## Novel Quantum Chemistry Algorithms Based on the Variational Quantum Eigensolver

Harper R. Grimsley

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

Nicholas J. Mayhall, Chair T. Daniel Crawford Sophia E. Economou Eduard F. Valeyev

> December 2, 2022 Blacksburg, Virginia

Keywords: Quantum Chemistry, Quantum Computing, Variational Quantum Eigensolver, Unitary Coupled Cluster Copyright 2022, Harper R. Grimsley

## Novel Quantum Chemistry Algorithms Based on the Variational Quantum Eigensolver

Harper R. Grimsley

(ABSTRACT)

The variational quantum eigensolver (VQE) approach is currently one of the most promising strategies for simulating chemical systems on quantum hardware. In this work, I will describe a new quantum algorithm and a new set of classical algorithms based on VQE. The quantum algorithm, ADAPT-VQE, shows promise in mitigating many of the known limitations of VQEs: Ansatz ambiguity, local minima, and barren plateaus are all addressed to varying degrees by ADAPT-VQE. The classical algorithm family, O2DX-UCCSD, draws inspiration from VQEs, but is classically solvable in polynomial time. This group of algorithms yields equations similar to those of the linearized coupled cluster theory (LCCSD) but is more systematically improvable and, for X = 3 or  $X = \infty$ , can break single bonds, which LCCSD cannot do. The overall aim of this work is to showcase the richness of the VQE algorithm and the breadth of its derivative applications.

## Novel Quantum Chemistry Algorithms Based on the Variational Quantum Eigensolver

Harper R. Grimsley

#### (GENERAL AUDIENCE ABSTRACT)

A core goal of quantum chemistry is to compute accurate ground-state energies for molecules. Quantum computers promise to simulate quantum systems in ways that classical computers cannot. It is believed that quantum computers may be able to characterize molecules that are too large for classical computers to treat accurately. One approach to this is the variational quantum eigensolver, or VQE. The idea of a VQE is to use a quantum computer to measure the molecular energy associated with a quantum state which is parametrized by some classical set of parameters. A classical computer will use a classical optimization scheme to update those parameters before the quantum computer measures the energy again. This loop is expected to minimize the quantum resources needed for a quantum computer to be useful, since much of the work is outsourced to classical computers. In this work, I describe two novel algorithms based on the VQE which solve some of its problems.

# Dedication

To my sister, mother, and grandmother. Your unconditional support has meant the world to me.

# Acknowledgments

I must begin by thanking my mother, Ginger Grimsley, who was my first and loudest advocate throughout my education. Whether it was "encouraging" me to do extra math homework or fighting school bureacracy to get me in more advanced classes, she was a constant force in pushing me to achieve all I could in school. I am also grateful to her and the rest of my family for their constant emotional support. I want to thank my high school chemistry teacher, Aakash Kumar, for instilling in me an early love of chemistry and physics. I do not believe I would have pursued a career in chemistry without his teaching and friendship. I also want to thank my undergraduate advisor, Stefan Kilvanek, for tolerating my insufferable college persona while teaching me to be a researcher, and for encouraging me to switch to theory after the first lithium fire. His mentorship was invaluable in pushing me to attend graduate school. In a similar vein, I wish to think Peter Pulay, who was the first person to teach me any quantum chemistry, and who also steered me into theory. I want to thank my labmates, current and former, in the Mayhall group for their support, mentorship, and comaraderie. Day in and day out, your group members are the ones who push you forward. I finally wish to thank my advisor, Nicholas Mayhall, for his unwavering support. An outstanding family man and human being, he was always the first to nominate me for any award or help me with the job search that accompanies the end of a PhD. He was, at all times, understanding of my personal and professional limitations, and pushed me just the right amount to lead a scientifically productive and balanced life. Like most of the people that have worked for Nick, I am a better person and chemist for having known him. There are other friends, teachers, and professors that deserve my gratitude, but I will stop here for brevity. I would not be here if it were not for all of you, and so: Thank you.

# Contents

1	Ger	General Background and Introduction to Quantum Algorithms for Chem-				
	istry					
	1.1	The Electronic Schrödinger Equation	1			
	1.2	Hartree-Fock Theory	3			
	1.3	Electron Correlation	4			
	1.4 Quantum Information Basics		8			
		1.4.1 Qubits and Gates	8			
		1.4.2 Efficient Qubit Representation of States	9			
	1.5	The Variational Quantum Eigensolver	12			
	1.6	Attribution	16			
2	The Necessity and Ambiguity of Trotterization in VQE		19			
3	The ADAPT-VQE Method		31			
4	Local Minima and Barren Plateaus in the ADAPT-VQE Landscape		48			
5	Intr	Introduction to Classically Tractable Approximations				
	5.1	Classical Approximations to UCC	62			
	5.2	Linearized Coupled Cluster Theory	64			

6	The O2DX Framework	67
7	Conclusions	78
Bi	bliography	81

# Chapter 1

# General Background and Introduction to Quantum Algorithms for Chemistry

### 1.1 The Electronic Schrödinger Equation

A fundamental goal of electronic structure theory is the exact or approximate solution of the time-independent Schrödinger equation given in equation 1.1.

$$\hat{H} |\Psi\rangle = E |\Psi\rangle \tag{1.1}$$

where  $\hat{H}$  is the molecular Hamiltonian,  $|\Psi\rangle$  is a ground or excited state of the molecule, and E is the associated molecular energy. Let us introduce two assumptions which are typical in quantum chemistry:

- 1. The Born-Oppenheimer approximation holds, i.e. nuclei can be treated as static relative to the electrons in a molecule.
- 2. Relativistic effects beyond the existence of electronic spin can be safely ignored.

# Chapter 1. General Background and Introduction to Quantum Algorithms for Chemistry

Within these approximations, only the non-relativistic, electronic part of the Hamiltonian is of interest to us. This Hamiltonian,  $\hat{H}$ , is given in equation 1.2 in atomic units. [1]

$$\hat{H} = -\frac{1}{2} \sum_{i}^{N_e} \nabla_i^2 - \sum_{i}^{N_e} \sum_{j}^{N_N} \frac{Z_j}{|\hat{r}_i - \hat{R}_j|} + \frac{1}{2} \sum_{i \neq j}^{N_e} \frac{1}{|\hat{r}_i - \hat{r}_j|}$$
(1.2)

where  $N_e$  and  $N_N$  are the number of electrons and nuclei,  $Z_j$  is the charge of nucleus j,  $\hat{r}_i$  is the position of electron i, and  $\hat{R}_j$  is the position of nucleus j. To be clear, nuclearnuclear interactions are important for computing accurate energies, even within the Born-Oppenheimer approximation, but they are trivially included. The electronic Hamiltonian in equation 1.2 is cumbersome, and the first two terms are often collected into a set of one-electron terms  $\{h_i\}$  as in equation 1.3. [1]

$$\hat{H} = \sum_{i} h_i + \frac{1}{2} \sum_{i \neq j} \frac{1}{r_{ij}}$$
(1.3)

The notation has been streamlined as there is now no explicit nuclear dependence in the Hamiltonian. Note that we could have included all nuclear-nuclear interactions as a single, constant term, and the eigenstates of  $\hat{H}$  would not change. The energy would simply be shifted by this constant. We now have an explicit, simple form of the electronic Hamiltonian, which defines the electronic structure problem in terms of one- and two-electron interactions. We note in passing that the Schrödinger equation is trivial to solve for one electron and one nucleus (a hydrogenic atom). These solutions constitute atomic orbitals, which will be useful in the next section.

### **1.2** Hartree-Fock Theory

Let us now introduce a new approximation: A set of M spin orbitals can represent where electrons might be in real space. These spin orbitals are typically approximations to a finite subset of the infinite number of atomic orbitals obtained from treating each nucleus in a molecule independently and as a one-electron system. We call this set of spin orbitals the atomic orbital basis,  $\{|\chi_i\rangle\}$ . We can orthogonalize and normalize the atomic orbitals so that for all i and j,  $\langle \chi_i | \chi_j \rangle = \delta_{ij}$ .

We will now consider the form of a minimally complex wavefunction in terms of  $N_e$  orbitals. Naïvely, we could simply use

$$|\Psi\rangle = \chi_1(\mathbf{1})\chi_2(\mathbf{2})\dots\chi_{N_e}(\mathbf{N_e}),\tag{1.4}$$

which is called a Hartree product. While it is normalized, this tensor product of spin orbitals does not account for the antisymmetry of fermions. The simplest way to do so is to instead use a Slater determinant:

$$|\Psi\rangle = \frac{1}{\sqrt{N!}} \det \begin{vmatrix} \chi_1(\mathbf{1}) & \chi_2(\mathbf{1}) & \dots & \chi_{N_e}(\mathbf{1}) \\ \chi_1(\mathbf{2}) & \chi_2(\mathbf{2}) & \dots & \chi_{N_e}(\mathbf{2}) \\ \dots & \dots & \dots & \dots \\ \chi_1(\mathbf{N_e}) & \chi_2(\mathbf{N_e}) & \dots & \chi_{N_e}(\mathbf{N_e}) \end{vmatrix}.$$
(1.5)

From now on, we will use  $|\chi_1\chi_2\ldots\chi_{N_e}\rangle$  to refer to the associated Slater determinant.

Hartree-Fock (HF) theory attempts to answer the following question: What is the best way to mix the basis orbitals so that the occupied set of  $N_e$  molecular orbitals gives the best (i.e. lowest energy) determinant? This mixing can be understood from a physicist's perspective

#### Chapter 1. General Background and Introduction to Quantum Algorithms for Chemistry

as using unitary rotations to create the lowest-energy product state. From a chemist's perspective, one is minimizing a Lagrangian which is the sum of the single-determinant energy as a function of the orbitals and constraints enforcing orthonormality of the orbitals themselves:

$$\mathcal{L} = \langle \chi_1 \chi_2 \dots \chi_N | \hat{H} | \chi_1 \chi_2 \dots \chi_N \rangle - \sum_{ij} \epsilon_{ij} \left( \delta_{ij} - \langle \chi_i | \chi_j \rangle \right)$$
(1.6)

Minimizing this Lagrangian gives the HF energy, but the orbitals which minimize it are not well-defined. Arbitrary mixing between occupied orbitals or between virtual (unoccupied) orbitals will not change the wavefunction energy. When one creates the stationary condition  $\delta \mathcal{L} = 0$ , the multipliers  $\epsilon_{ij}$  appear in the equations. The specific details of these equations are beyond the scope of this work, but it is possible to solve the stationary condition with the constraint that  $\epsilon_{ij} = 0$  if  $i \neq j$ . Performing this constrained minimization gives the HF energy and a set of well-defined, "canonical" HF orbitals.

Hartree-Fock is frequently used interchangably with the self-consistent field (SCF) approach. We will use HF to refer specifically to the canonical orbitals, while SCF refers to any choice of orbitals which yields the HF energy. For more details on this process, the reader is encouraged to consult the textbook of Szabo and Ostlund. [2] The end result of HF is that one obtains a set of well-defined orbitals to be used in post-HF calculations, as well as a crude, variational approximation to the energy.

### **1.3** Electron Correlation

Unfortunately, Hartree-Fock is not generally able to reproduce experimental quantities such as reaction energies. Hartree-Fock is unable to reproduce experimental energies because

#### **1.3. Electron Correlation**

it fails to explicitly capture the instantaneous interactions between electrons, only treating electron-electron interactions in an average manner. Equivalently, we can say that even if one chooses the best single determinant to represent the wavefunction, they ignore all the other Slater determinants. These other determinants can be expressed in terms of excitations away from the reference determinant  $|0\rangle$ .

Excitations are represented as strings of creation operators  $\hat{a}_j^{\dagger}$  and annihilation operators  $\hat{a}_k$ . These operators create an electron in spin orbital j and annihilate an electron from spin orbital k respectively. Any Slater determinant can be represented as a string of only creation operators acting on the vacuum state  $|\rangle$ . The vacuum state has no electrons in it. We follow the convention of reference [2] and define the action of a creation operator by how it affects a Slater determinant:

$$\hat{a}_{j}^{\dagger} |\chi_{k}\chi_{l}\dots\chi_{n}\rangle = |\chi_{j}\chi_{k}\chi_{l}\dots\chi_{n}\rangle$$
(1.7)

In addition to representing Slater determinants this way, it is convenient to decompose operators into sums of excitations. These excitation operators typically have annihilation operators in them, whose action is defined as

$$\hat{a}_j |\chi_j \chi_k \dots \chi_n \rangle = |\chi_k \dots \chi_n \rangle.$$
(1.8)

It is important to note that these operators create or delete an electron in the first position of the Slater determinant. This importance is a result of the antisymmetry of the wavefunction. Rearranging the Slater determinant to a different order with a different first position can change the sign. For example:

$$|\chi_i \chi_j \rangle = -|\chi_j \chi_i \rangle \tag{1.9}$$

As a result of the Pauli exclusion principle, a creation operator cannot create two electrons in the same orbital. Additionally, an annihilation operator cannot destroy an electron which

# Chapter 1. General Background and Introduction to Quantum Algorithms for Chemistry

is not present in a Slater determinant. These requirements are manifested in the useful rules:

$$\hat{a}_i^{\dagger} \hat{a}_i^{\dagger} = 0 \tag{1.10}$$

$$\hat{a}_j \hat{a}_j = 0 \tag{1.11}$$

Due to the antisymmetric nature of the wavefunction, the operators also obey the canonical anticommutation relations:

$$\hat{a}_j \hat{a}_k = -\hat{a}_k \hat{a}_j \tag{1.12}$$

$$\hat{a}_j^{\dagger} \hat{a}_k^{\dagger} = -\hat{a}_k^{\dagger} \hat{a}_j^{\dagger} \tag{1.13}$$

$$\hat{a}_j \hat{a}_k^{\dagger} = \delta_{jk} - \hat{a}_k^{\dagger} \hat{a}_j \tag{1.14}$$

$$\hat{a}_j^{\dagger} \hat{a}_k = \delta_{jk} - \hat{a}_k \hat{a}_j^{\dagger} \tag{1.15}$$

These rules will be important later in the discussion of implementing operators on quantum computers.

For now, it is sufficient to understand the notation and meaning of excitation operators. For example:

$$|\phi_{ij}^{ab}\rangle = \hat{a}_{ij}^{ab} |0\rangle = \hat{a}_a^{\dagger} \hat{a}_b^{\dagger} \hat{a}_j \hat{a}_i |0\rangle.$$
(1.16)

We have used the traditional convention that  $i, j, \ldots$  refer to occupied orbitals and  $a, b, \ldots$ refer to virtual orbitals. Note that in this case, the operators are acting on  $|0\rangle$ , not the vacuum state. (The vacuum state can be trivially transformed into the reference state by a series of creation operators.) An exact wavefunction can be expressed as a linear combination

#### **1.3. Electron Correlation**

of all of these determinants:

$$|\Psi\rangle = c_0 |0\rangle + \sum_{ia} c_i^a |\phi_i^a\rangle + \sum_{\substack{i < j \\ a < b}} c_{ij}^{ab} |\phi_{ij}^{ab}\rangle + \dots$$
(1.17)

One can obtain the molecular ground state by diagonalizing the Hamiltonian in the basis of all determinants. The lowest eigenvalue is the ground state energy, while the corresponding eigenvector is the ground state wavefunction. Excited states correspond to higher eigenvalues of the Hamiltonian and their eigenvectors. This diagonalization approach is referred to as full configuration interaction, or FCI, and represents the exact energy within the chemistry model of the problem. Unfortunately, there are combinatorially many determinants for a molecule as a function of system size. This limits the systems that can be treated to very small active spaces of molecules. The largest FCI calculation ever performed was on 22 spatial orbitals with 22 electrons. [3]. Larger systems can be treated approximately in polynomial time by truncating the Hamiltonian to some subset of the determinants based on excitation rank and then diagonalizing. This is referred to as truncated CI. While the resulting energy in truncated CI is bounded below by the FCI energy, i.e., it is a variational method, truncated CI is not size-extensive. That is, the energy error of a system does not scale linearly as a function of system size. This means that as the system of interest becomes larger, we expect increasingly poor results. There are other methods which are sizeextensive but non-variational, i.e., not bounded below by the exact energy, including density functional theory (DFT), perturbation theory, and coupled cluster theory. Size-extensivity is widely regarded as more important than a variational bound on the energy, so these methods are used much more commonly than truncated CI. Unfortunately, non-variational methods tend to break down in the presence of significant multi-configurational character of the wavefunction. When  $|\left<0|\Psi\right>|^2$  is small, i.e., the reference is a poor approximation to the true wavefunction, we say that the system exhibits strong or static correlation. This

# Chapter 1. General Background and Introduction to Quantum Algorithms for Chemistry

is in contrast to weak or dynamical correlation which is generally well-described by the non-variational methods. Small systems can be addressed by FCI, and weakly correlated systems can be treated by the polynomially scaling methods. Large, strongly correlated systems, however, remain an unsolved problem.

Classic, real-world examples of strong correlation include transition metals, excited states, and bond-breaking. [4] A specific example of an open strong correlation problem is identifying the mechanism used by the iron-molybdenum cofactor (FeMoco) in nitrogenase. It is a strongly correlated system which has a conservative active space of 54 electrons in 54 spatial orbitals. However, a recent estimate suggests that an active space of 113 electrons in 76 orbitals is likely needed to obtain a reasonable ground-state energy. [5] Characterizing this molecule is of immense chemical interest, as nitrogenase uses it to catalyze ambient nitrogen fixation (That is, the high temperatures and pressures of the Haber-Bosch process are not required). Such molecules, as well as the earlier examples, are enough to motivate the development of novel methods that can treat large systems in the presence of strong correlation.

### **1.4 Quantum Information Basics**

#### 1.4.1 Qubits and Gates

The core difference between classical and quantum computing lies in the elementary unit of computation. Classical computers store information as bits, binary units of information that can be viewed as either  $|0\rangle$  or  $|1\rangle$ . Quantum computers, on the other hand, store information as quantum bits, or qubits. These can exist as superpositions of the form  $|\Psi\rangle = c_0 |0\rangle + c_1 |1\rangle$  where  $c_0$  and  $c_1$  are complex numbers which give a normalized one-qubit state (We will

ignore the possibility of many-level qudits in this work). By normalized, we mean that  $\langle \Psi | \Psi \rangle = 1$ . One qubit can theoretically store far more information than a classical bit, with the caveat that it must be prepared and measured multiple times to extract that information. However, the real power of quantum computing comes from the ability to entangle qubits. For two bits, the allowed states are  $|00\rangle$ ,  $|11\rangle$ ,  $|01\rangle$ , and  $|10\rangle$ . However, two quantum bits can exist as a linear combination of all four of these states. This capacity for quantum entanglement mirrors the multi-configurational correlation observed in electrons, making quantum computers an attractive approach for modeling electronic systems.

#### 1.4.2 Efficient Qubit Representation of States

Let us imagine a half-filled, closed shell active space with N electrons. There are  $\binom{N}{N/2}^2$  configurations which obey  $S_z$  symmetry and have the correct particle number. Each of these configurations will need eight bytes to store a double precision floating point number. This implies an exponential scaling of memory with respect to system size. The Jordan-Wigner transformation offers a scheme to map a set of electrons and spin orbitals to a set of spins which can be modeled by qubits. The appeal of this approach is that one can represent the state of M spin orbitals with only M qubits, since the M qubits can exist as a linear combination of several different configurations. [6] This implies a linear scaling of memory with respect to system size.

The realization of this linear scaling is, on paper, quite simple. Qubits in state  $|0\rangle$  correspond to unoccupied molecular spin orbitals, while qubits in state  $|1\rangle$  correspond to occupied molecular spin orbitals. Let us index our qubits from left to right to correspond to orbitals as  $|\chi_1\chi_2\ldots\chi_N\rangle$ . For example, the HF state of H<sub>2</sub> in a minimal basis would be  $|1100\rangle$ . The two bonding spin orbitals are occupied, and the two antibonding spin orbitals are unoccupied.

# Chapter 1. General Background and Introduction to Quantum Algorithms for Chemistry

Linear combinations like  $c_1 |1100\rangle + c_2 |0011\rangle$  can be used to express multi-configurational electronic states, without introducing more qubits than spin orbitals.

We still need to be able to express operators in the qubit basis, but the rules for converting the fermionic creation operators to qubit operators are straightforward. Creation and annihilation operators obey the transformation equations 1.18 and 1.19: [7]

$$a_k^{\dagger} = \frac{1}{2} \left( X_k - i Y_k \right) \otimes_{j < k} Z_j$$
 (1.18)

$$a_{k} = \frac{1}{2} \left( X_{k} + i Y_{k} \right) \otimes_{j < k} Z_{j}$$
(1.19)

Here, we use the elementary Pauli gates, which are defined as:

$$X = |1\rangle \langle 0| + |0\rangle \langle 1| \tag{1.20}$$

$$Y = i \left( \left| 1 \right\rangle \left\langle 0 \right| - \left| 0 \right\rangle \left\langle 1 \right| \right) \tag{1.21}$$

$$Z = |0\rangle \langle 0| - |1\rangle \langle 1| \tag{1.22}$$

To motivate equations 1.18 and 1.19, let us consider an arbitrary qubit product state where each qubit has  $b_j$  equal to 0 or 1:

$$|b_1\rangle |b_2\rangle \dots |b_n\rangle \equiv \prod_{j=1}^n \left( b_j \hat{a}_j^{\dagger} + (1-b_j) \mathbf{1} \right) |\rangle$$
(1.23)

The left-hand side of equation 1.23 is the qubit representation of a single configuration. The right-hand side is the Fermionic representation, expressed as a string of creation operators acting on the vacuum state.

Now let us consider a creation operator,  $\hat{a}_k^{\dagger}$  and its action on this configuration,  $|\phi\rangle$ . Assuming

#### **1.4. QUANTUM INFORMATION BASICS**

that  $j \neq k$ , we have

$$a_k^{\dagger} \left| \phi \right\rangle = a_k^{\dagger} \prod_{j=1}^n \left( b_j \hat{a}_j^{\dagger} + (1 - b_j) \mathbf{1} \right) \left| \right\rangle \tag{1.24}$$

$$= \left(-b_1 \hat{a}_1^{\dagger} + (1-b_1) \mathbf{1}\right) \hat{a}_k^{\dagger} \prod_{j=2}^n \left(b_j \hat{a}_j^{\dagger} + (1-b_j) \mathbf{1}\right) |\rangle.$$
(1.25)

Let us consider what has happened when we moved  $\hat{a}_k^{\dagger}$  to the right of  $\left(b_1 \hat{a}_1^{\dagger} + (1 - b_1) \mathbf{1}\right)$ . If  $b_1$  was 0, i.e., if qubit 1 was in state  $|0\rangle$ , then this rearrangement is described as

$$\hat{a}_{k}^{\dagger}\hat{a}_{1}^{\dagger} = -\hat{a}_{1}^{\dagger}\hat{a}_{k}^{\dagger}.$$
(1.26)

If  $b_1$  was 1, i.e., if qubit 1 was in state  $|1\rangle$ , then the rearrangement would be described as

$$\hat{a}_k^{\dagger} \mathbf{1} = \mathbf{1} \hat{a}_k^{\dagger}. \tag{1.27}$$

In a slight abuse of notation, we will now use Fermionic representations of the operators and qubit representations of the spin orbitals:

$$\hat{a}_{k}^{\dagger} \left| 0 \right\rangle_{1} = \left| 0 \right\rangle_{1} \hat{a}_{k}^{\dagger} \tag{1.28}$$

$$\hat{a}_k^{\dagger} \left| 1 \right\rangle_1 = - \left| 1 \right\rangle_1 \hat{a}_k^{\dagger} \tag{1.29}$$

The implication here is that until  $\hat{a}_k^{\dagger}$  "reaches" qubit k, it will do nothing to qubits in state  $|0\rangle$ , and will introduce a sign change to qubits in state  $|1\rangle$ . This is precisely the action of the Pauli Z operator. Consequently, we know that

$$\hat{a}_{k}^{\dagger} |\phi\rangle = \left( \bigotimes_{j=1}^{k-1} Z_{j} |b_{j}\rangle \right) \hat{a}_{k}^{\dagger} \left( \bigotimes_{j=k}^{n} |b_{j}\rangle \right)$$
(1.30)

#### Chapter 1. General Background and Introduction to Quantum Algorithms for Chemistry

Until this point, all manipulations would have worked exactly the same way if we had used  $\hat{a}_k$  instead of  $\hat{a}_k^{\dagger}$ . It is now important to distinguish between the two. We first consider the action of  $\hat{a}_k^{\dagger}$  on  $|b_k\rangle_k$ . The operator  $\hat{a}_k$  will have action defined by:

$$\hat{a}_{k}^{\dagger}\left|0\right\rangle_{k}=\left|1\right\rangle_{k}\tag{1.31}$$

$$\hat{a}_k^{\dagger} \left| 1 \right\rangle_k = 0 \tag{1.32}$$

The operator  $\hat{a}_k$  can be described with respect to qubit k as  $|1\rangle_k \langle 0|_k$ , which can be expressed with the Pauli operators as  $\frac{1}{2}(X - iY)$ . Note that in both cases, the operator  $\hat{a}_k$  disappears, so that we will not need a "Z-string" on the remaining qubits. Substitution into equation 1.30 yields equation 1.18. If we instead consider the action of  $\hat{a}_k$  on  $|b_k\rangle_k$ , we have:

$$\hat{a}_k \left| 0 \right\rangle_k = 0 \tag{1.33}$$

$$\hat{a}_k \left| 1 \right\rangle_k = \left| 0 \right\rangle_k \tag{1.34}$$

The operator  $\hat{a}_k$  is described with respect to qubit k as  $|0\rangle_k \langle 1|_k$ , which can be expressed with the Pauli operators as  $\frac{1}{2}(X+iY)$ . Again, the operator disappears so that there will be no Z-strings on the remaining qubits, yielding equation 1.19. Our states and operators can all be described in terms of excitation operators, so we can now prepare arbitrary states, measure expectation values of the Hamiltonian, and apply our own operators through these basic rules.

### 1.5 The Variational Quantum Eigensolver

A quantum algorithm known as quantum phase estimation (QPE) has been designed which can obtain the exact eigenvalues of a Hamiltonian on a quantum computer. [8, 9] Unfor-

#### **1.5. The Variational Quantum Eigensolver**

tunately, the algorithm requires deep circuits and therefore long coherence times of qubits. The qubits available today are not particularly robust, making the deep circuits of QPE impossible to realize on existing hardware. An alternative was proposed and demonstrated in 2013 in the form of the variational quantum eigensolver, or VQE. [10, 11]

The essential idea of a VQE is to introduce a classical set of parameters  $\boldsymbol{\theta}$  which will parametrize a unitary operator  $\mathcal{U}(\boldsymbol{\theta})$ . These parameters will be classically optimized, while the parameterized unitary is executed on a quantum processor, acting on some normalized reference  $|0\rangle$ . The expectation value of the problem Hamiltonian  $\hat{H}$  will be measured on the ansatz  $\mathcal{U}(\boldsymbol{\theta}) |0\rangle$  on the quantum device to yield its energy. This energy is fed back to a classical optimizer, which will update the parameters so that a new unitary can be enacted on the reference. This process is illustrated diagramatically in Fig. 1.1.

The variational aspect of a VQE comes from the fact that one is minimizing a Rayleigh quotient, given as:

$$E\left(\boldsymbol{\theta}\right) = \frac{\langle 0|\mathcal{U}^{\dagger}\left(\boldsymbol{\theta}\right) \, H\mathcal{U}\left(\boldsymbol{\theta}\right) \left|0\right\rangle}{\langle 0|\,\mathcal{U}^{\dagger}\left(\boldsymbol{\theta}\right) \, \mathcal{U}\left(\boldsymbol{\theta}\right) \left|0\right\rangle} \tag{1.35}$$

The denominator in equation 1.35 is unity regardless of  $\boldsymbol{\theta}$ , so one simply optimizes  $\langle 0|\mathcal{U}^{\dagger}(\boldsymbol{\theta}) \hat{H}\mathcal{U}(\boldsymbol{\theta}) |0\rangle$ . Aside from enforcing the variational principle, unitary operations are used because quantum operations correspond to physical evolutions of the quantum register, and therefore must be unitary processes.

