

# Exploring Entanglement and Optimization within the Hamiltonian Variational Ansatz

Roeland Wiersema,<sup>1,2,\*</sup> Cunlu Zhou<sup>3,1,†</sup> Yvette de Sereville,<sup>2,4</sup> Juan Felipe Carrasquilla,<sup>1,2</sup>  
Yong Baek Kim,<sup>5</sup> and Henry Yuen<sup>6</sup>

<sup>1</sup>Vector Institute, MaRS Centre, Toronto, Ontario M5G 1M1, Canada

<sup>2</sup>Department of Physics and Astronomy, University of Waterloo, Ontario N2L 3G1, Canada

<sup>3</sup>Department of Computer Science, University of Toronto, Ontario M5T 3A1, Canada

<sup>4</sup>Institute for Quantum Computing, University of Waterloo, Ontario N2L 3G1, Canada

<sup>5</sup>Department of Physics, University of Toronto, Ontario M5S 1A7, Canada

<sup>6</sup>Department of Computer Science and Department of Mathematics, University of Toronto, Ontario M5T 3A1, Canada



(Received 10 August 2020; accepted 16 November 2020; published 8 December 2020)

Quantum variational algorithms are one of the most promising applications of near-term quantum computers; however, recent studies have demonstrated that unless the variational quantum circuits are configured in a problem-specific manner, optimization of such circuits will most likely fail. In this paper, we focus on a special family of quantum circuits called the Hamiltonian variational ansatz (HVA), which takes inspiration from the quantum approximate optimization algorithm and adiabatic quantum computation. Through the study of its entanglement spectrum and energy-gradient statistics, we find that the HVA exhibits favorable structural properties such as mild or entirely absent barren plateaus and a restricted state space that eases their optimization in comparison to the well-studied “hardware-efficient ansatz.” We also numerically observe that the optimization landscape of the HVA becomes almost trap free, i.e., there are no suboptimal minima, when the ansatz is overparametrized. We observe a size-dependent “computational phase transition” as the number of layers in the HVA circuit is increased where the optimization crosses over from a hard to an easy region in terms of the quality of the approximations and the speed of convergence to a good solution. In contrast to the analogous transitions observed in the learning of random unitaries, which occur at a number of layers that grows exponentially with the number of qubits, our variational-quantum-eigensolver experiments suggest that the threshold to achieve the overparametrization phenomenon scales at most polynomially in the number of qubits for the transverse-field Ising and XXZ models. Lastly, as a demonstration of its entangling power and effectiveness, we show that the HVA can find accurate approximations to the ground states of a modified Haldane-Shastry Hamiltonian on a ring, which has long-range interactions and has a power-law entanglement scaling.

DOI: [10.1103/PRXQuantum.1.020319](https://doi.org/10.1103/PRXQuantum.1.020319)

## I. INTRODUCTION

With the advent of noisy intermediate-scale quantum (NISQ) computers [1], near-term quantum algorithms, such as variational quantum eigensolvers (VQEs), may offer computational capabilities beyond those of the best

current classical computers and algorithms for approximating ground states of quantum many-body systems. A VQE algorithm contains three ingredients: a variational-quantum-circuit ansatz specified by a set of parameters  $\theta$ ; an energy function given by the expectation value of a local Hamiltonian  $H$ , composed of local measurements on the variational circuit state; and a classical optimizer. A natural first approach is the random-quantum-circuit ansatz [2–4], which is capable of expressing a wide variety of states. However, this has been shown to be ineffective for gradient-based optimization strategies due to the barren-plateau phenomenon [5–8], which causes the optimization of randomly initialized circuits to get stuck on flat areas in the cost landscape, where gradients are exponentially small. These observations suggest that an effective

\*rwiersema@uwaterloo.ca

†czhou@cs.toronto.edu

‡These authors contributed equally to this work.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

ansatz for a VQE requires a circuit that is problem specific, such that the optimization landscape of the problem is not hindered by barren plateaus. For quantum many-body problems, Ref. [9] suggests a novel variational circuit that is now called the Hamiltonian variational ansatz (HVA). While there is no rigorous proof that the HVA will be an effective ansatz, recent work has demonstrated that it is rather effective for several one- and two-dimensional (1D and 2D, respectively) quantum many-body models [10,11]. It is thus an intriguing question to further understand the empirically observed effectiveness of the HVA.

For the purpose of understanding the effectiveness of such *Ansätze*, it is useful to note that quantum entanglement provides a window into the capabilities of several families of numerical techniques and algorithms aimed at understanding the properties of quantum many-body states, as well as helping us delineate the boundary between quantum states that can be simulated classically and those that call for quantum simulators and quantum computers for their accurate description. For instance, for a 1D gapped local Hamiltonian, the entanglement entropy of the ground state obeys an area law, i.e., the entanglement entropy grows proportional to the boundary area of the system instead of the system size [12]. This remarkable result allows us to combat the exponential scaling of the Hilbert space, since this area law provides evidence that the relevant physics of a system only takes place in a restricted part of the full state space. These observations have inspired a variety of variational numerical methods, most notably, tensor-network approaches such as the matrix-product state (MPS), multiscale entanglement renormalization, and projected entangled pair states [13] but also deep-learning-inspired variational approaches, which have been successful at representing quantum many-body states [14–17].

In this paper, we study various entanglement properties of the HVA and present several results on the favorable features of the HVA that shed light on the underlying reasons for its effectiveness for solving natural many-body problems. Our findings suggest that the HVA is highly expressive but yet structured enough to allow for efficient optimization. Through the study of two prototypical models in condensed-matter physics, namely the 1D transverse-field Ising model (TFIM) and the XXZ model, we find that entanglement entropy and entanglement spectrum can shed light on the initialization and optimization properties of the HVA in the context of the VQE algorithm. Whereas the HVA provides a restricted and effective state space for the TFIM, which yields ground-state approximations largely insensitive to the circuit initialization, the 1D-XXZ-model ansatz requires a careful parameter initialization for its successful optimization. Through the study of the dynamics of the entanglement spectrum during the optimization of the XXZ model, we find that initializing the HVA near the identity operator enables a restricted and

effective subspace during optimization that yields accurate approximations to the ground state with fast convergence. Furthermore, we show evidence that the gradient vanishing problem in the HVA, especially if the HVA is initialized near the identity operator, is mild or entirely absent in comparison to the random circuit ansatz, where barren plateaus in the energy landscape cause gradients to decay exponentially with increasing system size. We also explore the overparametrization phenomena in the HVA and observe a “computational phase transition” between an underparametrized and an overparametrized regime, where the optimization landscape of the HVA crosses over to a regime with faster convergence and an absence of low-quality solutions. Lastly, as a demonstration of the entangling power and effectiveness of the HVA, we study a modified Haldane-Shastry (MHS) Hamiltonian, which has long-range interactions and a power-law scaling entanglement entropy [18]. We observe that the HVA can find approximations to the ground state of the MHS Hamiltonian reaching fidelities  $> 99\%$  for system sizes  $N = 4$ ,  $N = 8$ ,  $N = 12$ , and  $N = 16$  and circuit depths  $p = N$ . Our findings point to important features of the HVA that will lead to a deeper understanding of its effectiveness and point the way to developing more sophisticated *Ansätze* for other many-body problems, as well as more informed optimization strategies. Moreover, we establish a substantial connection between quantum entanglement and the efficacy of the HVA and show how entanglement properties such as the entanglement spectrum can be used to study variational quantum circuits. Furthermore, the surprising phenomenon of overparametrization in the HVA signals a nontrivial connection with deep neural networks, which merits further investigation.

In Sec. II, we introduce the basic concepts of the VQE and the HVA. In Sec. III, we introduce two paradigmatic quantum many-body models that we use in our study, the TFIM and the XXZ model, as well as their respective *Ansätze*. We also introduce the necessary entanglement concepts used in this paper. Our main results are presented in Sec. IV and we conclude in Sec. V. In the appendices, we include the computational details in Appendix A, some additional numerical results in Appendix B, and extra results on the dynamics of entanglement entropy in Appendix C.

## II. VARIATIONAL QUANTUM EIGENSOLVER AND HAMILTONIAN VARIATIONAL ANSATZ

The VQE [19] is a hybrid classical-quantum algorithm for finding eigenstates of a quantum many-body Hamiltonian. According to the variational principle of quantum mechanics, a parametrized wave function  $|\psi(\boldsymbol{\theta})\rangle$  provides an upper bound on the ground-state energy,

$$E_{\text{ground}} \leq \langle \psi(\boldsymbol{\theta}) | H | \psi(\boldsymbol{\theta}) \rangle = E(\boldsymbol{\theta}), \quad (1)$$

where  $H$  is a  $k$ -local lattice Hamiltonian. Hence, we can approximate the ground state by minimizing  $E(\boldsymbol{\theta})$  with respect to the parameters  $\boldsymbol{\theta}$ . In the case of the VQE, the wave function  $|\psi(\boldsymbol{\theta})\rangle$  corresponds to a depth- $p$  quantum circuit specified by a unitary matrix  $U(\boldsymbol{\theta})$ , i.e.,  $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle$ , where a number of  $m$  parameters specify the unitary  $\boldsymbol{\theta} \in \mathbb{R}^m$ . We can estimate the variational energy  $E(\boldsymbol{\theta})_p$ , where  $E(\boldsymbol{\theta})_p$  denotes the energy at the  $p$ -level circuit, by measuring the observables that compose the Hamiltonian of the system over the quantum state  $U(\boldsymbol{\theta})|0\rangle$ . We use a classical optimization procedure to find the optimal parameters  $\boldsymbol{\theta}^*$  that minimize the energy.

