TETRIS-ADAPT-VQE: An adaptive algorithm that yields shallower, denser circuit *Ansätze*

Panagiotis G. Anastasiou,[1,2,*] Yanzhu Chen,[1,2] Nicholas J. Mayhall,[3,2] Edwin Barnes,[1,2] and Sophia E. Economou[1,2,†]

[1]*Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA*

[2]*Virginia Tech Center for Quantum Information Science and Engineering, Blacksburg, Virginia 24061, USA*

[3]*Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, USA*

(Received 1 October 2022; revised 7 November 2023; accepted 3 January 2024; published 7 March 2024)

Adaptive quantum variational algorithms are particularly promising for simulating strongly correlated systems on near-term quantum hardware, but they are not yet viable due, in large part, to the severe coherence time limitations on current devices. In this paper, we introduce an algorithm called TETRIS-ADAPT-VQE (tiling efficient trial circuits with rotations implemented simultaneously adaptive derivative-assembled problem-tailored *Ansatz* variational quantum eigensolver), which iteratively builds up variational *Ansätze* a few operators at a time in a way dictated by the problem being simulated. This algorithm is a modified version of the ADAPT-VQE algorithm, in which the one-operator-at-a-time rule is lifted to allow for the addition of multiple operators with disjoint supports in each iteration. TETRIS-ADAPT-VQE results in denser but significantly shallower circuits, without increasing the number of controlled-NOT gates or variational parameters. Its advantage over the original algorithm in terms of circuit depths increases with the system size. Moreover, the expensive step of measuring the energy gradient with respect to each candidate unitary at each iteration is performed only a fraction of the time compared with ADAPT-VQE. These improvements bring us closer to the goal of demonstrating a practical quantum advantage on quantum hardware.

DOI: 10.1103/PhysRevResearch.6.013254

I. INTRODUCTION

It is widely hoped that noisy intermediate-scale quantum (NISQ) devices can find applications in the field of quantum simulation, where the archetypal problem is to find the eigenvalues of a given Hamiltonian [1]. To avoid the large resource overhead of the quantum phase estimation algorithm (PEA) [2], which involves a large number of qubits coherently evolving under very deep circuits [3] and is thus not expected to offer an advantage over classical simulation in the near future, a hybrid quantum-classical algorithm, the variational quantum eigensolver (VQE), was introduced and experimentally realized for small chemical systems; this algorithm aims to shorten quantum circuits by leveraging the power of classical optimization [4].

Based on the variational principle, the VQE prepares a parametrized guess wave function (known as an *Ansatz*) using a quantum circuit on the quantum processor. It then iteratively optimizes the variational parameters in order to minimize an objective function, usually the energy of the system being simulated. In each iteration, any evaluation of the objective function and/or its gradient is realized by measurements on the quantum processor, and some classical optimization scheme is chosen to update the parameters on the classical processor. Since in the VQE the quantum processor is only needed for measuring observables of the *Ansatz* state, the quantum circuits involved are shorter compared with the PEA. The success of VQEs crucially depends on a good choice of *Ansatz*. On the one hand, the *Ansatz* should be expressive enough so that the ground state of the Hamiltonian can be accurately approximated, and on the other hand it should be special enough so that classical optimization of the variational parameters can be efficiently carried out. Finally, the *Ansatz* should also be hardware friendly so that the trial state can be successfully prepared as a quantum circuit on a NISQ processor.

Depending on which of the above properties they are designed to address, the most commonly used *Ansätze* can be roughly divided into two families: the hardware-efficient [5–7] and the chemically inspired *Ansätze* [8–10]. The former include circuits consisting of repeating layers of parametrized single-qubit rotation gates and entangling gates, which are easy to implement on a given quantum processor. Such designs take very general forms to ensure expressivity. Nevertheless, it has been shown that a completely problem-agnostic VQE *Ansatz* would hinder the classical optimization [11]. In contrast, the latter are inspired by classical computational chemistry and consist of much more complicated circuits. A widely used *Ansatz* in this category is the unitary extension of coupled-cluster singles and doubles (UCCSD), which forms the basis of many others. They are known from classical computational chemistry to possess several desirable features (such as being both variational and size extensive); however, they correspond to relatively deep circuits that are difficult to

*panastasiou@vt.edu

†economou@vt.edu

realize on devices with limited coherence times [12], and their performance is sensitive to the specific Trotterization used in their implementation [13].

Among the numerous strategies to find a middle ground between the two extremes [5,14], some adaptive approaches seek to construct the *Ansatz* based on information gathered while running the algorithm. The earliest such algorithm was developed by Grimsley *et al.* [15], where, in contrast to conventional VQEs, the *Ansatz* is grown dynamically instead of chosen *a priori*. This algorithm, adaptive derivative-assembled problem-tailored *Ansatz* variational quantum eigensolver (ADAPT-VQE), directly incorporates problem features into the *Ansatz*, foregoing the maximum level of expressivity in favor of greater efficiency in the classical optimization, while at the same time retaining some degree of flexibility to avoid deep circuits.

The central idea of ADAPT-VQE is to add one operator at a time to the *Ansatz*, with each new addition followed by classical optimization of the current *Ansatz* as in conventional VQEs. Each new operator is selected from a predefined operator pool based on the gradient of the objective function, which is determined by the problem Hamiltonian together with the previously optimized state. This makes the *Ansatz* inherently specific to the problem. It was shown that ADAPT-VQE results in shorter *Ansätze* with fewer variational parameters than UCCSD, except for certain strongly correlated molecules for which UCCSD fails to reach chemical accuracy. Using a more hardware-friendly operator pool, qubit-ADAPT-VQE [16] requires even shallower circuits and fewer controlled-NOT (CNOT) gates. Moreover, by constructing a problem-specific *Ansatz*, ADAPT-VQE is less prone to obstacles in the classical optimization compared with more hardware-efficient but problem-agnostic *Ansätze* [17].

Despite the considerable reduction in circuit depth it offers compared with other VQEs, ADAPT-VQE still requires circuit depths on the order of tens of thousands to accurately estimate ground-state energies of small molecules, which lies beyond the capacity of most quantum devices to this day [18,19]. This is due in part to the fact that each pool operator acts on only a small subset of qubits, leaving most qubits idle and unexploited at any given point in the construction of the state preparation circuit. Moreover, the *Ansatz* construction introduces additional overhead costs. Specifically, in each iteration of the algorithm, the gradient of the objective function needs to be measured for each operator in the pool. While this overhead is unlikely to be the bottleneck in near-term applications, it is desirable to reduce it in the long run, when the system size of the computation task grows significantly.

In this paper, we propose a variant of ADAPT-VQE, named tiling efficient trial circuits with rotations implemented simultaneously (TETRIS)-ADAPT-VQE. We achieve the goal of further reducing the circuit depth by allowing unitary operators that act on disparate sets of qubits to be added to the *Ansatz* simultaneously at each iteration. This results in a faster reduction of the objective function. We demonstrate this protocol using numerical simulations for a range of molecules with different geometries. We show that despite the much shallower quantum circuits, TETRIS-ADAPT can successfully produce results with the same accuracy as the original ADAPT-VQE.

The rest of the paper is organized as follows. In Sec. II, we review in detail the procedure of ADAPT-VQE, as it makes up the backbone of the algorithm we propose. We then present the strategy proposed here, including a discussion of what operator pools to use and how to translate the sequence of unitaries in the *Ansatz* to a circuit with quantum gates. Section III contains the details and results of the numerical simulations that demonstrate the improvement TETRIS-ADAPT-VQE can provide. Compared with ADAPT-VQE, it further reduces the resources needed to reach a certain level of accuracy while keeping the advantages of ADAPT-VQE. Finally, we conclude in Sec. IV.

