ADAPT-VQE: An exact variational algorithm for fermionic simulations on a quantum computer

Harper R. Grimsley, Sophia E. Economou, Edwin Barnes, and Nicholas J. Mayhall

Department of Chemistry, Virginia Tech, Blacksburg, VA 24061, USA
Department of Physics, Virginia Tech, Blacksburg, VA 24061, USA

Quantum simulation of chemical systems is one of the most promising near-term applications of quantum computers. The variational quantum eigensolver, a leading algorithm for molecular simulations on quantum hardware, has a serious limitation in that it typically relies on a pre-selected wavefunction ansatz that results in approximate wavefunctions and energies. Here we present an arbitrarily accurate variational algorithm termed ADAPT-VQE. Instead of fixing an ansatz upfront, this algorithm grows it systematically one operator at a time in a way dictated by the system being simulated. This generates an ansatz with a small number of parameters, leading to shallow-depth circuits. We present numerical simulations, including for a prototypical strongly correlated molecule, which show that our algorithm performs much better than a unitary coupled cluster approach, in terms of both circuit depth and chemical accuracy. Our results highlight the potential of our adaptive algorithm for exact simulations with present-day and near-term quantum hardware.

I. INTRODUCTION

Anticipation that a useful quantum computer will be realized in the near future has motivated intense research into developing quantum algorithms which can potentially make progress on classically intractable computational problems. While many research areas expect to see transformative change with the development of such quantum devices, computational chemistry is poised to be among the first domains to significantly benefit from such new technologies. Due to the exponential growth in the size of the Hilbert space with increasing orbitals, a quantum computer with tens of qubits could potentially surpass classical algorithms [1]. Achieving such a capability depends not only on the quality of the qubits, but also critically on the efficiency of the algorithms.

The phase estimation algorithm (PEA) [2] was the first algorithm proposed for simulating electronic structure problems on a quantum computer [3-8]. PEA provides a path for obtaining the exact ground state electronic energy for a molecule by evolving in time a quantum state with significant overlap with the ground state using the molecular Hamiltonian of interest. Due to the very long circuit depths and complex quantum gates required by PEA, the coherence times needed to simulate interesting electronic states would exceed the coherence times available on any existing or near-term quantum device. Improvements to PEA still require significant resources and experimental demonstrations to date only involve a few qubits [4,5].

In order to reduce the significant hardware demands required by PEA and exploit the capabilities of noisy intermediate-scale quantum (NISQ) devices [7], the variational quantum eigensolver (VQE) algorithm was proposed and demonstrated using photonic qubits by Peruzzo et al. [8]. This was followed by several theoretical studies on VQE [9-15] and demonstrations on other hardware such as superconducting qubits [5,14,16] and trapped ions [17,18].

VQE is a hybrid quantum-classical algorithm, because the computational work is shared between classical and quantum hardware. VQE starts with an assumption about the form of the target wavefunction. Based on this form, an ansatz with several tunable parameters is constructed, and a quantum circuit capable of producing this ansatz is designed. The ansatz parameters are variationally adjusted until they minimize the expectation value of the molecular Hamiltonian. Classical hardware is used to precompute all the Hamiltonian terms and to update the parameters during the circuit optimization. The quantum hardware is only used to prepare a state (defined by its current set of ansatz parameter values) and to perform measurements of the various interaction terms in the molecular Hamiltonian, \( \hat{H} = \sum g_{i} \hat{\alpha}_{i} \). Because the individual operator terms, \( \hat{\alpha}_{i} \), generally do not commute, the state preparation has to be repeated multiple times, until all the individual operators have been measured enough times to get sufficient statistics on their mean value. Details on all these steps can be found in Ref. [19].

Compared to PEA, VQE is much more suitable for NISQ devices, trading in the long circuit depths for shorter state preparation circuits, at the expense of a much higher number of measurements. Although VQE has been demonstrated to be more efficient and error-tolerant [4,10,14], this comes with the compromise that the ansatz generally only allows one to obtain approximations to the ground state. Because the choice of ansatz determines the variational flexibility of the trial state, the quality of a VQE simulation is only as good as the ansatz.

Several approaches have been explored with the goal of creating a compact ansatz which provides high accu-
racy with few parameters and shallow circuits. The first ansatz explored \cite{19} was based on the unitary variant of coupled cluster theory truncated at single and double excitations (UCCSD), inspired by early efforts in computational chemistry to improve coupled cluster theory \cite{19,22}. In UCCSD, trial states are generated by applying to a reference state a unitary operator in the form of an exponential of a sum of single and double fermion operators with their coefficients taken as free parameters. More recent proposals based on UCCSD include the unitary Bogoliubov coupled cluster theory which takes a generalized Hartree-Fock (HF) state as the reference \cite{23} and the $k$-UpCCGSD approach of Lee et al. \cite{15} which uses $k$ products of unitary paired generalized double excitations, along with the full set of generalized single excitations. The $k$-UpCCGSD approach builds on early work by Nakatsuji \cite{24,27} and Nooijen \cite{28} studying the use of generalized excitation terms in classical quantum chemistry algorithms, but prunes the expansive operator list by restricting the two-particle terms to only paired interactions. Ryabinkin et al. \cite{29} recently proposed a coupled cluster-like ansatz which is constructed directly in the qubit representation with the goal of achieving shallower circuits. While not directly a variation of the UCCSD ansatz itself, Ref. \cite{14} developed an approach (termed the quantum subspace expansion) to extract not just the expectation value of $H$, but all the matrix elements $\langle I | H | J \rangle$ in a small subspace consisting of single excitations from the trial state. This Hamiltonian matrix is then diagonalized on a classical computer, which reduces the impact of decoherence and gives access to excited states.

Despite these considerable improvements to the UCCSD ansatz for VQE, this remains an approximate approach that works best for systems that are not strongly correlated. However, strongly correlated systems are the hardest to simulate classically, and this is precisely the motivation for performing simulations using quantum computers. While an exact VQE simulation could in principle be performed by adding higher rank excitations to the ansatz, this would be prohibitively expensive for both the classical subroutines and NISQ devices. To overcome these challenges, we need to avoid imposing an ad hoc ansatz and instead allow the system to determine its own compact, quasi-optimal ansatz.

In this paper, we achieve this by introducing a simple algorithm termed Adaptive Derivative-Assembled Pseudo-Trotter ansatz Variational Quantum Eigensolver (ADAPT-VQE). ADAPT-VQE determines a quasi-optimal ansatz with the minimal number of operators for a desired level of accuracy. The key idea is to systematically grow the ansatz by adding fermionic operators one-at-a-time, such that the maximal amount of correlation energy is recovered at each step. This results in a wavefunction ansatz that is discovered by the algorithm, and which cannot be predicted a priori from a traditional excitation-based scheme like UCCSD. While intuitive, this approach can also be derived more rigorously as a particular approach to Full Configuration Interaction (FCI) VQE. We demonstrate the power of ADAPT-VQE through numerical simulations of three molecules of increasing complexity: LiH, BeH$_2$, and H$_6$. In each case, we find vastly improved performance compared to UCCSD, both in terms of the number of operators needed to form the trial states and in terms of chemical accuracy. Therefore, we believe that ADAPT-VQE is an ideal hybrid algorithm for NISQ devices.