As with other variational methods for approximating the ground state, a key ingredient in the success of the method is finding a good parametrization scheme for the wave function. Ideally, the manifold of states parametrized by the ansatz of choice contains the ground state of interest and this ground state can be reached using a numerical optimization. The HVA [9] is a quantum circuit ansatz inspired by the quantum approximate optimization algorithm (QAOA) [20] and adiabatic computation [21]. Instead of using only two (noncommuting) operators as in QAOA, the HVA uses more terms of the Hamiltonian. More specifically,

$$H = \sum_s H_s, \quad (2)$$

where we assume that each pair of  $H_s$  and  $H_{s'}$  does not commute, i.e.,  $[H_s, H_{s'}] \neq 0$ . A depth- $p$  HVA is given by

$$|\psi_p\rangle = \prod_{l=1}^p \left( \prod_s \exp -i\theta_{s,l} H_s \right) |\psi_0\rangle, \quad (3)$$

where  $|\psi_0\rangle$  is the ground state of one of the terms in Eq. (2), i.e.,  $H_{s_0}$ . When ordering the unitaries, we make sure that  $H_{s_0}$  is not the first  $H_s$  acting on  $|\psi_0\rangle$ . Note that due to the periodicity of the complex exponent, we can restrict the parameters to  $[0, 2\pi]$ , although in the case of certain symmetries, this restriction can be made tighter without losing expressive power [10]. Since these circuits are model specific, the properties of the circuit can vary per problem. We give some concrete examples of the HVA in the next section.

### III. METHODS AND MODELS

#### A. Models

##### 1. Transverse-field Ising model

The TFIM is a paradigmatic model for studies of quantum magnetism. The Hamiltonian for the 1D chain is

given by

$$H_{\text{TFIM}} = - \sum_{i=1}^N [\sigma_i^z \sigma_{i+1}^z + g \sigma_i^x] = H_{zz} + g H_x, \quad (4)$$

with  $H_{zz} = - \sum_{i=1}^N \sigma_i^z \sigma_{i+1}^z$  and  $H_x = - \sum_{i=1}^N \sigma_i^x$ , where we assume that  $g > 0$  and use periodic boundary conditions  $\sigma_{N+1}^z \equiv \sigma_1^z$ . Here,  $\sigma_i^\alpha$  corresponds to a Pauli matrix  $\alpha = x, y, z$  acting on a site  $i$ , where the Pauli matrices are defined as follows:

$$\sigma_x \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The Hamiltonian has a  $\mathbb{Z}_2$  symmetry, so it is invariant under the operation of flipping all spins.

For  $g < 1$ , the system is in a ferromagnetic phase in which the Hamiltonian favors spin alignment along the  $z$  direction. For  $g > 1$ , the system transitions to a disordered paramagnetic phase. In the limit as  $g \rightarrow \infty$ , the  $\sigma^x$  term dominates the Hamiltonian and the ground state becomes  $|+\rangle^{\otimes N}$ . At  $g = 1$ , there is a critical point and the system becomes gapless in the thermodynamic limit.

A depth- $p$  HVA circuit for the TFIM corresponds to

$$U_{\text{TFIM}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{l=1}^p \exp \left\{ -i \frac{\beta_l}{2} H_{zz} \right\} \exp \left\{ -i \frac{\gamma_l}{2} H_x \right\}. \quad (5)$$

Hence, for a depth- $p$  circuit, we have  $2p$  parameters. Figure 1 illustrates the corresponding quantum circuit for  $N = 4$  and  $p = 1$ . Note that we choose  $|\psi_0\rangle$  in Eq. (3) to be the ground state of  $H_x = - \sum_{i=1}^N \sigma_i^x$ , i.e.,  $|\psi_0\rangle = |+\rangle^{\otimes N}$ . The HVA circuit of Eq. (5) is the same as the QAOA ansatz used in Ref. [20] for solving the MaxCut problem.

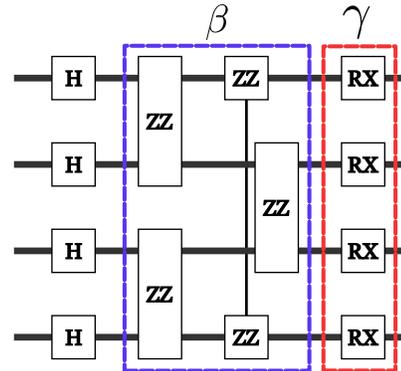


FIG. 1. The HVA quantum circuit for the TFIM with  $p = 1$ . The first layer of Hadamard gates, represented by  $H$ , are used to construct the initial  $|+\rangle$  state. The ZZ gates are 2-local qubit rotation gates of the form  $ZZ = \exp\{i\beta_l/2\sigma_i^z\sigma_j^z\}$ . The RX gates are single-qubit rotation gates  $RX = \exp\{i\gamma_l/2\sigma_i^x\}$ .

By using the Jordan-Wigner transformation, it has been shown that the ground state can be represented accurately with a depth  $p = N/2$  circuit for the case in which  $g = 0$  [22]. For the case in which  $g \neq 0$ , there is only numerical evidence to support this claim [10,23]. In Appendix B, we confirm that for the TFIM, one can consistently find the ground state for  $g \in \{0.5, 0.52, \dots, 1.5\}$  with a depth  $p = N/2$  circuit.

## 2. XXZ model

Another prototypical model for studying quantum magnetism is the XXZ model. For the 1D XXZ model, the Hamiltonian is given by

$$H_{\text{XXZ}} = \sum_{i=1}^N [\sigma_i^x \sigma_{i+1}^x + \sigma_i^y \sigma_{i+1}^y + \Delta \sigma_i^z \sigma_{i+1}^z] = H_{xx} + H_{yy} + \Delta H_{zz}, \quad (6)$$

with  $H_{xx} = \sum_{i=1}^N \sigma_i^x \sigma_{i+1}^x$ ,  $H_{yy} = \sum_{i=1}^N \sigma_i^y \sigma_{i+1}^y$  and  $H_{zz} = \sum_{i=1}^N \sigma_i^z \sigma_{i+1}^z$ . Again, we assume periodic boundary conditions. The parameter  $\Delta$  controls the spin anisotropy in the model. For  $\Delta = 1$ , this model has an  $SU(2)$  symmetry and is equivalent to the Heisenberg chain. For  $\Delta \neq 1$ , this symmetry gets reduced to a  $U(1) \times \mathbb{Z}_2$  symmetry. For  $1 < |\Delta|$ , the system is in the XY quasi-long-range ordered state and becomes gapless in the thermodynamic limit. At  $|\Delta| = 1$ , there is a phase transition to the Néel ordered state. This model can be solved exactly using the Bethe ansatz for  $N \rightarrow \infty$  [24].

Inspired by Ref. [10], we decompose the 1D chain into even and odd links and separate the Hamiltonian into two parts,

$$H^{\text{even}} = H_{xx}^{\text{even}} + H_{yy}^{\text{even}} + H_{zz}^{\text{even}}, \\ H^{\text{odd}} = H_{xx}^{\text{odd}} + H_{yy}^{\text{odd}} + H_{zz}^{\text{odd}},$$

where the indices only run over nonoverlapping bonds:

$$H_{\alpha\alpha}^{\text{even}} = \sum_{i=1}^{N/2} \sigma_{2i-1}^{\alpha} \sigma_{2i}^{\alpha} \quad \text{and} \quad H_{\alpha\alpha}^{\text{odd}} = \sum_{i=1}^{N/2} \sigma_{2i}^{\alpha} \sigma_{2i+1}^{\alpha}$$

for  $\alpha = x, y, z$ . Our numerical experiments indicate that separately parametrizing these bonds gives better performance when studying the anisotropic system  $\Delta \neq 1$ . Additionally, we parametrize the  $H_{xx}, H_{yy}$  and  $H_{zz}$  terms with their own respective parameters. The reason for this is that for  $\Delta \neq 1$ , the anisotropy in the model cannot be accounted for by a single parameter. A depth- $p$  HVA

circuit for the XXZ model corresponds to

$$U_{\text{XXZ}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{l=1}^p G(\theta_l, H_{zz}^{\text{odd}}) G(\phi_l, H_{xx}^{\text{odd}}) G(\phi_l, H_{yy}^{\text{odd}}) \\ G(\beta_l, H_{zz}^{\text{even}}) G(\gamma_l, H_{xx}^{\text{even}}) G(\gamma_l, H_{yy}^{\text{even}}), \quad (7)$$

where

$$G(x, H) = \exp -i \frac{x}{2} H.$$

Hence for a depth- $p$  circuit, we have  $4p$  parameters. Figure 2 illustrates a quantum circuit for  $N = 4$  and  $p = 1$ . We choose the initial state  $|\psi_0\rangle$  in Eq. (3) to be the ground state of  $H^{\text{even}}$ , i.e.,  $|\psi_0\rangle = \bigotimes_{i=1}^{N/2} \frac{1}{\sqrt{2}} (|01\rangle - |10\rangle)_{2i-1, 2i} = \bigotimes_{i=1}^{N/2} |\Psi^-\rangle$ . It has been shown in Ref. [10] that the Heisenberg chain (i.e.,  $\Delta = 1$ ) can be solved accurately using the HVA with  $p = N/2$ . Note that for the case of  $\Delta = 1$ , one can use a single parameter for  $H_{xx} + H_{yy} + H_{zz}$ . In Appendix B, we find that for  $\Delta \in \{0.5, 0.52, \dots, 1.5\}$  and  $\Delta \neq 1$ , a depth  $p = N/2$  HVA circuit is sufficient to find a close approximation to the ground state.