II. ALGORITHM DETAILS

Although our algorithm is quite general, in this paper we focus on its application to problems in quantum computational chemistry. Computational chemistry is largely concerned with finding the ground-state energy of the electronic part of molecular Hamiltonians:

$$\hat{H} = \sum_{p,q} h_q^p a_p^\dagger a_q + \frac{1}{2} \sum_{p,q,r,s} h_{sr}^{pq} a_p^\dagger a_q^\dagger a_r a_s, \tag{1}$$

where h_q^p and h_{sr}^{pq} are single- and two-electron integrals, respectively, and $a^\dagger$ and a are the second-quantized fermionic creation and annihilation operators, respectively. Our algorithm takes the same adaptive procedure as in ADAPT-VQE, with an alternative strategy of selecting the operators that contribute to the VQE *Ansatz*. Like ADAPT-VQE, the operator pools draw inspiration from the UCCSD *Ansatz*, which is briefly reviewed in Sec. II A. Readers familiar with the topic can jump to Sec. II B.

A. Unitary coupled cluster

The UCCSD *Ansatz* has the form

$$|\Psi_{\text{UCCSD}}\rangle = e^{\hat{T}_1 + \hat{T}_2} |\Psi_{\text{ref}}\rangle , \tag{2}$$

where

$$\hat{T}_1 = \sum_{i,a} \hat{\tau}_i^a = \sum_{i,a} \tau_i^a (a_a^\dagger a_i - a_i^\dagger a_a), \tag{3}$$

$$\hat{T}_2 = \sum_{i,j,a,b} \hat{\tau}_{ij}^{ab} = \sum_{i,j,a,b} \tau_{ij}^{ab} (a_a^\dagger a_b^\dagger a_i a_j - a_j^\dagger a_i^\dagger a_b a_a). \tag{4}$$

In practice the unitary is Trotterized to first order:

$$|\Psi_{\text{tUCCSD}}\rangle = \prod_{c \in \{ia\}} e^{\hat{\tau}_c} \prod_{d \in \{ijab\}} e^{\hat{\tau}_d} |\Psi_{\text{ref}}\rangle , \tag{5}$$

where i, j (a, b) represent occupied (virtual) orbitals and the reference state $|\Psi_{\text{ref}}\rangle$ is usually the Hartree-Fock state. From classical simulations, UCCSD is known to be a reliable *Ansatz*, although it was recently shown [13] that the low-order Trotterized form is not chemically well defined as it fails to reach chemical accuracy for certain operator orderings. Furthermore, due to the large number of exponential factors, it is too expensive to prepare on quantum processors. Lastly, for strongly correlated systems, UCCSD often fails to reach

chemical accuracy (error < 1 kcal/mol) for the ground-state energy [20].

B. ADAPT-VQE

The ADAPT-VQE algorithm grows problem-tailored *Ansätze* by appending one unitary at a time to the trial state [15]. The user begins by defining an operator pool $\mathcal{P} = \{P_i\}$, a collection of anti-Hermitian generators, such as the $\hat{\tau}_i^a$ and $\hat{\tau}_{ij}^{ab}$ in Eqs. (3) and (4), from which the *Ansatz* is to be constructed. In addition, a reference state is chosen, usually the Hartree-Fock (HF) ground state. At each iteration, the expectation value of the commutator of the Hamiltonian and each pool operator is measured. This is equivalent to adding each candidate operator to the *Ansatz*, setting its variational parameter equal to zero, and computing the energy gradient with respect to it. The operator with the largest gradient norm is appended to the *Ansatz*. For example, at the $(k+1)$th layer, with the previously optimized state $|\Psi^{(k)}\rangle$, the energy gradient for the candidate operator P_i with respect to its parameter θ_i is

$$\begin{aligned}\left.\frac{\partial E}{\partial \theta_i}\right|_{\theta_i=0} &= \left[\frac{\partial}{\partial \theta_i}\langle\Psi^{(k)}|\, e^{-\theta_i P_i} H e^{\theta_i P_i}\,|\Psi^{(k)}\rangle\right]\Bigg|_{\theta_i=0} \\ &= \langle\Psi^{(k)}|\,[H, P_i]\,|\Psi^{(k)}\rangle\,, \end{aligned} \tag{6}$$

using the anti-Hermiticity $P_i^\dagger = -P_i$. The algorithm proceeds as follows:

(1) On the classical device, compute one- and two-body integrals and transform the fermionic Hamiltonian to a qubit Hamiltonian using a suitable mapping, e.g., Jordan-Wigner (JW) mapping.

(2) On the quantum device, prepare the current *Ansatz* and measure the energy gradient $\frac{\partial E}{\partial \theta_i}|_{\theta_i=0}$ for every candidate pool operator P_i with respect to its variational parameter θ_i.

(3) If the pool gradient norm is smaller than a predetermined threshold, ADAPT-VQE has converged, and the algorithm terminates.

(4) Add the operator with the largest gradient norm from step (2) to the *Ansatz*, with its variational parameter set to zero.

(5) Perform a VQE subroutine to update *all* parameters in the current *Ansatz*.

(6) Repeat steps (2)–(5) until convergence.

C. TETRIS-ADAPT-VQE

In TETRIS-ADAPT-VQE, we add not only the operator associated with the largest gradient, but also the operator associated with the next-largest gradient norm that is supported on qubits *different* from the support of the first added operator. We continue this process of adding operators with successively smaller (or equal) gradients and with support that is disjoint from that of all previously added operators, until no further operators satisfying these criteria are found. Note that this can be done without measuring any additional gradients compared with ordinary ADAPT-VQE since the gradient of each pool operator need only be measured once per iteration of the algorithm as usual. This modification of the algorithm amounts to replacing the original step (4) from above with the following subroutine for adding operators to the *Ansatz*:

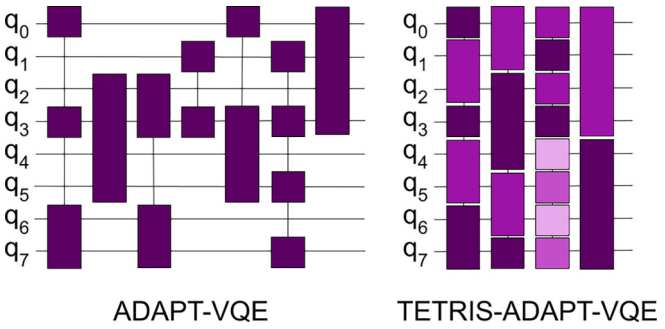

FIG. 1. Schematic diagram of the *Ansatz* preparation circuits produced by ADAPT-VQE (left) and TETRIS-ADAPT-VQE (right). TETRIS-ADAPT-VQE grows *Ansätze* prepared by shallower and denser circuits compared with ADAPT-VQE, by adding multiple operators at each iteration. The dark violet rectangles represent gates implementing the largest-gradient operator, whereas lighter tones correspond to operators beyond the highest-gradient one, added in TETRIS-ADAPT-VQE.

(a) Sort pool operators according to the norms of their gradients, in descending order.

(b) Identify the operator with the highest gradient norm acting on qubits not acted on by any previously added operator in the current ADAPT iteration. Add the operator to the *Ansatz*, with its variational parameter set to zero.

