# High-Accuracy Neural-Network Quantum States via Randomized Real-Time Dynamics

**John M. Martyn**
Pacific Northwest National Laboratory,
Harvard Quantum Initiative, Harvard University,
NSF AI Institute for Artificial Intelligence
and Fundamental Interactions
john_martyn@fas.harvard.edu

**Di Luo**
University of California, Los Angeles
diluo@g.ucla.edu

## Abstract

We introduce a new algorithm to enhance the accuracy of neural-network quantum states (NQSs) in approximating ground states of quantum systems. Our method, termed *dynamical averaging*, estimates expectation values by sampling a NQS as it evolves in real-time. By using insights from quantum information theory, we establish an up-to-quadratic suppression of error in estimating arbitrary ground-state observables. Importantly, this improved accuracy requires neither further energy minimization nor an enhanced parameterization, but rather exploits the influence of randomness on quantum states. We demonstrate the advantage of dynamical averaging on an important quantum spin model in many-body physics, where it reduces the relative error in correlation functions from roughly $10\%$ to below $1\%$.

## 1 Introduction

Machine learning has emerged as a powerful tool for addressing large-scale scientific challenges [1–4]. A notable such challenge is the quantum many-body problem [5, 6], which seeks to characterize relevant states of interacting quantum systems (e.g., ground states, thermal states) and holds wide-reaching applications in chemistry, condensed matter physics, and materials science [7, 8]. However, this problem is notoriously difficult to solve due to the complexity of quantum many-body interactions and the exponentially large Hilbert space of a quantum system.

A prominent line of work has leveraged machine learning to tackle the quantum many-body problem by introducing neural-network quantum states (NQSs) [9], a powerful variational ansatz that parameterizes a quantum state by neural networks. Supported by the rich expressivity of neural-networks [10, 11], NQSs have been shown to accurately model complex quantum systems, including ground states, real-time dynamics, quantum field theories, and quantum materials [12–16].

Despite their success, NQSs still incur a small, but non-negligible, systematic error in approximating a target state, which limits their ability to accurately resolve quantities of interest. Conventional wisdom in machine learning suggests that reducing this error requires additional optimization or a more expressive model. Remarkably, this is unnecessary for NQSs: we introduce *dynamical averaging* as a quantum-inspired algorithm that provably reduces systematic error in approximating ground states, without requiring further energy minimization nor increasing model complexity.

In this work, we make three key contributions as follows:

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Machine Learning and the Physical Sciences Workshop, NeurIPS 2025.

- We draw upon results from quantum information to prove that randomly evolving a NQS in real-time provides an up-to-quadratic suppression of systematic errors (i.e. $\epsilon \to \mathcal{O}(\epsilon^2)$) in estimating arbitrary ground-state observables.
- We develop dynamical averaging as a simple algorithm that realizes this error suppression by sampling a NQS as it evolves in real-time.
- We showcase the benefits of dynamical averaging on an important quantum spin model.

## 2 Theoretical Background and Motivation

**Background** A quantum system on $N$ spins is described by a Hamiltonian $H$, which is a $2^N \times 2^N$ Hermitian matrix. Its eigenvalues and eigenvectors, denoted by $E_n$ and $|E_n\rangle$, respectively, represent the energy levels of the system. The lowest level $|E_0\rangle$ describes the system's ground state, which encodes many relevant physical properties. However, approximating the ground state is a notoriously challenging problem, and is generally intractable exactly for interacting systems [17].

To address this, computational methods often invoke the variational principle, which states that the energy of a quantum state always exceeds the ground state energy: $\langle\psi|H|\psi\rangle \geq E_0$. Indeed, NQSs parameterize a state vector, denoted by $|\psi_\theta\rangle$ for some parameters $\theta$, and minimize its energy to approximate the ground state. The energy of a NQS is estimated using variational Monte Carlo, which equates the energy to a classical expectation value [18]:

$$\langle\psi_\theta|H|\psi_\theta\rangle = \mathop{\mathbb{E}}_{x\sim|\psi_\theta(x)|^2}\big[E_{\text{loc}}(x)\big], \tag{1}$$

where $\psi_\theta(x) = \langle x|\psi_\theta\rangle$ is the NQS in a basis $|x\rangle$, and $E_{\text{loc}}(x) = \frac{\langle x|H|\psi_\theta\rangle}{\langle x|\psi_\theta\rangle}$ is known as the *local energy*. Thus, the energy can be estimated by sampling $|\psi_\theta(x)|^2$ and computing the empirical mean of $E_{\text{loc}}(x)$. Analogous expressions hold for other observables, e.g. $\langle\psi_\theta|O|\psi_\theta\rangle$, upon replacing $E_{\text{loc}}$ with the appropriate local observable $O_{\text{loc}}(x) = \frac{\langle x|O|\psi_\theta\rangle}{\langle x|\psi_\theta\rangle}$. This approach is also used to evaluate the gradient of the energy with respect to $\theta$, which is used to perform optimization.

