}_n(\lambda,\lambda+\epsilon), \tag{9}$$

where $\mathcal{F}(\lambda,\lambda+\epsilon) = |\langle\Psi_n(\lambda)|\Psi(\lambda+\epsilon)\rangle|^2$. In the thermodynamic limit, $\chi_{\mathcal{F}}$ diverges at criticality, following universal scaling laws [63,64], and can be estimated from neighboring fidelities as $\chi_{\mathcal{F}} \approx -\frac{1}{\epsilon^2}\ln(\mathcal{F}(\lambda-\epsilon)\mathcal{F}(\lambda+\epsilon))$. As shown in Figs. 3(c) and 3(d), this estimate reproduces the expected divergence in both one and two dimensions, with the scaling collapse visible in the insets. In one dimension, the TFIM can be solved through a Jordan-Wigner transformation to a free fermion model [65], and our results are in excellent agreement with the exact solution [solid lines in panel (c)]. See the Supplemental Material [58] for more details.

At criticality, the physics is governed by CFT. The spectra of $(2+1)$-dimensional CFTs on a spatial torus are expected to be universal [49], but unlike their two-dimensional counterparts, they are not fixed by radial quantization and are analytically inaccessible [66].

In Fig. 2(d), we show that adiabatic transport reproduces universal torus gap ratios of the 3D Ising and Ising* CFTs. Using the lowest excitation with fully periodic boundary conditions as the reference gap, $\Delta_\lambda^{\text{ref}} = E_\lambda^{(\mathbf{k}=0,\text{PP})} - E_\lambda^{(\text{GS,PP})}$, the gap ratio is defined as

$$R_\lambda^X = \frac{E_\lambda^X - E_\lambda^{(\text{GS,PP})}}{\Delta_\lambda^{\text{ref}}}. \tag{10}$$

Here, $X$ denotes either a finite-momentum excitation with periodic boundary conditions (PP) or the lowest-energy state obtained by imposing antiperiodic boundary conditions along one (AP) or both (AA) torus cycles. The former is associated with a conformal descendant, while AP and AA boundary conditions implement $\mathbb{Z}_2$ defect lines wrapping one or both cycles of the torus.

The estimates of the CFT gap ratios at the critical point for the $8\times 8$ results are presented in Table I. The finite-momentum and AP ratios agree well with the estimates of Ref. [49], while the AA ratio shows a larger deviation, consistent with stronger finite-size effects from two intersecting defect lines. Together with the critical-exponent estimates, these results show that the variational NQS approach resolves both momentum and defect excitations near the transition.

## IV. CONCLUSION

We have introduced an accurate transport method to access NQS excited states, naturally extending the existing reli-

TABLE I. Critical exponents $z$ and $\nu$ and CFT gap ratios from adiabatic transport. The critical exponents are compared with exact values calculated by conformal bootstrap [59–61], while the gap ratios are compared with the literature values of Ref. [49].

| Dimension | Exponent | Exact | Transport |
|---|---|---|---|
| 1D | $z$ | 1 | 0.99(1) |
| | $\nu$ | 1 | 1.024(3) |
| 2D | $z$ | 1 | 1.03(2) |
| | $\nu$ | 0.629 97 | 0.6315(7) |
| Size | Ratio ($\boldsymbol{X}$) | Literature | Transport |
| $8 \times 8$ | $(\mathbf{k} = \pi/4, PP)$ | 5.305 | 5.26(3) |
| | $(\mathbf{k} = 0, AP)$ | 0.352 | 0.345(6) |
| | $(\mathbf{k} = 0, AA)$ | 0.484 | 0.507(7) |

able VMC toolbox. Generalized stochastic parameter update equations based on the shift-and-invert procedure, used with simple perturbative estimates of target energies, enable the parallel preparation of NQS excited states. Without sacrificing accuracy, this feature eliminates the quadratic computational overhead introduced by enforcing orthogonality constraints. Benchmarks against the integrable (one-dimensional) and large nonintegrable (two-dimensional) TFIM reveal that adiabatic transport, coupled with NQS, can access both low-lying excited states and universal critical physics through accurate estimates of the many-body gap and the critical exponents.