The paper is organized as follows. In Sec. \ref{sec:background} we review the standard UCCSD ansatz and its implementation on quantum processors. In Sec. \ref{sec:background} we describe ADAPT-VQE in detail and show how it can be derived from a FCI ansatz. In Sec. \ref{sec:results} we apply ADAPT-VQE to three molecules and compare its performance to that of UCCSD. In Sec. \ref{sec:results} we discuss our results and future research directions it enables. We conclude in Sec. \ref{sec:conclusion}.

\section{Background}

In order to define the approach, several definitions and notations need to be established. First, molecular orbital indices $i$ and $j$ denote occupied orbitals, $a$ and $b$ denote virtual orbitals, and $p$, $q$, $r$, and $s$ denote arbitrary molecular orbitals. In coupled cluster theory, in particular CCSD, an expansion based on the HF state $|\psi_{\text{HF}}\rangle$ is created by using an exponential ansatz involving single and double excitation operators:

$$
|\psi_{\text{CCSD}}\rangle = e^{\hat{T}_1 + \hat{T}_2} |\psi_{\text{HF}}\rangle,
$$

where the excitation operators are defined as:

$$
\hat{T}_1 = \sum_{i\alpha} \hat{t}_{i\alpha} = \sum_{i\alpha} v_{i\alpha} \hat{a}_\alpha^{\dagger} \hat{a}_i,
$$

$$
\hat{T}_2 = \sum_{i<j,a<b} \hat{t}_{ij}^{ab} = \sum_{i<j,a<b} t_{ij}^{ab} \hat{a}_a^{\dagger} \hat{a}_b \hat{a}_j \hat{a}_i.
$$

For closed shell molecules near equilibrium, CCSD provides a robust ansatz for molecular simulations. However, for strongly correlated molecules CCSD fails to describe the necessary correlation, and variational methods such as MRCI \cite{30,32} and DMRG \cite{33,35} become necessary. Early efforts to combine size extensivity and variationality were pioneered by Bartlett, Kutzelnigg, and coworkers \cite{10,21}. In this context, a unitary variant of coupled cluster theory (UCCSD) was defined by replacing the excitation operators with a sum of excitation and de-excitation operators:

$$
\hat{t}_{ij}^{ab} \rightarrow \hat{t}_{ij}^{ab} - \hat{t}_{ij}^{ij} = \hat{z}_{ij}^{ab}.
$$

Because UCCSD is based on a unitary operator, the adjoint is the inverse, and the expectation value of the UCCSD wavefunction can be expanded using the Baker-Campbell-Hausdorff (BCH) formula to obtain a normalized Hamiltonian expectation value (Rayleigh quotient)
for variational optimization. Unfortunately, the BCH expansion does not truncate at finite order, making UCCSD computationally intractable on classical hardware. However, the unitary nature of UCCSD is actually a benefit for quantum algorithms as it corresponds to a coherent time evolution, and this was the original motivation for using UCCSD in VQE \cite{CC}.  

In addition to a unitary form, CCSD can also be generalized by including excitation operators which immediately annihilate the HF state. These would include excitations from occupied to occupied, virtual to virtual, etc. Generalized excitations or interactions of this form have been considered previously, and have been used in the context of VQE recently by Lee and coworkers \cite{Lee}. In this case the cluster operators are further generalized to remove the HF-based subspace restriction: \(\hat{\tau}_{pq} \rightarrow \hat{\tau}_{pq}^{rs}\), where \(p, q, r, \) and \(s\) refer to any arbitrary orbital.

Although UCCSD is perhaps a natural ansatz for VQE, it cannot be implemented directly as written or as explored previously in the quantum chemistry context. Because the gate model of quantum computation is realistically bound to using gates acting on only a few qubits at a time, the UCCSD operator must be broken up into a time-ordered sequence of few (one or two) particle operators. This is achieved by using a Trotter expansion of a matrix exponential \cite{trotter}

\[
e^{A+B} = \lim_{n \to \infty} \left(e^{(A/n)}e^{(B/n)}\right)^n.
\]

While the use of a truncated Trotter expansion represents an approximation to the underlying UCCSD ansatz, recent work has clearly shown that this does not strongly affect the results because the variational flexibility is sufficient to absorb this error \cite{CC}, and that even a single Trotter number \((n = 1)\) is sufficient to reproduce the results of UCCSD. As a result, a unitary, generalized, Trotterized ansatz becomes:

\[
|\psi^{\text{UCC}}\rangle = \prod_{s \in \{pq\}} e^{\ell_s} \prod_{d \in \{pqrs\}} e^{\ell_d} |\psi^{\text{HF}}\rangle,
\]

where notation is introduced such that the generalized singles index, \(s\), runs over all unique pairs of \(p, q\) and the doubles index, \(d\), over unique combinations of \(p, q, r, s\).

### III. ADAPT-VQE

The above discussion described the Trotter expansion as an approximation to UCCSGD. However, as recognized previously \cite{CC, Bravyi}, if the parameters are optimized after the Trotterization, this is not so much an approximation to UCC as it is a wholly unique ansatz. In fact, the exact FCI solution could be obtained by simply going to an \(n\)th order Trotterized form of UCCSD and allowing the different parameter replicas to vary independently. This is due to the fact that \(n\)-body interactions can be described as products of one- and two-body interactions.

The exact (FCI) quantum state can thus be represented as an arbitrarily long product of one- and two-body operators,

\[
|\psi^{\text{FCI}}\rangle = \prod_{k} \prod_{pq} e^{\tau_{pq}^{\ell_{pq}}} \prod_{pqrs} e^{\tau_{pqrs}^{\ell_{pqrs}}} |\psi^{\text{HF}}\rangle,
\]

where \(\tau_{pq}^{\ell_{pq}}(k)\) is the \(k\)th instance, or “replica”, of the operator in \(\tau_{pq}^{\ell_{pq}} - \tau_{pq}^{\ell_{pq}}\). It is important to note that this is not a Trotter approximation to any simple two-body ansatz, as each replica can assume different parameter values, e.g., \(\tau_{pq}^{\ell_{pq}}(k) \neq \tau_{pq}^{\ell_{pq}}(j)\).

The main goal in this paper is to approximate FCI with arbitrary accuracy using a maximally compact sequence of unitary operators. The basic outline of the algorithm is drawn schematically in Fig. 1 and is as follows:

1. On classical hardware, compute one- and two-electron integrals, and transform the fermionic Hamiltonian into a qubit representation using an appropriate transformation: Jordan-Wigner, Bravyi-Kitaev, etc. This is a standard step in regular VQE.

2. Define an “Operator Pool”. This is simply a collection of operator definitions which will be used to construct the ansatz. For the examples presented in the next section, we consider the set of all unique spin-adapted one- and two-body operators, but one might imagine adding a few three-body or four-body terms as well.

3. Initialize qubits to an appropriate reference state, ideally one with the correct number of electrons. The HF state would be a sensible choice here. Initialize the ansatz to the identity operator.

4. On a quantum computer, prepare a trial state with the current ansatz and measure the commutator of the Hamiltonian with each operator in the pool. Repeating this multiple times and averaging gives the gradient of the expectation value of the Hamiltonian with respect to the coefficient of each operator.

