

## Benchmarking variational quantum eigensolvers for the square-octagon-lattice Kitaev model

Andy C. Y. Li<sup>1</sup>, M. Sohaib Alam,<sup>2</sup> Thomas Iadecola<sup>3,4</sup>, Ammar Jahin<sup>5</sup>, Joshua Job<sup>6</sup>, Doga Murat Kurkcuoglu,<sup>1</sup> Richard Li,<sup>7</sup> Peter P. Orth<sup>3,4</sup>, A. Barış Özgüler<sup>1</sup>, Gabriel N. Perdue<sup>1</sup>, and Norm M. Tubman<sup>8</sup>

<sup>1</sup>Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA

<sup>2</sup>Rigetti Computing, Berkeley, California 94701, USA

<sup>3</sup>Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA

<sup>4</sup>Ames National Laboratory, Ames, Iowa 50011, USA

<sup>5</sup>Department of Physics, University of Florida, 2001 Museum Road, Gainesville, Florida 32611, USA

<sup>6</sup>Lockheed Martin Advanced Technology Center, Sunnyvale, California 94089, USA

<sup>7</sup>Department of Physics, Yale University, New Haven, Connecticut 06520, USA

<sup>8</sup>Quantum Artificial Intelligence Laboratory, Exploration Technology Directorate, NASA Ames Research Center, Moffett Field, California 94035, USA



(Received 16 December 2022; accepted 1 April 2023; published 1 August 2023)

Quantum spin systems may offer the first opportunities for beyond-classical quantum computations of scientific interest. While general quantum simulation algorithms likely require error-corrected qubits, there may be applications of scientific interest prior to the practical implementation of quantum error correction. The variational quantum eigensolver (VQE) is a promising approach to finding energy eigenvalues on noisy quantum computers. Lattice models are of broad interest for use on near-term quantum hardware due to the sparsity of the number of Hamiltonian terms and the possibility of matching the lattice geometry to the hardware geometry. Here, we consider the Kitaev spin model on a hardware-native square-octagon qubit connectivity map, and examine the possibility of efficiently probing its rich phase diagram with VQE approaches. By benchmarking different choices of variational *Ansatz* states and classical optimizers, we illustrate the advantage of a mixed optimization approach using the Hamiltonian variational *Ansatz* (HVA) and the potential of probing the system's phase diagram using VQE. We further demonstrate the implementation of HVA circuits on Rigetti's Aspen-9 chip with error mitigation.

DOI: [10.1103/PhysRevResearch.5.033071](https://doi.org/10.1103/PhysRevResearch.5.033071)

### I. INTRODUCTION

In the context of quantum computation there is reason to believe the quantum simulation of spin systems may offer early results in the search for beyond classical computations of real scientific interest [1]. Although beyond-classical calculations that offer truly new insights into scientific problems will likely require quantum error correction (QEC) [2,3], it may be possible to find approachable questions of scientific interest even before the advent of full QEC if we carefully pair the computing hardware and problem. Lattice models are a natural class of systems to consider in this regard, and they also tend to have sparse Hamiltonian representations that require fewer quantum resources than other Hamiltonian models [4]. In this paper, we consider a spin model [5,6] that maps naturally onto the square-octagon qubit connectivity of a superconducting quantum processor and perform an open study of the most efficient ways of finding the ground state of this system using variational quantum algorithms based on parametrized quantum circuits.

The near-term, so-called noisy intermediate scale quantum (NISQ) [7–9] computing era is defined by the conditions imposed by noisy quantum hardware. Decoherence errors mandate the use of “shallow” quantum circuit programs, i.e., quantum circuits with a small number of consecutive operations on the qubit register array. Quantum noise also currently limits the effective “width” of a circuit, which is governed by the total number of qubits deployed. A wide circuit consists of more quantum gates than a narrow circuit, and thus the gate error rates effectively limit how wide a circuit could be before the result becomes no longer useful with the decreasing circuit fidelity.

The variational quantum eigensolver (VQE) [10–18] is an algorithm that seeks to implement low-depth variational *Ansatz* circuits to find eigenstates of complex many-body Hamiltonians. Different approaches to VQE entail different choices of *Ansatz* circuits, which take into account different figures of merit. These can include reducing the circuit depth, increasing the accuracy or wave-function fidelity, or minimizing the number of variational parameters. To match these varied goals, there are many different approaches which include hardware efficient *Ansätze* (HEA) [13], adaptive *Ansätze* [14,16,17,19–22], and *Ansätze* inspired by classical simulations [13,17,23]. One particular *Ansatz* of interest is the Hamiltonian variational *Ansatz* (HVA) [24]. This *Ansatz* draws from ideas expressed by the quantum approximate

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.

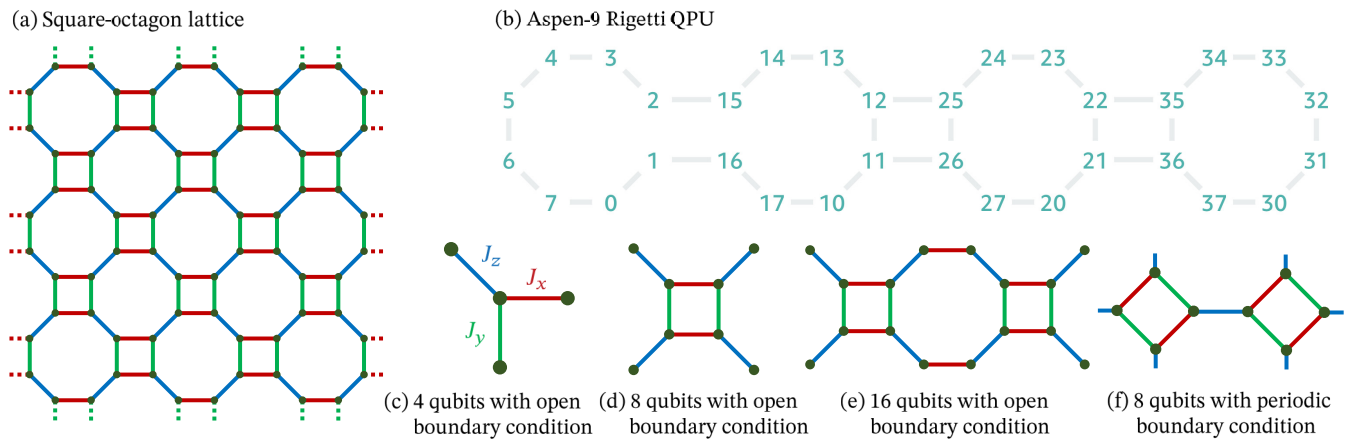


FIG. 1. Lattice geometry. A square-octagon lattice with the Kitaev couplings is shown in (a). The Aspen-9 QPU is arranged with a square-octagon connectivity shown in (b). Note that qubits 1 and 2 and qubits 15 and 16 are not connected due to a hardware issue. The Aspen-9 geometry allows us to pick an appropriate sublattice to simulate a square-octagon-lattice Kitaev model with open boundary conditions, for example, the four-qubit setup in (c), the eight-qubit lattice in (d), and the 16-qubit lattice in (e). We also consider an eight-qubit setup with periodic boundary conditions, shown in (f), to demonstrate VQE calculations of expectation values. The  $J_x$  (red),  $J_y$  (green), and  $J_z$  (blue) couplings are assigned to each linkage.

optimization algorithm and adiabatic quantum computation [25,26]. Recent work [27] suggests the HVA may be less prone to problems with “barren plateaus” [28] and therefore easier to optimize than the HEA (see, however, Ref. [29]).

Kitaev spin models [5,30,31], which are a family of frustrated quantum spin models with bond-dependent interactions, provide an intriguing testbed for NISQ quantum simulation. Kitaev models can be defined on arbitrary trivalent graphs and are appealing due to their exact solvability via a mapping to free Majorana fermions [5]. Despite their simplicity, they yield a rich variety of phases, including gapped  $\mathbb{Z}_2$  and gapless  $U(1)$  spin liquids [32]. In a magnetic field, Kitaev models support non-Abelian Ising anyons [5], which (in addition to the aforementioned  $\mathbb{Z}_2$  spin-liquid phase) makes them a promising platform for topological quantum computation [33–35]. Kitaev-like models are predicted to arise in spin-orbit-coupled Mott insulators [36], and a plethora of materials candidates exist [31]. Putative signatures of spin-liquid physics have been observed in neutron scattering [37,38] and thermal transport measurements [39,40].

In realistic systems, the desired “Kitaev interactions” compete with more mundane (e.g., Heisenberg) interactions and external magnetic fields, all of which spoil the Kitaev model’s exact solvability and often favor magnetically ordered ground states. One therefore must resort to numerical methods [41–55], such as exact diagonalization (ED), tensor-network techniques like the density-matrix renormalization group [56–58], and Monte Carlo [59,60], to study the ground-state phase diagram. Thus, Kitaev models provide a useful benchmark for near-term quantum algorithms for the study of interacting quantum systems.

In this paper, we focus on the Kitaev model on the square-octagon lattice [6,61], which maps natively with minimal compilation overhead onto Rigetti’s Aspen architecture featuring 32 transmons with a square-octagon topology on the Aspen-9 chip used in this paper [62–64]. This native mapping makes the quantum simulation of the square-octagon Kitaev

model a potential NISQ application of the Aspen-9 quantum processing unit (QPU) to carry out quantum computing calculations without full QEC. We will benchmark the variational *Ansätze* and optimization algorithms to determine the ground state of the Kitaev model with and without a magnetic field. Supported by the results of classical simulations, we will illustrate a mixed optimization approach using both a local optimizer and a nonlocal optimizer started with multiple initial values. This opens up the possibility of using VQE calculations to probe the phase diagram of the Kitaev model in the presence of a magnetic field or other perturbations away from the solvable limit. Together with an experimental test run of classically optimized VQE circuits on the QPU, this paper provides insights into the appropriate VQE approach for further VQE experiments with a system size beyond the capability of classical computation.

This paper is organized as follows. In Sec. II A, we review the Kitaev model on the square-octagon lattice and its phase diagram in the presence of a magnetic field. This is followed by a review of VQE approaches in Sec. II B. We discuss our methodology for investigating the circuit *Ansätze* (Sec. III B) and optimizers (Sec. III C), and for implementing the classically optimized circuits on Aspen-9 with noise mitigation (Sec. III E). The results of the paper are then discussed in Sec. IV.

## II. BACKGROUND

### A. Square-octagon-lattice Kitaev model

The ferromagnetic Kitaev model on the square-octagon lattice has the Hamiltonian

$$H_K = -J_x \sum_{x \text{ bonds}} X_i X_j - J_y \sum_{y \text{ bonds}} Y_i Y_j - J_z \sum_{z \text{ bonds}} Z_i Z_j, \quad (2.1)$$

where  $J_{x,y,z} > 0$  and we partition the nearest-neighbor bonds of the lattice into  $x$ ,  $y$ , and  $z$  sets depending on their orientation [see Fig. 1(a)], and  $\{X_i, Y_i, Z_i\}$  are the Pauli matrices for site

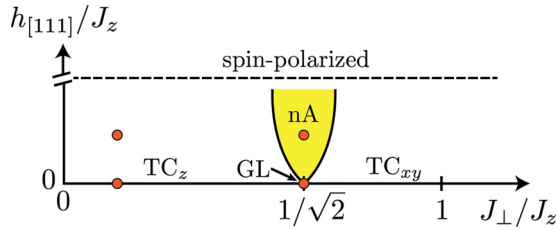


FIG. 2. Phase diagram. Schematic phase diagram of the Kitaev model (2.2) on the square-octagon lattice as a function of spin exchange anisotropy  $J_{\perp}/J_z$  with  $J_{\perp} \equiv J_x = J_y$  and magnetic field in [111] direction  $h_{[111]}$ . It includes gapped toric code phases ( $TC_z$ ,  $TC_{xy}$ ) that are stable with respect to small fields, the gapless line (GL) at  $J_{\perp}/J_z = 1/\sqrt{2}$ , and a phase with non-Abelian (nA) Majorana excitations that emerges in field above the gapless line. At large magnetic fields the system enters a spin-polarized paramagnetic phase. The red circles denote the different representative model parameter points that are studied in our benchmark simulations.

*i.* Like the original honeycomb-lattice version, the Kitaev model  $H_K$  on the square-octagon lattice is exactly solvable by mapping to free fermions [6]. To move beyond the exactly solvable limit and explore the rich phase diagram emerging in finite magnetic field (see Fig. 2), we consider the model

$$H = H_K + \sum_i (h_x X_i + h_y Y_i + h_z Z_i), \quad (2.2)$$

where  $\vec{h} = (h_x, h_y, h_z)$  is the magnetic field. This Hamiltonian and variants thereof have been studied in Refs. [6,61,65–69] using a variety of methods, and the general features of its phase diagram are well understood.

