Classically efficient quantum scalable Fermi-Hubbard benchmark

Bryan T. Gard [*] and Adam M. Meier

Georgia Tech Research Institute, Atlanta, Georgia 30332, USA

In order to quantify the relative performance of different test-bed quantum computing devices, it is useful to benchmark them using a common protocol. While some benchmarks rely on the performance of random circuits and are generic in nature, here we instead propose and implement a practical, application-based benchmark. In particular, our protocol calculates the energy of the ground state in the single-particle subspace of a one-dimensional (1D) Fermi Hubbard model, a problem which is efficient to solve classically. We provide a quantum ansatz for the problem that is provably able to probe the full single-particle subspace for a general-length 1D chain and scales efficiently in number of gates and measurements. Finally, we demonstrate and analyze the benchmark performance on superconducting and ion-trap test-bed hardware from three hardware vendors and with up to 24 qubits.

I. INTRODUCTION

The performance of noisy intermediate-scale quantum (NISQ) devices is impacted by a variety of noise sources [1], which limit the sizes of quantum circuits that can be implemented effectively. In order to characterize the generic performance of NISQ devices, benchmarks based on random circuits have been used widely [2–8]. These benchmarks, including randomized benchmarking (RB), quantum volume (QV), and algorithmic qubits (AQs), all rely on both the notion that random circuits are useful stand-ins for applications of interest and the constructions of the underlying circuits that allow them to be efficiently simulated classically [6,9]. However, the driving interest in NISQ devices is their promise for improving solutions to practical problems that are not efficient classically, including quantum chemistry [10,11], combinatorial optimization [12–16], quantum simulation [17,18], and machine learning [19,20]. In order to probe these use cases more directly, recent research has explored "application-specific" benchmarks that directly mimic practical applications of interest [16,21–23]. We focus on the variational quantum eigensolver (VQE), which has received significant attention as a hybrid algorithm that has the potential to improve on classical solutions to practical quantum simulation and chemistry problems in the near term [24–30].

A NISQ application-specific benchmark benefits from several key properties. It should be scalable in the sense that it can be applied to arbitrarily sized quantum systems. For near-term implementations, it is also desirable for the benchmark to have low circuit depth. Similarly, it should require few unique circuits and measurement settings, thus lessening the required run time and cost on commercial test-bed devices. The benchmark should be straightforward to implement efficiently on a broad class of hardwares and architectures, which generally favors constructions based on standard one- and two-qubit gates with nearest-neighbor connectivity. Finally, the benchmark should discriminate between the performances of currently available devices.

In this work, we introduce and demonstrate an efficient application-based benchmark focused on finding low-energy states of the Fermi-Hubbard model, in a spirit similar to work from Dallaire-Demers *et al.* [21]. We construct a scalable quantum ansatz for the problem which uses relevant symmetries in the problem to reduce the required quantum resources [28,31–33]. This ansatz is similar to ones using Givens rotations [34–36].

We show that the construction of the quantum circuit for use in the benchmark has a gate count and optimization parameter count which scale linearly in the number of qubits. Additionally, the benchmark also requires only a constant number of noncommuting sets of measurements (independent of problem size). Additionally, our ansatz assumes only nearest-neighbor connectivity and uses a common gate set consisting of controlled-NOT (CNOT) entangling gates and parametrized single-qubit rotations. We also implement the benchmark on nine different test-bed devices from three different providers to demonstrate how broadly approachable it is for current hardware. In addition, one can employ compression techniques which reduce the qubit requirements [37,38], although we do not apply this technique in our benchmark.

Any quantum computer benchmark reports the performance of a composite of the quantum hardware and the software required to program the hardware. For each device we benchmark, QISKIT or Amazon BRAKET software submits quantum circuit jobs to the respective quantum hardware. This software also maps virtual qubits and gates in the quantum circuits to specific physical qubits and gates on the device; this mapping can be optimized by the software to highlight the best-performing qubits and interactions. Therefore, the resulting performance is a consequence of both the hardware performance and the automated selection done by the software.

The primary result of our benchmark is a single number representing the largest size (twice the number of fermionic

---

*Bryan.gard@gtri.gatech.edu

sites) of the Fermi-Hubbard calculation implemented in the protocol that returns a result below an error threshold. This output is similar in kind to the "algorithmic qubits" benchmark [9] in that it represents the maximum number of qubits in the device that can be effectively utilized to solve a problem. We note that QV and AQs both require that classical simulation of the random circuits is possible for these benchmarks and therefore are inherently not scalable unless restricted to gates chosen from the Clifford group. In our benchmark, we do not directly require classical simulation, although the single-particle problem is exactly efficiently solvable, nevertheless. We discuss extending the benchmark to more particles in Sec. A 3.

We begin in Sec. II with a description of our benchmark and its required components. In Sec. III we show the results of our benchmark run on IBM, IonQ, and Rigetti hardware. We discuss our results and conclude in Sec. IV.

II. BENCHMARK

While benchmarks like RB and QV attempt to capture generic performance based on random circuits constructed within a rubric, it is also interesting to consider the performance of quantum devices based on their ability to solve practical and specific problems. We focus on the problem of finding the ground state of a physically relevant Hamiltonian, which is a common underlying task in NISQ applications. In particular, we choose the Hamiltonian for the one-dimensional (1D) Fermi-Hubbard model because of its simplicity and familiarity, as well as its symmetry properties, which lead to simpler circuits. The general 1D Fermi-Hubbard model is given by the equation

$$\hat{H} = -t \sum_{\langle i,j \rangle, \sigma} (\hat{a}_{i,\sigma}^\dagger \hat{a}_{j,\sigma} + \hat{a}_{j,\sigma}^\dagger \hat{a}_{i,\sigma}) + U \sum_i \hat{a}_{i,\uparrow}^\dagger \hat{a}_{i,\uparrow} \hat{a}_{i,\downarrow}^\dagger \hat{a}_{i,\downarrow},$$
(1)

where $\hat{a}_{i,\sigma}^\dagger$ ($\hat{a}_{i,\sigma}$) is the creation (annihilation) operator associated with site i and spin σ. We map this fermionic problem to a qubit problem by way of the Jordan-Wigner mapping, which directly maps spin orbitals to qubits while maintaining the anticommutation rules of the original Hamiltonian [39]. This maps a problem defined for an L-site chain to $N = 2L$ qubits (since each site has two spins).