**Motivation** Our work is motivated by a simple observation: when optimizing a NQS to approximate a ground state, the energy is often far more accurate than other observables. To illustrate this, Fig. 1 shows the energy minimization process of a NQS applied to a quantum Ising model, where we see that the error in energy decays faster than that of a correlation function, and reaches values a few orders of magnitude smaller. This discrepancy is notable because the energy does not entirely characterize the ground state; other observables like correlation functions, magnetization, etc. are often of equal physical interest.

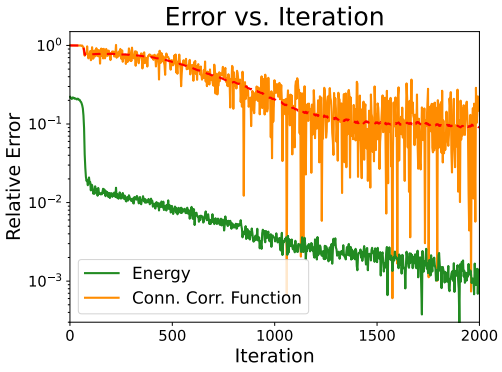


Figure 1: Application of a NQS to the mixed field Ising model on $N = 12$ spins (see Sec. 4 for details). We show relative errors in the energy and a connected correlation function $\langle\sigma^x_{N/2}\sigma^x_{N/2+1}\rangle - \langle\sigma^x_{N/2}\rangle\langle\sigma^x_{N/2+1}\rangle$.

This phenomenon is generic in NQSs, and stems from the ground state being an eigenstate of the Hamiltonian. To see this, consider a NQS that approximates the ground state as

$$|\psi_\theta\rangle = \sqrt{1-\epsilon^2}|E_0\rangle + \epsilon|E_\perp\rangle, \tag{2}$$

where $|E_\perp\rangle$ is a linear combination of excited states, and $\epsilon \ll 1$ characterizes the systematic error.[1] Then, the relative error in an observable $O$ is:

$$\left|\frac{\langle\psi_\theta|O|\psi_\theta\rangle - \langle E_0|O|E_0\rangle}{\langle E_0|O|E_0\rangle}\right| = \left|2\epsilon\sqrt{1-\epsilon^2}\frac{\text{Re}\big(\langle E_0|O|E_\perp\rangle\big)}{\langle E_0|O|E_0\rangle} + \epsilon^2\left(1 - \frac{\langle E_\perp|O|E_\perp\rangle}{\langle E_0|O|E_0\rangle}\right)\right|. \tag{3}$$

[1] Any NQS approximation to the ground state may be expressed in this form. We may always take $\epsilon \in \mathbb{R}$ by absorbing a complex phase into the definition of $|E_\perp\rangle$.

For an arbitrary observable, where $\langle E_0|O|E_\perp\rangle \neq 0$ in general, the relative error is $\mathcal{O}(\epsilon)$. In contrast, for the energy ($O = H$), orthogonality of the eigenstates dictates that $\langle E_0|H|E_\perp\rangle = 0$ and thus reduces the relative error to $\mathcal{O}(\epsilon^2)$.

This observation can be formalized through a metric from quantum information known as the *trace distance*, which measures the deviation between two density matrices $\rho$ and $\sigma$: $d_{\text{tr}}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$. The trace distance is especially useful as it bounds the difference in expectation values [19]:

$$|\operatorname{tr}(\rho O) - \operatorname{tr}(\sigma O)| \leq \|O\|\|\rho - \sigma\|_1 = 2\|O\| d_{\text{tr}}(\rho, \sigma), \tag{4}$$

which is tight for arbitrary observables $O$. For the NQS $|\psi_\theta\rangle$, a straightforward calculation shows that its trace distance with the ground state is $d_{\text{tr}}(|\psi_\theta\rangle\langle\psi_\theta|,\ |E_0\rangle\langle E_0|) = \mathcal{O}(\epsilon)$, which implies that $|\psi_\theta\rangle$ generally incurs $\mathcal{O}(\epsilon)$ errors in expectation values. However, because Eq. (4) is only an upper bound, certain observables can attain smaller errors; as we saw above, the energy enjoys $\mathcal{O}(\epsilon^2)$ error scaling.

At first glance, this discrepancy might appear fundamental because the energy is the quantity being minimized, and may naturally converge faster than other observables. However, in applying machine learning to physics, we can often surpass our naive beliefs by using physical insights, like symmetries and equations of motion. This raises the question: *can we use physical insights to estimate observables with the same accuracy as the energy?* We answer yes, showing how randomized real-time dynamics reduces the trace distance to $\mathcal{O}(\epsilon^2)$, thus improving the accuracy of arbitrary observables.