At zero field, the model features two gapped  $\mathbb{Z}_2$  spin-liquid phases, which we dub  $TC_z$  and  $TC_{xy}$ . The  $TC_z$  phase can be understood perturbatively in the limit  $J_z^2 \gg J_x^2 + J_y^2$ , where an effective Hamiltonian equivalent to the well-known toric code [70] emerges [6,61]. The  $TC_{xy}$  phase can be understood in the opposite limit  $J_x^2 + J_y^2 \gg J_z^2$ , where perturbation theory yields [61] a Hamiltonian equivalent to the so-called Wen-plaquette model [71], which is in turn equivalent to the toric code [72,73]. In both TC phases, the model has  $\mathbb{Z}_2$  topological order and supports Abelian anyonic excitations with nontrivial mutual statistics; nevertheless, the phases are distinct [61,67]. A critical line separating these two gapped phases appears at  $J_z^2 = J_x^2 + J_y^2$ .

TABLE I. Definition of studied model parameter values. This table contains the model parameter values that are used in our benchmark simulations below. They correspond to positions in the phase diagram in the toric code phase without a magnetic field ( $TC_z$ ) and with a magnetic field ( $TC_z + h$ ), on the gapless line (GL) and above the gapless line in a magnetic field (GL+h). The ground-state energy  $E_g$  is determined by ED for the four-qubit setup ( $N = 4$ ), the eight-qubit lattice ( $N = 8$ ), and the 16-qubit lattice ( $N = 16$ ) with open boundary conditions shown in Figs. 1(b)–1(d).

Label	Model parameters						Ground-state energy $E_g$		
	$J_x$	$J_y$	$J_z$	$h_x$	$h_y$	$h_z$	$N = 4$	$N = 8$	$N = 16$
$TC_z$	0.1	0.1	1	0	0	0	-1.0100	-4.0100	-8.0250
$TC_z+h$	0.1	0.1	1	$\frac{0.05}{\sqrt{3}}$	$\frac{0.05}{\sqrt{3}}$	$\frac{0.05}{\sqrt{3}}$	-1.1723	-4.2476	-8.5002
GL	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	1	0	0	0	-1.4142	-4.4721	-9.3002
GL+h	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	1	$\frac{0.05}{\sqrt{3}}$	$\frac{0.05}{\sqrt{3}}$	$\frac{0.05}{\sqrt{3}}$	-1.5831	-4.7011	-9.7008

At small but finite field, the TC phases persist [66], as is expected due to the gapped nature of these phases. Provided that none of  $h_{x,y,z} = 0$ , the gapless line at  $J_z^2 = J_x^2 + J_y^2$  gives way to a gapped phase with non-Abelian Majorana excitations [6]. At large field, the system enters a trivial spin-polarized paramagnetic phase. The phase diagram of the square-octagon Kitaev model in a field has been studied in Refs. [6,65,66], but a detailed understanding of the location of all transitions is lacking. A schematic of the phase diagram for  $J_x = J_y = J_{\perp}$  in a [111]-oriented field  $\vec{h} = h_{[111]}(1, 1, 1)/\sqrt{3}$  is shown in Fig. 2.

In the following, we focus on a number of representative points in the phase diagram, which are highlighted in Fig. 2 and defined in Table I. The table also includes the exact ground-state energies for these parameters that are obtained using ED, which we will use to benchmark our VQE results. We consider lattices with open boundary conditions shown in Figs. 1(b)–1(d) for the benchmarks.

### B. VQE

VQE algorithms prepare the eigenstates of the system Hamiltonian  $H$  by optimizing the cost function associated with a trial state  $|\psi(\vec{\theta})\rangle$ , which is prepared by a parametrized circuit *Ansatz*  $U(\vec{\theta})$  such that  $|\psi(\vec{\theta})\rangle = U(\vec{\theta})|0\rangle$ , where  $|0\rangle$  is a chosen reference state [11,13]. To prepare the ground state, the VQE algorithm minimizes, using a classical optimizer, the energy of the trial state, i.e.,

$$E(\vec{\theta}) = \langle \psi(\vec{\theta}) | H | \psi(\vec{\theta}) \rangle. \quad (2.3)$$

The cost function is measured on the QPU, and the result is fed into the classical optimizer, making the VQE a hybrid quantum-classical approach. Several different approaches are possible for calculating excited states [10,12,16,74]. One approach is to minimize a modified cost function including the overlaps with the lower-energy eigenstates determined in the previous iterations [75,76].

Similar to other variational algorithms, the output of VQE only approximates the eigenstates of  $H$ , and the quality of the solution depends on the choice of *Ansatz*. The true ground state cannot be expressed by the trial state  $|\psi(\vec{\theta})\rangle$  if an inappropriate *Ansatz* is chosen, resulting in an approximated energy much higher than the true ground-state energy. The success of VQE algorithms hence strongly depends on the

circuit *Ansatz*. Optimizing a generic HEA is a challenging task for several reasons, including barren plateaus [28] and sensitivity to the optimizer metaparameters [77]. The challenge of optimization grows even bigger with the rapidly increasing number of variational parameters with the system size. The HVA, built using rotations generated by Hamiltonian terms, has been suggested to be a more efficient option than HEA in certain cases [27], as has been observed in several numerical simulations [77–79]. Later in this paper, we will compare the effectiveness of these two *Ansätze* in preparing the ground state of the square-octagon Kitaev model with 8 and 16 qubits.

The choice of classical optimization algorithm plays an important role in VQE implementations as well. Picking an inappropriate optimizer results in a slow convergence rate (and thus a longer QPU runtime), if it ever converges to a sufficiently good minimum at all [77,80–82]. We will also benchmark the efficiency of a few optimizers using the Kitaev model to provide some insights into the VQE optimization strategy.

### III. METHODOLOGY

#### A. Overview of approach

In the following sections we review our theoretical and experimental approaches to studying VQE simulations of the Kitaev model. Our paper first aims, via classical simulations, to understand the modeling requirements for obtaining different phases of the Kitaev model with various *Ansätze* and optimizers, and then tests these approaches on present-day quantum hardware. In this paper, we classically simulate the quantum programs using PYQUIL [83], CIRQ [84], and QISKIT [85], and we execute classically optimized quantum circuits on Rigetti’s Aspen-9 QPU.

#### B. VQE *Ansatz*

The QPUs are initialized in a chosen reference state  $|0\rangle$  with all physical qubits being at zero in the computational basis. The *Ansatz* trial state  $|\psi(\vec{\theta})\rangle = U(\vec{\theta})|0\rangle$  is prepared by two classes (HVA and HEA) of parametrized circuit *Ansätze*  $U(\vec{\theta})$  to be discussed in this section.

We construct the HVA using the Kitaev couplings  $J_x$ ,  $J_y$ , and  $J_z$  in Eq. (2.1), and the magnetic fields  $h_x$ ,  $h_y$ , and  $h_z$  in Eq. (2.2). The HVA represented by the unitary matrix  $U(\vec{\theta})$  has  $L$  layers such that

$$U(\vec{\theta}) = U_{\text{HVA}}(\vec{\theta}_L)U_{\text{HVA}}(\vec{\theta}_{L-1})\cdots U_{\text{HVA}}(\vec{\theta}_1). \quad (3.1)$$

Each layer  $U_{\text{HVA}}(\vec{\theta}_\ell)$  consists of the exponentiation of the Kitaev couplings and the magnetic fields multiplied by the parameters  $\vec{\theta}_\ell$  such that

$$U_{\text{HVA}}(\vec{\theta}_\ell) = e^{-i\theta_{\ell,6} \sum_i Z_i} e^{-i\theta_{\ell,5} \sum_{z \text{ bonds}} Z_i Z_j} \\ \times e^{-i\theta_{\ell,4} \sum_i Y_i} e^{-i\theta_{\ell,3} \sum_{y \text{ bonds}} Y_i Y_j} \\ \times e^{-i\theta_{\ell,2} \sum_i X_i} e^{-i\theta_{\ell,1} \sum_{x \text{ bonds}} X_i X_j}. \quad (3.2)$$

The two-qubit gate count  $N_b$  of each HVA layer scales linearly with the number  $N$  of qubits. (For an infinite square-octagon lattice,  $N_b = \frac{3}{2}N$ .) The Aspen-9 connectivity shown

in Fig. 1(b) is natively the square-octagon lattice. One HVA layer can then be executed with  $N_b \sim N$  CPHASE gates supported natively by Aspen-9 [64]. If we execute the circuit on a QPU with a different two-dimensional connectivity, an overhead of  $\mathcal{O}(N)$  SWAP gates will be required for implementing the HVA. Although this overhead does not change the overall gate complexity  $\mathcal{O}(N)$  of the HVA implementation, it can still be quite demanding for NISQ devices. Using the Aspen-9 QPU with the native connectivity thus provides a significant advantage in the near term.

There are six parameters for each HVA layer, independent of the system size  $N$ . This makes the optimization of the HVA straightforward to analyze with increasing system size. We can make a rough estimate of how many times the cost function has to be evaluated based on the results on smaller lattices with a similar number of layers. Nonetheless, if the ground state of a larger system exhibits longer-range entanglement, more layers will typically be needed to achieve the same degree of accuracy as in a smaller system. Hence, the required number of layers and the total number of parameters depend on the system size, and usually can only be determined by testing. This makes the prediction nontrivial especially when we consider small lattices in Fig. 1 with noticeable finite-size effects.

Other than HVA, we consider the HEA which is widely used in NISQ applications [13,86,87]. The general principle of the HEA is to construct circuit *Ansätze* using the native gates supported by the QPU, and hence different specific forms of the HEA are constructed targeting different QPUs. For Aspen-9, the native gate set consists of the XY gate [63], the CPHASE and CZ gates [64], as well as the single-qubit gates  $R_z(\theta)$  and  $R_x(k\pi/2)$  with  $k = \pm 1, 2$ . Moreover, the QUIL programming language [83] and its accompanying optimizing compiler QUIL-C [88] on the Rigetti stack [89] admit parametric compilation, which allows for the *Ansatz* to be compiled only once, so that the numerical values of the *Ansatz* parameters are only updated at runtime, without the need of incurring the compilation overhead at every step of the optimization process. This allows for faster execution times and feedback loops between the quantum and classical processors in the hybrid computation.

Roughly speaking, a HEA usually consists of layers of parametrized single-qubit rotations and two-qubit gates to create entanglement between qubits. Adopting this idea with respect to the Aspen-9 native gate set and its connectivity, we consider two *Ansätze*, namely, the HEA-CZ using CZ gates and single-qubit rotation gates, and the HEA-XY using XY gates and single-qubit rotation gates. Here, we focus on the native two-qubit gates (CZ and XY) and slightly relax the native restriction on the single-qubit gates since the single-qubit gates have a much lower error rate than the two-qubit gates. The unitary representation  $U$  of an  $N$ -layer HEA is given by

$$U(\vec{\theta}) = U_{\text{HEA}}(\vec{\theta}_L) \cdots U_{\text{HEA}}(\vec{\theta}_1) U_0(\vec{\theta}_0). \quad (3.3)$$

Here, the layer  $U_0$  preparing each qubit in an arbitrary unentangled state is given by

$$U_0(\vec{\theta}_0) = \prod_i R_{z,i}(\theta_{0,1,i}) R_{x,i}(\theta_{0,0,i}). \quad (3.4)$$

TABLE II. Classical optimization algorithms and the different types of optimizers tested for the VQE algorithm. These include gradient-free, gradient-based, and nonlocal optimizers, and a genetic algorithm. We use the SCIPY [90] implementation for BFGS and dual annealing, a FORTRAN implementation developed by Powell [91,92] for BOBYQA, the PYCMA [93] implementation for CMA-ES, and a PYTHON implementation developed by Gomez-Dans [94] for SPSA.