On-demand access to excited states removes important roadblocks in precision many-body calculations. They are a key ingredient of quasiparticle excitation spectra on top of highly correlated states. The screening of candidate materials for a target property, such as the estimation of optical band gaps [67,68], can benefit significantly from precise access to higher states. Similarly, finite-temperature and real-time solvers indirectly rely on excited states to push our understanding of thermal properties as well as prethermalization. Adiabatic transport itself has become a key computational benchmark for near-term practical quantum advantage [69].

Having recovered universal CFT gap ratios for the Ising and Ising* sectors, adiabatic transport has shown promise as a numerical probe of universal physics. As topological states are characterized by robust ground-state degeneracies on the torus, NQS transport methods may provide a route to resolving small gaps in other topologically ordered systems. Exploiting the high expressive power of NQS to represent correlated *spectra* is well positioned to be the next frontier of computationally driven physical insight. Adiabatic transport offers a universal and rapidly scalable computational framework to reach this goal. We are excited to see which challenging open problems it will attack next.

## ACKNOWLEDGMENTS

M.M. acknowledges many useful discussions with Jannes Nys. A.O. acknowledges support from the Pierre Hohenberg Graduate Scholar Fellowship and computational resources provided by the Flatiron Institute. The Flatiron Institute is a division of the Simons Foundation. D.S. thanks AFOSR for support through Award No. FA9550-25-1-0067.

## DATA AVAILABILITY

All simulations were performed on graphical processing units using the JAX [70] library for array manipulation and automatic differentiation. Equinox [71] was used for neural network design and Optax for optimization. Data were post-processed using NumPy [72] and SciPy [73]. The plots were produced using the Matplotlib [74] library. The code needed to reproduce the results in this work and explore new ones can be found in the repository in Refs. [75,76].

## APPENDIX A: INVERSE POWER ITERATION

In this Appendix, we give a more detailed treatment of the IPI. As noted in the main text, given an estimate of the target eigenvalue $\omega$, we produce a sequence of eigenstate estimates $|\Psi_k\rangle$ such that $|\Psi_{k+1}\rangle \propto (H-\omega)^{-1}|\Psi_k\rangle$. After sufficiently many iterations, this procedure extracts a target eigenstate $|m\rangle$ with energy $E_m$ if the starting state $|\Psi_0\rangle = \sum_n c_n|n\rangle$ has a nonzero overlap with it,

$$\begin{aligned}\lim_{k\to\infty}|\Psi_k\rangle &= \lim_{k\to\infty}(H-\omega)^{-k}|\Psi_0\rangle \\ &= \lim_{k\to\infty}\sum_n c_n(E_n-\omega)^{-k}|n\rangle \propto |m\rangle. \end{aligned} \tag{A1}$$

To project one IPI onto the variational manifold, we require that parameter updates $\delta\theta$ satisfy

$$|\Psi_{\theta+\delta\theta}\rangle \approx |\Psi_\theta\rangle + \sum_\mu \delta\theta^\mu |\partial_\mu\Psi_\theta\rangle \propto (H-\omega)^{-1}|\Psi_\theta\rangle. \tag{A2}$$

We multiply Eq. (A2) by $(H-\omega)$ from the left and expand the solution in the tangent space basis $|\partial_\mu\Psi_\theta\rangle$, recovering

$$\sum_\nu \mathrm{Re}\,\{\langle\partial_\mu\Psi_\theta|(H-\omega)|\partial_\nu\Psi_\theta\rangle\}\delta\theta^\nu = -\mathrm{Re}\,\{\langle\partial_\mu\Psi_\theta|H|\Psi_\theta\rangle\}, \tag{A3}$$

after making use of $\partial_\mu\langle\Psi_\theta|\Psi_\theta\rangle = 2\mathrm{Re}\langle\partial_\mu\Psi_\theta|\Psi_\theta\rangle = 0$ and the assumption that the parameters are real. The last equation leads to the parameter update rule $\delta\theta = -\mathbf{G}^{-1}f$, with definitions in Eq. (4). The matrix $\mathbf{G}$ and the vector $f$ must be evaluated by Monte Carlo methods as