## 3 Methods

**Increasing the Accuracy of NQSs** Recent work in quantum information has shown that randomization can help in approximating unitary transformations [20–22]. Specifically, randomizing over many approximate transformations can outperform any single approximation of the same cost. Here we show that this principle also extends to NQSs: by randomizing over many NQSs, we can better capture the ground state. This amounts to taking a pure state approximation, i.e. $|\psi_\theta\rangle$, and promoting it to a corresponding mixed state $\rho_\theta$, which more accurately approximates the ground state. Crucially, we do not seek to optimize over NQS representations of mixed states, which would be prohibitively expensive. Rather, we introduce a simple procedure that maps a NQS to a mixed state.

The key idea of our procedure is to randomly apply operations that behave differently on the eigenstates of $H$, enabling us to selectively suppress contributions from excited states. We realize this by evolving $|\psi_\theta\rangle$ in real-time for random durations, thus promoting it to the following mixed state:

$$\rho_\theta = \int dt\ p(t) e^{-iHt}|\psi_\theta\rangle\langle\psi_\theta|e^{iHt}, \tag{5}$$

where $p(t)$ is a probability distribution specified later. Intuitively, this procedure imparts a random phase on the excited state components in $|\psi_\theta\rangle$, which suppresses their contributions and produces a more accurate approximation of the ground state. We make this statement rigorous as follows:

**Theorem 1.** *Suppose a NQS $|\psi_\theta\rangle$ approximates a ground state $|E_0\rangle$ with trace distance error $d_{tr}(|\psi_\theta\rangle\langle\psi_\theta|,\ |E_0\rangle\langle E_0|) \leq \epsilon$. Then, the mixed state $\rho_\theta$ (Eq. (5)) suffers error*

$$d_{tr}\big(\rho_\theta,\ |E_0\rangle\langle E_0|\big) \leq \mathcal{O}(\epsilon^2) + \mathcal{O}\big(\epsilon \cdot \max_{n\geq 1}|C(\Delta_n)|\big), \tag{6}$$

*where $\Delta_n = E_n - E_0$ are the excitation energies of the Hamiltonian, and $C(\Delta_n) = \int dt\ p(t)e^{i\Delta_n t}$ is the characteristic function of $p(t)$.*

For brevity, the proof of this theorem is included in App. A. This result shows that randomizing over real-time evolution provides a more accurate ground state approximation. If the characteristic functions are sufficiently small, i.e. $C(\Delta_n) = \mathcal{O}(\epsilon)$, then the trace distance is suppressed to $\mathcal{O}(\epsilon^2)$. If instead $C(\Delta_n) = \Omega(\epsilon)$, then the trace distance is reduced to $\mathcal{O}(\epsilon \cdot |C(\Delta_n)|)$, which is still an improvement because $|C(\Delta_n)| \leq 1$. Therefore, $\rho_\theta$ achieves an up-to-quadratic reduction in trace distance, which by Eq. (4) translates to an equivalent reduction in the error of arbitrary observables.

A remaining question is how to select a distribution $p(t)$ that ensures a small characteristic function. In practice, $\max_{n\geq 1}|C(\Delta_n)|$ often evaluates to $|C(\Delta_1)|$, and so bounding this quantity generally requires either knowledge of the spectral gap $\Delta_1 = E_1 - E_0$ (or a lower bound thereof), or the ability to adaptively tune the distribution. For instance, if $p(t)$ is chosen to be a zero-mean Gaussian with

variance $\sigma^2$, then $C(\Delta_1) = e^{-\frac{1}{2}\sigma^2\Delta_1^2}$, and setting $\sigma^2 = \mathcal{O}\big(\frac{1}{\Delta_1^2}\log(1/\epsilon)\big)$ ensures $C(\Delta_1) = \mathcal{O}(\epsilon)$. However, if $\Delta_1$ is unknown, one can adaptively tune $\sigma^2$ to achieve a desired accuracy; for example, one can successively increase $\Delta_1$ until $C(\Delta_1) = \mathcal{O}(\epsilon^2)$ and no further improvement in accuracy is observed.

**Realization through Dynamical Averaging** To realize $\rho_\theta$, we follow the framework of variational Monte Carlo (see Eq. (1)) and express an expectation value as

$$\mathrm{tr}(O\rho_\theta) = \int dt\; p(t)\langle\psi_\theta|e^{iHt}Oe^{-iHt}|\psi_\theta\rangle = \mathop{\mathbb{E}}_{t\sim p(t)}\left[\mathop{\mathbb{E}}_{x\sim|\psi_\theta(x,t)|^2}\Big[O_{\mathrm{loc}}(x,t)\Big]\right], \tag{7}$$

where $\psi_\theta(x,t) = \langle x|e^{-iHt}|\psi_\theta\rangle$ is the time-evolved NQS in a basis $|x\rangle$, and $O_{\mathrm{loc}}(x,t) = \langle x|Oe^{-iHt}|\psi_\theta\rangle/\langle x|e^{-iHt}|\psi_\theta\rangle$ is the time-evolved local observable. This introduces a nested sampling procedure: instead of sampling from only $|\psi_\theta(x)|^2$, we sample from time-evolved states $|\psi_\theta(x,t)|^2$, drawn according to $p(t)$.