5. If the norm of the gradient vector is smaller than some threshold, \(\epsilon\), exit.

6. Identify the operator with the largest gradient and add this single operator to the left end of the ansatz, with a new variational parameter. Perform a VQE experiment to re-optimize all parameters in the ansatz.

7. Go to step 4.

As described above and illustrated in Fig. 1 each iteration starts as a series of uncoupled experiments to obtain the parameter gradients via measurements of operator commutators. The purpose of these gradient measurements is to determine the best operator with which to
grow the ansatz, as the operator with the largest gradient is likely to recover the most correlation energy in the subsequent VQE minimization. This process is continued iteratively, until a convergence threshold is met. In the classical numerical examples presented below, we chose to consider the $L^2$ norm of the gradient vector to determine convergence. This is just one possibility, and alternative convergence indicators could be used instead in step 5. At convergence, the ADAPT-VQE algorithm obtains the following ansatz:

$$\psi^{\text{ADAPT}}(e) = \langle e^{\hat{p}_1} \cdots e^{\hat{p}_{N-1}} e^{\hat{q}_N} | \psi_{\text{HF}} \rangle$$  \hspace{1cm} (8)$$

where the identity of each $\hat{q}_n$ is determined by the algorithm. Here, we note that some of the steps in ADAPT-VQE can be implemented in more than one way. For instance, the computation of the gradient in step 4 could be done by measuring the Hamiltonian itself instead of the commutator of the Hamiltonian with operators from the pool. In this approach, the gradient associated with a given operator can be obtained by measuring the expectation value of the Hamiltonian with and without the additional parameter included in the state preparation circuit. This comes at the cost of increasing the circuit depth slightly, requiring many more sampling points to obtain better statistics for calculating finite derivatives, and this may or may not be preferable depending on the complexity of the Hamiltonian-pool operator commutators. The re-optimization subroutine in step 6 can be implemented on either a classical or quantum processor using any of the gradient- or non-gradient-based optimization routines that have been proposed or demonstrated for VQE \cite{10, 13, 14, 16}. Note that this subroutine is distinct from the gradient computed in step 4 of the algorithm. Additional possible modifications to the algorithm are discussed in Sec. V.

### A. Derivation as a sparse-update FCI ansatz

In addition to the intuitive definition of the algorithm presented above, one can also view it as a version of FCI VQE (using the ansatz definition in Eq. 7), but with a sparsity constraint on the gradient-based update step, such that only single parameter updates are allowed per iteration. We will adopt a slightly more generic notation at this point to simplify the discussion:

$$\psi^{\text{FCI}} = \prod_k \prod_n e^{\theta_k^{\text{FCI}} \hat{A}_n} | \psi_{\text{HF}} \rangle, \hspace{1cm} (9)$$

where $\hat{A}_n \in \{\hat{a}_p^{\dagger} \hat{a}_q - \hat{a}_q^{\dagger} \hat{a}_p, \hat{a}_p^{\dagger} \hat{a}_q^{\dagger} \hat{a}_s - \hat{a}_s^{\dagger} \hat{a}_q^{\dagger} \hat{a}_p \}$. The goal is to find parameters $\theta_n$ such that the energy functional is minimized to a value that is as close as possible to the FCI energy:

$$\frac{\partial}{\partial \theta_k^{\text{FCI}}} \langle \psi_{\text{HF}} \rangle | e^{-\theta_1^{\text{FCI}} \hat{A}_1} e^{-\theta_2^{\text{FCI}} \hat{A}_2} \cdots \hat{H} \cdots e^{\theta_2^{\text{FCI}} \hat{A}_2} e^{\theta_1^{\text{FCI}} \hat{A}_1} | \psi_{\text{HF}} \rangle = 0. \hspace{1cm} (10)$$
1. First iteration

If we initialize all of the parameters, $\theta_m^k$, to zero, then the expectation value of the 0th iteration state, $|\psi^{(0)}\rangle$, would simply return the uncorrelated HF energy:

$$|\psi^{(0)}\rangle = |\psi^{HF}\rangle. \quad (11)$$

Using the gradient to update all the parameters in the wavefunction would immediately lead to an exponentially complicated wavefunction which could no longer be simulated either classically or quantum mechanically. The reason is that all $k$ replicas have identical gradients:

$$\frac{\partial E^{(0)}}{\partial \theta_m^k} = \frac{\partial E^{(0)}}{\partial \theta_m^j} \quad \text{for all } k, j. \quad (13)$$

This would mean that the ansatz would have an exponentially large number of terms after the first iteration, and also the circuit depth would be exponentially large. Alternatively, we choose to impose a sparsity constraint on the parameter update, such that only a single parameter may be updated at a time. This has the consequence that after updating the single chosen parameter, the ansatz has only increased in complexity by a single non-zero parameter, and the resulting circuit has only grown in depth to accommodate the single extra one- or two-body operator:

$$|\psi^{(1)}\rangle = e^{\theta_p^1 \hat{A}_p} |\psi^{HF}\rangle. \quad (14)$$

To recover the greatest amount of electron correlation with the maximally compact wavefunction, we choose to update only the operator with the largest gradient. As a result of this update, the ansatz has grown from HF to a simple UCC-type ansatz with only a single excitation operator.

To determine the stepsize for this iteration (essentially a line-search), we perform a variational minimization via VQE:

$$E^{(1)} = \min_{\theta_p^1 \theta_q^1} \langle \psi^{HF}| e^{-\theta_p^1 \hat{A}_p} \hat{H} e^{\theta_p^1 \hat{A}_p} |\psi^{HF}\rangle. \quad (15)$$

The result of this relatively simple VQE experiment then provides the first iteration energy and wavefunction.

2. Second iteration

In the second iteration, the computation of the gradient of all the one- and two-body operators is required once again, and can be performed in a NISQ-friendly parallel execution on uncoupled quantum computers:

$$\frac{\partial E^{(1)}}{\partial \theta_m^k} = \langle \psi^{(1)}|[H, \hat{A}_m]|\psi^{(1)}\rangle. \quad (16)$$

However, because $\theta_q^1$ has already converged in the previous VQE step, its gradient is zero. Just as in the previous iteration, a sparse update is performed to identify and insert only the operator with the largest magnitude gradient. This grows the ansatz by one parameter, $\theta_q^1$. The second iteration energy is then obtained by a new VQE simulation for a UCC ansatz containing two excitation operators:

$$E^{(2)} = \min_{\theta_p^1 \theta_q^1} \langle \psi^{HF}| e^{-\theta_p^1 \hat{A}_p} e^{-\theta_q^1 \hat{A}_q} \hat{H} e^{\theta_p^1 \hat{A}_p} e^{\theta_q^1 \hat{A}_q} |\psi^{HF}\rangle. \quad (17)$$

We emphasize that it is not only the new parameter which is optimized, but all previously added parameters are updated as well. Future iterations follow similarly.