(c) If the operators added in the current ADAPT iteration collectively cover all qubits, proceed to step (5).

(d) If the list of operators with nonzero gradients and disjoint support with operators already added in the current ADAPT iteration has been exhausted, proceed to step (5).

(e) Go to substep (b).

We illustrate this strategy in Fig. 1, showing the locations of the gates implementing the operators chosen by the original ADAPT-VQE and TETRIS-ADAPT-VQE. Note that the highest-gradient operators coincide for the two versions of the algorithm in the first layer but differ in the subsequent layers, as a consequence of the diverging *Ansätze*.

D. Operator pools

The original ADAPT-VQE paper [15] compared ADAPT *Ansätze* with the commonly used, fixed UCCSD *Ansatz*, and as such it used pools of anti-Hermitian sums of single and double fermionic excitation-deexcitation operators of the forms $\hat{T}_{ij} = a_i^\dagger a_j - a_j^\dagger a_i$ and $\hat{T}_{ijkl} = a_i^\dagger a_j^\dagger a_k a_l - a_l^\dagger a_k^\dagger a_j a_i$. In the Jordan-Wigner mapping, which we use in the remainder of this paper, $\hat{T}_{ij}$ and $\hat{T}_{ijkl}$, for spin-orbital (and qubit) indices $i < j < k < l$, can be written in terms of sums of Pauli strings as

$$\hat{T}_{ij} = \frac{i}{2}(X_i Y_j - Y_i X_j) \prod_{p=i+1}^{j-1} Z_p, \tag{7}$$

$$\begin{aligned}\hat{T}_{ijkl} = \frac{i}{8}&(X_i Y_j X_k X_l + Y_i X_j X_k X_l + Y_i Y_j Y_k X_l \\ &+ Y_i Y_j X_k Y_l - X_i X_j Y_k X_l - X_i X_j X_k Y_l \\ &- Y_i X_j Y_k Y_l - X_i Y_j Y_k Y_l) \prod_{p=i+1}^{j-1} Z_p \prod_{r=k+1}^{l-1} Z_r, \end{aligned} \tag{8}$$

where the Pauli strings are tensor products of Pauli matrices on the corresponding qubits, and identities are implied on qubits with omitted indices. Although chemically sound, the large number of terms and linearly increasing number of qubits involved in these operators translate into deep circuits with large numbers of two-qubit gates (see Sec. II E) and make fermionic operators challenging for NISQ simulation.

In an effort to construct even more hardware-efficient *Ansätze*, Tang *et al.* [16] showed that decomposing the Jordan-Wigner mapped fermionic operators into individual Pauli strings and using those as pool operators results in shallower circuits, an approach dubbed qubit-ADAPT-VQE. In the same work, it was found that omitting the trailing Pauli-Zs from the Pauli strings did not affect convergence. At the expense of additional variational parameters, qubit-ADAPT-VQE yielded *Ansätze* with up to an order of magnitude fewer CNOT gates for the systems studied.

In a similar spirit, Yordanov *et al.* [21] introduced the qubit-excitation-based (QEB)-ADAPT-VQE, which uses single and double qubit excitations, obtained by replacing the fermionic creation and annihilation operators by the corresponding qubit operators $Q_i^\dagger = (X_i - iY_i)/2$ and $Q_i = (X_i + iY_i)/2$ satisfying

$$\{Q_i, Q_i^\dagger\} = 1, \quad [Q_i, Q_j^\dagger] = 0 \quad \text{for } i \neq j, \tag{9}$$

$$[Q_i, Q_j] = [Q_i^\dagger, Q_j^\dagger] = 0. \tag{10}$$

In light of Eq. (9), in the Jordan-Wigner mapping, this is equivalent to omitting the Pauli-Z strings in Eqs. (7) and (8). Using the optimized circuits in Ref. [22] to implement the qubit excitation (QE) evolutions resulted in *Ansätze* with CNOT counts similar to or lower than those of qubit-ADAPT-VQE and numbers of variational parameters comparable to those of fermionic ADAPT-VQE.

In this paper we use the qubit and QE pools because their operators act nontrivially on fewer qubits compared with those of the fermionic pool. In general, more operators can therefore be added at each TETRIS-ADAPT-VQE iteration. The qubit pool has generators of the form

$$iX_iY_j,\ iX_iX_jX_kY_l,\ iY_iY_jY_kX_l, \tag{11}$$

and we include all combinations of qubit indices such that their sum is an even number; that is, each qubit operator acts nontrivially on an even number of α (β) spin-orbitals. In the QE pool, the qubit-excitation generators for $\hat{\tilde{T}}_{ij} = Q_i^\dagger Q_j - Q_j^\dagger Q_i$ and $\hat{\tilde{T}}_{ijkl} = Q_i^\dagger Q_j^\dagger Q_k Q_l - Q_l^\dagger Q_k^\dagger Q_j Q_i$ in the Jordan-Wigner mapping have the form

$$\hat{\tilde{T}}_{ij} = \frac{i}{2}(X_iY_j - Y_iX_j), \tag{12}$$

$$\begin{aligned}\hat{\tilde{T}}_{ijkl} = \frac{i}{8}(&X_iY_jX_kX_l + Y_iX_jX_kX_l \\ &+ Y_iY_jY_kX_l + Y_iY_jX_kY_l - X_iX_jY_kX_l \\ &- X_iX_jX_kY_l - Y_iX_jY_kY_l - X_iY_jY_kY_l),\end{aligned} \tag{13}$$

where we include all combinations of indices such that no spin flips are allowed. For both pools it is implied that $i \neq j \neq k \neq l$, and many index permutations in the operators above differ only by an overall sign and need not be included in practice.

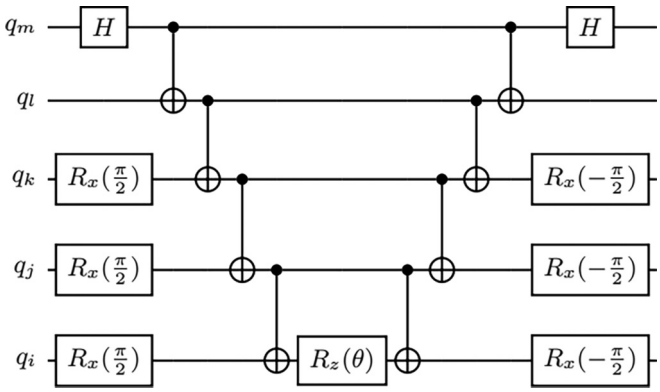

FIG. 2. Example circuit of the exponential map implementing $e^{-i\frac{\theta}{2}Y_iY_jY_kZ_lX_m}$.

E. Quantum circuits

To obtain the CNOT gate counts and circuit depths, we compile circuits effecting the operator evolutions in QISKIT [23]. For the qubit pool, which consists of individual Pauli strings, the exponentiation is performed with the usual exponential map circuit. For example, suppose that the operator $Z_a \otimes Z_b \otimes Z_c$ is to be exponentiated. Then a "CNOT staircase" computes the parities of the three qubits in the computational basis, an $R_z(\theta)$ gate applies the phase shift, and an inverse CNOT staircase uncomputes the parity. If the tensor product in the exponent contains X or Y Paulis, then those are rotated to the Z basis by H and $R_x(\frac{\pi}{2})$ gates, respectively. Figure 2 shows the circuit performing $e^{-i\frac{\theta}{2}Y_iY_jY_kZ_lX_m}$.