A convenient way to realize this double expectation value is to evolve $|\psi_\theta(t)\rangle = e^{-iHt}|\psi_\theta\rangle$ in real-time, and collect expectation values $\mathbb{E}_{x\sim|\psi_\theta(x,t)|^2}\big[O_{\mathrm{loc}}(x,t)\big]$ during the evolution. These expectation values can then be aggregated together according to Eq. (7). We call this algorithm *dynamical averaging*, and we will use it to improve the accuracy of NQSs.

In practice, when time-evolution is implemented as a sequence of discrete time steps (using e.g., the time-dependent variational principle), we realize dynamical averaging by evolving $|\psi_\theta(t)\rangle$ in time and estimating the target observables at each time step. Importantly, real-time evolution methods for NQSs already require many samples from the current state $|\psi_\theta(x,t)\rangle$, so computing observables thereof is no more expensive. These expectation values computed at time $t$ are then weighted by $p(t)$ and summed to estimate $\mathrm{tr}(O\rho_\theta)$. As an alternative method, one could also sample many times $t_j \sim p(t)$, pause at these times to estimate the target expectation value, and aggregate these values together to evaluate $\mathrm{tr}(O\rho_\theta)$.

The main overhead of dynamical averaging lies in simulating real-time evolution. While this can be expensive for arbitrary states, it is considerably more efficient when evolving states close to the ground state, as required here. From a theoretical perspective, recent results show that time evolution is significantly less expensive in the low energy subspace [23–26]. On the algorithmic side, numerous accurate and fast methods for evolving NQSs have recently been introduced [13, 14, 27, 28].

These methods inevitably incur errors from two main sources: (1) time discretization, due to finite time step sizes (e.g., some $\Delta t$), and (2) the ability to accurately represent the time-evolved state $|\psi_\theta(t)\rangle$ as an NQS. Fortunately, these errors are not expected to be so severe, as time evolution affects only the excited state components in $|\psi_\theta\rangle$ (beyond a global phase), which have small amplitude $\mathcal{O}(\epsilon)$. Letting $\delta$ denote the discretization error, this adds only a small error $\mathcal{O}(\epsilon\delta)$ to the trace distance in Eq. (6), such that choosing $\delta = \mathcal{O}(\epsilon)$ maintains quadratic error suppression. Representational errors are also expected to be limited, as NQSs can capture highly-entangled states in time evolution [13, 29]. In addition, the trace distance between $|\psi_\theta(0)\rangle$ and $|\psi_\theta(t)\rangle$ never exceeds $\mathcal{O}(\epsilon)$, suggesting that an NQS that already represents $|\psi_\theta(0)\rangle$ should also be able to capture the closely-related state $|\psi_\theta(t)\rangle$.

## 4 Experiments

Dynamical averaging provides high-accuracy ground-state observables, and is particularly useful when further energy minimization only provides diminishing returns or even flatlines. As such, it is most effective when deployed post-energy minimization (rather than in the training process), as a means of refining estimates of ground-state observables.

As introduced in the previous section, the dynamical averaging algorithm is broadly applicable to quantum systems and straightforward to implement in existing NQS frameworks. Here we apply it to the 1D mixed field quantum Ising model on $N$ spins:

$$H = -J\sum_{j=1}^{N}\sigma_j^z\sigma_{j+1}^z - h_x\sum_{j=1}^{N}\sigma_j^x - h_z\sum_{j=1}^{N}\sigma_j^z, \tag{8}$$

We set $J = 1$, $h_x = 1.05$, and $h_z = 0.05$, at which point the model is non-integrable, and consider $N = 12$ spins to allow comparison with the exact solution. We use a NQS parametrized

by restricted Boltzmann machines with feature density $\alpha = 1$ (using NetKet [30, 31]), and employ stochastic reconfiguration to minimize energy. For dynamical averaging, we choose $p(t)$ to be a zero-mean Gaussian with variance $\sigma^2 = 1$, truncated at $|t| \leq 2$, and perform time-evolution using time-dependent VMC [32] via the Euler method with time-step size $dt \in [10^{-4}, 10^{-3}]$.