Solving for the ground state of the generic Fermi-Hubbard Hamiltonian is computationally hard. A brute-force classical solution for a two-dimensional (2D) grid-based Fermi-Hubbard model requires diagonalization of a size $2^{2n_x n_y} \times 2^{2n_x n_y}$ matrix for n_x and n_y sites in the horizontal and vertical directions [30]. The model is analytically solvable in the limiting cases of $\frac{U}{t} \to 0$ and $\frac{U}{t} \to \infty$ and in the 1D case using the Bethe ansatz, but general exact solutions are not known [40–42]. Here, to further simplify the problem for current NISQ hardware and to make a more intuitive benchmark, we restrict our focus to the single-particle ground state. In this regime, we can take $U = 0$ without loss of generality because, for the single-particle ground state, there are no possible interactions, and therefore, Eq. (1) is invariant to the choice of U. This choice leads to a simpler form of the Hamiltonian that can be solved exactly [43]. We further fix $t = 1$ and choose open (nonperiodic) boundary conditions.
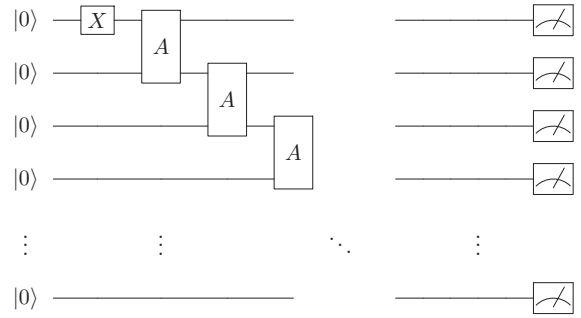

FIG. 1. Circuit used to produce the single-particle ansatz. Each A gate preserves the particle number and spin projection. A gates are present on the top half of qubits, and measurements are taken on all qubits.

Under these conditions, the analytic, single-particle ground-state energy of the 1D Fermi Hubbard model is

$$E_{\text{gs}} = 2 \cos[L\pi/(L+1)],$$
(2)

where L is the number of sites in the chain. This simple solution for the restricted problem gives an easy target for the benchmark at any length L. Note that considering $U = 0$ was only a tool to clarify the ideal exact solution; for the measurement of energies on hardware, states which are not only single particles are present due to noise. In order to capture these errors in the benchmark results and tie the benchmark implementation to the more general problem, we use the full Hamiltonian with $U = 2$.

To test the ability of NISQ devices to find this target ground state, we employ the VQE algorithm. This algorithm relies on the fact that, for any parameterized quantum state $\psi(\theta_i)$, minimizing $\langle \psi(\theta_i)|H|\psi(\theta_i) \rangle$ over θ_i bounds the ground-state energy. If the parametrization allows for the creation of exactly the space of states allowed by the problem constraints, e.g., a fixed particle number, then optimizing the energy in this way solves for a ground state subject to those constraints (there may be many due to degeneracy).

To implement the VQE algorithm we also need to specify an ansatz (sometimes also called a variational quantum circuit). For resource efficiency, we choose an ansatz which is built to enforce the symmetries in the Hamiltonian. This kind of ansatz has been used in previous works related to symmetry-preserving circuits for chemical ground states [28,31,33,42,44]. Our ansatz, depicted in Fig. 1, is built from a single primitive gate, A, with a single parameter, θ, where A has the following computational-basis matrix representation:

$$A(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$
(3)

Constructing an ansatz from this gate has several beneficial features. First, an ansatz built from parameterized SWAP-type gates naturally conserves the particle number. In addition, when Eq. (1) is mapped to qubit operators using a spin-block-based Jordan-Wigner mapping, our ansatz also preserves spin
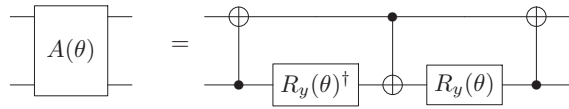
FIG. 2. Decomposition of the A gate in terms of elementary single- and two-qubit gates. $R_y(\phi) = \exp(-i\phi\sigma_y/2)$.

FIG. 3. Example symmetry-based ansatz for $L = 4$, composed of three simplified A gates and three parameters. While gates are required on only the first L qubits, measurement of all $N = 2L$ qubits is performed to calculate the energy of the Hamiltonian defined on N qubits.

projection. The ansatz also maps real states to real states, and we note that all ground states of our chosen problem are real valued. Since these symmetries are encoded into the ansatz, it fundamentally requires fewer quantum resources than a more general ansatz, e.g., Refs. [25,45]. Since the problem of interest is a single-particle ground state, we start with a single excited qubit and construct a simple ladderlike circuit for an arbitrary number of sites, as shown in Fig. 1. Intuitively, the proposed circuit can swap the initial single-particle excitation into any qubit, controlled by the parameters θ_i for qubit i.

In the standard gate set for quantum computing and for a generic input state, each A gate can be decomposed into three CNOT and two single-qubit gates, as shown in Fig. 2. However, since we always begin the benchmark circuit in a fixed initial state, we can make small simplifications to the general circuit, as shown in Figs. 7 and 8 in the Appendix.

Based on these simplifications, the ansatz requires $2L - 3$ CNOT gates and $L - 1$ parameters for $L \geqslant 2$, the number of sites in the Fermi-Hubbard chain. Since the ansatz requires only 1D nearest-neighbor connectivity, this gate count can be achieved by any device with nearest-neighbor CNOT gates, avoiding the overhead frequently required to map arbitrarily connected circuits to a specific hardware connectivity. As a result, the benchmark does not significantly benefit from the arbitrary connectivity allowed in some quantum systems, e.g., trapped ions. This is an example of the idiosyncrasies introduced by application-specific benchmarks that differentiate them from random-circuit benchmarks. Finally, we prove in Sec. A 2 that the ansatz can identify the single-particle ground state for any L.

Figure 3 presents an example of the ansatz for $L = 4$. Notably, the circuit fills only the top half of the quantum register. This is a direct consequence of the single-particle ground state necessarily having a spin-projection value of $s_z = \pm 0.5$. Because we have mapped the original fermionic problem to qubit operators using the Jordan-Wigner mapping and we have used a block-based spin encoding, the top half of qubits encodes spin-up particles, while the remaining qubits encode spin down. The two choices of s_z are degenerate with one another and are related by a simple bit-flip operation on all qubits, so we need only generate one of these choices, and we choose the $s_z = 0.5$ state. Note that the general Hamiltonian itself requires $2L$ qubits to express, even though gates are present on only the first L qubits in our restricted problem. In our benchmark, we still choose to measure all $2L$ qubits to calculate the energy of the single-particle state, and this choice means that measurement errors on the unused qubits still impact the result. In principle we could allow devices which employ midcircuit reset to reset the empty qubits prior to measurement, but we do not use this in our benchmark. Writing the protocol in this way preserves our ability to extend
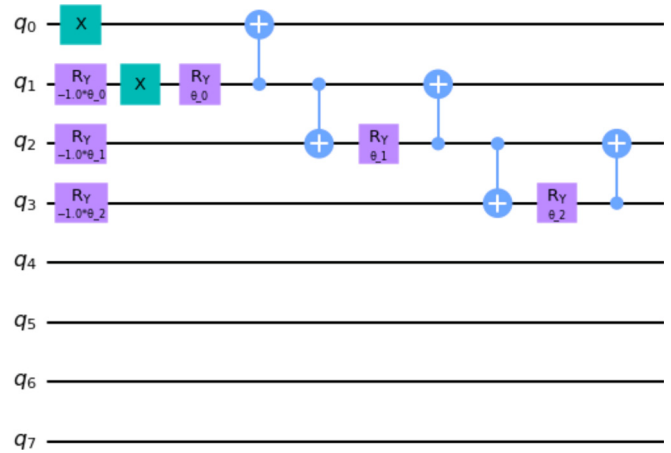
the Hamiltonian to more than one particle without changing the construction. We further discuss extension of the benchmark to more particles in Sec. A 3.