In this work, we consider the problem of approximating the ground state at the critical points  $g = 1$  and  $\Delta = 1$  for the TFIM and the XXZ model, respectively, since their particular entanglement scaling properties makes them harder to approximate using classical methods [25], such as the density-matrix renormalization group (DMRG). Due to the criticality of the aforementioned systems at these order values, the energy spectrum becomes gapless in the thermodynamic limit and hence there is a logarithmic correction of  $S \propto \log N$  to the area law of entanglement entropy. A matrix-product state with bond dimension  $D$  bounds the entanglement of the state to  $S \leq 2 \log D$ , so the necessary bond dimension to express the ground state grows polynomially in a DMRG calculation [25].

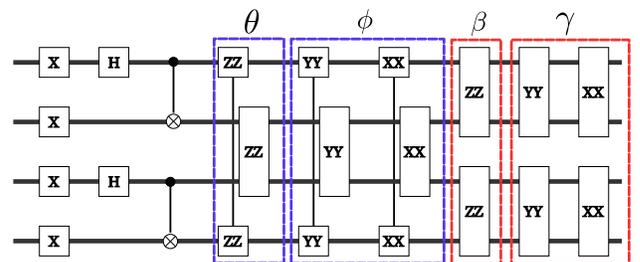


FIG. 2. The HVA quantum circuit for the XXZ model with  $p = 1$ . Here, the X gates are given by  $X = \sigma_i^x$ . Together with a single Hadamard gate and a controlled-NOT on even links, we prepare the  $|\Psi^-\rangle$  Bell state. The 2-local qubit rotations are all of the form  $AA = \exp\{-ix/2\sigma_i^{\alpha}\sigma_j^{\alpha}\}$ , with  $x = \theta, \phi, \beta, \gamma$  depending on whether the links are even or odd and ZZ or XX, YY [see Eq. (7)].

### 3. Performance metrics

We use the *fidelity*  $\mathcal{F}$  between the VQE optimized state  $|\psi(\theta^*)\rangle$  and the true ground state  $|\psi_{\text{ground}}\rangle$  obtained from exact diagonalization:

$$\mathcal{F} = |\langle \psi(\theta^*) | \psi_{\text{ground}} \rangle|.$$

Note that for the models studied in this work,  $|\psi_{\text{ground}}\rangle$  is always nondegenerate. If the fidelity is  $> 99.9\%$ , we assume that we have successfully found the ground state. When assessing the quality of an optimized HVA circuit, the fidelity is a strong indicator of the success for solving the ground-state problem, since the infidelity upper bounds the difference between the ground state and variational expectation value of any observable. Letting  $1 - \mathcal{F} < \epsilon$ ,

$$|\langle O \rangle_{\text{ground}} - \langle O \rangle_{\theta}| \leq 2c\sqrt{\epsilon(1-\epsilon)} + \epsilon,$$

where  $c$  is the operator norm of  $O$  [26].

### B. Entanglement

In the context of quantum many-body physics, quantum correlations play a central role in our current understanding of the equilibrium and out-of-equilibrium properties of several systems in condensed matter. The source of these correlations is inherently nonlocal and can be traced back to the presence of entanglement in the quantum state. In this section, we introduce several commonly used entanglement quantities in quantum many-body physics.

In classical systems, one uses entropy to quantify the lack of knowledge of the state of the system due to thermal fluctuations. However, for a quantum system at zero temperature, the entropy of a subsystem has a different origin: entanglement. To quantify it, we use the bipartite *entanglement entropy* [12], which is defined as the *von Neumann entropy* of the reduced density matrix  $\rho_A$ . To obtain this reduced density matrix, we divide the system into two subsystems  $A$  and  $B$  and trace out subsystem  $B$ ,

$$\rho_A(|\psi\rangle) = \text{Tr}_B(|\psi\rangle\langle\psi|), \quad (8)$$

where  $|\psi\rangle$  is a pure state. For example, for an eight-spin model on a ring, a typical bipartition is given in Fig. 3.

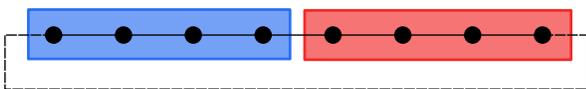


FIG. 3. The division of the full system into two subsystems  $A$  (blue) and  $B$  (red) on a 1D chain.

The von Neumann entropy generalizes the concept of Shannon entropy to quantum states and is given by

$$S(\rho_A) \triangleq -(\rho_A \log \rho_A). \quad (9)$$

Since a bipartite quantum state can always be rewritten using the Schmidt decomposition,

$$|\psi\rangle = \sum_{k=0}^K e^{-(1/2)\xi_k} |\psi_A^k\rangle \otimes |\psi_B^k\rangle, \quad (10)$$

with  $\langle \psi_A^k | \psi_A^m \rangle = \langle \psi_B^k | \psi_B^m \rangle = \delta_{km}$  and where  $K$  is the size of the smallest subsystem, the von Neumann entropy reduces to [27]

$$S(\rho_A) = \sum_{k=0}^K \xi_k \exp -\xi_k. \quad (11)$$

In recent years, the importance of entanglement in condensed-matter physics has been elucidated in several systems through the study of the scaling behavior of the entanglement entropy, which has enabled the identification and characterization of exotic phases of matter such as topological quantum states [28] and quantum spin liquids [29,30].

Full characterization of the entanglement properties of a system cannot be done by looking solely at the entanglement entropy [27,31,32]. The so-called *entanglement spectrum* has a much richer structure and has been used to study many-body localization [31], observable thermalization [33], irreversibility in quantum circuits [32], and preparation of ground states of nonintegrable quantum models [34]. In addition, the entanglement spectrum has been used to study the properties of variational methods such as the restricted Boltzmann machine [18]. The entanglement spectrum is defined as the eigenvalue spectrum of the *entanglement Hamiltonian*

$$H_{\text{ent}} \triangleq -\log \rho_A. \quad (12)$$

From Eq. (10), it follows directly that this Hamiltonian has eigenvalues  $\xi_k$ . For random quantum states distributed according to the Haar measure, the entanglement spectrum follows the *Marchenko-Pastur distribution* [35,36]. This distribution describes the asymptotic average density of eigenvalues of Wishart matrices, i.e., matrices of the form  $XX^*$ , where the  $X$  are  $m \times n$  random matrices.

Finally, the *Page entropy* [37] describes the average entanglement entropy over randomly drawn pure states in

the entire Hilbert space, and is given by

$$S_{\text{Page}} = -\frac{d_A - 1}{2d_B} + \sum_{k=d_B+1}^{d_A d_B} \frac{1}{k} \approx \log(d_A) - \frac{d_A}{2d_B}, \quad (13)$$

where  $d_A$  and  $d_B$  are the dimensions of subsystems  $A$  and  $B$ , respectively.

## IV. MAIN RESULTS

### A. The ansatz space through the lens of entanglement spectrum

The effectiveness of a VQE optimization is determined by two factors. First, one requires an expressive enough *ansatz space* that contains the ground state. The ansatz space of a specific model  $H$  and depth  $p$  refers to the set of all possible quantum states that can be reached by applying a depth- $p$  HVA circuit corresponding to  $H$  to a fixed initial state  $|\psi_0\rangle$ , which depends on the model. Second, the non-convex cost landscape induced by the variational energy of Eq. (1) must be favorable, in the sense that the optimization does not get stuck in local minima and can reliably reach the ground state.

Here, we investigate the properties of the ansatz space by examining the entanglement spectra of HVA quantum states generated with random parameters sampled uniformly in the range  $[0, \pi]$  for the TFIM and  $[0, 2\pi]$  for the XXZ model. For each model, we sample 5000 sets of parameters and calculate the entanglement spectrum of the resulting state. If the spectrum of the sampled states follows a distribution close to the Marchenko-Pastur (MP) distribution, a random HVA state has an entanglement spectrum that resembles that of a Haar random state. On the contrary, a distribution far away from the MP distribution indicates a restricted manifold of states that has a nonrandom structure. We hypothesize that the shape of the average entanglement spectrum can give an insight into the performance of the VQE optimization by revealing the structure of the ansatz space.