In Fig. 2, we compare connected correlation functions $\langle \sigma^x_{N/2} \sigma^x_{N/2+j} \rangle_C := \langle \sigma^x_{N/2} \sigma^x_{N/2+j} \rangle - \langle \sigma^x_{N/2} \rangle \langle \sigma^x_{N/2+j} \rangle$ (defined analogously for $\sigma^z$), evaluated with both standard NQS methods and dynamical averaging. We see that dynamical averaging consistently achieves smaller errors, both during optimization and at the final iteration. The improvement is most pronounced in the $\sigma^x$ correlations, where the error decreases from $\gtrsim 10\%$ to $\lesssim 1\%$. This highlights the value of dynamical averaging in improving estimates of important ground-state observables.

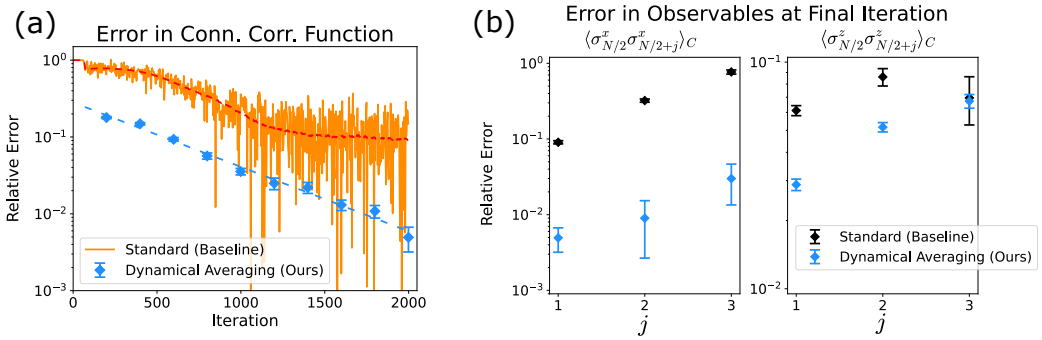


Figure 2: (a) Error in the connected correlation function $\langle \sigma^x_{N/2} \sigma^x_{N/2+1} \rangle - \langle \sigma^x_{N/2} \rangle \langle \sigma^x_{N/2+1} \rangle$. (b) Errors in various connected correlation functions at the final iteration.

## 5 Conclusion and Discussion

We have used insights from quantum information to develop a new NQS algorithm that improves ground state approximation. At a conceptual level, dynamical averaging exploits a fundamental feature of quantum mechanics: a quantum state is quadratic in the state vector, i.e. $\rho_\psi = |\psi\rangle\langle\psi|$, while typical NQS methods optimize only the state vector $|\psi_\theta\rangle$. By averaging over many such state vectors, we promote the variational object to a mixed state that better approximates the ground state.

We note for completeness that dynamical averaging has its limitations. While it achieves $\mathcal{O}(\epsilon^2)$ errors for arbitrary observables, it does not uniformly reduce the error of every observable. As we noted in the Sec. 2, the error in the energy is already $\mathcal{O}(\epsilon^2)$ small when using standard NQS methods, so dynamical averaging cannot improve it further. In fact, errors in real-time evolution can slightly increase the error in energy. For this reason, dynamical averaging is best suited for estimating observables other than the energy, where it achieves genuine improvements over standard methods.

Looking ahead, our work suggests several extensions. For one, dynamical averaging could be applied to excited state NQSs [33, 34] to more accurately target higher energy states. Furthermore, dynamical averaging can be generalized beyond real-time dynamics. In general, any operator that acts differently on the eigenstates of $H$ can, when appropriately randomized, suppress contributions from undesired states. For instance, one could randomize over symmetry operations to selectively target states in certain symmetry sectors. In the broader context of computational physics, this work illustrates that NQSs can benefit greatly from a synthesis with quantum information, which provides both the conceptual and technical tools to design better machine-learning-based simulation algorithms.

*Acknowledgments*: This research was supported by PNNL's Quantum Algorithms and Architecture for Domain Science (QuAADS) Laboratory Directed Research and Development (LDRD) Initiative. The Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830.

## References


[1] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021. URL: https://doi.org/10.1038/s41586-021-03819-2.

[2] Dan Guest, Kyle Cranmer, and Daniel Whiteson. Deep learning and its application to LHC physics. *Annual Review of Nuclear and Particle Science*, 68(1):161–181, October 2018. URL: http://dx.doi.org/10.1146/annurev-nucl-101917-021019, doi:10.1146/annurev-nucl-101917-021019.

[3] Joe G Greener, Shaun M Kandathil, Lewis Moffat, and David T Jones. A guide to machine learning for biologists. *Nature reviews Molecular cell biology*, 23(1):40–55, 2022. doi:10.1038/s41580-021-00407-0.

[4] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Rev. Mod. Phys.*, 91:045002, Dec 2019. URL: https://link.aps.org/doi/10.1103/RevModPhys.91.045002, doi:10.1103/RevModPhys.91.045002.