The problem Hamiltonian, once mapped to qubit operators, can be written as a sum of Pauli strings. For instance, the Hamiltonian of the simplest case at $L = 2$ is

$$\hat{H} = IIII + \tfrac{1}{2}(IZIZ + ZIZI - IZYY - IZII - ZIII$$
$$- IZXX - YYZI - IIIZ - IIZI - XXZI).$$

For most devices, measuring any operators that contain X or Y requires that we rotate the qubits prior to a Z-basis measurement. We can group these operators into a small number of tensor-product-basis (TPB) sets where all operators within each set are diagonal in the same tensor product basis and can be measured simultaneously [25]. In this example case, we can group the Hamiltonian into five groups of terms given by

$$\hat{H} = H_0 + H_1 + H_2 + H_3 + H_4,$$
$$H_0 = IIII + \tfrac{1}{2}(IZIZ + ZIZI - ZIII$$
$$\qquad - IZII - IIZI - IIIZ),$$
$$H_1 = -\tfrac{1}{2}IZXX,$$
$$H_2 = -\tfrac{1}{2}IZYY,$$
$$H_3 = -\tfrac{1}{2}YYZI,$$
$$H_4 = -\tfrac{1}{2}XXZI.$$

Note that this grouping is not unique and many other choices exist but do not change the resulting measurement of $\langle\hat{H}\rangle$. In Ref. [30], the authors showed that these Hamiltonian terms can always be grouped into five TPB sets at most. As a result, measuring the Hamiltonian requires five different measurement settings that can each be implemented with only single-qubit gates. Because the evaluation of the Hamiltonian requires a fixed number of simple measurement settings and the ansatz requires only linear scaling of both qubits and gates,
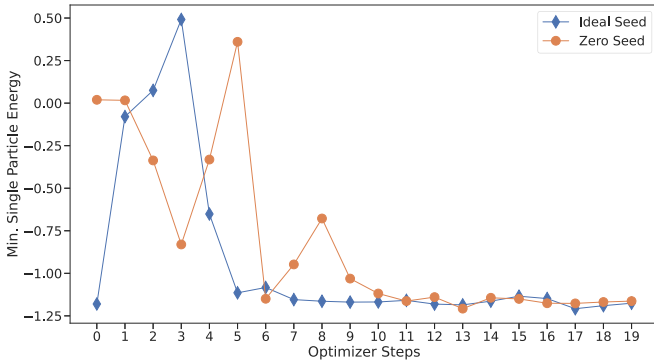
FIG. 4. VQE convergence using the COBYLA optimizer over 20 steps on IBM Guadalupe hardware, using an initial point seeded from an ideal simulation (blue diamonds) and using a fixed initial point of [0,0] (orange circles).

the benchmark is both simple to implement at small sizes and simple to scale to larger sizes.

The motivating application for our benchmark is the use of the VQE hybrid algorithm in order to find minimum-energy solutions on quantum hardware. To find the minimum energies, the algorithm optimizes a parameterized circuit (creating a parameterized state). While the VQE algorithm can, in principle, start with any seed parameters, we can assist it by preoptimizing the parameters classically. In the case of the hardware devices tested, we do not actually perform the hybrid algorithm at all. Instead, we evaluate only the energy of the preoptimized parameters. This choice reduces the computational (and monetary) cost of the benchmark, which is only formally dependent on the accuracy of the energy of the optimized parameters. Through comparison of the energy associated with the preoptimized parameters with the ideal result, we can show that the preoptimization does not appreciably impact the benchmark results: the maximum-energy difference we find in this comparison is 2.946×10^{-10} at ($L = 12$). This demonstrates that the VQE algorithm itself along with classical optimization is able to very accurately find the desired minimal-energy states and does not limit the benchmark results we report.

In Fig. 4 we show an example of the convergence behavior of the full VQE algorithm over 20 steps using IBM Guadalupe for the three-site Fermi Hubbard benchmark. We plot the convergence for two different sets of initial parameters: one with initial parameters arbitrarily fixed to zero and one with the classically preoptimized parameters. The use of the ideal parameters assists by improving the minimum number of steps before convergence of the optimizer but is not necessary. This simple case with only two parameters indicates the expected result that the VQE algorithm is able to converge on hardware eventually without the need of any specific initial point derived from simulation.

The issue of how to optimize hybrid performance with respect to splitting resources between VQE evaluations and classical preoptimization for large application instances is an interesting and important problem. Performing poorly at this optimization could impact very large instances of our Fermi length benchmark and is an example of how the benchmark evaluates both hardware and software aspects of NISQ

TABLE I. Table of devices used in our benchmark along with their maximum usable qubits, quantum volume (where available), and the result of our benchmark. L^* lists the maximum number of qubits (twice the chain length) used which pass the benchmark for both the raw (R) and mitigated (M) results. An asterisk (*) denotes benchmarks which were terminated because they exceeded the threshold error score; otherwise, the benchmark was run up to the maximum size allowed by each device or postprocessing. We used a relaxed-value score of 10 as our error-score threshold for passing the benchmark as described in the text.

Device	Maximum Qubits	log(QV)	L^* (R)	L^* (M)
Jakarta	7	4	0*	6
Casablanca	7	5	0*	6
Guadalupe	16	5	0*	10*
Toronto	27	5	0*	6*
Mumbai	27	7	0*	16*
Montreal	27	7	0*	24
Brooklyn	65	5	0*	20*
IonQ	11		5*	5*
Rigetti Aspen-9	32		0*	0*

devices; however, the instances here are too small for the difficulty of classical optimization to have an impact.

III. RESULTS

We ran the benchmark described in the previous section on nine devices provided by IBM, Rigetti Computing, and IonQ. We accessed private IBM devices (seven total) through IBM QISKIT software [46]. We accessed IonQ and Rigetti devices through Amazon Web Services' (AWS) BRAKET software.

For each choice of chain length, we seed the hardware evaluations by first running a simulator of the VQE algorithm using the high-performance QULACS software package [47]. These simulations provide optimal parameters θ_i for each choice of L that we use to initialize the same problem on the quantum devices. For all hardware, we evaluate the energy at the fixed seed-parameter values in order to reduce cost. The benchmark evaluation relies only on the energy evaluation of this single point for each L. The QULACS classical simulation is not a required step for the actual VQE algorithm; in principle, the same results can be obtained by implementing more steps of the hybrid algorithm on the quantum hardware at a greater cost to the quantum devices.

For each device, we run the benchmark for increasing chain lengths starting with $L = 2$ and terminating when the benchmark performance is severely degraded or when the maximum system size is reached. For all hardware devices, we fixed the number of repetitions (shots) to 8192. In Table I we capture the qubit number, the quantum volume reported elsewhere in the literature, and our benchmark result for each device. Of the nine devices we benchmarked, three were able to run to their full physical limit, while the remaining six were terminated early (note that the protocol requires an even number of qubits).