At this point, one might ask what form  $\mathcal{U}(\boldsymbol{\theta})$  takes. Proposed answers are extremely diverse [12], but we will group them into three broad categories, neglecting adaptive methods for now:

- 1. Chemically-inspired unitaries. [10, 13–15]
- 2. Hardware-efficient unitaries. [16, 17]



Figure 1.1: A diagram showing the VQE paradigm.

#### 1.5. THE VARIATIONAL QUANTUM EIGENSOLVER

#### 3. Gate-free unitaries. [18]

We will focus on the chemically inspired ansätze in this work. Most chemically inspired ansätze are motivated by unitary coupled cluster (UCC) theory, an approach proposed in the 1970s by Werner Kutzelnigg. [19] The unitary coupled cluster ansatz, as originally proposed, takes the form:

$$\mathcal{U}(\boldsymbol{\theta})|0\rangle = e^{\hat{T} - \hat{T}^{\dagger}}|0\rangle \tag{1.36}$$

The  $\hat{T}$  operator is borrowed from coupled cluster theory. Typically,  $\hat{T}$  is truncated to the single and double excitations, i.e.  $\hat{T} = \hat{T}_1 + \hat{T}_2$ . Using the notation of  $\theta$  parameters rather than t-amplitudes, this operator takes the form:

$$\hat{T} = \sum_{ia} \theta_i^a \hat{a}_i^a + \sum_{\substack{i < j \\ a < b}} \theta_{ij}^{ab} \hat{a}_{ij}^{ab}$$
(1.37)

Unfortunately, the expectation value of the Hamiltonian on the ansatz in equation 1.36 cannot be efficiently evaluated on a classical computer. This is because the Baker-Campbell-Hausdorff (BCH) expansion, which truncates in traditional coupled cluster theory, does not truncate in UCC. On quantum hardware, however, such an ansatz could be efficiently prepared, with a caveat: It is not obvious how to implement the complex operator  $e^{\hat{T}-\hat{T}^{\dagger}}$ in an efficient way. Decomposition of this many-body operator into elementary one- and two-qubit gates yields an extremely large gate count and circuit depth. [20, 21] Typically, implementations of UCCSD will use a first-order Trotter approximation [22] of the form:

$$e^{\sum_{ia}\theta_i^a(\hat{a}_i^a - \hat{a}_a^i) + \sum_{ijab}\theta_{ij}^{ab}\left(\hat{a}_{ij}^{ab} - \hat{a}_{ab}^{ij}\right)} \approx \prod_{ia} e^{\theta_i^a\left(\hat{a}_i^a - \hat{a}_a^i\right)} \prod_{\substack{i < j \\ a < b}} e^{\theta_{ij}^{ab}\left(\hat{a}_{ij}^{ab} - \hat{a}_{ab}^{ij}\right)}$$
(1.38)

Note that the ordering of the right-hand side of equation 1.38 is not well-defined. The singles are placed in front of the doubles for notational simplicity, but there is no reason in particular

#### Chapter 1. General Background and Introduction to Quantum Algorithms for Chemistry

to choose such an ordering. The singles and doubles could be placed in any order. In chapter 2, we will discuss the importance of the choice of operator ordering, but will note in passing that the ambiguity of ordering is a significant problem, a solution of which will be addressed in chapter 3.

As a final note on VQEs in general, there are significant problems with classical optimization. Both local minima [23] and so-called barren plateaus [24] tend to appear in the parameter landscapes of VQEs. Local minima should be familiar to the reader. The basic idea of a barren plateau is that the gradient variance of the parameter landscape vanishes exponentially as a function of problem size, i.e. the number of qubits. This is equivalent to the appearance of a "narrow gorge" in the cost function (energy, in our case) landscape. [25] Inside of a very small region of the landscape, the energy gradient is large and a good minimum exists. Outside of that region, however, the energy gradient is very small. As the narrow gorge becomes more and more narrow, the odds of a given initialization having a good search direction become vanishingly small, making classical optimization impossible without some way of specifically targeting the gorge. Both barren plateaus and local minima are significant problems for the classical optimization of VQEs. We will address our approach to dealing with them in chapter 4.

### 1.6 Attribution

16

The work in chapter 2 had the following author contributions:

- Theory was developed by all authors.
- Code implementation was done by Harper Grimsley and Daniel Claudino.
- Simulations were performed by Harper Grimsley and Daniel Claudino.

#### **1.6.** Attribution

• All authors contributed to writing of the manuscript.

It is reprinted with permission from [20]. Copyright 2020 American Chemical Society. The work in chapter 3 had the following author contributions:

- Theory was developed by all authors.
- Code implementation was done by Harper Grimsley and Nicholas Mayhall.
- Simulations were performed by Nicholas Mayhall.
- All authors contributed to writing of the manuscript.

It is reprinted with permission from [26]. The creative commons license is available here. The work in chapter 4 had the following author contributions:

- Theory was developed by all authors.
- Code implementation was done by Harper Grimsley.
- Simulations were performed by Harper Grimsley.
- All authors contributed to writing of the manuscript.

It is a newer draft of the work in [27]. The non-exclusive license to distribute the draft on arXiv is available here.

The work in chapter 6 had the following author contributions:

- Theory was developed by both authors.
- Code implementation was done by Harper Grimsley.

Chapter 1. General Background and Introduction to Quantum Algorithms for Chemistry

• Simulations were performed by Harper Grimsley.

18

• Both authors contributed to writing of the manuscript.

It is reprinted with permission from [28]. Copyright 2022 American Chemical Society.

# Chapter 2

# The Necessity and Ambiguity of Trotterization in VQE

Letter

### Is the Trotterized UCCSD Ansatz Chemically Well-Defined?

Harper R. Grimsley,<sup>†</sup> Daniel Claudino,<sup>†®</sup> Sophia E. Economou,<sup>‡</sup> Edwin Barnes,<sup>‡</sup> and Nicholas J. Mayhall<sup>\*,†®</sup>

<sup>†</sup>Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, United States <sup>‡</sup>Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, United States

**Supporting Information** 

**ABSTRACT:** The variational quantum eigensolver (VQE) has emerged as one of the most promising near-term quantum algorithms that can be used to simulate many-body systems such as molecular electronic structures. Serving as an attractive ansatz in the VQE algorithm, unitary coupled cluster (UCC) theory has seen a renewed interest in recent literature. However, unlike the original classical UCC theory, implementation on a quantum computer requires a finite-order Suzuki-Trotter decomposition to separate the exponentials of the large sum of Pauli operators. While previous literature has recognized the nonuniqueness of different orderings of the operators in the Trotterized form of UCC methods, the question of whether or not different orderings matter at the *chemical scale* has not been



addressed. In this Letter, we explore the effect of operator ordering on the Trotterized UCCSD ansatz, as well as the much more compact k-UpCCGSD ansatz recently proposed by Lee et al. [J. Chem. Theory Comput., 2019, 15, 311. arXiv, 2019, quant-ph:1909.09114. https://arxiv.org/abs/1909.09114]. We observe a significant, system-dependent variation in the energies of Trotterizations with different operator orderings. The energy variations occur on a chemical scale, sometimes on the order of hundreds of kcal/mol. This Letter establishes the need to define not only the operators present in the ansatz but also the order in which they appear. This is necessary for adhering to the quantum chemical notion of a "model chemistry", in addition to the general importance of scientific reproducibility. As a final note, we suggest a useful strategy to select out of the combinatorial number of possibilities, a single well-defined and effective ordering of the operators.

The ability to accurately simulate chemistry at the subatomic level can provide deeper scientific insights and further reaching predictions than through experiment alone. Although exact simulation requires computational resources which increase exponentially with system size, many stable molecules can be accurately modeled using polynomially scaling techniques, providing accurate and interpretable results. Examples of such approximations include density-functional theory, perturbation theory, or coupled-cluster theory. To study more complicated systems with many strongly correlated electrons such as those involved in numerous catalytic systems or materials applications, more general modeling solutions are needed.

Quantum simulation, which has recently seen a dramatic increase in activity due to rapid developments in both hardware and algorithms, provides an exciting possibility for performing approximation-free simulations without the exponential computational cost plaguing classical simulations. Because the Hilbert space of a single spin—orbital can be mapped to the Hilbert space of a single qubit, the exponential growth of the molecular Hamiltonian is matched by the exponential growth of a quantum computer's Hilbert space. Consequently, a quantum computer with only tens of logical qubits could potentially demonstrate a quantum advantage.<sup>1–3</sup> While full error-correction is not

expected to be realized in the near future, so-called Noisy Intermediate Scaled Quantum (NISQ) devices<sup>4</sup> have interesting properties that might still offer important computational advantages.

While the first quantum algorithm proposed for simulating many-body systems, the Phase Estimation Algorithm (PEA),<sup>1,5</sup> provides a path for achieving arbitrarily accurate simulations, it does so at the cost of incredibly deep circuits. Because device noise and errors limit the number of gates that can be applied in sequence, PEA is not viable on NISQ devices. In 2014, Peruzzo and co-workers proposed and demonstrated an alternative algorithm termed the Variational Quantum Eigensolver (VQE)<sup>7</sup> which offers unique advantages for NISQ devices. Unlike PEA, VQE limits the depth of the circuit, which makes it possible to implement on current and near-term devices. However, this comes at the cost of an increased number of measurements and the introduction of a wave function ansatz that can limit the accuracy of the simulation (although our recent approach, ADAPT-VQE, can remove the ansatz error).8 The initial demonstration of VQE7 was followed by several theoretical

Received: October 29, 2019 Published: December 16, 2019 studies<sup>9–15</sup> and demonstrations on other hardware such as superconducting qubits<sup>10,14,16</sup> and trapped ions.<sup>17,18</sup>

A key ingredient in VQE is the ansatz, which is implemented as a quantum circuit which constructs trial wave functions that are measured and then updated in a classical optimization loop. The quality of the ansatz ultimately determines the accuracy of the simulated ground state energy and properties. In the original proposal, the unitary variant of coupled-cluster theory was chosen as an ansatz due to several attractive features:

- Accurate: Coupled-cluster theory is among the most accurate classical methods for many-body simulation.
- Well studied: The unitary variant of coupled-cluster singles and doubles (UCCSD) has been analyzed in detail in the context of classical simulations.<sup>19–22</sup>
- Unitary: Because a quantum circuit implements unitary operations, the unitary nature of UCCSD makes the approach natural in a VQE context.

The UCCSD ansatz is obtained by replacing the traditional Hermitian cluster operator terms in coupled cluster theory with anti-Hermitian operators

$$\begin{aligned} |\Psi_{\text{UCCSD}}\rangle &= e^{\mathcal{T}_1 + \mathcal{T}_2} |0\rangle \\ \hat{\mathcal{T}}_1 &= \sum_{ia} \theta_{ia} (a_a^{\dagger} a_i - a_i^{\dagger} a_a) \\ \hat{\mathcal{T}}_2 &= \sum_{ijab} \theta_{ijab} (a_a^{\dagger} a_b^{\dagger} a_i a_j - a_j^{\dagger} a_i^{\dagger} a_b a_a) \end{aligned}$$
(1)

where  $|0\rangle$  is the uncorrelated reference state, usually Hartree– Fock,  $a_p^{\dagger}(a_p)$  is a creation (annihilation) operator for the orbital indexed by p, and  $\{\theta_{ia}, \theta_{ijab}\}$  are the parameters to be variationally optimized.

Although the unitarity of UCCSD implies an ease of implementation on quantum hardware, gate-based quantum computing requires a decomposition of operations into one- and two-qubit gates, such as single-qubit rotations and CNOT gates. In contrast, complicating direct implementation, the  $\hat{\mathcal{T}}_n$  operators simultaneously act on N qubits. In principle, any unitary operation can be decomposed into one- and two-qubit gates.<sup>23</sup> However, the number of gates produced from such a decomposition grows rapidly with the number of qubits acted on by the unitary, making it desirable to use an approximation scheme such as Suzuki-Trotter<sup>24</sup> when implementing *N*-qubit unitary operators.

The first-order Suzuki-Trotter approximation is given by eq 2.

$$e^{\hat{A}+\hat{B}} \approx e^{\hat{A}}e^{\hat{B}} \tag{2}$$

This becomes exact in infinite order:

$$e^{\hat{A}+\hat{B}} = \lim_{n \to \infty} \left( e^{\frac{\hat{A}}{n}} e^{\frac{\hat{B}}{n}} \right)^n \tag{3}$$

To approximate UCCSD accurately using a product form, large Trotter numbers, *n*, could in principle be used. This would, of course, create extremely deep circuits, making quantum simulation intractable. Alternatively, one could choose an aggressive truncation such as that in eq 2. In general, this would provide a very poor approximation to the UCCSD wave function but would provide a relatively shallow circuit that is better for NISQ realization. Note the stark difference between the effect of "Trotterizing" the ansatz in VQE and Trotterizing the time-evolution operator for algorithms like PEA. In Trotterizing the evolution operator, the goal is to reproduce the dynamics of the original Hamiltonian. Any Trotter error destroys the dynamics, and thus convergence with respect to Trotter error is sought.<sup>25–27</sup> In contrast, when Trotterizing the ansatz in VQE it is generally accepted that the variational optimization can, in practice, absorb most of the energy difference between the conventional UCCSD and the Trotterized form.<sup>12,13,28</sup>

At this point we want to clarify some of the language used above. Despite having used the Suzuki-Trotter approximation as a motivation for separating out the ansatz into a product form, it is no longer appropriate to call this a Trotter approximation. The reason is that the Trotterization occurs before parameter optimization. Thus, one is actually variationally optimizing the parameters of the product form, and it no longer relates to the conventional UCCSD (in ref 29 Evangelista et al. refer to this as the disentangled form of UCCSD, opting to avoid the Trotterization language altogether). In fact, if one were to use the optimized parameters from the product form and insert them into the conventional UCCSD ansatz, the result would necessarily be higher in energy. Therefore, it is important to note that the term "Trotterized form" referred to throughout this Letter is not an approximation to the conventional UCCSD ansatz. It is instead a different ansatz altogether, a point easily made by recognizing that the Trotterized form can sometimes yield a lower energy than the conventional, yet variational, UCCSD.

Unfortunately, a problem of definition arises during Trotterization. Reordering the product approximation in eq 2 does not generally give the same result, except in the trivial case where the operators commute. With the number of operator orderings being a path enumeration problem, the number of possible ansatzes produced during Trotterization (and potentially reported in the literature) is exponentially large. This, of course, is not an issue in UCCSD, as a sum of operators has no dependence on the order in which they are summed.

The objective of this Letter is to determine if the term "Trotterized UCCSD" is sufficiently well-defined, such that the range of energies coming from different operator orderings falls within some notion of chemical accuracy (e.g., 1 kcal/mol), a term referred to in the title as chemically well-defined. If that were the case, then the term "Trotterized UCCSD" would be welldefined, as an arbitrary operator ordering would produce practically similar results. However, if changing the operator ordering significantly changes the accuracy on a chemical scale, then it proves necessary to provide more information to fully define an ansatz and to provide reproducible results. To answer this question, we perform classical simulations with randomly shuffled operators using a custom code built with Open-Fermion<sup>30</sup> and Psi4,<sup>31</sup> which uses the gradient algorithm we developed, which is outlined in the Appendix of ref 8. The results using various operator orderings are compared to both UCCSD and Full CI (FCI).

#### NUMERICAL EXAMPLES

We consider four molecules in the context of the UCCSD ansatz, LiH,  $H_6$ , BeH<sub>2</sub>, and  $N_2$  with its 1s and 2s orbitals frozen. All molecules are arranged in uniform, linear geometries with varying interatomic distances. For each system, we classically simulate the calculation of a potential energy curve using a large number of random operator orderings.

For each system, the minimal STO-3G basis is used to minimize computational cost (the implementations use the full



Figure 1. Potential energy curves relative to the FCI dissociation limit of each system for, from left to right,  $H_6$ , LiH, BeH<sub>2</sub>, and N<sub>2</sub> (top) and errors from FCI (bottom) for the UCCSD ansatz.

Hilbert space of the orbitals), and the restricted Hartree-Fock (RHF) singlet state is chosen as the reference state. The oneand two-electron integrals are computed with the Psi4 quantum chemical package.<sup>31</sup> The Hamiltonian, anti-Hermitian operators in the UCCSD ansatz, and reference state are formed in the qubit basis using the Jordan-Wigner transform in Open-Fermion.<sup>30</sup> At this point, the various orderings of ansatzes are constructed, and their parameters  $\{\theta_{ia}, \theta_{ijab}\}$  are optimized by the SciPy implementation of BFGS.<sup>32</sup> The potential energy curves are displayed in Figure 1 along with standard deviation plots and range plots. We additionally compare the random Trotter orderings to a "sequential gradient ordering" (SGO), a quasideterministic method where one operator with the largest gradient is added at a time, according to the prescription followed by the ADAPT-VQE ansatz construction. However, in contrast to the ADAPT-VQE, the SGO approach refrains from allowing inclusion of more than one instance of the same operator so that a direct comparison to the original UCCSD results can be made.

A cursory evaluation of the data suggests that the variance among different ansatzes increases with static correlation of the chemical system. Because these tend to be the systems of greatest chemical interest for VQE since they represent classically hard problems, the ability to choose good Trotter orderings is critical.

The UCCSD results for the first molecular PES,  $H_{67}$  are characterized by an accurate description near the equilibrium region, a quick increase in error upon bond breaking, and then a similarly rapid decrease in error as the bond is further stretched to dissociation. With five "bonds" being broken simultaneously, it is expected that UCCSD should fail to accurately describe this system. One interesting observation from this plot is that the *ordering variance* (the statistical variance of the energies computed with randomly shuffled operators) increases as the UCCSD error increases. In contrast to  $H_{67}$ , LiH is a relatively simple system, and we observe negligible ordering variance.

Regardless of Trotter ordering, the curves are all extremely good approximations.

Similar to  $H_6$ ,  $BeH_2$  exhibits a simultaneous quick rise in the ordering variance and UCCSD energy error. However, unlike  $H_6$ , the ordering variance decreases again after bond breaking, along with the UCCSD energy error. The range of values obtained from different orderings is of the same order of magnitude as the actual absolute error of the UCCSD energy.

Unlike both  $H_6$  and  $BeH_2$ , the UCCSD curve for  $N_2$  does not decrease in error after bond breaking but rather flattens out to a nearly constant error of around 10 kcal/mol. The ordering variance increases alongside the UCCSD error and also levels out, despite a significant jump occurring around 3.5 Å in the range of energy values obtained from the Trotterized ansatzes. This is due to at least one of the operator orderings getting stuck in a local minimum (the variational parameters are initialized to 0), which is a consequence of the highly nonlinear nature of the optimization.

Overall, we find that when static correlation appears, the energy differences between orderings increase. This can be understood from the fact that the differences between operator orderings depend on the commutators of the operators, and these in turn depend on the optimal parameter values, which tend to be larger when the electron correlation is stronger. (A system with no electron correlation would have an optimal solution with all parameters equal to zero.) As such it makes sense that for more strongly correlated systems, the differences between operator orderings increase. While uniquely welldefined (up to orderings of operators with degenerate gradients), the sequential gradient ordering scheme does not appear to be reliably better or worse than other orderings.

Alternative Ways To Reorder Operators. In Figure 1, a comparison is made between the un-Trotterized ansatz and a series of randomly shuffled Trotterizations. However, one could group the operators by excitation rank before Trotterization. This would result in a significantly reduced sampling space and

#### Journal of Chemical Theory and Computation

potentially provide more consistently accurate results. To address this possibility, we have computed the performance of multiple different orderings, such as grouping singles first and doubles second or doubles first and singles second. From these results, we find that it is generally favorable to apply double excitations to the reference first, followed by singles. This data is provided in the Supporting Information.

*k*-UpCCGSD. From the results in Figure 1, we notice that the ordering variance increases with error in the associated un-Trotterized ansatz. It seems then that when UCCSD is accurate, there may be an excess of operators, such that the extra operators (while not necessary for accurate energy estimates) are useful in minimizing the differences between different Trotterization orderings. To test this hypothesis, we have additionally considered the more compact *k*-UpCCGSD ansatz by Lee et al.,<sup>15</sup> which has far fewer parameters (for small *k*) than UCCSD, where *k* controls the number of variational parameters by considering *k* products of the ansatz with all generalized paired doubles and orbital rotations:

$$|\Psi_{k-\text{UpCCGSD}}\rangle = \prod_{i=1}^{k} (e^{\hat{T}^{(i)} - \hat{T}^{(i)\dagger}})|0\rangle$$
(4)

The k-UpCCGSD ansatz is a more economical parametrization where only the operators which are expected to be most important are included. This translates into having fewer excess parameters, such that higher accuracy can be reached with a comparable circuit depth by increasing k. Based on our results above, we would anticipate a higher ordering variance for small values of k (larger than UCCSD), but that by increasing k, one can make the energy error (and thus the ordering variance) arbitrarily small. On the other hand, it is worth mentioning that the improvement attained by increasing k is accompanied by placing a heavier burden on the classical optimizer, as it tends to exacerbate the highly nonlinear character of the underlying optimization, making it difficult to locate the global minimum. Moreover, the minima found by the optimizer show strong dependence on the initialization of the variational parameters. One way this can be circumvented in the cases involving the un-Trotterized version of the *k*-UpCCGSD, as presented in eq 4, is to perform many simulations with the variational parameters  $\hat{\theta}$ initialized at random, as suggested in ref 15 and carried out here by repeating the simulations at each bond length 100 times and taking the lowest energy value as the global minimum for each geometry. This leads to potential energy curves for k = 1, 2 that are smooth in the energy scale relevant in the current context. The variational parameters are initialized at 0 for all Trotterized ansatzes constructed based on eq 4, in line with what is detailed for the UCCSD ansatz and whose results are displayed in Figure 1.

Figure 2 shows simulation results for  $H_6$  with k = 1, 2 for 100 randomly sampled operator groupings. Several features of the performance of the different Trotterized versions of 1- and 2-UpCCGSD agree with the results for the Trotterized versions of the UCCSD ansatz. For short bond distances (<1.1 Å), there is an evident insensitivity of the energy with respect to a specific sampling of the operators. Despite being already fairly small in this regime with k = 1, this distinction is largely quenched when k = 2, rendering the results with differently sampled ansatzes visually identical on the scale of the plots.

The most remarkable divergences among the operator groupings and the size of the generator, that is, k = 1 vs k = 2, are observed when moving toward the limit of H<sub>6</sub> dissociating



**Figure 2.** Potential energy curves relative to the FCI dissociation limit of  $H_6$  into six hydrogen atoms (top) and errors from FCI (bottom) for the 1-UpCCGSD (left) and 2-UpCCGSD (right) ansatzes.

into six noninteracting hydrogen atoms. These results are consistent with the observations found in ref 33, which noticed that for k = 1 there were large differences in energy depending on whether one grouped or split the singles and doubles excitations. Ansatzes with different operator groupings start to deviate in the vicinity of the Coulson-Fischer point. In this region, none of the orderings that were sampled for the 1-UpCCGSD operators approach the corresponding un-Trotterized and FCI energies. Some of the ansatzes are able to get back on track in closely approaching the FCI dissociation limit, along with the un-Trotterized 1-UpCCGSD energies. These ansatzes happen to be largely comprised of double excitation operators flocked closer to the reference determinant, which is in line with the findings from the simulations with the SD orderings, provided in the Supporting Information. The 1-UpCCGSD ansatz tracks well the FCI results, being able to provide the correct qualitative behavior along the PES. However, this ansatz is quite compact, and its limited number of parameters impairs its ability to variationally achieve results that are quantitatively comparable to FCI. The operator ordering originated from the SGO construction closely follows the lowest energies from random operator samplings. It is worth pointing out that it is able to overcome the deficiencies around the Coulson-Fischer point and asymptotically recover the exact (FCI) dissociation limit, whereas the corresponding un-Trotterized ansatz cannot account for all the missing correlation as the H-H bonds are stretched. Because in the SGO ansatz the operators are added according to the magnitude of their gradient component, an ansatz with identical operators, such as k-UpCCGSD with k > 1cannot be unambiguously defined and that is why we do not report such results in Figure 2.

The disparities among operator groupings are largely removed all throughout the potential energy curves by doubling the number of variational parameters, accomplished by setting k = 2. We preserve the same orderings studied for k = 1, that is, we have a product of two Trotterized exponential generators wherein operators are not shuffled across the two instances of  $e^{\hat{T}^{(i)} - \hat{T}^{(i)\dagger}}$ . Except for a slight spread surrounding the Coulson-Fischer point which is the region most strongly correlated in the

#### Journal of Chemical Theory and Computation

potential energy curve, all of the different ansatzes behave in a strikingly similar fashion. The errors are largest in this region, and keeping in mind the different scales in the plots when changing k, they are significantly mitigated in comparison with k = 1, with all orderings approaching the FCI energy in the dissociation limit. The advantage due to a larger set of variational parameters is also reflected in the un-Trotterized version of the ansatz, 2-UpCCGSD, whose dissociation curve practically overlays with the FCI results. The significant improvement in the results with k = 2, accompanied by a virtually absent spread in the computed energies, is in agreement with the findings of Lee et al.,<sup>15</sup> which implies that these ansatzes are relatively insensitive to the ordering of the operators.

#### CONCLUSIONS

In this Letter, we sought to determine if the operator ordering in Trotterized UCCSD impacts the results in a "chemically meaningful" way, such that the differences between unique operator orderings produce results which differ on a chemical scale, i.e., greater than 1 kcal/mol. Our numerical simulations clearly demonstrate that the operator ordering has a significant effect (large energy differences between orderings) only when there is a significant amount of electron correlation. However, the renewed interest in UCCSD (and the relevance of the Trotterized form) is due to the use of the UCCSD ansatz in VQE simulations on quantum computers. Strongly correlated molecules are the primary target of quantum simulations, and so this makes the issue of operator ordering even more important. Consequently, the results in this paper emphasize that to ensure scientific reproducibility, it is necessary for authors to report the specific orderings used in simulations involving Trotterized ansatzes. These results strongly advocate for the use of a dynamic ansatz which uniquely determines the operator ordering, such as ADAPT-VQE,<sup>8</sup> or adopting an ansatz which does not require trotterization (such as the Jastrow-based approach in ref 33). Our findings also suggest that there are systematic patterns to which Trotter orderings will give the lowest energy, offering a useful route to defining useful and unique operator orderings.

#### ASSOCIATED CONTENT

#### **Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.9b01083.

Plots for simulations with operators shuffled only within excitation rank sets and details on how to access code used to obtain results (PDF)

Specific operator orderings and corresponding energy values for data used in Figure 2 (XLSX)

#### AUTHOR INFORMATION

#### **Corresponding Author**

\*E-mail: nmayhall@vt.edu.

#### ORCID 💿

Daniel Claudino: 0000-0002-8860-0689 Nicholas J. Mayhall: 0000-0002-1312-9781

#### Notes

The authors declare no competing financial interest.

#### ACKNOWLEDGMENTS

This research was supported by the U.S. 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 U.S. Department of Energy.

#### REFERENCES

(1) Aspuru-Guzik, A.; Dutoi, A. D.; Love, P. J.; Head-Gordon, M. Simulated Quantum Computation of Molecular Energies. *Science* **2005**, 309, 1704–1707.