Figure 4 shows the average entanglement spectrum for a state in the ansatz space of circuits with depths ranging from  $1, 2, \dots, N$  for the  $N = 16$  qubit TFIM and the XXZ model. From the insets, we see that both *Ansätze* have enough entangling power to express the ground state, even for low-depth circuits. For the TFIM with 16 qubits [Fig. 4(a)], we see that for all  $p$ , the HVA spectrum is further away from the MP distribution and the HVA space corresponding to the TFIM appears to be a manifold of states with a restricted entanglement structure. In contrast, for the XXZ model, we see that the average spectra gets closer to the MP distribution as  $p$  increases. This suggests that the HVA space for the XXZ model is not as restricted as for the TFIM. This can be understood directly by looking at the circuit complexity, which for the XXZ model

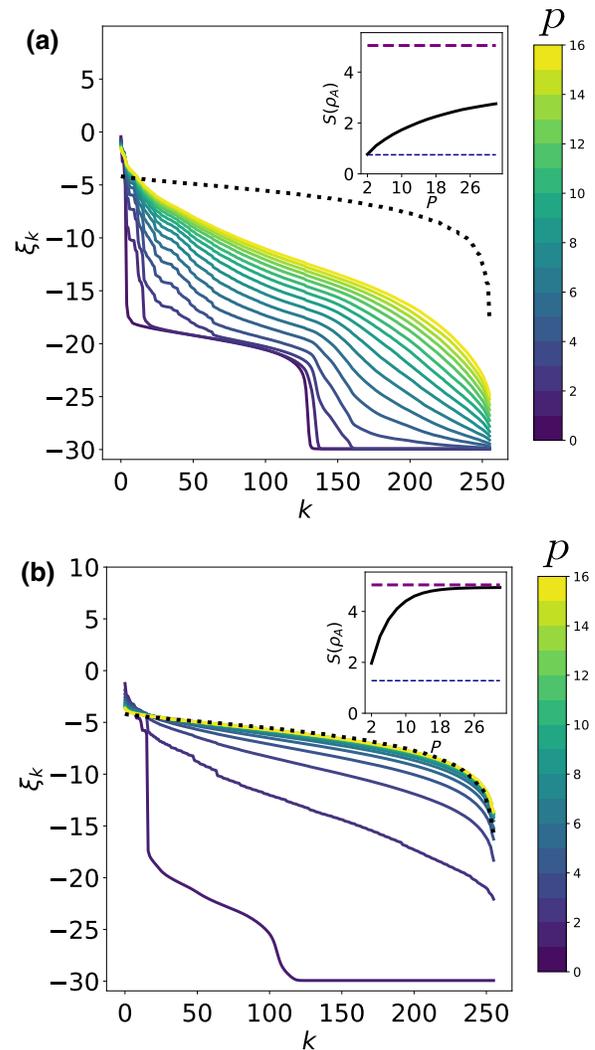


FIG. 4. The average entanglement spectrum of HVA quantum states from layer  $p = 1$  (bottom line in purple) to  $p = N$  (top line in yellow) over 5000 random parameter initializations: (a) TFIM; (b) XXZ model.  $\xi_k$  denotes the  $k$ th eigenvalue of  $H_{\text{ent}}$ . The eigenvalues are arranged in descending order and cut off at  $\xi_k = -30$ . The black lines in the insets show how close the average entanglement entropy is to the Page entropy (purple dashed line) as a function of the increasing circuit depth. The lower blue dashed line in the inset indicates the entanglement entropy of the ground state. We see that the average HVA state is more entangled than the ground states of interest.

contains more gates and parameters per layer. However, this is necessary because the XXZ model is inherently a much richer model in terms of physics and it may be necessary for the HVA space to accommodate a greater variety of states.

We now turn to examining the entanglement features of the XXZ-model HVA states explored during optimization. For the variational minimization of Eq. (1), we use a gradient-descent algorithm (for details, see Appendix A). Since the cost function is nonconvex, the quality of the

solution will vary significantly between different starting points in parameter space. We compare the following initialization strategies:

1. A completely random-state initialization, where all parameters are sampled as  $\theta \sim \mathcal{U}(0, 2\pi)$ .
2. An identity initialization. We set all parameters equal to  $\pi$ , so that our circuit is equal to the identity circuit and a global phase  $i$ .

Our approach of starting close to the identity is similar to the block-identity initialization strategy discussed in Ref. [38]; however, we study a simpler version by setting all parameters equal to  $\pi$ . For both parameter initializations, we extract the final layer state from the circuit at multiple times during the optimization and calculate its entanglement spectrum using Eq. (12). Not surprisingly, our experiments indicate that a random start is prone to getting stuck in a local minimum, due to our local optimization strategies combined with a nonconvex energy landscape. The identity start, on the other hand, allows us to consistently find a high-fidelity state for both systems with a depth  $p = N/2$  circuit (see Appendix B).

To study this finding in more detail, we investigate the dynamics of the entanglement spectrum for different initialization strategies. In Fig. 5(a), we see that an identity-state initialization stays far away from the MP distribution at all times, indicating that we are accessing a highly structured restricted subspace of the full HVA space. Additionally, this initialization reaches a state with a  $> 99.9\%$  fidelity state. On the contrary, the random-state initialization in Fig. 5(b) starts close to the MP distribution and then moves to a more structured local minimum with 70% fidelity. We conclude that even though the shape of the entanglement spectrum from Fig. 4(b) indicates a possible large unstructured ansatz space, a local optimization is still capable of finding the ground state if we choose a suitable parameter initialization. We further investigate the qualitative properties of the optimization dynamics in Appendix C. In the next section, we will see that the disadvantage of starting at a bad initial point can be overcome by making the circuit sufficiently deep, a process known as *overparametrization*.

### B. Overparametrization in the HVA

Overparametrization is a phenomenon in certain types of nonconvex optimization problems. For an overparametrized model, the optimization landscape becomes dramatically better (e.g., almost trap free or almost convex) as the number of parameters reaches some threshold. In most cases, the rate of convergence also becomes better, sometimes even exponentially faster after passing this threshold.

Overparametrization has been studied extensively in the classical deep-neural-network literature [39–42]. For

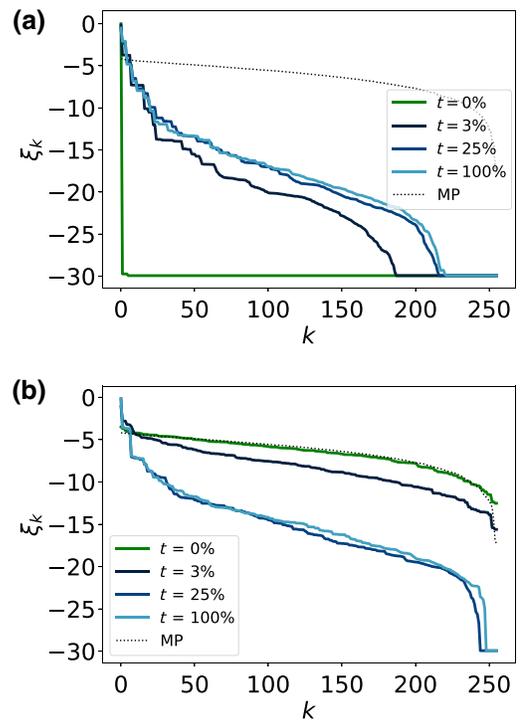


FIG. 5. The change of the entanglement spectrum of the final layer during the optimization. Both figures are for a 16-qubit XXZ model with a depth  $p = N/2$  circuit and the times are percentages of the total optimization time: (a) corresponds to a converged state of fidelity, whereas (b) corresponds to an approximately 70% fidelity state. (a) The identity-state initialization remains far away from the MP distribution at all times during the optimization and convergence to state with  $> 99.9\%$  with the ground state. Since this initialization strategy starts with the identity circuit, we find the  $t = 0\%$  state to be a product state, as indicated by the single eigenvalue. (b) The random initialization starts close to the MP distribution and converges to a local minimum with approximately 70% fidelity.

example, in Ref. [40], it has been shown that under certain mild assumptions, the optimization landscape of a deep neural network is almost convex in a large neighborhood of a random starting point. As a consequence, the stochastic gradient-descent algorithm can almost always find an accurate solution.

Although for VQE algorithms it is clear that we have  $\min E(\theta)_{p+1} \leq \min E(\theta)_p$ , it is not clear if this minimum can be found consistently due to the nonconvexity of the energy landscape. Hence, a deeper understanding of the energy landscape with increasing depth is required. There is some work on overparametrization in the context of controllable quantum systems with *unconstrained* time-varying controls [43–45], where the authors show that there are no suboptimal local minima in the optimization landscape. For the case of a *constrained* controllable quantum system, a recent work [46] considers the problem of learning  $d$ -dimensional Haar random unitaries  $U(d)$

by gradient descent using a general alternating-operator ansatz of the form  $e^{-i\gamma_p A} e^{-i\beta_p B} \dots e^{-i\gamma_1 A} e^{-i\beta_1 B}$ , where  $A$  and  $B$  are matrices sampled from the Gaussian unitary ensemble [47]. The authors show that gradient descent always converges to an accurate solution when the number of parameters is  $d^2$  or greater and that a “computational phase transition” is observed between an underparametrization ( $< d^2$ ) and an overparametrization ( $> d^2$ ) regime.