We show the raw energies computed using each device in Fig. 5. In the left panel, we show the raw results, and we observe that the superconducting devices typically have larger
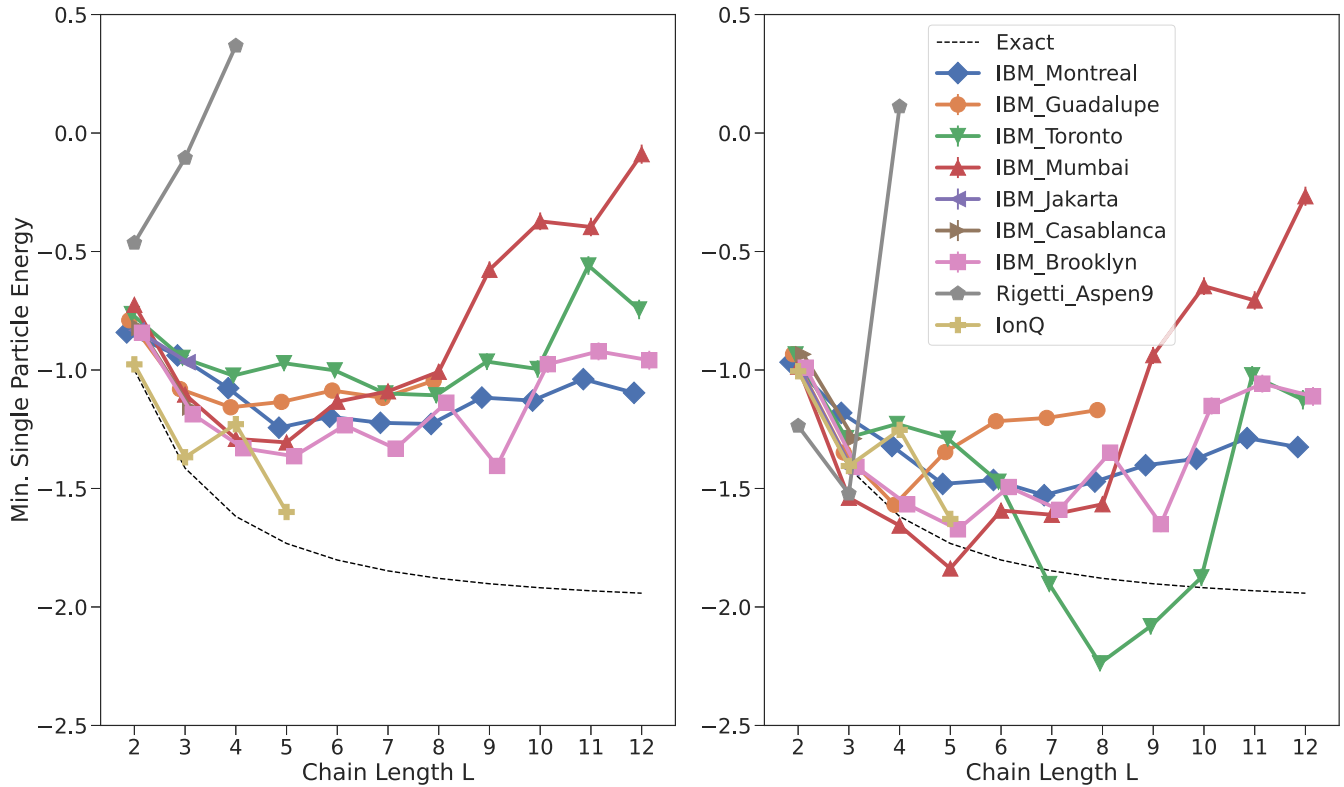
FIG. 5. Left: Raw data of the single-particle energy as a function of Fermi-Hubbard chain length ($N = 2L$). In all cases, error bars are smaller than marker size. Right: Mitigated data of the single-particle energy after SPAM error mitigation. Data points are horizontally offset slightly from their integer values to reduce clutter.

errors than the IonQ device, which performs well through $L = 5$ (with the exception of $L = 4$). Most devices visually follow the trend of the ideal energy at short chain lengths before crossing a threshold and drifting farther away from the exact energy. This makes the benchmark easy to evaluate qualitatively. However, either due to noise in the hardware or differences in software optimization, the threshold can be difficult to identify exactly, e.g., for the IBM Brooklyn device, which performs better at length 9 than at length 8. For this reason we establish in the following paragraphs a criterion that condenses the benchmark results to a single number for report.

We find that the explanatory power of the benchmark results benefits from the option of including standardized state preparation and measurement (SPAM) error mitigation in the protocol. We describe our approach to SPAM mitigation in Sec. A 4. SPAM mitigation leads to significantly better energy estimation across most devices tested. The results of applying the SPAM mitigation are shown in the right panel of Fig. 5, and here the hardware results track the ideal results more closely than in the left panel, although significant errors are still present. When characterizing the performance of a device with our benchmark, we consider the performance obtained after SPAM mitigation to be the primary result. However, we note that this mitigation technique can require significant classical computing resources for large problem sizes. For this reason, we include only benchmark results up to 24 qubits. We also present the raw (unmitigated) results in both Fig. 5 and Table I for comparison.

In order to distill the benchmark to a single number, we choose a benchmark score inspired by the accuracy requirement in the LINPACK benchmark [48]. We define our error score by

$$E_s = \sqrt{2M} \times |\langle E \rangle - E_{gs}|/L, \qquad (4)$$

where $\langle E \rangle$ is the measured energy from the quantum device after SPAM mitigation, E_{gs} is the known classical result, L is the chain length, and M is the number of shots in each measurement. M plays the role of the factor representing machine precision in the LINPACK benchmark as here it controls the accuracy with which we can measure energy, up to statistical variance.

Consistent with the intuition for shot noise of a single measurement, we find empirically from simulations that the standard error of the measured energy scales as $\sqrt{2M}$ for $M > 100$, and therefore, the prefactor in Eq. (4) effectively removes the error score's dependence on M when $M > 100$. To avoid small sample effects, we propose that our benchmark always be implemented with $M > 100$, and we note that comparisons are simplest when M is fixed. In practice, IBM hardware limited $M \leqslant 8192$, and so we fix $M = 8192$ for all devices for the benchmark results we present in this paper.