[5] U. Schollwöck. The density-matrix renormalization group. *Rev. Mod. Phys.*, 77:259–315, Apr 2005. URL: https://link.aps.org/doi/10.1103/RevModPhys.77.259, doi:10.1103/RevModPhys.77.259.

[6] Guifré Vidal. Efficient simulation of one-dimensional quantum many-body systems. *Phys. Rev. Lett.*, 93:040502, Jul 2004. URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.040502, doi:10.1103/PhysRevLett.93.040502.

[7] F. Verstraete, V. Murg, and J.I. Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in Physics*, 57(2):143–224, 2008. doi:10.1080/14789940801912366.

[8] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum monte carlo simulations of solids. *Rev. Mod. Phys.*, 73:33–83, Jan 2001. URL: https://link.aps.org/doi/10.1103/RevModPhys.73.33, doi:10.1103/RevModPhys.73.33.

[9] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017. URL: https://www.science.org/doi/abs/10.1126/science.aag2302, doi:10.1126/science.aag2302.

[10] Xun Gao and Lu-Ming Duan. Efficient representation of quantum many-body states with deep neural networks. *Nature Communications*, 8(1):662, September 2017. URL: http://dx.doi.org/10.1038/s41467-017-00705-2.

[11] Or Sharir, Amnon Shashua, and Giuseppe Carleo. Neural tensor contractions and the expressive power of deep neural quantum states. *Phys. Rev. B*, 106:205136, Nov 2022. URL: https://link.aps.org/doi/10.1103/PhysRevB.106.205136, doi:10.1103/PhysRevB.106.205136.

[12] Juan Carrasquilla and Giacomo Torlai. How to use neural networks to investigate quantum many-body physics. *PRX Quantum*, 2:040201, Nov 2021. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.040201, doi:10.1103/PRXQuantum.2.040201.

[13] Irene López Gutiérrez and Christian B. Mendl. Real time evolution with neural-network quantum states. *Quantum*, 6:627, January 2022. URL: http://dx.doi.org/10.22331/q-2022-01-20-627, doi:10.22331/q-2022-01-20-627.

[14] Alessandro Sinibaldi, Clemens Giuliani, Giuseppe Carleo, and Filippo Vicentini. Unbiasing time-dependent variational monte carlo by projected quantum evolution. *Quantum*, 7:1131, October 2023. URL: http://dx.doi.org/10.22331/q-2023-10-10-1131, doi:10.22331/q-2023-10-10-1131.

[15] John M. Martyn, Khadijeh Najafi, and Di Luo. Variational neural-network ansatz for continuum quantum field theory. *Physical Review Letters*, 131(8), August 2023. URL: http://dx.doi.org/10.1103/PhysRevLett.131.081601, doi:10.1103/PhysRevLett.131.081601.

[16] Di Luo, David D. Dai, and Liang Fu. Pairing-based graph neural network for simulating quantum materials, 2023. arXiv:2311.02143.

[17] Alexei Yu Kitaev, Alexander Shen, and Mikhail N Vyalyi. *Classical and quantum computation*. Number 47. American Mathematical Soc., 2002.

[18] Federico Becca and Sandro Sorella. *Quantum Monte Carlo approaches for correlated systems*. Cambridge University Press, 2017.

[19] Mark M Wilde. *Quantum information theory*. Cambridge university press, 2013.

[20] M. B. Hastings. Turning gate synthesis errors into incoherent errors, 2016. arXiv:1612.01011.

[21] Earl Campbell. Shorter gate sequences for quantum computing by mixing unitaries. *Phys. Rev. A*, 95:042306, Apr 2017. URL: https://link.aps.org/doi/10.1103/PhysRevA.95.042306, doi:10.1103/PhysRevA.95.042306.

[22] Earl Campbell. Random compiler for fast hamiltonian simulation. *Physical Review Letters*, 123(7):070503, August 2019. URL: http://dx.doi.org/10.1103/PhysRevLett.123.070503.

[23] Guang Hao Low and Isaac L. Chuang. Hamiltonian simulation by uniform spectral amplification, 2017. arXiv:1707.05391.

[24] Burak Şahinoğlu and Rolando D. Somma. Hamiltonian simulation in the low-energy subspace. *npj Quantum Information*, 7(1), July 2021. URL: http://dx.doi.org/10.1038/s41534-021-00451-w.

[25] Alexander Zlokapa and Rolando D. Somma. Hamiltonian simulation for low-energy states with optimal time dependence. *Quantum*, 8:1449, August 2024. URL: http://dx.doi.org/10.22331/q-2024-08-27-1449, doi:10.22331/q-2024-08-27-1449.

[26] Kaoru Mizuta and Tomotaka Kuwahara. Trotterization is substantially efficient for low-energy states, 2025. arXiv:2504.20746.