$$\begin{aligned}\mathbf{G}_{\mu\nu} &= 2\,\mathrm{Re}\mathbb{E}_{\mathbf{x}\sim|\Psi_\theta|^2}\left[\frac{\langle\partial_\mu\Psi_\theta|\mathbf{x}\rangle}{\langle\Psi_\theta|\mathbf{x}\rangle}\frac{\langle\mathbf{x}|(H-\omega)|\partial_\nu\Psi_\theta\rangle}{\langle\mathbf{x}|\Psi_\theta\rangle}\right], \\ f_\mu &= 2\mathrm{Re}\mathbb{E}_{\mathbf{x}\sim|\Psi_\theta|^2}\left[\frac{\langle\partial_\mu\Psi_\theta|\mathbf{x}\rangle}{\langle\Psi_\theta|\mathbf{x}\rangle}\frac{\langle\mathbf{x}|H|\Psi_\theta\rangle}{\langle\mathbf{x}|\Psi_\theta\rangle}\right]. \end{aligned} \tag{A4}$$

We identify familiar ingredients of traditional VMC calculations in Eq. (A4) as the local energy $E_{\mathrm{loc}}(\mathbf{x})$ and the NQS *Jacobian* $\mathcal{J}_\mu(\mathbf{x})$, and define the *projected Jacobian* $\mathcal{P}_\mu(\mathbf{x})$ as

$$\begin{aligned} E_{\mathrm{loc}}(\mathbf{x}) &= \frac{\langle\mathbf{x}|H|\Psi_\theta\rangle}{\langle\mathbf{x}|\Psi_\theta\rangle}, \quad \mathcal{J}_\mu(\mathbf{x}) = \frac{\langle\mathbf{x}|\partial_\mu\Psi_\theta\rangle}{\langle\mathbf{x}|\Psi_\theta\rangle}, \\ \text{and} \quad \mathcal{P}_\mu(\mathbf{x}) &= \frac{\langle\mathbf{x}|(H-\omega)|\partial_\mu\Psi_\theta\rangle}{\langle\mathbf{x}|\Psi_\theta\rangle}. \end{aligned} \tag{A5}$$

All quantities defined in Eq. (A5) have efficient Monte Carlo estimators provided that the Hamiltonian $H$ is sparse enough to have only polynomially many connected basis elements to $|\mathbf{x}\rangle$, which is a requirement for ground-state VMC calculations as well. We refer readers interested in the full treatment of

local energies and NQS Jacobians to Refs. [1,3–5]. After algebraic manipulation, we write the estimator for the projected Jacobian as

$$\mathcal{P}_\mu(\mathbf{x}) = \partial_\mu E_{\mathrm{loc}}(\mathbf{x}) + \mathcal{J}_\mu(\mathbf{x})(E_{\mathrm{loc}}(\mathbf{x}) - \omega). \tag{A6}$$

We note that all of the derivatives in the resulting expressions can be evaluated using efficient automatic differentiation. The intractable norm of the variational state cancels out in all expressions in Eqs. (A5) and (A7).

In the finite-sample approximation using Monte Carlo samples $\{\mathbf{x}_1, \ldots, \mathbf{x}_{N_s}\}$, drawn so that $\mathbb{E}_{\mathbf{x}\sim|\Psi_\theta|^2}[A(\mathbf{x})] \approx \frac{1}{N_s}\sum_{i=1}^{N_s} A(\mathbf{x}_i)$ holds, the parameter update equation reduces to a simple linear system. Defining matrices

$$J_{i\mu} = \frac{\mathcal{J}_\mu(\mathbf{x}_i)}{\sqrt{N_s}}, \quad P_{i\mu} = \frac{\mathcal{P}_\mu(\mathbf{x}_i)}{\sqrt{N_s}}, \quad \text{and} \quad \varepsilon_i = \frac{E_{\mathrm{loc}}(\mathbf{x}_i)}{\sqrt{N_s}}, \tag{A7}$$

that linear system reads $J^\top P\,\delta\theta = -J^\top\varepsilon$.

For any given inverse power iteration, it is beneficial to employ the Woodbury identity [47,77] whenever the number of variational parameters $P$ exceeds the number of Monte Carlo samples $N_s$,

$$\delta\theta = -(J^\top P + \gamma\mathbb{1}_P)^{-1}J^\top\varepsilon = -J^\top(PJ^\top + \gamma\mathbb{1}_{N_s})^{-1}\varepsilon, \tag{A8}$$

to invert larger matrices, with an optional diagonal shift $\gamma$ for numerical stability.