Optimizer	Gradient-free	Genetic	Local
BFGS	×	×	✓
BOBYQA [91]	✓	×	✓
CMA-ES [95]	✓	✓	×
Dual Annealing [96]	✓	×	×
SPSA [97]	×	×	✓

Note that an arbitrary single-qubit rotation is represented by the gate sequence  $R_Z R_X R_Z$ . The rightmost  $R_Z$  can be omitted here since the qubits are initialized in the zero state and  $R_Z|0\rangle$  just gives an irrelevant global phase. The other layers  $U_{\text{HEA}}$  then entangle the qubits using the two-qubit gates. For HEA-CZ, we have

$$U_{\text{HEA}}(\vec{\theta}_\ell) = \prod_{j,k \in G} R_{Z,j}(\theta_{\ell,4,(j,k)}) R_{Z,k}(\theta_{\ell,3,(j,k)}) \\ \times R_{X,j}(\theta_{\ell,2,(j,k)}) R_{X,k}(\theta_{\ell,1,(j,k)}) CZ_{j,k},$$

where  $G$  represents the hardware-native connectivity. Similarly, for HEA-XY, we have

$$U_{\text{HEA}}(\vec{\theta}_\ell) = \prod_{j,k \in G} R_{Z,j}(\theta_{\ell,6,(j,k)}) R_{Z,k}(\theta_{\ell,5,(j,k)}) \\ \times R_{X,j}(\theta_{\ell,4,(j,k)}) R_{X,k}(\theta_{\ell,3,(j,k)}) \\ \times R_{Z,j}(\theta_{\ell,2,(j,k)}) R_{Z,k}(\theta_{\ell,1,(j,k)}) XY_{j,k}(\theta_{\ell,0,(j,k)}).$$

Once again, the sequence  $R_Z R_X R_Z$  represents an arbitrary single-qubit rotation. For HEA-CZ, since  $R_z$  commutes with CZ, the first  $R_z$  can be combined with the last  $R_z$  in the previous layer.  $U_{\text{HEA}}(\vec{\theta}_\ell)$  thus requires two fewer single-qubit gates for HEA-CZ than for HEA-XY.

In the present case, with the native connectivity being the same as the lattice geometry, the number of two-qubit gates of the HEA scales as  $N$ , which is the same as for the HVA, as expected. On the other hand, the number of parameters per layer also scales linearly with  $N$ , in contrast to the constant scaling of the HVA. The much larger number of parameters makes the HEA more expressive especially when a small number of layers is used. However, this also makes the HEA hard to scale up since optimizing a large number of parameters is challenging for nonconvex cost functions.

### C. Optimizers

The cost function associated with VQE algorithms is in general nonconvex with many local extrema. Similar to optimizing other classical nonconvex cost functions, it is typically hard to decide on an efficient optimizer *a priori* [98]. To make our benchmark less dependent on a specific choice of the optimizer, we will test the list of optimizers shown in Table II. This list consists of a few well-known gradient-free algo-

rithms (BOBYQA, CMA-ES, dual annealing), gradient-based optimizers (SPSA, BFGS), nonlocal optimizers (CMA-ES, dual annealing), and a genetic algorithm (CMA-ES). Even though this is far from a comprehensive list, we can gain insight into how different types of optimizers work for our specific VQE cost function.

We compute the gradients needed for the gradient-based optimizers by finite-difference methods, which require an additional measurement of the cost function using the QPU for each gradient  $\frac{\partial \langle E(\vec{\theta}) \rangle}{\partial \theta_j}$ . This makes the cost of running these gradient-based optimizers larger than for the gradient-free algorithms, making them suboptimal given the limited clock rate of present-day QPUs. More efficient quantum algorithms to evaluate gradients have attracted great interest recently [99–102]. For instance, the parameter-shift rule allows the gradients to be evaluated with an analytic expression without introducing overhead in circuit depth or circuit width [101]. The analytical evaluation will improve the performance of the optimization with noise. However, using the parameter-shift rule for the HVA *Ansatz* requires a generator decomposition [103], which makes the number of required expectation-value measurements scale with the system size  $N$  (compared to constant scaling for finite-difference methods). Other algorithms usually require ancilla qubits and/or extra gate operations, and thus are challenging to implement on NISQ devices. Nonetheless, once QPUs with better gate fidelities and coherence properties are available, further studies will be required to revisit these results and strike a balance between quantum resource requirements and optimization efficiency.

### D. Determining energy gap by VQE

The energy gap between the ground state and the first excited state is a useful quantity in probing the system's phase diagram. VQE algorithms can obtain excited states by modifying the cost function with an additional term penalizing the overlap between the trial state and the lower-energy state obtained previously [76]. In particular, to determine the energy gap, we use the following cost function to determine the next excited state beyond the ground state:

$$C(\vec{\theta}) = \langle \psi(\vec{\theta}) | H | \psi(\vec{\theta}) \rangle + p |\langle \psi(\vec{\theta}_{\sigma,0}) | \psi(\vec{\theta}) \rangle|^2, \quad (3.5)$$

where  $p$  is the penalty weight and  $\vec{\theta}_{\sigma,0}$  are the optimal parameters corresponding to the ground state previously determined by VQE. The overlap can be experimentally determined on QPUs by executing the circuit corresponding to  $U^\dagger(\vec{\theta}_{\sigma,0})U(\vec{\theta})|0\rangle$  and then measuring the population of the zero state  $|0\rangle$ . The penalty weight  $p$  is chosen to be larger than any relevant energy gaps, and we choose it to be 0.3 in our numerical experiment.

### E. Experiment and noise mitigation

We run preoptimized HVA circuits with one layer (HVA-1) on four- and eight-qubit sublattices of Rigetti's Aspen-9 QPU. The variational optimization of the HVA is first performed on a classical simulator, and the circuits are then run on the QPU for the optimal choice of parameters. While this approach circumvents the hybrid quantum/classical computation, it nevertheless serves as a benchmark for how well the

quantum processor performs in proof-of-concept experiments. For comparison, recent state of the art results for many body Hamiltonian simulations have opted for either purely classical optimization instead of a VQE step [104] or running their algorithms on smaller numbers of qubits [23,105].

We perform digital zero-noise extrapolation (ZNE) [106–108] by first exaggerating the errors in the circuit by replacing a subset of CZs by an odd multiple of them. We choose such subsets to increase the total CZ count in the circuit by a factor of  $\lambda$ . For each scale factor  $\lambda$ , we run 25–100 (four qubits) and 100 (eight qubits) different circuit implementations for 1000 shots each. For the four-qubit case, we pick  $\lambda = 1, 1.3, 1.6, 2, 2.3$ . The bare HVA-1 circuit, with six CZs, corresponds to  $\lambda = 1.0$ . To obtain  $\lambda > 1.0$ , we note that the exponentials of each of the two-local terms ( $XX, YY$ , and  $ZZ$ ) are compiled to two CZs, along with a few one-qubit gates. For  $\lambda = 1.3$ , we replace the first of the two CZs compiling the exponential of the  $XX$  term with three CZs, then again for the  $YY$  and  $ZZ$  terms, then take the average of all three types of circuits to get an estimated expectation value of the cost Hamiltonian. For  $\lambda = 1.6$ , we replace both the CZs with three CZs each, once for each of the three two-local terms. For  $\lambda = 2.0$ , we replace both CZs in the  $XX$  term, and the first of the two CZs in the  $YY$  term with three CZs each for one experiment, and for another we replace both CZs in the  $ZZ$  term and the second of the two CZs in the  $YY$  term with three CZs each. Finally, for  $\lambda = 2.3$ , we perform three experiments, in which we replace the CZs in the  $XX$  and  $YY$ ,  $XX$  and  $ZZ$ , or  $YY$  and  $ZZ$  exponentials with three CZs each.

For the eight-qubit case, we choose  $\lambda = 1, 1.5, 2, 3, 5$ . For  $\lambda = 1$ , we have a total of 16 CZs, two for each of the edges. For  $\lambda = 1.5$ , we run 25 different circuits for each of the scenarios where we replace (a) every CZ in the  $XX$  links with three CZs, (b) every CZ in the  $YY$  links with three CZs, (c) every CZ in two of the four  $ZZ$  links with three CZs, and (d) every CZ in the remaining two of the four  $ZZ$  links with three CZs. For  $\lambda = 2$ , we run 50 different circuits for each of the scenarios where we replace (i) every CZ in the  $XX$  and  $YY$  links with three CZs and (ii) every CZ in the  $ZZ$  links with three CZs. For  $\lambda = 3$  ( $\lambda = 5$ ), we simply take 100 different

circuits where we replace every CZ in the circuit with three (five) CZs.

For each of the circuit implementations in each scale factor family, we estimate the cost function (2.3) by measuring the expectation values of all Pauli strings given in (2.2). Each of these expectation values associated with  $N$  qubits is determined experimentally by applying suitable single-qubit gates to transform any Pauli  $X$  or Pauli  $Y$  to Pauli  $Z$  followed by a projective measurement over the computational basis. The measured probability distributions are subjected to readout errors which can be mitigated as follows. The readout errors can be modeled by the classical Markovian process [109] described by the equation

$$p_{\text{read}} = R p_{\text{prepared}}. \tag{3.6}$$

Here,  $R$  is the  $2^N \times 2^N$  confusion matrix that relates the probability distribution  $p_{\text{prepared}}$  of the prepared state without readout error and the probability distribution  $p_{\text{read}}$  experimentally measured with readout error. Each element of  $R$  can be interpreted as the conditional probability of measuring the bitstring  $x_{\text{read}}$  given that the bitstring  $x_{\text{prepared}}$  is prepared. In our experiment,  $R$  is estimated by preparing a bitstring, then computing the fraction of each of  $2^N$  bitstrings from 10 000 shots to estimate the conditional probabilities, then repeating this procedure for all  $2^N$  bitstrings. By inverting  $R$ , we obtain a readout error mitigated estimate of the probability of outcomes of all bitstrings using

$$p_{\text{prepared}} = R^{-1} p_{\text{read}}. \tag{3.7}$$

In turn, these lead to readout-error-mitigated estimates of the cost function (2.3) in each of the 100 circuits for any given family associated with a scale factor. We then average these expectation values over all 100 circuits to produce a single value representing the family of circuits associated with a given scale factor. These representative values are then plotted against the scale factors, and a polynomial fit is found to extrapolate to the  $\lambda = 0$  (zero noise) limit to obtain an estimate of the cost function mitigated for both gate and readout errors.

For the eight-qubit case, we also perform randomized compilation [110] to twirl the physical error channels into

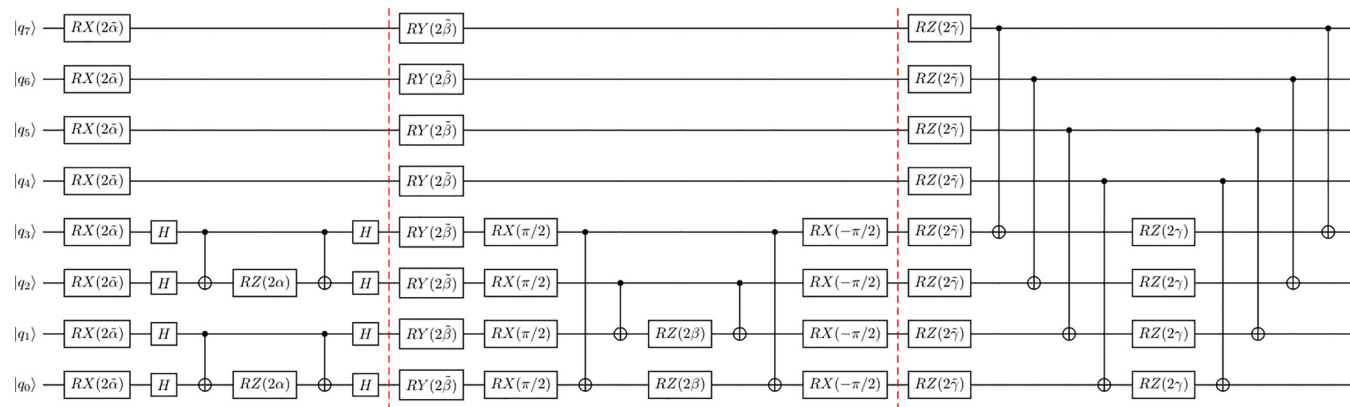


FIG. 3. HVA with one layer on eight qubits. The Hamiltonian variational Ansatz (HVA) with one layer on eight qubits splits into commuting blocks. The first block corresponds to the operation  $e^{-i\tilde{\alpha} \sum_q X_q} e^{-i\tilde{\alpha} \sum_{(i,j) \in X \text{ links}} X_i X_j}$ , the second to  $e^{-i\tilde{\beta} \sum_q Y_q} e^{-i\tilde{\beta} \sum_{(i,j) \in Y \text{ links}} Y_i Y_j}$ , and the third to  $e^{-i\tilde{\gamma} \sum_q Z_q} e^{-i\tilde{\gamma} \sum_{(i,j) \in Z \text{ links}} Z_i Z_j}$ . For the circuit shown here, we used  $X$  links =  $\{(q_0, q_1), (q_2, q_3)\}$ ,  $Y$  links =  $\{(q_0, q_3), (q_1, q_2)\}$ , and  $Z$  links =  $\{(q_0, q_4), (q_1, q_5), (q_2, q_6), (q_3, q_7)\}$ . To execute this circuit, we map qubits  $(q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7)$  to qubits  $(12, 25, 26, 11, 13, 24, 27, 10)$ , respectively, on Aspen-9 shown in Fig. 1(b).