For the QE pool, we use the optimized circuits introduced in Refs. [21,22], derived by sequentially decomposing into opposite half rotations the multiqubit-controlled gates in exchange-interaction-type circuits, noting that single and double qubit excitation operators $\hat{\tilde{T}}_{ij}$ and $\hat{\tilde{T}}_{ijkl}$ continuously exchange states: $|1_i0_j\rangle$ with $|0_i1_j\rangle$ and $|1_i1_j0_k0_l\rangle$ with $|0_i0_j1_k1_l\rangle$. We show two example circuits in Fig. 3 taken from Ref. [21]. Throughout this paper, we assume all-to-all qubit connectivity; so CNOT gates between non-neighboring qubits can be performed.

III. RESULTS

We numerically simulate the performances of TETRIS-ADAPT-VQE and the original ADAPT-VQE for the tasks of finding the ground-state energy and preparing the ground state of the following molecules: H_4 (1.0–5.0 Å), LiH (1.0–4.0 Å), H_6 (1.0–5.0 Å), and BeH_2 (2.0–3.0 Å) at 1.0 Å bond length increments, all in the linear configuration, and using the STO-3G basis set (Slater-type orbital with three Gaussian orbitals). We will show that TETRIS-ADAPT-VQE provides additional improvement on both circuit depth and measurement overhead, without sacrificing the merits of the original ADAPT-VQE. The results in this section were produced using PYTHON code written by the authors. The classical optimization scheme used throughout this paper is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method as implemented in SCIPY [24], with a gradient norm tolerance of 10^{-10}. The ADAPT convergence criterion was a pool gradient norm threshold of 10^{-7}. All energies are in units of hartrees.

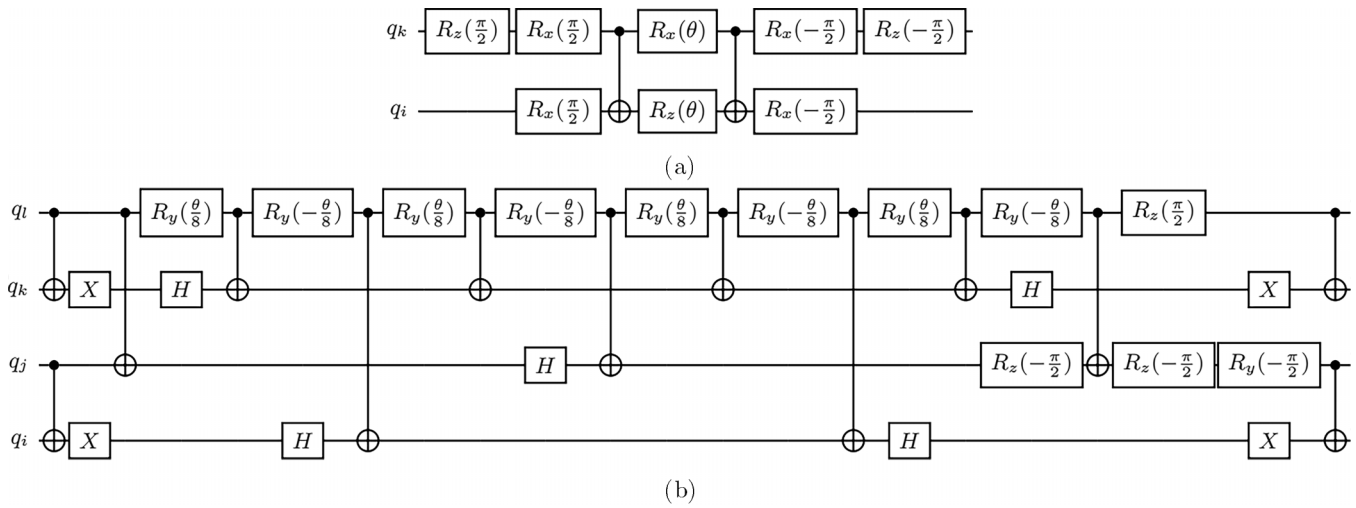

FIG. 3. Example circuits performing (a) a single qubit excitation, $e^{i\frac{\theta}{2}(X_iY_k-Y_iX_k)}$, and (b) a double qubit excitation evolution, $e^{\theta\hat{T}_{ijkl}}$, as in Eq. (13).

A. Circuit depth reduction

The main motivation of TETRIS-ADAPT-VQE is circuit depth reduction. Since ADAPT-VQE only adds the largest-gradient operator at each iteration, one may speculate that adding operators with smaller gradients alongside it would result in suboptimal *Ansätze* with a greater total number of parameters and CNOT gates. On the other hand, the magnitude of the operator gradient, which is measured about the point in parameter space where the VQE converged in the previous iteration, contains only local information about the ability of a candidate operator to reduce the value of the cost function. Put another way, the operator with the highest gradient is not guaranteed to be the operator that causes the greatest energy reduction. In Ref. [21] the authors explored the possibility of separately appending the n largest-gradient operators to the *Ansatz* at each ADAPT iteration, performing the VQE on all n trial *Ansätze*, and updating the main *Ansatz* with only the operator that resulted in the lowest energy. Moreover, adding operators beyond the highest-gradient one may enable further adjustments to the *Ansatz* that would otherwise appear in later

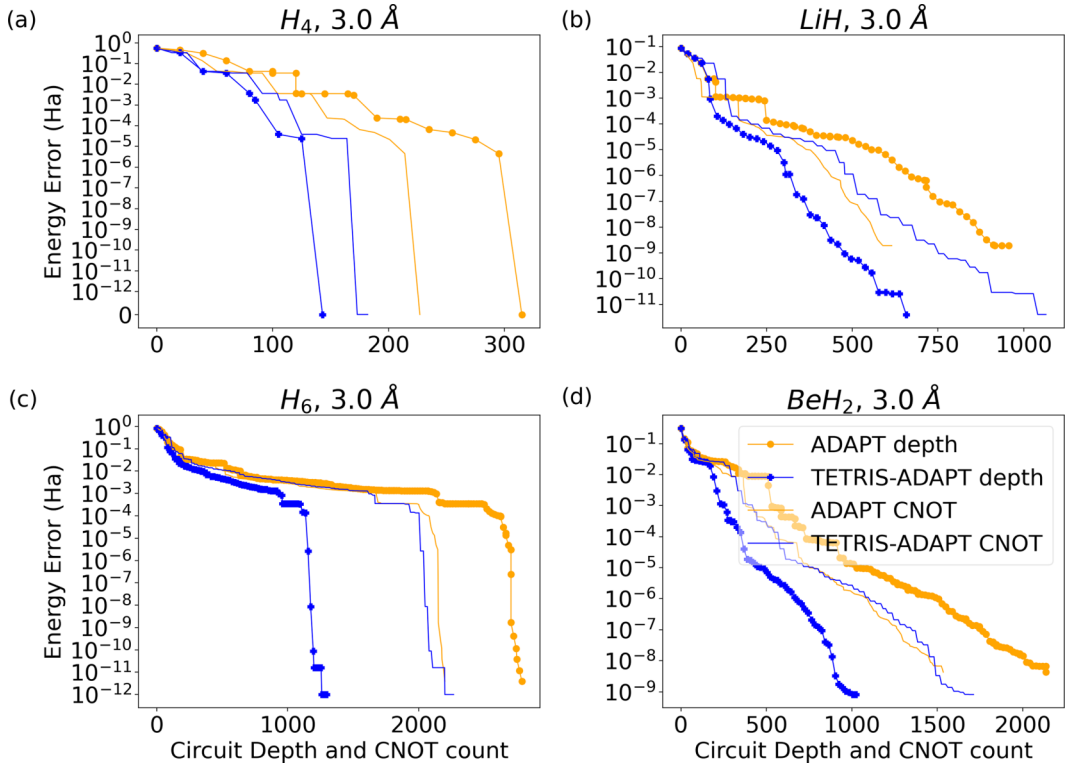

FIG. 4. Resources required by ADAPT-VQE (orange lines) and TETRIS-ADAPT-VQE (blue lines). Circuit depths and CNOT gate counts as functions of energy error for linear (a) H_4, (b) LiH, (c) H_6, and (d) BeH_2 with internuclear distances of 3.0 Å and the QE pool.