Although the choice of the linear solver is arbitrary in principle as long as the parameter update takes the form $\theta' = \theta - \eta\,\mathbf{G}^{-1}f$, in practice, we found that a modified pseudoinverse offers the best balance of accuracy and speed. The solver we use has been developed in previous research by some of the authors in Ref. [32], building on insights in Ref. [30]. More details can be found in the Supplemental Material [58].

Shift-and-invert is guaranteed to converge exponentially for dense matrices. In the case of stochastically estimated $J$ and $P$, we observe that convergence depends on the number of Monte Carlo samples and the conditioning of $J$ as well. Our tests and those of Ref. [50] suggest that the stochastic version inherits IPI convergence properties of the dense version under sufficient Monte Carlo samples. Determining exact conditions for convergence and more detailed measurements is left for future work.

## APPENDIX B: NEURAL NETWORK ARCHITECTURE

The variational wavefunction is defined as the residual RBM [7,56,57] multiplied by the initial condition wavefunction, as outlined in the main text. The architecture is based on a simple, shallow RBM trial state commonly used in VMC calculations. We use an adapted version of the RBM that is invariant with respect to translational lattice symmetries,

$$\mathtt{RBM}_{(w,b)}(\mathbf{x}) = \sum_{k,\mathbf{r}} \ln\cosh(b^k + (w^k * \mathbf{x})_{\mathbf{r}}), \tag{B1}$$

due to the appropriate convolution operation $*$ over lattice coordinates $\mathbf{r}$. In Eq. (B1), biases $b$ and weights $w$ can be optimized to represent different states of lattice spins $\mathbf{x}$. The RBM on its own is a parameter-efficient ansatz capable of capturing many interesting correlated states.

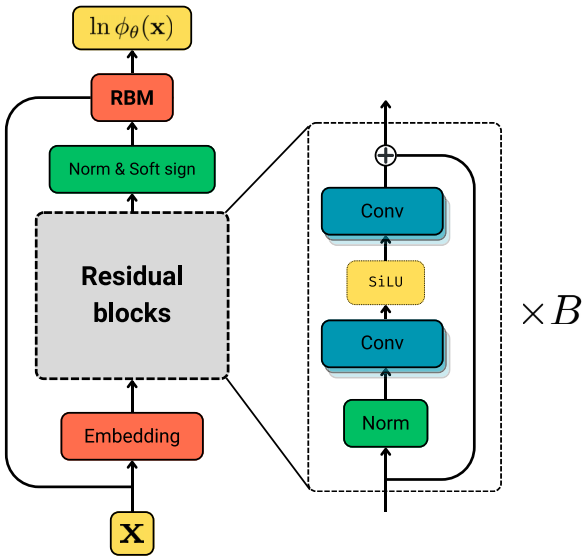


FIG. 4. The sublayer internal connectivity of the residual RBM NQS architecture used in the adiabatic transport simulations. This diagram shows the NQS amplitude factor of the full wavefunction.

However, we found that capturing excited states requires increased expressivity beyond RBMs. Therefore, we supplement the RBM architecture with a custom convolutional encoder [54] $f_\alpha$ parametrized by $\alpha$. The encoder outputs additional *hidden* lattice spin configurations $\mathbf{y} = f_\alpha(\mathbf{x})$, which are stacked on top of the input configuration $\mathbf{x}$. The stacked spin configurations are passed into an RBM with complex parameters, producing the final log-amplitude $\ln\psi_\theta(\mathbf{x})$. Augmenting the existing RBM architecture with hidden spins in this way is inspired by *backflow* methods for fermionic Hamiltonians and allows us to make the overall model more expressive without sacrificing any of the components that make the RBM a useful inductive bias in the first place. Mathematically, we replace the RBM input $\mathbf{x}$ in Eq. (B1) with an augmented spin variable $\tilde{\mathbf{x}}$ with $d+1$ *channels*,

$$\tilde{\mathbf{x}} = \left[\mathbf{x},\ f_\alpha^{(1)}(\mathbf{x}),\ \ldots,\ f_\alpha^{(d)}(\mathbf{x})\right], \tag{B2}$$

where $[\cdots]$ denotes the concatenation along a new *feature* or *channel* dimension. The internal connectivity of the resulting neural network is shown in Fig. 4.