B. Quantum and Classical Hardware requirements

The evaluation of all the gradient terms could in principle be achieved in a NISQ-friendly, highly parallel manner with a large number of uncoupled quantum computers all tasked with preparing the same state and measuring a different operator. This is the same potential for parallelization that the underlying VQE subroutine has. Just as with the original motivation for VQE, ADAPT-VQE decreases the circuit depth at the expense of a larger number of measurements. In our case a sequence of VQE experiments is performed, with the most resource-demanding experimental steps happening at the end. This constitutes a rather large prefactor which would scale with the size of the system, but the crucial advantage is controllability over the ansatz accuracy (in principle approaching FCI). Because the number of non-zero parameters equals the number of iterations, in order...
to discover an ansatz for a large system, an equal number of VQE re-optimizations will need to be performed. One strategy to minimize this prefactor could simply be to add a few operators at a time.

Determining resource requirements for adaptive procedures is rather difficult. The classical resources are not expected to be significant in the foreseeable future. However, as quantum technology progresses toward deeper circuits, the parameter manipulation and updating on a classical computer could become costly. However, we expect the dependence between parameters at the beginning and end of the ADAPT-VQE circuit to decay with circuit depth, such that one could imagine freezing the early parameters after a certain number of iterations. This would possibly establish an approach for FCI with only a polynomial number of variables, completely avoiding any exponential cost for the classical hardware.

IV. NUMERICAL EXAMPLES

In this section, we explore the convergence properties of the ADAPT-VQE algorithm with a few small molecular systems, LiH, BeH₂, and linear H₆. The former two molecules have been simulated using quantum hardware [10,18]. H₆ is included as a prototypical strongly correlated molecule, which allows us to test the ADAPT-VQE approach for systems which are not well described with unitary coupled cluster.

In order to perform the simulations, an in-house code was written, using Psi4 [38,39] for the integral calculation (via the OpenFermion [40] interface) and OpenFermion was used for the Jordan-Wigner operator transformation. All calculations used the Broyden-Fletcher-Goldfarb-Shannon (BFGS) minimization implemented within Scipy [41]. To classically simulate the re-optimization subroutine in step 6 of ADAPT-VQE, we could use a standard numerical gradient method. However, in order to improve the efficiency and allow accurate gradients for tight convergence, we derived and implemented an efficient analytic gradient function, which is detailed in Appendix A. By reusing intermediates between individual parameter gradients, this algorithm retains the full gradient vector for all parameters at a cost which is only roughly 2x that of the base energy evaluation. For a comparison, Table I lists the times required for a few VQE iterations of H₂O with the STO-3G basis set using both numerical and analytical gradients.

As discussed in the previous section, the ADAPT ansatz uses a convergence threshold to determine when the calculation should terminate (step 5). Here we use the norm of the gradient vector and compare it to threshold $\epsilon_m$, which we define as

$$\epsilon_m = 10^{-m}. \tag{18}$$

For example, an ADAPT-VQE calculation where the norm of the operator pool gradient is converged to less than 0.001 would be denoted as ADAPT($\epsilon_3$). In what follows we present numerical results for bond-dissociation curves for LiH, BeH₂ and H₆ for three different choices of the threshold ($m = 1, 2, 3$). We also investigate alternate protocols for the ansatz growth and demonstrate the superiority of the ADAPT ansatz.

A. LiH Bond Dissociation

Here, we study the LiH bond dissociation computed using several methods, including FCI, UCCSD (un-Trotterized), HF, ADAPT($\epsilon_1$), ADAPT($\epsilon_2$), and ADAPT($\epsilon_3$), all with the STO-3G basis set. In this basis set, LiH has 6 spatial orbitals and a Hilbert space of dimension 4096. By starting with the HF state with two $\alpha$ (spin-up) and two $\beta$ (spin-down) electrons and using only number conserving operators, the relevant subspace to explore has a dimension of $\binom{6}{3} \cdot \binom{6}{2} = 225$. In this basis, the occupied orbitals are $\{1,2\}$, and the virtual orbitals are $\{3,4,5,6\}$.

The bond dissociation curves are shown in Fig. 2(a:ii), where all the curves, with the exception of HF, cannot really be distinguished on this scale. However, as shown in Fig. 2(a:ii), when the FCI energy is subtracted and the scale is adjusted, significant differences become evident. Shading is used to indicate chemical accuracy, which is achieved in all cases other than HF. LiH has only a single pair of electrons (a $\sigma$ bond) breaking along the dissociation coordinate, and UCCSD exhibits chemical accuracy throughout the curve. While ADAPT($\epsilon_1$) is not as accurate as UCCSD, ADAPT($\epsilon_2$) is comparable to UCCSD at short bond distances and comfortably outperforms it at longer distances. This is also evident in Table II, where the average error across the potential energy surface (PES) is shown. Remarkably, ADAPT($\epsilon_3$) outperforms UCCSD throughout the whole curve by at least an order of magnitude and in some cases up to four orders of magnitude.

| Iteration | Finite Differences, s | Analytical, s | x Speedup |
|-----------|-----------------------|--------------|-----------|
| 1         | 379.0                 | 2.6          | 145.7     |
| 2         | 405.2                 | 3.1          | 130.5     |
| 3         | 375.7                 | 2.7          | 140.4     |
| 4         | 377.6                 | 2.8          | 134.1     |
| 5         | 369.7                 | 2.6          | 144.7     |
| 6         | 383.2                 | 2.5          | 150.6     |
| 7         | 374.6                 | 3.0          | 126.9     |
| 8         | 369.0                 | 2.6          | 141.3     |
| 9         | 378.0                 | 3.1          | 122.7     |
| 10        | 397.4                 | 3.4          | 118.5     |
| Average   | 380.9                 | 2.8          | 134.6     |
Even more impressive is how few parameters are needed to achieve this level of accuracy. As shown in Fig. 2(a:iii), in all three cases and for all bond distances, ADAPT is much more compact than UCCSD. UCCSD has 92 parameters, which can be reduced to 64 by spin-adapting. In all three ADAPT calculations, less than half of the parameters are needed compared to UCCSD. Although UCCSD is noticeably more accurate than the sim-
ples ADAPT calculation with a gradient norm threshold of 0.1, the ADAPT\((\epsilon_1)\) ansatz is incredibly compact, consisting of less than 10 parameters across the curve. For example, the ADAPT\((\epsilon_1)\) ansatz for LiH at bond distance 2.39 Å is

\[
|\psi^{\text{ADAPT}(\epsilon_1)}\rangle = e^{\hat{z}_{01}^{12}} e^{\hat{z}_{06}^{22}} e^{\hat{z}_{33}^{34}} e^{\hat{z}_{33}^{34}} e^{\hat{z}_{36}^{66}} |\psi^{\text{HF}}\rangle,
\]

which includes a mixture of both double excitations and correlated single excitations \((\hat{n}_i \hat{a}_j^\dagger \hat{a}_k)\). The indices denote spatial orbitals, overbar on an index denotes \(\beta\) spin, and spin-adapted interactions are implied. For example \(\hat{z}_{01}^{12}\) is really \(\bar{\tau}_{01}^{12}\). An interesting feature of the ansatz returned by ADAPT-VQE, Eq. (19), is that the HOMO-LUMO double excitation \((e^{\hat{z}_{22}^{33}})\) is not the first operator, but instead the third. This is different from what one might expect if classical MP2 or CCSD amplitudes were used to order the ansatz. The reason is that in choosing the next operator no state energy information is used, for instance in the form of a denominator penalizing high energy terms. Interestingly, at convergence it is not the HOMO-LUMO term or the first operator with the largest amplitude, but rather the second operator, \(e^{\hat{z}_{36}^{66}}\).