TABLE I. Summary of circuit depth and CNOT count results for the four molecules and the two pools we study. The ADAPT-VQE to TETRIS-ADAPT-VQE average (over different bond lengths) circuit depth and CNOT count ratios calculated at ADAPT-VQE convergence are given in the fourth and fifth columns. Geometries for which either version of the algorithm failed to reach the ground state were excluded.[a] The number of qubits required to simulate each molecule is given in column 2. It is evident that the circuit depth ratio grows with the system size and is consistently higher for the qubit pool. We note that although TETRIS-ADAPT-VQE requires a higher number of CNOT gates in the case of LiH, it also tends to converge to a lower energy error and state infidelity as seen in Figs. 4(b) and 5(a).

Molecule	Qubits	Pool	Depth ratio	CNOT ratio
H_4	8	qubit	1.64	0.99
H_4	8	QE	1.58	1.00
LiH	12	qubit	2.08	0.90
LiH	12	QE	1.76	0.80
H_6	12	qubit	2.32	1.02
H_6	12	QE	2.25	1.02
BeH_2	14	qubit	2.73	1.11
BeH_2	14	QE	2.56	1.09

[a]For the qubit pool, these are as follows: H_4 at 3.0–5.0 Å and H_6 at 4.0–5.0 Å. For the QE pool, this is H_6 at 5.0 Å.

layers. In fact, we show below that TETRIS-ADAPT-VQE achieves circuit depth reductions with numbers of variational parameters and CNOT gates similar to those required by the original ADAPT-VQE.

In this section, we compare ADAPT-VQE and TETRIS-ADAPT-VQE in terms of CNOT counts and circuit depths. Figure 4 shows the results for linear H_4, LiH, H_6, and BeH_2 at an internuclear distance of 3.0 Å, using the QE pool. CNOT counts and circuit depths are obtained from circuits constructed as described in Sec. II E using built-in QISKIT functions, and QISKIT optimization level 1 (back-to-back gate cancellation). We note that for a fair comparison, when transpiling ADAPT-VQE circuits, unitaries added in consecutive layers but acting on disparate sets of qubits are performed concurrently.

It is evident that TETRIS-ADAPT-VQE leads to *Ansätze* prepared by much shallower circuits, while requiring roughly the same number of CNOT gates as ADAPT-VQE. Furthermore, the larger the system size, the greater the advantage of TETRIS-ADAPT-VQE as shown in Table I, because more operators fit in the same ADAPT layer, in general.

Beyond ground-state energy estimation, ADAPT-VQE can be used for ground-state preparation. The faster energy descent of TETRIS-ADAPT does not come at the expense of *Ansatz* quality, as the two versions of the algorithm yield equally faithful *Ansätze*. Figure 5 shows that the TETRIS version of the algorithm approximates the true ground state significantly faster than the original implementation.

B. Measurement overhead reduction

In the NISQ era, the bottleneck for VQEs is the circuit depth and number of gates required for state preparation, rather than the number of state preparations and measurements. Although ADAPT-VQE yields more compact and accurate *Ansätze* compared with static ones such as the Trotterized UCCSD (tUCCSD), it does so at the additional cost of repeatedly measuring the pool operator gradient expectation values. The qubit and QE pools consist of operators acting on up to four distinct orbitals; therefore they contain $O(N^4)$ operators each, where N is the number of orbitals (or qubits). Since the molecular Hamiltonian, Eq. (1), contains up to N^4 terms, $O(N^8)$ observable expectation values are required to measure the operator gradients for the entire pool, as can be seen from Eq. (6) [25]. This step can be fully parallelized, assuming the availability of multiple quantum processors. Still, we would like to reduce the quantum resource requirements associated with the gradient measurement step, by reducing either the cost of a single iteration or the total number of iterations. One way to go about the first is to reduce the number of operators in the pool by removing redundancies and exploiting known symmetries of the system under study [16,27]. Furthermore, methods for estimating the gradients by approximately reconstructing the three-body reduced density matrix (3-RDM) from the 2-RDM [26] and by reusing informationally complete positive operator-valued measure (IC-POVM) data obtained for energy estimation [28] have been used to remove the overhead altogether, albeit by sacrificing *Ansatz* compactness to various degrees. Another recently proposed scheme that markedly reduces the runtime of a single iteration

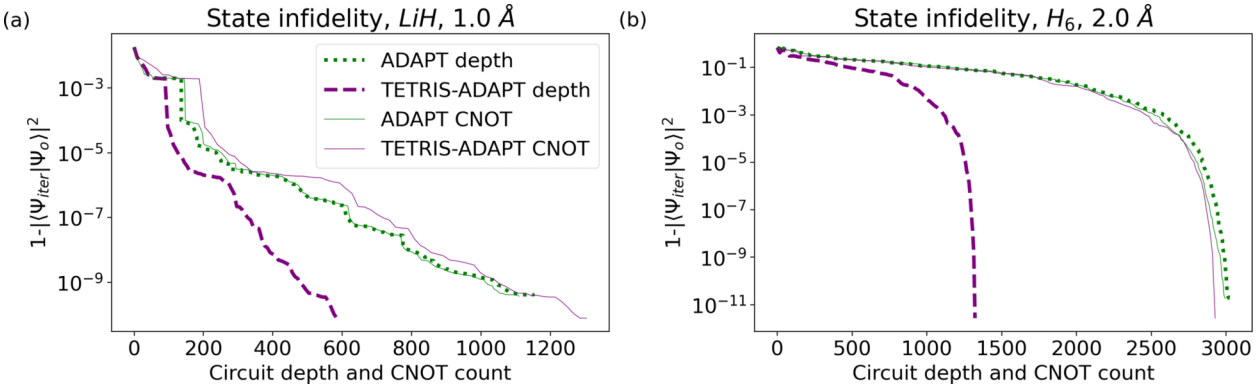

FIG. 5. Performances of ADAPT-VQE (green dotted and solid lines) and TETRIS-ADAPT-VQE (purple dashed and solid lines) in terms of state infidelity as a function of circuit depth and CNOT count for (a) LiH, with 1.0 Å bond length, and (b) linear H_6, with 2.0 Å bond length and the qubit pool.

TABLE II. Ratio of the ADAPT iterations and operator gradient measurements of ADAPT-VQE to those of TETRIS-ADAPT-VQE calculated at ADAPT-VQE convergence. Results are averaged over all molecular geometries simulated, excluding those for which either version of the algorithm failed to reach the ground state.