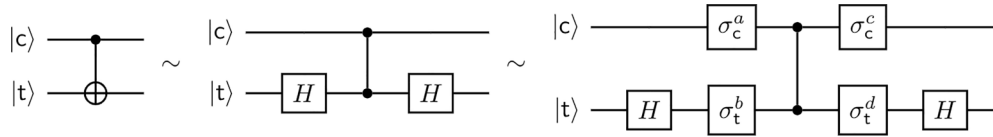


FIG. 4. Twirled CZ used to compile a CNOT. Every use of the CNOT gate in Fig. 3 is compiled to a native CZ with some single-qubit gates. Assuming that the major source of noise lies in the two-qubit gates, we apply random Pauli operators before every use of the CZ, and then apply another pair of Pauli operators in order to twirl the noise channel associated with the CZ into a stochastic Pauli channel. The relationship between the indices  $a, b, c, d \in \{0, 1, 2, 3\}$  (denoting respectively the Pauli operators  $\{I, X, Y, Z\}$ ) is given in Eq. (3.8). The subscripts  $c$  and  $t$  on the Pauli matrices refer to “control” and “target” qubits. All three circuits depicted above are logically equivalent, but generally lead to different noise characteristics.

stochastic errors to improve the result. In the HVA, the exponential of every two-body term can be compiled using two native CZs. Before each CZ gate, we apply random Pauli operators  $\sigma_c^a \sigma_t^b$ , where  $a, b = 0, \dots, 3$  and the subscripts  $c$  and  $t$  denote the control and target qubits, respectively. After applying the CZ unitary  $\Lambda = |0\rangle\langle 0|_c \otimes \mathbb{I}_t + |1\rangle\langle 1|_c \otimes Z_t$  [111], we apply Pauli operators  $\sigma_c^c \sigma_t^d$  such that  $\sigma_c^c \sigma_t^d = \Lambda \sigma_c^a \sigma_t^b \Lambda^\dagger$ . Following Ref. [106], this is given by

$$\begin{aligned} c &= a + b(3 - b)(3 - 2a)/2, \\ d &= b + a(3 - a)(3 - 2b)/2. \end{aligned} \tag{3.8}$$

This noise tailoring makes the errors more well behaved and better suited for error mitigation techniques such as ZNE [106,112–114]. A schematic circuit for the HVA with one layer on an eight-qubit sublattice is shown in Fig. 3. The twirling operation is depicted in Fig. 4.

For twirled experiments we take point estimates of the expectation values for many circuit implementations for each scale factor and feed them into a Bayesian linear regression model implemented in TURING.JL [115] with Gaussian priors for slope and intercept and a truncated-at-zero Gaussian prior for the variance, and we run Markov chain Monte Carlo (MCMC) until well converged. Hyperparameters of the priors were all set to mean zero and standard deviation 1000 (i.e., no prior information or preference over the reasonable range of

the parameters), but these hyperparameters had negligible impact on the posterior distribution so long as they were not set to effectively exclude the region of the observed data. MCMC was run with a no U-turn sampler [116] for  $10^4$  iterations, we exclude the first 3000 to exclude any warmup period, and for all runs we had an effective remaining sample size  $> 1000$ . For untwirled experiments, we follow the same basic procedure but incorporate shot noise for each circuit implementation by performing a Bayesian bootstrap [117] over the shot data to produce an estimate and uncertainty for the expectation value from each circuit implementation. We then feed those estimates and uncertainties into the Bayesian linear regression defined above.

## IV. RESULTS

### A. State vector simulations

We test the optimizers listed in Table II using the cost function (2.3) associated with the four-layer HVA for the eight-qubit and 16-qubit lattices shown in Fig. 1 and the parameter choice (GL +  $h$ ) in Table I. This cost function has 24 parameters to be optimized. The results using a state vector simulator are summarized in Table III.

BFGS and BOBYQA are susceptible to local extrema and their performance highly depends on the initial choice of the parameters. We employ a multiple-initial-condition strategy

TABLE III. Optimizer performance with state vector simulator. We test the optimizers with eight qubits and 16 qubits using a four-layer HVA and the (GL+ $h$ ) parameter set. We employ a multiple-initial-values strategy to BFGS and BOBYQA to avoid the local optimizers being trapped at a bad local extreme. The lowest energy is reported for the multiple-initial-values strategy. While the genetic algorithm (CMA-ES) is less susceptible to local extrema, a multiple-initial-values strategy with a smaller iteration cutoff also improves its performance. SPSA and dual annealing converge slowly and reach the maximum iteration (cutoff). Note that the error in energy is the difference between the optimized energy and the ground-state energy computed by ED.

Qubits	Optimizer	Error in energy	Cost function evaluations
8	BFGS, 501 initial values	0.00094	Mean, 5352; max, 10865
16	BFGS, 501 initial values	0.02672	Mean, 5569; max, 10186
8	BOBYQA, 501 initial values	0.00045	Mean, 1099; max, 1920
16	BOBYQA, 501 initial values	0.01744	Mean, 1255; max, 3049
8	CMA-ES	0.00015	43290
16	CMA-ES	0.04036	49335
8	CMA-ES, 80 initial values	0.00005	Mean, 20528; max, 83590
16	CMA-ES, 80 initial values	0.01327	Mean, 27504; max, 73138
8	Dual annealing	0.00252	100000 (cutoff)
16	Dual annealing	0.04634	100106 (cutoff)
8	SPSA	0.04500	100000 (cutoff)
16	SPSA	0.08145	100000 (cutoff)

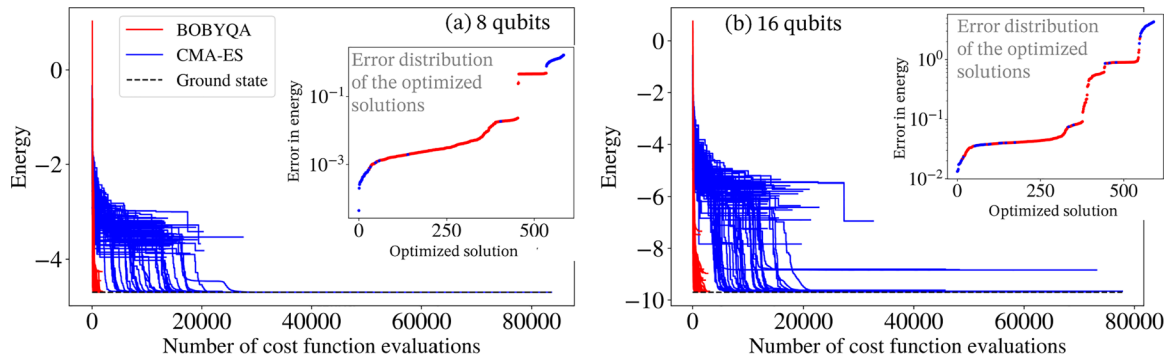


FIG. 5. Noiseless optimization with BOBYQA and CMA-ES. The cost function associated with the four-layer HVA and the parameter set  $(GL + h)$  is optimized by BOBYQA (red) with 501 random initial values and CMA-ES (blue) with 80 random initial values. For both (a) eight qubits and (b) 16 qubits, BOBYQA converges faster than CMA-ES. The error distribution of the optimized solutions (see insets showing the error in energy associated with each solution optimized by BOBYQA and CMA-ES) shows that the solutions of CMA-ES give overall better results but are also less consistent. The mixed usage of the two optimizers provides a complementary cross-check of their results.

and randomly pick 501 initial parameter values to start with. In general, the number of initial values required for a good result will increase with the number of parameters of the cost function. BOBYQA gives a smaller error in energy compared to BFGS in both eight-qubit and 16-qubit tests, and BOBYQA also converges much faster than BFGS, which requires the costly computation of gradients by finite-difference methods. The other local optimizer, SPSA, is less susceptible to local extrema due to its stochastic nature. However, its convergence is slow and will require many iterations in order to achieve good accuracy. Among the local optimizers, BOBYQA with multiple initial values gives the best performance in terms of both the accuracy of the result and the speed of convergence.

CMA-ES and dual annealing are nonlocal optimizers that are able to escape from local extrema. Table III shows that both optimizers converge toward a minimum with energy close to the exact value even starting from just one initial value. CMA-ES consistently converges faster among the two. Obtaining the best performance from CMA-ES requires fine tuning of the optimizer metaparameters, which depends on the neighborhood of the initial value and is challenging to determine *a priori*. To avoid becoming overly sensitive to the choice of metaparameters, we also employ a multiple-initial-value strategy with 80 random initial values. For our VQE application, the CMA-ES optimization with multiple initial values consistently shows improvement over that with a single initial value.

To benchmark different *Ansätze*, we employ both BOBYQA with 501 random initial values and CMA-ES with 80 random initial values and report the better result from the two methods. Different *Ansätze* (and different layer numbers) result in vastly different cost function landscapes and numbers of parameters. A mixed usage of different optimization algorithms, which is widely used in black-box optimization, helps us better navigate through the vast varieties of cost functions. In Fig. 5, we study the noiseless optimization for each optimizer with different initial values. For both eight-qubit and 16-qubit cases, CMA-ES needs more cost function evaluations to converge, as expected for a nonlocal optimizer. As indicated by the error distribution of the optimized solutions shown in the insets of Fig. 5, some optimization runs with CMA-ES give the best result but others can be far off. The runs

also converge to different optimized parameters with the two different optimizers and also different initial values, revealing a complex energy landscape with many local minima. This makes the inclusion of the BOBYQA results helpful, since it allows us to crosscheck the optimizations. A similar conclusion can be made by resampling the distribution shown in the insets of Fig. 5 to investigate the error as a function of the number of initial values of optimization. As shown in Fig. 6, the CMA-ES optimizer gives a better result than BOBYQA after a sufficient number of initial values are used. Hence, BOBYQA serves as a useful cross-check for an insufficient number of initial values. We expect that VQE optimization carried out on QPUs will also benefit from this approach of

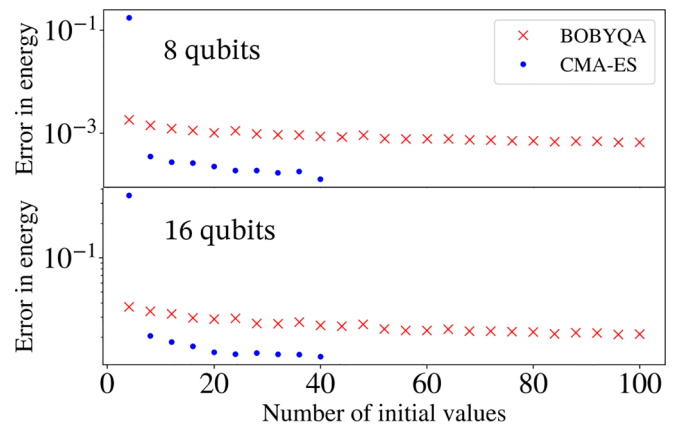


FIG. 6. Error as a function of the number of initial parameter values for the BOBYQA and CMA-ES optimizers. The error decreases with increasing number of initial values. CMA-ES converges faster compared to BOBYQA at the price of a much higher number of cost function evaluations per initial value. The performance of CMA-ES exceeds that of BOBYQA after a threshold number of initial values. The performance of both methods improves gradually beyond that threshold. We resample the error distribution shown in Fig. 5 to obtain this result. Each data point in this plot is obtained by averaging the minimum energies of 50 different random choices of initial values. The statistical error (not shown) is negligible except for the leftmost CMA-ES data point, and this is consistent with the fact that CMA-ES is unreliable with a small number of initial values.