Since the HVA also has the form of an alternating-operator ansatz and the problem of finding the ground states can also be seen as a constrained quantum control problem, we expect a similar overparametrization phenomenon in our setting. To investigate this, we randomly sample 100 initial parameters  $\theta$  (uniformly drawn from the interval  $[0, \pi]$  for the TFIM and  $[0, 2\pi]$  for the XXZ model) and perform the optimization for increasing values of  $p$ . Here, we set the stopping criterion for the optimization to  $\epsilon_{\text{res}} = E(\theta)_p - E_{\text{ground}} < 1e-4$  and the maximum number of iterations to 3000. Indeed, Fig. 6 shows that the overparametrization phenomenon also occurs in the HVA

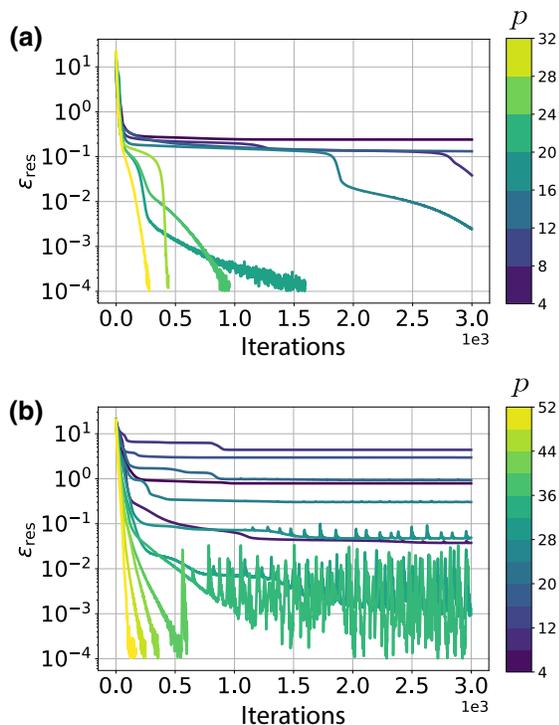


FIG. 6. Overparametrization in the HVA: (a) TFIM; (b) XXZ model. Each line corresponds to the VQE optimization at depth  $p$  that takes the most iterations to converge out of 100 random initializations. Both figures are for  $N = 12$  qubits. The rapid oscillations in (b) are artifacts of the ADAM optimizer and are less severe as the circuit depth increases. Due to our stopping criterion, we know that if the number of iterations is smaller than 3000, then  $\epsilon_{\text{res}} \leq 1e-4$  and so the model does converge to a good ground-state approximation.

for the 12-qubit TFIM and the XXZ model. We find that for both the TFIM and the XXZ model, gradient descent from all 100 random starting points converges to an accurate solution once the depth  $p$  reaches a certain threshold  $\tilde{p}(N)$ .

Moreover, we also observe a “computational phase transition” around this threshold where the convergence speed becomes exponentially fast, i.e., the decrease of the residue energy as a function of the number of iterations. However, this threshold  $\tilde{p}(N)$  is not tight, i.e., for depth  $p < \tilde{p}(N)$ , it is possible that the gradient descent still converges to a high-fidelity state. This indicates that in the setting of finding ground states using the HVA, the problem is more structured and gradient descent is effective. In Fig. 7, we see that for all system sizes, the mean number of iterations eventually converges to about 100 iterations. In addition, we can find the overparametrization thresholds  $\tilde{p}(N)$  in Table I for the TFIM and the XXZ model with different system sizes. Our data suggest that  $\tilde{p}(N)$  has at most a polynomial scaling, which is compatible with the

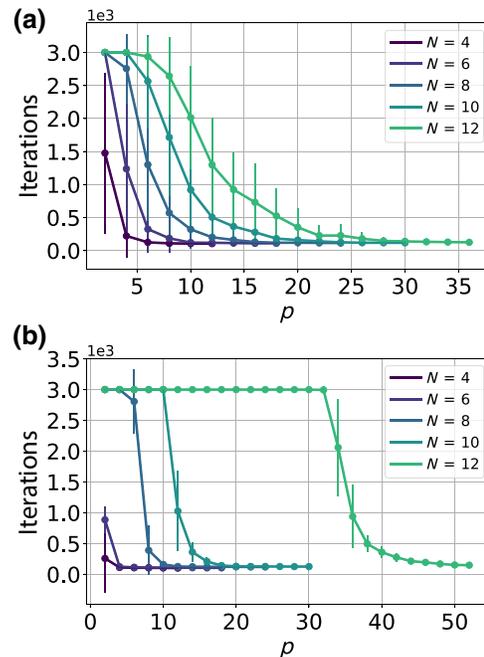


FIG. 7. The mean iteration time to convergence as a function of depth: (a) TFIM; (b) XXZ model. The error bars indicate the standard deviation over 100 different initializations. For both models, there is a clear cutoff where the number of iterations saturates. Note that if the number of iterations is smaller than 3000, then we know that  $\epsilon_{\text{res}} \leq 1e-4$ , indicating that the optimization has converged to a good ground-state approximation. We see that the error bars decrease systematically with depth. For both models, there is a critical  $p$  after which all random initializations converge to a good ground-state approximation. Moreover, for depth  $p = 34$  and  $p = 52$  for the TFIM and the XXZ model, respectively, the number of iterations to find the ground state is of the order of 100 iterations for every starting point.

TABLE I. The overparametrization threshold  $\tilde{p}(N)$  for the TFIM and the XXZ model with different system sizes  $N$ . By “threshold,” we mean that when  $p \geq \tilde{p}(N)$ , all the random initializations converge to an accurate solution.

$N$	TFIM $\tilde{p}(N)$	XXZ model $\tilde{p}(N)$
4	6	4
6	6	4
8	8	8
10	10	12
12	14	36

analogous parameter count required to express critical 1D ground states with an MPS. A more detailed view of these data can be found in Fig. 11 in the Appendix, which shows that all random initializations converge to the ground state after a certain threshold  $\tilde{p}(N)$ .

This is a striking difference compared with Ref. [46], where the number of parameters to achieve overparametrization is  $(2^N)^2$ . From Fig. 7, we can also see that the iteration time decreases substantially as  $p$  increases, saturating to around 100 iterations after a certain  $p$  for all  $N$ .

The overparametrization phenomenon in the HVA shows a clear difference between the HVA and parametrized random quantum circuits (RQC), because there is no indication or evidence that the landscape of RQC gets better as one increases the depth. On the contrary, in our experiments with random circuits of comparable depths to our HVA circuits, we are unable to observe the same overparametrization phenomenon. This can be explained from the barren-plateau point of view and the lack of structure in the ansatz space.

### C. Ameliorated barren plateaus in the HVA

In Ref. [5], a *barren-plateau phenomenon* has been observed for the VQE on random quantum circuits, where all gradients are exponentially close to zero with overwhelmingly high probability, making local optimization within the ansatz space extremely challenging. The barren-plateau phenomenon is due to the fact that RQCs consisting of single- and two-qubit gates form a 2-design, which means that the gradients of the energy objective function will obey the same concentration of measure properties as if the circuits are Haar random unitaries.

In contrast to the RQC ansatz, we show that the optimization landscape of the HVA is much more favorable. This is clearly illustrated when optimizing the HVA corresponding to the TFIM: to begin with, as discussed in Sec. IV A, the manifold of states has a much more restricted entanglement structure than a typical Haar random state—this already indicates that the HVA circuits do not form 2-designs and thus do not obey the same kind of

concentration-of-measure phenomenon as RQCs. On the other hand, the entanglement spectrum of the ansatz space corresponding to the XXZ model does not immediately rule out the same barren-plateau behavior as exhibited by RQCs.

Nonetheless, we determine that the barren-plateau problem is significantly ameliorated in the TFIM and mild in the XXZ model. In Fig. 8, we calculate the variance of gradients as a function of the number of qubits  $N$  and the depth  $p$  over 20 random points per  $N$  and per  $p$ . For the TFIM, the flatness of the variance curve indicates no barren-plateau problem. However, for the XXZ model, we see an exponential decay, but this decay is not as strong as in RQCs [5]. The scaling of the mean-gradient magnitudes follows a similar pattern. Nonetheless, we can reliably find an accurate solution when choosing an identity start (see Appendix B), where the barren-plateau problem is absent. Indeed, sampling gradients close to the identity initialization gives a constant gradient variance for all  $N$ . This indicates that the vanishing-gradient problem can be circumvented by choosing a suitable initialization strategy.

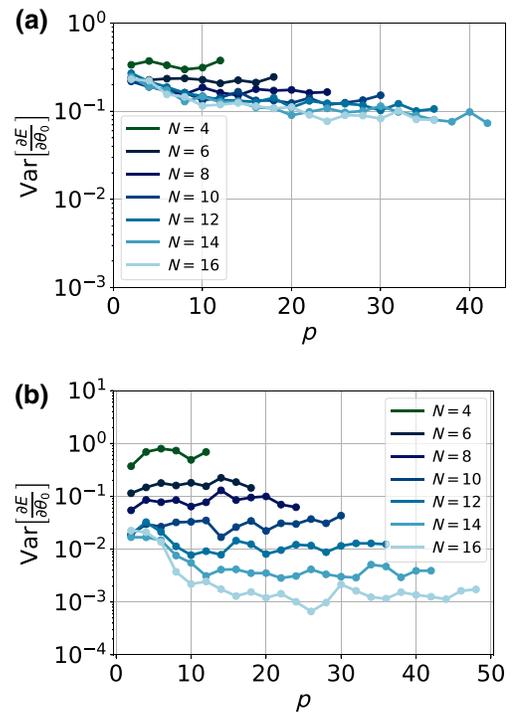


FIG. 8. The variance of the gradients of a single  $Z_1Z_2$  term with respect to  $\theta_0$  as a function of the number of qubits at initialization. The number of samples used per  $N$  for each  $p$  is 20. (a) For the TFIM, the gradient-variance decay is almost constant for all  $N$ . (b) The XXZ-model gradient variance is still exponential, although the effect is not as pronounced as for the RQCs of Ref. [5], where the  $N = 16$  variance is 2 orders of magnitude smaller.