Molecule	Qubits	Pool	Ratio
H_4	8	qubit	2.1
H_4	8	QE	2.2
LiH	12	qubit	2.8
LiH	12	QE	2.2
H_6	12	qubit	3.1
H_6	12	QE	2.9
BeH_2	14	qubit	3.4
BeH_2	14	QE	3.3

without restricting the size of the pool or resorting to approximations is to simultaneously measure commuting observables [29]. It was shown that the $O(N^8)$ qubit and QE pool gradient observables have a straightforward partitioning into commuting sets of $O(N^3)$. That is, the entire pool gradient can be measured in $O(N^5)$ measurements, and evaluating the qubit and QE pool gradients is $O(N)$ times as expensive as a single (naïve) energy evaluation. TETRIS-ADAPT-VQE, while being compatible with all aforementioned approaches, further reduces the overall measurement cost of ADAPT-VQE by reducing the number of times the pool gradients are measured. Assuming that the *Ansatz* is dominated by double-excitation-like operators, which act on four qubits each for the qubit and QE pools in the JW mapping, TETRIS-ADAPT-VQE adds roughly $\lfloor \frac{N}{4} \rfloor$ operators per iteration, while keeping the total number of operators for a given level of accuracy roughly the same. By extension, the number of times the operator pool gradients are measured is reduced by about $\lfloor \frac{N}{4} \rfloor$, an advantage over ADAPT-VQE, linear in the system size. In Table II, we list the ADAPT iteration and measurement overhead reduction observed in the simulation of the four molecules.

C. Enhanced *Ansatz* expressivity

Simply by adding multiple operators acting on qubits different from the support of the highest-gradient operator, the TETRIS strategy lends ADAPT-VQE the ability to explore a larger part of the Hilbert space, at least at earlier stages of the algorithm. The *Ansatz* can therefore overlap with some components in the true ground state that are otherwise inaccessible to a shallow *Ansatz*.

We now use a concrete example to illustrate this. We consider the *Ansätze* grown for H_4 at 3.0 Å using the qubit pool, with the HF reference state, $|\Psi_{\text{ref}}\rangle = |11110000\rangle$, where orbitals are assigned to qubits in order of increasing energy, and even (odd) indices correspond to α (β) spin-orbitals. The optimized qubit-ADAPT *Ansätze* after one and two iterations of the algorithm are

$$\begin{aligned}|\phi_1\rangle &= e^{0.5652 i X_2 X_3 X_6 Y_7} |\Psi_{\text{ref}}\rangle \\ &= 0.8445\,|11110000\rangle - 0.5356\,|11000011\rangle\end{aligned}$$

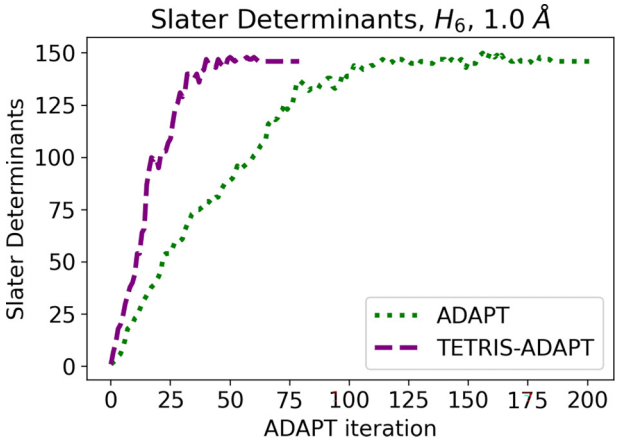

FIG. 6. Number of Slater determinants in the *Ansatz* (equivalently, computational basis states in the state of the quantum processor) with coefficients of absolute value greater than or equal to 10^{-3}, as a function of ADAPT iteration for linear H_6, bond length of 1.0 Å, and the QE pool.

and

$$\begin{aligned}|\phi_2\rangle &= e^{-1.5708 i X_0 X_3 X_5 Y_6} e^{1.5708 i X_2 X_3 X_6 Y_7} |\Psi_{\text{ref}}\rangle \\ &= |01010101\rangle\,,\end{aligned}$$

respectively. In this specific example, qubit-ADAPT arrives in the all-spin-down state, an excited state of the system, and the *Ansatz* growth stops, as the pool gradient vanishes at eigenstates of the Hamiltonian. For the same system, the corresponding TETRIS-qubit-ADAPT-VQE *Ansatz* after just one iteration is

$$\begin{aligned}|\psi_1\rangle &= e^{0.6757 i X_0 X_1 X_4 Y_5 + 0.6748 i X_2 X_3 X_6 Y_7} |\Psi_{\text{ref}}\rangle \\ &= 0.6092\,|11110000\rangle - 0.4884\,|00111100\rangle \\ &\quad - 0.4875\,|11000011\rangle + 0.3908\,|00001111\rangle\,,\end{aligned}$$

where adding two double-excitation-like qubit operators side by side generates a quadruple excitation subterm, i.e., the $|00001111\rangle$ component. The quadruply excited Slater determinant turns out to be the most dominant term beyond the reference determinant in the ground state:

$$\begin{aligned}|\Psi_0\rangle &= 0.5182\,|11110000\rangle + 0.4429\,|00001111\rangle \\ &\quad - 0.3658\,|11001100\rangle - 0.3394\,|00110011\rangle + \cdots\,,\end{aligned}$$

where $|\Psi_0\rangle$ is the ground state and the computational basis states are ordered according to the magnitude of their coefficients. In this case, the TETRIS strategy enables the *Ansätze* to include this dominant term within just the first layer, whereas the original qubit-ADAPT *Ansatz* with one layer is orthogonal to $|00001111\rangle$. TETRIS-qubit-ADAPT-VQE converges to the exact ground state within 13 iterations.

We argue that the increase in expressivity here is still *problem aware*, as the operators are added in the order of the magnitude of their gradients. This is in contrast to the problem-agnostic *Ansatz* where entangling gates are blindly added to enlarge the accessible Hilbert space. Figure 6 shows the number of Slater determinants in the *Ansatz* as a function of ADAPT layer for H_6 at 1.0 Å and both versions of the

algorithm. TETRIS-ADAPT-VQE explores the Hilbert space faster, but just the right amount, as both curves plateau at the same number. We present numerical evidence in Sec. III D to demonstrate that the TETRIS strategy does not lose the remarkable optimizability of ADAPT-VQE.

D. Optimizability

The optimizability of ADAPT-VQE and the parameter landscape it generates were recently investigated in Ref. [17] by searching for local minima via random parameter initializations and subsequent optimizations at each ADAPT iteration. By virtue of adding a large-gradient operator to the *Ansatz* at each layer, ADAPT-VQE generates a steep parameter landscape near the energy minimum found in the previous iteration. Furthermore, using the parameters found in the previous VQE step as the starting point in the new iteration ensures that ADAPT-VQE focuses on one (possibly not global) energy minimum, thus "burrowing" through the parameter landscape. Because the extra operators TETRIS-ADAPT-VQE adds to the *Ansatz* at each layer do not interfere with the largest-gradient one, they do not disrupt the steep parameter landscape. Instead, they provide additional directions in which the optimizer can march, thus increasing the chance of escaping a high-energy trap. For these reasons, we expect TETRIS-ADAPT-VQE to inherit the favorable features of the original ADAPT-VQE.