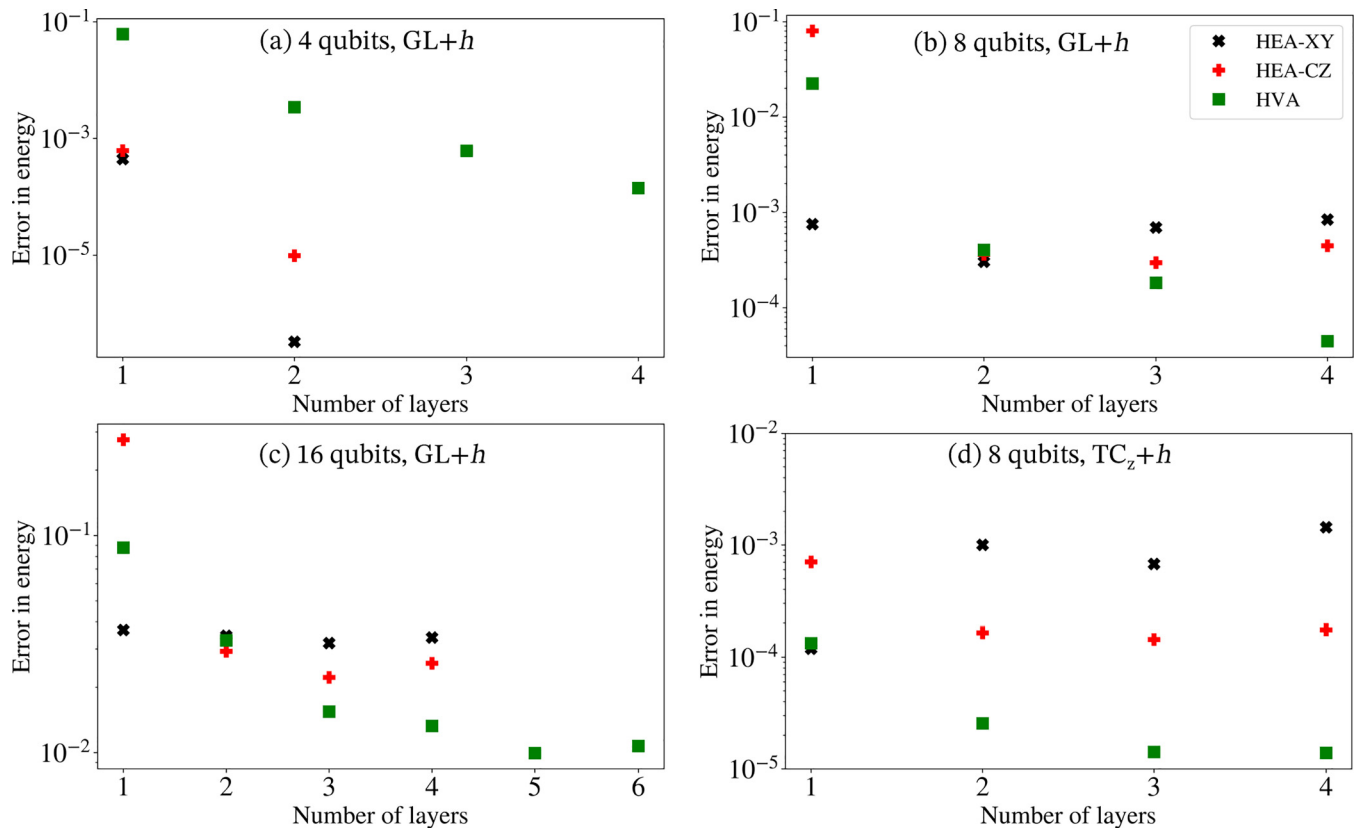


FIG. 7. Ansatz benchmark by state vector simulation. The three different *Ansätze* are optimized with respect to three different setups, (a) four-qubit lattice with  $(GL+h)$ , (b) eight-qubit lattice with  $(GL+h)$ , (c) 16-qubit lattice with  $(GL+h)$ , and (d) eight-qubit lattice with  $(TC_z+h)$ . The HEA-CZ (red pluses) and the HEA-XY (black crosses) show a bigger error in energy compared to that of the HVA (green squares) except for the four-qubit case. This performance difference is mainly due to the difficulty in optimizing the two HEAs with a large number of parameters, as suggested by the minimal improvement of the error when increasing the number of layers of the HEAs.

mixed optimizers with multiple initial values, especially when the QPUs' clock rate becomes sufficiently high to support a large number of cost function evaluations.

We apply this mixed optimization approach to optimize the three different *Ansätze* (HVA, HEA-CZ, and HEA-XY) with different numbers of layers using a state vector simulator. To make our observation more representative, we run the optimization test with three different setups, namely, four-qubit  $(GL+h)$ , eight-qubit  $(GL+h)$ , 16-qubit  $(GL+h)$ , and eight-qubit  $(TC_z+h)$  (see Table I for the parameter values). Except for the simple four-qubit case, the results in Fig. 7 show that the HVA in general outperforms the other two HEAs when a deeper *Ansatz* is used to reduce the error in energy. The main reason for the HEAs' worse performance is due to the extreme difficulty in the optimization, which is directly related to the barren plateau phenomenon [28]. The number of parameters increases rapidly for HEAs; for example, the four-layer HEA-XY for 16 qubits has 536 parameters. This makes it excessively challenging to scale up the VQE with HEAs to larger circuit depths and system sizes. This is also revealed by the fact that two HEAs show nearly no improvement with increasing number of layers in Fig. 7. Hence, the VQE using HVA will be a more promising approach toward quantum advantage for the square-octagon-lattice Kitaev model.

The optimization approach and the HVA *Ansatz* can be used to probe the phase diagram of the model. We demon-

strate this potential by determining the single-spin expectation values and static spin-spin correlators in Fig. 8 for an eight-qubit lattice with periodic boundary conditions as shown in Fig. 1(f). We choose this lattice structure and boundary condition so that the system's behavior is less susceptible to boundary effects. Note that the lattice contains four  $x$ ,  $y$ , and  $z$  bonds each. To adapt to this different setup, we have to increase the number of layers of the HVA *Ansatz* to 8 to achieve reasonable accuracy. To characterize the system's properties as a function of Hamiltonian parameters, we measure the averages of the single-site expectation values  $\langle X \rangle$ ,  $\langle Y \rangle$ , and  $\langle Z \rangle$  over all sites, as well as the averages of the correlators  $C_{\alpha\alpha} = \langle \alpha_j \alpha_k \rangle - \langle \alpha_j \rangle \langle \alpha_k \rangle$  ( $\alpha = X, Y, Z$ ) over all pairs of sites for which  $J_\alpha \neq 0$ . While this small system does not exhibit the exact same phase transitions shown in Fig. 2 for large lattices in the thermodynamic limit, the same set of measurements can be used to study the phase diagram of larger systems. Increasing  $J_\perp$  with a fixed  $h_{[111]}$ , the eight-qubit lattice shows an increasing average value of  $\langle Z \rangle$  while  $\langle X \rangle$  and  $\langle Y \rangle$  remain close to zero. Note that the asymmetry between  $\langle Z \rangle$  and  $\langle X \rangle$  or  $\langle Y \rangle$  at  $J_\perp/J_z = 1$  is due to the asymmetry in the lattice geometry. Meanwhile the average correlators  $C_{\alpha\alpha}$  all increase gradually from zero as  $J_\perp$  grows. This behavior is captured by the VQE optimization results (data points) as compared to the simulation using ED (dashed lines). One can observe a small deviation between the VQE and ED results for  $C_{ZZ}$

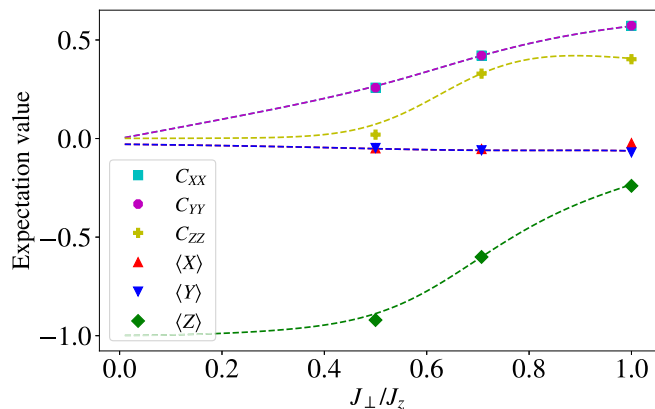


FIG. 8. Expectation values determined by VQE for an eight-qubit periodic lattice. The VQE results (data points) are consistent with those obtained from ED (dashed lines). Here, the single-qubit expectation values  $\langle X \rangle$ ,  $\langle Y \rangle$ , and  $\langle Z \rangle$  are averaged over all qubits, and the correlators  $C_{\alpha\alpha} = \langle \alpha_j \alpha_k \rangle - \langle \alpha_j \rangle \langle \alpha_k \rangle$  ( $\alpha = X, Y, Z$ ) are averaged over the qubit pairs for which  $J_{\alpha} \neq 0$ . The VQE results are determined using an eight-layer HVA *Ansatz* with the optimization using 480 initial values for CMA-ES and 3001 initial values for BOBYQA, and the expectation values are calculated using the optimized parameters corresponding to the lowest energy among all optimized solutions. We use the parameters  $J_z = 1$  and  $h_x = h_y = h_z = 0.05/\sqrt{3}$ .

at  $J_{\perp}/J_z = 0.5$ . This could be improved by increasing the number of layers used with the tradeoff of higher resource requirements in the optimization.

The phase diagram can also be probed by searching for parameter values where the energy gap between the ground and first excited states closes. By adding the overlap with the ground state to the VQE cost function as discussed in Sec. III D, we determined the excited-state energies using the same HVA *Ansatz* and optimization approach. For the eight-qubit periodic lattice, we determine the optimal parameters  $\bar{\theta}_{o,0}$  corresponding to the ground states for different values of  $J_{\perp}$  using the eight-layer HVA *Ansatz* as discussed in the previous paragraph. For excited states, we optimize the modified cost function in (3.5) with the same eight-layer HVA *Ansatz*. For the optimization, we use CMA-ES with 960 ( $J_{\perp} = 0.5, 1/\sqrt{2}$ ) and 480 ( $J_{\perp} = 1$ ) random initial values, and BOBYQA with 6001 ( $J_{\perp} = 0.5, 1/\sqrt{2}$ ) and 3001 ( $J_{\perp} = 1$ ) random initial values. We double the number of initial values for  $J_{\perp} = 0.5, 1/\sqrt{2}$  as an attempt to improve the accuracy of the excited-state energies determined by the optimization.

The ground- and excited-state energies determined by VQE are shown in Fig. 9. To better visualize the results, we offset the energy shown in the figure by the ground-state energy  $E_{0,ED}$  determined using ED. We also plot the ED result ( $E_{j,ED}$  where  $j = 0, 1, \dots, 5$ ) for up to the fifth excited state for comparison. As expected, the VQE results (red crosses) for the ground-state energy are consistent with the ED result (red dashed line) with a small energy error. On the other hand, the excited-state energies (blue pluses) determined by VQE are closest to those of the third excited state instead of the nearly degenerate first and second excited states (except for  $J_{\perp} = 1$  where the first three excited states are close to each other).

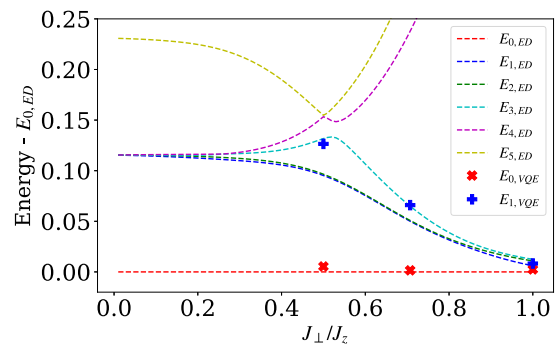


FIG. 9. Excited-state and ground-state energies determined by VQE for an eight-qubit periodic lattice. The ground-state energies  $E_{0,VQE}$  (red crosses) and the first excited-state energies  $E_{1,VQE}$  (blue pluses) determined by VQE are compared to the low-lying energy spectrum (dashed lines) determined by ED. VQE correctly predicts the qualitative result of the closing of the energy gap near  $J_{\perp} = J_z$ . However, the excited-state energies determined by VQE are closest to the third excited state instead of the nearly degenerate first and second excited states. We use the parameters  $J_z = 1$  and  $h_x = h_y = h_z = 0.05/\sqrt{3}$ .

This means that while the VQE result still qualitatively predicts the closing of the energy gap, it does not give the correct energy gap between the ground state and the first excited state. We tried multiple values of  $p$  and observed a similar difficulty in getting to the lower-energy excited states. Future study of alternative *Ansätze* better tailored to the excited states will be needed to improve the accuracy.

## B. Simulations with shot noise

It is crucial to take the shot noise into consideration when running variational algorithms on QPUs. If too many measurement shots are requested, the long runtime may be beyond the QPU capacity. If too few shots are requested, the stronger noise may induce undesirable barren plateaus to the cost function landscape [118]. Investigating optimization strategies that tolerate a small number of shots is important for practical applications of the variational algorithms [77].

In this subsection, we evaluate the energy in the simulations with a finite number of shots and then test the optimizers listed in Table II once again with the noisy cost function. This allows us to examine the robustness of our optimization approach in the presence of the shot noise. We run the test with 8000 shots using the cost function associated with the four-layer HVA for the eight-qubit lattice and the parameter choice (GL +  $h$ ) to facilitate the comparison with the noiseless results. The results are summarized in Table IV. To make the effect of the shot noise clear, we determine the energy  $E_{\text{noisy}}$  measured with shot noise and its noiseless counterpart  $E_{\text{noiseless}}$  obtained using a state vector simulator for the optimized parameters obtained by minimizing the noisy cost function. The error ( $E_{\text{noiseless}} - E_0$ ) from the exact ground-state energy  $E_0$  shows the performance of the VQE since this quantity is directly related to how close the optimized state is to the ground state. The measured deviation ( $E_{\text{noisy}} - E_0$ ) gives us additional information regarding the performance of the optimizers in the presence of the shot noise.