The rest of this Appendix is devoted to laying out the structure of the convolutional encoder $f_\alpha$. First, we lift the internal representation of each spin from $\sigma_i \in \{-1, +1\}$ to $\mathbf{h}_i = \sigma_i \otimes \mathbf{u}$, where $\mathbf{u} \in \mathbb{R}^d$ is a trainable embedding vector. The embedding vector is shared between all lattice sites to preserve any lattice symmetries.

After embedding, the lifted spin representation is passed through $B$ *residual* blocks in sequence. Each block is a two-layer convolutional network with a $\mathtt{SiLU}$ [78] nonlinearity and a skip connection [55]. We also apply layer normalization [79] across the lattice dimensions, as illustrated in Fig. 4. Mathematically, the output $\mathbf{h}'$ of each block is

$$\mathbf{h}' = \mathbf{h} + \mathtt{Conv}_{\downarrow}(\mathtt{SiLU}(\mathtt{Conv}_{\uparrow}(\mathtt{Norm}(\mathbf{h})))) \tag{B3}$$

for a given input $\mathbf{h}$. Convolutions in Eq. (B3) are defined as $\mathtt{Conv}(\mathbf{h})^k = b^k + \sum_l w^{kl} * \mathbf{h}^l$, with $*$ being the periodic

(antiperiodic) convolution operation in the case of periodic (antiperiodic) boundary conditions on $\mathbf{x}$ or the zero-padded convolution in the case of open boundary conditions. By choosing the appropriate convolution, we keep each of the residual blocks and the overall model translationally invariant. The first convolution, $\texttt{Conv}_{\uparrow}$, increases the number of features to $\alpha\, d$ before the second convolution $\texttt{Conv}_{\downarrow}$ decreases it back to $d$.

Layer normalization is applied across the lattice dimension for each feature or channel independently as

$$(\texttt{Norm}(\mathbf{h}))^k = \frac{\mathbf{h}^k - \mathbb{E}[\mathbf{h}^k]}{\sqrt{\text{Var}[\mathbf{h}^k] + \epsilon}} \times \gamma^k + \beta^k, \tag{B4}$$

where $\mathbb{E}[\cdot]$ and $\text{Var}[\cdot]$ are the empirical mean and variance, respectively, and the scale $\gamma$ and the shift $\beta$ are trainable parameters, with $\epsilon$ fixed to a small constant for numerical stability in cases of vanishing variance.

We produce the final hidden spin values as

$$\mathbf{y} = \texttt{SoftSign}(\texttt{Norm}(\mathbf{h})) \tag{B5}$$

from the output residual blocks $\mathbf{h}$. We emphasize that each component of $\mathbf{y}$ is restricted to $-1 < y_i < 1$ by this operation. The `SiLU` [78] and `SoftSign` nonlinearities in Eqs. (B3) and (B5) are defined as

$$\texttt{SiLU}(x) = \frac{x}{1+e^{-x}} \quad \text{and} \quad \texttt{SoftSign}(x) = \frac{x}{1+|x|}, \tag{B6}$$

and are always applied elementwise. Numerical values of all hyperparameters can be found in the Supplemental Material [58].


[1] J. Carrasquilla and G. Torlai, How to use neural networks to investigate quantum many-body physics, PRX Quantum **2**, 040201 (2021).

[2] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[3] M. Medvidović and J. R. Moreno, Neural-network quantum states for many-body physics, Eur. Phys. J. Plus **139**, 631 (2024).

[4] H. Lange, A. Van de Walle, A. Abedinnia, and A. Bohrdt, From architectures to applications: A review of neural quantum states, Quantum Sci. Tech. **9**, 040501 (2024).

[5] A. Dawid, J. Arnold, B. Requena, A. Gresch, M. Płodzień, K. Donatella, K. A. Nicoli, P. Stornati, R. Koch, M. Büttner, *et al.*, *Machine Learning in Quantum Sciences* (Cambridge University Press, Cambridge, 2025).