When we run our benchmark for increasing L, we note the first length occurrence $L^* + 1$ for which each device has an error score greater than 10, and we report that the device passes the benchmark up to a length L^*. The threshold value of 10 was chosen arbitrarily, although it is consistent with bounding the LINPACK requirement that the error score should
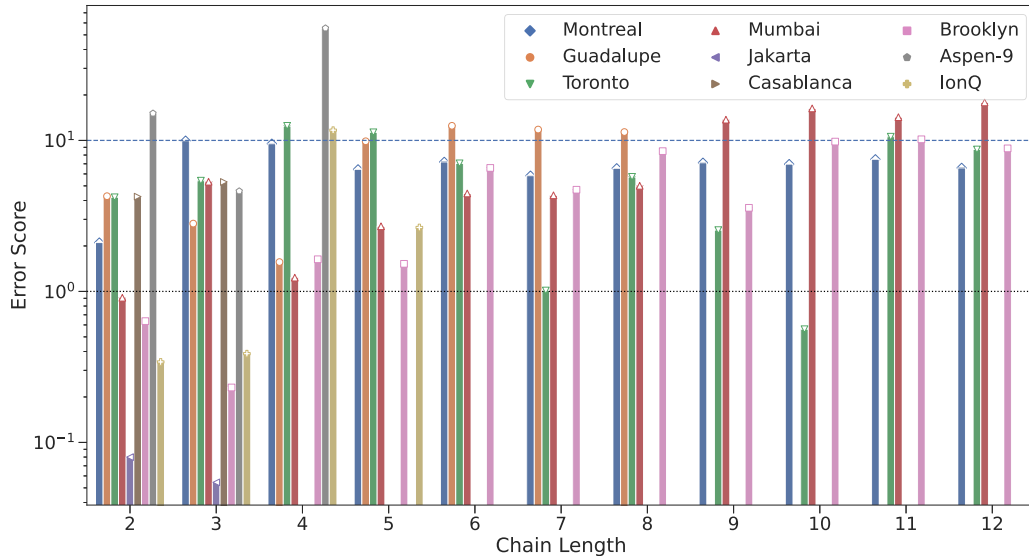
FIG. 6. Error score of each device, grouped by chain length (a lower error score indicates a better result). A relaxed error threshold (blue dashed line) is shown at a value of 10, and an aggressive error threshold (black dashed line) is shown at a value of 1. For each device, when the first violation of the chosen threshold is found, the device is said to have failed the benchmark. Symbols above each bar are for easy identification of their associated bar along with the accompanying legend.

have $O(1)$. It may be useful to consider other, tighter threshold values in the future; however, this alteration of the benchmark should be reported clearly for the sake of fair comparison. While we choose Eq. (4) to be defined in terms of the classically known solution, this is not necessarily required. For a ground-state energy calculation where the finite-size solutions are known to monotonically approach the asymptotic value, one could instead define a threshold requirement based on the monotonically decreasing behavior of the experimentally determined energy, even without knowing the exact asymptotic target value.

We capture the error score for each device and chain length in Fig. 6. This rescaling of the data presented in Fig. 5 helps to visualize when each device exceeds our chosen threshold. While not all devices increase in error monotonically, Fig. 6 shows that they nevertheless slowly trend toward and past the error-threshold line. Notably, most devices are able to pass the suggested score threshold of 10, but very few approach a score of 1 after more than a handful of chain lengths. This marks a significant goal for improvement for current NISQ devices.

IBM's Montreal device, which has one of the largest quantum volumes of all devices we tested, also performs very well with respect to our score threshold, based on a combination of large qubit number and low noise. The Montreal device results improved significantly upon applying SPAM mitigation, but others, e.g., IonQ, improve less noticeably. This is expected as the IonQ device typically has significantly smaller SPAM errors than the superconducting devices. See SI A 4 for further details regarding SPAM mitigation. We also wish to highlight the smaller IBM devices, Jakarta and Casablanca, which perform well within this benchmark up to their maximum allowed size. We suggest that the lacking performance of the Rigetti Aspen-9 device is due to errors which are not corrected through simple SPAM mitigation techniques and would require more rigorous error-mitigation strategies.

IV. CONCLUSION

We have proposed and demonstrated an application-specific quantum benchmark that is well suited for current and near-term NISQ devices. We have shown that our proposed benchmark scales in small increments of the required quantum resources, giving it the ability to discriminate devices finely. The choice to restrict the problem to a single-particle subspace of the 1D Fermi Hubbard model also allows us to exactly solve for the energy at arbitrary L, making the benchmark easy to verify classically at any size.

We applied a well-conditioned symmetry-based ansatz and SPAM error-mitigation strategy in order to improve the results of the energy estimation. This ansatz is hardware agnostic and easily defined for arbitrary L, and we showed analytically that it is always able to find the single-particle ground state of the supplied 1D Fermi Hubbard model. The primary result of our chosen benchmark clearly discriminates the performance of available quantum test beds in a way that tracks closely, but not exactly, with the more abstract "algorithmic qubits" metric. Analyses of the impact of SPAM error mitigation and the scaling of the ansatz performance as a function of length provided some clarity about the primary limitations of the devices. The result is a benchmark that is simple to implement and analyze and yet provides good quantitative discrimination between current-generation devices for a problem similar to a commercially relevant NISQ application.
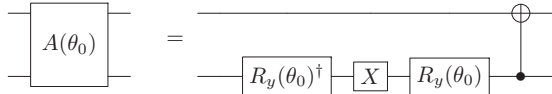
FIG. 7. Since the initial state is fixed, we can simplify the first A gate (the topmost one) in the general circuit shown in Fig. 1.

APPENDIX

1. A gate simplification

Our efficient ansatz is constructed from a single X gate and many primitive A gates. Due to the construction of the circuit, we can always simplify the general decomposition of the A gate presented in the main text without changing the overall calculation implemented by the circuit. Since the circuit always begins with an X gate on the first qubit, we can simplify the first A gate as shown in Fig. 7. We can simplify all other A gates as shown in Fig. 8. These simplifications are significant improvements for current NISQ devices, which are heavily limited by two-qubit gate error rates.

2. Exact ansatz

In the main text we used an ansatz that is composed of particle-conserving gates that can be decomposed into two or fewer CNOT gates for the fixed input state we have considered. It is straightforward to show that the ansatz described in the main text is always able to produce any single-particle state and, consequently, that optimization over the parameters of the ansatz can always find the single-particle ground-state energy. Consider the first application of an A gate onto the fixed input state defined on n qubits. This first interaction results in the state (up to a global phase)

$$A_{0,1}|10\cdots0\rangle = \sin\theta_0|10\cdots0\rangle + \cos\theta_0|01\cdots0\rangle \equiv |s_0\rangle,$$

where $A_{i,i+1}$ is the gate acting on qubits i, $i+1$ and has corresponding parameter θ_i. The second A gate has a similar action and produces the state

$$\begin{aligned}A_{1,2}|s_0\rangle = {}& \sin\theta_0|100\cdots0\rangle \\ & + \cos\theta_0\sin\theta_1|010\cdots0\rangle \\ & + \cos\theta_0\cos\theta_1|001\cdots0\rangle.\end{aligned}$$

For the general case of our ansatz on n qubits, we can compactly write the single-particle (in lexicographical order)
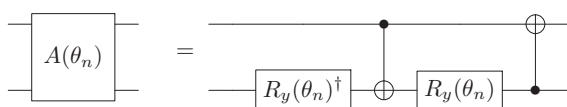

FIG. 8. Simplification of all A ($n > 0$) gates in the general circuit except for the first gate, which follows the simplification shown in Fig. 7.

coefficients as

$$\begin{aligned}c_0 &= \sin\theta_0, \\ c_1 &= \cos\theta_0\sin\theta_1, \\ c_2 &= \cos\theta_0\cos\theta_1\sin\theta_2, \\ &\vdots \\ c_{n-1} &= \cos\theta_0\cdots\cos\theta_{n-1}\sin\theta_n, \\ c_n &= \cos\theta_0\cdots\cos\theta_{n-1}\cos\theta_n\end{aligned}$$

for $n \geqslant 2$. We can see that repeated applications of the ladder of A gates in the ansatz leads to a state whose coefficients have the form of the coordinates of an n-dimensional hypersphere. Since this construction can create any valid hyperspherical coordinates, the ansatz is able to specify any superposition of single-particle states with real coefficients. This set of states is isomorphic to the target (restricted) subspace for our Fermi-Hubbard problem, so we guarantee that a correct choice of θ_i is capable of creating the ground state of the subspace. Assuming the optimizer software works in an idealized way, it will be capable of finding this ground state and the corresponding minimum energy.

