Phenomenological theory of variational quantum ground-state preparation

Nikita Astrakhantsev,[1,*] Guglielmo Mazzola,[2,3] Ivano Tavernelli,[2] and Giuseppe Carleo[4]
[1]*Department of Physics, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[2]*IBM Quantum, IBM Research—Zurich, CH-8803 Rüschlikon, Switzerland*
[3]*Institute for Computational Science, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*
[4]*Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

(Received 1 December 2022; accepted 28 August 2023; published 28 September 2023)

The variational approach is a cornerstone of computational physics, considering both conventional and quantum computing computational platforms. The variational quantum eigensolver algorithm aims to prepare the ground state of a Hamiltonian exploiting parametrized quantum circuits that may offer an advantage compared to classical trial states used, for instance, in quantum Monte Carlo or tensor network calculations. While, traditionally, the main focus has been on developing better trial circuits, we show that the algorithm's success, if optimized within stochastic gradient descent (SGD) or quantum natural gradient descent (QNGD), crucially depends on other parameters such as the learning rate, the number N_s of measurements to estimate the gradient components, and the Hamiltonian gap Δ. Within the standard SGD or QNGD, we first observe the existence of a finite N_s value below which the optimization is impossible, and the energy variance resembles the behavior of the specific heat in second-order phase transitions. Second, when N_s is above such threshold level, and learning is possible, we develop a phenomenological model that relates the fidelity of the state preparation with the optimization hyperparameters and Δ. More specifically, we observe that the computational resources scale as $1/\Delta^2$, and we propose a symmetry enhancement of the variational ansatz as a way to increase the closing gap. We test our understanding on several instances of two-dimensional frustrated quantum magnets, which are believed to be the most promising candidates for near-term quantum advantage through variational quantum simulations.

DOI: 10.1103/PhysRevResearch.5.033225

I. INTRODUCTION

The task of preparing a quantum state on a qubit register is of fundamental importance in quantum computing. To this end, the variational quantum eigensolver (VQE) [1–3] algorithm, of widespread use in several quantum application domains, from chemistry and physics [4–6], machine learning [7], combinatorial optimization [8], and finance [9] aims at variationally preparing the ground state of a Hamiltonian $\hat{H}$ [3,4,9–16]. While the VQE algorithm has already been implemented on existing noisy devices, its importance will persist in the future fault-tolerant regime as it is also necessary for more advanced quantum algorithms such as eigenstate projection methods [17,18].

The two essential ingredients of the method are (i) a parametrized quantum circuit with variational parameters $\boldsymbol{\theta}$, which produces a wave function $|\psi(\boldsymbol{\theta})\rangle$ expressing the ground state of the problem, and (ii) a *learning* procedure aimed at optimizing the circuit variational parameters $\boldsymbol{\theta}$. This last step involves a feedback loop between the quantum and classical resources.

The VQE is among the most widely used quantum algorithms, and it has been adapted to many particular contexts and models. Significant effort has been spent so far in devising better variational forms, tailored to various application domains [3,6,12,19–23] or hardware architectures [4,6]. In addition, the optimization part of the algorithm has been studied [3], including works discussing the distribution of local minima in the energy landscape [24,25], its behavior in the overparametrized regime [26], and the concept of the barren plateaus [27], manifested in gradients vanishing exponentially fast with system size [28,29]. Importantly, this phenomenon was reported only in the context of random circuit variational forms. Recent works have proven an absence of such plateaus in the quantum convolutional neural network architecture [30] and circuits aimed at the study of quantum magnets [31,32].

However, significantly less attention has been paid to a realistic scenario of optimization with sizable shot noise, and to the relation between the optimization efficiency and the problem's physical regime, including the dependency on the system gap Δ [28].

This paper aims to provide a phenomenological theory underlying the efficiency of the training: (i) within several optimization algorithms, (ii) in various regimes of model parameters, and (iii) as a function of optimization hyperparameters such as the learning rate η or the number of gradient-estimating samples N_s per optimization step.

*nikita.astrakhantsev@physik.uzh.ch

We adopt the best-case scenario of a noise-free hardware setting, i.e., we consider perfect gates and readout, and study the stochastic gradient descent (SGD) and the stochastic reconfiguration (SR) approach [33], as implemented on quantum computers in the quantum natural gradient descent (QNGD) approach [34]. In such a hardware-free noise setting, the remaining obstacle for the state preparation is the inherent statistical quantum measurement noise in the gradient estimation $[\nabla\langle\hat{H}\rangle]$ induced by a finite budget N_s of shots, i.e., circuit repetitions [3,12,35].

In this paper, we only consider gradient-based optimization. These methods are likely the only scalable approach to VQE as, at least in the classical setting, they represent the only practical strategy that provides stable optimization of a large number of parameters [36–38]. Applications of other gradient-free methods such as SPSA are restricted to the problems of quantum chemistry with only a small number of variational parameters [3].

In this scenario, we analyze how the performance of variational ground-state preparation is affected by the sample size N_s per gradient component and the learning rate η of a gradient descent step. First, we observe that the resulting fidelity shows a critical behavior as a function of a suitably defined stochastic energy fluctuations measure $\epsilon \propto 1/N_s$, namely, we show that the fidelity vanishes when this measure is larger than a certain threshold value, $\epsilon > \epsilon^c$, and shows a rapid growth instead at $\epsilon < \epsilon^c$. Further in the text, we refer to this sharp change in the algorithm performance as an *algorithmic phase transition* and, somewhat loosely, call this behavior *critical*. We emphasize that these algorithm performance regimes, as well as the transition between them and threshold (critical) ϵ_c, stem primarily from the optimization hyperparameters such as η, N_s, and are not related to conventional phase transitions in a medium.

To understand the mechanism behind this transition, in this paper, we attempt to qualitatively reproduce some its features for a given circuit with a simple distribution over the parameter space given by the Boltzmann distribution $\Pi(\boldsymbol{\theta}) \propto \exp(-E(\boldsymbol{\theta})/T)$, with T being the effective temperature of the system and $E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|\hat{H}|\psi(\boldsymbol{\theta})\rangle$ being the energy expectation. Second, we address estimating the sample size required to reach a certain overlap with the ground state. For sufficiently large samples, we observe that the circuit infidelity behaves as $A\epsilon + \mathcal{I}_0$, with $\mathcal{I}_0$ representing the circuit's inability to exactly express the quantum state.

By considering numerical simulations of different two-dimensional frustrated spin-1/2 magnets, we observe that the prefactor A has a universal behavior and, in the case of such systems, grows as $\sim 1/\Delta^2$ with the spectral gap Δ. Such dependence imposes a constraint on the minimum circuit shots number N_s required to reach certain fidelity and on the class of quantum systems addressable with VQE. Based on this observation, we also discuss symmetry-based strategies to effectively increase Δ, thus, in some cases, significantly reducing the required resources for the algorithm.

Before moving forward, let us comment on the choice of two-dimensional frustrated spin–1/2 models as the research focus. These models, to our belief, are among the most promising candidates for near-term quantum advantage for two reasons. (i) First, among the spin models, they represent the most challenging systems for classical calculations [39]. This is in contrast to the one-dimensional or nonfrustrated models in which the ground state can be efficiently computed. (ii) Second, their qubit Hamiltonian remains local [40] (involves only Pauli strings of $O(1)$ length, which does not scale with the number of degrees of freedom, provided that the hardware qubit layout is consistent with the spin lattice).[1] This enables quantum circuits compatible with the current superconducting quantum platforms [43]. Moreover, the gradient descent optimization in such problems requires measurement of a relatively small number of Hamiltonian terms scaling linearly, $O(N_q)$, with the number of qubits.[2]

Therefore, we believe the assessments in this paper are directly related to a highly relevant use-case of variational quantum approaches, which at the same time has a large potential to reach quantum advantage [45].

This paper is organized as follows. In Sec. II, we outline the phenomenological theory for the observed algorithmic phase transition and residual infidelity scaling. In Sec. III, we introduce a model for two-dimensional frustrated magnets and symmetry-enhanced circuits. Finally, in Sec. IV we demonstrate numerical evidence for these claims, and we draw conclusions in Sec. V.

II. PHENOMENOLOGY OF STATE PREPARATION

A. An effective stochastic temperature and algorithmic phase transition

In stochastic optimization methods, such as variational Monte Carlo or gradient-based machine learning, it is known that the power spectrum of statistical noise, under certain assumptions, defines an effective temperature [46–48]. Concretely, we consider a variational circuit parametrized with a vector of parameters $\boldsymbol{\theta}$ and consider the update law representing the SGD approach $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \eta \boldsymbol{f}^t$ with η being the learning rate. In the ideal case, the force is given as the average over the full Hilbert space $\boldsymbol{f}^t = \nabla_{\boldsymbol{\theta}(t)}\langle\psi(\boldsymbol{\theta}(t))|\hat{H}|\psi(\boldsymbol{\theta}(t))\rangle$, while in the SGD algorithm the force is obtained by averaging over N_s circuit shots. In the actual multivariate VQE optimization, noise is governed by a $\boldsymbol{\theta}$–dependent and nondiagonal covariance matrix. Here, to analytically examine the stochastic VQE optimization dynamics, we assume that $f_k \sim \mathcal{N}(\nabla_k\langle\hat{H}\rangle, \sigma)$, i.e., that the forces on the parameters, f_k, are distributed normally with a diagonal and uniform variance, $\sigma^2 \simeq \frac{1}{N_p}\sum_k \mathrm{Var}\, f_k/N_s$. As mentioned above, in this paper, N_s stands for the number of shots used for the estimation of each gradient component. The effective parameter's pseudodynamics

[1] In quantum chemistry, there also exist the ansätze that only utilize qubit-local operations such as the low-depth circuit ansatz (LDCA) [41] or the quantum number preserving ansatz (QNP) [42]. In this paper, importantly, locality of the qubit Hamiltonian stems from the locality of the physical Hamiltonian.

[2] Quantum-chemistry applications are also among the important candidates to reach quantum advantage thanks to the efficient measurement grouping techniques such as proposed in Ref. [44], that allow us to improve the scaling of the required number of measured Hamiltonian terms from $O(N_q^4)$, which naturally arise from the second-quantized chemistry Hamiltonian, to $O(N_q^3)$.

is therefore given by a Langevin equation of the form

$$\theta_k^{t+1} = \theta_k^t - \eta(\nabla_k \langle \hat{H} \rangle + \mathcal{N}(0, \sigma)), \tag{1}$$

where index k enumerates components of the variational parameters vector of length N_p. In addition, we assume that the gradient variances, Var f_k, are approximately equal for all parameters.[3] Then, the stationary solution of Eq. (1) is the Boltzmann distribution $\Pi(\boldsymbol{\theta}) \propto \exp(-E(\boldsymbol{\theta})/T)$, where we defined

$$T = \frac{1}{N_p} \sum_k \mathrm{Var}\, f_k \eta / N_s \tag{2}$$

as an effective temperature of the system. Note that since the learning rate has the dimension of inverse energy, the temperature has the dimension of energy, as expected. This definition agrees with the well-known expression for the effective temperature in variational Monte Carlo optimization [47,48]. Without loss of generality, we assume that the ground-state energy value is $E = 0$, and is also reached at $\boldsymbol{\theta} = \mathbf{0}$. We thus write to the second order $E(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \hat{D}\, \boldsymbol{\theta}/2 = (1/2) \sum_k D_k \tilde{\theta}_k^2$, where we consider a basis $\tilde{\boldsymbol{\theta}}$ delivering a diagonal form to $\hat{D}$ [49–51]. This implies that energy fluctuations are proportional to the number of parameters N_p with $T/2$ per degree of freedom,[4] namely, $\langle E(\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}} = (1/2) N_p T$, where $\langle \ldots \rangle_{\boldsymbol{\theta}}$ denotes averaging over the parameters' distribution.