(2) McArdle, S.; Endo, S.; Aspuru-Guzik, A.; Benjamin, S.; Yuan, X. Quantum computational chemistry. 2018, arXiv:1808.10402. https://arXiv.org/abs/1808.10402 (accessed Dec 17, 2019).

(3) Cao, Y.; Romero, J.; Olson, J. P.; Degroote, M.; Johnson, P. D.; Kieferová, M.; Kivlichan, I. D.; Menke, T.; Peropadre, B.; Sawaya, N. P. D.; Sim, S.; Veis, L.; Aspuru-Guzik, A. Quantum Chemistry in the Age of Quantum Computing. *Chem. Rev* **2019**, *119*, 10856–10915.

(4) Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79.

(5) Kitaev, A. Y. Quantum measurements and the Abelian Stabilizer Problem. 1995. quant-ph/9511026. arXiv e-prints. https://arxiv.org/abs/quant-ph/9511026 (accessed Dec 17, 2019).

(6) Lloyd, S. Universal Quantum Simulators. Science 1996, 273, 1073–1078.

(7) Peruzzo, A.; McClean, J.; Shadbolt, P.; Yung, M.-H.; Zhou, X.-Q.; Love, P. J.; Aspuru-Guzik, A.; O'Brien, J. L. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* 2014, *5*, 4213.
(8) Grimsley, H. R.; Economou, S. E.; Barnes, E.; Mayhall, N. J. An

adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nat. Commun.* **2019**, *10*, 3007.

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

(10) O'Malley, P.; et al. Scalable Quantum Simulation of Molecular Energies. *Phys. Rev. X* 2016, *6*, 031007.

(11) McClean, J. R.; Kimchi-Schwartz, M. E.; Carter, J.; de Jong, W. A. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A: At., Mol., Opt. Phys.* **2017**, 95, 042308.

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

(13) Romero, J.; Babbush, R.; McClean, J. R.; Hempel, C.; Love, P. J.; Aspuru-Guzik, A. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Sci. Technol.* **2019**, *4*, 014008.

(14) Colless, J. I.; Ramasesh, V. V.; Dahlen, D.; Blok, M. S.; Kimchi-Schwartz, M. E.; McClean, J. R.; Carter, J.; de Jong, W. A.; Siddiqi, I. Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm. *Phys. Rev. X* **2018**, *8*, 011021.

(15) Lee, J.; Huggins, W. J.; Head-Gordon, M.; Whaley, K. B. Generalized Unitary Coupled Cluster Wave functions for Quantum Computation. J. Chem. Theory Comput. **2019**, 15, 311–324.

(16) Kandala, A.; Mezzacapo, A.; Temme, K.; Takita, M.; Brink, M.; Chow, J. M.; Gambetta, J. M. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **2017**, *549*, 242–246.

(17) Shen, Y.; Zhang, X.; Zhang, S.; Zhang, J.-N.; Yung, M.-H.; Kim, K. Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure. *Phys. Rev. A: At., Mol., Opt. Phys.* **2017**, *95*, 020501.

(18) Hempel, C.; Maier, C.; Romero, J.; McClean, J.; Monz, T.; Shen, H.; Jurcevic, P.; Lanyon, B. P.; Love, P.; Babbush, R.; Aspuru-Guzik, A.; Blatt, R.; Roos, C. F. Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator. *Phys. Rev. X* **2018**, *8*, 031022.

#### Journal of Chemical Theory and Computation

(20) Kutzelnigg, W. Error analysis and improvements of coupledcluster theory. *Theoret. Chim. Acta* **1991**, *80*, 349–386.

(21) Taube, A. G.; Bartlett, R. J. New perspectives on unitary coupledcluster theory. *Int. J. Quantum Chem.* **2006**, *106*, 3393–3401.

(22) Harsha, G.; Shiozaki, T.; Scuseria, G. E. On the difference between variational and unitary coupled cluster theories. *J. Chem. Phys.* **2018**, *148*, 044107.

(23) Nielsen, M. A.; Chuang, I. L. Quantum Computation and Quantum Information, 10th Anniversary ed.; Cambridge University Press: 2010.

(24) Hatano, N.; Suzuki, M. Finding Exponential Product Formulas of Higher Orders. In *Quantum Annealing and Other Optimization Methods*; Springer: Berlin, Heidelberg, 2005; pp 37–68, DOI: 10.1007/11526216 2.

(25) Babbush, R.; McClean, J.; Wecker, D.; Aspuru-Guzik, A.; Wiebe, N. Chemical Basis of Trotter-Suzuki Errors in Quantum Chemistry Simulation. *Phys. Rev. A: At., Mol., Opt. Phys.* **2015**, *91*, 022311.

(26) Heyl, M.; Hauke, P.; Zoller, P. Quantum Localization Bounds Trotter Errors in Digital Quantum Simulation. *Science Advances* **2019**, *5*, No. eaau8342.

(27) Sieberer, L. M.; Olsacher, T.; Elben, A.; Heyl, M.; Hauke, P.; Haake, F.; Zoller, P. Digital Quantum Simulation, Trotter Errors, and Quantum Chaos of the Kicked Top. *npj Quantum Information* **2019**, *5*, 78.

(28) Rubin, N. C. A Hybrid Classical/Quantum Approach for Large-Scale Studies of Quantum Systems with Density Matrix Embedding Theory. 2016. arXiv:1610.06910 [cond-mat, physics:quant-ph]. arXiv: 1610.06910. https://arxiv.org/abs/1610.06910 (accessed Dec 17, 2019).

(29) Evangelista, F. A.; Chan, G. K.-L.; Scuseria, G. E. Exact Parameterization of Fermionic Wave Functions via Unitary Coupled Cluster Theory. 2019. arXiv:1910.10130 [cond-mat, physics:physics, physics:quant-ph]. https://arxiv.org/abs/1910.10130 (accessed Dec 17, 2019).

(30) McClean, J. R. et al. OpenFermion: The Electronic Structure Package for Quantum Computers. 2017, arXiv:1710.07629 [physics, physics:quant-ph]. arXiv: 1710.07629. https://arxiv.org/abs/1710.07629 (accessed Dec 17, 2019).

(31) Parrish, R. M.; et al. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185–3197.

(32) Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; Wiley-Interscience: New York, NY, USA, 1987; DOI: 10.1002/9781118723203.

(33) A Jastrow-Type Decomposition in Quantum Chemistry for Low-Depth Quantum Circuits. [1909.12410v1]. https://arxiv.org/abs/ 1909.12410v1 (accessed Dec 17, 2019).

### Supporting Information: Is the Trotterized UCCSD ansatz chemically well-defined?

Harper R. Grimsley,<sup>1</sup> Daniel Claudino,<sup>1</sup> Sophia E. Economou,<sup>2</sup> Edwin Barnes,<sup>2</sup> and Nicholas J. Mayhall<sup>1,\*</sup>

<sup>1</sup>Department of Chemistry, Virginia Tech, Blacksburg, VA 24061, USA <sup>2</sup>Department of Physics, Virginia Tech, Blacksburg, VA 24061, USA (Dated: October 2019)

#### I. OPERATOR GROUPINGS

The single and double excitation operators  $\{a_a^{\dagger}a_i\}$  and  $\{a_a^{\dagger}a_b^{\dagger}a_ia_j\}$  (or  $\{a_p^{\dagger}a_q\}$  and  $\{a_p^{\dagger}a_q^{\dagger}a_ra_s\}$  in the case of generalized excitations) are labeled from looping over the last index, which is nested in a loop over the previous index, and so on, starting with singles and then indexing double excitations. For example, the first operator, with index 0, is  $\{a_{nacc}^{\dagger}+1a_0\}$ , where  $n_{occ}$  is the number of occupied orbitals.

Here we provide a brief description of the different ways to group operators in the Trotterized ansatz:

- 1. Fully Shuffled (FS) Operators All operators in the Trotter product are mixed at random. We sample 100 random orderings for each point.
- 2. Singles-Doubles (SD) Shuffled Operators The single and double excitations are mixed separately. The double excitations are applied to the reference state first. We sample 100 random orderings for each point.
- 3. Doubles-Singles (DS) Shuffled Operators The double and single excitations are mixed separately. The single excitations are applied to the reference state first. We sample 100 random orderings for each point.
- 4. Gradient Order (GO) The gradient of the reference energy with respect to each operator parameter is computed and used to sort the operators. Those with the largest gradients are applied first to the reference. This ordering is nearly deterministic, with the ordering of operators with equivalent gradients being chosen by lexical ordering.
- 5. Sequential Gradient Order (SGO) The gradient of each operator not yet applied is evaluated at the current ansatz, similarly to our ADAPT method.<sup>1</sup> The operator of largest gradient at each step is applied to the current ansatz, before reoptimizing all parameters. This ordering is also deterministic up to degeneracies in the gradient.
- 6. Unitary Coupled Cluster with Singles and Doubles (UCCSD) The un-Trotterized UCCSD energy is computed as a reference. This method is deterministic.
- 7. k products of Unitary pair Coupled Cluster with Generalized Singles and Doubles (k-UpCCGSD) The un-Trotterized k-UpCCGSD energy is computed as a reference. This method is deterministic.

We illustrate the effect of these distinct operator arrangements in comparison with UCCSD along with the data in Figure 1 in the main text in Figure 1.

Similar results follow from the ansatzes investigated in Figure 2 of the main text, which are portrayed in Figure 2. In general, Figures 1 and 2 suggest that placing double excitations closest to the reference gives lower energies than random operator arrangements, with tighter variance. The opposite effects are observed when the single excitations are all placed closest to the reference. This is fairly intuitive, given that double excitations tend to be more important than single excitations in methods like CCSD and CISD.

#### **II. DATA AND IMPLEMENTATION**

The data in the UCC plots were obtained using two in-house codes available at:

https://github.com/hrgrimsl/MG\_VQE\_Tools/tree/SI

and

https://github.com/hrgrimsl/spock

The SI\_File folder found there contains a variety of scripts which were used to write input files, and "seeding" scripts which can be used to manage calling and executing the "writing" scripts with different seeds for the random ordering of the operators. The workflow is outlined in the table below:



FIG. 1. Results for the different ansatzes derived from UCCSD for  $H_6$ , LiH, BeH<sub>2</sub> and N<sub>2</sub>.



FIG. 2. Results for the different ansatzes derived from 1-UpCCGSD (top) and 2-UpCCGSD (bottom) for  $H_6$ , where PEC refers to potential energy curve.

Seeding Script	Writing Script(s)	Purpose
h6inter.py	$Write_H6.py$	$FS H_6 Data$
h6intra.py	Write_H6B.py	$SD H_6 Data$
h6intra2.py	Write_H6C.py	$DS H_6 Data$
H6_misc.py	h6ucc.py	UCC $H_6$ Data
	h6sGO.py	$sGO H_6 Data$
	h6GO.py	GO $H_6$ Data
lihinter.py	Write_LiH.py	FS LiH Data
lihintra.py	Write_LiHB.py	SD LiH Data
lihintra2.py	Write_LiHC.py	DS LiH Data
LiH_misc.py	lihucc.py	UCC LiH Data
	lihsGO.py	sGO LiH Data
	lihGO.py	GO LiH Data
beh2inter.py	$Write_BeH2.py$	FS BeH2 Data
beh2intra.py	$Write_BeH2B.py$	SD BeH2 Data
beh2intra2.py	$Write\_BeH2C.py$	DS BeH2 Data
BeH2_misc.py	beh2ucc.py	UCC BeH2 Data
	beh2sGO.py	$sGO BeH_2 Data$
	beh2GO.py	$GO BeH_2 Data$
n2inter.py	Write_N2.py	$FS N_2 Data$
n2intra.py	Write_N2.py	$SD N_2 Data$
n2intra2.py	Write_N2.py	$DS N_2 Data$
N2_misc.py	n2ucc.py	UCC $N_2$ Data
	n2sGO.py	sGO $N_2$ Data
	n2GO.py	$GO N_2 Data$

Much of the data was assembled piecemeal, so the settings in a given template script may not replicate the whole
dissociation curve. Reselecting the radii in the writing scripts should enable exact replication of any arbitrary point on the curve, as the 0-99 random seeds are the correct ones. The code will require some additional minor work to function on a setup other than the cluster we used.

For the data used in Figure 2, specific operator orderings, along with the corresponding energy values, can be found in the .xlsx file also included in the Supporting Information.

<sup>\*</sup> nmayhall@vt.edu

<sup>&</sup>lt;sup>1</sup> Grimsley, H. R.; Economou, S. E.; Barnes, E.; Mayhall, N. J. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature Communications* **2019**, *10*, 3007.

## Chapter 3

## The ADAPT-VQE Method



#### ARTICLE

https://doi.org/10.1038/s41467-019-10988-2

OPEN

# An adaptive variational algorithm for exact molecular simulations on a quantum computer

Harper R. Grimsley<sup>1</sup>, Sophia E. Economou<sup>2</sup>, Edwin Barnes<sup>2</sup> & Nicholas J. Mayhall <sup>1</sup>

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 that, instead of fixing an ansatz upfront, grows it systematically one operator at a time in a way dictated by the molecule 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.

<sup>1</sup> Department of Chemistry, Virginia Tech, Blacksburg, VA 24061, USA. <sup>2</sup> Department of Physics, Virginia Tech, Blacksburg, VA 24061, USA. Correspondence and requests for materials should be addressed to N.J.M. (email: nmayhall@vt.edu)

A nticipation 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<sup>1–3</sup>. 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)<sup>4</sup> was the first algorithm proposed for simulating electronic structure problems on a quantum computer<sup>1,5</sup>. 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<sup>6–8</sup>.

In order to reduce the significant hardware demands required by PEA and exploit the capabilities of noisy intermediate-scale quantum (NISQ) devices<sup>9</sup>, the variational quantum eigensolver (VQE) algorithm was proposed and demonstrated using photonic qubits by Peruzzo et al.<sup>10</sup>. This was followed by several theoretical studies on VQE<sup>7,11–17</sup> and demonstrations on other hardware such as superconducting qubits<sup>7,16,18</sup> and trapped ions<sup>19,20</sup>. Other approaches have been pursued as well, including methods for adiabatic quantum computation<sup>21</sup> and quantum machine learning<sup>22</sup>.

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,  $\hat{H} = \sum_i g_i \hat{o}_i$ . Because the individual operator terms,  $\hat{o}_i$ , 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.<sup>12</sup>.

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<sup>7,12,16</sup>, 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 accuracy with few parameters and shallow circuits. The first ansatz explored<sup>10</sup> 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<sup>23–26</sup>. 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<sup>27</sup> and the k-UpCCGSD approach of Lee et al.<sup>17</sup> which uses k products of unitary paired generalized doubles excitations, along with the full set of generalized single excitations. The k-UpCCGSD approach builds on early work by Nakatsuji<sup>28-31</sup> and Nooijen<sup>32</sup> 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, which provides a systematic way to converge to FCI without introducing higher excitation rank operators. Ryabinkin et al.<sup>33</sup> 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. <sup>16</sup> developed an approach (termed the quantum subspace expansion) to extract not just the expectation value of  $\hat{H}$  but all the matrix elements  $\langle I | \hat{H} | I \rangle$  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. Even further from the original UCC ansatz, Kandala et al.<sup>18</sup> have used an alternative ansatz for their VQE experiments based on the native entangling gate in their superconducting qubit device, referred to as a "hardware-efficient ansatz". This allows entanglement to be created directly from a device-wide unitary instead of through a more traditional gate decomposition of a fermionic operator.

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 optimization procedure for Full Configuration Interaction (FCI) VQE and is more thoroughly discussed in Section 1 of the Supplement. We demonstrate the power of ADAPT-VQE through numerical simulations of three molecules of increasing complexity: LiH, BeH<sub>2</sub>, and H<sub>6</sub>. 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.

#### Results

**Specification of the adopted notation**. 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:

$$\langle \psi^{\text{CCSD}} \rangle = e^{\hat{T}_1 + \hat{T}_2} | \psi^{\text{HF}} \rangle,$$
 (1)

where the excitation operators are defined as:

$$\hat{T}_1 = \sum_{ia} \hat{t}_i^a = \sum_{ia} t_i^a \hat{a}_a^\dagger \hat{a}_i$$
(2)

$$\hat{T}_{2} = \sum_{i < j, a, b} \hat{t}_{ij}^{ab} = \sum_{i < j, a < b} t_{ij}^{ab} \hat{a}_{a}^{\dagger} \hat{a}_{b}^{\dagger} \hat{a}_{i} \hat{a}_{j}.$$
(3)

For closed shell molecules near equilibrium, CCSD provides a robust ansatz for molecular simulations. Early efforts to combine size extensivity and variationality were pioneered by Bartlett, Kutzelnigg, and coworkers<sup>23–25</sup>. In this context, a unitary variant of coupled cluster theory (UCCSD) was defined by replacing the excitation operators with an anti-Hermitian sum of excitation and de-excitation operators:

$$\hat{t}_{ij}^{ab} \to \hat{t}_{ij}^{ab} - \hat{t}_{ab}^{ij} = \hat{\tau}_{ij}^{ab}.$$
(4)

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<sup>10</sup>.

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<sup>17</sup>. In this case the cluster operators are further generalized to remove the HF-based subspace restriction:  $\hat{\tau}_{ij}^{ab} \rightarrow \hat{\tau}_{pq}^{rs}$ , 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<sup>34</sup>,

$$e^{A+B} = \lim_{n \to \infty} \left( e^{(A/n)} e^{(B/n)} \right)^n.$$
 (5)

Because the generalized single and double excitation operators do not commute, the use of a truncated Trotter expansion represents an approximation to the underlying UCCSD ansatz, and recent work has shown clearly that this does not strongly affect the results because the variational flexibility is sufficient to absorb this error<sup>7</sup>, 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:

$$|\psi^{\text{tUCC}}\rangle = \prod_{s \in \{pq\}} e^{\hat{t}_s} \prod_{d \in \{pqrs\}} e^{\hat{t}_d} |\psi^{\text{HF}}\rangle, \tag{6}$$

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.

**ADAPT-VQE algorithm**. The above discussion described the Trotter expansion as an approximation to UCCGSD. However, as recognized previously<sup>7,35</sup>, 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. The exact (FCI) quantum state can thus be represented as an arbitrarily long product of one- and two-body operators,

$$\Psi^{\text{FCI}}\rangle = \prod_{k}^{\infty} \prod_{pq} e^{\hat{r}_{p}^{q}(k)} \prod_{pqrs} e^{\hat{r}_{pq}^{rs}(k)} |\Psi^{\text{HF}}\rangle \tag{7}$$

where  $\hat{\tau}_{pq}^{rs}(k)$  is the *k*th instance, or "replica", of the operators in  $\hat{t}_{pq}^{rs} - \hat{t}_{rs}^{pq}$ . 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.,  $\tau_{pq}^{rs}(k) \neq \tau_{pq}^{rs}(j)$ .

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:

- 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-complemented 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. If multiple quantum computers are available, perform this step on all devices simultaneously.
- (5) Measure the commutator of the Hamiltonian with each operator in the pool to get the gradient. Repeating this multiple times and averaging gives the gradient of the expectation value of the Hamiltonian with respect to the coefficient of each operator. This can be done in parallel.
- (6) If the norm of the gradient vector is smaller than some threshold,  $\varepsilon$ , exit.
- (7) Identify the operator with the largest gradient and add this single operator to the left end of the ansatz, with a new variational parameter. Note that this does not "drain" the pool in the sense that choosing an operator does not remove it from the pool so it can be used again later.
- (8) Perform a VQE experiment to re-optimize all parameters in the ansatz.
- (9) Go to step 4.



**Fig. 1** Schematic depiction of the ADAPT-VQE algorithm described presented. Since step 1 occurs on classical hardware, it is not included in the illustration.  $\vec{\theta}^{(n)}$  is the list of ansatz parameters at the *n*th iteration. The number of parameters,  $len(\vec{\theta}^{(n)})$ , is equal to the number of operators in the ansatz. "Operator Pool" refers to the collection of operators which are used to grow the ansatz one-at-a-time. Each  $\tau_p^a$  represents a generalized single or double excitation, and these operators are then spin-complemented. The orbital indices refer to spatial orbitals, and the overbar indicates  $\beta$  spin. Orbital indices without overbars have  $\alpha$  spin. Note that growing the ansatz does not drain the pool, and so operators can show up multiple times if selected by the algorithm

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 gradient expression in step 5 is derived in section IB of the Supplemental Information). 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 6. At convergence, the ADAPT-VQE algorithm obtains the following ansatz:

$$\left|\psi^{\text{ADAPT}(\varepsilon)}\right\rangle = \left(e^{\hat{\tau}_{N}}\right)\left(e^{\hat{\tau}_{N-1}}\right)\cdots\left(e^{\hat{\tau}_{2}}\right)\left(e^{\hat{\tau}_{1}}\right)\left|\psi^{\text{HF}}\right\rangle \tag{8}$$

where the identity of each  $\hat{\tau}_i$  is determined by the algorithm.

The re-optimization subroutine in step 8 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<sup>12,15,16,18</sup>. Note that this subroutine is distinct from the gradient computed in step 5 of the algorithm. Additional possible modifications to the algorithm are mentioned in the "Discussion" section.

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 reoptimizations 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.

**Molecular dissociation simulation results.** In this section, we explore the convergence properties of the ADAPT-VQE algorithm with a few small molecular systems, LiH, BeH<sub>2</sub>, and linear H<sub>6</sub>. The former two molecules have been simulated using quantum hardware<sup>18,20</sup>. H<sub>6</sub> 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<sup>36,37</sup> for the integral calculation (via the OpenFermion-Psi4<sup>38</sup> interface) and OpenFermion was used for the Jordan–Wigner operator transformation. All calculations used the Broyden–Fletcher–Goldfarb–Shannon (BFGS) minimization implemented within Scipy<sup>39</sup>. To classically simulate the re-optimization subroutine in step 8 of ADAPT-VQE, we could use a standard numerical gradient method. However, in order to improve the efficiency and allow precise gradients for tight convergence, we derived and implemented an efficient analytic

gradient function, which is detailed in Section D of the Supplement. By reusing intermediates between individual parameter gradients, this algorithm obtains the full gradient vector for all parameters at a cost which is only roughly  $2\times$  that of the base energy evaluation. A table with timing data is included in Table 1 of the Supplement.

As discussed in the previous section, the ADAPT ansatz uses a convergence threshold to determine when the calculation should terminate (step 6). Here we use the norm of the gradient vector and compare it to threshold  $\varepsilon_m$ , which we define as

$$\varepsilon_m = 10^{-m}.\tag{9}$$

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( $\varepsilon_3$ ). In what follows we present numerical results for bond-dissociation curves for LiH, BeH<sub>2</sub>, and H<sub>6</sub> 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.

Here, we study the LiH bond dissociation computed using several methods, including FCI, UCCSD (un-Trotterized), HF, ADAPT( $\varepsilon_1$ ), ADAPT( $\varepsilon_2$ ), and ADAPT( $\varepsilon_3$ ), 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  $\binom{6}{1} \cdot \binom{6}{1} = 225$ . In this basis, the

has a dimension of  $\binom{6}{2} \cdot \binom{6}{2} = 225$ . 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. 2a, where all the curves, with the exception of HF, cannot really be distinguished on this scale. However, as shown in Fig. 2b, 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( $\varepsilon_1$ ) is not as accurate as UCCSD, ADAPT( $\varepsilon_2$ ) is comparable to UCCSD at short bond distances and comfortably outperforms it at longer distances. This is also evident in Table 1, where the average error across the potential energy surface (PES) is shown. Remarkably, ADAPT( $\varepsilon_3$ ) outperforms UCCSD throughout the whole curve by at least an order of magnitude and in some cases up to four orders of magnitude.

Even more impressive is how few parameters are needed to achieve this level of accuracy. As shown in Fig. 2c, 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 combining spin-complements. In all three ADAPT calculations, fewer than half of the parameters are needed compared to UCCSD. Although UCCSD is noticeably more accurate than the simplest ADAPT calculation with a gradient norm threshold of 0.1, the ADAPT( $\varepsilon_1$ ) ansatz is incredibly compact, consisting of fewer than 10 parameters across the curve. For example, the ADAPT( $\varepsilon_1$ ) ansatz for LiH at bond distance 2.39 Å is

$$\left|\psi^{\text{ADAPT}(\varepsilon_{1})}\right\rangle = e^{\hat{\tau}_{12}^{16}} e^{\hat{\tau}_{22}^{55}} e^{\hat{\tau}_{22}^{44}} e^{\hat{\tau}_{12}^{13}} e^{\hat{\tau}_{11}^{33}} e^{\hat{\tau}_{22}^{33}} e^{\hat{\tau}_{22}^{36}} e^{\hat{\tau}_{22}^{66}} \left|\psi^{\text{HF}}\right\rangle, \tag{10}$$

which includes a mixture of both double excitations and correlated single excitations  $(\hat{n}_{j}\hat{a}_{a}^{\dagger}\hat{a}_{i})$ . The indices denote spatial orbitals, overbar on an index denotes  $\beta$  spin, and spin-complemented interactions are implied. For example  $\hat{\tau}_{01}^{06}$  is really  $\hat{\tau}_{01}^{06} + \hat{\tau}_{\overline{01}}^{\overline{06}}$ . An interesting feature of the ansatz returned by

ADAPT-VQE, Eq. (10), is that the HOMO–LUMO double excitation  $\left(e^{\hat{\tau}_{22}^{33}}\right)$  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,  $e^{\hat{\tau}_{22}^{36}}$ .

In Fig. 2d–f, the dissociation curves for BeH<sub>2</sub> are shown. In the STO-3G basis, BeH<sub>2</sub> has 7 spatial orbitals, for a total Hilbert space dimension of 16,384, and a neutral molecule subspace of dimension  $\begin{pmatrix} 7 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 7 \\ 3 \end{pmatrix} = 1225$ . Unlike with LiH, UCCSD does not provide chemically accurate results across the full PES. UCCSD and ADAPT( $\varepsilon_1$ ) are comparable at smaller bond distances. Beyond ~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( $\varepsilon_2$ ) and ADAPT( $\varepsilon_3$ )

listed in Table 1. Now we move our focus to the  $H_6$  data. 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  $H_{62}$ , with the results presented in Fig. 2g-i.

provide nearly exact results, with average deviations from FCI

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. 2h, and (2) the increased number of parameters selected in the ADAPT calculations in Fig. 2i. Despite being strongly correlated, such that higher excitation rank operators should be needed, both ADAPT( $\varepsilon_2$ ) and ADAPT( $\varepsilon_3$ ) provide accurate results with only one- and two-body operators. Moreover, in the case of ADAPT( $\varepsilon_2$ ) this is achieved with fewer operators than UCCSD for most bond distances. ADAPT( $\varepsilon_3$ ) 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. 2h, first at R(H-H) = 1.8 Å where ADAPT( $\varepsilon_1$ ) experiences a large jump in energy, and second at 2.5 Å where ADAPT( $\varepsilon_2$ ) increases in energy. Figure 2i shows that these energy jumps correspond to sudden drops in parameter counts.

The cause of the discontinuities in the  $H_6$  data can be explained from the convergence data provided in the Supplement (see Supplement Fig. 1). For larger bond lengths, as additional operators are added to the ansatz, the energy flattens out before dropping substantially again. If the convergence criterion is too lenient, then the ADAPT-VQE optimization will abort at such "false gradient troughs". In the ADAPT( $\varepsilon_2$ ) data of Fig. 2h, i, the jump in energy error and drop in parameter number, respectively, are caused by the 2.5 Å optimization aborting at a false gradient trough, while the optimizations at other bond lengths do not. Of course, if a tighter threshold is used (such as 0.001), the ADAPT-VQE algorithm does not prematurely abort, and ultimately yields high-accuracy results, even for this strongly correlated system. More sophisticated convergence checks in step 6 might avoid these situations and will be one focus of future work.