3. Benchmark extensions

We have chosen to restrict the Fermi length benchmark to the single-particle and spin-up symmetry subspace. The minimum-energy single-particle state may not be the minimum energy over all particle numbers. Note that our SWAP-like ansatz and fixed initial state ensure that the ansatz cannot generate any states other than single-particle states. Therefore, generation of multiparticle states occurs only due to quantum errors, and, in their absence, optimizing over the allowed parameters optimizes over only the restricted subspace. This restricted problem requires fewer quantum resources and has a simple closed-form solution that we use to simplify the benchmark evaluation. For the one-particle subspace we require only that the top half of the qubits is subjected to gates. This simplification is allowed because the one-particle subspace is made up of two degenerate parts corresponding to spin up and spin down, and either one requires only half of the qubits in our chosen representation.

In the fully general case, we can segregate any Hamiltonian which has a well-defined particle-number symmetry into spaces based on these particle numbers. The size of the segregated subspaces naturally obeys the form of $\sum_{m=0}^{n}\binom{n}{m} = 2^n$. Therefore, the full Hilbert space of size 2^n is broken up into sets of size $\binom{n}{m}$ for $m \leqslant n$ particles. It is easy to see that the single-particle subspace is linear in the system size $\binom{n}{1} = O(n)$, while other subspaces grow much faster. Notably, the largest subspace is given by the so-called half-filling case $m = n/2$, which grows exponentially in n.

The subspace size gives us a rough characterization of the classical computational complexity if the solution method is assumed to be brute-force matrix diagonalization within each subspace. Using sparse methods to diagonalize a size $2^n \times 2^n$ matrix within a smaller block of size $O(n^m)$ takes classical resources (time and memory) that scale polynomially in n for a fixed particle-number subspace m and exponentially as m increases, for example, if we fix $m = n/2$.

For each choice of subspace, we can also quantify the minimum number of parameters required to fully specify a quantum state within this space. The single-particle subspace can be written as a linear combination of single-particle states defined by

$$|s\rangle = \alpha |10 \cdots 0\rangle + \beta |01 \cdots 0\rangle + \cdots + \zeta |00 \cdots 1\rangle.$$

Therefore, any real state of the above form can be fully specified with a minimal parameter count of $\binom{n}{m} - 1 = \binom{n}{1} - 1 = n - 1$ real parameters (minus one due to the normalization requirement). We are not aware of a quantum ansatz which is capable of matching this minimal parameter count in general. However, an ansatz for less trivial spaces remains an active area of research. Further symmetries of a given Hamiltonian may allow further reductions of these parameter counts, in principle.

If we extend the benchmark to the two-particle case and a 2D lattice structure, there are now two nondegenerate spin subspaces: one where both particles have the same spin (e.g., $s_z = \pm 1$) and one where they have opposite spins ($s_z = 0$). In the $s_z = 1$ case, we still require gates on only the top half of the qubits, but the $s_z = 0$ case requires gates on all qubits.

Results from previous investigations of small problems in quantum chemistry show that a similar ansatz performs well in this two-particle subspace [28,32]. Therefore, we can extend the protocol used for the benchmark to the evaluation of symmetry subspaces for which an exact classical solution is not necessarily known. Because we want to make this extension straightforward, we insist on using the entire space and not just the spin-up subspace in our benchmark protocol. The cost of this is that the benchmark requires twice as many qubits as are truly necessary to solve the restricted problem. This choice is typical of the trade-off between simplification and application specificity that must be negotiated for any benchmark.

4. SPAM mitigation

In order to mitigate errors that occur during state preparation and measurement, we employ a simple construction that maps ideal, intended states to noisy output states using the equation

$$M_{ij}x_k = y_k, \tag{A1}$$

where M_{ij} is a scattering matrix relating the ideal input states x_k to their measured noisy output states y_k [49]. Constructing this scattering matrix without restrictions would require measuring independently all 2^N input bit strings for an N-qubit system. This is possible, and very effective, for small systems but becomes untenable rapidly as the number of qubits exceeds 10. However, we can reduce this overhead by assuming that SPAM errors for each qubit are uncorrelated with other qubits. Under this assumption, we need to prepare only two measurement circuits, the all-zero state and the all-one state, independent of the number of qubits in the system. The results allow us to populate N sets of size 2×2 scattering matrices M_k. We then construct the full scattering matrix by $M_{ij} = M_1 \bigotimes M_2 \bigotimes \cdots M_N$.

For SPAM mitigation, we intend to relate the noisy output states back to ideal input states; therefore, we need to calculate M_{ij}^{-1} (or its pseudoinverse in the case of singular values). For the factorized case described above this inversion is not a challenging task. However, actually carrying out the $M_{ij}^{-1}y_k$ operation is a large calculation in our implementation, even though we avoid explicitly constructing the full $2^N \times 2^N$ matrix. This calculation is the current bottleneck in our classical postprocessing and limits SPAM mitigation to problems with fewer than 26 qubits.

To further illustrate the impact of the technique, we show in Fig. 9 the energy correction (difference) between our raw results and the SPAM mitigated results. Figure 9 captures some information about the nature of the errors on the devices;
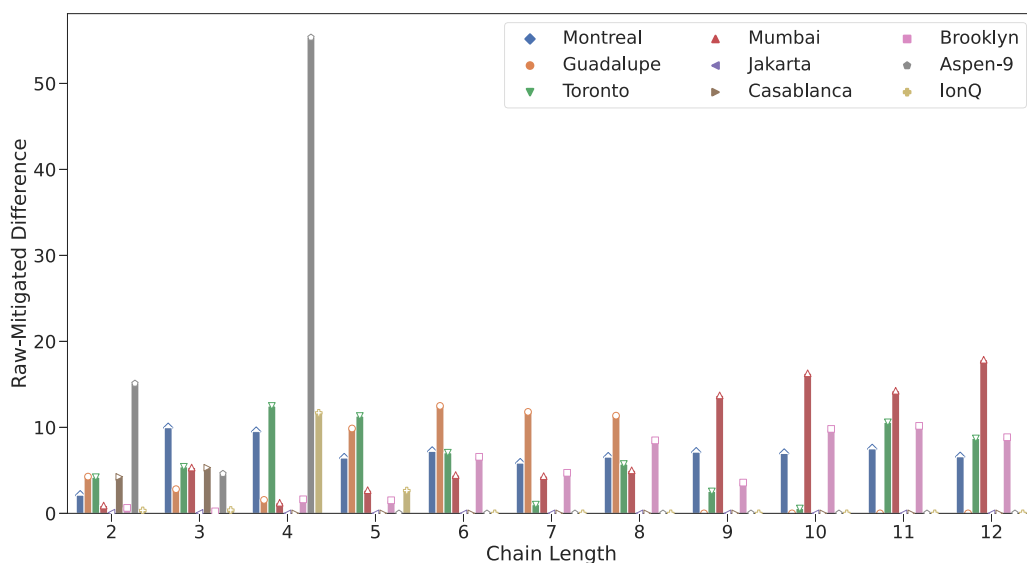

FIG. 9. Energy differences between each device's raw results and its SPAM-mitigated results. This difference characterizes the magnitude of errors which arise due to uncorrelated SPAM errors. Large, positive differences indicate that the mitigation results in a largely improved energy measurement after mitigation. Not all devices are present at each chain length due to varying physical system size. Symbols above each bar are for easy identification of their associated bar along with the accompanying legend.