\[
\begin{array}{|c|c|c|c|}
\hline
\text{LiH} & \text{ADAPT(}\epsilon_1)\text{)} & \text{ADAPT(}\epsilon_2\text{)} & \text{ADAPT(}\epsilon_3\text{)} \\
\hline
0.0480 & 0.3000 & 0.0058 & 0.0002 \\
2.2384 & 0.8023 & 0.0907 & 0.0041 \\
3.7387 & 4.5297 & 0.3023 & 0.0047 \\
\hline
\end{array}
\]

TABLE II: Average errors across the PES scan for the different methods assessed. Units in kcal/mol.

\[\text{UCCSD} \quad \text{ADAPT(}\epsilon_1\text{)} \quad \text{ADAPT(}\epsilon_2\text{)} \quad \text{ADAPT(}\epsilon_3\text{)}\]

\[\text{LiH} \quad 0.0480 \quad 0.3000 \quad 0.0058 \quad 0.0002\]

\[\text{BeH}_2 \quad 2.2384 \quad 0.8023 \quad 0.0907 \quad 0.0041\]

\[\text{H}_6 \quad 3.7387 \quad 4.5297 \quad 0.3023 \quad 0.0047\]

B. BeH\(_2\) double dissociation

In Fig. 2(b:i-iii), the dissociation curves for BeH\(_2\) are shown. In the STO-3G basis, BeH\(_2\) has 7 spatial orbitals, for a total Hilbert space dimension of 16384, and a neutral molecule subspace of dimension \((\binom{7}{2}) = 1225\). Unlike with LiH, UCCSD does not provide chemically accurate results across the full PES. UCCSD and ADAPT\((\epsilon_1)\) are comparable at smaller bond distances. Beyond \(\sim 3\) Å, they both go above 1 kcal/mol in absolute error. However, still with a small fraction of the number of parameters in UCCSD, both ADAPT\((\epsilon_2)\) and ADAPT\((\epsilon_3)\) provide nearly exact results, with average deviations from FCI listed in Table II.

C. Linear H\(_6\) dissociation

At bond-breaking, the previous two molecules involved strong correlation between only two and four electrons, respectively. In order to evaluate the ability of ADAPT-VQE to converge to FCI in the presence of much stronger correlations, we have computed the simultaneous stretching of H\(_6\), with the results presented in Fig. 2(c:i-iii).

The complexity of this strongly correlated system is reflected in two obvious ways: 1) the failure of UCCSD to achieve chemical accuracy across the curve in Fig. 2(c:ii), and 2) the increased number of parameters selected in the ADAPT calculations in Fig. 2(c:iii). Despite being strongly correlated, such that higher excitation rank operators should be needed, both ADAPT\((\epsilon_2)\) and ADAPT\((\epsilon_3)\) provide accurate results with only one- and two-body operators. Moreover, in the case of ADAPT\((\epsilon_2)\) this is achieved with fewer operators than UCCSD for most bond distances. ADAPT\((\epsilon_3)\) also uses fewer parameters than UCCSD up to the distance where UCCSD fails to reach chemical accuracy.

Because the algorithm is adaptive, during the course of a chemical event (bond breaking, isomerization, etc) the number of parameters can change abruptly, leading to discontinuous potential energy curves. Two notable examples of this can be seen in Fig. 2(c:ii), first at \(R(H-H) = 1.8\) Å where ADAPT\((\epsilon_1)\) experiences a large jump in energy, and second at 2.5 Å where ADAPT\((\epsilon_2)\) increases in energy. Fig. 2(c:iii) shows that these energy jumps correspond to sudden drops in parameter counts. Analyzing the convergence properties of the H\(_6\) system, we see that a local minimum in the gradient norm begins to appear with increasing bond distance. This local minimum first crosses the 0.1 threshold at 1.8 Å, and then crosses the 0.01 threshold at 2.5 Å. This is the cause of the discontinuities in the H\(_6\) data, and a plot of this data is provided in Appendix B. Of course, if a tighter threshold is used (such as 0.001), the ADAPT-VQE algorithm can climb out of this local minimum, and ultimately yield high-accuracy results, even for this strongly correlated system. More sophisticated convergence checks in step 5 might avoid these situations and will be one focus of future work.

D. Dependence of convergence on operator ordering

To demonstrate the importance of the gradient-based operator ordering chosen by ADAPT-VQE, we compare it to a few alternate procedures for growing the ansatz: a) Random (ijab): Randomly select from a pool of \(\tau_{ij}^{ab}\), where the indices are restricted to those which do not annihilate the HF reference state. b) Random (pqrs): Randomly select from a pool of \(\tau_{pq}^{rs}\), where the indices are not restricted. c) Lexical (ijab): Select from an ordered pool of \(\tau_{ij}^{ab}\), where the indices are restricted to those which do not annihilate the HF reference state. d) Lexical (pqrs): Select from an ordered pool of \(\tau_{pq}^{rs}\), where the indices are not restricted.

In Fig. 3 we show the convergence of each of these orderings and compare them to ADAPT using BeH\(_2\) as a typical example. What stands out is that the ADAPT ansatz converges dramatically faster than the other four
As the name suggests, ADAPT-VQE could be classified as one member of a family of adaptive-basis strategies that has seen success in constructing compact many-electron wavefunctions \[ \text{49,55} \]. Of these, the ADAPT ansatz is perhaps most closely related to the @-CC method of Lyakh and Bartlett \[ \text{49} \], in which a compact set of cluster operators is iteratively determined to describe the state of interest on a classical computer. Our approach is distinct in that it is not only designed for a quantum computer implementation, but also defined for a different wavefunction form (product of unitary operators vs. coupled cluster) and a different importance metric (operator gradient of the many-electron state vs. a single electron-defined importance function) for determining new parameters.

An important aspect of ADAPT-VQE is that several steps of the algorithm can be implemented in multiple ways, lending it still greater versatility across a wide landscape of problems and suggesting that it should perhaps be thought of as a class of algorithms rather than a specific one. In Secs. \text{[III]} and [IV], we already discussed a few algorithmic options, including different ways to perform the gradient-based parameter update and to determine convergence. We also mentioned the possibility of freezing early parameters at later stages of the algorithm in order to speed up the re-optimization steps. Below, we discuss a few more modifications to explore.

Although the ADAPT-VQE algorithm is notably not a perturbative approach, it still has a perturbative flavor in that the suitability of the next iteration’s best operator only involves the interaction of that operator with the Hamiltonian. As such, the algorithm may not be able to recognize the best quadruple excitation (for example) during one update. That being said, the physics described by quadruple excitations is ultimately captured after multiple iterations through the product of at least two two-body interactions. The consequence of this is that convergence will likely not be as fast for strongly correlated systems because the algorithm can only “see” two body operators at a time.

Fortunately, however, multiple strategies can be pursued to address any possible slow convergence issues. One possible approach would be to add a selection of three- or four-body interactions into the operator pool, such that these could be inserted when needed. Alternatively, one might imagine trying to update the ansatz with two (or more) operators in each iteration, such that the best set of operators is added. The operator pool would still consist of only one- and two-body interactions, but higher-body interactions could be incorporated through products of operators. Even further, one might imagine computing the second derivative and using Hessian matrix elements to identify cooperative effects between operators in the pool. We will explore each of these approaches in future work, with the aim of determining the fastest converging algorithm in different chemical scenarios.