Dependence of convergence on operator ordering. To demonstrate the importance of the gradient-based operator



**Fig. 2** Dissociation curves for LiH, BeH<sub>2</sub>, and H<sub>6</sub>. Potential energy as a function of nuclear coordinate, Hartree units (**a**, **d**, **g**). Absolute energy differences from FCI, kcal/mol units (**b**, **e**, **h**). Shaded blue region represents area within "chemical accuracy" as 1 kcal/mol. Number of variational operators in associated ansatz (**c**, **f**, **i**). Notation:  $\epsilon$  indicates gradient norm threshold used such that  $\epsilon_m = 10^{-m}$ . In all curves, the FCI curve lies directly underneath the ADAPT( $\epsilon_3$ ) curve, and so is not visible

Table 1 Average errors across the PES scan for the different methods assessed. Units in kcal/mol

	UCCSD	ADAPT(E1)	<b>Α<b>ΔΑΡΤ</b>(ε<sub>2</sub>)</b>	ADAPT(E3)
LiH	0.0480	0.3000	0.0058	0.0002
$BeH_2$	2.2384	0.8023	0.0907	0.0041
H <sub>6</sub>	3.7387	4.5297	0.3023	0.0047



Fig. 3 Convergence of the energy as a function of parameter count for  $BeH_2$  with a Be-H distance of 2.39 Å. The black dot indicates the UCCSD point

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 cases considered. While the two random-growth ansätze converge relatively similar to each other regardless of whether restricted indices are used or not, the lexically ordered ansatz shows a clear distinction between the restricted index (singles and doubles) and un-restricted index (generalized singles and doubles) ordering. This is due to the fact that the first operators in the ansatz involve creation operators on the occupied orbitals, and these do not contribute until the wavefunction has become entangled. The un-Trotterized UCCSD result is also marked for reference. Overall, the data in Fig. 3 demonstrate that an iterative gradient minimization algorithm yields a highly compact ansatz for a given state.

#### Discussion

An obvious metric for evaluating the performance of any simulation algorithm can be simply described as some accuracy measure vs. some cost measure. While the accuracy measure in a simulation is often easy to define, the cost measure is more nuanced. For variational quantum simulations, there are two factors which largely determine the overall cost: circuit depth and number of measurements (or shot count). Shot count is important as it determines the time to solution. It is possible that due to the sheer number of measurements, a particular quantum simulation becomes intractable. However, for NISQ devices in which coherence times (and thus number of gates) are limited, circuit depth is usually the most critical cost metric, as it determines whether or not a simulation can occur at all. By taking circuit depth as the most important cost metric to address, the original VQE has been successful by minimizing circuit depth at the cost of increased number of measurements. Similar to the original VQE, our new ADAPT-VQE algorithm seeks to further minimize the circuit depth with an increased number of measurements.

In this direction, the data clearly demonstrates that ADAPT-VQE succeeds in creating a more compact and accurate wavefunction ansatz than 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.

In terms of shot-count, ADAPT-VQE will likely have an increased number of measurements compared to UCCSD-based VQE due to the necessary gradient measurements. However, this is perhaps an easier problem to address (compared to circuit depth) as the individual runs can in principle occur simultaneously if several devices exist. Further, the shot count also depends on the number of iterations required for the classical optimization of the ansatz parameters. For strongly correlated systems where perturbation theory fails, the existing approach of using classical MP2 amplitudes to initialize the UCCSD parameters<sup>12</sup> is not likely to provide much improvement in the UCCSD-based VQE. Alternatively, each iteration of ADAPT-VQE only adds a single new parameter, with the previously optimized parameters already being initialized to rather sensible values. This might ultimately decrease the number of iterations needed for the VQE subroutine in ADAPT-VQE, thus decreasing the shot count (although this is not likely to fully compensate for the large number of measurements for the gradient). 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<sup>40-46</sup> and single-electron wavefunctions<sup>47-53</sup>, or as a relative of methods using sequential transformations which have been explored in the context of multireference coupled-cluster theory<sup>54,55</sup>. Of these, the ADAPT ansatz is perhaps most closely related to the @-CC method of Lyakh and Bartlett<sup>47</sup>, 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, see ref. 47) 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 the "Results" section, 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 twobody 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. Because only local knowledge of the FCI energy landscape is used to update the ADAPT-VQE ansatz construction, the "true optimally compact ansatz" is not guaranteed. As a result, flat energy landscapes (associated with "false gradient troughs") are possible. Further classical simulations and device implementations are needed to provide better insight into the numerical behavior.

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.

In this paper, we presented ADAPT-VQE, a novel variational hybrid quantum-classical algorithm designed to achieve exact results at convergence. Unlike typical ansätze, which tends 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 routines for compiling state preparation circuits and quantum-parallelism should make ADAPT-VQE extremely useful for simulations of molecules on both currently available and future quantum computers.

#### Data availability

The data for the numerical simulations is available upon reasonable request.

#### Code availability

The code for the numerical simulations is available upon reasonable request.

Received: 4 January 2019 Accepted: 6 June 2019 Published online: 08 July 2019

#### References

- Aspuru-Guzik, A., Dutoi, A. D., Love, P. J. & Head-Gordon, M. Simulated quantum computation of molecular energies. *Science* 309, 1704–1707 (2005).
- McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S. & Yuan, X. Quantum computational chemistry. http://arxiv.org/abs/1808.10402 (2018).
- Cao, Y. et al. Quantum chemistry in the age of quantum computing. http:// arxiv.org/abs/1812.09976 (2018).
- 4. Kitaev, A. Y. Quantum measurements and the Abelian stabilizer problem. *arXiv e-prints*. https://arxiv.org/abs/quant-ph/9511026 (1995).
- 5. Lloyd, S. Universal quantum simulators. Science 273, 1073-1078 (1996).
- Lanyon, B. P. et al. Towards quantum chemistry on a quantum computer. *Nat. Chem.* 2, 106–111 (2010).
- O'Malley, P. et al. Scalable quantum simulation of molecular energies. *Phys. Rev. X* 6, 031007 (2016).
- Paesani, S. et al. Experimental Bayesian quantum phase estimation on a silicon photonic chip. *Phys. Rev. Lett.* 118, 100503 (2017).
- 9. Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* 2, 79 (2018).
- 10. Peruzzo, A. et al. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* 5, 4213 (2014).
- Whitfield, J. D., Biamonte, J. & Aspuru-Guzik, A. Simulation of electronic structure Hamiltonians using quantum computers. *Mol. Phys.* 109, 735–750 (2011).
- McClean, J. R., Romero, J., Babbush, R. & Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* 18, 023023 (2016).
- McClean, J. R., Kimchi-Schwartz, M. E., Carter, J. & de Jong, W. A. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A* 95, 042308 (2017).
- 14. Barkoutsos, P. K. et al. Quantum algorithms for electronic structure calculations: particle/hole Hamiltonian and optimized wavefunction expansions. *Phys. Rev. A* **98**, 022322 (2018).
- Romero, J. et al. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Sci. Technol.* 4, 014008 (2018).
- 16. Colless, J. I. et al. Computation of molecular spectra on a quantum processor with an error-resilient algorithm. *Phys. Rev. X* 8, 011021 (2018).
- Lee, J., Huggins, W. J., Head-Gordon, M. & Whaley, K. B. Generalized unitary coupled cluster wavefunctions for quantum computation. *J. Chem. Theory Comput.* 15, 311–324 (2019).
- Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* 549, 242–246 (2017).
- Shen, Y. et al. Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure. *Phys. Rev. A* 95, 020501 (2017).
- 20. Hempel, C. et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Phys. Rev. X* **8**, 031022 (2018).
- Xia, R., Bian, T. & Kais, S. Electronic structure calculations and the Ising Hamiltonian. J. Phys. Chem. B 122, 3384–3395 (2017).
- 22. Xia, R. & Kais, S. Quantum machine learning for electronic structure calculations. *Nat. Commun.* 9, 4195 (2018).
- Bartlett, R. J., Kucharski, S. A. & Noga, J. Alternative coupled-cluster ansätze II. The unitary coupled-cluster method. *Chem. Phys. Lett.* 155, 133–140 (1989).
- 24. Kutzelnigg, W. Error analysis and improvements of coupled-cluster theory. *Theor. Chim. Acta* **80**, 349–386 (1991).
- Taube, A. G. & Bartlett, R. J. New perspectives on unitary coupled-cluster theory. Int. J. Quantum Chem. 106, 3393–3401 (2006).
- Harsha, G., Shiozaki, T. & Scuseria, G. E. On the difference between variational and unitary coupled cluster theories. J. Chem. Phys. 148, 044107 (2018).
- Dallaire-Demers, P.-L., Romero, J., Veis, L., Sim, S. & Aspuru-Guzik, A. Lowdepth circuit ansatz for preparing correlated fermionic states on a quantum computer. http://arxiv.org/abs/1801.01053 (2018).
- Nakatsuji, H. & Ehara, M. Iterative CI general singles and doubles (ICIGSD) method for calculating the exact wave functions of the ground and excited states of molecules. *J. Chem. Phys.* **122**, 194108 (2005).
- 29. Nakatsuji, H. Structure of the exact wave function. J. Chem. Phys. 113, 2949 (2000).
- Nakatsuji, H. & Ehara, M. Structure of the exact wave function. V. Iterative configuration interaction method for molecular systems within finite basis. J. Chem. Phys. 117, 9–12 (2002).
- Nakatsuji, H. & Davidson, E. R. Structure of the exact wave function. II. Iterative configuration interaction method. J. Chem. Phys. 115, 2000–2006 (2001).
- Nooijen, M. Can the eigenstates of a many-body Hamiltonian be represented exactly using a general two-body cluster expansion? *Phys. Rev. Lett.* 84, 2108–2111 (2000).
- Ryabinkin, I. G., Yen, T.-C., Genin, S. N. & Izmaylov, A. F. Qubit coupledcluster method: a systematic approach to quantum chemistry on a quantum computer. J. Chem. Theory Comput. 14, 6317–6326 (2018).

- Hatano, N. & Suzuki, M. Finding Exponential Product Formulas of Higher Orders. 37–68 (Springer, Berlin, Heidelberg, 2005).
- Barkoutsos, P. K. et al. Quantum algorithms for electronic structure calculations: particle-hole Hamiltonian and optimized wave-function expansions. *Phys. Rev. A* 98, 022322 (2018).
- Turney, J. M. et al. Psi4: an open-source ab initio electronic structure program. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2, 556–565 (2012).
- Smith, D. G. A. et al. Psi4NumPy: an interactive quantum chemistry programming environment for reference implementations and rapid development. J. Chem. Theory Comput. 14, 3504–3511 (2018).
- McClean, J. R. et al. OpenFermion: the electronic structure package for quantum computers. http://arxiv.org/abs/1710.07629 (2017).
- Jones, E., Travis, O. & Peterson, P. SciPy: open source scientific tools for Python. http://www.scipy.org/ (2001).
- Harrison, R. J. Approximating full configuration interaction with selected configuration interaction and perturbation theory. *J. Chem. Phys.* 94, 5021–5031 (1991).
- Povill, A., Rubio, J. & Illas, F. Treating large intermediate spaces in the CIPSI method through a direct selected CI algorithm. *Theor. Chim. Acta* 82, 229–238 (1992).
- 42. Peris, G., Planelles, J., Malrieu, J.-P. & Paldus, J. Perturbatively selected CI as an optimal source for externally corrected CCSD. *J. Chem. Phys.* **110**, 11708 (1999).
- Evangelista, F. A. Adaptive multiconfigurational wave functions. J. Chem. Phys. 140, 124114 (2014).
- Schriber, J. B. & Evangelista, F. A. Communication: an adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy. J. Chem. Phys. 144, 161106 (2016).
- Holmes, A. A., Tubman, N. M. & Umrigar, C. J. Heat-bath configuration interaction: an efficient selected configuration interaction algorithm inspired by heat-bath sampling. J. Chem. Theory Comput. 12, 3674–3680 (2016).
- Xu, E., Uejima, M. & Ten-No, S. L. Full coupled-cluster reduction for accurate description of strong electron correlation. *Phys. Rev. Lett.* **121**, 113001 (2018).
- Lyakh, D. I. & Bartlett, R. J. An adaptive coupled-cluster theory: @CC approach. J. Chem. Phys. 133, 244112 (2010).
- Bischoff, F. A. & Valeev, E. F. Computing molecular correlation energies with guaranteed precision. J. Chem. Phys. 139, 114106 (2013).
- Laikov, D. N. Intrinsic minimal atomic basis representation of molecular electronic wavefunctions. *Int. J. Quantum Chem.* 111, 2851–2867 (2011).
- Lu, W. C. et al. Molecule intrinsic minimal basis sets. I. Exact resolution of ab initio optimized molecular orbitals in terms of deformed atomic minimalbasis orbitals. J. Chem. Phys. 120, 2629–2637 (2004).
- Berghold, G., Parrinello, M. & Hutter, J. Polarized atomic orbitals for linear scaling methods. J. Chem. Phys. 116, 1800–1810 (2002).
- Lee, M. S. & Head-Gordon, M. Polarized atomic orbitals for self-consistent field electronic structure calculations. J. Chem. Phys. 107, 9085 (1998).
- Schütt, O. & VandeVondele, J. Machine learning adaptive basis sets for efficient large scale density functional theory simulation. J. Chem. Theory Comput. 14, 4168–4175 (2018).

- Evangelista, F. A., Hanauer, M., Köhn, A. & Gauss, J. A sequential transformation approach to the internally contracted multireference coupled cluster method. *J. Chem. Phys.* 136, 204108 (2012).
- Evangelista, F. A. Alternative single-reference coupled cluster approaches for multireference problems: the simpler, the better. *J. Chem. Phys.* 134, 224102 (2011).

#### 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 the Department of Energy (Award No. DE-SC0019318).

#### **Author contributions**

N.J.M., S.E.E., and E.B. conceived the project. H.R.G. wrote the code. H.R.G. and N.J.M. performed the calculations. N.J.M., S.E.E., and E.B. wrote the paper.

#### **Additional information**

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-10988-2.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/ reprintsandpermissions/

Peer review information: Nature Communications thanks Peter Haynes and other anonymous reviewers for their contribution to the peer review of this work.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/ licenses/by/4.0/.

© The Author(s) 2019

#### Supplementary Information for: An adaptive variational algorithm for exact molecular simulations on a quantum computer Grimsley et al.

1

#### I. SUPPLEMENTARY DISCUSSION

#### A. Derivation of ADAPT-VQE 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 full CI (FCI) VQE (using the ansatz definition in Eq. (7) of the main text), but with a sparsity constraint on the gradientbased 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:

$$\left|\psi^{\text{FCI}}\right\rangle = \prod_{k}^{\infty} \prod_{n}^{N^{4}} e^{\theta_{n}^{k} \hat{A}_{n}} \left|\psi^{\text{HF}}\right\rangle,\tag{1}$$

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

$$\frac{\partial}{\partial \theta_m^k} \left\langle \psi^{\mathrm{HF}} \right| e^{-\theta_1^1 \hat{A}_1} e^{-\theta_2^1 \hat{A}_2} \cdots \hat{H} \cdots e^{\theta_2^1 \hat{A}_2} e^{\theta_1^1 \hat{A}_1} \left| \psi^{\mathrm{HF}} \right\rangle = 0.$$
<sup>(2)</sup>

#### B. 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:

$$\left|\psi^{(0)}\right\rangle = \left|\psi^{\rm HF}\right\rangle.\tag{3}$$

A gradient descent step from the HF state could then provide an improved set of parameters, lowering the expected energy. The gradient at the 0th iteration is quite simple due to the fact that all the  $\theta_m^k$  vanish:

$$\frac{\partial E^{(0)}}{\partial \theta_m^k} = \frac{\partial}{\partial \theta_m^k} \left\langle \psi^{\rm HF} \right| e^{-\theta_1^1 \hat{A}_1} e^{-\theta_2^1 \hat{A}_2} \cdots \hat{H} \cdots e^{\theta_2^1 \hat{A}_2} e^{\theta_1^1 \hat{A}_1} \left| \psi^{\rm HF} \right\rangle \Big|_{\theta_m^k = 0} = \left\langle \psi^{\rm HF} \right| [H, \hat{A}_m] \left| \psi^{\rm HF} \right\rangle. \tag{4}$$

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, because each of the terms that were initially set to zero become non-zero, meaning that each term now has to appear in the circuit. The reason is that all k replicas have identical gradients:

$$\frac{\partial E^{(0)}}{\partial \theta_m^k} = \frac{\partial E^{(0)}}{\partial \theta_m^j} \text{for all } k, j.$$
(5)

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

<sup>\*</sup>Electronic address: nmayhall@vt.edu

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

Supplementary Table I: Timing comparison between finite difference and analytical gradients for several VQE iterations. System being studied is  $H_2O$  with STO-3G basis set. Ansatz has 140 parameters.

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:

$$\psi^{(1)}\rangle = e^{\theta_p \hat{A}_p} \left|\psi^{\rm HF}\rangle.$$
(6)

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:

$$E^{(1)} = \min_{\theta_p^1} \left\langle \psi^{\mathrm{HF}} \right| e^{-\theta_p^1 \hat{A}_p} \hat{H} e^{\theta_p^1 \hat{A}_p} \left| \psi^{\mathrm{HF}} \right\rangle.$$
(7)

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

#### C. 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:

$$\frac{\partial E^{(1)}}{\partial \theta_m^k} = \left\langle \psi^{(1)} \big| [H, \hat{A}_m] \big| \psi^{(1)} \right\rangle.$$
(8)

However, because  $\theta_p^1$  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:

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

$$\tag{9}$$

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.

#### D. 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 the main text. 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  $|\psi_0\rangle$ :

$$|\phi\rangle = \prod_{i=1}^{N} e^{\theta_i T_i} |\psi_0\rangle, \qquad (10)$$

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.,  $\prod_{i=1}^{3} \hat{o}_i = \hat{o}_3 \hat{o}_2 \hat{o}_1$ . Assuming  $|\psi_0\rangle$  is normalized, the functional to be minimized is then

$$E_{\phi} = \langle \phi | \hat{H} | \phi \rangle \,. \tag{11}$$

The state  $|\phi\rangle$  can be obtained without direct matrix exponentiation, [1, 2] and in this work we use the Scipy expm\_multiply function to compute the action of a matrix exponential on a [3] vector, which is relatively efficient since the operator in each exponent in Supp. Eq. 10 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 [4] then follow directly as

$$\frac{\partial E_{\phi}}{\partial \theta_{i}} = \langle \phi | \hat{H} \prod_{j=i+1}^{N} \left( e^{\theta_{j} T_{j}} \right) T_{i} \prod_{k=1}^{i} \left( e^{\theta_{k} T_{k}} \right) | \psi_{0} \rangle$$
$$- \langle \psi_{0} | \prod_{k=i}^{1} \left( e^{-\theta_{k} T_{k}} \right) T_{i} \prod_{j=N}^{i+1} \left( e^{-\theta_{j} T_{j}} \right) \hat{H} | \phi \rangle , \qquad (12)$$

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.,  $\prod_{j=3}^{1} \hat{o}_j = \hat{o}_1 \hat{o}_2 \hat{o}_3$ . Recognizing that the second line is just the adjoint of the first, Supp. Eq. 12 can be simplified to

$$\frac{\partial E_{\phi}}{\partial \theta_{i}} = 2\Re \left( \left\langle \phi \right| \hat{H} \prod_{j=i+1}^{N} \left( e^{\theta_{j} T_{j}} \right) T_{i} \prod_{k=1}^{i} \left( e^{\theta_{k} T_{k}} \right) \left| \psi_{0} \right\rangle \right).$$
(13)

In order to obtain a more compact expression for the *i*th parameter derivative, we partially expand the wavefunction expression in Supp. Eq. 10, grouping the unitaries up to and after *i* to define the notation for a state partially rotated with unitaries from 1 to *i* as  $|\psi_{1,i}\rangle$ :

$$|\phi\rangle = \left[\prod_{j=i+1}^{N} e^{\theta_j T_j}\right] \left[\prod_{k=1}^{i} e^{\theta_k T_k}\right] |\psi_0\rangle \tag{14}$$

$$= \left[\prod_{j=i+1}^{N} e^{\theta_j T_j}\right] |\psi_{1,i}\rangle \tag{15}$$

$$= |\psi_{1,N}\rangle. \tag{16}$$

Using this notation and defining  $|\sigma_i\rangle = \prod_{j=N}^{i+1} \left(e^{-\theta_j T_j}\right) \hat{H} |\phi\rangle$  where  $|\sigma_N\rangle = \hat{H} |\phi\rangle$ , Supp. Eq. 13 becomes

$$\frac{\partial E_{\phi}}{\partial \theta_{i}} = 2\Re \left\langle \sigma_{i} \right| T_{i} \left| \psi_{1,i} \right\rangle.$$
(17)

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 the left-most operator, i = N, where we need to calculate

$$\frac{\partial E_{\phi}}{\partial \theta_N} = 2\Re \left\langle \sigma_N \right| T_N \left| \psi_{1,N} \right\rangle. \tag{18}$$

At the second step, we need to compute

$$\frac{\partial E_{\phi}}{\partial \theta_{N-1}} = 2\Re \left\langle \sigma_{N-1} | T_{N-1} | \psi_{1,N-1} \right\rangle, \tag{19}$$

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:

$$\left|\sigma_{N-1}\right\rangle = e^{-\theta_N T_N} \left|\sigma_N\right\rangle,\tag{20}$$

$$|\psi_{1,N-1}\rangle = e^{-\theta_N T_N} |\psi_{1,N}\rangle. \tag{21}$$

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:

$$|\sigma_i\rangle = e^{-\theta_{i+1}T_{i+1}} |\sigma_{i+1}\rangle, \qquad (22)$$

$$|\psi_{1,i}\rangle = e^{-\theta_{i+1}T_{i+1}} |\psi_{1,i+1}\rangle.$$
(23)

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. For a comparison, Table I lists the times required for a few VQE iterations of H<sub>2</sub>O with the STO-3G basis set using both numerical and analytical gradients.

#### E. 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 common shortcut in classical electronic structure theory which allows one to cancel systematic errors 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<sub>2</sub>, and H<sub>6</sub> in Fig. 1.

While LiH and BeH<sub>2</sub> both exhibit rapid convergence, the strongly correlated H<sub>6</sub> 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<sub>6</sub> is intrinsically 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 operators that might be needed to make the ansatz sufficiently accurate. In Fig. 1(c), the difference between R = 2.4 and 2.5 Å involves a significant reduction in the number of parameters for ADAPT( $\epsilon_2$ ), 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.



Supplementary Figure 1: Convergence of the gradient max (top), gradient norm (middle), and energy error (bottom) at different bond lengths for (a) LiH, (b) BeH<sub>2</sub>, and (c) H<sub>6</sub>. Each line corresponds to a distinct bond distance, ranging from light to dark as the distance increases. For both LiH ( $R_{eq} = 2.39$  Å) and BeH<sub>2</sub> ( $R_{eq} = 1.342$  Å), the bond distance ranges from 0.5  $R_{eq}$  - 2.5  $R_{eq}$ . For H<sub>6</sub>, the bond distance ranges from 0.5 Å- 2.5 Å.

- Higham, N. J. & Al-Mohy, A. H. Computing matrix functions. <u>Acta Numerica</u> 19, 159-208 (2010). URL http://www.journals.cambridge.org/abstract{\\_}S0962492910000036.
- [2] Al-Mohy, A. H. & Higham, N. J. Computing the Action of the Matrix Exponential, with an Application to Exponential Integrators. <u>SIAM Journal on Scientific Computing</u> 33, 488–511 (2011). URL http://epubs.siam.org/doi/10.1137/ 100788860.
- [3] Jones, E., Travis, O. & Peterson, P. SciPy: Open source scientific tools for Python (2001–). URL http://www.scipy.org/.
- [4] Romero, J. et al. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. <u>Quantum Science and Technology</u> 4, 014008 (2018). URL http://stacks.iop.org/2058-9565/4/i=1/a=014008?key= crossref.0e74114499ad0be8c7997e37f733ca0d.

Chapter 4

## Local Minima and Barren Plateaus in the ADAPT-VQE Landscape

#### ADAPT-VQE mitigates rough parameter landscapes and barren plateaus

Harper R. Grimsley and Nicholas J. Mayhall\* Chemistry Department, Virginia Tech

George S. Barron, Edwin Barnes, and Sophia E. Economou *Physics Department, Virginia Tech* (Dated: October 31, 2022)

Variational quantum eigensolvers (VQEs) represent a powerful class of hybrid quantum-classical algorithms for computing molecular energies. Various numerical issues exist for these methods, however, including barren plateaus and large numbers of local minima. In this work, we consider Adaptive, Problem-Tailored (ADAPT)-VQE ansätze, and examine how they are impacted by these local minima. We find that while ADAPT-VQE does not remove local minima, the gradient-informed, one-operator-at-a-time circuit construction accomplishes two things: First, it provides an initialization strategy that can yield solutions with over an order of magnitude smaller error compared to random initialization, and which is applicable in situations where chemical intuition cannot help with initialization, i.e., when Hartree-Fock is a poor approximation to the ground state. Second, even if an ADAPT-VQE iteration converges to a local trap at one step, it can still "burrow" toward the exact solution by adding more operators, which preferentially deepens the occupied trap. This same mechanism helps highlight a surprising feature of ADAPT-VQE: It should not suffer optimization problems due to barren plateaus and random initialization. Even if such barren plateaus appear in the parameter landscape, our analysis suggests that ADAPT-VQE avoids such regions by design.

#### I. INTRODUCTION

Quantum computers have long been viewed as a promising technology for quantum simulation.<sup>1</sup> However, the limited capabilities of Noisy, Intermediate-Scale Quantum (NISQ) devices restrict the types of algorithms that can be implemented at present.<sup>2</sup> While quantum phase estimation (QPE) provides a route to efficient molecular simulation,<sup>3</sup> the presence of both noise and errors on NISQ devices make near-term implementation of large-scale phase estimation intractable.

In response to the intractability of QPE, the variational quantum eigensolver (VQE) was introduced by Peruzzo et. al.<sup>4</sup> as a hybrid quantum-classical approach to finding approximate eigenvalues of a Hamiltonian,  $\mathcal{H}$ . In VQE, a quantum processor is used to apply a parameterized unitary transformation expressed as a quantum circuit (or even a direct pulse<sup>5-7</sup>),  $\mathcal{U}(\boldsymbol{\theta})$ , to some easily prepared reference state,  $|0\rangle$ .<sup>4,8–11</sup> The target Hamiltonian is then measured with the prepared state to obtain the energy as a function of circuit parameters:

$$E(\boldsymbol{\theta}) = \langle 0 | \mathcal{U}^{\dagger}(\boldsymbol{\theta}) \mathcal{H}\mathcal{U}(\boldsymbol{\theta}) | 0 \rangle.$$
 (1)

Using such quantum resources to prepare states and measure observables, a VQE will classically optimize  $\theta$  in order to minimize  $E(\theta)$ . The quality of the optimal energy for a given VQE is naturally dependent on the quality of the parameterization  $\mathcal{U}(\theta)$ , but because unitary operators are norm-preserving, the energy in Eq. (1) is variationally bounded from below by the ground-state energy of  $\mathcal{H}$ . The main advantage of VQEs is relatively low circuit depth,<sup>4</sup> avoiding the long, coherent evolutions of QPE.<sup>12</sup> This makes VQEs more appealing in the absence of fault-tolerant quantum computers. The circuit depth of a VQE is defined by the choice of  $\mathcal{U}$ , so that there is generally a trade-off between accuracy and circuit depth.