We employ the same technique to show that TETRIS-ADAPT-VQE possesses the same desirable features as ADAPT-VQE. Choosing the molecules LiH (at 3.0 Å), linear H_6 (at 1.0 Å), and linear BeH_2 (at 2.0 Å) as examples, we simulate ADAPT-VQE and TETRIS-ADAPT-VQE with the QE pool and different parameter initialization schemes. At each ADAPT iteration, 300 random parameter initializations are drawn from a uniform distribution from $-\pi$ to $+\pi$ as the initial parameters in the *Ansatz*, which lead to local energy minima found in subsequent optimization using the BFGS algorithm. These local minima are compared with the minimum energies found using two other initializations: setting the newly added parameters to zero while recycling the previously found optimal parameters, which we refer to as "warm start," and setting all the parameters in the *Ansatz* to zero, referred to as "cold start."

From Fig. 7, it is evident that as the dimension of the parameter space increases, it becomes increasingly improbable for random initializations to find the minima reached by both versions of ADAPT-VQE and warm starting, which reliably lands at minima with errors one to two orders of magnitude smaller. In Figs. 7(c) and 7(d), and especially in Figs. 7(b) and 7(f), this ceases to be true in the final iterations, where the cost function landscape becomes increasingly trap-free. This is a signature of the phenomenon of overparametrization, in which the *Ansatz* has more than a critical number of parameters such that all relevant directions in state space can be explored [30]. In this regime, the deep ADAPT burrow appears to persist as a local minimum and in some cases perform worse than random initializations as seen in Figs. 7(b) and 7(f). However, achieving overparametrization generally requires an unfavorably high number of parameters and unitary rotations (up to exponentially many in the system size, depending on the *Ansatz* [31]), and it is not a regime in which the compact ADAPT-VQE *Ansätze* are expected to find use [17,32]. We also note that although cold starting is remarkably good at finding the ADAPT-VQE minimum, it also requires a greater number of function evaluations, and its energy does not decrease monotonically.

This demonstrates the efficacy of the usual initialization (i.e., warm start) in ADAPT-VQE, regardless of whether or not the TETRIS strategy is used. With this practical initialization scheme, the classical optimization in TETRIS-ADAPT-VQE produces a result equally as good as that of ADAPT-VQE with a similar number of variational parameters, even though at a given layer the dimension of the parameter space increases by more than 1.

IV. CONCLUSIONS

Quantum simulation algorithms designed for NISQ devices should account for limited coherence times and noisy gates. In prior work, ADAPT-VQE, which iteratively grows problem-tailored *Ansätze* based on local energy gradient information, was shown to yield accurate trial wave functions requiring fewer variational parameters, prepared by shallower circuits and fewer gates compared with UCCSD [15,16]. In this paper, we introduce a pool-independent variation of the ADAPT-VQE algorithm, called TETRIS-ADAPT-VQE, in order to cut down on the resources needed to carry out a given simulation task. TETRIS-ADAPT-VQE achieves this in three ways:

(*a*) *Curtailing circuit depths.* The original ADAPT-VQE algorithm adds a single operator per iteration, and more often than not, the sets of qubits that two consecutive operators act on intersect. That is, operators added in consecutive layers cannot be implemented simultaneously at the circuit level, but instead must be applied in succession. In contrast, TETRIS-ADAPT-VQE by design adds multiple operators in the same iteration, each acting on different sets of qubits, allowing them to be implemented simultaneously in the circuit. This results in significantly shallower circuits with roughly the same number of CNOT gates and variational parameters compared with the original ADAPT-VQE.

(*b*) *Slashing the measurement overhead.* Although the operator gradient measurement step can in principle be performed in parallel, there will be a finite number of quantum processors available for a given experiment, and reducing the total number of shots is desirable. Because TETRIS-ADAPT-VQE adds multiple operators to the *Ansatz* at each iteration while keeping the total number of operators roughly the same, the number of times the pool gradient needs to be measured is only a fraction of that of the original algorithm. In the JW mapping, with the qubit and QE pools, the factor by which TETRIS-ADAPT-VQE reduces the gradient measurement overhead grows roughly linearly with the system size.

(*c*) *Exploring the Hilbert space faster.* By adding multiple operators in tandem, TETRIS-ADAPT-VQE samples the Hilbert space faster and in more directions, especially in the early iterations. Because operator addition is still gradient based and problem aware, the *Ansatz* does not leave the subspace where the solution lives.

Note added. Recently, follow-up work which further compares the differences and analyzes the properties of ADAPT, TETRIS-ADAPT, and related schemes appeared in a preprint [33]. The authors confirm our conclusions regarding circuit