V. DISCUSSION

As demonstrated above, ADAPT-VQE provides a compact and accurate wavefunction ansatz that easily outperforms UCCSD. The algorithm achieves this by systematically identifying the optimal set and ordering of operators to use in the wavefunction ansatz for a given problem. The efficiency of ADAPT-VQE makes it very promising for quantum chemistry simulations on NISQ devices, where circuit depth limitations remain a significant challenge. As hardware capabilities continue to increase, in terms of both the size and number of quantum processors available, ADAPT-VQE will offer an ideal quantum-parallel approach to performing nontrivial quantum chemistry simulations. We therefore expect this algorithm to have a strong impact on these efforts in the near term.

As the name suggests, ADAPT-VQE could be classified as one member of a family of adaptive-basis strategies that has seen success in constructing compact many-electron wavefunctions \[ \text{42,48} \] and single-electron wave-
VI. CONCLUSIONS

In this paper, we presented ADAPT-VQE, a novel variational hybrid quantum-classical algorithm designed to achieve exact results at convergence. Unlike typical ansatze, which tend to be ad hoc, our approach is based on an ansatz that is determined by the system being simulated, and it features a well-defined, built-in convergence criterion. Moreover, the parameter count, and thus the gate depth, is kept to a minimum. A detailed description of the algorithm is given, and numerical examples are provided to demonstrate the performance of the ADAPT method with both weakly and strongly correlated systems. Based on these results, we find the ADAPT-VQE algorithm to be an operator- and parameter-efficient method capable of high accuracy, with controllable errors, that routinely outperforms UCCSD. Its compatibility with classical pre-compiling and quantum-parallelism should make ADAPT-VQE extremely useful for simulations of molecules on both currently available and future quantum computers.

VII. ACKNOWLEDGEMENTS

This research was supported by the US Department of Energy (Award No. DE-sc0019199) and the National Science Foundation (Award No. 1839136). S.E.E. also acknowledges support from Award No. DE-sc0019318 from the Department of Energy.

Appendix A: Analytic Gradients

This section presents a derivation of the analytic gradient method used to improve both the computational efficiency and the numerical precision of the classical simulation of the parameter re-optimization that occurs in step 6 of the ADAPT-VQE algorithm, as described in Sec. [III] It should be noted, however, that the gradient algorithm we present here is in fact completely general and could be used for an arbitrary Trotterized ansatz, not just ADAPT.

Consider a general Trotterized ansatz, $|\phi\rangle$, obtained by applying $N$ excitation operators on a given a reference state $|\psi_0\rangle$:

$$|\phi\rangle = \prod_{i=1}^{N} e^{\theta_i T_i} |\psi_0\rangle,$$  \hspace{1cm} (A1)

where the $T_i = t_i - t_i^\dagger$ are a set of anti-Hermitian operators, the $\theta_i$ are free parameters, and the product is defined such that the terms with smaller values of the index $i$ appear on the right, i.e., $\prod_{i=1}^{N} \theta_i = \theta_N \theta_{N-1} \ldots \theta_1$. Assuming $|\psi_0\rangle$ is normalized, the functional to be minimized is then

$$E_\phi = \langle \phi | \hat{H} | \phi \rangle.$$  \hspace{1cm} (A2)

The state $|\phi\rangle$ can be obtained without direct matrix exponentiation and in this work we use the Scipy `expm_multiply` function to compute the action of a matrix exponential on a vector, which is relatively efficient since the operator in each exponent in Eq. A1 is incredibly sparse.

In order to perform a gradient-based minimization, the derivative of the energy with respect to each parameter in the ansatz is needed. The analytical derivatives then follow directly as

$$\frac{\partial E_\phi}{\partial \theta_i} = \langle \phi | \hat{H} \prod_{j=i+1}^{N} (e^{\theta_j T_j}) \prod_{k=1}^{i} (e^{\theta_k T_k}) |\psi_0\rangle - \langle \psi_0 | \prod_{j=i}^{i+1} (e^{-\theta_j T_j}) \prod_{k=i}^{N} (e^{-\theta_k T_k}) \hat{H} | \phi \rangle,$$  \hspace{1cm} (A3)

where in the second line, we use the convention where products with indices starting at higher values correspond to a reverse-order product in which higher-index terms are on the right, i.e., $\prod_{j=1}^{i} \theta_j = \theta_1 \theta_2 \theta_3 \ldots \theta_i$. Recognizing that the second line is just the adjoint of the first, Eq. A3 can be simplified to

$$\frac{\partial E_\phi}{\partial \theta_i} = 2\Re \left( \langle \phi | \hat{H} \prod_{j=i+1}^{N} (e^{\theta_j T_j}) \prod_{k=1}^{i} (e^{\theta_k T_k}) |\psi_0\rangle \right).$$  \hspace{1cm} (A4)

In order to obtain a more compact expression for the $i$th parameter derivative, we partially expand the wavefunction expression in Eq. A1 grouping the unitaries up to and after $i$ to define the notation for a state partially rotated with unitaries from 1 to $i$ as $|\psi_{1,i}\rangle$:

$$|\phi\rangle = \prod_{j=i+1}^{N} e^{\theta_j T_j} \left[ \prod_{k=1}^{i} e^{\theta_k T_k} \right] |\psi_0\rangle,$$  \hspace{1cm} (A5)

$$= \prod_{j=i+1}^{N} e^{\theta_j T_j} |\psi_{1,i}\rangle,$$  \hspace{1cm} (A6)

$$= |\psi_{1,N}\rangle.$$  \hspace{1cm} (A7)

Using this notation and defining $|\sigma_i\rangle = \prod_{j=i}^{N} (e^{-\theta_j T_j}) \hat{H} | \phi \rangle$ where $|\sigma_N\rangle = \hat{H} | \phi \rangle$, Eq. A4 becomes

$$\frac{\partial E_\phi}{\partial \theta_i} = 2\Re \langle \sigma_i | T_i | \psi_{1,i}\rangle.$$  \hspace{1cm} (A8)

While this analytical gradient formula implemented $N$ times (one for each parameter) is already an improvement over its numerical counterpart in terms of precision, it can be defined recursively to provide a significant improvement for computational efficiency. The gradient algorithm proceeds by computing one parameter gradient using precomputed quantities from the previous parameter’s gradient evaluation, significantly reducing CPU cost. We start the algorithm, the base case, at

$$|\phi\rangle = \prod_{i=1}^{N} e^{\theta_i T_i} |\psi_0\rangle,$$  \hspace{1cm} (A9)
the left-most operator, $i = N$, where we need to calculate
\[ \frac{\partial E_\phi}{\partial \theta_N} = 2\mathbf{R} \langle \sigma_{N} | T_{N} | \psi_{1,N} \rangle. \]  
(A9)