An outstanding challenge with many VQE ansätze is that the cost function, Eq. (1), creates a rough parameter landscape full of local minima, complicating the parameter optimization. Bittel and Kliesch have identified situations where there are so many far-from-optimal local minima that VQEs must be NP-hard in general.<sup>13</sup> The problem of local minima can be ameliorated through overparametrization in both quantum optimal control<sup>5,14</sup> and classical neural network settings.<sup>15,16</sup> This idea of overparametrization avoiding local minima has since been applied to VQEs: Rivera-Dean et. al. used this philosophy by employing a neural network to distort their cost function landscape mid-VQE.<sup>17</sup> This enabled them, in some cases, to escape from local minima.<sup>18</sup> Alternative strategies for avoiding local minima include collectively optimizing an ansatz for several Hamiltonians at the same time with a "snake" algorithm<sup>19</sup> and a "sweeping" approach to energy minimization called Unitary Block Optimization.<sup>20</sup>

A recent theoretical analysis by Larocca et. al. suggests that quantum neural networks (of which VQEs are a special case) undergo a sort of phase transition where local minima cease to be a problem.<sup>21</sup> This transition tends to occur when the number of parameters surpasses the dimension of the associated ansatz's dynamical Lie algebra, or DLA. The DLA for an ansatz of the form  $|\Psi\rangle = e^{\theta_1 A_1} e^{\theta_2 A_2} \dots e^{\theta_M A_M} |\phi_0\rangle$  is defined as the span of the set of repeated commutators of  $\{\hat{A}_i\}$ . As the authors point out, their results imply that this desirable overparametrization is likely to be unachievable for ansätze due to the exponential scaling of the DLA dimension with ansatz length. Perhaps even more alarmingly, Wierichs

et. al. were able to identify situations where adding additional parameters actually hurts the performance of gradient descent methods.<sup>22</sup>

In addition to the problems with local traps, it has recently been recognized that VQEs might also become impossible to optimize (even to a local mininum) as the system size increases. For sufficiently flexible or expressive VQE ans $\ddot{a}tze^{23}$ , it has been found that the energy landscape flattens (as quantified by the variance in the parameter gradients) exponentially fast as the system size increases.<sup>24</sup> The exponential growth of these flat landscapes (so-called "barren plateaus"), means that only a vanishingly small region of parameter space exists which has gradients large enough to measure with high enough precision to perform gradient descent. This region of concentrated cost has been termed a "narrow gorge."<sup>25</sup> As a result, initializing the optimization from a random point in parameter space is bound to land in a barren plateau, meaning that the number of circuit executions (shots) needed to resolve the search direction increases exponentially with the number of qubits, preventing any opportunity for quantum advantage. While intelligent heuristics for parameter initializations might help protect an optimization from getting stuck in a barren plateau (e.g., starting from a Hartree-Fock solution in molecular VQEs), the success is largely determined on a case-bycase basis.<sup>24</sup>

In this work, we present arguments and numerical simulations that indicate that our recently introduced adaptive variational algorithm, ADAPT-VQE,<sup>26</sup> is expected to be effectively immune to local minima and barren plateaus in the parameter landscape, at least in the noisefree case. Both issues are avoided because the algorithm systematically "burrows" a deep well in the landscape until the global minimum is reached. In other words, ADAPT-VQE dynamically modifies its parameter landscape in such a way that problematic regions are never explored. This phenomenon can be understood directly from the gradient criterion used to iteratively update the wavefunction ansatz. We illustrate this behavior with simulations of several different molecules. In Appendix B, we also show that the smoothness of the landscape can be controlled by intentionally overparameterizing the ansatz. In Appendix F, we show how the fidelity (overlap with the target state) is affected by number of parameters.

#### II. ADAPT-VQE ALGORITHM AND NUMERICAL SIMULATIONS

#### A. ADAPT-VQE

In recent work, we developed a dynamic framework for constructing ansätze that have much faster energy convergence with respect to circuit depth. This approach, referred to as ADAPT-VQE,<sup>26,27</sup> uses measurements of the molecular energy gradient to dynamically grow an ansatz, operator by operator, creating a highly compact ansatz that quickly converges to the exact solution. Defining a pool of anti-Hermitian operators,  $\mathcal{A} = \{A_i\}$ , we outline the steps in Algorithm 1.

Algorithm 1: ADAPT-VQE Algorithm			
$k \leftarrow 0;$			
$\mathcal{U}_{0}\left(oldsymbol{ heta} ight) \leftarrow 1;$			
while $Converged = False do$			
for $A_i \in \mathcal{A}$ do			
$  \mathbf{g}_i \leftarrow$			
$\left  \frac{\partial}{\partial \theta_i} \left\langle 0   \mathcal{U}_k^{\dagger} \left( \boldsymbol{\theta} \right) e^{-\theta_i A_i} \mathcal{H} e^{\theta_i A_i} \mathcal{U}_k \left( \boldsymbol{\theta} \right)   0 \right\rangle \right _{\theta_i = 0, \boldsymbol{\theta}};$			
end			
$i \leftarrow \text{index of largest element of } \mathbf{g};$			
$k \leftarrow k+1;$			
$\mathcal{U}_{k}\left(oldsymbol{ heta} ight)\leftarrow e^{ heta_{i}A_{i}}\mathcal{U}_{k}\left(oldsymbol{ heta} ight);$			
$\boldsymbol{\theta} \leftarrow \operatorname{argmin}_{\boldsymbol{\theta}} \langle 0   \mathcal{U}_{k}^{\dagger}(\boldsymbol{\theta}) \mathcal{H} \mathcal{U}_{k}(\boldsymbol{\theta})   0 \rangle;$			
$E_{k} \leftarrow \langle 0   \mathcal{U}_{k}^{\dagger}(\boldsymbol{\theta}) \mathcal{H} \mathcal{U}_{k}(\boldsymbol{\theta})   0 \rangle;$			
end			
Return $E_k, \theta$ ;			

At each ADAPT-VQE iteration, the gradient,  $\frac{\partial E}{\partial \theta_i}$ , is measured with respect to all operators in the pool. The operator with the largest gradient magnitude is then added to the ansatz with the associated parameter initialized to zero. The other parameters in the ansatz are initialized using the optimal values from the previous step (we refer to this as parameter "recycling"). At this point, an ordinary VQE is performed using some classical optimization algorithm. In this work, we exclusively use the Brovden–Fletcher–Goldfarb–Shanno (BFGS) method.<sup>28</sup> a quasi-Newton strategy, because we are explicitly seeking information about local minima, and because we are not including any noise models in our simulations. In all cases, a gradient norm of  $1 \times 10^{-8}$  was pursued, but not necessarily achieved, by the solver. In cases where the solver could not achieve this accuracy, its output was still used. Because we initialize the new parameter added during each ADAPT-VQE iteration to zero, the new trial circuit is equivalent to the previous one during the first VQE iteration. Consequently, the energy can only improve during this VOE, i.e., the energy decreases monotonically. Parameters are added one-by-one in this fashion until some convergence criteria are achieved. Reasonable choices include the norm (either  $l^2$  or  $l^{\infty}$ ) of the vector of gradients, g, or the number of operators in the ansatz.

All simulations were conducted using a locally developed code which can be found on GitHub at https: //github.com/hrgrimsl/adapt. OpenFermion<sup>29</sup> was used to construct matrix representations of operators under the Jordan-Wigner transformation and PySCF<sup>30</sup> was used to obtain molecular integrals. Because our focus in this work is to first understand the noise-free parameter landscapes associated with ADAPT-VQE, all simulations are performed without any noise models. Future work will explore how the presence of noise affects the landscapes. For all the ADAPT-VQE calculations in this work, the unitary coupled cluster with singles and doubles (UCCSD) operator pool is used,<sup>26</sup> without spincomplemented or spin-adapted operators. While many different pools can be used for ADAPT-VQE calculations, in this paper we focus primarily on the original fermionic pool due to its robustness in that it seems to consistently converge to an exact eigenstate and has a connection with the stationary conditions of the Anti-Hermitian Contracted Schrödinger equation.<sup>31</sup> Details of this pool are provided in Appendix A.

#### B. Prevalence and Distribution of Local Minima

In this section, we numerically explore the parameter landscapes of several example systems using ADAPT-VQE. Our aim is to characterize the way in which the number and distribution of local minima change as ADAPT-VQE gradually increases the length of the ansatz (and thus the depth of the circuit). For each molecule and bond distance considered, we first run ADAPT-VQE normally, where the initial parameter values used in the VQE at each iteration of the algorithm are chosen to be the "recycled" parameters, i.e., the optimal values obtained from the previous iteration. This yields an ansatz that reproduces the target ground state with high accuracy.

After using ADAPT-VQE to define the ansatz, we then use this ansatz to search for local minima by repeatedly reinitializing each VQE with randomly chosen parameters, and reoptimizing.<sup>32</sup> In this work, we performed 1000 such random initializations for each ansatz considered unless otherwise specified. The numbers of samples were chosen due to computational considerations, and tests were performed to verify that increasing the number of random initializations does not change the results qualitatively. For each layer of the ansatz and each random initialization, we record the minimum energy obtained by the VQE subroutine. These values correspond to the energies of local minima in the landscape associated with each ansatz.

In addition to these random initializations, we also include both the "recycled" parameters from the previous VQE (the default initialization in ADAPT-VQE<sup>26</sup>) and the **0** parameter vector associated with the Hartree-Fock (HF) reference. All 1002 initializations of a given ansatz are then optimized with BFGS, and the resulting energy errors are shown with rainbow-colored bars in each figure. The colors indicate relative energy ordering at a given ansatz, such that red corresponds to the highest energy and violet to the lowest energy. The recycled initialization's outcome is of particular interest since this is the default, deterministic initialization for ADAPT-VQE, and the approach used when growing the ansätze used in the data. These conventions will be used throughout this work.

We consider linear  $H_4$  (8 qubits) at 1 and 3 Å and



FIG. 1. ADAPT-VQE results for  $H_4$  at 1 Å. The *x*-axis corresponds to the number of ADAPT-VQE iterations, i.e. the number of operators in the ansatz at a given step. The *y*-axis corresponds to the error from the exact FCI energy. The red curve corresponds to the energy obtained through BFGS minimization using an HF guess, i.e. one where all parameters are zero. The green curve corresponds to the energy obtained through BFGS minimization using the standard ADAPT-VQE in which optimal parameter values in one iteration are recycled as initial guesses in the next iteration, and with the new parameter initialized to zero. The colored dots correspond to all the energies obtained through BFGS optimizations, with red being the highest energy and violet the lowest.

linear  $H_6$  (12 qubits) at 1, 2, and 3 Å as toy models exhibiting varying degrees of electron correlation (and entanglement in the target wavefunction). While not interesting as chemistry agents, the fictitious molecules  $H_4$ and  $H_6$  provide an excellent testbed for quantifying the effect of strong correlation. Such "molecules" are often used as surrogates for real strongly correlated systems such as ones involving transition metals, which are too large to simulate classically. In addition, we study LiH (12 qubits) at 1.62 Å and BeH<sub>2</sub> (14 qubits) at 1.33 Å as examples of real molecules at equilibrium geometries. These geometries were obtained through optimization at the B3LYP<sup>33</sup>/6-31G<sup>\*34-37</sup> level of theory in PySCF,<sup>30</sup> and are given in the Supplementary Information. All ADAPT-VQE calculations were performed in the STO- $3G^{38,39}$  basis. No symmetries were used to reduce the number of qubits. In cases where the exact solution was not obtained, the number of ADAPT-VQE iterations was determined by computational considerations.

 $H_4$  molecule In Fig. 1 we show the energies (relative to the global minimum obtained from a full configuration interaction (FCI) calculation) of the various local minima as a function of ansatz length (as defined by the ADAPT-VQE algorithm). After a short period without local minima, the random initializations begin to diverge to an increasing number of distinct local mini-



FIG. 2. ADAPT-VQE Results for  $H_4$  at 3 Å. The axes and colors are as in Fig. 1.

ima as the number of parameters increases. In contrast to the random initializations, both the HF and the recycled initializations converge to the same minimum for  $H_4$  at 1 Å, which is consistently better than the average random initialization. This is our first indication that good initializations can reliably avoid high-energy traps. Interestingly, even though ADAPT-VQE doesn't always find the lowest energy trap, it does eventually converge. Additionally, we observe that there are still many local minima even after these "chemically informed" guesses are able to reach the exact ground state. In Appendix B, we consider the prospect of removing local minima through systematic overparameterization for  $H_4$  at 1 Å. While we are successful in removing local minima using a novel "ADAPT<sup>N</sup>" approach, deeper circuits are actually required to achieve the overparameterization than to simply add operators until ADAPT-VQE reaches the ground state in spite of local minima.

In Fig. 2, we see that for the more strongly correlated 3 Å bond distance, the HF and recycled initializations differ. The recycled initialization is able to reach the ground state with fewer parameters than the HF initialization, though this behavior is not consistently observed in other systems. Again, we see ADAPT-VQE converging to the exact solution far faster than a typical (yellow-green) random initialization.

 $H_6$  molecule In Fig. 3, we begin to see the true power of an intelligent guess by simulating  $H_6$  at 1Å. As the ansatz grows longer, a massive gap opens up between the random guesses and the HF/recycled ones. This gap implies that in practice, it is very difficult to do better than simply recycling the previous parameters in ADAPT. This gap is further numerical evidence of a "narrow gorge", in which the exact solution is hypothesized to exist.<sup>25</sup> Although such a landscape is often associated with optimization difficulties, here we see that ADAPT-VQE is able to stay very close to the narrow gorge, avoiding such issues. We emphasize that this feature is not only a result of good initialization,<sup>40</sup> but rather a cooperative effect between initialization and the gradient-guided ansatz construction. In Appendix D, we demonstrate this explicitly by performing simulations using the recycled initialization, but on randomized (not gradient-guided) ansätze. We finally notice a sharp increase in the median around 140 parameters. This indicates that as the number of parameters increases, so too does the number of local traps. Furthermore, these new traps are preferentially high in energy, thus moving the median solution to higher energies. This further implies that as the system grows in size, the overwhelming num-

In Fig. 3, the same gap appears for  $H_6$  at 2 Å that appeared at 1 Å. As the ansatz grows in depth (i.e., around 50 parameters), we notice an earlier rise in the median energy of the traps found.

ber of solutions will be high in energy, making random

sampling of VQE initializations intractable.

In Fig. 3, for  $H_6$  at 3Å, the energy distribution of the local traps significantly increases at the beginning, but chokes up around 100 parameters where the large gap is seen again. The HF and recycled initializations are still far better than random ones. We see the sharp increase in the median again here.

LiH molecule In Fig. 4 we see similar behavior for LiH to that of  $H_6$  at 1 Å. While the solution gap is less pronounced, both HF and the recycled initialization are always significantly better than nearly every random initialization.

 $BeH_2$  molecule We observe similar behavior once again in Fig. 5 for BeH<sub>2</sub>, with the exception that a large gap is observed.

In all cases, we observe that for more than a few parameters, local minima emerge, and for large numbers of parameters, these minima often dominate the energy landscape. In many cases initializing all parameters to 0 (HF) is a reasonable choice that leads to low energy minima.

Trap "Burrowing" The problem of local minima seems to be partially mitigated by ADAPT-VQE itself. Even in cases where the recycled initialization converges to a high-energy trap, ADAPT-VQE progresses by adding an operator which is chosen to preferentially deepen the current trap (via the gradient criterion). As such, over a sequence of ADAPT-VQE iterations, the current trap becomes increasingly deep relative to the other parameter traps, such that a gap can open up between the current minimum (which approaches the global minimum) and all other local minima. Thus ADAPT-VQE appears to "burrow" into the parameter landscape, creating a single deep well as opposed to stabilizing all local minima (i.e., reaching overparameterization). This burrowing effect is depicted graphically in Fig. 6.

5



FIG. 3. ADAPT-VQE Results for  $H_6$  at 1 Å. The axes and colors are as in Fig. 1.



FIG. 4. ADAPT-VQE Results for LiH at 1.62 Å. The axes and colors are as in Fig. 1.



FIG. 5. ADAPT-VQE Results for  $BeH_2$  at 1.33Å. The axes and colors are as in Fig. 1.



FIG. 6. Schematic cartoon of how the parameter landscapes change as parameters are added using a) ADAPT-VQE, and b) a "controllable" or overparameterized ansatz which has a greater number of parameters than the rank of the DLA. Here the *y*-axis is meant to convey error, and the *x*-axis is meant to convey a generalized coordinate in parameter space.

#### C. Sensitivity to Parameter Landscape Gradients

#### 1. Insensitivity to Barren Plateaus

In the previous section, we demonstrated that while the parameter landscapes exhibit a large number of local traps that are high in energy, ADAPT-VQE is robust due to the fact that any local minimum in early stages of the algorithm can often be deepened into a global minimum at later stages. This same mechanism implies a similar robustness to the presence of barren plateaus. As mentioned above, the barren plateau phenomenon has been recently recognized as a serious obstacle to the use of VQEs in practical settings. The problem arises from the observation that highly expressive ansätze (more specifically, circuits which form a 2-design), which are attractive from an accuracy perspective, exhibit an exponentially decreasing gradient variance with increasing system size. This means that the vast majority of parameter space becomes essentially flat. In the course of optimizing the parameters of such an expressive ansatz, a randomly chosen initialization will (with overwhelming probability) correspond to a point in parameter space where the gradient of the cost function is so small that an exponentially large number of measurements are needed to resolve a

meaningful search direction in the presence of noise. As a result, the ability to optimize or train such expressive circuits is suspect at best. While a physically inspired parameter initialization can be effective (e.g., HF initialization), difficult cases (like those exhibiting strong correlation) may prevent efficient initialization.

Unlike the non-adaptive situation in which a static ansatz is first defined and then optimized, ADAPT-VQE slowly brings a given stationary point (initially the reference state) to the exact solution, via this burrowing mechanism. As such, each VQE subroutine performed along the way is "warm-started", in that one already has a decent initialization coming from the previous optimization. Using this recycled initialization, we have a clear characterization of the parameter landscape about the initial point: all previous parameters are optimized, and thus have zero gradients, and the newly added operator has a large gradient by design, since we specifically add the operator with the largest gradient. This means that each VQE subroutine in the ADAPT-VQE algorithm is initialized with a single parameter which is guaranteed to be greater than  $\epsilon$  (the ADAPT-VQE convergence threshold). Based on this argument, we do not expect difficulty due to barren plateaus when training ADAPT-VQE ansätze as system sizes are scaled up. We emphasize that this argument does not suggest that the ansätze constructed by ADAPT-VQE are free from barren plateaus, only that our algorithm remains localized to a region in parameter space with significant gradients.

We note that our analysis focuses exclusively on barren plateaus that arise from highly expressive circuits. ADAPT-VQE may still suffer from noise-induced barren plateaus (NIBP's),<sup>41</sup> which present problems for any VQE ansatz that scales polynomially in depth with system size, since they are a direct consequence of decoherence. Due to the problem-tailored nature of ADAPT-VQE and the computational difficulty of simulating increasingly large system sizes classically, we do not yet know how ADAPT-VOE ansätze scale with system size. Extrapolations from small system simulations will likely provide an overly pessimistic estimation due to the fact that correlation length will not simultaneously increase (at least for gapped systems). For a constant accuracy threshold, we expect the ansatz length to scale at least linearly (and thus ultimately suffer from NIBP's), though a detailed study of this is not yet available. However, even if we assume that ADAPT-VQE might have an exponential scaling asymptotically, the problems of interest to chemistry are far from the asymptotic limit (around 100 logical qubits), and it is possible that a quantum advantage could still be demonstrated on finite problem instances. As such, further investigation into ADAPT-VQE's performance in the presence of noise in general is indeed warranted.

 $\mathbf{6}$ 



FIG. 7. ADAPT-VQE with recycled parameters for  $H_4$  at 3 Å. The x-axis corresponds to the ADAPT-VQE iteration. The green curve depicts the gradient associated with the operator to be added at each step, while the red curve depicts the energy at each ADAPT-VQE step. The blue lines depict the excited FCI eigenstates which are lower than the HF energy.

#### 2."Gradient Troughs"

Although barren plateaus seem to pose no threat to the ability to scale up ADAPT-VQE based on the arguments in the previous section, there is still a related issue that might prevent ADAPT-VQE from converging to accurate solutions. As described above, at each ADAPT-VQE step, the ansatz is extended using the operator with the largest gradient:

$$\frac{\partial E}{\partial \theta_i} = \langle \psi(\theta) | \left[ \hat{H}, \hat{A}_i \right] | \psi(\theta) \rangle \,. \tag{2}$$

The ansatz is then repeatedly extended until the largest gradient in the operator  $pool^{42}$  is smaller than some threshold,  $\epsilon$ . Noise on a NISQ device, however, defines some lowest possible threshold,  $\epsilon_{\min}$ , that can be resolved using a given shot allowance. In our earlier work,<sup>26</sup> we sometimes observed non-monotonic convergence of the gradients as a function of ansatz length (although the energy convergence is guaranteed to be monotonic), such that as the ansatz is extended, the pool gradients might first decrease, then increase again before finally converging. This "gradient trough", therefore presents a challenge in the presence of noise. If a gradient trough appears and drops below the NISQ resolvable threshold,  $\epsilon_{\min}$ , then the ADAPT-VQE algorithm may halt prematurely.

How do these gradient troughs grow with system size? If we were to find that they grow exponentially fast, meaning that the largest gradient in the operator pool is exponentially suppressed as the number of qubits increases, then this would suggest concern for the scalability of ADAPT-VQE. However, this does not need to be the case. Choosing a local orbital  $basis^{43}$  one can imagine trivial situations where the gradients not only avoid exponential suppression, but any suppression at all. Consider the  $n^{th}$  iteration of an ADAPT-VQE calculation of



FIG. 8. Gradient troughs at eigenstates. Overlay of the ADAPT-VQE error vs. iteration (red line, right axis) with the FCI excited states (blue horizontal lines, right axis) for H<sub>6</sub> at 2, 3, 4, and 5 Å. x-axis corresponds to the ADAPT-VQE iteration. The largest pool gradient associated with the operator to be added at each step is shown in green (left axis).

a molecular wavefunction,  $|\psi_n\rangle$ . If one were to double the number of qubits by adding another molecule (at infinite distance so as to remove interactions between the systems), the total wavefunction at iteration 2n would have a product form,  $|\psi_{2n}^{AB}\rangle = |\psi_n^A\rangle |\psi_n^B\rangle$ . Any pool operator  $\hat{O}_i$  that is local to either subsystem has the exact same gradient in the supersystem,  $|\psi_{2n}^{AB}\rangle$ , as it does in the subsystem,  $|\psi_n^A\rangle$ . For example, consider an operator,  $\hat{O}_i^A$ , local to subsystem A:

$$\frac{\partial E^{AB}}{\partial \theta_{i}} = \left\langle \psi_{2n}^{AB} \middle| [\hat{H}, \hat{O}_{i}^{A}] \middle| \psi_{2n}^{AB} \right\rangle \\
= \left\langle \psi_{2n}^{AB} \middle| [\hat{H}^{A} + \hat{H}^{B}, \hat{O}_{i}^{A}] \middle| \psi_{2n}^{AB} \right\rangle \\
= \left\langle \psi_{n}^{A} \middle| [\hat{H}^{A}, \hat{O}_{i}^{A}] \middle| \psi_{n}^{A} \right\rangle \left\langle \psi_{n}^{B} \middle| \psi_{n}^{B} \right\rangle \\
= \frac{\partial E^{A}}{\partial \theta_{i}}.$$
(3)

The additive separability of non-interacting subsystems is referred to as "size-consistency" in the chemistry literature. However, in addition to additive separability of the energy, size-extensive wavefunctions (like UCCSD) also demonstrate "size-intensivity" for intensive properties (e.g., density, optical gaps, etc). As shown in Eq. 3, the gradient with respect to a local rotation is not affected by the presence of an additional non-interacting system, thus demonstrating size-intensivity.

In the limit of a large system, any further additions to the system size will necessarily be too far away from a given subsystem to interact. Based on this argument, we don't expect gradient troughs to deepen asymptotically with system size. However, more work is needed to characterize the behavior of gradient troughs as the system size increases in the presence of interactions.

#### 3. Effect of Low-Lying FCI Eigenstates

In order to understand the nature of the "gradient troughs" discussed in Sec. II C 2, and shown in Figures 7 and 8, we superimposed the low-lying FCI energies with the ADAPT-VQE energies computed. The FCI spectrum is plotted as a set of blue horizontal lines. We only plot  $H_4$  and  $H_6$ , as the other systems studied have no nearby excited states, nor do they exhibit any gradient troughs. In the region of the gradient trough, the energy also becomes very flat, (i.e., consider operators 9-16 in Fig. 7 and operators 50-100 in Fig. 8).

By plotting the exact eigenstates on top of these curves, one readily sees that the gradient troughs occur when ADAPT-VQE falls inside of a nearly degenerate manifold of FCI excited states. Should the ADAPT-VQE threshold be chosen loose enough (or if there is too much device noise to measure the gradient below this value) that the algorithm is aborted in this region, then ADAPT-VQE will be unable to advance further toward the ground state, remaining stuck as an approximation to an excited state (or in general some arbitrary superposition of the nearly degenerate eigenstates). This appearance of gradient troughs was first noticed in the paper that introduced ADAPT-VQE,<sup>26</sup> however the origin of the onset and the interpretation was not clear at that time.

As a consequence, although ADAPT-VQE isn't expected to suffer from the more general problem of barren plateaus, more work is needed to understand how to escape any gradient troughs to ensure smooth convergence to the exact solution, particularly when noise is included. This remains an outstanding problem associated with ADAPT-VQE, warranting more research.

#### III. CONCLUSIONS

Underparameterized ansätze are difficult to optimize due to large numbers of local minima, while highly expressive ansätze are difficult to optimize due to barren plateaus. In this paper, we find that ADAPT-VQE does not necessarily suffer from these challenges. We have studied the parameter landscapes arising from various ADAPT-VQE generated ansätze and have arrived at the following conclusions:

- 1. Chemically informed initialization helps avoid traps: ADAPT-VQE's process of re-using parameters at each step focuses the search space on a local region, keeping the algorithm relatively easy to train despite the rough overall landscape. The parameter vector from the previous iteration tends to be a relatively good initial guess for the following ADAPT-VQE iteration. This means that by simply "recycling" the parameters from one ADAPT-VQE iteration to the next, the vast majority of parameter traps are entirely avoided. Similarly, it seems that the chemical intuition granted by the HF state avoids most traps.
- 2. Trap burrowing corrects local minima: Even if the early iterations get stuck in a trap, the adaptive construction iteratively extends the ansatz in a direction that is guaranteed to improve the cost function near the current stationary point. By continuously focusing on a local point in parameter space, ADAPT-VQE can "burrow" into a given local minimum, even if the vast majority of traps remain high in energy.
- 3. Barren plateau avoidance: The nature of the ADAPT-VQE algorithm suggests that barren plateaus should not prove problematic in the parameter optimization step. This originates from the fact that ADAPT-VQE specifically adds a large gradient operator, generating a steep landscape, such that a search direction is resolvable without an exponential number of shots.
- 4. Gradient troughs: ADAPT-VQE can still exhibit numerical challenges. An exponentially vanishing *pool* operator gradient could potentially arise, resulting in ADAPT-VQE becoming stuck during the operator addition step (in contrast to the parameter optimization step). Numerical evidence suggests that these gradient troughs appear when the ADAPT-VQE energy starts to converge near one or more excited states. Heuristics for diag-

nosing and addressing such issues will be the focus of future work.

Despite the presence of local minima and the possibility of barren plateaus in standard ADAPT-VQE ansatze, we conclude that ADAPT-VQE can be optimized reasonably well through parameter recycling. Consequently, in addition to being parameter- and gate- efficient, ADAPT-VQE appears to be relatively immune to the problems of both local minima and barren plateaus in VQEs.

#### IV. DATA AVAILABILITY

All data generated with code available at https://github.com/hrgrimsl/adapt.