TABLE IV. Optimizer performance with shot noise for eight qubits. We test the optimizers with 8000 measurement shots and eight qubits using a four-layer HVA and the (GL+h) parameter set similar to the noiseless case. A modified version of BOBYQA, BOBYQA-noisy, is introduced to better handle the noisy cost function. Using the optimized parameters obtained by minimizing the noisy cost function, we determine the energy  $E_{\text{noisy}}$  measured with shot noise and its noiseless counterpart  $E_{\text{noiseless}}$  obtained using a noiseless state vector simulator. The standard deviation of 100 energy evaluations with 8000 shots is about 0.02 for the optimized parameters obtained by these optimizers. Note that  $E_{\text{noisy}}$  can be lower than  $E_0$  due to the shot noise.

Optimizer	$E_{\text{noiseless}} - E_0$	$E_{\text{noisy}} - E_0$	Cost function evaluations
BFGS, 501 initial values	0.45069	0.42052	Mean, 747; max, 1994
BOBYQA, 501 initial values	0.27485	0.21843	Mean, 471; max, 610
BOBYQA-noisy, 501 initial values	0.07989	-0.00453	Mean, 3532; max, 4004
CMA-ES	0.02416	-0.06462	37570
CMA-ES, 80 initial values	0.01610	-0.07125	Mean, 21042; max, 52000
Dual annealing	0.04534	-0.01631	60101
SPSA	0.00612	0.00879	100000 (cutoff)

BFGS shows significantly worse performance compared to the noiseless case. The errors introduced by the shot noise are amplified by the small finite step size used by finite-difference methods to compute gradients. The amplified errors make the BFGS much more likely to be trapped. Using analytical methods to determine the gradients directly through measurements [99,102] could avoid the amplification and would be important when implementing BFGS (and similar gradient-based methods) on QPUs.

The performance of BOBYQA is also worse than that in the noiseless case. This observation is consistent with previous results indicating that BOBYQA typically requires further modifications to optimize the noisy cost function [119]. Here, we adopt one of the simplest modifications to increase the number of interpolation points from  $(2n + 1)$  to  $(n + 1)(n + 2)/2$  where  $n$  is the number of parameters. The modified version, BOBYQA-noisy, gives a much improved result as expected. Note that the negative measured deviation suggests that the measured energy associated with the optimized parameters is lower than the exact ground-state energy, which is only possible due to the shot noise. This means that the

optimization has converged within the limits set by shot noise. (The standard deviation of 100 energy evaluations with 8000 shots is  $\approx 0.02$  for all optimized parameters.)

CMA-ES and dual annealing, together with BOBYQA-noisy, have a negative measured deviation for the optimized parameters. (The measured energy is lower than the ground-state energy due to the shot noise.) Among the three optimizers, CMA-ES has the best performance ranked by the error determined by the state vector simulator. Figure 10 suggests that this superior performance is related to the ability to converge to a better solution by CMA-ES than that by BOBYQA-noisy with the cost of a larger number of cost function evaluations. While CMA-ES may still converge to a poor solution for some initial values, a significant portion of the optimization trajectories converge to solutions close to the ground state. This also suggests the mixed optimization approach is still useful in this noisy case with the BOBYQA-noisy serving as a cross-check of the CMA-ES result.

SPSA shows the best performance among the optimizers being tested. In particular, the error (0.006 12) associated with its optimized parameters is even smaller than the noiseless optimization performed by SPSA with an error of 0.045 00. This result should not be overinterpreted since these are just two specific optimization trajectories. Nonetheless, this result suggests that the performance of SPSA may be less

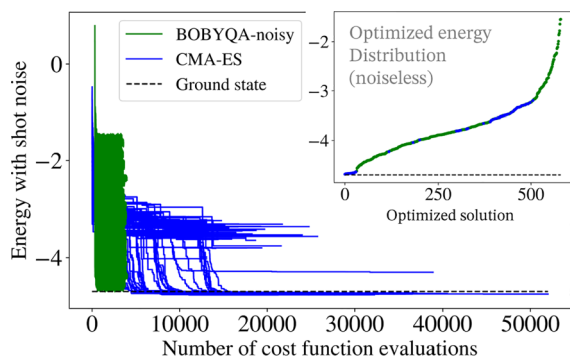


FIG. 10. Eight-qubit optimization with shot noise using BOBYQA-noisy and CMA-ES. The cost function associated with the four-layer HVA and the parameter set (GL + h) is optimized by BOBYQA-noisy (green) with 501 random initial values and CMA-ES (blue) with 80 random initial values. Similar to the noiseless optimization, BOBYQA-noisy converges faster than CMA-ES but the solutions of CMA-ES give overall better results as shown by the optimized energy distribution (see insets).

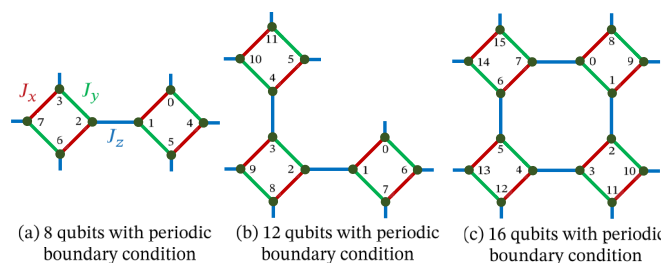


FIG. 11. Lattice geometry with periodic boundary conditions. The  $J_x$  (red),  $J_y$  (green), and  $J_z$  (blue) couplings are assigned to each linkage. The qubits at the edge are connected by  $J_z$  coupling to the one on the other side of the lattice. In particular, for the periodic boundary conditions, the following qubit pairs are  $J_z$  coupled: (a) (0, 5), (3, 6), and (4, 7); (b) (0, 7), (5, 10), (6, 9), and (8, 11); (c) (8, 11), (9, 14), (10, 13), and (12, 15).

TABLE V. Optimizer performance with shot noise using SPSA. We test the optimizers with various measurement shots and qubits using a one-layer HVA and the  $(TC_z)$  parameter set. We use the same definition of the error (noiseless) and measured deviation as in Table IV. The standard deviation of 100 noisy energy evaluations is evaluated at the optimal parameters determined by the optimization. All optimizations have a cutoff of cost function evaluation at 10000.

Qubits	Shots	Error (noiseless)	Measured deviation	Standard deviation of 100 noisy evaluations
8	8192	0.06977	-0.00246	0.00440
12	4096	0.00191	0.00406	0.00756
16	2048	0.14701	-0.01185	0.01103

susceptible to shot noise, and it would be useful to include SPSA to the mixed optimization approach together with other optimizers for VQE with shot noise.

We further explore the performance of SPSA with various magnitudes of shot noise and lattices shown in Fig. 11 with periodic boundary conditions, and SPSA converges within the shot noise limit. QPU experiments are more difficult to be carried out with these lattices since the overhead of SWAP gates to simulate the periodic boundary conditions will be substantial for the existing QPU’s connectivity. Nonetheless, boundary effects on these lattices are less significant compared to those of the lattices with open boundary conditions. It is thus interesting to investigate numerically how the SPSA performs for these lattices. We summarize the results in Table V associated with one-layer HVA and the  $TC_z$  parameter set for the eight-qubit, 12-qubit, and 16-qubit lattices shown in Fig. 11. We decrease the number of shots for larger lattices since sampling

is computationally expensive for large lattices. Similarly, to limit the required computational resources, we fix the cutoff of the number of cost function evaluations to be 10 000, which is one tenth of that used in the main text. In general, the SPSA optimizer converges to a good solution within the limits set by shot noise. The error of the optimized solution determined using a noiseless state vector simulator spans a much wider range of values for different setups. Further studies with more systematic choices of the number of shots and other lattice and parameter setups will be useful to understand the spread of the errors while SPSA shows a similar level of convergence.

### C. QPU experiment

The experimental method has been described in Sec. III E. As discussed there, we perform a proof-of-concept experiment using the HVA with one layer on four-qubit and eight-qubit sublattices of Aspen-9 consisting of qubits (10, 11, 12, 13, 24, 25, 26, 27); see Fig. 1(b). A schematic HVA-1 circuit for eight qubits is depicted in Fig. 3; in order to execute this on Aspen-9, we identify qubits  $(q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7)$  with qubits (12, 25, 26, 11, 13, 24, 27, 10), respectively, on Aspen-9. To perform digital ZNE, we increase the total CZ count in the circuit by various scale factors and the expectation values measured at each scale factor are mitigated for readout errors. To improve the results for the eight-qubit case, we twirl the corresponding circuits with random Pauli operators to tailor the noise into a stochastic channel, which may allow ZNE to work better.

Figure 12 plots the Hamiltonian expectation values in the classically optimized variational state for various scale factors, and the zero-noise limit is extrapolated via Bayesian linear regression. (For the eight-qubit case, the final two scale

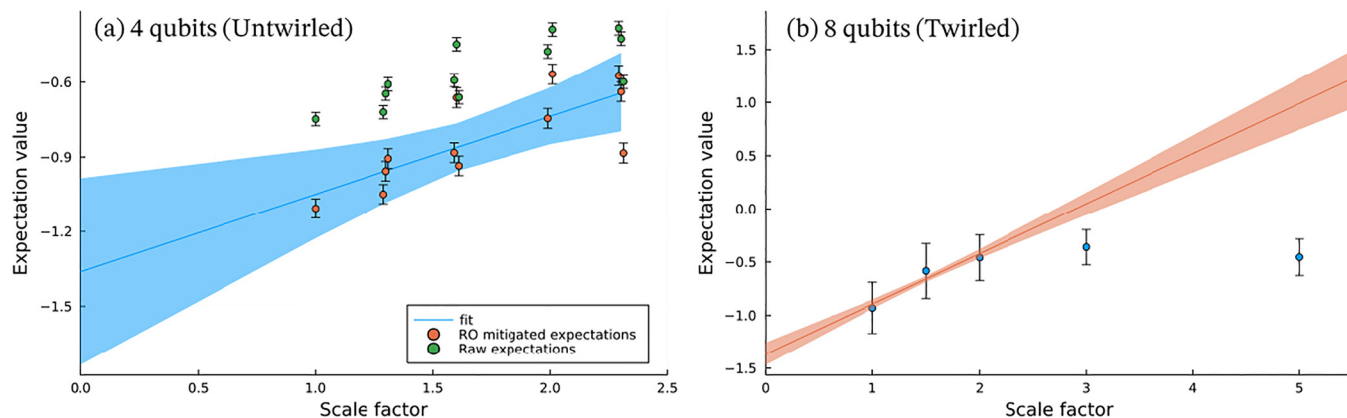


FIG. 12. QPU experiments of optimized HVA circuits (twirled and untwirled) on Aspen-9 with digital ZNE and readout error mitigation. The  $x$  axis represents the scale by which the total CZ count in the circuit is increased. QPU data are plotted as the mean and standard deviation of the distribution over expectation values for circuit implementations. A line of best fit is obtained via Bayesian linear regression, and we plot the expectation value of the posterior distribution and a 95% confidence ribbon. The Hamiltonian expectation value, plotted on the  $y$  axis, is then extrapolated from the regression to the  $x = 0$  limit to obtain a fitted noiseless estimate with the depicted uncertainty. (a) The four-qubit experiment for various ZNE scale factors uses the HVA-1 Ansatz with no (Pauli) twirling. The results extrapolate to roughly  $-1.3623 (\pm 0.1896)$ , which is within roughly one standard deviation of the true eigenvalue of  $-1.583$ . There are one to three circuits for each scale factor (plotted with jitter on the  $x$  axis for better visibility) for raw and mitigated expectations, representing several different ways to implement the scaling factor. (b) For the eight-qubit experiment, the QPU data points corresponding to scale factors of 3 and 5 are dominated by noise and are not used for ZNE. The results extrapolate to  $\approx -1.36$  (with a standard deviation of about 0.054) which is deviated from the best value of about  $-4.23$  that the one-layer HVA can produce.

factors are omitted from the extrapolation because the noise in the system overwhelms any meaningful patterns one could exploit in the extrapolation scheme beyond a scale factor of 2.) The four-qubit experiment gives the extrapolated energy to be  $-1.3623 \pm 0.1896$  which is within one standard deviation of the one-layer-HVA best value  $-1.5217$  and, roughly speaking, the true ground-state energy  $-1.5831$ . The eight-qubit circuit involves more two-qubit gates and a larger infidelity is expected. Anticipating the larger error to be mitigated, we perform randomized compilation or Pauli twirling to improve the performance of ZNE as explained in Sec. III E. However, even with the additional twirling, we find the extrapolated value for the expectation value to be about  $-1.36$  (with a standard deviation of about 0.054), which is far from the best value of about  $-4.23$  that the one-layer HVA can produce in noiseless simulation. This shows that the execution of a one-layer HVA circuit for eight qubits, which contains  $16\lambda$  CZ gates for scale factor  $\lambda \geq 1$ , is close to the QPU coherence time. We note that two-qubit gates were executed sequentially in this paper, and parallel execution may lead to an improvement of the results. The relatively large discrepancy between the optimal value obtained on the QPU versus the exact energy suggests the presence of systematic shifts in the gate parameters when they are executed on the QPU. This may be mitigated when performing the optimization entirely on the QPU. Although this is a proof-of-concept experiment, it motivates the use of sophisticated error mitigation techniques along with better gate fidelities to achieve better results on the QPU in the future.