[6] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, Nat. Phys. **13**, 431 (2017).

[7] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[8] D. Luo and B. K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, Phys. Rev. Lett. **122**, 226401 (2019).

[9] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, Proc. Natl. Acad. Sci. USA **119**, e2122059119 (2022).

[10] A. Malyshev, J. M. Arrazola, and A. I. Lvovsky, Autoregressive neural quantum states with quantum number symmetries, arXiv:2310.04166.

[11] Y. Gu, W. Li, H. Lin, B. Zhan, R. Li, Y. Huang, D. He, Y. Wu, T. Xiang, M. Qin, L. Wang, and D. Lv, Solving the Hubbard model with neural quantum states, Nat. Commun. **17**, 7838 (2026).

[12] A. Chen, Z.-Q. Wan, A. Sengupta, A. Georges, and C. Roth, Neural network-augmented Pfaffian wave-functions for scalable simulations of interacting fermions, arXiv:2507.10705.

[13] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks, Phys. Rev. Res. **2**, 033429 (2020).

[14] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic Schrödinger equation, Nat. Chem. **12**, 891 (2020).

[15] I. von Glehn, J. S. Spencer, and D. Pfau, A self-attention ansatz for *ab-initio* quantum chemistry, arXiv:2211.13672.

[16] A. Lovato, C. Adams, G. Carleo, and N. Rocco, Hidden-nucleons neural-network quantum states for the nuclear many-body problem, Phys. Rev. Res. **4**, 043178 (2022).

[17] G. Pescia, J. Han, A. Lovato, J. Lu, and G. Carleo, Neural-network quantum states for periodic systems in continuous space, Phys. Rev. Res. **4**, 023138 (2022).

[18] G. Pescia, J. Nys, J. Kim, A. Lovato, and G. Carleo, Message-passing neural quantum states for the homogeneous electron gas, Phys. Rev. B **110**, 035108 (2024).

[19] C. Smith, Y. Chen, R. Levy, Y. Yang, M. A. Morales, and S. Zhang, Unified variational approach description of ground-state phases of the two-dimensional electron gas, Phys. Rev. Lett. **133**, 266504 (2024).

[20] J. Kim, G. Pescia, B. Fore, J. Nys, G. Carleo, S. Gandolfi, M. Hjorth-Jensen, and A. Lovato, Neural-network quantum states for ultra-cold Fermi gases, Commun. Phys. **7**, 148 (2024).

[21] A. Foster, Z. Schätzle, P. B. Szabó, L. Cheng, J. Köhler, G. Cassella, N. Gao, J. Li, F. Noé, and J. Hermann, An *ab initio* foundation model of wavefunctions that accurately describes chemical bond breaking, arXiv:2506.19960.

[22] O. Sharir, A. Shashua, and G. Carleo, Neural tensor contractions and the expressive power of deep neural quantum states, Phys. Rev. B **106**, 205136 (2022).

[23] S. R. White, Density matrix formulation for quantum renormalization groups, Phys. Rev. Lett. **69**, 2863 (1992).

[24] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Ann. Phys. **326**, 96 (2011).

[25] D. Wu, R. Rossi, F. Vicentini, N. Astrakhantsev, F. Becca, X. Cao, J. Carrasquilla, F. Ferrari, A. Georges, M. Hibat-Allah, *et al.*, Variational benchmarks for quantum many-body problems, Science **386**, 296 (2024).

[26] Z. Li, H. Xie, X. Dong, and L. Wang, Deep variational free energy calculation of hydrogen Hugoniot, Phys. Rev. Lett. **136**, 076504 (2026).

[27] J. Nys, Z. Denis, and G. Carleo, Real-time quantum dynamics of thermal states with neural thermofields, Phys. Rev. B **109**, 235120 (2024).

[28] M. Schmitt and M. Heyl, Simulating dynamics of correlated matter with neural quantum states, arXiv:2506.03124.

[29] M. Schmitt, M. M. Rams, J. Dziarmaga, M. Heyl, and W. H. Zurek, Quantum phase transition dynamics in the two-dimensional transverse-field Ising model, Sci. Adv. **8**, eabl6850 (2022).