#### V. CODE AVAILABILITY

Code developed for this project is open-source and available at https://github.com/hrgrimsl/adapt.

#### VI. COMPETING INTERESTS

The authors declare no competing financial or non-financial interests.

#### VII. AUTHOR CONTRIBUTIONS

H.R.G. wrote the code used in this work and ran all simulations. All authors contributed to the design of simulations, theoretical developments, and writing in the paper.

#### VIII. ACKNOWLEDGEMENTS

N.J.M., S.E.E., and E.B. are grateful for financial support provided by the U.S. Department of Energy. N.J.M. and E.B. acknowledge Award No. DE-SC0019199. S.E.E. acknowledges the DOE Office of Science, National Quantum Information Science Research Centers, Co-design Center for Quantum Advantage (C2QA), Contract No. DE-SC0012704. H.R.G. acknowledges support provided by the Institute for Critical Technology and Applied Science at Virginia Tech. The authors thank the Advanced Research Computing at Virginia Tech for the computational infrastructure.

- R. P. Feynman, International Journal of Theoretical Physics 21, 467 (1982).
- [2] J. Preskill, Quantum 2, 79 (2018), publisher: Verein zur Förderung des Open Access Publizierens in den Quan-

tenwissenschaften.

- [3] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Science 309, 1704 (2005), publisher: American Association for the Advancement of Science.
- [4] 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).
- [5] A. Asthana, C. Liu, O. R. Meitei, S. E. Economou, E. Barnes, and N. J. Mayhall, arXiv:2203.06818 [physics, physics:quant-ph] (2022), arXiv: 2203.06818.
- [6] O. R. Meitei, B. T. Gard, G. S. Barron, D. P. Pappas, S. E. Economou, E. Barnes, and N. J. Mayhall, npj Quantum Information 7, 1 (2021), number: 1 Publisher: Nature Publishing Group.
- [7] A. B. Magann, C. Arenz, M. D. Grace, T.-S. Ho, R. L. Kosut, J. R. McClean, H. A. Rabitz, and M. Sarovar, PRX Quantum 2, 010101 (2021), publisher: American Physical Society.
- [8] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, Chemical Reviews **119**, 10856 (2019), publisher: American Chemical Society.
- [9] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Nature Reviews Physics 3, 625 (2021).
- [10] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, Physics Reports The Variational Quantum Eigensolver: a review of methods and best practices, 986, 1 (2022).
- [11] D. A. Fedorov, B. Peng, N. Govind, and Y. Alexeev, Materials Theory 6, 2 (2022).
- [12] A. Y. Kitaev, arXiv:quant-ph/9511026 (1995), arXiv: quant-ph/9511026.
- [13] L. Bittel and M. Kliesch, Physical Review Letters 127, 120502 (2021), publisher: American Physical Society.
- [14] G. Riviello, K. M. Tibbetts, C. Brif, R. Long, R.-B. Wu, T.-S. Ho, and H. Rabitz, Physical Review A 91, 043401 (2015), publisher: American Physical Society.
- [15] D. Lopez-Paz and L. Sagun, in International Conference on Learning Representations (2018).
- [16] S. S. Du and X. Zhai, in International Conference on Learning Representations (2019).
- [17] J. Rivera-Dean, P. Huembeli, A. Acín, and J. Bowles, arXiv:2104.02955 [quant-ph] (2021), arXiv: 2104.02955.
- [18] The neural network temporarily adds additional "weight" parameters to the optimization. Even when the neural network is then reset to the identity, a better set of parameters  $\boldsymbol{\theta}$  was sometimes found for the undistorted cost function.
- [19] D.-B. Zhang and T. Yin, Physical Review A 101, 032311 (2020).
- [20] L. Slattery, B. Villalonga, and B. K. Clark, Physical Review Research 4, 023072 (2022), publisher: American Physical Society.
- [21] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, arXiv:2109.11676 [quant-ph, stat] (2021), arXiv: 2109.11676.
- [22] D. Wierichs, C. Gogolin, and M. Kastoryano, Physical Review Research 2, 043246 (2020).
- [23] Formal arguments have largely been restricted to 2design structures.

- [24] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nature Communications 9, 4812 (2018), number: 1 Publisher: Nature Publishing Group.
- [25] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, Quantum Science and Technology 7, 045015 (2022), publisher: IOP Publishing.
- [26] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, Nature Communications 10, 3007 (2019).
- [27] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, PRX Quantum 2, 020310 (2021), publisher: American Physical Society.
- [28] R. Fletcher, Practical Methods of Optimization, 2nd ed. (Wiley, Chichester, 2000).
- [29] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. O'Brien, B. O'Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. P. D. Sawaya, B. Senjean, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush, Quantum Science and Technology 5, 034014 (2020), publisher: IOP Publishing.
- [30] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, WIREs Computational Molecular Science 8, e1340 (2018).
- [31] D. A. Mazziotti, Physical Review Letters 97, 143002 (2006), publisher: American Physical Society.
- [32] Each parameter was randomly initialized on an interval of length  $2\pi$  in order to coincide with the period of  $e^{A_i\theta_i}$  for the chosen pool.
- [33] A. D. Becke, The Journal of Chemical Physics 98, 5648 (1993), publisher: American Institute of Physics.
- [34] J. D. Dill and J. A. Pople, The Journal of Chemical Physics 62, 2921 (1975), publisher: American Institute of Physics.
- [35] R. Ditchfield, W. J. Hehre, and J. A. Pople, The Journal of Chemical Physics 54, 724 (1971), publisher: American Institute of Physics.
- [36] P. C. Hariharan and J. A. Pople, Theoretica chimica acta 28, 213 (1973).
- [37] W. J. Hehre, R. Ditchfield, and J. A. Pople, The Journal of Chemical Physics 56, 2257 (1972), publisher: American Institute of Physics.
- [38] W. J. Hehre, R. F. Stewart, and J. A. Pople, The Journal of Chemical Physics 51, 2657 (1969), publisher: American Institute of Physics.
- [39] J. B. Collins, P. von R. Schleyer, J. S. Binkley, and J. A. Pople, The Journal of Chemical Physics 64, 5142 (1976), publisher: American Institute of Physics.
- [40] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, Quantum Machine Intelligence 3, 5 (2021).
- [41] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Nature Communications 12, 6961 (2021).
- [42] In the first paper the convergence criterion was taken to be the norm of the gradients in the pool, rather than the maximum.
- [43] One is always free to rotate occupied or virtual orbitals without changing the associated Slater determinant, due to orbital subspace rotational invariance.

- [44] F. A. Evangelista, G. K.-L. Chan, and G. E. Scuseria, The Journal of Chemical Physics 151, 244112 (2019).
- [45] M. D. Sapova and A. K. Fedorov, Communications Physics 5, 1 (2022), number: 1 Publisher: Nature Publishing Group.

#### Appendix A: Operator Pool

In this work, with the exception of Appendix E, all experiments were performed with a pool of particle-hole (i.e., not generalized) Fermionic operators. These operators were restricted to single excitations of the form  $\hat{a}^a_i - \hat{a}^i_a$  and double excitations of the form  $\hat{a}^{ab}_{ij} - \hat{a}^{ij}_{ab}$ , where i and j are occupied spin orbitals and a and b are virtual spin orbitals. While these operators are not required to respect  $\hat{S}^2$  symmetry, the pool does not include "spin flips," operators which change the total numbers of alpha and beta electrons. This pool grows quartically with system size. If these operators are used as generators for an ansatz of the form  $|\Psi\rangle = e^{\theta_1 A_1} e^{\theta_2 A_2} \dots e^{\theta_M A_M} |\phi_0\rangle$ with infinite repetition of the operators and correct ordering, the FCI solution should be reachable.<sup>44</sup> However, we have not yet proven that the ADAPT-VQE convergence criterion goes to zero *only* when converged to an exact eigenstate.

#### Appendix B: ADAPT<sup>N</sup>

In this section we turn our attention to the following question: can we remove all local traps from ADAPT-VQE without necessarily reaching "controllability" where all local minima are exact? It follows from Ref. 21 that one can systematically overparametrize a given ansatz of the form  $|\Psi\rangle = e^{\theta_1 A_1} e^{\theta_2 A_2} \dots e^{\theta_M A_M} |\phi_0\rangle$ by simply repeating all the operators N times such that the number of parameters approaches or exceeds the dimension of the associated DLA. This ansatz repetition can be defined as:

$$|\Psi\rangle = \prod_{i=1}^{N} \left(\prod_{j=1}^{M} e^{\theta_{j}^{i} A_{j}}\right) |\phi_{0}\rangle.$$
 (B1)

We use the superscript on  $\theta_j^i$  to indicate the *i*th independent parameter associated with the same operator  $A_j$ . Because the same generators are being used, the DLA remains the same, but the number of parameters has now been multiplied by N. We apply this idea to ADAPT-VQE in what we call ADAPT<sup>N</sup>. As with normal ADAPT, we begin with a reference  $|\phi_0\rangle$ . We choose the first operator  $A_1$  to add as before, but apply it to the reference N times with N distinct parameters: B2.

$$|\Psi_1\rangle = e^{\theta_1^1 A_1} e^{\theta_1^2 A_1} \dots e^{\theta_1^N A_1} |\phi_0\rangle.$$
 (B2)

The energy of  $|\Psi_1\rangle$  is variationally minimized with respect to  $\boldsymbol{\theta}$ , beginning from  $\boldsymbol{\theta} = \mathbf{0}$ . For each operator  $A_i$ 



FIG. 9. ADAPT<sup>N</sup> for N = 1, 2, 3, 4 on H<sub>4</sub> at 1 Å separation. The axes and colors are as in Fig. 1. 300 Random initializations are performed for each ansatz.

Number of Parameters

in the operator pool, we now measure the energy gradient associated with adding it to the front of the ansatz with a parameter of 0. As with the original ADAPT-VQE, we choose the next operator in our ansatz,  $A_2$ , to be the operator with largest gradient magnitude. We then add the operator in a collated fashion:

$$|\Psi_{2}\rangle = \left(e^{\theta_{2}^{1}A_{2}}e^{\theta_{1}^{1}A_{1}}\right)\left(e^{\theta_{2}^{2}A_{2}}e^{\theta_{1}^{2}A_{1}}\right)\dots\left(e^{\theta_{2}^{N}A_{2}}e^{\theta_{1}^{N}A_{1}}\right)|\phi_{0}\rangle.$$
(B3)

This procedure of adding a new operator to the "core" ansatz, and replicating N times, is then repeated until convergence. The intuitive way to understand ADAPT<sup>N</sup> is that ADAPT-VQE is adding N parameters at a time, rather than the usual 1. However, unlike the "batched" approach of Ref. 45, our method does not add N different operators, a choice made to ensure that the DLA dimension does not increase with N. Because the ratio of the number of parameters to DLA dimension increases with N, we hypothesize that ADAPT<sup>N</sup> should exhibit fewer local minima with increased N.

To test this hypothesis, we consider first the numerically challenging linear H<sub>4</sub> at 1Å bond distance in the STO-3G basis. In Figure 9, we show the convergence of ADAPT<sup>N</sup> for  $N \leq 4$  as a function of the number of parameters. We consider 300 random initializations at each ansatz length in addition to the recycled initialization from the previous VQE subroutine, and the HF initialization. We then plot the optimized results of all 302 initializations.

For ADAPT<sup>1</sup> (i.e. standard ADAPT) and ADAPT<sup>2</sup>, we see similar behavior. Random guesses are frequently better than recycled ones, with random initializations of  $\theta$  landing in numerous local minima. For ADAPT<sup>1</sup> and ADAPT<sup>2</sup>, the landscape is dominated by local minima except when there are very few operators or when FCI has almost been reached. For ADAPT<sup>3</sup>, most of the local minima seem to be gone, and local minima never outperform the recycled initialization. For ADAPT<sup>4</sup>, every VQE subroutine appears to be free of local minima. This gives us a simple way to make ADAPT-VQE (or similar VQEs) trap-free. The downside is that more parameters and deeper circuits are required to introduce the same number of distinct operators. Based on these simulations, the lower values of N seem to be more depthefficient at achieving any given accuracy at any given circuit depth, even if only the recycled initialization is tried. Furthermore, based on Ref. 21, the necessary N to keep ADAPT-VQE trap-free will increase with the number of operators needed to describe a system, which will generally grow with system size. We therefore expect this overparametrization strategy to scale poorly. Contrary to the prediction of Ref. 22, however, we have shown that an ADAPT-like method can avoid local traps via overparametrization.

#### Appendix C: Difficulty of Exhaustive Sampling

While sampling over parameter initializations provides a path for enumerating different local traps in the landscape, the computational cost of such sampling is high. and any such study will necessarily exhibit artifacts arising from insufficient sampling. While our best efforts were made to avoid such artifacts, we have noticed a few. Take for example, the 3 Å plot in Figure 3, where the best solution at 60 parameters is very slightly worse than the best solution at 59 parameters. We compared using 300 and 1000 random initializations at 60 parameters (Fig. 10) for an ADAPT-VQE calculation on  $H_6$  at 3 Å and found little reason to believe that 1000 initializations are enough to exhaustively find all minima. While these small artifacts persist, the general trends seem converged. For instance, in Fig. 10, increasing the number of initializations increases the number of solutions found. The distribution of the solutions is essentially unchanged, with the median values (yellow-green) staying in roughly the same place.

#### Appendix D: Importance of Gradient-Guided Ansatz

In Figs. 1-5, we demonstrated that the recycled (and HF) initialization performed far better than the average local minimum. In order to ascertain the role that the ADAPT-VQE ansatz construction has on the ability of the ansatz to burrow, we carry out similar simulations to those presented in Fig. 3, except that before performing the VQE optimizations, we shuffle the operators in the ansatz. As a result, the ansatz is no longer constructed by ADAPT-VQE. The results are shown in Fig. 11. In each

of the 3 random reorderings, we find that the large gap that appeared in Fig. 3 between the recycled (and HF) minimum and the local traps, largely disappears when considering a randomly shuffled ansatz. This is clear ev-





FIG. 10. Results of various initializations for the 60parameter ADAPT-VQE ansatz for  $H_6$  at 3Å. 300 random initializations are shown on the left. 1000 random initializations are depicted on the right.

idence that it is not sufficient to only slowly optimize the parameters one-at-a-time via recycled initialization, but that the ansatz must be grown in an efficient manner as well if one is to observe the full burrowing effect.

#### Appendix E: Generalized Singles and Doubles Pool Simulations

When the generalized singles and doubles (i.e., those that destroy the reference state) were included in the pool, the results were basically the same for H<sub>6</sub> at 1-3Å. We include them here for completeness in Fig. 12.

#### Appendix F: Infidelity Versus Number of Parameters

Throughout the manuscript, we have included plots of energy error vs the number of ansatz parameters (or ADAPT-VQE iteration). While this is generally the target metric used for chemistry purposes, one might also be interested in understanding the convergence of the the fidelity, or overlap with the ground state, as a function of ansatz parameters, when being used as a input state generator for quantum phase estimation. In Fig. 13, we plot the infidelity, defined as  $1 - |\langle \psi^{\text{Exact}} | \psi^{\text{ADAPT}} \rangle|^2$ . Here we more directly see that the ADAPT-VQE convergence becomes more difficult as the gap decreases.



FIG. 11. 3 Random reorderings of the first 71 operators identified in an ADAPT-VQE simulation on  $H_6$  at 1 Å. (Computer time was exhausted before all 71 operators were added in this case.) The x-axis corresponds to the number of operators in the ansatz at a given step. The y-axis corresponds to the error from the exact FCI energy. The red curve (HF) corresponds to the energy obtained through BFGS minimization using an all-zero initialization. The green curve corresponds to the energy obtained through BFGS minimization using the standard ADAPT-VQE recycled initialization. The colored dashes correspond to all the energies obtained through BFGS optimizations, with red being the highest energy and violet the lowest. 300 random initializations were performed for each ansatz.



FIG. 12. GSD Pool. ADAPT-VQE results for  $H_6$  at 1, 2, and 3 Å with the generalized singles and doubles (GSD) pool. The *x*-axis corresponds to the number of ADAPT-VQE iterations, i.e. the number of operators in the ansatz at a given step. The *y*-axis corresponds to the error from the exact FCI energy. The red curve corresponds to the energy obtained through BFGS minimization using an HF guess, i.e. one where all parameters are zero. The green curve corresponds to the energy obtained through BFGS minimization using the standard ADAPT-VQE in which optimal parameter values in one iteration are recycled as initial guesses in the next iteration, and with the new parameter initialized to zero. The colored dots correspond to all the energies obtained through BFGS optimizations, with red being the highest energy and violet the lowest.



FIG. 13. Infidelity vs parameter count. ADAPT-VQE results for  $H_6$  at 1, 2, and 3 Å with the singles and doubles (SD) pool. The x-axis corresponds to the number of ADAPT-VQE iterations, i.e. the number of operators in the ansatz at a given step. The y-axis corresponds to the infidelity. The red curve corresponds to the infidelity obtained through BFGS minimization using an HF guess, i.e. one where all parameters are zero. The green curve corresponds to the infidelity obtained through BFGS minimization using the standard ADAPT-VQE in which optimal parameter values in one iteration are recycled as initial guesses in the next iteration, and with the new parameter initialized to zero. The colored dots correspond to all the fidelities obtained through BFGS optimizations of the energy, with red being the highest energy and violet the lowest. Note that the colored dots are not generally in chromatic order, as a wavefunction with a better energy can have a lower fidelity.

## Chapter 5

## Introduction to Classically Tractable Approximations

### 5.1 Classical Approximations to UCC

As discussed in the previous section, unitary coupled cluster (UCC) is intractable on a quantum computer when  $\hat{T}$  is truncated to any excitation rank higher than singles. [17] UCC with singles only, UCCS, is equivalent to the SCF procedure when performed on any Slater determinant reference. One can expand the UCC energy as an infinite BCH series, as in equation 5.1.

$$\langle 0|e^{\hat{T}^{\dagger}-\hat{T}}\hat{H}e^{\hat{T}-\hat{T}^{\dagger}}|0\rangle = \langle 0|\hat{H}|0\rangle + \langle 0|\left[\hat{H},\hat{T}-\hat{T}^{\dagger}\right]|0\rangle + \frac{1}{2!}\langle 0|\left[\left[\hat{H},\hat{T}-\hat{T}^{\dagger}\right],\hat{T}-\hat{T}^{\dagger}\right]|0\rangle + \dots$$
(5.1)

This expansion can be truncated to yield an energy functional which is tractable to evaluate on a classical computer. Unfortunately, truncating the method will destroy the variational character (in the sense that the energy functional is no longer bounded below by the groundstate energy) of the functional, and can destroy the size-consistency of UCC depending on the truncation scheme.

The two main truncation schemes are as follows:

- 1. Perturbative truncation of the BCH series based on the full  $\hat{T}$ , which is not truncated by rank [29]
- 2. Truncation of the BCH series based on power of  $\hat{T}$  for a fixed excitation rank of  $\hat{T}$  [30]

We restrict our focus to the latter strategy, and will only concern ourselves with the case where  $\hat{T}$  is restricted to the singles and doubles, i.e. truncated UCCSD. As Kutzelnigg has computed [30], the terms introduced by the first three orders of the expansion are given as:

$$E^{(0)} = \langle 0|\hat{H}|0\rangle \tag{5.2}$$

$$E^{(1)} = 2Re \left\langle 0|\hat{H}\hat{T}|0\right\rangle \tag{5.3}$$

$$E^{(2)} = Re \left\langle 0 | \hat{H}\hat{T}^2 + \hat{T}^{\dagger}\hat{H}\hat{T} - \hat{T}^{\dagger}\hat{T}\hat{H} | 0 \right\rangle$$
(5.4)

Summing  $E^{(0)}$ ,  $E^{(1)}$ , and  $E^{(2)}$  gives an energy functional which is quadratic in  $\hat{T}$ . We call the procedure of minimizing this functional the Order-2 (O2-) UCCSD algorithm. The choice of notation will be explained in chapter 6, where we expand on our contributions to this type of method. When one differentiates these equations, they obtain a set of stationary conditions [28], given as:

$$0 = \langle \phi_i^a | \hat{H} | 0 \rangle + \langle \phi_i^a | \hat{H}_N \hat{T}_1 | 0 \rangle + \langle \phi_i^a | \hat{T}_1^\dagger \hat{V}_N | 0 \rangle + \frac{1}{2} \langle \phi_i^a | \hat{F}_N \hat{T}_2 | 0 \rangle$$
(5.5)

$$0 = \langle \phi_{ij}^{ab} | \hat{H} | 0 \rangle + \langle \phi_{ij}^{ab} | \hat{H} \hat{T}_2 | 0 \rangle + \frac{1}{2} \langle \phi_{ij}^{ab} | \hat{F}_N \hat{T}_1 | 0 \rangle$$
(5.6)

We have assumed that the wavefunction is real, and use the standard definitions of  $\hat{F}_N$  and  $\hat{V}_N$  as components of the Hamiltonian. [31] When these stationary conditions are met, the O2-UCCSD energy (the sum of  $E^{(0)}$ ,  $E^{(1)}$ , and  $E^{(2)}$ ) becomes

$$E = E_0 + \langle 0 | \hat{H}_N \hat{T} | 0 \rangle \tag{5.7}$$

We will see in the next section that these equations are very similar to those of another approach. As a final note, one can compute the second-order truncation of a Trotterized UCCSD energy functional. [32] In chapter 6 we consider the energy functionals that result from certain Trotter orderings and how they lead to new properties in O2-UCCSD and in our extensions of the approach.

### 5.2 Linearized Coupled Cluster Theory

In the previous section, we introduced the O2-UCCSD method. In this section we will introduce an older method: linearized coupled cluster theory. When truncated to the doubles, the method is interchangably referred to as linearized coupled cluster with doubles, LCCD, and doubly-excited many-body perturbation theory to infinite order. (D-MBPT( $\infty$ )). When truncated to the singles and doubles, the method is referred to as LCCSD. The coupled electron pair approximation, CEPA, is sometimes used to denote both LCCD and LCCSD, so we will refrain from using the term to avoid ambiguity. The different names for the same method come from various approaches to obtaining the same equations. This type of formalism was originally proposed by Wilfried Meyer in 1972. [33]

To understand linearized coupled cluster theory, let us first consider traditional (projective) coupled cluster theory. The Schrödinger equation tells us that

$$\hat{H} |\Psi\rangle = E |\Psi\rangle,$$
 (5.8)

where  $|\Psi\rangle$  is the exact wavefunction of the system, and E is the ground state energy. We

#### 5.2. LINEARIZED COUPLED CLUSTER THEORY

can introduce an approximate wavefunction defined as

$$|\Psi\rangle = e^{\hat{T}} |0\rangle. \tag{5.9}$$

If  $\hat{T}$  includes *N*-electron excitations for an *N*-electron problem, this is actually an exact parametrization of the wavefunction. For our purposes, we will truncate  $\hat{T}$  to  $\hat{T}_1 + \hat{T}_2$ , the single and double excitation operators, giving coupled cluster with singles and doubles, or CCSD. Substituting equation 5.9 into equation 5.8 gives us an approximation to the true Schrödinger equation:

$$\hat{H}e^{\hat{T}}\left|0\right\rangle = Ee^{\hat{T}}\left|0\right\rangle \tag{5.10}$$

We can multiply each side of equation 5.10 by  $e^{-\hat{T}}$  and project onto the reference determinant and each singly or doubly excited determinant:

$$E = \langle 0|e^{-T}\hat{H}e^{T}|0\rangle \tag{5.11}$$

$$0 = \langle \phi_i^a | e^{-\hat{T}} \hat{H} e^{\hat{T}} | 0 \rangle \tag{5.12}$$

$$0 = \langle \phi_{ij}^{ab} | e^{-\hat{T}} \hat{H} e^{\hat{T}} | 0 \rangle \tag{5.13}$$

In CCSD, equations 5.12 and 5.13 are solved to determine the t-amplitudes that inform the energy in equation 5.11. In LCCSD, we introduce the approximation that  $e^{-\hat{T}}\hat{H}e^{\hat{T}}$  can be approximated by ignoring terms which are quadratic in  $\hat{T}$  or higher, i.e., we assume that

$$e^{-\hat{T}}\hat{H}e^{\hat{T}} \approx \hat{H} + \left[\hat{H}, \hat{T}\right].$$
(5.14)
Using this approximation in equations 5.11-5.13 gives the LCCSD equations [34], given as:

$$E = \langle 0|\hat{H}|0\rangle + \langle 0|\hat{H}\hat{T}|0\rangle \tag{5.15}$$

$$0 = \langle \phi_i^a | \hat{H} | 0 \rangle + \langle \phi_i^a | \hat{H} \hat{T} | 0 \rangle$$
(5.16)

$$0 = \langle \phi_{ij}^{ab} | \hat{H} | 0 \rangle + \langle \phi_{ij}^{ab} | \left[ \hat{H}, \hat{T} \right] | 0 \rangle$$
(5.17)

These equations are very similar, but not identical, to those of O2-UCCSD. Most notably, the LCCSD equations contain only connected terms where  $\hat{H}$  contracts with  $\hat{T}$ . This is sufficient to make LCCSD size-extensive, while O2-UCCSD is not. [31] The similarities and differences of these methods are interesting, and are discussed further in chapter 6.

# Chapter 6

# The O2DX Framework

### New Local Explorations of the Unitary Coupled Cluster Energy Landscape

Harper R. Grimsley\* and Nicholas J. Mayhall

Cite This: https://doi.org/10.1021/acs.jctc.2c00751



Article Recommendations

ACCESS

III Metrics & More

**ABSTRACT:** The recent quantum information boom has effected a resurgence of interest in the unitary coupled cluster (UCC) theory. Our group's interest in local energy landscapes of unitary ansätze prompted us to investigate the *classical* approach of truncating the Taylor series expansion (instead of a perturbative expansion) of the UCC with singles and doubles (UCCSD) energy at the second order. This amounts to an approach where the electron correlation energy is estimated by taking a single Newton–Raphson step from Hartree–Fock toward UCCSD. Such an approach has been explored previously, but the accuracy was not extensively studied. In this paper, we investigate the performance and observe similar pathologies to the linearized coupled cluster with singles and doubles. We introduce the use of derivatives of order three or greater to help partially recover the variational lower bound of true UCCSD, restricting these derivatives to those of the "unmixed" category in order to simplify the model. By testing the approach on several potential energy surfaces and reaction energies, we find



this "diagonal" approximation to higher order terms to be effective at reducing sensitivity near singularities for strongly correlated regimes, while not significantly diminishing the accuracy of weakly correlated systems.

#### INTRODUCTION

The unitary coupled cluster (UCC) has intrigued chemists for the past 4 decades as a variational form of the already powerful coupled cluster theory.<sup>1-6</sup> It has garnered significant attention in recent years due to its use in variational quantum eigensolvers (VQEs) and hybrid quantum-classical algorithms.<sup>7-9</sup> Quantum gates must correspond to physical evolutions, making the unitary parameterization of the wavefunction a natural choice. Various approximations and modifications of UCC have seen use, including more general ansätze based on UCC-like generators.<sup>10–16</sup> A comprehensive review of this topic has recently been prepared by Anand et al.<sup>17</sup> Dynamical ansätze such as the adaptive, derivativeassembled, pseudo-Trotterized VQE (ADAPT-VQE) method<sup>18,19</sup> build a sequential, UCC-like ansatz one anti-Hermitian operator at a time, the order of which is determined by the energy gradient associated with that operator's introduction to the ansatz. ADAPT-VQE gives a quasi-optimal operator ordering that is informed by the problem Hamiltonian but only includes information about the first energy derivative with respect to the energy parameters.