TABLE II. List of benchmark devices along with some of their specifications taken from their respective cloud service providers. Gate errors are measured from standard randomized benchmarking techniques and are averaged over all qubits (or pairs for 2-QB gates). All specifications other than qubit number may drift over time, and therefore, these values are only a snapshot, taken on 20 August 2021, which is near the same date as our Fermi-Hubbard-based benchmark. A layout of "HH" denotes a heavy-hexagonal topology and is the preferred topology of all current IBM Quantum (IBMQ) devices. We use standard abbreviations for qubit (QB) and readout (RO). In addition, T1 and T2 describe decoherence and dephasing time, respectively.

	Jakarta	Casablanca	Guadalupe	Toronto	Mumbai	Montreal	Brooklyn	IonQ	Aspen-9
Qubits	7	7	16	27	27	27	65	11	32
1-QB gate error (%)	0.03	0.03	0.04	0.05	0.03	0.03	0.04	0.27	1.12
2-QB gate error (%)	0.98	0.97	1.02	1.06	0.83	1.31	1.09	3.04	11.87
Average RO error (%)	2.58	2.07	2.08	3.77	3.96	1.96	2.37	0.02	11.32
Average T1 (μs)	113.88	100.26	97.71	105.99	115.04	106.16	65.79	10000	29.81
Average T2 (μs)	34.23	115.76	107.87	117.30	114.47	78.96	79.21	0.2	18.09
Service	IBMQ	IBMQ	IBMQ	IBMQ	IBMQ	IBMQ	IBMQ	AWS	AWS
Layout	HH	HH	HH	HH	HH	HH	HH	Linear	Ring
Software	QISKIT	QISKIT	QISKIT	QISKIT	QISKIT	QISKIT	QISKIT	PENNYLANE	PENNYLANE

for example, the IonQ device corrections are much smaller than those for the Rigetti device, and this is consistent with the difference in self-reported measurement error rates between the two devices. We also see that there is a significant component of the error in each case that our mitigation does not remove. These include gate errors and SPAM errors due to correlating or drifting conditions.

5. Experimental details

In our experimental runs we accessed nine different quantum devices from three different vendors through two different online, cloud-based services. Specifications for each IBM device can be found in Ref. [50]. However, we note that some specifications of the devices change from day to day as calibrations are updated unpredictably. Therefore, we include here a summary of specifications from each device obtained during the time period of our experiments. Table II shows a summary of device specifications for each tested device.

For brevity and to follow the reporting of IBM Quantum, we list average qubit specifications, individual qubit specifications would require an exceedingly large number of statistics to report and would still not represent the variability over time. We note that the IonQ device is capable of directly entangling any pair of ions in its chain, without any associated overhead of swap operations. This all-to-all connectivity does not manifest as an advantage in our benchmark because we utilize only nearest-neighbor connectivity, which all tested devices support. In principle, all-to-all connectivity should provide an advantage for related Fermi-Hubbard simulations, but attempting to penalize nearest-neighbor connectivity for this specific simulation would complicate the benchmark artificially.

For each device we utilized the packaged VQE algorithm in each code base with our custom ansatz, discussed in the main text. In all cases, we fix the number of shots to 8192, and

the required evaluations in order to measure $\langle \hat{H} \rangle$ is five; therefore, each quantum hardware is queried $8192 \times 5 = 40\,960$ times for each choice of L (independent of L). Note that one can measure $\langle H \rangle$ with only single-qubit measurements as H is factorized into five tensored product basis sets and does not generally require multiqubit measurements [25]. In Fig. 10 we provide an example layout of a typical IBM chip layout configuration (Ibmq Guadalupe) and the shape of the heavy-hexagonal layout. Also included in Fig. 10 is the ring layout of Rigetti's Aspen-9 chip showing their connected ring configuration.
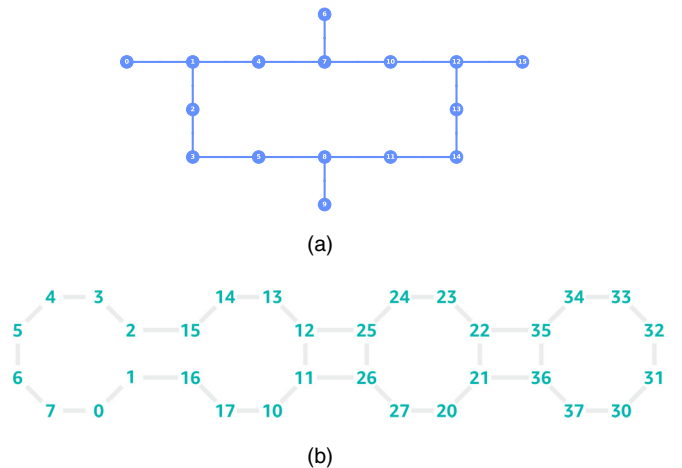

FIG. 10. Layout of two example quantum devices. (a) Layout connectivity of the IBMQ Guadalupe device in a heavy hexagonal configuration. Other IBM devices follow a similar layout but vary in the number of qubits. (b) Layout of the Rigetti Aspen-9 device in a connected ring layout.

[1] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[2] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B **7**, S347 (2005).

[3] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).

[4] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, Phys. Rev. A **77**, 012307 (2008).

[5] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, Characterization of Addressability by Simultaneous Randomized Benchmarking, Phys. Rev. Lett. **109**, 240504 (2012).

[6] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, Phys. Rev. A **100**, 032328 (2019).

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

[8] R. Blume-Kohout and K. C. Young, A volumetric framework for quantum computer benchmarks, Quantum **4**, 362 (2020).

[9] Ionq.com (2022), https://ionq.com/posts/december-09-2020-scaling-quantum-computer-roadmap.

[10] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, Quantum chemistry in the age of quantum computing, Chem. Rev. **119**, 10856 (2019).

[11] F. Arute *et al.*, Hartree-Fock on a superconducting qubit quantum computer, Science **369**, 1084 (2020).

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

[13] E. Farhi and A. W. Harrow, Quantum supremacy through the quantum approximate optimization algorithm, arXiv:1602.07674.