## V. CONCLUSION

In this paper, we investigated and benchmarked the appropriate *Ansätze* and optimization strategies to find the ground state of the square-octagon-lattice Kitaev model with a VQE approach, and demonstrated the ability to determine the system's properties in different parameter regimes using VQE. We showed that the HVA generally outperforms the two HEAs using CZ or XY gates in the entangling layers. The HEAs require an expensive optimization process due to their rapidly increasing number of parameters with system size and circuit depth, and the unstructured nature of HEAs also makes them susceptible to barren plateaus. The HVA is associated with a relatively easy optimization and is thus better suited to the QPU experiments. The VQE solutions with HVA allow us to probe static expectation values and the energy gap as a

function of Hamiltonian parameters, making it possible to probe the phase diagram using VQE. We also conducted an experimental test to run classically optimized four-qubit and eight-qubit HVA circuits on the Aspen-9 QPU. With the help of readout error mitigation and ZNE, the four-qubit experiment gives a result roughly within one standard deviation of the true ground-state energy. For the eight-qubit experiment, the wider and deeper circuit increases the infidelity of the results. There, we observe a substantial difference from the state vector simulation even with the additional randomized compilation. This suggests that extra error mitigation techniques and performing the VQE optimization directly on the QPU would be required to extract useful results from the QPU for the future implementation of Kitaev-model VQE experiments.

Multiple optimization algorithms have been tested to optimize the cost function associated with our VQE application. We find that a mixed usage of different optimizers with multiple initial values gives a much more consistent and better result than just using one specific optimizer with one initial value. In particular, a local optimizer (BOBYQA) mixed with a nonlocal optimizer (CMA-ES) is shown to be an appropriate strategy for VQE calculations of the Kitaev model. SPSA is also a useful addition to the optimization with shot noise. Although we performed our tests with the noiseless simulator and simulation with shot noise, we expect that QPU experiments will also benefit from this approach as long as the QPU's clock rate is high enough to support the required number of cost function evaluations.

## ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Superconducting Quantum Materials and Systems Center (SQMS) under contract number DE-AC02-07CH11359. We would like to thank the entire SQMS algorithms team for fruitful and thought provoking discussions around these works. We would also like to thank the Rigetti team for assistance with running experiments on the Aspen-9 processor, particularly Matt Reagor, Mark Hodson, Bram Evert, Alex Hill, Eric Hulburd, and Dylan Anthony. P.P.O. acknowledges useful discussions with Alexander Huynh and Anirban Mukherjee. Some calculations were performed as part of the XSEDE computational Project No. TG-MCA93S030 on Bridges-2 at the Pittsburgh super-computer center.

- 
- [1] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Toward the first quantum simulation with quantum speedup, *Proc. Natl. Acad. Sci. USA* **115**, 9456 (2018).
  - [2] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A* **52**, R2493 (1995).
  - [3] J. Roffe, Quantum error correction: An introductory guide, *Contemp. Phys.* **60**, 226 (2019).
  - [4] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, A. Bengtsson, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, Y.-A. Chen, B. Chiaro, R. Collins, S. J. Cotton, W. Courtney *et al.*, Observation of separated dynamics of charge and spin in the Fermi-Hubbard model, [arXiv:2010.07965](https://arxiv.org/abs/2010.07965).
  - [5] A. Kitaev, Anyons in an exactly solved model and beyond, *Ann. Phys. (NY)* **321**, 2 (2006).
  - [6] S. Yang, D. L. Zhou, and C. P. Sun, Mosaic spin models with topological order, *Phys. Rev. B* **76**, 180404(R) (2007).
  - [7] J. Preskill, Quantum Computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
  - [8] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).

- [9] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, *Rev. Mod. Phys.* **94**, 015004 (2022).
- [10] 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).
- [11] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding *et al.*, Scalable Quantum Simulation of Molecular Energies, *Phys. Rev. X* **6**, 031007 (2016).
- [12] 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).
- [13] 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).
- [14] 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).
- [15] W. J. Huggins, J. Lee, U. Baek, B. O'Gorman, and K. B. Whaley, A non-orthogonal variational quantum eigensolver, *New J. Phys.* **22**, 073009 (2020).
- [16] F. Zhang, N. Gomes, Y. Yao, P. P. Orth, and T. Iadecola, Adaptive variational quantum eigensolvers for highly excited states, *Phys. Rev. B* **104**, 075159 (2021).
- [17] H. L. Tang, V. O. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, Qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor, *PRX Quantum* **2**, 020310 (2021).
- [18] O. R. Meitei, B. T. Gard, G. S. Barron, D. P. Pappas, S. E. Economou, E. Barnes, and N. J. Mayhall, Gate-free state preparation for fast variational quantum eigensolver simulations, *npj Quantum Info.* **7**, 155 (2021).
- [19] N. Gomes, A. Mukherjee, F. Zhang, T. Iadecola, C.-Z. Wang, K.-M. Ho, P. P. Orth, and Y.-X. Yao, Adaptive variational quantum imaginary time evolution approach for ground state preparation, *Adv. Quantum Technol.* **4**, 2100114 (2021).
- [20] J. S. Van Dyke, G. S. Barron, N. J. Mayhall, E. Barnes, and S. E. Economou, Scaling adaptive quantum simulation algorithms via operator pool tiling, [arXiv:2206.14215](https://arxiv.org/abs/2206.14215).
- [21] A. Mukherjee, N. F. Berthussen, 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).
- [22] P. G. Anastasiou, Y. Chen, N. J. Mayhall, E. Barnes, and S. E. Economou, TETRIS-ADAPT-VQE: An adaptive algorithm that yields shallower, denser circuit ansätze, [arXiv:2209.10562](https://arxiv.org/abs/2209.10562).
- [23] Y. Zhang, L. Cincio, C. F. A. Negre, P. Czarnik, P. Coles, P. M. Anisimov, S. M. Mniszewski, S. Tretiak, and P. A. Dub, Variational quantum eigensolver with reduced circuit complexity, *npj Quant. Inform.* **8**, 96 (2022).
- [24] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, *Phys. Rev. A* **92**, 042303 (2015).
- [25] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, Quantum computation by adiabatic evolution, [arXiv:quant-ph/0001106](https://arxiv.org/abs/quant-ph/0001106).
- [26] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, [arXiv:1411.4028](https://arxiv.org/abs/1411.4028).
- [27] 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).
- [28] 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).
- [29] 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).
- [30] M. Hermanns, I. Kimchi, and J. Knolle, Physics of the Kitaev model: Fractionalization, dynamic correlations, and material connections, *Annu. Rev. Condens. Matter Phys.* **9**, 17 (2018).
- [31] H. Takagi, T. Takayama, G. Jackeli, G. Khaliullin, and S. E. Nagler, Concept and realization of Kitaev quantum spin liquids, *Nat. Rev. Phys.* **1**, 264 (2019).
- [32] C. Hickey and S. Trebst, Emergence of a field-driven  $u(1)$  spin liquid in the Kitaev honeycomb model, *Nat. Commun.* **10**, 530 (2019).
- [33] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma, Non-Abelian anyons and topological quantum computation, *Rev. Mod. Phys.* **80**, 1083 (2008).
- [34] Y.-C. Lee, C. G. Brell, and S. T. Flammia, Topological quantum error correction in the Kitaev honeycomb model, *J. Stat. Mech.* (2017) 083106.
- [35] M. H. Freedman, M. Larsen, and Z. Wang, A modular functor which is universal for quantum computation, *Commun. Math. Phys.* **227**, 605 (2002).
- [36] G. Jackeli and G. Khaliullin, Mott Insulators in the Strong Spin-Orbit Coupling Limit: From Heisenberg to a Quantum Compass and Kitaev Models, *Phys. Rev. Lett.* **102**, 017205 (2009).
- [37] A. Banerjee, C. A. Bridges, J.-Q. Yan, A. A. Aczel, L. Li, M. B. Stone, G. E. Granroth, M. D. Lumsden, Y. Yiu, J. Knolle *et al.*, Proximate Kitaev quantum spin liquid behaviour in a honeycomb magnet, *Nat. Mater.* **15**, 733 (2016).
- [38] A. Banerjee, P. Lampen-Kelley, J. Knolle, C. Balz, A. A. Aczel, B. Winn, Y. Liu, D. Pajerowski, J. Yan, and C. A. Bridges, Excitations in the field-induced quantum spin liquid state of  $\alpha$ -RuCl<sub>3</sub>, *npj Quantum Mater.* **3**, 8 (2018).
- [39] Y. Kasahara, K. Sugii, T. Ohnishi, M. Shimozawa, M. Yamashita, N. Kurita, H. Tanaka, J. Nasu, Y. Motome, T. Shibauchi *et al.*, Unusual Thermal Hall Effect in a Kitaev Spin Liquid Candidate  $\alpha$ -RuCl<sub>3</sub>, *Phys. Rev. Lett.* **120**, 217205 (2018).
- [40] T. Yokoi, S. Ma, Y. Kasahara, S. Kasahara, T. Shibauchi, N. Kurita, H. Tanaka, J. Nasu, Y. Motome, C. Hickey *et al.*, Half-integer quantized anomalous thermal Hall effect in the Kitaev material candidate  $\alpha$ -RuCl<sub>3</sub>, *Science* **373**, 568 (2021).
- [41] J. Chaloupka, G. Jackeli, and G. Khaliullin, Kitaev-Heisenberg Model on a Honeycomb Lattice: Possible Exotic Phases in Iridium Oxides A<sub>2</sub>IrO<sub>3</sub>, *Phys. Rev. Lett.* **105**, 027204 (2010).
- [42] H.-C. Jiang, Z.-C. Gu, X.-L. Qi, and S. Trebst, Possible proximity of the Mott insulating iridate Na<sub>2</sub>IrO<sub>3</sub> to a topological phase: Phase diagram of the Heisenberg-Kitaev model in a magnetic field, *Phys. Rev. B* **83**, 245104 (2011).