Below we will provide numerical evidence supporting the existence of an algorithmic phase transition as a function of the energy fluctuations' measure $\epsilon = (1/2) N_p T$, separating the $\epsilon > \epsilon^c$ regime where the optimization is impossible from the $\epsilon < \epsilon^c$ regime where the algorithm finds a sizable finite overlap with the ground state. Notably, the switch between the two regimes often occurs not through a smooth crossover but is instead characterized by a sharp phase transition with a well-defined ϵ^c, also marked, similar to the heat capacity behavior in second-order phase transitions, by a peak in the energy variance $\langle (E - \langle E \rangle_{\boldsymbol{\theta}(t)})^2 \rangle_{\boldsymbol{\theta}(t)}$, where we average over the parameters $\boldsymbol{\theta}(t)$ obtained during the optimization process. We will also show that ϵ^c does not decrease exponentially with the system size, keeping the state preparation feasible while approaching the thermodynamic limit.

We emphasize that our numerical experiment setup is different from the standard thermodynamic case, where both the energy fluctuations and the full problem Hamiltonian energy scale Λ are proportional to the number of system degrees of freedom, i.e., are both extensive quantities. In our setup, Λ, being a characteristic of a problem and not of the ansatz, is independent of N_p. In the meantime, the energy fluctuations ϵ, depending on the ansatz, are linear in N_p. Intuitively, the trainable phase should satisfy $\epsilon \ll \Lambda$. Since the latter is independent of N_p, the magnitude of the energy fluctuations $\epsilon = (1/2) N_p T$, and not temperature T, sets a measure detecting the algorithmic phase transition.

[3]This assumption is numerically well verified in the case of the ansatz states considered in the following.

[4]Strictly speaking, the energy fluctuation should be the sum of effective temperatures $\langle E(\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}} = (1/2) \sum_k T_k$ associated to the system parameters. In our case, however, for simplicity, we assume all temperatures equal.

Substituting the nondiagonal $\boldsymbol{\theta}$–dependent covariance matrix with the diagonal uniform variance σ^2 is a major approximation. We show, however, that sampling from the Boltzmann distribution $\exp(-E(\boldsymbol{\theta})/T)$ defining the partition function

$$\mathcal{Z}_{\boldsymbol{\theta}} = \int \mathrm{d}\boldsymbol{\theta} \exp\left(-E(\boldsymbol{\theta})/T\right) \tag{3}$$

can reproduce some features of the original stochastic VQE optimization, such as the mentioned parametrical regions with vanishing training fidelity and the pronounced separation from the trainable phase with the transition point marked by a peak of energy variance. Thus, the reported behavior manifests itself also within sampling the variational circuit parameters from the thermal Boltzmann distribution $\Pi(\boldsymbol{\theta}) \propto \exp(-E(\boldsymbol{\theta})/T)$.

B. A phenomenological scaling law for the residual infidelity

In the $\epsilon < \epsilon^c$ regime, when the optimization reaches sizable fidelity, we propose the following empirical scaling law for the residual infidelity $\mathcal{I} = 1 - |\langle \psi(\boldsymbol{\theta}) | \psi_0 \rangle|^2$:

$$\mathcal{I} = A\epsilon + \mathcal{I}_0, \tag{4}$$

where $\mathcal{I}_0$ describes the ideal representational ineffectiveness of a given variational circuit. It has been demonstrated in an ideal exact gradient setup that $\mathcal{I}_0$ can be made arbitrarily small with a suitable choice of the ansatz and the corresponding circuit depth [31,52,53]. However, in addition to the limitations associated with the expressivity of the circuit, the measured state infidelity also stems from the finite sample size of gradient estimates per component, N_s.

We argue that this residual infidelity depends on the same effective energy fluctuations' measure ϵ introduced above and on the spectral properties of the system of interest, such as the gap to the first excited state Δ. Our numerical experiments suggest a fast scaling with the inverse gap, which to a reasonable degree follows the $A \propto 1/\Delta^2$ behavior. This proportionality implies an increasing optimization complexity for systems with a closing gap. This dependence is similar to the one of the adiabatic theorem, which imposes the relation $\tau \propto 1/\Delta^2$ on the smallest time extent of the quantum annealing [54]. In the following, we also provide a recipe allowing one to ameliorate this problem by employing symmetry projections [23,55].

III. MODEL AND QUANTUM CIRCUITS

We perform numerical experiments on the $j_1 - j_2$ Heisenberg spin–1/2 model on a series of two-dimensional lattices. We place both model and geometries in focus since (i) the ground state of this model under certain conditions realizes a quantum spin liquid, an exotic and long-sought phase of matter [56,57], (ii) tuning the couplings ratio j_1/j_2 allows one, on each of these geometries, to explore different model regimes, and to trigger gap closing, (iii) lattice models allow for a controllable study of the thermodynamic limit.

The spin-1/2 Heisenberg model is described by the Hamiltonian

$$\hat{H} = j_1 \sum_{\langle i,j \rangle} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j + j_2 \sum_{\langle\langle i,j \rangle\rangle} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j, \tag{5}$$

with $\langle \ldots \rangle$ representing the j_1 bonds and $\langle\langle \ldots \rangle\rangle$ representing the j_2 bonds. The range of ratios j_2/j_1 is chosen, in each particular case, to interpolate between the frustrated and magnetically ordered regimes. See Appendix A 1 for the definition of the lattice geometries and for the description of the parameter regimes for which a gap closing is expected.

For numerical experiments with the SGD and SR techniques, we symmetrize our wave function such that it transforms as a specific irreducible representation of the lattice symmetry group. Among the possible approaches to symmetrization, one can highlight working with the gate set that respects the desired symmetry [58,59] and working with an arbitrary gate set and symmetrizing the wave function after the application of all gates [23]. In this paper, we investigate the former approach, as it provides higher expressive power to the ansatz at a relatively shallow depth [60].

The spin-spin interaction in the Hamiltonian [see Eq. (5)] can be replaced by the *SWAP operator* $\hat{P}_{ij} = \frac{1}{2}(\hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j + \hat{\mathbb{1}})$, which induces an exchange of spin states between sites i and j. Importantly, if the initial state is prepared in the total spin-zero $S = 0$ sector, such as the dimer product state

$$|\psi_D\rangle = \bigotimes_{0 \leqslant i < N/2} \frac{1}{\sqrt{2}} (|\uparrow_{2i} \downarrow_{2i+1}\rangle - |\downarrow_{2i} \uparrow_{2i+1}\rangle), \tag{6}$$

action of the eSWAP operators $\exp(i\theta \hat{P}_{ij})$ preserves the total spin and the wave function reads

$$|\psi(\boldsymbol{\theta})\rangle = \left(\prod_{\alpha} e^{i\theta_\alpha \hat{P}_{i_\alpha j_\alpha}} \right) |\psi_D\rangle, \tag{7}$$

where to define the $(i_\alpha,\ j_\alpha)$ pairs we employ the checkerboard decomposition of the Hamiltonian Eq. (5). Specifically, in the case of a square lattice with periodic boundary conditions (PBCs) in both directions, the Hamiltonian can—up to a constant—be written as a sum of $L \times 4$ SWAP operators acting between pairs of qubits. We split this set of pairs into eight layers of full coverings (when each qubit belongs to exactly one pair), and treat these pairs as $(i_\alpha,\ j_\alpha)$. Note that all eSWAP operators within one layer can be applied simultaneously. In addition to the fixed total spin quantum number, the ansatz also allows one to fix the spatial point-group symmetry representation of the wave function. We use the postcircuit symmetrization protocol introduced by Ref. [23] (for details of such fixation, optimization protocol and the definition $(i_\alpha,\ j_\alpha)$ in the triangular, hexagonal, and open boundary condition (OBC) cases, see Appendix A 2 for more details about the gate layout).

This ansatz allows one to effectively change Δ (if the symmetry is imposed, effective Δ is the energy of the first excited state *in the respective irreducible representation*) and improves the ability to express the ground state (decreases $\mathcal{I}_0$), keeping N_p relatively low. Importantly, the introduced ansatz only shows a mild decay of the gradient norms with system size, signaling the absence of the so-called barren plateau issue, proven to emerge in a generic quantum circuit setup [29]. The proposed circuit has a subexponential decay of gradients, similarly to the case of the transverse field Ising model reported in Ref. [61], as we demonstrate in Appendix A 3. Other barren plateau-free circuits aimed at the study of frustrated two-dimensional magnets were also previously reported [31,32]. Finally, we note that the eSWAP gates are simply the restriction of the two-qubit rotation gates $\hat{U}_{ij}(\theta, \varphi) = \exp(i\theta(\hat{X}_i\hat{X}_j + \hat{Y}_i\hat{Y}_j) + i\varphi\hat{Z}_i\hat{Z}_j)$ to the SU(2)-symmetric case $\varphi = \theta$. These two-qubit gates are the standard choice for the VQE studies of spin–1/2 Heisenberg magnets [23,31,62–64]. Therefore, the proposed ansatz Eq. (A2) represents a rather general family of circuits with the additional advantage of the imposed spin quantum number conservation.

IV. RESULTS

A. Small sample size regime: Critical behavior

The optimization of the circuit parameters requires the estimation of the energy gradient with respect to the gate parameters $\boldsymbol{\theta}$, for which a circuit is being executed N_s times per gradient component. Therefore, the total number of circuit shots is $N_s N_{\text{par}}$. Larger N_s leads to higher fidelity $\mathcal{F} = |\langle\psi(\boldsymbol{\theta})|\psi_0\rangle|^2$ (i.e., the squared overlap with the ground state), as well as for expectation values of the form $\langle\psi(\boldsymbol{\theta})|\hat{O}|\psi(\boldsymbol{\theta})\rangle$ for any given observable $\hat{O}$. A naively expected behavior in such a case would be a gradual growth of fidelity with N_s. However, Ref. [60], which studies neural quantum states [65] (an ansatz class used in *classical* computing) reported a sharply different behavior with nearly zero fidelity for $N_s < N_s^c$ smaller than some critical N_s^c number of samples.[5]

1. *Manifestation of the critical behavior*

To test which scenario (a smooth transition or a critical behavior) is realized with the quantum state preparation, we perform optimization with SGD in the small-N_s regime.[6] We plot the circuit fidelity (or, similarly, infidelity) throughout the Results section on the vertical axis. To obtain the circuit fidelity, we employ the following protocol: Starting from a random set of parameters, we optimize a circuit using the SGD or SR approach until the sliding mean of fidelity with a window of several hundred steps stabilizes.[7]

After such convergence, we use the mean over the next thousand iteration steps as the circuit fidelity. The procedure is repeated ten times to estimate the error bars. In Fig. 1(a), we show the state fidelity as a function of N_s for various j_2/j_1 on the 4×4 square lattice with the OBC in the first and PBC in the second dimension to avoid geometrical frustrations in the $j_2/j_1 = 0$ pure nearest-neighbor case. One can see that

[5]Another prominent example of an algorithmic phase transition occurring in an optimization problem is the so-called *jamming transition* reported for an artificial neural network training [66].

[6]In this regime of small N_s, the metric tensor inversion requires a very large regularization (see Appendix A 2), effectively turning SR into SGD.

[7]By stabilization, we assume that the relative fluctuations of the sliding mean do not exceed 5% of the current sliding-mean value.