[27] Anka Van de Walle, Markus Schmitt, and Annabelle Bohrdt. Many-body dynamics with explicitly time-dependent neural quantum states, 2024. URL: https://arxiv.org/abs/2412.11830, arXiv:2412.11830.

[28] Luca Gravina, Vincenzo Savona, and Filippo Vicentini. Neural projected quantum dynamics: a systematic study. *Quantum*, 9:1803, July 2025. URL: http://dx.doi.org/10.22331/q-2025-07-22-1803, doi:10.22331/q-2025-07-22-1803.

[29] Markus Schmitt and Markus Heyl. Quantum many-body dynamics in two dimensions with artificial neural networks. *Phys. Rev. Lett.*, 125:100503, Sep 2020. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.100503, doi:10.1103/PhysRevLett.125.100503.

[30] Giuseppe Carleo, Kenny Choo, Damian Hofmann, James E. T. Smith, Tom Westerhout, Fabien Alet, Emily J. Davis, Stavros Efthymiou, Ivan Glasser, Sheng-Hsuan Lin, Marta Mauri, Guglielmo Mazzola, Christian B. Mendl, Evert van Nieuwenburg, Ossian O'Reilly, Hugo Théveniaut, Giacomo Torlai, Filippo Vicentini, and Alexander Wietek. Netket: A machine learning toolkit for many-body quantum systems. *SoftwareX*, 10:100311, 2019. URL: http://www.sciencedirect.com/science/article/pii/S2352711019300974, doi:10.1016/j.softx.2019.100311.

[31] Filippo Vicentini, Damian Hofmann, Attila Szabó, Dian Wu, Christopher Roth, Clemens Giuliani, Gabriel Pescia, Jannes Nys, Vladimir Vargas-Calderón, Nikita Astrakhantsev, and Giuseppe Carleo. NetKet 3: Machine Learning Toolbox for Many-Body Quantum Systems. *SciPost Phys. Codebases*, page 7, 2022. URL: https://scipost.org/10.21468/SciPostPhysCodeb.7, doi:10.21468/SciPostPhysCodeb.7.

[32] Giuseppe Carleo, Federico Becca, Marco Schiró, and Michele Fabrizio. Localization and glassy dynamics of many-body quantum systems. *Scientific Reports*, 2(1):243, February 2012. URL: http://dx.doi.org/10.1038/srep00243.

[33] M. T. Entwistle, Z. Schätzle, P. A. Erdman, J. Hermann, and F. Noé. Electronic excited states in deep variational monte carlo. *Nature Communications*, 14(1):274, January 2023. URL: http://dx.doi.org/10.1038/s41467-022-35534-5, doi:10.1038/s41467-022-35534-5.

[34] David Pfau, Simon Axelrod, Halvard Sutterud, Ingrid von Glehn, and James S. Spencer. Accurate computation of quantum excited states with neural networks. *Science*, 385(6711):eadn0137, August 2024. URL: http://dx.doi.org/10.1126/science.adn0137.

## A Proof of Theorem 1

*Proof.* For simplicity, we will take the ground state energy to be zero, $E_0 = 0$, and make appropriate replacements at the end of the proof. Consequently, the excitation energies are $\Delta_n = E_n - E_0 = E_n$. This effectively shifts the energy by a constant, which only changes physically-irrelevant global phases.

To begin, let us expand $|\psi_\theta\rangle$ in the eigenbasis of $H$ as $|\psi_\theta\rangle = \sum_n a_n |E_n\rangle$, where we take $a_0 \in \mathbb{R}$ and $a_0 > 0$ without loss of generality. Then, for pure states $|\psi_\theta\rangle$ and $|E_0\rangle$, their trace distance evaluates to [19]

$$d_{\text{tr}}(|\psi_\theta\rangle\langle\psi_\theta|, |E_0\rangle\langle E_0|) = \sqrt{1 - |\langle E_0|\psi_\theta\rangle|^2} \leq \epsilon. \tag{9}$$

This implies that $a_0 = \langle E_0|\psi_\theta\rangle \geq \sqrt{1-\epsilon^2} = 1 - \mathcal{O}(\epsilon^2)$.