- [43] J. Chaloupka, G. Jackeli, and G. Khaliullin, Zigzag Magnetic Order in the Iridium Oxide  $\text{Na}_2\text{IrO}_3$ , *Phys. Rev. Lett.* **110**, 097204 (2013).
- [44] J. Osorio Iregui, P. Corboz, and M. Troyer, Probing the stability of the spin-liquid phases in the Kitaev-Heisenberg model using tensor network algorithms, *Phys. Rev. B* **90**, 195102 (2014).
- [45] J. G. Rau, Eric Kin-Ho Lee, and H.-Y. Kee, Generic Spin Model for the Honeycomb Iridates beyond the Kitaev Limit, *Phys. Rev. Lett.* **112**, 077204 (2014).
- [46] K. Shinjo, S. Sota, and T. Tohyama, Density-matrix renormalization group study of the extended Kitaev-Heisenberg model, *Phys. Rev. B* **91**, 054401 (2015).
- [47] S. M. Winter, Y. Li, H. O. Jeschke, and R. Valentí, Challenges in design of Kitaev materials: Magnetic interactions from competing energy scales, *Phys. Rev. B* **93**, 214431 (2016).
- [48] R. Yadav, N. A. Bogdanov, V. M. Katukuri, S. Nishimoto, J. Van Den Brink, and L. Hozoi, Kitaev exchange and field-induced quantum spin-liquid states in honeycomb  $\alpha\text{-RuCl}_3$ , *Sci. Rep.* **6**, 1 (2016).
- [49] M. Gohlke, R. Verresen, R. Moessner, and F. Pollmann, Dynamics of the Kitaev-Heisenberg Model, *Phys. Rev. Lett.* **119**, 157203 (2017).
- [50] D. Gotfryd, J. Rusnačko, K. Wohlfeld, G. Jackeli, J. c. v. Chaloupka, and A. M. Oleś, Phase diagram and spin correlations of the Kitaev-Heisenberg model: Importance of quantum effects, *Phys. Rev. B* **95**, 024426 (2017).
- [51] Z. Zhu, I. Kimchi, D. N. Sheng, and L. Fu, Robust non-Abelian spin liquid and a possible intermediate phase in the antiferromagnetic Kitaev model with magnetic field, *Phys. Rev. B* **97**, 241110(R) (2018).
- [52] N. D. Patel and N. Trivedi, Magnetic field-induced intermediate quantum spin liquid with a spinon fermi surface, *Proc. Natl. Acad. Sci. USA* **116**, 12199 (2019).
- [53] J. S. Gordon, A. Catuneanu, E. S. Sørensen, and H.-Y. Kee, Theory of the field-revealed Kitaev spin liquid, *Nat. Commun.* **10**, 1 (2019).
- [54] P. Czarnik, A. Francuz, and J. Dziarmaga, Tensor network simulation of the Kitaev-Heisenberg model at finite temperature, *Phys. Rev. B* **100**, 165147 (2019).
- [55] H.-Y. Lee, R. Kaneko, L. E. Chern, T. Okubo, Y. Yamaji, N. Kawashima, and Y. B. Kim, Magnetic field induced quantum phases in a tensor network study of Kitaev magnets, *Nat. Commun.* **11**, 1 (2020).
- [56] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [57] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Ann. Phys. (NY)* **326**, 96 (2011).
- [58] M. Fishman, S. R. White, and E. Miles Stoudenmire, The ITensor Software library for tensor network calculations, *SciPost Phys. Codebases* **4** (2022).
- [59] P. A. Mishchenko, Y. Kato, and Y. Motome, A quantum Monte Carlo method on asymptotic Lefschetz thimbles for quantum spin systems: An application to the Kitaev model in a magnetic field, *Phys. Rev. D* **104**, 074517 (2021).
- [60] M. Kurita, Y. Yamaji, S. Morita, and M. Imada, Variational Monte Carlo method in the presence of spin-orbit interaction and its application to Kitaev and Kitaev-Heisenberg models, *Phys. Rev. B* **92**, 035122 (2015).
- [61] G. Kells, J. Kailasvuori, J. K. Slingerland, and J. Vala, Kaleidoscope of topological phases with multiple majorana species, *New J. Phys.* **13**, 095014 (2011).
- [62] S. S. Hong, A. T. Papageorge, P. Sivarajah, G. Crossman, N. Didier, A. M. Polloreno, E. A. Sete, S. W. Turkowski, M. P. da Silva, and B. R. Johnson, Demonstration of a parametrically activated entangling gate protected from flux noise, *Phys. Rev. A* **101**, 012302 (2020).
- [63] D. M. Abrams, N. Didier, B. R. Johnson, M. P. da Silva, and C. A. Ryan, Implementation of XY entangling gates with a single calibrated pulse, *Nature Electronics* **3**, 744 (2020).
- [64] M. Reagor *et al.*, Demonstration of universal parametric entangling gates on a multi-qubit lattice, *Sci. Adv.* **4**, eaao3603 (2018).
- [65] C. Berke, S. Trebst, and C. Hickey, Field stability of majorana spin liquids in antiferromagnetic Kitaev models, *Phys. Rev. B* **101**, 214442 (2020).
- [66] C. Hickey, M. Gohlke, C. Berke, and S. Trebst, Generic field-driven phenomena in Kitaev spin liquids: Canted magnetism and proximate spin liquid physics, *Phys. Rev. B* **103**, 064417 (2021).
- [67] M. G. Yamada, Topological  $\mathbb{Z}_2$  invariant in Kitaev spin liquids: Classification of gapped spin liquids beyond projective symmetry group, *Phys. Rev. Res.* **3**, L012001 (2021).
- [68] T. A. Bespalova and O. Kyriienko, Quantum simulation and ground state preparation for the honeycomb Kitaev model, [arXiv:2109.13883](https://arxiv.org/abs/2109.13883).
- [69] A. Jahin, A. C. Y. Li, T. Iadecola, P. P. Orth, G. N. Perdue, A. Macridin, M. S. Alam, and N. M. Tubman, Fermionic approach to variational quantum simulation of Kitaev spin models, *Phys. Rev. A* **106**, 022434 (2022).
- [70] A. Kitaev, Fault-tolerant quantum computation by anyons, *Ann. Phys. (NY)* **303**, 2 (2003).
- [71] X.-G. Wen, Quantum Orders in an Exact Soluble Model, *Phys. Rev. Lett.* **90**, 016803 (2003).
- [72] Z. Nussinov and G. Ortiz, A symmetry principle for topological quantum order, *Ann. Phys. (NY)* **324**, 977 (2009).
- [73] B. J. Brown, W. Son, C. V. Kraus, R. Fazio, and V. Vedral, Generating topological order from a two-dimensional cluster state using a duality mapping, *New J. Phys.* **13**, 065010 (2011).
- [74] D.-B. Zhang, B.-L. Chen, Z.-H. Yuan, and T. Yin, Variational quantum eigensolvers by variance minimization, *Chinese Phys. B* **31**, 120301 (2022).
- [75] K. M. Nakanishi, K. Mitarai, and K. Fujii, Subspace-search variational quantum eigensolver for excited states, *Phys. Rev. Res.* **1**, 033062 (2019).
- [76] O. Higgott, D. Wang, and S. Brierley, Variational Quantum Computation of Excited States, *Quantum* **3**, 156 (2019).
- [77] A. Gu, A. Lowe, P. A. Dub, P. J. Coles, and A. Arrasmith, Adaptive shot allocation for fast convergence in variational quantum algorithms, [arXiv:2108.10434](https://arxiv.org/abs/2108.10434).
- [78] W. W. Ho and T. H. Hsieh, Efficient variational simulation of non-trivial quantum states, *SciPost Phys.* **6**, 029 (2019).
- [79] C. Cade, L. Mineh, A. Montanaro, and S. Stanisic, Strategies for solving the Fermi-Hubbard model on near-term quantum computers, *Phys. Rev. B* **102**, 235122 (2020).
- [80] W. Lavrijsen, A. Tudor, J. Müller, C. Iancu, and W. de Jong, Classical optimizers for noisy intermediate-scale

- quantum devices, in *Proceedings of the 2020 IEEE International Conference on Quantum Computing and Engineering* (IEEE, Piscataway, NJ, 2020), pp. 267–277.
- [81] A. Anand, M. Degroote, and A. Aspuru-Guzik, Natural evolutionary strategies for variational quantum computation, *Machine Learning: Science and Technology* **2**, 045012 (2021).
- [82] A. W. Harrow and J. C. Napp, Low-Depth Gradient Measurements Can Improve Convergence in Variational Hybrid Quantum-Classical Algorithms, *Phys. Rev. Lett.* **126**, 140502 (2021).
- [83] R. S. Smith, M. J. Curtis, and W. J. Zeng, A practical quantum instruction set architecture, [arXiv:1608.03355](https://arxiv.org/abs/1608.03355).
- [84] Cirq Developers, Cirq (v1.1.0) (2022), doi:[10.5281/zenodo.7465577](https://doi.org/10.5281/zenodo.7465577).
- [85] Qiskit contributors, Qiskit An open-source framework for quantum computing (2023), doi:[10.5281/zenodo.2573505](https://doi.org/10.5281/zenodo.2573505).
- [86] 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).
- [87] E. Peters, J. Caldeira, A. Ho, S. Leichenauer, M. Mohseni, H. Neven, P. Spentzouris, D. Strain, and G. N. Perdue, Machine learning of high dimensional data on a noisy quantum processor, *npj Quantum Info.* **7**, 161 (2021).
- [88] R. S. Smith, E. C. Peterson, M. G. Skilbeck, and E. J. Davis, An open-source, industrial-strength optimizing compiler for quantum programs, *Quantum Sci. Technol.* **5**, 044001 (2020).
- [89] P. J. Karalekas, N. A. Tezak, E. C. Peterson, C. A. Ryan, M. P. da Silva, and R. S. Smith, A quantum-classical cloud platform optimized for variational hybrid algorithms, *Quantum Sci. Technol.* **5**, 024003 (2020).
- [90] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey *et al.*, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods* **17**, 261 (2020).
- [91] M. J. D. Powell, The BOBYQA algorithm for bound constrained optimization without derivatives, Cambridge University Report No. DAMTP 2009/NA06, Centre for Mathematical Sciences, University of Cambridge, UK, p. 26.
- [92] Michael J. D. Powell, Bobyqa, <https://www.zhangzk.net/software.html>.
- [93] N. Hansen, Y. Akimoto, and P. Baudis, CMA-ES/pycma on Github, Zenodo (2019).
- [94] J. Gomez-Dans, Spsa, <https://github.com/jgomezdans/spsa> (2012).
- [95] N. Hansen and A. Ostermeier, Completely derandomized self-adaptation in evolution strategies, *Evol. Comput.* **9**, 159 (2001).
- [96] Y. Xiang, D. Sun, W. Fan, and X. Gong, Generalized simulated annealing algorithm and its application to the Thomson model, *Phys. Lett. A* **233**, 216 (1997).
- [97] J. C. Spall, Multivariate stochastic approximation using a simultaneous perturbation gradient approximation, *IEEE Trans. Automat. Contr.* **37**, 332 (1992).
- [98] L. M. Rios and N. V. Sahinidis, Derivative-free optimization: a review of algorithms and comparison of software implementations, *J. Global Optim.* **56**, 1247 (2013).
- [99] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, *Quantum Sci. Technol.* **4**, 014008 (2018).
- [100] A. Gilyén, S. Arunachalam, and N. Wiebe, Optimizing quantum optimization algorithms via faster quantum gradient computation, in *Proceedings of the 2019 Annual ACM-SIAM Symposium on Discrete Algorithms* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2019), pp. 1425–1444.
- [101] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* **99**, 032331 (2019).
- [102] K. Mitarai, Y. O. Nakagawa, and W. Mizukami, Theory of analytical energy derivatives for the variational quantum eigensolver, *Phys. Rev. Res.* **2**, 013129 (2020).
- [103] A. F. Izmaylov, R. A. Lang, and T.-C. Yen, Analytic gradients in variational quantum algorithms: Algebraic extensions of the parameter-shift rule to general unitary transformations, *Phys. Rev. A* **104**, 062443 (2021).
- [104] W. J. Huggins, B. A. O’Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, Unbiasing fermionic quantum Monte Carlo with a quantum computer, *Nature* **603**, 416 (2022).
- [105] K. Klymko, C. Mejuto-Zaera, S. J. Cotton, F. Wudarski, M. Urbanek, D. Hait, M. Head-Gordon, K. B. Whaley, J. Moussa, N. Wiebe *et al.*, Real-time evolution for ultracompact Hamiltonian eigenstates on quantum hardware, *PRX Quantum* **3**, 020323 (2022).
- [106] Y. Li and S. C. Benjamin, Efficient Variational Quantum Simulator Incorporating Active Error Minimization, *Phys. Rev. X* **7**, 021050 (2017).
- [107] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [108] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, Digital zero noise extrapolation for quantum error mitigation, in *Proceedings of the 2020 IEEE International Conference on Quantum Computing and Engineering* (IEEE, Piscataway, NJ, 2020), pp. 306–316.
- [109] M. R. Geller, Rigorous measurement error correction, *Quantum Sci. Technol.* **5**, 03LT01 (2020).
- [110] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, *Phys. Rev. A* **94**, 052325 (2016).
- [111] Note that the same unitary can also be written  $\mathbb{I}_c \otimes |0\rangle\langle 0| + \mathbb{Z}_c \otimes |1\rangle\langle 1|_t$ , so that the CZ gate remains invariant if we swap the labels c and t. Therefore, it does not matter how we assign the labels “control”/“target” to the two qubits participating in this interaction.
- [112] K. Schultz, R. LaRose, A. Mari, G. Quiroz, N. Shammah, B. D. Clader, and W. J. Zeng, Impact of time-correlated noise on zero-noise extrapolation, *Phys. Rev. A* **106**, 052406 (2022).
- [113] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O’Brien, Quantum error mitigation, [arXiv:2210.00921](https://arxiv.org/abs/2210.00921).



- [114] V. Russo, A. Mari, N. Shammah, R. LaRose, and W. J. Zeng, Testing platform-independent quantum error mitigation on noisy quantum computers, [arXiv:2210.07194](https://arxiv.org/abs/2210.07194).
- [115] H. Ge, K. Xu, and Z. Ghahramani, Turing: A language for flexible probabilistic inference, in *Proceedings of the International Conference on Artificial Intelligence and Statistics* (JMLR, Cambridge MA, 2018), pp. 1682–1690.
- [116] M. D. Hoffman, A. Gelman *et al.*, The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo, *J. Mach. Learn. Res.* **15**, 1593 (2014).
- [117] D. B. Rubin, The bayesian bootstrap, *Ann. Statist.* **9**, 130 (1981).
- [118] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, *Nat Commun.* **12**, 6961 (2021)
- [119] C. Cartis, J. Fiala, B. Marteau, and L. Roberts, Improving the flexibility and robustness of model-based derivative-free optimization solvers, *ACM Trans. Math. Softw.* **45**, 1 (2019).