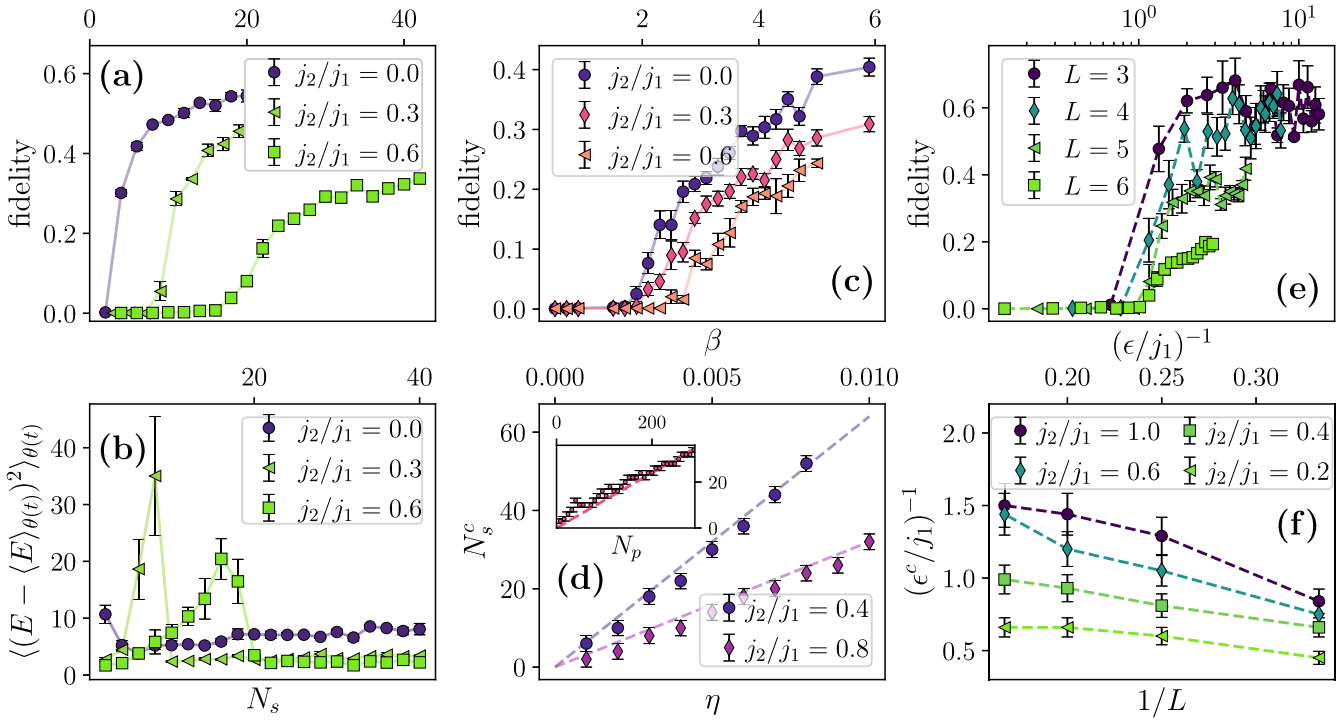

FIG. 1. (a), (b) Fidelity and energy variance as a function of N_s for various j_2/j_1 measured on the 4×4 square lattice with periodic boundary conditions (PBCs) and with the circuit depth $D = 8$ (the number of parameters $N_p = 64$). (c) Fidelity as a function of inverse temperature β within simulations of the 4×4 square lattice at various j_2/j_1 within direct sampling from the thermal partition function Eq. (3). (d) Critical number of samples N_s^c as a function of learning rate η on the 4×4 triangular lattice with PBC. (d) Inset: Critical number of samples N_s^c as a function of number of parameters N_p on the 4×4 square lattice with PBC at $j_2/j_1 = 0.4$. (e), (f) Circuit fidelity as a function of ϵ/j_1 within the $L \times 4$ setup at $j_2/j_1 = 0.4$ and the inverse critical fluctuation measure $(\epsilon^c/j_1)^{-1}$ as a function of $1/L$.

the fidelity remains vanishingly small for $N_s < N_s^c$, followed by a rapid growth afterwards. Such drastic change of pattern is suggestive of a *critical behavior* (see Appendix A 4 for similar behavior in other frustrated lattices). In Appendix A 4, we also show that at the transition, some optimization attempts manage to obtain nonzero fidelity, which nevertheless widely oscillates during optimization. Note that the transitions happen in a small-N_s regime with tens of samples per gradient component. There, the QNGD method requires large regularization due to the ill conditioning and effectively implements the SGD approach. This justifies only considering the SGD optimization in Sec. II and in the present simulations.

To explore this transition in N_s, in Fig. 1(b) we show $c_V/N_s^2 = \mathrm{Var}\, E$ as a function of N_s, where c_V is the effective specific heat, and the variance is computed over the energy values that were obtained during a long (converged) optimization process. We observe that the peak of $\mathrm{Var}\, E$ coincides precisely with N_s^c, defined as the departure point from the zero-fidelity regime. The effective specific heat c_V also shows a peak in this region. This quantity generalizes the thermal specific heat $c_V^\beta = \beta^2\, \mathrm{Var}\, E$, where we assume that the inverse number of samples, $1/N_s$, plays the role of an effective temperature, and β is the inverse temperature. We observe a similar critical behavior across other cluster dimensions and geometries under consideration. Notably, the reported algorithmic phase transition is also accompanied by a qualitative change in the distribution of the overlaps with higher excited states $O_k = |\langle \psi(\boldsymbol{\theta}) | \psi_k \rangle|^2$, namely, it changes from being peaked at $1/|\mathcal{H}|$ (with $|\mathcal{H}|$ being the size of the Hilbert space) to be much wider (see Appendix A 4). Thus, at small N_s, the circuit learns *uniform* overlap with all eigenstates, while at $N_s > N_s^c$ the circuit favors only several low-lying excitations. This is consistent with the argument outlined in Sec. II, based on the thermal partition function.

To substantiate this, we simulate the parametric partition function $\mathcal{Z}_{\boldsymbol{\theta}}$ defined in Eq. (3) using a Metropolis algorithm that performs random walks in the parameter space. In Fig. 1(c), we show the average fidelity computed along the generated Markov chain at different temperatures for the 4×4 square lattice. We keep the number of parameters fixed and thus present the data as a function of $\beta = 1/T$. Even though the growth after the critical value β_c is not as sharp as observed in Fig. 1(a), we notice also in this case a region of nearly zero fidelity for $\beta < \beta_c$. Similarly to the case in Fig. 1(b), the specific heat c_V^β has a pronounced peak at β_c. The histogram of overlaps O_k shows a similar qualitative change between the two regimes (see Appendix A 4 for details).

We note that the abruptness of growth at $N_s > N_s^c$ ($\beta > \beta_c$) is clearly dependent on j_2/j_1, as seen from contrasting the $j_2/j_1 = 0.6$ and $j_2/j_1 = 0.0$ curves demonstrated in Figs. 1(a) and 1(c). However, the transition at $j_2/j_1 = 0$ appears at $N_s \to 0$, but at finite β. Furthermore, this sampling fails to reproduce the maximum fidelities at large β. We believe this is due to the sampling from the thermal partition function and neglecting correlation between the noise components. Nevertheless, the clear separation between the two

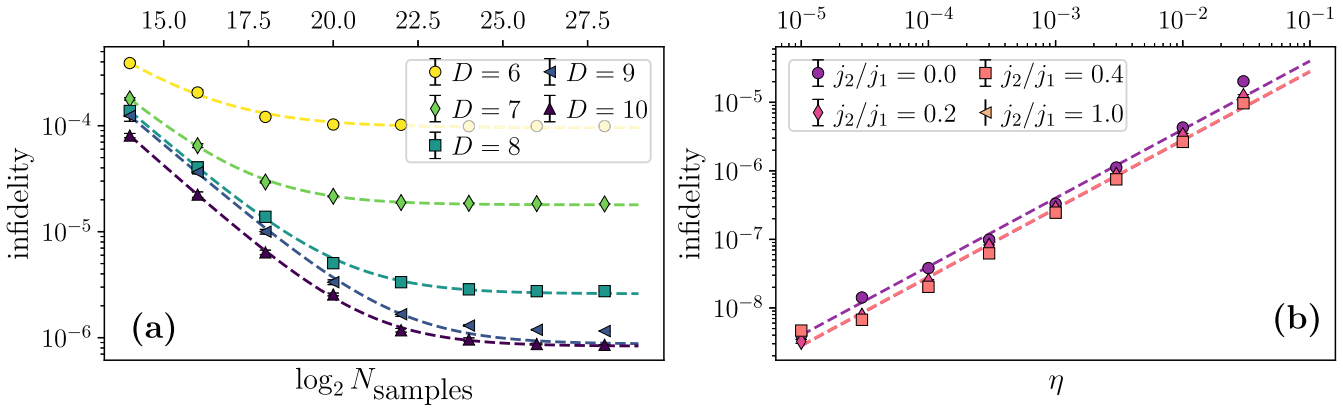

FIG. 2. Large-N_s regime study of the 4×4 square lattice. (a) Infidelity as a function of N_s for a set of depths $D = N_p/8$ at $j_2/j_1 = 0.4$ computed with $\eta = 10^{-2}$. The dashed line shows fit with the ansatz Eq. (4). (b) Infidelity as a function of η for a set of j_2/j_1 at $D = 8$. The data was generated at $N_s = 2^{20}$ to ensure convergence and well conditioning of the metric tensor at any value of η under consideration.

algorithmic regimes marked by a peak of energy variance, remains present.

To verify the effective temperature expression Eq. (2), in Fig. 1(d), we plot N_s^c as a function of η for the 4×4 triangular lattice. To estimate the transition position N_s^c, we take the smallest N_s such that the averaged fidelity exceeds the double inverse Hilbert space size. From Fig. 1(d), we observe a linear dependence on the learning rate η. This supports the linear η proportionality in Eq. (2). The inset shows N_s^c as a function of N_p. To vary $N_p = 8D$, we change the circuit depth D, i.e., the number of parameter layers, as discussed in Appendix A 2. As expected, we observe a linear dependence, while the variance per parameter, $\frac{1}{N_p}\sum_k \text{Var}\, f_k$, remains mostly unchanged. This behavior is also observed in other lattice geometries. This shows a pronounced criticality in the energy fluctuations measure $\epsilon = (1/2)TN_p$, rather than in the mere effective temperature T.

2. *Critical behavior in the thermodynamic limit*

We aim to> extrapolate the critical energy fluctuations measure ϵ^c to the thermodynamic limit. To this end, we consider $L \times 4$ square lattices with the x extent of the lattice $L = 3$, 4, 5, and 6. As the circuit architecture, we select the eight checkerboard decomposition layers of the Hamiltonian in Eq. (5). Thus, due to the high computational cost otherwise, the depth of the circuit does not scale with L. Additionally, OBCs in the L direction and PBCs in the second direction are imposed, ensuring the absence of geometric frustrations. In Fig. 1(e), we report the state fidelity as a function of ϵ, while in Fig. 1(f), we show the dependence of ϵ^c on the lattice dimension L^{-1}. We observe a saturation with $1/L$ and note that the critical energy fluctuations do now grow exponentially with system size.[8] Note that the curves in Fig. 1(e) approach different final asymptotic values $\mathcal{F}_\infty$ for $\epsilon < \epsilon^c$, while fair extrapolation to the thermodynamic limit should ensure a nonvanishing (preferably same) saturation. Thus, as L increases, one needs to employ deeper circuits, which would correspondingly increase ϵ^c at a given L. However, due to the high computational cost, we refrain from scaling up the circuit depth at large cluster volumes $L = 5$, 6. Nevertheless, we believe that such an increase in the circuit depth would not change the overall subexponential scaling character of $\epsilon^c(L)$. For instance, Ref. [67] reports that, to reach a certain fidelity $\mathcal{F}$ with the ground state on the frustrated N-site kagome lattice, only $\propto \sqrt{N}$ gates are necessary. Thus, with such depth adjustment, the saturating behavior of $\epsilon^c(L)$ shown in Fig. 1(f) can only change to a polynomial dependence on L and not to an exponential growth.

The main practical quantity of interest is $N_{\text{tot}}(L) = N_{\text{SGD}}(N_s^c) \times N_s^c$, the total number of samples required to train a circuit in the course of $N_{\text{SGD}}(N_s^c)$ SGD iterations, which is proportional to the required hardware resources. We define $N_{\text{SGD}}(N_s)$ as the number of SGD steps needed to reach 90% of the saturated fidelity. We observe that $N_{\text{SGD}}(N_s)$ required to saturate $\mathcal{F}_\infty$ [as defined in Fig. 1(e)] shows no growth with L (see Appendix A 5 for details). Thus, the obtained scaling for $\epsilon^c(L)$ results in a polynomial scaling of $N_{\text{tot}}(L)$. Note that reducing η will allow reaching ϵ^c with a smaller number of samples per SGD iteration. However, such convergence would require proportionally more SGD steps N_{SGD}, making $N_{\text{tot}}(L)$ a true lower bound on the required total number of samples.