At the second step, we need to compute
\[ \frac{\partial E_\phi}{\partial \theta_{N-1}} = 2\mathbf{R} \langle \sigma_{N-1} | T_{N-1} | \psi_{1,N-1} \rangle, \]  
(A10)

but this can be facilitated by using the fact that the new states $|\sigma_{N-1}\rangle$ and $|\psi_{1,N-1}\rangle$ needed at this stage can be obtained simply by applying a single matrix exponential to the states from the first step:
\[ |\sigma_{N-1}\rangle = e^{-\theta_{N} T_{N}} |\sigma_{N}\rangle, \]  
(A11)
\[ |\psi_{1,N-1}\rangle = e^{-\theta_{N} T_{N}} |\psi_{1,N}\rangle. \]  
(A12)

This continues to be the case at all later steps, where the states needed at each step can be similarly obtained from those at the previous step:
\[ |\sigma_{i}\rangle = e^{-\theta_{i+1} T_{i+1}} |\sigma_{i+1}\rangle, \]  
(A13)
\[ |\psi_{1,i}\rangle = e^{-\theta_{i+1} T_{i+1}} |\psi_{1,i+1}\rangle. \]  
(A14)

The benefit of this is that instead of applying all the Trotter operators in the ansatz for each single parameter, each parameter gradient only requires two `expm_multiply` calls and one sparse matrix vector multiplication. Stacking the $|\sigma_{i}\rangle$ and $|\psi_{1,i}\rangle$ into a two column matrix and computing the exponential action on that offers a numerical speedup in practice. As a consequence, obtaining the gradient with respect to all parameters is only roughly twice the cost of a single energy evaluation. Assuming the computational cost of other operations to be negligible, this recursive algorithm reduces the cost of computing a gradient from $N^2$ to $3N$. In the case of water in the STO-3G basis (140 operators) this represents a nearly 47-fold improvement. This was approximately reflected in practice during the first ten optimization steps of water in Table I.

**Appendix B: Energy discontinuities**

As an adaptive algorithm, ADAPT-VQE self-consistently determines a quasi-optimal ansatz to describe the ground state of a particular Hamiltonian. However, as the nuclei move, a new Hamiltonian is obtained. This means that any two distinct points on a potential energy surface will generally be treated with different ansatze. Not only will the two points generally have a different number of operators, but the order and composition of the ansatz will also generally differ. This means that the typical reliance on energy cancellation is not expected to work, and therefore, one must converge the ansatz growth tightly enough for the desired application.

As discussed in the text, two notable examples of this discontinuous behaviour occur for $H_6$. In order to better illustrate how these discontinuities are created, we plot the norm of the gradient vector as a function of the iteration (i.e., number of parameters). A separate convergence plot is shown for each of the molecules, LiH, BeH$_2$, and $H_6$ in Fig. A.4.

While LiH and BeH$_2$ both exhibit rapid convergence, the strongly correlated $H_6$ has a distinctly different behavior. For short bond lengths (lighter colored lines) similar convergence behavior as the previous molecules is observed. In contrast, for longer bond lengths, a local minimum in the gradient norm begins to appear, requiring an increase in the number of parameters for tight convergence. This does not necessarily mean that the ansatz grown by ADAPT-VQE is not compact. $H_6$ is in-
trinsically more complicated, especially starting from a HF reference, and so it is expected to require more operators (and parameters). However, the consequence of this local minimum is that sometimes the algorithm may signal convergence prematurely, thus ignoring the extra parameters that might be needed to make the ansatz sufficiently accurate. In Fig. A.4(c), the difference between R = 2.4 and 2.5 Å involves a significant reduction in the number of parameters for ADAPT($\varepsilon_2$), due to the fact that the 2.5 curve has a local minimum which crosses the convergence threshold before actual convergence is achieved. The 2.4 Å local minimum is less deep, and doesn’t cross the threshold. As such, the algorithm needs to climb out of the local minimum to find the “true” convergence point where the gradient finally drops below the threshold with a larger number of parameters. The consequence of this is the significant drop in the number of parameters (and accuracy) going from 2.4 to 2.5 Å. Future work will focus on analyzing difficult cases such as these to determine better convergence criteria that are less susceptible to local minimum convergence.

[1] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Science 309 (2005), URL http://science.sciencemag.org/content/309/5747/1704
[2] Y. Y. Kitaev, arXiv e-prints quant-ph/9511026 (1995), quant-ph/9511026.
[3] S. Lloyd, Science 273, 1073 (1999), ISSN 0036-8075, http://science.sciencemag.org/content/273/5278/1073.full.pdf, URL http://science.sciencemag.org/content/273/5278/1073/
[4] B. P. Lanyon, J. D. Whitfield, G. G. Gillett, M. E. Goggin, M. F. Almeida, J. Kassal, J. D. Biamonte, M. Moshini, B. J. Powell, M. Barbieri, et al., Nat. Chem. 2, 106 (2010).
[5] O’Malley, R. Babbush, I. Kivlichan, J. Romero, J. McClean, R. Barends, J. Kelly, P. Roushan, A. Trant, N. Ding, et al., Physical Review X 6, 031007 (2016), ISSN 2160-3308, URL https://link.aps.org/doi/10.1103/PhysRevX.6.031007
[6] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O’Brien, and M. G. Thompson, Phys. Rev. Lett. 118, 100503 (2017), URL https://link.aps.org/doi/10.1103/PhysRevLett.118.100503
[7] J. Preskill, Quantum 2, 79 (2018), ISSN 2521-327X, URL https://doi.org/10.22331/q-2018-08-06-79
[8] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, Nature communications 5, 4213 (2014), ISSN 2041-1723, 1304.3061, URL http://arxiv.org/abs/1304.3061
[9] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, Molecular Physics 109, 735 (2011), ISSN 0026-8976, URL http://www.tandfonline.com/doi/abs/10.1080/00268976.2011.552441
[10] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New Journal of Physics 18, 023023 (2016), ISSN 1367-2630, URL http://stacks.iop.org/1367-2630/18/i=2/a=023023?key=crossref.3f09e5324aab86cb9ff9fee03d285927
[11] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, Physical Review A 95, 042308 (2017), ISSN 2469-9926, URL http://link.aps.org/doi/10.1103/PhysRevA.95.042308
[12] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, et al. (2018), 1805.04340, URL http://arxiv.org/abs/1805.04340
[13] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Quantum Science and Technology 4, 014008 (2018), ISSN 2058-9565, URL http://stacks.iop.org/2058-9565/4/i=1/a=014008?key=crossref.0e7411499ad08e8c7997e37f733ca0d
[14] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, Physical Review X 8, 011021 (2018), ISSN 2160-3308, URL https://link.aps.org/doi/10.1105/PhysRevX.8.011021
[15] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley (2018), 1810.02327, URL http://arxiv.org/abs/1810.02327
[16] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature 549, 242 (2017), ISSN 0028-0836, URL http://www.nature.com/doi/10.1038/nature23879
[17] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, Phys. Rev. A 95, 020501 (2017), URL https://link.aps.org/doi/10.1103/PhysRevA.95.020501
[18] A. Mezzacapo, et al. (2018), 1805.04340, URL http://arxiv.org/abs/1801.01053
[19] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, et al., Phys. Rev. X 8, 031022 (2018), URL https://link.aps.org/doi/10.1103/PhysRevX.8.031022
[20] W. Kutzelnigg, Theoretica Chimica Acta 80, 349 (1991), ISSN 0040-5744, URL http://link.springer.com/10.1007/BF01174181
[21] A. G. Taube and R. J. Bartlett, International Journal of Quantum Chemistry 106, 3393 (2006), ISSN 0021-9606, URL http://doi.wiley.com/10.1002/qua.21198
[22] G. Harsha, T. Shiozaki, and G. E. Scuseria, The Journal of Chemical Physics Letters 155, 133 (1989), ISSN 0009-2614, URL https://www.sciencedirect.com/science/article/pii/S0009261489873725
[23] M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, Physical Review Letters 113, 2949 (2000), ISSN 0021-9606, URL http://aip.scitation.org/doi/abs/10.1063/1.1287275
[26] H. Nakatsuji and M. Ehara, The Journal of Chemical Physics 117, 9 (2002), ISSN 0021-9606, URL http://aip.scitation.org/doi/10.1063/1.1487830