[14] G. Nannicini, Performance of hybrid quantum-classical variational heuristics for combinatorial optimization, Phys. Rev. E **99**, 013304 (2019).

[15] J. Rajakumar, J. Moondra, S. Gupta, and C. D. Herold, Generating target graph couplings for qaoa from native quantum hardware couplings, arXiv:2011.08165.

[16] S. Martiel, T. Ayral, and C. Allouche, Benchmarking quantum coprocessors in an application-centric, hardware-agnostic, and scalable way, IEEE Trans. Quantum Eng. **2**, 3102011 (2021).

[17] I. Buluta and F. Nori, Quantum simulators, Science **326**, 108 (2009).

[18] J. W. Britton, B. C. Sawyer, A. C. Keith, C.-C. J. Wang, J. K. Freericks, H. Uys, M. J. Biercuk, and J. J. Bollinger, Engineered two-dimensional Ising interactions in a trapped-ion quantum simulator with hundreds of spins, Nature (London) **484**, 489 (2012).

[19] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2012).

[20] M. Schuld, I. Sinayskiy, and F. Petruccione, An introduction to quantum machine learning, Contemp. Phys. **56**, 172 (2015).

[21] P.-L. Dallaire-Demers, M. Stęchły, J. F. Gonthier, N. T. Bashige, J. Romero, and Y. Cao, An application benchmark for fermionic quantum simulations, arXiv:2003.01862.

[22] A. H. Karamlou, W. A. Simon, A. Katabarwa, T. L. Scholten, B. Peropadre, and Y. Cao, Analyzing the performance of variational quantum factoring on a superconducting quantum processor, npj Quantum Inf. **7**, 156 (2021).

[23] T. Lubinski, S. Johri, P. Varosy, J. Coleman, L. Zhao, J. Necaise, C. H. Baldwin, K. Mayer, and T. Proctor, Application-oriented performance benchmarks for quantum computing, arXiv:2110.03137.

[24] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[25] 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).

[26] D. Wang, O. Higgott, and S. Brierley, Accelerated Variational Quantum Eigensolver, Phys. Rev. Lett. **122**, 140504 (2019).

[27] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martínez, Quantum Computation of Electronic Transitions Using a Variational Quantum Eigensolver, Phys. Rev. Lett. **122**, 230401 (2019).

[28] B. T. Gard, L. H. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm, npj Quant. Inf. **6**, 10 (2020).

[29] A. J. McCaskey, Z. P. Parks, J. Jakowski, S. V. Moore, T. D. Morris, T. S. Humble, and R. C. Pooser, Quantum chemistry as a benchmark for near-term quantum computers, npj Quantum Inf. **5**, 99 (2019).

[30] 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).

[31] G. S. Barron, B. T. Gard, O. J. Altman, N. J. Mayhall, E. Barnes, and S. E. Economou, Preserving Symmetries for Variational Quantum Eigensolvers in the Presence of Noise, Phys. Rev. Appl. **16**, 034003 (2021).

[32] K. Yeter-Aydeniz, B. T. Gard, J. Jakowski, S. Majumder, G. S. Barron, G. Siopsis, T. S. Humble, and R. C. Pooser, Benchmarking quantum chemistry computations with variational, imaginary time evolution, and Krylov space solver algorithms, Adv. Quantum Technol. **4**, 2100012 (2021).

[33] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, S. Filipp, and I. Tavernelli, Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions, Phys. Rev. A **98**, 022322 (2018).

[34] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, Quantum Simulation of Electronic Structure with Linear Depth and Connectivity, Phys. Rev. Lett. **120**, 110501 (2018).

[35] Z. Jiang, K. J. Sung, K. Kechedzhi, V. N. Smelyanskiy, and S. Boixo, Quantum Algorithms to Simulate Many-Body Physics of Correlated Fermions, Phys. Rev. Appl. **9**, 044036 (2018).

[36] T. Takeshita, N. C. Rubin, Z. Jiang, E. Lee, R. Babbush, and J. R. McClean, Increasing the Representation Accuracy of Quantum Simulations of Chemistry without Extra Quantum Resources, Phys. Rev. X **10**, 011004 (2020).

[37] A. Montanaro and S. Stanisic, Compressed variational quantum eigensolver for the Fermi-Hubbard model, arXiv:2006.01179.

[38] E. Kökcü, D. Camps, L. Bassman, J. K. Freericks, W. A. de Jong, R. Van Beeumen, and A. F. Kemper, Algebraic compression of quantum circuits for Hamiltonian evolution, Phys. Rev. A **105**, 032420 (2022).

[39] E. Fradkin, Jordan-Wigner Transformation for Quantum-Spin Systems in Two Dimensions and Fractional Statistics, Phys. Rev. Lett. **63**, 322 (1989).

[40] E. H. Lieb and F. Wu, The one-dimensional Hubbard model: A reminiscence, Physica A (Amsterdam, Neth.) **321**, 1 (2003).

[41] V. E. Korepin and F. H. L. Eßler, *Exactly Solvable Models of Strongly Correlated Electrons* (World Scientific, Singapore, 1994).

[42] J. S. Van Dyke, G. S. Barron, N. J. Mayhall, E. Barnes, and S. E. Economou, Preparing Bethe Ansatz Eigenstates on a Quantum Computer, PRX Quantum **2**, 040329 (2021).

[43] A. Osipov, Evaluation of small elements of the eigenvectors of certain symmetric tridiagonal matrices with high relative accuracy, Appl. Comput. Harmonic Anal. **43**, 173 (2017).

[44] A. Uvarov, J. D. Biamonte, and D. Yudin, Variational quantum eigensolver for frustrated quantum systems, Phys. Rev. B **102**, 075104 (2020).

[45] A. G. Rattew, S. Hu, M. Pistoia, R. Chen, and S. Wood, A domain-agnostic, noise-resistant, hardware-efficient evolutionary variational quantum eigensolver, arXiv:1910.09694.

[46] Qiskit/qiskit: Qiskit is an open-source SDK for working with quantum computers at the level of circuits, algorithms, and application modules (2022), https://github.com/qiskit/qiskit.

[47] Y. Suzuki, Y. Kawase, Y. Masumura, Y. Hiraga, M. Nakadai, J. Chen, K. M. Nakanishi, K. Mitarai, R. Imai, S. Tamiya, T. Yamamoto, T. Yan, T. Kawakubo, Y. O. Nakagawa, Y. Ibe, Y. Zhang, H. Yamashita, H. Yoshimura, A. Hayashi, and K. Fujii, Qulacs: A fast and versatile quantum circuit simulator for research purpose, Quantum **5**, 559 (2021).

[48] J. J. Dongarra, P. Luszczek, and A. Petitet, The LINPACK benchmark: Past, present and future, Concurrency Comput.: Pract. Exper. **15**, 803 (2003).

[49] Qiskit.org. (2022), Qiskit Measurement Error Mitigation, Available at: https://qiskit.org/textbook/ch-quantum-hardware/measurement-error-mitigation.html.

[50] IBM Quantum (2022), https://quantum-computing.ibm.com/.