### D. The entangling power of HVA circuits

For a 1D gapped quantum system, the entanglement entropy of the ground state obeys an area law [48–50], i.e., the entanglement entropy grows proportionally to the boundary area  $\partial I$  of the subsystem  $\rho_A$ :

$$S(\rho_A) = \mathcal{O}(|\partial I|).$$

In 1D, the boundary area  $\partial I$  is either 1 (for an open chain) or 2 (for a closed chain) and the area law simply says that the entanglement entropy should be constant as  $N$  increases. For a 1D conformally invariant gapless (critical) system, the entanglement entropy of the ground state has a logarithmic scaling instead [51], i.e.,

$$S(\rho_A) = \mathcal{O}(\log(n)).$$

The entangling power is an important factor for characterizing the expressiveness and efficiency of many variational *Ansätze* in condensed-matter physics. It characterizes how much entanglement (measured by the entanglement entropy) can be generated by the variational circuit. For example, in the matrix-product state representation, the entangling power is limited by the so-called *bond dimension*  $D$ , which affects the expressive power and computational cost of the ansatz. For a 1D gapped system with energy gap  $\epsilon$ , the ground state can be approximated well by an MPS with sublinear bond dimension  $D = \exp[\tilde{O}(\log^{3/4} n / \epsilon^{1/4})]$  [52]. In the case of the HVA, the amount of entanglement generated by the circuit depends on the depth  $p$  of the circuit. Indeed, we observe numerically in Fig. 4 that the HVA circuits for the TFIM and the

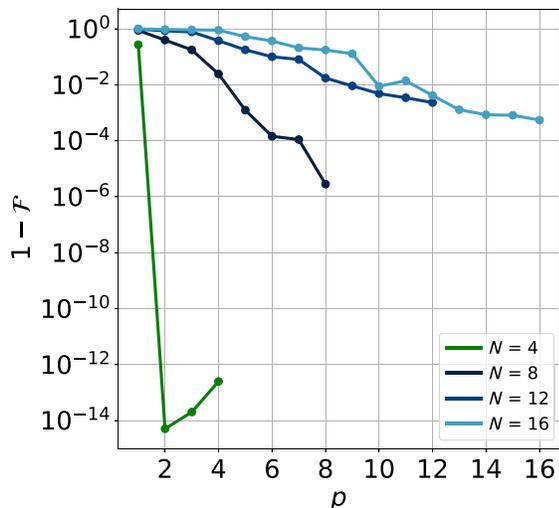


FIG. 9. The infidelities found after optimization for the MHS Hamiltonian. The circuit is initialized with an identity start. For the four-qubit case, we get close to machine precision and hence the fidelities are unstable.

XXZ model have enough entangling power to express the ground states.

As a demonstration that the full entangling power of the HVA can be utilized effectively, we solve for the ground state of the so-called modified Haldane-Shastry (MHS) Hamiltonian. This model has long-range interactions and is expected to have power-law entanglement scaling in the ground state [53,54]. The MHS Hamiltonian is given by

$$H_{\text{MHS}} = \sum_{j < k}^N \frac{1}{d_{jk}^2} (-\sigma_x^j \sigma_x^k - \sigma_y^j \sigma_y^k + \sigma_z^j \sigma_z^k),$$

where  $d_{jk} = \frac{N}{\pi} |\sin(\pi(j-k)/N)|$ . Due to the form of the Hamiltonian, we can use the same HVA, given in Eq. (7), as for the XXZ model. In Fig. 9, we see that it is possible to find the ground state with  $> 99.7\%$  fidelity using a depth  $p = N$  circuit for  $N = 4$ ,  $N = 8$ ,  $N = 12$ , and  $N = 16$ .

### V. CONCLUSION

In this work, we shed light on some of the desirable properties of the HVA as a critical ingredient in the variational quantum eigensolver algorithm. In particular, we show evidence that there are only mild or entirely absent barren plateaus in the HVA. This is strikingly different from the commonly used random quantum circuits. Moreover, we also observe an overparametrization phenomenon in the HVA. Similar to what has been observed in deep neural networks, the optimization landscape of the HVA becomes increasingly better as the ansatz is overparametrized and eventually becomes trap free as the overparametrization reaches a certain threshold. In contrast to the case of learning Haar random unitaries, we observe that such threshold in the HVA scales at most polynomially with the system size. Finally, we provide numerical evidence that the HVA can be used to find the ground state of the MHS Hamiltonian, which has a power-law scaling entanglement. We believe that our findings point to important features of the HVA that will lead to a deeper understanding of its effectiveness and that point the way to the development of more sophisticated *Ansätze* for other many-body problems, as well as better-informed optimization and/or initialization strategies.

As for future work, since most 1D quantum many-body systems can be simulated efficiently using classical methods, the crucible for the HVA will be 2D systems. If low-depth circuits are capable of reproducing nontrivial 2D quantum states, then one can start thinking about when a quantum advantage can be reached for systems where classical methods are computationally expensive or even ineffective. The effectiveness of the identity initialization, both in terms of the absence of vanishing gradients and the reliability of finding a good ground-state approximation, is striking. Scrutinizing the mechanism for why

this is the case will require a deeper understanding of the energy landscape of the HVA. Our preliminary results for the XXZ model and the TFIM on rectangular lattices show that this initialization strategy remains effective even for 2D systems.

Lastly, the overparametrized regime is a double-edged sword. On the one hand, it implies that we can improve the energy landscape by increasing the depth of the circuit, ameliorating the effects of local minima. On the other hand, the growth in circuit depth may well nullify this increase in performance due to the longer coherence times required and multiplicative gate errors. In order to assess how useful this regime is for hardware implementations, we would require an understanding of the effect that noise has on the optimization in the overparametrized regime. The recent work of Wang *et al.* [55] indicates that for a class of VQE *Ansätze*, including the quantum alternating-operator ansatz, there could be severe noise-induced barren plateaus when the number of layers scales polynomially. However, for the practical performance of the general HVA, a more careful analysis of the trade-off between the benefits of overparametrization and the detrimental effects of noise-induced barren plateaus is needed. Moreover, research on the design of more effective variational quantum circuits based on the HVA should also be pursued.

### ACKNOWLEDGMENTS

H.Y. is supported by the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant 2019-06636 and a Google Quantum Research Award. Y.B.K. is supported by the NSERC of Canada and the Killam Research Fellowship from the Canada Council of the Arts. J.C. acknowledges support from NSERC, the Shared Hierarchical Academic Research Computing Network

(SHARCNET), Compute Canada, a Google Quantum Research Award, and the CIFAR AI chair program. The resources used in preparing this research were provided, in part, by the Province of Ontario, Government of Canada through the Canadian Institute for Advanced Research (CIFAR) and by companies sponsoring the Vector Institute [www.vectorinstitute.ai/#partners](http://www.vectorinstitute.ai/#partners). C.Z. acknowledges partial support from the Postgraduate Affiliate Award from the Vector Institute.

### APPENDIX A: COMPUTATIONAL DETAILS

For the implementation of our quantum circuits, we use ZYGLOX [56], a powerful TensorFlow-based quantum simulator. For the classical optimization process, we use ADAM (adaptive moment estimation) [57], a gradient-descent-based optimizer, which is widely used in the machine-learning community. Compared to vanilla gradient descent and its other variants, ADAM updates the learning rates adaptively on a per-parameter basis by using estimates of the first and second moments of the gradients. In our own investigation for solving the ground-energy problem with the HVA, ADAM outperforms all the other optimizers available in TensorFlow, with respect to fidelity and convergence times.

Unless stated otherwise, the stopping criterion for our optimization is defined as  $|E(\theta_t) - E(\theta_{t+1})| < 1 \times 10^{-13}$ , where  $t$  is the iteration number. The maximum number of iterations is set to 15 000. We use an initial learning rate of  $r = 0.01$  for ADAM, which gives reasonably consistent results across all the models. Through our own investigation into initial ADAM learning rates, we find a learning rate of  $1 \times 10^{-3} \leq r \leq 4 \times 10^{-2}$  to be a good choice for the optimization for both the TFIM and the XXZ model,

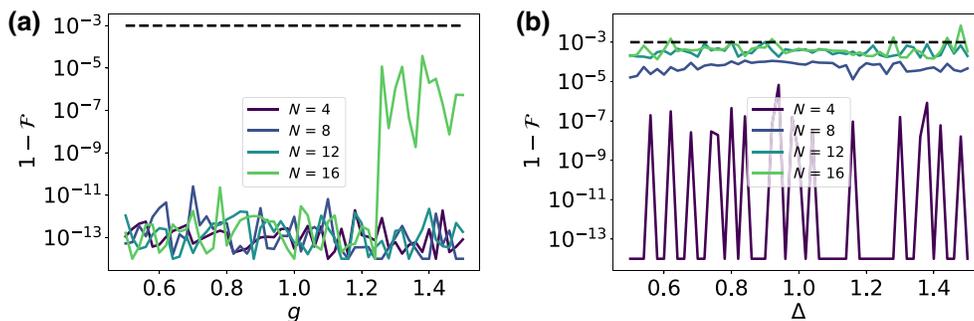


FIG. 10. The infidelities as a function of the order values  $g$  and  $\Delta$  for (a) the TFIM and (b) the XXZ model, respectively. For these results, we use the identity initialization combined with a depth  $p = N/2$  circuit. The dashed black line indicates the cutoff for 99.9% fidelity. (a) For the TFIM, we can obtain machine-precision results, except in the region  $g > 1.24$  for  $N = 16$ . In this region, the optimization has not fully converged but is stopped after 15 000 iterations. We note that for increasing  $N$ , the time until convergence is polynomial in  $N$  (not shown here), similar to what has been observed in Ref. [23]. Additionally, we observe a worsening of this scaling with increased  $g$ . (b) For the XXZ model, we are unable to consistently reach machine-precision fidelities. In addition, the fidelities that we find become worse as  $N$  increases. However, except for a couple of outliers, we are able to obtain  $> 99.9\%$  fidelities for  $N \leq 16$  or all order values.