[30] M. Schmitt and M. Heyl, Quantum many-body dynamics in two dimensions with artificial neural networks, Phys. Rev. Lett. **125**, 100503 (2020).

[31] I. L. Gutiérrez and C. B. Mendl, Real time evolution with neural-network quantum states, Quantum **6**, 627 (2022).

[32] M. Medvidović and D. Sels, Variational quantum dynamics of two-dimensional rotor models, PRX Quantum **4**, 040302 (2023).

[33] A. Sinibaldi, C. Giuliani, G. Carleo, and F. Vicentini, Unbiasing time-dependent variational Monte Carlo by projected quantum evolution, Quantum **7**, 1131 (2023).

[34] J. Nys, G. Pescia, A. Sinibaldi, and G. Carleo, *Ab-initio* variational wave functions for the time-dependent many-electron Schrödinger equation, Nat. Commun. **15**, 9404 (2024).

[35] A. Van de Walle, M. Schmitt, and A. Bohrdt, Many-body dynamics with explicitly time-dependent neural quantum states, Mach. Learn. Sci. Techn. **6**, 045011 (2025).

[36] A. Sinibaldi, D. Hendry, F. Vicentini, and G. Carleo, Time-dependent neural Galerkin method for quantum dynamics, Phys. Rev. Lett. **136**, 120402 (2026).

[37] A. Georges and G. Kotliar, Hubbard model in infinite dimensions, Phys. Rev. B **45**, 6479 (1992).

[38] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, Rev. Mod. Phys. **68**, 13 (1996).

[39] F. Aryasetiawan and O. Gunnarsson, The GW method, Rep. Prog. Phys. **61**, 237 (1998).

[40] L. Reining, The GW approximation: Content, successes and limitations, WIREs Comput. Mol. Sci. **8**, e1344 (2018).

[41] D. Golze, M. Dvorak, and P. Rinke, The GW compendium: A practical guide to theoretical photoemission spectroscopy, Front. Chem. **7**, 377 (2019).

[42] D. Pfau, S. Axelrod, H. Sutterud, I. Von Glehn, and J. S. Spencer, Accurate computation of quantum excited states with neural networks, Science **385**, eadn0137 (2024).

[43] A. Kahn, L. Gravina, and F. Vicentini, Variational subspace methods and application to improving variational Monte Carlo dynamics, Quantum **10**, 2082 (2026).

[44] D. Hendry, A. Sinibaldi, and G. Carleo, Grassmann variational Monte Carlo with neural wave functions, Phys. Rev. B (2026), doi:10.1103/l8mr-567g.

[45] M. Kolodrubetz, D. Sels, P. Mehta, and A. Polkovnikov, Geometry and non-adiabatic response in quantum and classical systems, Phys. Rep. **697**, 1 (2017).

[46] H. Kim, M. Fishman, and D. Sels, Variational adiabatic transport of tensor networks, PRX Quantum **5**, 020361 (2024).

[47] R. Rende, L. L. Viteritti, L. Bardone, F. Becca, and S. Goldt, A simple linear algebra identity to optimize large-scale neural network quantum states, Commun. Phys. **7**, 260 (2024).

[48] V. Hernandes, T. Spriggs, S. Khaleefah, and E. Greplova, Adiabatic fine-tuning of neural quantum states enables detection of phase transitions in weight space, arXiv:2503.17140.

[49] M. Schuler, S. Whitsitt, L.-P. Henry, S. Sachdev, and A. M. Läuchli, Universal signatures of quantum critical points from finite-size torus spectra: A window into the operator content of higher-dimensional conformal field theories, Phys. Rev. Lett. **117**, 210401 (2016).

[50] H. Zhang, V. Armegioiu, J. Carrasquilla, S. Mishra, J. Müller, J. Nys, and M. Zeinhofer, Projected inverse iteration: An eigenvalue approach to ground-state computation with neural quantum states, arXiv:2606.07825.

[51] S. Sorella, Green function Monte Carlo with stochastic reconfiguration, Phys. Rev. Lett. **80**, 4558 (1998).

[52] S. I. Amari, Natural gradient works efficiently in learning, Neural Comput. **10**, 251 (1998).

[53] H. W. J. Blöte and Y. Deng, Cluster Monte Carlo simulation of the transverse Ising model, Phys. Rev. E **66**, 066110 (2002).