B. Large sample-size regime

In this section, we provide numerical evidence to validate Eq. (4) in the regime of accurate gradients. We employ the symmetrized wave function and the SR approach proposed in Ref. [33]. In this section, N_s has the meaning of the number of shots per gradient component and per symmetry projection (see Appendix A 2 for projection definition).

We consider the 4×4 square lattice at $j_2/j_1 = 0.4$ with a variable circuit depth. The circuit is built of the repetitive applications of the checkerboard decomposition of the Hamiltonian in Eq. (5) (a single decomposition is described in Sec. III), restricting ourselves to a maximum of N_p optimization parameters. The circuit depth is then $D = N_p/8$.

[8] We note that ϵ_c/j_1 varies only mildly with j_2/j_1, while the system gap has a two orders of magnitude difference between $j_2/j_1 = 0.0$ and 0.6. This suggests that in the trainable phase, in the definition $\epsilon \ll \Lambda$ given above, Λ is not the energy gap of $\hat{H}$ but rather some other energy scale of the system.

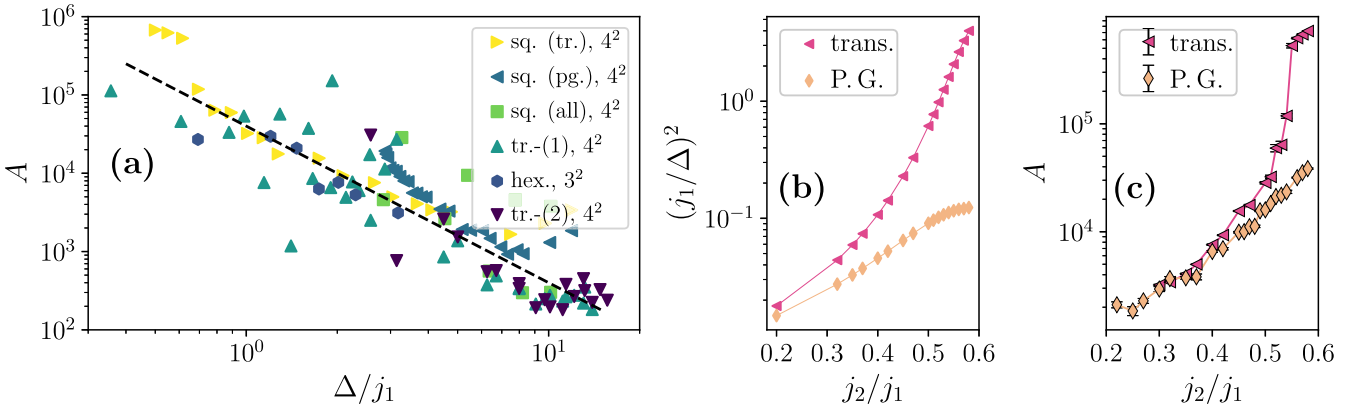

FIG. 3. (a) Dependence of the prefactor A introduced in Eq. (4) on the system spectral gap. The symbols denote sq. (tr.), 4^2: square 4^2 lattice case with only translational symmetry imposed; sq. (pg.), 4^2: same with only point symmetry group imposed;, sq. (all), 4^2: same with both translation and point group symmetries imposed; tr.-(1), 4^2: triangular 4^2 lattice with the full symmetry imposed with the starting dimerization along the j_1 bonds; tr.-(2): same with the dimerization along the j_2 bonds; and hex., 3^2: 3^2 hexagonal lattice with the full symmetry group imposed. (b) On the 4×4 square lattice, dependence the system gap $(j_1/\Delta)^2$ on j_2/j_1 in the two cases: (i) with only point group symmetry imposed, (ii) with only translational symmetry imposed. (c) Same for the resulting prefactor A.

In Fig. 2(a), we present the infidelity as a function of N_s. To verify the proposed functional form for $\mathcal{I}(N_s)$, the data are fitted with the expression in Eq. (4). To verify the power law, we replace $N_s \to N_s^\alpha$ and observe $\alpha \sim 1$ within the error bars consistently within the fitting procedure. The residual infidelity also decays exponentially with depth (for details, see Appendix A 6). Then, we study the dependence of A on circuit depth D (see Appendix A 6). We observe that the product $A(D)D^2$, which is proportional to the total number of measurements required to estimate the metric tensor, changes slowly with D. Therefore, as D grows, the individual components of G_{ij} can be measured with smaller precision, which is leveled out by the growing number of the matrix entries. Finally, in Fig. 2(b) we fit infidelity with a linear function $C\eta$, where C is a fit constant. These fits support the functional form Eq. (4) and provide an empirical way to verify the $1/N_s$ power law dependence of the residual infidelity.

In the following, we investigate the dependence of prefactor A on the spectral gap Δ. To this end, in Fig. 3(a) we present the offset A against $(\Delta/j_1)^{-1}$ for the different two-dimensional lattices considered in this paper. We consider different lattice geometries, different starting dimerization patterns, and symmetry projectors. The fit with $A_0\Delta^\alpha$ yields $\alpha \sim -2$. The black dashed line shows the A_0/Δ^2 fit of the data. We note that in the cases of other spin systems that we considered, such as the j_1–j_2 Heisenberg chain, we observed a scaling law significantly different from $1/\Delta^2$ and being highly depth dependent. We thus emphasize that our claim $\alpha \sim -2$ only concerns two-dimensional frustrated magnets.

From this observation, it follows that for increasingly smaller gaps, the systems become intractable, in the sense that it becomes harder to converge to the true ground state. However, imposing symmetrization of the ansatz (see Appendix A 2) can alleviate this problem. We illustrate this approach in the case of the 4×4 square lattice, where the parameter regime $j_2/j_1 \to 0.6$ leads to a nearly closing gap, accompanied by the emergence of a quantum spin liquid [68], namely, we contrast the cases of (i) only point-group symmetry imposed with eight terms in the projector against (ii) only translations imposed with 16 terms in the projector. As seen in Fig. 3(b) at $j_2/j_1 \to 0.6$, $(j_1/\Delta)^2$ diverges for case (ii), while it remains well-behaving for case (i). This results in a dramatically different behavior for $A(j_2/j_1)$, with the latter being moderate for (i) and exploding for (ii), as seen in Fig. 3(c). This improvement in the prefactor A magnitude emphasizes the importance of symmetries in the state preparation process of systems with a vanishing gap. Summarizing, the residual state infidelity, which is not due to the lack of representability of the chosen ansatz, but rather to the optimization process, scales as

$$\mathcal{I} - \mathcal{I}_0 \propto \frac{\epsilon}{\Delta^2}, \tag{8}$$

where $\epsilon = (1/2)TN_p$, with T defined in terms of sampling shots N_s and learning rate η, as in Eq. (2).

V. DISCUSSION

Variational state preparation is essential for various quantum computing algorithms in near-term and fault-tolerant regimes. Here we show that, besides the most apparent dependency on the circuit architecture, the fidelity of the prepared state strongly depends on the learning hyperparameters and the system-dependent properties. We lay out a phenomenology of state preparation as a function of two significant factors: (i) the number of samples N_s used to estimate the gradient of the variational parameters in an SGD step and (ii) the fundamental gap Δ of the model Hamiltonian. In particular, we explored the interplay of these two parameters, focusing on the two-dimensional spin-1/2 frustrated quantum magnets. These problems represent one of the most challenging cases for the classical approaches, while retaining locality in the qubit form. These problems are therefore marked as being among the most promising candidates to reach quantum advantage in the variational quantum studies. To express the ground state of such a model, we employ the SU(2)-protecting circuit consisting of eSWAP gates, which is only a mild restriction of the XXZ two-qubit rotations, standard in the variational studies of the spin-1/2 Heisenberg models.

This circuit is optimized using the SGD approach, a scalable optimization method in the limits of the large parameters number. Finally, we restrict our consideration to solely future fault-tolerant devices.

We observe that, in the regime of small N_s (noisy gradient), the stochastic optimization shows critical behavior with near-zero state fidelity for $N_s < N_s^c$ and rapid growth of fidelity at $N_s > N_s^c$. The pace of fidelity growth is highly geometry- and parameter-regime dependent. However, the zero-fidelity region is always present and is pronouncedly separate from the trainable phase. In addition, the point of transition, N_s^c, always features a peak of energy variance $\langle (E - \langle E \rangle_{\boldsymbol{\theta}(t)})^2 \rangle_{\boldsymbol{\theta}(t)}$, resembling heat capacity behavior in second-order phase transitions. Together with the notion of effective temperature, this separation allows us to discuss an effective algorithmic phase transition on the energy fluctuations measure $\epsilon = (1/2) N_p T$ axis. In the case of a two-dimensional square lattice, we found evidence that the critical energy fluctuation ϵ_c scales only polynomially with system size [see Fig. 1(f)], providing the basis for the possible applications of VQE in the study of larger-size frustrated magnets, inaccessible to classical algorithms.

To support the notion of the effective temperature, we show that the observed criticality and energy variance peak can be reproduced within sampling from a simpler parametric partition function Eq. (3) with the classical circuit parameters $\boldsymbol{\theta}$ distributed according to the Boltzmann weight. We note that, as mentioned in Sec. IV, this sampling does not reproduce the maximum circuit fidelity and the correspondence between β_c and N_s^c. We thus provide a simplified picture explaining some features of the reported algorithmic phase transition in ϵ.

The observed threshold N_s^c sets a minimum nonnegligible runtime of a successful VQE algorithm on realistic future quantum hardware. To see this, we consider a small yet interesting case of a $N_q = 10 \times 10$ square lattice. The current reasonable estimation for a single eSWAP gate runtime is in the order of 10^{-4} s in the fault-tolerant quantum regime [69]. With the depth of the circuit of order N_q, a single shot implies a wall time $\tau = 0.01$ s. Therefore, the total run time of each SGD iteration at $N_s = N_s^c$, is of the order $\tau \times N_s^c \times N_{\text{par}} = 0.01\,\text{s} \times 20 \times 5000 \sim 10\,\text{min}$. Here, we estimated the number of parameters as $5000 =$ depth $\times$ (50 eSWAP dimer pairs on each layer) and $N_s = 20$ per gradient component. Of course, translationally invariant architectures can be employed, to reduce N_{par} or the circuit depth [70]. Nevertheless, our finding suggests the existence of a nonvanishing minutes-order lower bound of quantum processing time per optimization step in a potential case of applying VQE to a classically-hard system.

The obtained timescale allows us to view the j_1–j_2 Heisenberg model on a two-dimensional lattice as a reasonable candidate for applying quantum variational methods. In addition, Ref. [39] shows that the frustrated j_1–j_2 models are the hardest spin models for the classical simulations, and therefore good candidates to reach quantum advantage.

We have also observed that the approach to the exact ground state in the large-N_s limit depends heavily on the system spectral properties; namely, in the case studies of two-dimensional frustrated magnets, the N_s-dependent contribution to the residual infidelity Eq. (4) scales as $1/\Delta^2$ with Δ being the energy gap of the studied system. We emphasize that this empirical dependence is only applicable to two-dimensional magnets that were explored in this paper. Moreover, even though we express A as the function of the gap to the first excited state, Δ, the $1/\Delta^2$ scaling is also related to the growing contributions to the infidelity coming from the higher-excited states other than the first excited state, namely, as we show in Appendix A 7 that approaching the gap-closing point on the j_2/j_1 axis is accompanied by the growing low-energy excited states density. The interplay between these two factors might lead to the observed A scaling.

Importantly, there exist numerous improvements of the QNGD method that allow us to reduce the shots budget, such as surrogate models that keep track of the energy landscape on the previous steps [71–73] or adaptive sampling that increases the shots budget in the case of large uncertainty [74,75]. We believe that while these methods will improve the overall prefactor, the problem of increasing hardness and divergence of N_s at the closing gap Δ would remain, which would still require the symmetry enhancement proposed in this paper. The corresponding numerical experiments of optimization with the improved QNGD are beyond the scope of this paper and are left to future works.