[27] H. Nakatsuji and E. R. Davidson, The Journal of Chemical Physics 115, 2000 (2001), ISSN 0021-9606, URL http://aip.scitation.org/doi/10.1063/1.1338302

[28] M. Nooijen, Physical Review Letters 84, 2108 (2000), ISSN 0031-9007, URL https://link.aps.org/doi/10.1103/PhysRevLett.84.2108

[29] I. G. Ryabinkin, T.-C. Yen, S. N. Genin, and A. F. Izmaylov (2018), 1809.03827, URL http://arxiv.org/abs/1809.03827

[30] C. F. Bender and E. R. Davidson, Physical Review 183, 23 (1969), ISSN 0031-899X, URL https://link.aps.org/doi/10.1103/PhysRev.183.23

[31] P. E. M. Siegbahn, International Journal of Quantum Chemistry 18, 1229 (1980), ISSN 0020-7608, URL http://doi.wiley.com/10.1002/qua.560180505

[32] H. Nakatsuji and E. R. Davidson, The Journal of Chemical Physics 115, 2000 (2001), ISSN 0021-9606, URL http://aip.scitation.org/doi/10.1063/1.1338302

[33] S. R. White, Physical Review Letters 69, 2863 (1992), ISSN 0031-9007, URL http://link.aps.org/doi/10.1103/PhysRevLett.69.2863

[34] O. K.-L. Chan and D. Zgid, Annual Reports in Computational Chemistry 5, 149 (2009), URL http://dx.doi.org/10.1016/978-0-12-373657-6

[35] G. K.-L. Chan and M. Head-Gordon, The Journal of Chemical Physics 116, 4462 (2002), ISSN 00219606, URL http://scitation.aip.org/content/aip/journal/jcp/116/11/10.1063/1.1449459

[36] N. Hatano and M. Suzuki (Springer, Berlin, Heidelberg, 2005), pp. 37–68, URL http://link.springer.com/10.1007/11526216f_12

[37] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, et al., Physical Review A 98, 022322 (2018), ISSN 2469-9926, URL http://doi.wiley.com/10.1002/qua.22322

[38] J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Ferrmann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, et al., Wiley Interdisciplinary Reviews: Computational Molecular Science 2, 556 (2012), ISSN 17590876, URL http://doi.wiley.com/10.1002/wcms.93

[39] D. G. A. Smith, L. A. Burns, D. A. Sirinanni, D. R. Nascimento, A. Kumar, A. M. James, J. B. Schirber, T. Zhang, B. Zhang, A. S. Abbott, et al., Journal of Chemical Theory and Computation 14, 3504 (2018), ISSN 15499618, URL http://pubs.acs.org/doi/10.1021/acs.jctc.8b00286

[40] J. R. McClean, I. D. Kivlichan, K. J. Sung, D. S. Steiger, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, T. Häner, et al. (2017), 1710.07629, URL http://arxiv.org/abs/1710.07629

[41] E. Jones, O. Travis, and P. Peterson, SciPy: Open source scientific tools for Python (2001–), URL http://www.scipy.org/

[42] R. J. Harrison, The Journal of Chemical Physics 94, 5021 (1991), ISSN 0021-9606, URL http://aip.scitation.org/doi/10.1063/1.460537

[43] A. Povil, J. Rubio, and F. Illas, Theoretica Chimica Acta 82, 229 (1992), ISSN 0040-5744, URL http://link.springer.com/10.1007/BF0113255

[44] G. Peris, J. P. Planelles, J.-P. Malrieu, and J. Paldus, The Journal of Chemical Physics 110, 11708 (1999), ISSN 0021-9606, URL https://aip.scitation.org/doi/10.1063/1.479116

[45] F. A. Evangelista, The Journal of Chemical Physics 140, 124114 (2014), ISSN 0021-9606, URL http://aip.scitation.org/doi/10.1063/1.4869192

[46] J. B. Schriber and F. A. Evangelista, The Journal of Chemical Physics 144, 161106 (2016), ISSN 0021-9606, URL http://aip.scitation.org/doi/10.1063/1.4948308

[47] A. A. Holmes, N. M. Tubman, and C. J. Umrigar, Journal of Chemical Theory and Computation 12, 3674 (2016), ISSN 1549-9618, URL http://pubs.acs.org/doi/10.1021/acs.jctc.6b00407

[48] E. Xu, M. Uejima, and S. L. Ten-no, Physical Review Letters 121, 113001 (2018), ISSN 0031-9007, URL http://link.aps.org/doi/10.1103/PhysRevLett.121.113001

[49] D. T. Lyakh and R. J. Bartlett, The Journal of Chemical Physics 133, 244112 (2010), ISSN 0021-9606, URL http://aip.scitation.org/doi/10.1063/1.3515476

[50] F. A. Bischoff and E. F. Valeev, The Journal of Chemical Physics 139, 114106 (2013), ISSN 0021-9606, URL http://aip.scitation.org/doi/10.1063/1.4820404

[51] D. N. Laikov, International Journal of Quantum Chemistry 111, 2851 (2011), ISSN 00207608, URL http://doi.wiley.com/10.1002/qua.22767

[52] W. C. Lu, C. Z. Wang, M. W. Schmidt, L. Bytautas, K. M. Ho, and K. Ruedenberg, The Journal of Chemical Physics 120, 2629 (2004), ISSN 0021-9606, URL http://aip.scitation.org/doi/10.1063/1.1638731

[53] G. Berghold, M. Parrinello, and J. Hutter, The Journal of Chemical Physics 116, 1800 (2002), ISSN 0021-9606, URL http://aip.scitation.org/doi/10.1063/1.1431270

[54] M. S. Lee and M. Head-Gordon, The Journal of Chemical Physics 107, 9085 (1998), ISSN 0021-9606, URL https://aip.scitation.org/doi/10.1063/1.475199

[55] O. Schütt and J. VandeVondele, Journal of Chemical Theory and Computation 14, 4168 (2018), ISSN 15499618, URL http://pubs.acs.org/doi/10.1021/acs.jctc.8b00378

[56] N. J. Higham and A. H. Al-Mohy, Acta Numerica 19, 159 (2010), ISSN 0962-4929, URL http://www.journals.cambridge.org/abstract_S0962492910000036

[57] A. H. Al-Mohy and N. J. Higham, SIAM Journal on Scientific Computing 33, 488 (2011), ISSN 1064-8275, URL http://epubs.siam.org/doi/10.1137/100788860