While the UCC has provided a useful framework for defining state preparation circuits in the VQE, as a classical approach, it also has several desirable attributes:

- 1. Variationality
- 2. Size extensivity<sup>20</sup>
- 3. Satisfaction of the generalized Hellmann–Feynman theorem<sup>2</sup>

Regardless of its various attractive features, computing the UCC energy requires evaluating an infinite series of commutators, even if the excitation rank of  $\hat{T}$  is restricted to singles and doubles (UCCSD). The UCCSD energy landscape, while complicated, is a smooth functional of the UCCSD wavefunction t-amplitudes. If we assume that we are sufficiently close to the energy minimum that the landscape is locally convex and that a finite solution exists which satisfies the stationary conditions, we can make a reasonable guess at the optimal values of the amplitudes by using a quadratic Taylor approximation to the energy functional in the tamplitudes. In general, we would need a higher order Taylor expansion to fully understand the landscape and recover the nice properties of the UCC. One way we can go beyond the quadratic approximation without much difficulty is to partially explore the cubic and higher characters of the landscape. Rather than including all higher order derivative tensor terms,

we consider only diagonal partial derivatives, for example,  $\frac{\partial^{k_{c}}}{(\partial t_{ii}^{ab})^{k}}$ 

beyond the second order. Such an approximation is based on the assumption that high order mixed derivatives are not

Received: July 21, 2022

important, a purely geometric idea, rather than one based on, for example, perturbation theory. We will explore this idea more concretely in the Theory section.

#### THEORY

We will begin this section by introducing the full UCCSD functional and proceed to its Taylor series truncation at the second order in the *t*-amplitudes. From there, we will introduce our "diagonal" correction strategy and give explicit equations for the third- and infinite-order situations. We will finally compare the third-order case to existing coupled electron pair approximation (CEPA) methods.

The UCC ansatz is given by eq 1.<sup>1</sup>

$$|\Psi_{\rm UCC}\rangle = e^{(\hat{T} - \hat{T}')} |\phi_0\rangle \tag{1}$$

where  $\hat{T}$  is defined in the usual way, as in eq 2.

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots = \sum_{ia} t_a^i \hat{a}_i^a + \sum_{\substack{i < j \\ a < b}} t_a^{ij} \hat{a}_{ij}^{ab} + \dots$$
(2)

We will define  $|\phi_0\rangle$  to be a normalized, single-determinant reference. We do not initially assume a Brillouin condition and will explore the consequences of the orbital choice later in the text. Because  $\hat{T} - \hat{T}^{\dagger}$  is anti-Hermitian, its exponential is unitary, and the quantity  $\epsilon_{\text{UCC}}[\mathbf{t}]$  in eq 3 is bounded below by the lowest eigenvalue of  $\hat{H}$  for any value of  $\mathbf{t}$ 

$$\epsilon_{\text{UCC}}[\mathbf{t}] = \langle \Psi_{\text{UCC}} | \hat{H} | \Psi_{\text{UCC}} \rangle \tag{3}$$

where t is the vector of t amplitudes. While  $\epsilon_{\rm UCC}$  is variational, symmetric, and size-extensive and satisfies the generalized Hellmann–Feynman theorem, it is classically intractable to evaluate. Unlike the traditional (non-unitary) coupled cluster energy, the Baker–Campbell–Hausdorff (BCH) expansion of  $\epsilon_{\rm UCC}$  will never terminate, regardless of the truncation of  $\hat{T}$ . Various artificial truncation schemes for  $\epsilon_{\rm UCC}$  exist, including truncation based on the perturbation order<sup>2</sup> and truncation based on the commutator order.<sup>1</sup> We will restrict our focus to the latter, which has a geometric interpretation and has been examined in far less detail. To denote the order (*n*) of Taylor series trunctation, we prepend an "On-" to the typical UCCSD. For example, restricting  $\hat{T}$  to singles and doubles and terminating the Taylor series after the second order, we obtain eq 4, which in this notation is given as O2-UCCSD.

$$\begin{split} \epsilon_{\text{O2-UCCSD}} &= E_0 + \sum_{ia} \left. \frac{\partial \epsilon_{\text{UCCSD}}}{\partial t_a^i} \right|_{\mathbf{0}} t_a^i + \sum_{\substack{i < j \\ a < b}} \left. \frac{\partial \epsilon_{\text{UCCSD}}}{\partial t_{ab}^{ij}} \right|_{\mathbf{0}} t_{ab}^{ij} \\ &+ \frac{1}{2} \sum_{ijab} \left. \frac{\partial^2 \epsilon_{\text{UCCSD}}}{\partial t_a^i \partial t_b^j} \right|_{\mathbf{0}} t_a^i t_b^j + \frac{1}{2} \sum_{\substack{i < j, k < l \\ a < b, c < d}} \left. \frac{\partial^2 \epsilon_{\text{UCCSD}}}{\partial t_{ab}^{ij} \partial t_{cd}^k} \right|_{\mathbf{0}} t_{ab}^{ij} t_c^{kl} \\ &+ \sum_{\substack{i < j \\ a < b, c < d}} \left. \frac{\partial^2 \epsilon_{\text{UCCSD}}}{\partial t_{ab}^{ij} \partial t_c^k} \right|_{\mathbf{0}} t_{ab}^{ij} t_c^k \\ &= E_0 + 2 \langle \phi_0 | \hat{H}_N \hat{T} | \phi_0 \rangle + \langle \phi_0 | \hat{T}^\dagger \hat{H}_N \hat{T} | \phi_0 \rangle \\ &+ \langle \phi_0 | \hat{H}_N \hat{T}_1^2 | \phi_0 \rangle - \langle \phi_0 | \hat{H}_N \hat{T}_1^\dagger \hat{T}_2 | \phi_0 \rangle \end{split}$$
(4)

(We use the standard definitions of  $\hat{F}_N$ ,  $\hat{V}_N$ , and  $\hat{H}_N$ , as used in the coupled cluster review of Crawford and Schaefer<sup>21</sup>.) Apart from the last two terms,  $\epsilon_{O2-UCCSD}$  is simply the linearized coupled cluster with singles and doubles (LCCSD) Lagrangian.<sup>22</sup> Similarly,  $\epsilon_{O2-UCCD}$  is precisely the LCCD Lagrangian. Minimizing  $\epsilon_{O2-UCCSD}$  amounts to solving a linear set of stationary equations, given by eqs 5 and 6.

$$\left\langle \phi_{i}^{a} \middle| \hat{H}_{N} \hat{T}_{1} + \hat{T}_{1}^{\dagger} \hat{V}_{N} + \frac{1}{2} \hat{F}_{N} \hat{T}_{2} \middle| \phi_{0} \right\rangle = -f_{i}^{a}$$
(5)

$$\left\langle \phi_{ij}^{ab} \middle| \hat{H}_{N} \hat{T}_{2} + \frac{1}{2} \hat{F}_{N} \hat{T}_{1} \middle| \phi_{0} \right\rangle = -\langle ij || ab \rangle \tag{6}$$

For non-Hartree–Fock orbitals, the lack of a Brillouin condition leads to a breakdown of size extensivity when differentiating the  $t_{ij}^{ab}t_i^{a}f_{jb}$  term with respect to the doubles amplitudes, resulting in a disconnected contribution.<sup>23</sup> This term ultimately persists due to an incomplete cancellation between diagrams in the  $\langle \phi_0 | \hat{T}^{\dagger} \hat{H}_N \hat{T} | \phi_0 \rangle$  and  $\langle \phi_0 | \hat{H}_N \hat{T}_1^{\dagger} \hat{T}_2 | \phi_0 \rangle$  terms. While this might not be expected to deteriorate performance too significantly, as the magnitude of the size inextensivity is determined only by the occupied-virtual block of the Fock matrix (i.e., the distance from an optimal set of orbitals), it is, in fact, possible to reformulate the problem slightly to recover the exact size extensivity, for both HF and non-HF orbitals. If we, instead, begin from the partially Trotterized (excitation rank-separated) energy in eq 7

$$\epsilon_{\rm tUCCSD} = \langle \phi_0 | e^{-\hat{k}_2} e^{-\hat{k}_1} \hat{H} e^{\hat{k}_1} e^{\hat{k}_2} | \phi_0 \rangle \tag{7}$$

where  $\hat{K}_i = \hat{T}_i - \hat{T}_i^{\dagger}$ , we will obtain the following second-order Taylor series approximation

$$\begin{aligned} \epsilon_{tO2-UCCSD} &= E_0 + 2\langle \phi_0 | \hat{H}_N \hat{T} | \phi_0 \rangle + \langle \phi_0 | \hat{T}^{\dagger} \hat{H}_N \hat{T} | \phi_0 \rangle \\ &+ \langle \phi_0 | \hat{H}_N \hat{T}_1^{\dagger} | \phi_0 \rangle - 2\langle \phi_0 | \hat{H}_N \hat{T}_1^{\dagger} \hat{T}_2 | \phi_0 \rangle \end{aligned} \tag{8}$$

This additional factor of 2 provides full cancellation of the disconnected terms in the gradient expression.

This is interesting from three perspectives: (i) any singledeterminant reference state leads to a size-extensive method that becomes equivalent to the first approach when HF orbitals are used, (ii) the disentangled form suggests the opportunity to develop a proper exact singles approach since the unitary formalism allows implementation by a simple orbital rotation, which will be considered in follow-up work, and (iii) a doubles-then-singles ordering of excitations is consistent with what we find to be accurate both in our ADAPT-VQE algorithm and with our previous direct study on Trotter ordering.<sup>16</sup> The resulting stationary conditions are given as

$$\langle \phi_i^a | \hat{H}_N \hat{T}_1 + \hat{T}_1^{\mathsf{T}} \hat{V}_N | \phi_0 \rangle = -f_i^a \tag{9}$$

$$\langle \phi_{ij}^{ab} | \hat{H}_{\rm N} \hat{T}_2 | \phi_0 \rangle = -\langle ij | | ab \rangle \tag{10}$$

At the time of writing, another group pointed out that one can perform a Newton step toward tUCCSD instead of UCCSD.<sup>24</sup> However, they do not make any argument for a specific Trotter ordering or take interest in the size extensivity of the solution.

One key advantage of the Taylor-truncated UCCSD approaches compared to CEPA approaches is that Taylortruncated approaches are systematically improvable. Inclusion of triple or higher excitations is obvious, if expensive, and we can work toward recovering the variational character of UCCSD by including higher ordered terms in the Taylor



Figure 1. Dissociation of HF in the  $6-31G^{**^{37-39}}$  basis.

series. We note that an analogous systematic improvability also exists within the perturbation-trunctated UCC schemes.

O2-UCCSD has another issue, unrelated to size extensivity. Linearized CC approximations have long been known to blow up due to singularities stemming from quasi-degeneracies.<sup>25</sup> This behavior is observed even in multireference formulations, though inclusion of non-linear terms avoids this issue.<sup>26</sup> As we will report, the same problem plagues O2-UCCSD. One of the most successful strategies for avoiding singularities historically has been the split-amplitude "almost linear CC" family of approaches, where some of the t-amplitudes are described by a large fixed part and a small variable component.27,28 The aforementioned singularity problem can alternatively be addressed by expanding  $\epsilon_{\mathrm{UCCSD}}$  to the third order in the tamplitudes, but such an expansion would significantly increase the complexity of the resulting expressions (albeit with no net increase in the asymptotic scaling). However, if we assume that the third derivative tensor is diagonally dominant, we can instead approximate these higher order terms with essentially no additional cost by including only the diagonal (or *unmixed*) third derivatives in the Taylor series, as in eq 11.

$$\begin{aligned} \epsilon_{\text{O2D3-UCCSD}} &= \\ \epsilon_{\text{O2-UCCSD}} + \frac{1}{6} \sum_{ia} \left. \frac{\partial^3 \epsilon_{\text{UCCSD}}}{(\partial t_a^{i_3})^3} \right|_{\mathbf{0}} (t_a^{i_j})^3 + \frac{1}{6} \sum_{\substack{i < j \\ a < b}} \left. \frac{\partial^3 \epsilon_{\text{UCCSD}}}{(\partial t_{ab}^{i_j})^3} \right|_{\mathbf{0}} (t_{ab}^{ij})^3 \\ &= \epsilon_{\text{O2-UCCSD}} - \frac{4}{3} \sum_{ia} f_i^a (t_a^i)^3 - \frac{4}{3} \sum_{\substack{i < j \\ a < b}} \langle ij || ab \rangle (t_{ab}^{ij})^3 \end{aligned}$$
(11)

Differentiating eq 11 introduces non-linear but diagonal terms. In practice, we minimize  $\epsilon_{\rm O2D3-UCCSD}$  directly to find the O2D3-UCCSD energy, but this could also be achieved by solving a series of "shifted" linear equations if desired. In general, including diagonal derivatives of higher order will only introduce new four-index contractions, which are negligible in an  $O(N^6)$  algorithm.

We will advocate, as an improved approach, the inclusion of diagonal derivatives to *infinite* order,  $O2D\infty$ -UCCSD. This energy is given in eq 12. The Einstein notation is used for the orbital indices to improve readability.

$$\begin{split} &\varepsilon_{\text{O2D}\infty-\text{UCCSD}} = \\ &\varepsilon_{\text{O2-UCCSD}} + \sum_{k=3}^{\infty} \frac{1}{k!} \Biggl( \frac{\partial^k \varepsilon_{\text{UCCSD}}}{(\partial t_a^{ij})^k} \Biggl|_{\mathbf{0}} (t_a^i)^k + \frac{1}{4} \frac{\partial^k \varepsilon_{\text{UCCSD}}}{(\partial t_{ab}^{ij})^k} \Biggl|_{\mathbf{0}} (t_{ab}^{ij})^k \Biggr) \\ &= E_0 + \frac{1}{2} \langle \phi_0 | [[\hat{H}, \hat{K}], \hat{K}] | \phi_0 \rangle + \sin(2t_a^i) f_i^a + \frac{1}{4} \sin(2t_{ab}^{ij}) \langle ij || ab \rangle \\ &+ (\sin^2(t_a^i) - (t_a^i)^2) \langle \phi_i^a | \hat{H}_N | \phi_i^a \rangle + \frac{1}{4} (\sin^2(t_{ab}^{ij}) - (t_{ab}^{ij})^2) \langle \phi_{ij}^a | \hat{H}_N | \phi_{ij}^{ab} \rangle \end{split}$$

$$(12)$$

A derivation of eq 12 is given in Appendix A. Our diagonal energies bear a natural resemblance to the third- and infiniteorder two-electron UCC energies computed by Kutzelnigg.<sup>3</sup> We note in passing that the diagonal corrections to the O2-UCCSD energy functional give a method that is no longer invariant to occupied–occupied and virtual–virtual orbital rotations.

To recapitulate, the O2(D2)-UCCSD method was previously described by Kutzelnigg<sup>1</sup> and has already been implemented for multiple reference determinants by Simons and Hoffmann.<sup>29</sup> This method corresponds to truncating the UCCSD functional at the second order in the *t*-amplitudes and then minimizing the truncated functional. Our O2D3- and O2D $\infty$ -UCCSD variants include the unmixed or "diagonal" derivatives to the third and infinite order, respectively (e.g., O2D3-UCCSD approximates the third derivative "jerk" tensor matrix by its diagonal). The deletion of extensivity-violating terms is not new, but our rationalization based on ansatz Trotterization is, and we denote the use of these deletions as O2DX-tUCCSD.

As a final note on the theory involved in these methods, we point out that the O2D3-UCCSD functional gives similar amplitude equations to performing a conventional CEPA derivation and treating only the exclusion principle-violating (EPV) terms in which *every* index, occupied and virtual, is exclusion principle-violating. (The authors recommend the review<sup>30</sup> by Wennmohs and Neese of the EPV-based CEPA derivations. See also Appendix B. which follows in their footsteps.) Our methods might be viewed as maximally simple approaches that still give unique, determinant-tailored shifts to individual excitations. This characterization is consistent with

pubs.acs.org/JCTC

Article



Figure 2. Dissociation of a single C–H bond in methane in the 6-31G\* basis.



Figure 3. Rotation of one  $CH_2$  group about the C–C bonding axis in ethylene in the CC-pVDZ<sup>45</sup> basis. The LCCSD and O2D2-UCCSD curves are essentially overlapping. CCSDT apparently converged to an excited state at 90°, so the point was omitted.

their improved capacity to break single bonds relative to LCCSD.<sup>31</sup> Our method might be compared to a minor complication of LCCSD,<sup>32</sup> a minor simplification of CEPA(2)/CEPA(3),<sup>33</sup> or a dramatic simplification of SC<sup>2</sup>CISD.<sup>34</sup>

#### RESULTS AND DISCUSSION

We consider three potential energy surfaces used by Malrieu et al.<sup>31</sup> to demonstrate the single bond-breaking ability of CEPA(3):

- 1. Dissociation of hydrogen fluoride
- 2. Dissociation of a single C-H bond in methane
- 3. Torsion of ethylene

We used optimized geometries from  $B3LYP^{35}/6-31G^{*36-39}$  calculations to determine the positions of atoms that were held

static in each curve. All O2-UCCSD energies, as well as LCCSD energies, were computed using a custom software package developed in-house, available at https://github.com/ hrgrimsl/taylor\_ucc. SCF, CCSD, CCSD(T), density functional theory (DFT), and FCI calculations were performed using PySCF.<sup>40</sup> CCSDT energies were obtained using the MRCC code of Kállay et al.<sup>41-43</sup> via a Psi4 interface.<sup>44</sup>

The divergence of LCCSD and O2D2-UCCSD in Figures 1-3 demonstrates the failure of the linear methods to break single bonds. In general, some excitation-specific correction of a higher order in t is required to treat this type of problem. For example, ACPF also fails to dissociate these molecules, despite having diagonal corrections of its own, since it still corrects every excitation uniformly.<sup>31</sup>

For HF dissociation (Figure 1), the O2D3-UCCSD approach gives a similar qualitative behavior to CCSD(T) if canonical orbitals are used. Canonical orbital O2D $\infty$ -UCCSD diverges for HF. We explain the worse performance of the infinite-order correction here by the strong orbital dependence of diagonal methods. When Kohn–Sham orbitals are used, the unphysical "hump" is eliminated from O2D3, and O2D $\infty$  becomes quite accurate for this dissociation.

The CH<sub>4</sub> dissociation is largely similar to that of HF, with two notable exceptions. First,  $O2D\infty$  does not completely diverge with canonical orbitals. Second, the KS orbitals fail to eliminate the unphysical "hump" at 2.5 Å entirely, with O2D3-UCCSD giving quantitatively better energies at dissociation. We suspect that this would not be the case with better orbitals.

In the case of ethylene torsion (Figure 3), LCCSD and O2D2-UCCSD's unphysical divergences are never fully corrected, though the O2D3- and O2D $\infty$ -UCCSD methods give a clear improvement. The KS orbitals appear considerably less helpful for this system, introducing very little difference to the diagonally corrected methods beyond easier numerical convergence of the algorithm. The difference between Trotterized and un-Trotterized methods is small for all three systems, as expected based on the full-order size extensivity of UCCSD.

The inconsistent utility of DFT orbitals motivates investigation into an optimal orbital choice for the O2D3and O2D $\infty$ -UCCSD methods. The matter of orbital "optimization" is complicated by the fact that divergence to –  $\infty$  is possible for some orbital choices, as seen in Figures 1 and 3. Minimizing the norm of t or some similar scheme might be appropriate, but we defer such questions to future work. This idea that using orbital rotations to avoid large tamplitudes is a viable strategy is somewhat corroborated by the relative performance of canonical and Kohn–Sham orbitals. For example, in the case of hydrogen fluoride dissociation, O2D $\infty$ -UCCSD diverges with canonical orbitals but not with Kohn–Sham orbitals. This implies that for the Kohn–Sham orbitals, the optimal amplitudes are finite, which is not the case with canonical orbitals.

The failure of our methods to fully deal with ethylene torsion suggests that we do lose some of the applicability of CEPA(3) as a single-reference method.<sup>31</sup> The primary difference between ethylene torsion and our single bondbreaking tests is that in ethylene, both the 2 – HOMO and 2 – LUMO begin to become degenerate as well as the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO). (As Malrieu et al. point out,<sup>31</sup> this is because it is a  $\pi/\pi^*$  orbital pair becoming degenerate.) While our methods include high order interactions between individual excited states and the reference, they neglect high order coupling between different excited states. Consequently, we expect our method to break down in situations where there are multiple coupled, excited determinants that are important.

The O2-UCCSD and O2-tUCCSD methods are extremely similar in their performance. This can be explained by the fact that UCCSD is fully connected at the full order<sup>23</sup> so that the affected terms occur at the third order or higher in the Taylor series expansion. A similar argument has been used previously to justify the manual deletion of similar types of "internally disconnected" terms from perturbatively truncated CC functional approaches.<sup>23</sup> Furthermore, in most situations, O2-UCCSD and O2-tUCCSD differ very little from LCCSD. All three methods involve solving similar sets of linear systems of equations and have similar pathologies involving singularities in those equations.<sup>4</sup>

As a broader test of applicability, we considered the CRE-31 reaction energy test set of Soydaş and Bozkaya,<sup>46</sup> motivated by its use for characterizing orbital-optimized LCCD (this test set is enumerated in Table 1). We elected to use the CC-pVTZ<sup>45,47</sup> basis for all systems, with geometries obtained from B3LYP/6-31G\*\* optimization in PySCF.

#### Table 1. Reaction Key for CRE-31

1)	$F_2O + H_2 \rightarrow F_2 + H_2O$
2)	$H_2O_2 + H_2 \rightarrow 2H_2O$
3)	$\rm CO + H_2 \rightarrow CH_2O$
4)	$CO + 3H_2 \rightarrow CH_4 + H_2O$
5)	$N_2 + 3H_2 \rightarrow 2NH_3$
6)	$N_2O + H_2 \rightarrow N_2 + H_2O$
7)	$HNO_2 + 3H_2 \rightarrow 2H_2O + NH_3$
8)	$C_2H_2 + H_2 \rightarrow C_2H_4$
9)	$CH_2CO + 2H_2 \rightarrow CH_2O + CH_4$
10)	$BH_3 + 3HF \rightarrow BF_3 + 3H_2$
11)	$HCOOH \rightarrow CO_2 + H_2$
12)	$CO + H_2O \rightarrow CO_2 + H_2$
13)	$C_2H_2 + HF \rightarrow CH_2CHF$
14)	$HCN + H_2O \rightarrow CO + NH_3$
15)	$HCN + H_2O \rightarrow HCONH_2$
16)	$\text{HCONH}_2 + \text{H}_2\text{O} \rightarrow \text{HCOOH} + \text{NH}_3$
17)	$HCN + NH_3 \rightarrow N_2 + CH_4$
18)	$CO + CH_4 \rightarrow CH_3CHO$
19)	$N_2 + F_2 \rightarrow trans - N_2F_2$
20)	$N_2 + F_2 \rightarrow cis - N_2 F_2$
21)	$2BH_3 \rightarrow B_2H_6$
22)	$CH_3ONO \rightarrow CH_3NO_2$
23)	$CH_2C \rightarrow C_2H_2$
24)	allene $\rightarrow$ propyne
25)	cyclopropene $\rightarrow$ propyne
26)	oxirane $\rightarrow$ CH <sub>3</sub> CHO
27)	vinyl alcohol $\rightarrow$ CH <sub>3</sub> CHO
28)	cyclobutene $\rightarrow$ 1, 3 – butadiene
29)	$2\mathrm{NH}_3 \rightarrow (\mathrm{NH}_3)_2$
30)	$2H_2O \rightarrow (H_2O)_2$
31)	$2\text{HF} \rightarrow (\text{HF})_2$

We begin our analysis of Figures 4 and 5 by noting that for some molecules, one or more of our approaches failed to find a local minimum on which to converge. This is a weakness of using an unconstrained optimizer (L-BFGS-B)<sup>48</sup> to minimize a functional that is not actually bounded below. One could imagine a different starting guess, allowing O2D3- or O2D∞-UCCSD to converge in these cases or settling for a minimized gradient as an approximate solution condition. These three molecules had three of the four highest CCSD  $\hat{T}_1$  diagnostics in the test set. (CH<sub>3</sub>NO<sub>2</sub>, CH<sub>3</sub>ONO, and HNO<sub>2</sub> had diagnostics 0.016, 0.018, and 0.018, respectively.) For context, Lee and Taylor consider a  $\hat{T}_1$  diagnostic of 0.02 indicative of an important multireference character.49 We exclude the two associated reactions from the main text. We note in passing that one can compute a  $\hat{T}_1$  diagnostic based on the amplitudes from the O2D2-UCCSD functional and that this O2D2-UCCSD  $\hat{T}_1$  diagnostic is able to predict the multireference character that causes the method to catastrophically over-

pubs.acs.org/JCTC



Figure 4. Reaction energy errors of the CRE-31 test set in the cc-pVTZ basis, with reactions 7 and 22 excluded due to non-convergence for some methods.



Figure 5. Gaussian fittings of the CRE-31 test set in the cc-pVTZ basis, with reactions 7 and 22 excluded due to non-convergence for some methods.

estimate the correlation energy. Individual reaction energies for each method are available in the Supporting Information, along with individual molecule error as a function of the O2D2-UCCSD  $\hat{T}_1$  diagnostic.

In general, our methods do not appear to be particularly helpful for the CRE-31 test set. Regardless of the orbital choice, Trotterization, or diagonal correction, we achieve a performance comparable to that of LCCSD and better than that of CCSD. We consider this middling performance useful overall—the truncated UCC framework overcomes one of the primary issues of LCCSD, its lack of systematic improvability.<sup>30</sup> Additionally, the amenability of these methods to multireference implementation should offer a route to avoiding problems with multiple quasi-degeneracies.

#### CONCLUSIONS

We have outlined a pedagogically simple way to correct Taylor-truncated UCC functionals without worsening their formal scaling, using only unmixed derivatives. While these corrections seem to be of minimal help for computing reaction energies, they dramatically improve the behavior of singlereference O2-UCCSD for single bond-breaking events, repairing one of the most prominent pathologies of LCCSD in an extremely simple, physically motivated way. We believe that further investigation into orbital optimization can only improve our method, given its strong orbital dependence, and plan to explore this in future work.

Additionally, we offer an alternative, Trotterized ansatz that eliminates the extensivity-violating diagrams from secondorder UCCSD, noting that tUCC possesses all the qualities that made UCCSD attractive to begin with. Furthermore, the exactness of certain Trotter orderings of tUCCSD...N has been rigorously proven by Evangelista et al., while UCCSD...N may not be exact in certain pathological situations.<sup>50</sup> It is worth noting that a "doubles-then-singles" operator ordering roughly corresponding to that used in our tUCCSD approaches was shown by Evangelista et al. to not be generally capable of representing arbitrary states, even in the case of only two electrons. A potentially exact (for two electrons) "singles-thendoubles" ordering would not give a size-extensive second-order approximation.