We note that the observed $1/\Delta^2$ infidelity scaling is fundamentally different from the well-known runtime scaling coincidentally appearing in the adiabatic quantum computing [76] and incoherent quantum tunneling [77–79]. Moreover, a simple two-level model would yield a $1/\Delta$ scaling, also observed in Ref. [80] [Eq. (25)]. This is because, if only the first excited state plays a role, then $\langle E \rangle_{\boldsymbol{\theta}} = \Delta \langle \mathcal{I} \rangle_{\boldsymbol{\theta}}$. Therefore, since we have shown that $\langle E \rangle_{\boldsymbol{\theta}} \sim 1/T \sim 1/N_s$, we arrive at $\langle \mathcal{I} \rangle_{\boldsymbol{\theta}} \sim 1/(\Delta N_s)$, which yields the $1/\Delta$ scaling of N_s with the gap. In contrast, in the case of the systems considered in this paper, higher excited states also contribute, which leads to a worse, $1/\Delta^2$, scaling.

This $1/\Delta^2$ scaling poses a significant obstacle for the study of systems with a closing gap. To address this problem, we showed how a symmetry-enhanced wave function—in addition to not being susceptible to the barren plateaus issue—can, in some cases, mitigate the effects of a closing gap, namely, imposing symmetry projection can increase the effective system gap Δ in this symmetry sector, relieving the exploding $A = 1/\Delta^2$ prefactor and providing a significant improvement in the algorithm efficiency, which can be quantified in several orders of magnitude. This development will open possibilities for using VQE to simulate complex many-body systems with near-term quantum computers, which could otherwise be intractable due to a closing gap.

ACKNOWLEDGMENTS

We are sincerely grateful to T. Westerhout for useful discussions and help with the technical simulation setup. We thank T. Neupert for helpful comments on our paper. Numerical simulations used the high-performance package LATTICE_SYMMETRIES [81] for quantum state vector manipulation. N.A. is funded by the Swiss National Science Foundation, Grant No. PP00P2_176877. G.M. acknowledges financial support from the Swiss National Science Foundation (Grant PCEFP2_203455). This research was supported by

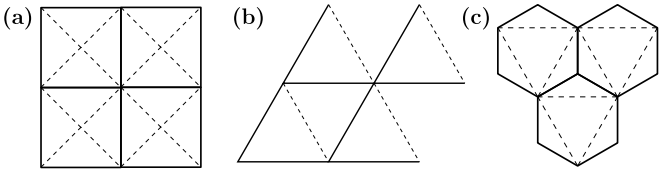

FIG. 4. Geometries of lattices considered in this paper. Solid lines represent j_1 bonds, while j_2 bonds are represented by dashed lines: (a) square lattice, (b) triangular lattice, (c) hexagonal lattice. In the cases of $L \times L$ equilateral 4×4 triangle and 3×3 kagome lattices considered in this paper, L is the number of one-site (triangle) and three-site (kagome) unit cells in each lattice dimension. In the case of the $L \times 4$ square lattice, L is the number of unit cells in the x-direction.

the NCCR MARVEL, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (Grant No. 205602).

APPENDIX

1. Two-dimensional geometries

In Fig. 4, we show the two-dimensional lattice geometries considered in this paper: square, triangle, and hexagonal. The j_1 bonds are marked as solid lines, while the j_2 bonds are shown as dashed lines. Each lattice shows a near-vanishing gap to the first excited state in the $S = 0$ sector in some region of j_2/j_1. For instance, in the case of a square lattice, the vicinity of the $j_2/j_1 \approx 0.55$ fluctuations melt magnetic orders and result in a frustrated phase, possibly gapless QSL [68]. On the triangular lattice, gap shrinks in the vicinity of $j_2/j_1 = 1.0$, where the model is most frustrated [82] and in the vicinity of $j_2/j_1 = 0.2$ on the hexagonal lattice [83].

2. Symmetric circuit routines

Symmetrized wave function: In the Heisenberg Hamiltonian, spin-spin interaction can be replaced with the *SWAP operator* $\hat{P}_{ij} = \frac{1}{2}(\hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j + \hat{\mathbb{1}})$, exchanging spin states on the site i and j. Notably, the SWAP operator commutes with the total spin operator $[\hat{P}_{ij}, \hat{S}^2] = 0$, allowing for working in the wave function sector with fixed total spin [23]. As a result, the symmetry-enhanced ansatz for the system of N spins is constructed by (1) first preparing the system in the simple *fully dimerized* state

$$|\psi_D\rangle = \bigotimes_{0 \leqslant i < N/2} \frac{1}{\sqrt{2}}(|\uparrow_{2i}\downarrow_{2i+1}\rangle - |\downarrow_{2i}\uparrow_{2i+1}\rangle), \tag{A1}$$

being the direct product of $N/2$ dimer pairs,[9] and (2) action of the string of parametrized eSWAP operators preserving the total spin $S = 0$:[10]

$$|\psi\rangle(\boldsymbol{\theta}) = \left(\prod_{\alpha} e^{i\theta_\alpha \hat{P}_{i_\alpha j_\alpha}}\right)|\psi_D\rangle. \tag{A2}$$

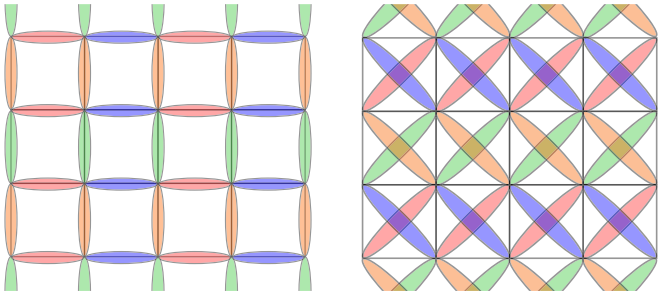

FIG. 5. Decomposition of the Hamiltonian terms into eight layers of full dimerization in the case of the 5×4 square lattice with PBCs (vertical) and OBCs (horizontal).

To define i_α and j_α, we employ checkerboard decomposition of the Hamiltonian Eq. (5); namely, in the case of a square lattice, we split the Hamiltonian terms into eight layers of full dimerizations, when all eSWAP operators within one layer can be applied simultaneously. Each of the eSWAP operators in a layer is parametrized with a distinct angle θ, making all rotation angles independent. First, we apply the four layers of the horizontal and vertical bonds (in the order red-blue-green-yellow) shown in Fig. 5 (left), and then the four layers of the diagonal bonds Fig. 5 (right).

In the case of the triangular lattice, to construct the (i_α, j_α) pairs, we effectively add the missing j_2 bonds to recover the square lattice setup and employ the same decomposition. In the case of the 3×3 hexagonal lattice, we first apply three layers of the j_1 bonds (each of the layers builds a full nearest-neighbor dimerization of the hexagonal lattice). Then, we note that the j_2 bonds act on two triangular sublattices. Therefore, we apply these j_2 bonds as the nearest-neighbor bonds of the two copies of the square lattice.

When all of the Hamiltonian terms are applied in the form of eSWAP (for the square lattice this happens at $D = 8$), to further grow D, we start the procedure over.

In this paper, we consider OBC-PBC and PBC-PBC. For instance, in the case of the 4×5 square lattice, the partition of the Hamiltonian terms into eight layers is shown in Fig. 5. The spatial symmetry projector operator is defined as

$$\hat{P} = \frac{1}{|G|}\sum_{g \in G} \chi_g \hat{g}, \tag{A3}$$

where G is the spatial symmetry group, consisting of the elementary unitary permutations $\hat{g}$ and χ_g are the characters, depending on the desired projection quantum number. The projected wave function $|\psi_P(\boldsymbol{\theta})\rangle = \frac{\hat{P}}{\sqrt{\mathcal{N}(\boldsymbol{\theta})}}|\psi(\boldsymbol{\theta})\rangle$ is normalized with $\mathcal{N}(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|\hat{P}|\psi(\boldsymbol{\theta})\rangle$. The energy gradient

[9] The exact dimerization pattern is chosen to maximize overlap with the ground state.

[10] To construct wave function in another spin sector, one needs to prepare one (or more) electron pairs in spin-triplet state $|\uparrow\uparrow\rangle$.

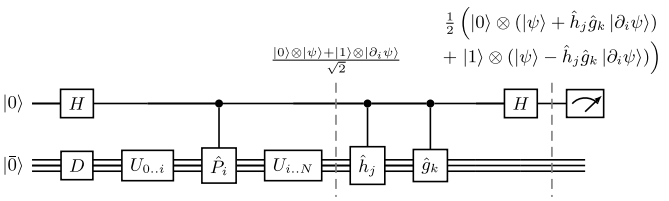

FIG. 6. Circuit for measurement of $\langle\psi|\hat{h}_j\hat{g}_k|\partial_i\psi\rangle$. The Hadamard scheme applied to the ancilla qubit allows us to obtain real and (if necessary for connection and metric tensor) imaginary parts of $\langle\psi|\hat{h}_j\hat{g}_k|\partial_i\psi\rangle$.

reads

$$\partial_i\langle E(\boldsymbol{\theta})\rangle = 2\,\mathrm{Re}\left[\frac{\langle\psi(\boldsymbol{\theta})|\hat{H}\hat{P}|\partial_i\psi(\boldsymbol{\theta})\rangle}{\mathcal{N}(\boldsymbol{\theta})} - \mathcal{A}_i(\boldsymbol{\theta})\langle E(\boldsymbol{\theta})\rangle\right], \tag{A4}$$

where $\mathcal{A}_i(\boldsymbol{\theta}) = \frac{1}{\mathcal{N}(\boldsymbol{\theta})}\langle\psi(\boldsymbol{\theta})|\hat{P}|\partial_i\psi(\boldsymbol{\theta})\rangle$ is the *connection*.

Finally, the metric tensor for natural gradient descent is defined as

$$G(\boldsymbol{\theta})_{ij} = \frac{\langle\partial_i\psi(\boldsymbol{\theta})|\hat{P}|\partial_j\psi(\boldsymbol{\theta})\rangle}{\mathcal{N}(\boldsymbol{\theta})} - \mathcal{A}_i^*(\boldsymbol{\theta})\mathcal{A}_j(\boldsymbol{\theta}), \tag{A5}$$

and is used to improve energy gradient $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \eta\sum_j(\mathrm{Re}\,G(\boldsymbol{\theta}))^{-1}_{ij}\partial_j\langle E(\boldsymbol{\theta})\rangle$ in the spirit of imaginary time evolution within SR [33]. The metric tensor obtained within sampling is regularized, $G_{\mathrm{reg}} = \sqrt{GG} + \beta\hat{\mathbb{1}}$, as suggested in Ref. [84].

Sample quantum circuit: In the course of optimization, it is required to measure expectation values of the kind $\langle\psi(\boldsymbol{\theta})|\hat{h}_j\hat{g}_k|\partial_i\psi(\boldsymbol{\theta})\rangle$, with $\hat{h}_j$ being the jth Hamiltonian term (unitary SWAP operator) and $\hat{g}_k$ being the kth unitary permutation. Note that any $\hat{g}_k$ permutation can be written as a product of note, more than $N-1$ pair SWAP operators. The quantum circuit used to measure such quantity is shown in Fig. 6. The real and imaginary parts of $\langle\psi(\boldsymbol{\theta})|\hat{h}_j\hat{g}_k|\partial_i\psi(\boldsymbol{\theta})\rangle$ are measured using the Hadamard test protocol.

3. Absence of barren plateaus

We investigate the possible presence or absence of barren plateaus in the case of our ansatz applied to Heisenberg models on two-dimensional frustrated spin-1/2 magnets. To this end, in Fig. 7 we plot the gradient magnitude per parameter and qubit $||\nabla E||/(j_1N_pN_q)$ (i) within the 4×4 square lattice setup at $j_2/j_1 = 0.4$ as a function of N_p and (ii) within the $L\times4$ setup on a square lattice at $j_2/j_1 = 0.4$ as a function of L. To obtain the gradient magnitude, we individually draw each angle from the uniform distribution $\theta_k \sim \mathcal{U}(-0.0, 0.1)$, while the starting state is fully dimerized as suggested in Sec. III. We then average over a hundred of such random starting points.