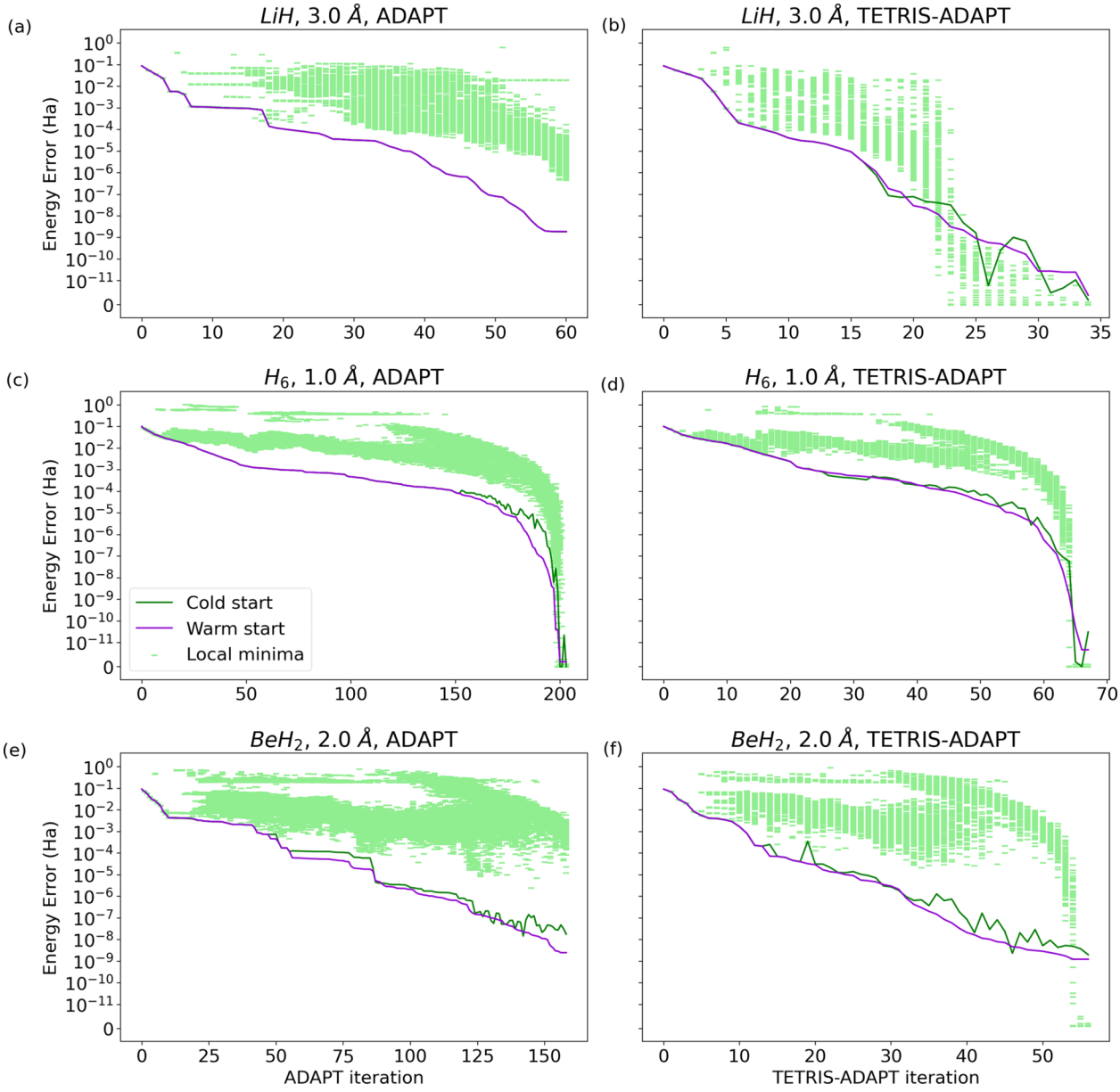

FIG. 7. Convergence curves for (a) and (b) LiH at 3.0 Å, (c) and (d) H_6 at 1.0 Å, and (e) and (f) BeH_2 at 2.0 Å, for ADAPT-VQE [(a), (c), and (e)] and TETRIS-ADAPT-VQE [(b), (d), and (f)] using the QE pool. Energy minima found by random parameter initialization and subsequent optimization at each ADAPT iteration are shown in light green. The warm-start initialization in which parameters are initialized in their optimal values from the previous ADAPT iteration is shown (purple line). The all-zero parameter initialization (cold start) is also shown (solid green line).

depth reduction without an increase in the number of variational parameters, as well as an $O(N)$ decrease in the number of operator gradient evaluations. Furthermore, they find that TETRIS-ADAPT-VQE is less susceptible than ADAPT-VQE to amplitude-damping and dephasing errors which affect idling and actively manipulated qubits alike (related to circuit depth), though not necessarily advantageous when it comes to depolarizing errors (related to the number of gates).

ACKNOWLEDGMENTS

This research was supported by the U.S. Department of Energy (Award No. DE-SC0019199). S.E.E. also acknowledges support by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Co-design Center for Quantum Advantage (C2QA) under Contract No. DE-SC0012704.

[1] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996).

[2] A. Y. Kitaev, Quantum measurements and the Abelian stabilizer problem, arXiv:quant-ph/9511026.

[3] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Quantum algorithms revisited, Proc. R. Soc. London A **454**, 339 (1998).
[4] 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).
[5] M. Ganzhorn, D.J. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, I. Tavernelli, and S. Filipp, Gate-efficient simulation of molecular eigenstates on a quantum computer, Phys. Rev. Appl. **11**, 044092 (2019).
[6] B. T. Gard, L. 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 Quantum Inf. **6**, 10 (2020).
[7] E. Fontana, M. Cerezo, A. Arrasmith, I. Rungger, and P. J. Coles, Non-trivial symmetries in quantum landscapes and their resilience to quantum noise, Quantum **6**, 804 (2022).
[8] S. E. Smart and D. A. Mazziotti, Quantum solver of contracted eigenvalue equations for scalable molecular simulations on quantum computing devices, Phys. Rev. Lett. **126**, 070504 (2021).
[9] W. Mizukami, K. Mitarai, Y. O. Nakagawa, T. Yamamoto, T. Yan, and Y. Y. Ohnishi, Orbital optimized unitary coupled cluster theory for quantum computer, Phys. Rev. Res. **2**, 033421 (2020).
[10] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, J. Chem. Theory Comput. **15**, 311 (2019).
[11] 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).
[12] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).
[13] H. R. Grimsley, D. Claudino, S. E. Economou, E. Barnes, and N. J. Mayhall, Is the Trotterized UCCSD ansatz chemically well-defined?, J. Chem. Theory Comput. **16**, 1 (2020).
[14] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, Quantum **3**, 214 (2019).
[15] 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).
[16] 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).
[17] H. R. Grimsley, G. S. Barron, E. Barnes, S. E. Economou, and N. J. Mayhall, Adaptive, problem-tailored variational quantum eigensolver mitigates rough parameter landscapes and barren plateaus, npj Quantum Inf. **9**, 19 (2023).
[18] 40 years of quantum computing, Nat. Rev. Phys. **4**, 1 (2022).
[19] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, Quantum computers, Nature (London) **464**, 45 (2010).
[20] A. Anand, P. Schleich, S. Alperin-Lea, P. W. K. Jensen, S. Sim, M. Dí az-Tinoco, J. S. Kottmann, M. Degroote, A. F. Izmaylov, and A. Aspuru-Guzik, A quantum computing view on unitary coupled cluster theory, Chem. Soc. Rev. **51**, 1659 (2022).
[21] Y. S. Yordanov, V. Armaos, C. H. W. Barnes, and D. R. M. Arvidsson-Shukur, Qubit-excitation-based adaptive variational quantum eigensolver, Commun. Phys. **4**, 228 (2021).
[22] Y. S. Yordanov, D. R. M. Arvidsson-Shukur, and C. H. W. Barnes, Efficient quantum circuits for quantum computational chemistry, Phys. Rev. A **102**, 062612 (2020).
[23] Qiskit Contributors, Qiskit: An open-source framework for quantum computing, 2023, 10.5281/zenodo.2573505.
[24] 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 Ccomputing in Python, Nat. Methods **17**, 261 (2020); **17**, 352 (2020).
[25] However, as was shown in Ref. [26], for fermionic pools, knowledge of the three-body reduced density matrix (3-RDM) which contains only $O(N^6)$ elements is sufficient for evaluating the pool gradients.
[26] J. Liu, Z. Li, and J. Yang, An efficient adaptive variational quantum solver of the Schrödinger equation based on reduced density matrices, J. Chem. Phys. **154**, 244112 (2021).
[27] V. O. Shkolnikov, N. J. Mayhall, S. E. Economou, and E. Barnes, Avoiding symmetry roadblocks and minimizing the measurement overhead of adaptive variational quantum eigensolvers, Quantum **7**, 1040 (2023).
[28] A. Nykänen, M. A. C. Rossi, E.-M. Borrelli, S. Maniscalco, and G. García-Pérez, Mitigating the measurement overhead of ADAPT-VQE with optimised informationally complete generalised measurements, arXiv:2212.09719.
[29] P. G. Anastasiou, N. J. Mayhall, E. Barnes, and S. E. Economou, How to really measure operator gradients in ADAPT-VQE, arXiv:2306.03227.
[30] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, Theory of overparametrization in quantum neural networks, Nat. Comput. Sci. **3**, 542 (2023).
[31] J. Kim, J. Kim, and D. Rosa, Universal effectiveness of high-depth circuits in variational eigenproblems, Phys. Rev. Res. **3**, 023203 (2021).
[32] It was shown in Ref. [30] that the critical number of parameters is upper bounded by the dimension of the dynamical Lie algebra (DLA) obtained by repeatedly taking nested commutators of the *Ansatz* generators. We observe that the two versions of the algorithm require similar numbers of variational parameters to reach the same energy accuracy, and for a given number of operators added to the *Ansatz* the dimension of the DLA for TETRIS-ADAPT-VQE is expected to be smaller since operators added in the same layer necessarily commute. Thus it is reasonable to expect TETRIS-ADAPT-VQE to reach overparametrization faster, as suggested by comparing Figs. 7(a) and 7(e) with Figs. 7(b) and 7(f).
[33] C. K. Long, K. Dalton, C. H. W. Barnes, D. R. M. Arvidsson-Shukur, and N. Mertig, Layering and subpool exploration for adaptive variational quantum eigensolvers: Reducing circuit depth, runtime, and susceptibility to noise, arXiv:2308.11708.