We can then prove the theorem by direct computation. Let us denote the time-evolved state as $|\psi_\theta(t)\rangle = e^{-iHt}|\psi_\theta\rangle$, and the density matrix of interest as $\rho_\theta = \mathbb{E}_{p(t)}\big[|\psi_\theta(t)\rangle\langle\psi_\theta(t)|\big]$. Let us also define $|\delta(t)\rangle := |\psi_\theta(t)\rangle - |E_0\rangle$. We can express the 1-norm that appears in $d_{\text{tr}}(\rho_\theta, |E_0\rangle\langle E_0|)$ as

$$\begin{aligned}
\big\|\rho_\theta - |E_0\rangle\langle E_0|\big\|_1 &= \Big\|\mathbb{E}\big[|\psi_\theta(t)\rangle\langle\psi_\theta(t)| - |E_0\rangle\langle E_0|\big]\Big\|_1 \\
&= \Big\|\mathbb{E}\big[|\delta(t)\rangle\langle\delta(t)| + |\delta(t)\rangle\langle E_0| + |E_0\rangle\langle\delta(t)|\big]\Big\|_1 \\
&\leq \Big\|\mathbb{E}\big[|\delta(t)\rangle\langle\delta(t)|\big]\Big\|_1 + \Big\|\mathbb{E}\big[|\delta(t)\rangle\big]\langle E_0|\Big\|_1 + \Big\||E_0\rangle \cdot \mathbb{E}\big[\langle\delta(t)|\big]\Big\|_1
\end{aligned} \tag{10}$$

where we have invoked the triangle inequality in obtaining the last line.

The first term is upper bounded as $\big\|\mathbb{E}\big[|\delta(t)\rangle\langle\delta(t)|\big]\big\|_1 \leq \mathbb{E}\big[\||\delta(t)\rangle\langle\delta(t)|\|_1\big]$. To evaluate this, recall that the 1-norm is the sum of the singular values. Noting that $|\delta(t)\rangle\langle\delta(t)|$ is itself an un-normalized singular value decomposition, its 1-norm is the squared 2-norm of $|\delta(t)\rangle$:

$$\big\||\delta(t)\rangle\langle\delta(t)|\big\|_1 = \langle\delta(t)|\delta(t)\rangle = 2 - 2a_0 \leq \mathcal{O}(\epsilon^2). \tag{11}$$

Therefore, we have that $\big\|\mathbb{E}\big[|\delta(t)\rangle\langle\delta(t)|\big]\big\|_1 \leq \mathcal{O}(\epsilon^2)$.

The next term in Eq. (10) similarly evaluates to

$$\Big\|\mathbb{E}\big[|\delta(t)\rangle\big]\langle E_0|\Big\|_1 = \big\|\mathbb{E}[\delta(t)\rangle]\big\|_2 = \Big\||E_0\rangle - \int dt p(t) e^{-iHt}|\psi_\theta\rangle\Big\|_2. \tag{12}$$

Inserting $|\psi_\theta\rangle = \sum_n a_n|E_n\rangle$, and using the orthogonality of the eigenstates, this norm evaluates to

$$\begin{aligned}
&\Big[1 - 2a_0 + \sum_n |a_n|^2 \int dt dt' p(t)p(t') e^{-iE_n(t-t')}\Big]^{1/2} = \Big[1 - 2a_0 + a_0^2 + \sum_{n\geq 1} |a_n|^2 |C(E_n)|^2\Big]^{1/2} \\
&\leq \Big[(1-a_0)^2 + \Big(\sum_{n\geq 1} |a_n|^2\Big) \cdot \max_{n\geq 1} |C(E_n)|^2\Big]^{1/2} \leq \Big[\mathcal{O}(\epsilon^4) + \mathcal{O}\big(\epsilon^2 \cdot \max_{n\geq 1} |C(E_n)|^2\big)\Big]^{1/2} \\
&= \mathcal{O}(\epsilon^2) + \mathcal{O}\big(\epsilon \cdot \max_{n\geq 1} |C(E_n)|\big)
\end{aligned} \tag{13}$$

where $C(E_n) = \int dt\ p(t)e^{iE_n t}$ is the characteristic function of $p(t)$, and we have used that $\sum_{n\geq 1} |a_n|^2 = 1 - a_0^2 \leq \mathcal{O}(\epsilon^2)$. Furthermore, the third term in Eq. (10) is analogously upper bounded.

Collecting all the above terms, we have:

$$\big\|\rho_\theta - |E_0\rangle\langle E_0|\big\|_1 \leq \mathcal{O}(\epsilon^2) + O\big(\epsilon \cdot \max_{n\geq 1} |C(E_n)|\big). \tag{14}$$

Then, reintroducing a non-zero ground state energy $E_0$ amounts to shifting the energies $E_n$ to excitation energies $E_n - E_0 = \Delta_n$. This yields the stated bound:

$$d_{\text{tr}}\big(\rho_\theta,\ |E_0\rangle\langle E_0|\big) = \frac{1}{2}\big\|\rho_\theta - |E_0\rangle\langle E_0|\big\|_1 \leq \mathcal{O}(\epsilon^2) + O\big(\epsilon \cdot \max_{n\geq 1} |C(\Delta_n)|\big). \tag{15}$$

For completeness, we note that the same result can alternatively be derived using the concept of *randomized compiling* [20, 21], as is often considered in quantum computing contexts.

□