We observe that in the first setup, the average gradient per parameter increases and saturates as a function of the number of circuit parameters N_p. This is because a deep circuit allows one to rapidly change the outcoming quantum state by only small variations of the circuit parameters. In the meantime, in the second setup, the gradients decay sub-exponentially

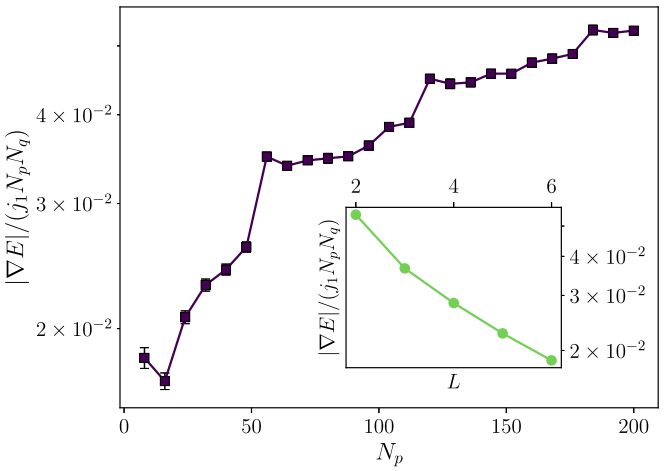

FIG. 7. The gradient magnitude per parameter and qubit $||\nabla E||/(j_1N_pN_q)$ within the 4×4 setup on the square lattice with $j_2/j_1 = 0.4$ as a function N_p. Inset: $L\times4$ setup on the square lattice at $j_2/j_1 = 0.4$ as a function of L.

with the number of qubits N_q. Such behavior was also reported in the case of the transverse field Ising model [61]. Also, our observation is in line with the circuit architectures applied to frustrated two-dimensional spin-1/2 Heisenberg magnets [31,32].

4. Algorithmic phase transition details

In this Appendix section, we provide additional details concerning the observed algorithmic phase transition. In Fig. 8, we plot fidelity as a function of N_s in two additional frustrated lattices: kagome and triangular. We observe similar clear separations of the regions with zero and nonzero fidelity. In Fig. 9, we plot energy variance $\langle(E-\langle E\rangle_{\boldsymbol{\theta}(t)})^2\rangle_{\boldsymbol{\theta}(t)}$ obtained on the 4×4 square lattice within sampling from the classical parametrical partition function $\mathcal{Z}_{\boldsymbol{\theta}}$. The maxima of the energy variance coincide precisely with the separation between the zero overlap and trainable phases observed in Fig. 1(c).

To get further insight into the algorithmic transition, in Fig. 10 we show histograms of fidelity of the first hundred $S=0$ low-lying excited states above and below the transition ($N_s = 6$ and $N_s = 12$) at $j_2/j_1 = 0.4$ on the 4×4 square

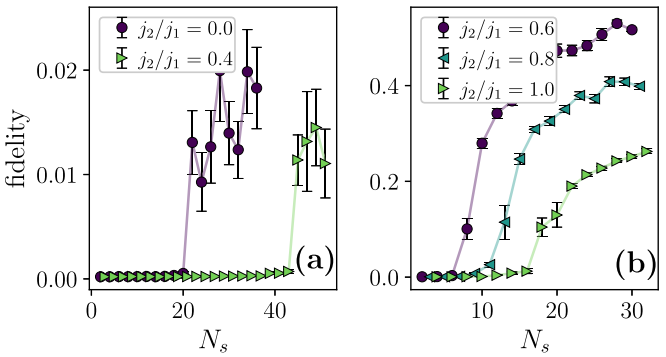

FIG. 8. Fidelity as a function of N_s for various j_2/j_1 measured on the (a) 2×3 kagome lattice (18 sites) with PBC and on the (b) 4×4 triangular lattice with PBC.

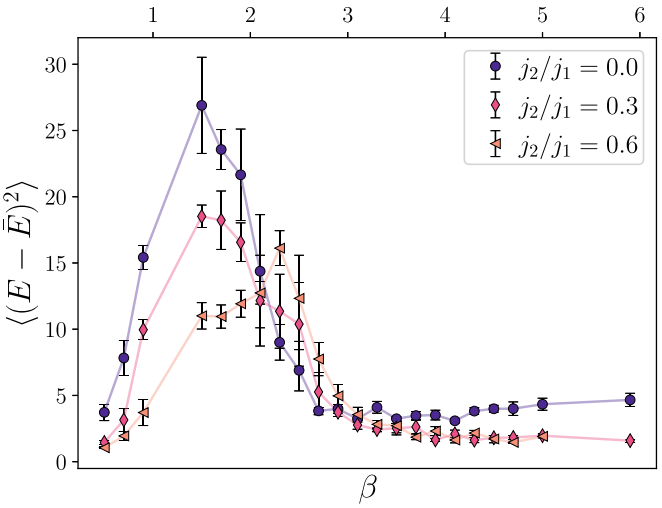

FIG. 9. Energy variance $\langle (E - \langle E \rangle_{\theta(t)})^2 \rangle_{\theta(t)}$ as a function of β on the 4×4 square lattice within sampling from the classical parametrical partition function $\mathcal{Z}_{\theta}$. The positions of maxima correspond to transition between the zero-overlap and trainable regimes.

lattice. We see that below the transition, the histogram is very narrow with the variance $\sigma = 10^{-3}$ and the small average overlap of $\mu = 10^{-3} \approx 1/|\mathcal{H}|$, with $|\mathcal{H}|$ being the $S = 0$ Hilbert space size of the problem. This indicates that all states that can possibly have overlap with our ansatz wave function $|\psi(\boldsymbol{\theta})\rangle$ are nearly of the same fidelity with a small variation. On the contrary, after the algorithmic transition, the histogram widens to $\sigma = 0.036$, showing *selectivity* and that only a few low-energy states dominate the fidelity. This dramatic change of the histogram between $N_s = 6$ and $N_s = 12$ highlights a significant change in the stochastic VQE optimization performance upon passing the algorithmic phase transition.

Additionally, in Fig. 11, we show fidelity as a function of the iteration number at $j_2/j_1 = 0.4$ on the 4×4 square lattice with PBCs and $D = 8$ for various $N_s = 7, 10$, and 14, when the system is, respectively, before, at, and after the algorithmic

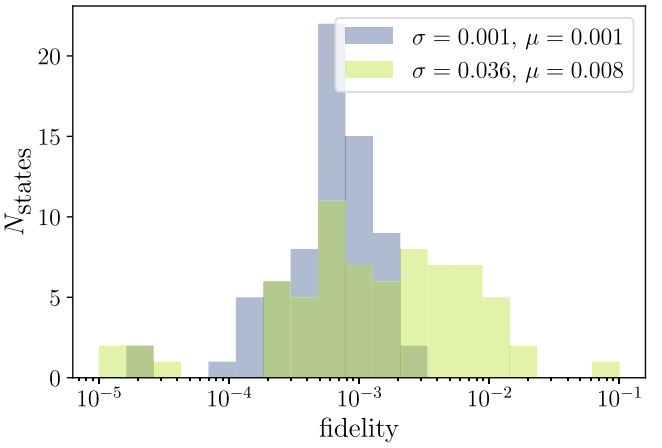

FIG. 10. Histogram of overlaps with excited states $O_k = |\langle \psi_k | \psi(\boldsymbol{\theta}) \rangle|^2$ for k corresponding to the lowest 100 states in the $S = 0$ sector measured at $j_2/j_1 = 0.4$ on 4×4 square lattice with PBC at $N_s = 6$ and $N_s = 12$.

transition. We see that below the transition, all optimization attempts lead to zero fidelity, while at the transition, some attempts are successful, but their fidelities widely oscillate over time. Finally, after the transition, all attempts manage to achieve stable fidelity.

5. Number of iterations to saturation

In Fig. 12, we investigate the number of SGD optimization steps required to reach 90% of the final fidelity F_{∞} within the $L \times 4$ setup on the square lattice at $j_2/j_1 = 0.4$ as a function of inverse energy fluctuation characteristic. We observe no pronounced dependence of $N_{\text{SGD}}(N_s)$ as a function of L at a given ϵ.

This contributes to our argument that the total required amount of computational resources $N_{\text{total}}(N_s) = N_s \times N_{\text{SGD}}(N_s)$ required to train a VQE wave function above the algorithmic phase transition, scales only polynomially with the system volume and thus the observed "noisy" phase with near-zero overlap can be avoided on larger clusters with an affordable circuit shots budget.

6. Large-N_s fit parameters

In the main text, we fit the overlap with the empirical ansatz $\mathcal{I}(N_s) = C/N_s^{\alpha} + \mathcal{I}_0$. The α power was introduced to verify the proposed $1/N_s$ scaling law. In Fig. 13, we show α and I_0 (inset) as functions of $D = N_p/8$ for $j_2/j_1 = 0.4$. The inability to express the ground state I_0, expectedly, decays exponentially with D. In the meantime, α remains close to 1 within error bars. Additionally, we explore the dependence of the A fit parameter on the depth D. In Fig. 14, we consider the case of the 4×4 square lattice with PBC, $j_2/j_1 = 0.0, 0.2, 0.4$ and point-group and translational symmetries applied. We plot $A(D)D^2$ as a function of D and observe that $A(D)D^2$ depends on D only slightly, in contrast to $\mathcal{I}_0(D)$ which decays exponentially. The product $A(D)D^2$ has the meaning of the total number of measurements required to estimate the metric tensor, since N_s in the expression for $\mathcal{I}$ has the meaning of required measurements per gradient component.

7. Phenomenology of the faster-than-$1/\Delta$ scaling

In this Appendix section, we present phenomenological and numerical considerations behind the faster-than-$1/\Delta$ scaling of the infidelity, which is well-described by the $1/\Delta^2$ power in case of the two-dimensional frustrated magnets.

As the first step, let us consider a two-level system, and a variational ansatz depending on a single parameter θ such that $\theta = 0$ corresponds to the ground state. The variational wave function is decomposed into the ground state $|\psi_0\rangle$ (with the energy $E_0 = 0$) and the first excited state $|\psi_1\rangle$ (with the energy $E_1 = \Delta$) as

$$\mathcal{N}|\psi(\theta)\rangle = |\psi_0\rangle + \theta A|\psi_1\rangle, \tag{A6}$$

where $A = \langle \psi_1 | \partial_{\theta} \psi(\theta = 0) \rangle$ and $\mathcal{N}$ is the normalization. Applied to this wave function, the SR optimization performs

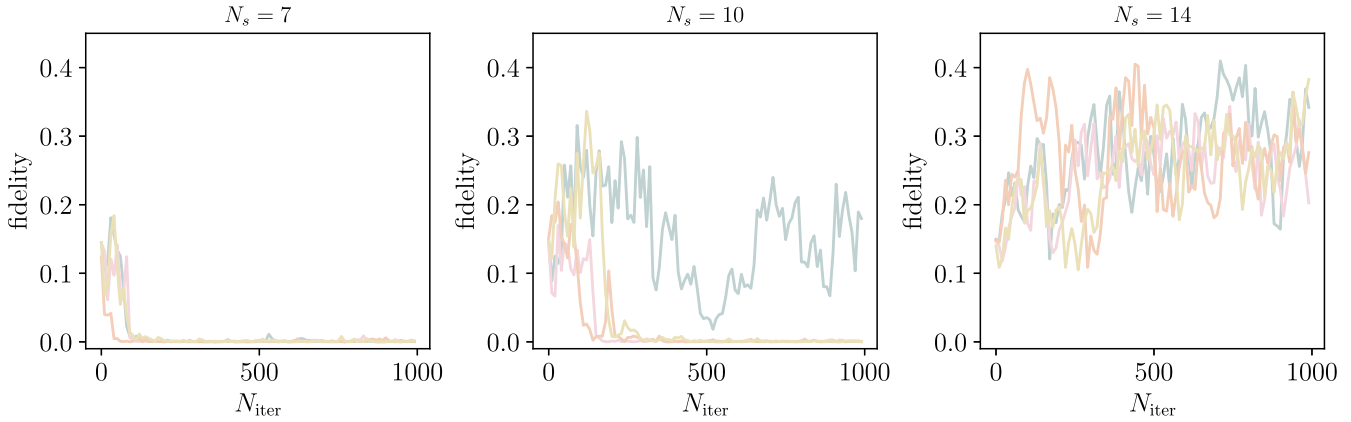

FIG. 11. Fidelity as a function of the iteration number at $j_2/j_1 = 0.4$ on the 4×4 square lattice with periodic boundary conditions and $D = 8$ for various $N_s = 7, 10$, and 14, when the system is, respectively, before, at, and after the algorithmic transition. Different curves correspond to several optimization attempts from random starting points.

imaginary-time evolution in the variational subspace

$$\mathcal{N}\dot{\theta}|\partial_\theta \psi(\theta)\rangle = -\Delta\theta A|\psi_1\rangle, \tag{A7}$$

which finally gives $\dot{\theta} = -\Delta\theta$.