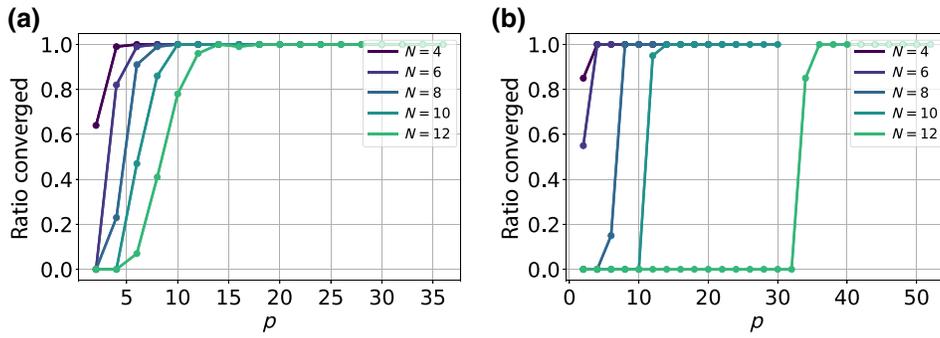


FIG. 11. The ratio of random initializations that converge to the ground state: (a) TFIM; (b) XXZ model. We consider a run converged if  $\epsilon_{\text{res}} \leq 1e - 4$ .

as it balances the optimization accuracy and convergence speed.

These simulations are performed on the University of Toronto Computer Science Department servers, which house either AMD Ryzen Threadripper 2990WX or Silicon Mechanics Rackform iServ R331.v4 with two 12-core Intel E5-2697v2 CPUs, with access to at most 10 GB of RAM. More computationally expensive simulations are done using Nvidia GeForce GTX 1050Ti graphics processing units (GPUs).

On the Vector Cluster, we use Intel(R) Xeon(R) Silver 4110 CPUs and Nvidia 480 T4 GPUs, with access to at most 32 GB of RAM.

## APPENDIX B: ADDITIONAL NUMERICAL RESULTS

In Fig. 10, we show the infidelities that we find after optimization for  $g \in \{0.5, 0.52, \dots, 1.5\}$  and  $\Delta \in \{0.5, 0.52, \dots, 1.5\}$  for the TFIM and the XXZ model, respectively. Our numerical results for the aforementioned ranges of order values are available in the data-set module of TensorFlow Quantum [58].

In Fig. 11, we show the ratio of the converged random initializations as a function of the circuit depth  $p$  for different system sizes  $N$ .

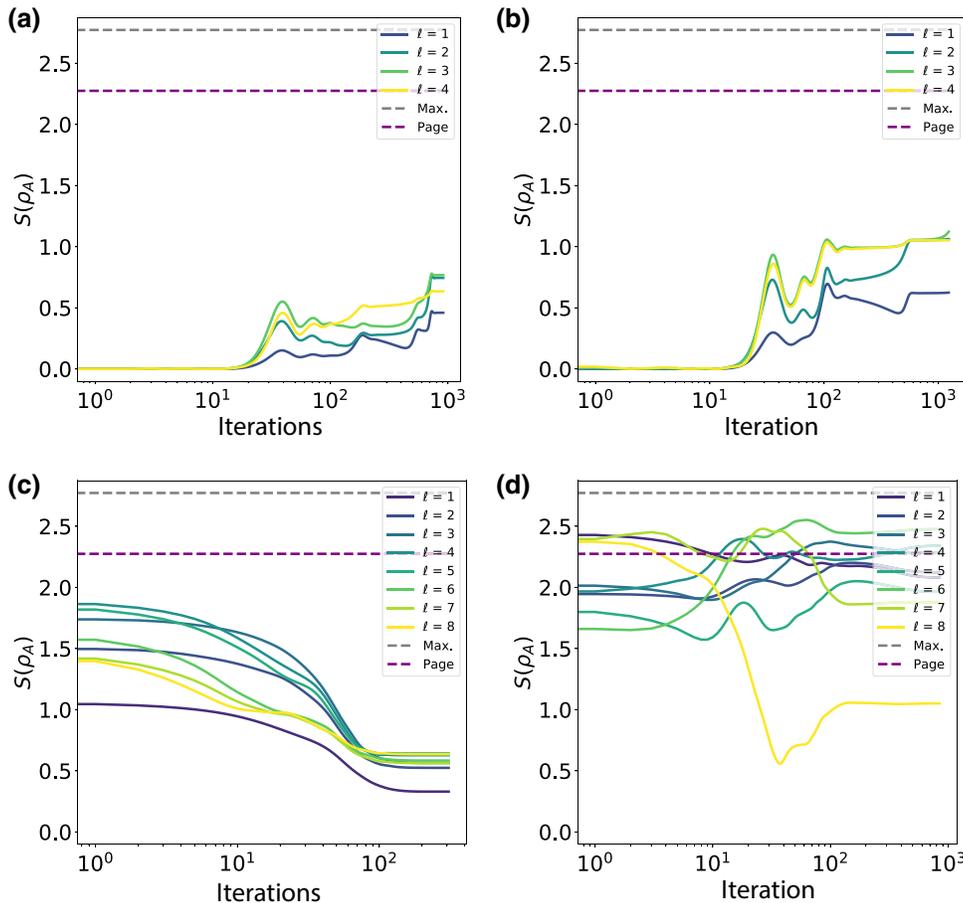


FIG. 12. The dynamics of the entanglement entropy at each layer during optimization: (a),(c) TFIM,  $p = 4$ ; (b),(d) XXZ model,  $p = 8$ . Each separate line indicates the entanglement entropy of the state in layer  $l$ . The gray dashed line denotes the maximum possible entanglement and the purple line gives the Page entropy. For all figures, the final state is a  $> 99.9\%$  fidelity state. (a) The identity initialization for an eight-qubit TFIM with  $g = 1.0$  and  $p = 4$ . (b) The same TFIM with a random-state initialization and overparametrization  $p = 8$ . (c) Random-state initialization for an eight-qubit XXZ model with  $\Delta = 1.0$  and  $p = 4$ . (d) Typical XXZ-model dynamics for a random-state initialization and overparametrization  $p = 8$ .

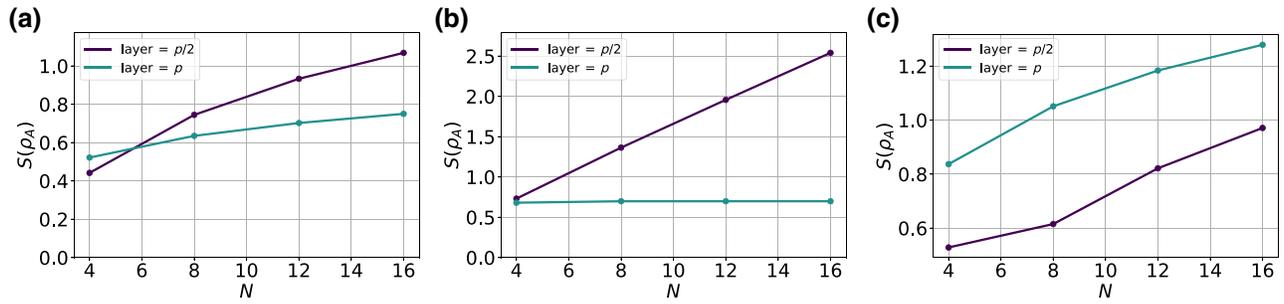


FIG. 13. The scaling of the entanglement entropy of the converged state after  $p/2$  and  $p$  layers: (a) TFIM,  $g = 1.0$ ; (b) TFIM,  $g = 0.5$ ; (c) XXZ model,  $\Delta = 1.0$ . (a) For the TFIM at the critical point, the ground-state entanglement entropy has a logarithmic correction with increasing  $N$ . The entanglement halfway through the circuit is larger than in the final layer. (b) For a noncritical point, the ground-state entanglement entropy is constant but the entanglement entropy halfway through the circuit scales linearly with system size. (c) For the XXZ model, in addition to the logarithmic scaling of the entanglement entropy, the final layer entanglement is consistently higher than in the  $p/2$  depth layer.

### APPENDIX C: DYNAMICS OF ENTANGLEMENT ENTROPY DURING OPTIMIZATION

To further elucidate the difference in initialization strategies, we qualitatively study the dynamics of the entanglement entropy during optimization. In Fig. 12, we calculate the entanglement entropy of  $\rho_A$  at each layer of the circuit during the optimization. Although not much can be said about the intermediate states for the random-state initialization, except that they are highly entangled, the entanglement entropy dynamics for the identity initialization have a distinct structure that is consistent as we increase the system size. In Fig. 13, we compare the scaling of the entanglement entropy for the identity start halfway through the circuit for different system sizes.