[54] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature (London) **521**, 436 (2015),.

[55] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, arXiv:1512.03385.

[56] R. Salakhutdinov, A. Mnih, and G. Hinton, Restricted Boltzmann machines for collaborative filtering, in *Proceedings of the 24th International Conference on Machine Learning* (ACM Press, Corvallis, OR, 2007), pp. 791–798.

[57] R. G. Melko, G. Carleo, J. Carrasquilla, and J. Ignacio Cirac, Restricted Boltzmann machines in quantum physics, Nat. Phys. **15**, 887 (2019).

[58] See Supplemental Material at http://link.aps.org/supplemental/10.1103/khy4-5v68 for details on the data collapse, the pseudoinverse solution to the update equations, the numerical values of all hyperparameters, and the analytical solution to the 1D TFIM.

[59] S. El-Showk, M. F. Paulos, D. Poland, S. Rychkov, D. Simmons-Duffin, and A. Vichi, Solving the 3d Ising model with the conformal bootstrap II. $c$-Minimization and precise critical exponents, J. Stat. Phys. **157**, 869 (2014).

[60] C.-H. Chang, V. Dommes, R. S. Erramilli, A. Homrich, P. Kravchuk, A. Liu, M. S. Mitchell, D. Poland, and D. Simmons-Duffin, Bootstrapping the 3d Ising stress tensor, J. High Energy Phys. 03 (2025) 136.

[61] M. Reehorst, Rigorous bounds on irrelevant operators in the 3d Ising model CFT, J. High Energy Phys. 09 (2022) 177.

[62] L. Wang, Y.-H. Liu, J. Imriška, P. N. Ma, and M. Troyer, Fidelity susceptibility made simple: A unified quantum Monte Carlo approach, Phys. Rev. X **5**, 031007 (2015).

[63] A. Polkovnikov, Universal adiabatic dynamics in the vicinity of a quantum critical point, Phys. Rev. B **72**, 161201(R) (2005).

[64] L. Campos Venuti and P. Zanardi, Quantum critical scaling of the geometric tensors, Phys. Rev. Lett. **99**, 095701 (2007).

[65] S. Sachdev, *Quantum Phase Transitions*, 2nd ed. (Cambridge University Press, Cambridge, 2011).

[66] P. Di Francesco, P. Mathieu, and D. Sénéchal, *Conformal Field Theory*, *Graduate Texts in Contemporary Physics* (Springer, New York, NY, 1997).

[67] L. Zhao and E. Neuscamman, Variational excitations in real solids: Optical gaps and insights into many-body perturbation theory, Phys. Rev. Lett. **123**, 036402 (2019).

[68] R. J. Hunt, M. Szyniszewski, G. I. Prayogo, R. Maezono, and N. D. Drummond, Quantum Monte Carlo calculations of energy gaps from first principles, Phys. Rev. B **98**, 075122 (2018).

[69] A. D. King, A. Nocera, M. M. Rams, J. Dziarmaga, R. Wiersema, W. Bernoudy, J. Raymond, N. Kaushal, N. Heinsdorf, R. Harris, *et al.*, Beyond-classical computation in quantum simulation, Science **388**, 199 (2025).

[70] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, Y. Katariya, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of Python+NumPy programs (2018), https://github.com/jax-ml/jax.

[71] P. Kidger and C. Garcia, Equinox: Neural networks in JAX via callable PyTrees and filtered transformations, arXiv:2111.00254.

[72] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, *et al.*, Array programming with NumPy, Nature (London) **585**, 357 (2020).

[73] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).

[74] J. D. Hunter, Matplotlib: A 2D graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[75] M. Medvidović and A. Orfi, Adiabatic transport, Zenodo, 2026, https://doi.org/10.5281/zenodo.21797112.

[76] https://github.com/Matematija/nqs-adiabatic

[77] A. Chen and M. Heyl, Empowering deep neural quantum states through efficient optimization, Nat. Phys. **20**, 1476 (2024).

[78] S. Elfwing, E. Uchibe, and K. Doya, Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, arXiv:1702.03118.

[79] J. L. Ba, J. R. Kiros, and G. E. Hinton, Layer normalization, arXiv:1607.06450.