In the noisy optimization considered in this paper, the Langevin equation therefore reads

$$\theta^n - \theta^{n-1} = -\Delta\theta^{n-1} + \xi^n, \tag{A8}$$

where $\xi^n \sim \mathcal{N}(0, \sigma)$ is the random noise and $\sigma^2 = \mathrm{Var}\,\dot{\theta}/N_s$.

The solution of this equation yields $\langle\theta^2\rangle = \sigma^2/\Delta = (\mathrm{Var}\,\dot{\theta})/(N_s\Delta)$. Since the infidelity in Eq. (A6) can be expressed as $\theta^2|\langle\psi_1|\partial_\theta\psi(\theta=0)\rangle|^2$, for the two-level model we obtain

$$\mathcal{I} = \frac{|\langle\psi_1|\partial_\theta\psi(\theta=0)\rangle|^2(\mathrm{Var}\,\dot{\theta})}{N_s\Delta}. \tag{A9}$$

Such analysis in the case of a general many-level system is cumbersome. However, to account for more excited states, we assume that the ground state is given by $\boldsymbol{\theta} = 0$, and each of the variational parameters $\theta_i \neq 0$ is responsible for obtaining overlap with only a single excited state $|\psi_i\rangle$. In such case, the infidelity generalizes the above expression as

$$\mathcal{I} = \frac{1}{N_s}\sum_i \frac{|\langle\psi_i|\partial_{\theta_i}\psi(\boldsymbol{\theta}=0)\rangle|^2(\mathrm{Var}\,\dot{\theta}_i)}{\Delta_i}, \tag{A10}$$

where Δ_i is the energy of the ith excited state. The products $|\langle\psi_i|\partial_{\theta_i}\psi(\boldsymbol{\theta}=0)\rangle|^2$ are ansatz specific, and remain near constant as $\Delta \to 0$. The main contribution to the increasing infidelity comes from the growing angles' amplitudes $\langle\theta_i^2\rangle$, which are controlled by the inverse energies $1/\Delta_i$.

As such, as the system enters the frustrated phase, the overall infidelity can grow both due to the shrinking system gap Δ and the accumulation of excited states in the region of small energies. These two processes could lead to the faster-than-$1/\Delta$ growth of the infidelity, as reported in this paper.

To substantiate this consideration, we consider the translation and symmetry-projected ansatz applied to the 4×4 square lattice with PBC. First, in Fig. 15 we plot the normalized spectral function $\rho(\omega) = \sum_k \delta(\omega - \Delta E_k)|\langle\psi(\boldsymbol{\theta})|\psi_k\rangle|^2$,

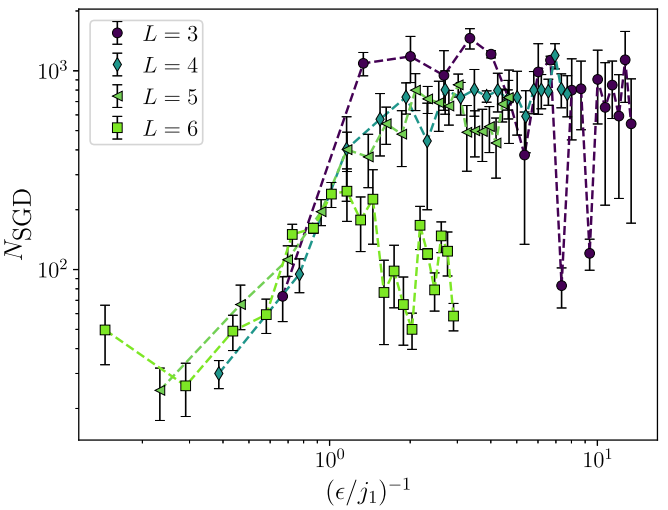

FIG. 12. The number of SGD steps required to reach 90% of the final fidelity F_∞ within the $L \times 4$ setup on the square lattice at $j_2/j_1 = 0.4$ as a function of inverse energy fluctuation $(\epsilon/j_1)^{-1}$.

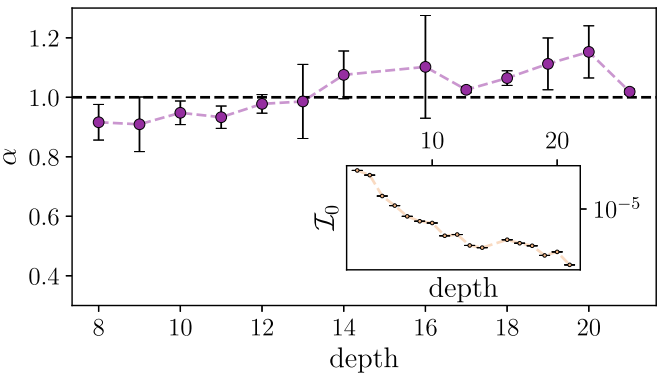

FIG. 13. Depth dependence of the power α at $j_2/j_1 = 0.4$ on a 4×4 square lattice. Inset: $\mathcal{I}_0$ as the function of depth.

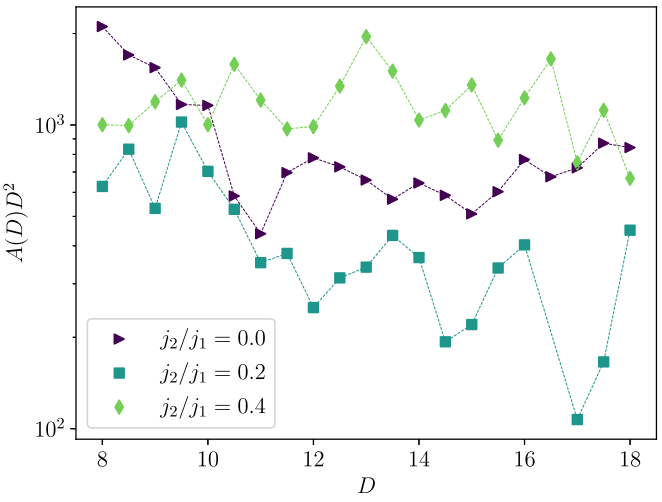

FIG. 14. Depth dependence of $A(D)D^2$ at $j_2/j_1 = 0.0, 0.2, 0.4$ on a 4×4 square lattice with symmetries applied.

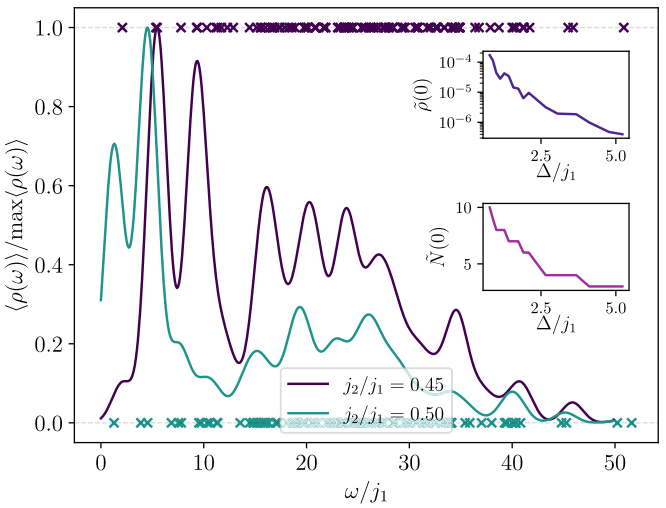

FIG. 15. The normalized spectral function $\rho(\omega)$ for $j_2/j_1 = 0.45,\ 0.5$ on the 4×4 square lattice. The delta-functions are smeared using the Gaussian approximation with the width $\Gamma = j_1$. For $j_2/j_1 = 0.45$, on the $y = 1$ horizontal axis we mark the excited states energies, for $j_2/j_1 = 0.5$ we do similarly at $y = 0$. The lower inset shows $\tilde{N}(0)$, which is the number of excited states below $10j_1$, while the upper inset shows $\tilde{\rho}(0) = \int_0^{10j_1} \rho(\omega)d\omega/\tilde{N}(0)$, which represents average overlap between the variational wave function and the low-lying excited states.

where $|\psi_k\rangle$ is the kth excited state of the system, ΔE_k is its energy distance to the ground state, $|\psi(\boldsymbol{\theta})\rangle$ is the variational state and the overlap is averaged over numerous SR optimization steps after convergence. This object resolves the total infidelity over energies of the excited states. The delta functions are smeared using the Gaussian approximation with the width $\Gamma = j_1$. For $j_2/j_1 = 0.45$, on the $y = 1$ horizontal axis, we mark the excited-state energies; for $j_2/j_1 = 0.5$, we do similarly at $y = 0$.

We see that the low-energy excited states have a much greater overlap with the variational wave function, in accord with our phenomenological theory.

We we mentioned above, at $\Delta \to 0$ the raise in the infidelity stems from excited states' gaps and accumulation of excited states. In the lower inset of Fig. 15, we plot $\tilde{N}(0)$, which is the number of excited states below $10j_1$. We indeed observe rapid growth of the number of low-lying excited states as $\Delta \to 0$. Then, in the upper inset of Fig. 15, we plot $\tilde{\rho}(0) = \int_0^{10j_1} \rho(\omega)d\omega/\tilde{N}(0)$, which represents the average overlap between the variational wave function and the low-lying excited states. Similarly, we observe that individual overlaps also grow at $\Delta \to 0$, which reflects an overall decrease of the excited states' energies.

This interplay of the two factors (shrinking system gap and the growing low-energy excited states density) in the two-dimensional frustrated magnets might lead to the observed approximate $1/\Delta^2$ growth of resource demand.

[1] 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).

[2] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Müller, W. Riess, G. Salis, J. Smolin, I. Tavernelli, and K. Temme, Quantum optimization using variational algorithms on near-term quantum devices, Quantum Sci. Technol. **3**, 030503 (2018).

[3] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio *et al.*, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[4] 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 (London) **549**, 242 (2017).

[5] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush *et al.*, Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator, Phys. Rev. X **8**, 031022 (2018).

[6] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, and P. Zoller, Self-verifying variational quantum simulation of lattice models, Nature (London) **569**, 355 (2019).

[7] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature (London) **567**, 209 (2019).

[8] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[9] S. Chakrabarti, R. Krishnakumar, G. Mazzola, N. Stamatopoulos, S. Woerner, and W. J. Zeng, A threshold

for quantum advantage in derivative pricing, Quantum **5**, 463 (2021).

[10] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, The variational quantum eigensolver: A review of methods and best practices, Phys. Rep. **986**, 1 (2022).

[11] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya *et al.*, Quantum chemistry in the age of quantum computing, Chem. Rev. **119**, 10856 (2019).

[12] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, Phys. Rev. A **92**, 042303 (2015).

[13] M. Cerezo, K. Sharma, A. Arrasmith, and P. J. Coles, Variational quantum state eigensolver, npj Quant. Inform. **8**, 113 (2022).

[14] M. Lubasch, J. Joo, P. Moinier, M. Kiffner, and D. Jaksch, Variational quantum algorithms for nonlinear problems, Phys. Rev. A **101**, 010301(R) (2020).