- [1] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [2] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature* **549**, 242 (2017).
- [3] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, *Adv. Quantum Technol.* **2**, 1900070 (2019).
- [4] C. Bravo-Prieto, J. Lumbreras-Zarapico, L. Tagliacozzo, and J. I. Latorre, Scaling of variational quantum circuit depth for condensed matter systems, *Quantum* **4**, 272 (2020).
- [5] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [6] K. Sharma, M. Cerezo, L. Cincio, and P. J. Coles, Trainability of dissipative perceptron-based quantum neural networks, [arXiv:2005.12458](https://arxiv.org/abs/2005.12458) (2020).

- [7] T. Volkoff and P. J. Coles, Large gradients via correlation in random parameterized quantum circuits, [arXiv:2005.12200](https://arxiv.org/abs/2005.12200) (2020).
- [8] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost-function-dependent barren plateaus in shallow quantum neural networks, [arXiv:2001.00550](https://arxiv.org/abs/2001.00550) (2020).
- [9] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, *Phys. Rev. A* **92**, 042303 (2015).
- [10] W. W. Ho and T. H. Hsieh, Efficient variational simulation of non-trivial quantum states, *SciPost Phys.* **6**, 29 (2019).
- [11] C. Cade, L. Mineh, A. Montanaro, and S. Stanisic, Strategies for solving the Fermi-Hubbard model on near-term quantum computers, [arXiv:1912.06007](https://arxiv.org/abs/1912.06007) (2019).
- [12] J. Eisert, M. Cramer, and M. B. Plenio, Colloquium: Area laws for the entanglement entropy, *Rev. Mod. Phys.* **82**, 277 (2010).
- [13] J. Eisert, in *Autumn School on Correlated Electrons: Emergent Phenomena in Correlated Matter*, edited by E. Pavarini, E. Koch, and U. Schollwöck (Forschungszentrum Jülich Zentralbibliothek, Verlag, Germany, 2013).
- [14] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [15] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, Deep Autoregressive Models for the Efficient Variational Simulation of Many-Body Quantum Systems, *Phys. Rev. Lett.* **124**, 020503 (2020).
- [16] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Reconstructing quantum states with generative models, *Nat. Mach. Intell.* **1**, 155 (2019).
- [17] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wavefunctions, *Phys. Rev. Res.* **2**, 023358 (2020).
- [18] D.-L. Deng, X. Li, and S. D. Sarma, Quantum Entanglement in Neural Network States, *Phys. Rev. X* **7**, 021021 (2017).
- [19] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien,

- A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 4213 (2014).
- [20] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, [arXiv:1411.4028](https://arxiv.org/abs/1411.4028) (2014).
- [21] E. Farhi, J. Goldstone, S. Gutmann, and Michael Sipser, Quantum computation by adiabatic evolution, [arXiv:quant-ph/0001106](https://arxiv.org/abs/quant-ph/0001106) (2000).
- [22] G. Bigan Mbeng, R. Fazio, and G. Santoro, Quantum annealing: A journey through digitalization, control, and hybrid quantum variational schemes, [arXiv:1906.08948](https://arxiv.org/abs/1906.08948) (2019).
- [23] D. Wierichs, C. Gogolin, and M. Kastoryano, Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer, *Phys. Rev. Res.* **2**, 043246 (2020).
- [24] F. Franchini, *An Introduction to Integrable Techniques for One-Dimensional Quantum Systems* (Springer International, Cham, Switzerland, 2017).
- [25] R. Orús, Tensor networks for complex quantum systems, *Nat. Rev. Phys.* **1**, 538 (2019).
- [26] M. J. S. Beach, R. G. Melko, T. Grover, and T. H. Hsieh, Making trotters sprint: A variational imaginary time ansatz for quantum many-body systems, *Phys. Rev. B* **100**, 094434 (2019).
- [27] H. Li and F. D. M. Haldane, Entanglement Spectrum as a Generalization of Entanglement Entropy: Identification of Topological Order in Non-Abelian Fractional Quantum Hall Effect States, *Phys. Rev. Lett.* **101**, 010504 (2008).
- [28] A. Kitaev and J. Preskill, Topological Entanglement Entropy, *Phys. Rev. Lett.* **96**, 110404 (2006).
- [29] S. V. Isakov, M. B. Hastings, and R. G. Melko, Topological entanglement entropy of a Bose-Hubbard spin liquid, *Nat. Phys.* **7**, 772 (2011).
- [30] Y. Zhang, T. Grover, and A. Vishwanath, Entanglement Entropy of Critical Spin Liquids, *Phys. Rev. Lett.* **107**, 067202 (2011).
- [31] Z.-C. Yang, C. Chamon, A. Hama, and E. R. Mucciolo, Two-Component Structure in the Entanglement Spectrum of Highly Excited States, *Phys. Rev. Lett.* **115**, 267206 (2015).
- [32] D. Shaffer, C. Chamon, A. Hama, and E. R. Mucciolo, Irreversibility and entanglement spectrum statistics in quantum circuits, *J. Stat. Mech.: Theory Exp.* **2014**, P12007 (2014).
- [33] S. D. Geraedts, R. Nandkishore, and N. Regnault, Many-body localization and thermalization: Insights from the entanglement spectrum, *Phys. Rev. B* **93**, 174202 (2016).
- [34] G. Matos, S. Johri, and Z. Papić, Quantifying the efficiency of state preparation via quantum variational eigensolvers (2020).
- [35] M. Žnidarič, Entanglement of random vectors, *J. Phys. A: Math. Theor.* **40**, F105 (2006).
- [36] V. A. Marčenko and L. A. Pastur, Distribution of eigenvalues for some sets of random matrices, *Math. USSR-Sbornik* **1**, 457 (1967).
- [37] D. N. Page, Average Entropy of a Subsystem, *Phys. Rev. Lett.* **71**, 1291 (1993).
- [38] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, *Quantum* **3**, 214 (2019).
- [39] Z. Allen-Zhu, Y. Li, and Y. Liang, Learning and generalization in overparameterized neural networks, going beyond two layers (2018).
- [40] Z. Allen-Zhu, Y. Li, and Z. Song, A convergence theory for deep learning via over-parameterization (2018).
- [41] Z. Chen, Y. Cao, D. Zou, and Q. Gu, How much overparameterization is sufficient to learn deep ReLU networks? (2019).
- [42] Z. Chen, Y. Cao, Q. Gu, and T. Zhang, Mean-field analysis of two-layer neural networks: Non-asymptotic rates and generalization bounds (2020).
- [43] H. A. Rabitz, M. M. Hsieh, and C. M. Rosenthal, Quantum optimally controlled transition landscapes, *Science* **303**, 1998 (2004).
- [44] H. Rabitz, M. Hsieh, and C. Rosenthal, Landscape for optimal control of quantum-mechanical unitary transformations, *Phys. Rev. A* **72**, 052337 (2005).
- [45] B. Russell, H. Rabitz, and R.-B. Wu, Control landscapes are almost always trap free: A geometric assessment, *J. Phys. A: Math. Theor.* **50**, 205302 (2017).
- [46] B. T. Kiani, S. Lloyd, and R. Maity, Learning unitaries by gradient descent (2020).
- [47] M. L. Mehta, *Random Matrices*, Pure and Applied Mathematics (Academic Press, 2004), 3rd ed.
- [48] M. B. Hastings, An area law for one-dimensional quantum systems, *J. Stat. Mech.: Theory Exp.* **2007**, P08024 (2007).
- [49] I. Arad, Z. Landau, and U. Vazirani, Improved one-dimensional area law for frustration-free systems, *Phys. Rev. B* **85**, 195145 (2012).
- [50] I. Arad, A. Kitaev, Z. Landau, and U. Vazirani, An area law and sub-exponential algorithm for 1D systems (2013).
- [51] P. Calabrese and J. Cardy, Entanglement entropy and quantum field theory, *J. Stat. Mech.: Theory Exp.* **2004**, P06002 (2004).
- [52] S. Gharibian, Y. Huang, Z. Landau, and S. W. Shin, Quantum Hamiltonian complexity, *Found. Trends<sup>o</sup> Theor. Comp. Sci.* **10**, 159 (2015).
- [53] F. D. M. Haldane, Exact Jastrow-Gutzwiller Resonating-Valence-Bond Ground State of the Spin- $\frac{1}{2}$  Antiferromagnetic Heisenberg Chain with  $1/r^2$  Exchange, *Phys. Rev. Lett.* **60**, 635 (1988).
- [54] B. S. Shastri, Exact Solution of an  $s = 1/2$  Heisenberg Antiferromagnetic Chain with Long-Ranged Interactions, *Phys. Rev. Lett.* **60**, 639 (1988).
- [55] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms (2020).
- [56] R. C. Wiersema, Zyglox. <https://github.com/therooler/zyglox> (2020).
- [57] D. Kingma and J. Ba, in *International Conference on Learning Representations*, edited by Y. Bengio and Y. LeCun (ICLR, San Diego, CA, USA, 2015).
- [58] M. Broughton, G. Verdon, T. McCourt, A. J. Martinez, J. H. Yoo, S. V. Isakov, P. Massey, M. Y. Niu, R. Halavati, E. Peters, M. Leib, A. Skolik, M. Streif, D. Von Dollen, J. R. McClean, S. Boixo, D. Bacon, A. K. Ho, H. Neven, and M. Mohseni, TensorFlow Quantum: A software framework for quantum machine learning (2020).