[15] G. Mazzola, Sampling, rates, and reaction currents through reverse stochastic quantization on quantum computers, Phys. Rev. A **104**, 022431 (2021).

[16] G. Scriva, N. Astrakhantsev, S. Pilati, and G. Mazzola, Challenges of variational quantum optimization with measurement shot noise, arXiv:2308.00044.

[17] D. S. Abrams and S. Lloyd, Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors, Phys. Rev. Lett. **83**, 5162 (1999).

[18] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, Proc. Natl. Acad. Sci. **114**, 7555 (2017).

[19] R. J. Bartlett and M. Musiał, Coupled-cluster theory in quantum chemistry, Rev. Mod. Phys. **79**, 291 (2007).

[20] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, Nat. Commun. **10**, 3007 (2019).

[21] H. R. Grimsley, D. Claudino, S. E. Economou, E. Barnes, and N. J. Mayhall, Is the trotterized UCCSD ansatz chemically well-defined? J. Chem. Theory Comput. **16**, 1 (2020).

[22] G. Mazzola, S. V. Mathis, G. Mazzola, and I. Tavernelli, Gauge-invariant quantum circuits for $u(1)$ and Yang-Mills lattice gauge theories, Phys. Rev. Res. **3**, 043209 (2021).

[23] K. Seki, T. Shirakawa, and S. Yunoki, Symmetry-adapted variational quantum eigensolver, Phys. Rev. A **101**, 052340 (2020).

[24] E. R. Anschuetz, Critical points in quantum generative models, arXiv:2109.06957.

[25] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, Nat. Commun. **13**, 7760 (2022).

[26] X. You, S. Chakrabarti, and X. Wu, A convergence theory for over-parameterized variational quantum eigensolvers, arXiv:2205.12481.

[27] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, Diagnosing barren plateaus with tools from quantum optimal control, Quantum **6**, 824 (2022).

[28] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

[29] 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).

[30] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, Absence of Barren Plateaus in Quantum Convolutional Neural Networks, Phys. Rev. X **11**, 041011 (2021).

[31] J. L. Bosse and A. Montanaro, Probing ground-state properties of the kagome antiferromagnetic Heisenberg model using the variational quantum eigensolver, Phys. Rev. B **105**, 094409 (2022).

[32] J.-G. Liu, Y.-H. Zhang, Y. Wan, and L. Wang, Variational quantum eigensolver with fewer qubits, Phys. Rev. Res. **1**, 023025 (2019).

[33] S. Sorella, Green Function Monte Carlo with Stochastic Reconfiguration, Phys. Rev. Lett. **80**, 4558 (1998).

[34] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, Quantum **4**, 269 (2020).

[35] G. Torlai, G. Mazzola, G. Carleo, and A. Mezzacapo, Precise measurement of quantum observables with neural-network estimators, Phys. Rev. Res. **2**, 022060(R) (2020).

[36] K. Nakano, C. Attaccalite, M. Barborini, L. Capriotti, M. Casula, E. Coccia, M. Dagrada, C. Genovese, Y. Luo, G. Mazzola *et al.*, Turborvb: A many-body toolkit for ab initio electronic simulations by quantum Monte Carlo, J. Chem. Phys. **152**, 204121 (2020).

[37] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems*, 1st ed. (Cambridge University Press, Cambridge, 2017).

[38] J. Lorraine, P. Vicol, and D. Duvenaud, Optimizing millions of hyperparameters by implicit differentiation, arXiv:1911.02590.

[39] D. Wu, R. Rossi, F. Vicentini, N. Astrakhantsev, F. Becca, X. Cao, J. Carrasquilla, F. Ferrari, A. Georges, M. Hibat-Allah, M. Imada, A. M. Läuchli, G. Mazzola, A. Mezzacapo, A. Millis, J. R. Moreno, T. Neupert, Y. Nomura, J. Nys, O. Parcollet *et al.*, Variational benchmarks for quantum many-body problems, arXiv:2302.04919 [quant-ph].

[40] J. Kempe, A. Kitaev, and O. Regev, The complexity of the local Hamiltonian problem, arXiv:quant-ph/0406180 [quant-ph].

[41] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer, arXiv:1801.01053.

[42] G.-L. R. Anselmetti, D. Wierichs, C. Gogolin, and R. M. Parrish, Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems, New J. Phys. **23**, 113010 (2021).

[43] F. Arute, K. Arya, R. Babbush, D. Bacon, J. Bardin, R. Barends, R. Biswas, S. Boixo, F. Brandao, D. Buell, B. Burkett, Y. Chen, J. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[44] T.-C. Yen, A. Ganeshram, and A. F. Izmaylov, Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates, arXiv:2201.01471.

[45] Y. Zhang, L. Cincio, C. F. A. Negre, P. Czarnik, P. J. Coles, P. M. Anisimov, S. M. Mniszewski, S. Tretiak, and P. A. Dub, Variational quantum eigensolver with reduced circuit complexity, npj Quantum Inf. **8**, 96 (2022).

[46] F. Becca, W.-J. Hu, Y. Iqbal, A. Parola, D. Poilblanc, and S. Sorella, Lanczos steps to improve variational wave functions, J. Phys.: Conf. Ser. **640**, 012039 (2015).

[47] S. Jastrzebski, Z. Kenton, D. Arpit, N. Ballas, A. Fischer, Y. Bengio, and A. Storkey, Three factors influencing minima in SGD, arXiv:1711.04623.

[48] F. He, T. Liu, and D. Tao, Control batch size and learning rate to generalize well: Theoretical and empirical evidence, in *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., Red Hook, New York, 2019), Vol. 32.

[49] X. Cheng, D. Yin, P. L. Bartlett, and M. I. Jordan, Stochastic gradient and Langevin processes, arXiv:1907.03215.

[50] S. Mandt, M. D. Hoffman, and D. M. Blei, A variational analysis of stochastic gradient algorithms, arXiv:1602.02666.

[51] P. Chaudhari and S. Soatto, Stochastic gradient descent performs variational inference, converges to limit cycles for deep networks, arXiv:1710.11029.

[52] A. Choquette, A. D. Paolo, P. K. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, Quantum-optimal-control-inspired ansatz for variational quantum algorithms, Phys. Rev. Res. **3**, 023092 (2021).

[53] L. Mineh and A. Montanaro, Solving the Hubbard model using density matrix embedding theory and the variational quantum eigensolver, Phys. Rev. B **105**, 125117 (2022).

[54] M. S. Sarandy, L. A. Wu, and D. A. Lidar, Consistency of the adiabatic theorem, arXiv:quant-ph/0405059.

[55] K. Seki and S. Yunoki, Spatial, spin, and charge symmetry projections for a Fermi-Hubbard model on a quantum computer, Phys. Rev. A **105**, 032419 (2022).

[56] L. Savary and L. Balents, Quantum spin liquids: A review, Rep. Prog. Phys. **80**, 016502 (2017).

[57] Y. Zhou, K. Kanoda, and T.-K. Ng, Quantum spin liquid states, Rev. Mod. Phys. **89**, 025003 (2017).

[58] L. Schatzki, M. Larocca, Q. T. Nguyen, F. Sauvage, and M. Cerezo, Theoretical guarantees for permutation-equivariant quantum neural networks, arXiv:2210.09974.

[59] J. J. Meyer, M. Mularski, E. Gil-Fuster, A. A. Mele, F. Arzani, A. Wilms, and J. Eisert, Exploiting symmetry in variational quantum machine learning, PRX Quantum **4**, 010328 (2023).

[60] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson, and A. A. Bagrov, Generalization properties of neural network approximations to frustrated magnet ground states, Nat. Commun. **11**, 1593 (2020).

[61] R. Wiersema, C. Zhou, Y. de Sereville, J. F. Carrasquilla, Y. B. Kim, and H. Yuen, Exploring entanglement and optimization within the Hamiltonian variational ansatz, PRX Quantum **1**, 020319 (2020).

[62] M. S. Jattana, F. Jin, H. D. Raedt, and K. Michielsen, Assessment of the variational quantum eigensolver: Application to the Heisenberg model, Front. Phys. **10**, 907160 (2022).

[63] V. Feulner and M. J. Hartmann, Variational quantum eigensolver ansatz for the J_1–J_2-model, Phys. Rev. B **106**, 144426 (2022).

[64] J. Kattemölle and J. van Wezel, Variational quantum eigensolver for the Heisenberg antiferromagnet on the kagome lattice, Phys. Rev. B **106**, 214429 (2022).

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

[66] S. Spigler, M. Geiger, S. d'Ascoli, L. Sagun, G. Biroli, and M. Wyart, A jamming transition from under- to over-parametrization affects generalization in deep learning, J. Phys. A: Math. Theor. **52**, 474001 (2019).

[67] J. L. Bosse and A. Montanaro, Probing ground state properties of the kagome antiferromagnetic Heisenberg model using the variational quantum eigensolver, arXiv:2108.08086.

[68] Y. Nomura and M. Imada, Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy, Phys. Rev. X **11**, 031034 (2021).

[69] A. G. Fowler and C. Gidney, Low overhead quantum computation using lattice surgery, arXiv:1808.06709.

[70] S. Oh, J. Choi, and J. Kim, A tutorial on quantum convolutional neural networks (QCNN), arXiv:2009.09423.

[71] R. Shaffer, L. Kocia, and M. Sarovar, Surrogate-based optimization for variational quantum algorithms, Phys. Rev. A **107**, 032415 (2023).

[72] C. N. Self, K. E. Khosla, A. W. R. Smith, F. Sauvage, P. D. Haynes, J. Knolle, F. Mintert, and M. S. Kim, Variational quantum algorithm with information sharing, npj Quantum Inf. **7**, 116 (2021).

[73] A. W. R. Smith, A. J. Paige, and M. S. Kim, Faster variational quantum algorithms with quantum kernel-based surrogate models, Quantum Sci. Technol. **8**, 045016 (2023).

[74] M. Majland, P. Ettenhuber, and N. T. Zinner, Fermionic adaptive sampling theory for variational quantum eigensolvers, arXiv:2303.07417.

[75] A. Mukherjee, N. F. Berthusen, J. C. Getelina, P. P. Orth, and Y.-X. Yao, Comparative study of adaptive variational quantum eigensolvers for multi-orbital impurity models, Commun. Phys. **6**, 4 (2023).

[76] T. Kadowaki and H. Nishimori, Quantum annealing in the transverse Ising model, Phys. Rev. E **58**, 5355 (1998).

[77] U. Weiss, H. Grabert, P. Hänggi, and P. Riseborough, Incoherent tunneling in a double well, Phys. Rev. B **35**, 9535 (1987).

[78] S. V. Isakov, G. Mazzola, V. N. Smelyanskiy, Z. Jiang, S. Boixo, H. Neven, and M. Troyer, Understanding Quantum Tunneling Through Quantum Monte Carlo Simulations, Phys. Rev. Lett. **117**, 180402 (2016).

[79] G. Mazzola, V. N. Smelyanskiy, and M. Troyer, Quantum Monte Carlo tunneling from quantum chemistry to quantum annealing, Phys. Rev. B **96**, 134305 (2017).

[80] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).

[81] T. Westerhout, Lattice-symmetries: A package for working with quantum many-body bases, J. Open Source Software **6**, 3537 (2021).

[82] Y. Iqbal, W.-J. Hu, R. Thomale, D. Poilblanc, and F. Becca, Spin liquid nature in the Heisenberg $J_1 - J_2$ triangular antiferromagnet, Phys. Rev. B **93**, 144411 (2016).

[83] S.-S. Gong, D. N. Sheng, O. I. Motrunich, and M. P. A. Fisher, Phase diagram of the spin-$\frac{1}{2}$ J_1-J_2 Heisenberg model on a honeycomb lattice, Phys. Rev. B **88**, 165138 (2013).

[84] J. Gacon, C. Zoufal, G. Carleo, and S. Woerner, Simultaneous perturbation stochastic approximation of the quantum fisher information, Quantum **5**, 567 (2021).