#### APPENDIX A

#### Derivation of eq 12

Computing the infinite-order correction to O2-UCCSD is somewhat involved. We will derive a more general form of eq 12 where one has *t*-amplitudes  $t_{\mu}$  associated with operator  $\hat{O}_{\mu}$ 

$$\hat{O}_{\mu} = \hat{a}^{\dagger}_{\mu} - \hat{a}_{\mu} \tag{A1}$$

We assume only that

$$\hat{a}_{\mu}^{\dagger} | \phi_0 \rangle = | \mu \rangle \tag{A2}$$

#### Journal of Chemical Theory and Computation

$$\hat{a}_{\mu}^{\dagger}|\mu\rangle = 0 \tag{A3}$$

$$\hat{a}_{\mu}|\phi_{0}\rangle = 0 \tag{A4}$$

$$\hat{a}_{\mu}|\mu\rangle = |\phi_0\rangle \tag{A5}$$

To find the infinite-order, unmixed part of the energy, we need to find (in the Einstein notation with respect to  $\mu$ )

$$\epsilon_{D\infty} = E_0 + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k \epsilon_{\text{UCC}}}{\left(\partial t_{\mu}\right)^k} \bigg|_{\mathbf{0}} t_{\mu}^k$$
(A6)

Using the BCH definition, we know that

$$\epsilon_{\rm UCC} = E_0 + \sum_{k=1}^{\infty} \frac{1}{k!} \langle \phi_0 | [\hat{H}_{\rm N}, \hat{K}]_k | \phi_0 \rangle \tag{A7}$$

where the subscript k denotes the k<sup>th</sup> nested commutator. Consequently

$$\frac{\partial^{k} \epsilon_{\text{UCC}}}{\left(\partial t_{\mu}\right)^{k}} \bigg|_{\mathbf{0}} = \langle \phi_{0} | [\hat{H}_{\text{N}}, \hat{O}_{\mu}]_{k} | \phi_{0} \rangle$$
(A8)

(The factor of 1/k! is cancelled by the k! terms that arise when differentiating). eq A8 can be simplified into two cases, where we are taking an even or odd derivative. We first consider the case where it is even. For  $k \in \mathbf{N}$ 

$$\langle \phi_{0} | [\hat{H}_{N}, \, \hat{O}_{\mu}]_{2k} | \phi_{0} \rangle = \sum_{j=0}^{2k} \binom{2k}{j} \langle \phi_{0} | (\hat{O}_{\mu}^{\dagger})^{j} \hat{H}_{N} (\hat{O}_{\mu})^{2k-j} | \phi_{0} \rangle$$
(A9)

For terms in this summand where *j* is even, we will get some multiple of  $\langle \phi_0 | \hat{H}_N | \phi_0 \rangle = 0$ , so we can simplify eq A9 to

$$\langle \phi_{0} | [\hat{H}_{N}, \hat{O}_{\mu}]_{2k} | \phi_{0} \rangle$$

$$= \sum_{j=0}^{k-1} \binom{2k}{2j+1} \langle \phi_{0} | (\hat{O}_{\mu}^{\dagger})^{2j+1} \hat{H}_{N} (\hat{O}_{\mu})^{2k-2j-1} | \phi_{0} \rangle$$
(A10)

$$\langle \phi_{0} | [\hat{H}_{N}, \, \hat{O}_{\mu}]_{2k} | \phi_{0} \rangle = \langle \mu | \hat{H}_{N} | \mu \rangle (-1)^{k-1} \sum_{j=0}^{k-1} \binom{2k}{2j+1}$$
(A11)

$$\langle \phi_0 | [\hat{H}_N, \, \hat{O}_\mu]_{2k} | \phi_0 \rangle = \langle \mu | \hat{H}_N | \mu \rangle (-1)^{k-1} 2^{2k-1}$$
 (A12)

$$\langle \phi_0 | [\hat{H}_{\rm N}, \, \hat{O}_{\mu}]_{2k} | \phi_0 \rangle \equiv \left. \frac{\partial^{2k} \epsilon_{\rm UCC}}{(\partial t_{\mu})^{2k}} \right|_0 \tag{A13}$$

using the binomial coefficient for expanding the nested commutator. We now consider the situation where we are taking an odd derivative. For  $k \ge 0$ :

$$\langle \phi_{0} | [\hat{H}_{N}, \hat{O}_{\mu}]_{2k+1} | \phi_{0} \rangle$$

$$= \sum_{j=0}^{2k+1} \binom{2k+1}{j} \langle \phi_{0} | (\hat{O}_{\mu}^{\dagger})^{j} \hat{H}_{N} (\hat{O}_{\mu})^{2k+1-j} | \phi_{0} \rangle$$
(A14)

The terms where j is odd are the complex conjugates of those where j is even. We will assume real-valued operators and molecular orbitals, so we can simplify eq A14 to

pubs.acs.org/JCTC

1.7.1

// IF A 1

$$\langle \phi_{0} | [H_{N}, O_{\mu}]_{2k+1} | \phi_{0} \rangle = 2 \sum_{j=0}^{k} {2k+1 \choose 2j} \langle \phi_{0} | (\hat{O}_{\mu}^{\dagger})^{2j} \hat{H}_{N} (\hat{O}_{\mu})^{2k+1-2j} | \phi_{0} \rangle$$
(A15)

$$\langle \phi_{0} | [\hat{H}_{N}, \hat{O}_{\mu}]_{2k+1} | \phi_{0} \rangle = 2 \langle \phi_{0} | \hat{H}_{N} | \mu \rangle (-1)^{k} \sum_{j=0}^{k} \binom{2k+1}{2j}$$
(A16)

$$\langle \phi_0 | [\hat{H}_{\mathrm{N}}, \, \hat{O}_{\mu}]_{2k+1} | \phi_0 \rangle = \langle \phi_0 | \hat{H}_{\mathrm{N}} | \mu \rangle (-1)^k 2^{2k+1} \tag{A17}$$

$$\langle \phi_0 | [\hat{H}_{\mathrm{N}}, \, \hat{O}_{\mu}]_{2k+1} | \phi_0 \rangle \equiv \left. \frac{\partial^{2k+1} \epsilon_{\mathrm{UCC}}}{\left( \partial t_{\mu} \right)^{2k+1}} \right|_{\mathbf{0}} \tag{A18}$$

Summing over all even terms in eq A6 gives

$$\langle \mu | \hat{H}_{\rm N} | \mu \rangle \sum_{k=1}^{\infty} \frac{1}{(2k)!} t_{\mu}^{2k} (-1)^{k-1} 2^{2k-1}$$
 (A19)

$$\langle \mu | \hat{H}_{\rm N} | \mu \rangle \sum_{k=1}^{\infty} \frac{1}{(2k)!} t_{\mu}^{2k} (-1)^{k-1} 2^{2k-1}$$
$$= \frac{1}{2} \langle \mu | \hat{H}_{\rm N} | \mu \rangle \left( \frac{(2t_{\mu})^2}{2!} - \frac{(2t_{\mu})^4}{4!} + \ldots \right)$$
(A20)

$$\langle \mu | \hat{H}_{N} | \mu \rangle \sum_{k=1}^{\infty} \frac{1}{(2k)!} t_{\mu}^{2k} (-1)^{k-1} 2^{2k-1}$$
  
=  $\frac{1}{2} \langle \mu | \hat{H}_{N} | \mu \rangle (1 - \cos(2t_{\mu}))$  (A21)

$$\langle \mu | \hat{H}_{\rm N} | \mu \rangle \sum_{k=1}^{\infty} \frac{1}{(2k)!} t_{\mu}^{2k} (-1)^{k-1} 2^{2k-1} = \sin^2(t_{\mu}) \langle \mu | \hat{H}_{\rm N} | \mu \rangle$$
(A22)

Summing over all odd terms in eq A6 gives

$$\langle \phi_0 | \hat{H}_N | \mu \rangle \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} (2t_{\mu})^{2k+1} (-1)^k$$
(A23)

$$\langle \phi_0 | \hat{H}_N | \mu \rangle \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} (2t_{\mu})^{2k+1} (-1)^k$$
  
=  $\langle \phi_0 | \hat{H}_N | \mu \rangle \left( 2t_{\mu} - \frac{1}{3!} (2t_{\mu})^3 + ... \right)$  (A24)

$$\langle \phi_0 | \hat{H}_N | \mu \rangle \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} (2t_{\mu})^{2k+1} (-1)^k$$
$$= \sin(2t_{\mu}) \langle \phi_0 | \hat{H}_N | \mu \rangle$$
(A25)

Combining  $E_0$  with lines A22 and A25 gives us  $e_{D\infty}$ . However, we still want to include the mixed second derivatives in O2D $\infty$ -UCC. Consequently, we add in the term

$$\frac{1}{2}\langle\phi_0|[[\hat{H}_{\rm N},\,\hat{K}],\,\hat{K}]|\phi_0\rangle - t_{\mu}^2\langle\mu|\hat{H}_{\rm N}|\mu\rangle \tag{A26}$$

for a total O2D $\infty$  energy of

$$\begin{aligned} \epsilon &= E_0 + \frac{1}{2} \langle \phi_0 | [[\hat{H}_N, \hat{K}], \hat{K}] | \phi_0 \rangle + \sum_{\mu} \sin(2t_{\mu}) \langle \phi_0 | \hat{H}_N | \mu \rangle \\ &+ \sum_{\mu} (\sin^2(t_{\mu}) - t_{\mu}^2) \langle \phi_{\mu} | \hat{H}_N | \phi_{\mu} \rangle \end{aligned}$$
(A27)

Restricting  $\mu$  to the singles and doubles yields eq 12.

#### APPENDIX B

#### Minimal EPV CEPA

We ignore the singles and triples for simplicity in making our point. Consider the doubles equations from CIDQ...N, where we use  $\hat{T}$  for consistency with our earlier definitions

$$E_{c}t_{ab}^{ij} = \langle \phi_{ij}^{ab} | \hat{H}_{N}(1 + \hat{T}_{2} + \hat{T}_{4}) | \phi_{0} \rangle$$
(B1)

$$E_{c} = \sum_{\substack{i < j \\ a < b}} \langle ij \| ab \rangle t_{ab}^{ij}$$
(B2)

Equations B1 and B2 are simply a statement of the CIDQ...N eigenvalue problem with intermediate normalization. We can simplify eq B1 by using a CC-type approximation

$$\hat{T}_4 \approx \frac{1}{2}\hat{T}_2^2 \tag{B3}$$

This leaves us with the quadratic CI term

$$\frac{1}{2} \langle \phi_{ij}^{ab} | \hat{H}_{\rm N} \hat{T}_2^2 | \phi_0 \rangle \tag{B4}$$

Further approximating the quadruples in this equation by

$$\hat{T}_2^2 |\phi_0\rangle \approx \{\hat{T}_2, \, \hat{a}_{ij}^{ab} t_{ab}^{ij}\} |\phi_0\rangle = 2\hat{T}_2 \hat{a}_{ij}^{ab} t_{ab}^{ij} |\phi_0\rangle \tag{B5}$$

lets us simplify term B4 to

$$\langle \phi_{ij}^{ab} | \hat{H}_N \hat{T}_2 | \phi_{ij}^{ab} \rangle t_{ab}^{ij} = E_c t_{ab}^{ij} - \sum_{\substack{k < l \\ c < d} \cup ijab} \langle kl \| cd \rangle t_{cd}^{kl} t_{ab}^{ij}$$
(B6)

The  $\cup$  summation is over the EPV terms where *c*, *d*, *k*, or *l* is equivalent to *a*, *b*, *i*, or *j*. Neglecting these terms entirely gives the LCCD equations

$$0 = \langle \phi_{ij}^{ab} | \hat{H}_N(1 + \hat{T}_2) | \phi_0 \rangle \tag{B7}$$

$$E_{c} = \sum_{\substack{i < j \\ a < b}} \langle ij \| ab \rangle t_{ab}^{ij}$$
(B8)

Something similar to our method emerges if one instead makes the approximation that

$$\sum_{\substack{k < l \\ c < d} \cup ijab} \langle kl \| cd \rangle t_{cd}^{kl} t_{ab}^{ij} \approx \langle ij \| ab \rangle (t_{ab}^{ij})^2$$
(B9)

that is, we only care about one EPV term, where every single index is EPV. This gives the amplitude equation

$$-\langle ij\|ab\rangle = \langle \phi_{ij}^{ab}|\hat{H}_{N}\hat{T}_{2}|\phi_{0}\rangle - \langle ij\|ab\rangle (t_{ab}^{ij})^{2}$$
(B10)

Including a factor of 2 in front of  $\langle ij || ab \rangle (t_{ab}^{ij})^2$  would give the stationary condition of the O2D3-UCCD method. This suggests that our diagonal corrections are similar in spirit to a simplified CEPA(3), where all EPV terms with k or l equal to i or j are considered.

#### ASSOCIATED CONTENT

#### **Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00751.

Molecular geometries, individual reaction energies for the CRE-31 test set, and the plot of individual reaction energy errors as a function of the O2D2-UCCSD  $\hat{T}_1$ diagnostic (ZIP)

#### AUTHOR INFORMATION

#### Corresponding Author

Harper R. Grimsley – Chemistry Department, Virginia Tech, Blacksburg, Virginia 24061, United States; o orcid.org/ 0000-0001-5672-468X; Email: hrgrimsl@vt.edu

#### Author

Nicholas J. Mayhall – Chemistry Department, Virginia Tech, Blacksburg, Virginia 24061, United States; Occid.org/ 0000-0002-1312-9781

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00751

#### Notes

The authors declare no competing financial interest.

#### ACKNOWLEDGMENTS

The authors express their gratitude to the Department of Energy for their support. (Award no. DE-SC0019199) H.R.G. thanks the Institute of Critical Technology and Applied Science for their financial support and Ayush Asthana for useful discussions on formal scaling. The authors also wish to thank the developers of Psi4, PySCF, and MRCC for the use of their software. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within this paper. URL: https://arc.vt.edu/.

#### REFERENCES

 Kutzelnigg, W.Methods of Electronic Structure Theory, 1st ed.; Schaefer, H. F., Ed.; Modern Theoretical Chemistry; Springer Science & Business Media: New York, NY, USA, 1977; Vol. 3; pp 170–172.
 Bartlett, R. J.; Kucharski, S. A.; Noga, J. Alternative coupledcluster ansätze II. The unitary coupled-cluster method. Chem. Phys. Lett. 1989, 155, 133–140.

(3) Kutzelnigg, W. Error analysis and improvements of coupledcluster theory. *Theor. Chim. Acta* **1991**, *80*, 349–386.

(4) Taube, A. G.; Bartlett, R. J. Rethinking linearized coupled-cluster theory. *Chem. Phys.* **2009**, *130*, 144112.

(5) Harsha, G.; Shiozaki, T.; Scuseria, G. E. On the difference between variational and unitary coupled cluster theories. *Chem. Phys.* **2018**, *148*, 044107.

(6) Chen, J.; Cheng, H.-P.; Freericks, J. K. Quantum-Inspired Algorithm for the Factorized Form of Unitary Coupled Cluster Theory. J. Chem. Theory Comput. 2021, 17, 841–847.

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

(8) Hempel, C.; Maier, C.; Romero, J.; McClean, J.; Monz, T.; Shen, H.; Jurcevic, P.; Lanyon, B. P.; Love, P.; Babbush, R.; Aspuru-Guzik, A.; Blatt, R.; Roos, C. F. Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator. *Phys. Rev. X* **2018**, *8*, 031022.

pubs.acs.org/JCTC

(9) Nam, Y.; et al. Ground-state energy estimation of the water molecule on a trapped-ion quantum computer. *Npj Quantum Inf.* **2020**, *6*, 33.

(10) Hatano, N.; Suzuki, M.Finding Exponential Product Formulas of Higher Orders. arXiv:math-ph/0506007, **2005**, 679, 37–68, arXiv: math-ph/0506007. DOI: 10.1007/11526216 2

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

(12) Romero, J.; Babbush, R.; McClean, J. R.; Hempel, C.; Love, P. J.; Aspuru-Guzik, A. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Sci. Technol.* **2018**, *4*, 014008.

(13) Ryabinkin, I. G.; Yen, T.-C.; Genin, S. N.; Izmaylov, A. F. Qubit Coupled Cluster Method: A Systematic Approach to Quantum Chemistry on a Quantum Computer. J. Chem. Theory Comput. 2018, 14, 6317–6326.

(14) Lee, J.; Huggins, W. J.; Head-Gordon, M.; Whaley, K. B. Generalized Unitary Coupled Cluster Wave functions for Quantum Computation. J. Chem. Theory Comput. **2019**, 15, 311–324.

(15) Ryabinkin, I. G.; Lang, R. A.; Genin, S. N.; Izmaylov, A. F. Iterative Qubit Coupled Cluster Approach with Efficient Screening of Generators. J. Chem. Theory Comput. **2020**, 16, 1055–1063.

(16) Grimsley, H. R.; Claudino, D.; Economou, S. E.; Barnes, E.; Mayhall, N. J. Is the Trotterized UCCSD Ansatz Chemically Well-Defined? *J. Chem. Theory Comput.* **2020**, *16*, 1–6.

(17) Anand, A.; Schleich, P.; Alperin-Lea, S.; Jensen, P. W. K.; Sim, S.; Díaz-Tinoco, M.; Kottmann, J. S.; Degroote, M.; Izmaylov, A. F.; Aspuru-Guzik, A. A quantum computing view on unitary coupled cluster theory. *Chem. Soc. Rev.* **2022**, *51*, 1659–1684.

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

(19) Tang, H. L.; Shkolnikov, V.; Barron, G. S.; Grimsley, H. R.; Mayhall, N. J.; Barnes, E.; Economou, S. E. Qubit-ADAPT-VQE: An Adaptive Algorithm for Constructing Hardware-Efficient Ansätze on a Quantum Processor. *PRX Quantum* **2021**, *2*, 020310.

(20) Taube, A. G.; Bartlett, R. J. New perspectives on unitary coupled-cluster theory. *Int. J. Quantum Chem.* **2006**, *106*, 3393–3401.

(21) Crawford, T. D.; Schaefer, H. F.*Reviews in Computational Chemistry*; John Wiley & Sons, Ltd, 2007, pp 33–136. DOI: 10.1002/9780470125915.ch2

(22) Pulay, P. Gradients in coupled-pair theories. J. Mol. Struct.: THEOCHEM 1983, 103, 57-66.

(23) Szalay, P. G.; Nooijen, M.; Bartlett, R. J. Alternative ansätze in single reference coupled-cluster theory. III. A critical analysis of different methods. *Chem. Phys.* **1995**, *103*, 281–298.

(24) Chen, J.; Cheng, H.-P.; Freericks, J. K. Low-Depth Unitary Coupled Cluster Theory for Quantum Computation. J. Chem. Theory Comput. **2022**, 18, 2193–2198.

(25) Jankowski, K.; Paldus, J. Applicability of coupled-pair theories to quasidegenerate electronic states: A model study. *Int. J. Quantum Chem.* **1980**, *18*, 1243–1269.

(26) Jankowski, K.; Paldus, J.; Grabowski, I.; Kowalski, K. Applicability of valence-universal multireference coupled-cluster theories to quasidegenerate electronic states. I. Models involving at most two-body amplitudes. *Chem. Phys.* **1992**, *97*, 7600–7612.

(27) Jankowski, K.; Kowalski, K. Approximate coupled cluster methods based on a split-amplitude strategy. *Chem. Phys. Lett.* **1996**, 256, 141–148.

(28) Jankowski, K.; Grabowski, I.; Kowalski, K. Approximate coupled-cluster methods employing split cluster amplitudes: Implementation of an almost-linear coupled-cluster formalism. *Chem. Phys.* **1998**, *109*, 6255–6263.

(29) Hoffmann, M. R.; Simons, J. A unitary multiconfigurational coupled-cluster method: Theory and applications. *Chem. Phys.* **1988**, 88, 993–1002.

(30) Wennmohs, F.; Neese, F. A comparative study of single reference correlation methods of the coupled-pair type. *Chem. Phys.* **2008**, 343, 217–230.

(31) Malrieu, J.-P.; Zhang, H.; Ma, J. Ability of a coupled electron pair approximation to treat single bond breakings. *Chem. Phys. Lett.* **2010**, 493, 179–184.

(32) Purvis, G. D.; Bartlett, R. J. The reduced linear equation method in coupled cluster theory. *Chem. Phys.* **1981**, 75, 1284–1292. (33) Ahlrichs, R. Many body perturbation calculations and coupled electron pair models. *Comput. Phys. Commun.* **1979**, *17*, 31–45.

(34) Daudey, J.-P.; Heully, J.-L.; Malrieu, J.-P. Size-consistent selfconsistent truncated or selected configuration interaction. *Chem. Phys.* **1993**, *99*, 1240.

(35) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *Chem. Phys.* **1993**, *98*, 5648–5652.

(36) Dill, J. D.; Pople, J. A. Self-consistent molecular orbital methods. XV. Extended Gaussian-type basis sets for lithium, beryllium, and boron. *Chem. Phys.* **1975**, *62*, 2921–2923.

(37) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *Chem. Phys.* **1971**, *54*, 724–728.

(38) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.

(39) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *Chem. Phys.* **1972**, *56*, 2257–2261.

(40) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1340.

(41) Kállay, M.; Surján, P. R. Higher excitations in coupled-cluster theory. *Chem. Phys.* **2001**, *115*, 2945–2954.

(42) Kállay, M.; et al. The MRCC program system: Accurate quantum chemistry from water to proteins. *Chem. Phys.* **2020**, *152*, 074107.

(43) MRCC, a quantum chemical program suite written by Kállay, M.; Nagy, P. R.; Mester, D.; Rolik, Z.; Samu, G.; Csontos, J.; Csóka, J.; SzabóGyevi-Nagy, P. B. L.; Hégely, B.; Ladjánszki, I.; Szegedy, L.; Ladóczki, B.; Petrov, K.; Farkas, M.; Mezei, P. D.; Ganyecz, Á. See www.mrcc.hu (accessed 2022-09-22).

(44) Parrish, R. M.; et al. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185–3197.

(45) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *Chem. Phys.* **1989**, *90*, 1007–1023.

(46) Soydaş, E.; Bozkaya, U. Assessment of the orbital-optimized coupled-electron pair theory for thermochemistry and kinetics: Improving on CCSD and CEPA(1). *J. Comput. Chem.* **2014**, *35*, 1073–1081.

(47) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *Chem. Phys.* **1994**, *100*, 2975–2988.

(48) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **1989**, *45*, 503–528.

(49) Lee, T. J.; Taylor, P. R. A diagnostic for determining the quality of single-reference electron correlation methods. *Int. J. Quantum Chem.* **1989**, *36*, 199–207.

(50) Evangelista, F. A.; Chan, G. K.-L.; Scuseria, G. E. Exact parameterization of fermionic wave functions via unitary coupled cluster theory. *Chem. Phys.* **2019**, *151*, 244112.

#### **Supplementary Information**

Harper R. Grimsley and Nicholas J. Mayhall Chemistry Department, Virginia Polytechnic Institute and State University (Dated: November 1, 2022)



FIG. 1: Plot of individual reaction energy errors as a function of the O2D2-UCCSD  $\hat{T}_1$  diagnostic,  $\frac{|\mathbf{t}_1|}{\sqrt{N_e/2}}$ .

## Chapter 7

### Conclusions

In chapter 2, the importance of operator ordering in a Trotterized VQE is established, and the need for an unambiguous choice of operator ordering is emphasized. The impact of operator ordering on the energy may be small for weakly correlated systems where the ansatz parameters are small, but it is extremely large for strongly correlated systems, which are the systems where classical methods break down and quantum algorithms are likely to help. Within the scope of this dissertation, this work helps to explain the success of ADAPT-VQE. In the broader scope of the field of VQE research at large, this work introduced the idea that researchers need to define their Trotterized ansätze more rigorously. It is also worth noting that this work is the first to suggest that higher-order Trotterization of the ansatz can yield an inferior energy. The conclusions of this paper are crucial to fully understanding gate-efficient Trotterized implementations of the impractical UCC ansatz.

In chapter 3, the ADAPT-VQE algorithm is introduced. ADAPT-VQE solves the ordering problem by using one-at-a-time operator addition based on energy gradients with respect to the VQE parameters. More importantly, it offers a dramatically more parameter-efficient approach to defining a VQE ansatz than the standard UCCSD ansatz or a randomly-defined ansatz using the same operators. ADAPT-VQE represents a massive improvement over traditional VQEs in terms of circuit depth and parameter counts. Numerous follow-up works to ADAPT-VQE have taken the form of modifications of the original algorithm and entirely new adaptive algorithms, both by its original developers and other groups. [35–

39] ADAPT-VQE was the first VQE to use a dynamically constructed ansatz, and remains popular in the QIS community.

In chapter 4, problems with the classical optimization in VQEs are addressed as they apply to ADAPT-VQE. The problem of local minima in the energy landscape seems to be largely circumvented by ADAPT-VQE, and a theoretical argument is made for ADAPT-VQE avoiding the problem of barren plateaus. As mentioned in the introduction, these classical optimization issues are considered major obstacles to the practical utility of VQEs. The fact that ADAPT-VQE seems to at least partially address them makes the highly parameter- and gateefficient algorithm even more attractive, and suggest that it will be one of the most practical VQEs going forward. At the same time, this work revealed a new problem with ADAPT-VQE in the form of premature convergence due to "gradient troughs" where ADAPT-VQE gets stuck in manifolds of excited states. It remains unclear whether ADAPT-VQE would have these problem cases if a greater degree of numerical precision was used in floating point operations, or if it is truly possible to have a zero operator gradient vector and still not have reached an eigenstate. Both the advantages and problems of ADAPT-VQE revealed in this paper motivate further research on the method.

In chapter 6, a family of VQE-inspired, classical chemistry methods is explored. Despite minimal additional complexity, superior results to the established O2D2 and LCCSD methods are obtained. In particular, these new approaches are far better at single bond-breaking. Interesting connections are made with existing classical methods in the CEPA family, despite totally different motivations behind the derivations. The fact that truncating certain Trotterized UCCSD ansätze yields a size-extensive method is fascinating and could potentially have implications in the Trotter ordering problem. These new methods are primarily interesting from the perspective that they are motivated by an energy functional that is only tractable in quantum algorithms. Future work on this project includes choosing better orbitals and (hopefully) developing useful insights about the parameter landscapes of UCCSD-based VQEs.

This work has revolutionized the implementations and refined the study of VQEs. Both quantum and classical algorithms have been designed here based on VQEs and their constituent parts. Despite pessimism about their practicality in the community, especially with respect to classical optimization, the research here is cause for optimism regarding VQEs. It is the hope of the author that this work motivates some interest by the reader in VQEs and, in particular, the UCC formalism.

## Bibliography

- Atkins, P. W.; Friedman, R. S., *Molecular Quantum Mechanics*, 5th edition; Oxford University Press: Oxford; New York, 2010.
- (2) Szabo, A.; Ostlund, N. S., Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, Reprint Edition; Dover Publications: Mineola, N.Y, 1996.
- (3) Vogiatzis, K. D.; Ma, D.; Olsen, J.; Gagliardi, L.; de Jong, W. A. J. Chem. Phys 2017, 147, 184111.
- (4) Abraham, V.; Mayhall, N. J. J. Chem. Phys **2021**, 155, 054101.
- Li, Z.; Li, J.; Dattani, N. S.; Umrigar, C. J.; Chan, G. K.-L. J. Chem. Phys 2019, 150, Publisher: American Institute of Physics, 024302.
- (6) Jordan, P.; Wigner, E. Z. Phys. **1928**, 47, 631–651.
- (7) Tranter, A.; Love, P. J.; Mintert, F.; Coveney, P. V. J. Chem. Theory Comput. 2018, 14, 5617–5630.
- (8) Kitaev, A. Y. 1995, DOI: 10.48550/arXiv.quant-ph/9511026.
- (9) Abrams, D. S.; Lloyd, S. *Phys. Rev. Lett.* **1999**, *83*, 5162–5165.
- (10) Peruzzo, A.; McClean, J.; Shadbolt, P.; Yung, M.-H.; Zhou, X.-Q.; Love, P. J.; Aspuru-Guzik, A.; O'Brien, J. L. Nat. Commun. 2014, 5, 4213.
- (11) McClean, J. R.; Romero, J.; Babbush, R.; Aspuru-Guzik, A. New J. Phys. 2016, 18, 023023.
- (12) Fedorov, D. A.; Peng, B.; Govind, N.; Alexeev, Y. Mater. theory 2022, 6, 2.

- (13) Lee, J.; Huggins, W. J.; Head-Gordon, M.; Whaley, K. B. J. Chem. Theory Comput. **2019**, 15, 311–324.
- Mizukami, W.; Mitarai, K.; Nakagawa, Y. O.; Yamamoto, T.; Yan, T.; Ohnishi, Y.-y.
   *Phys. rev. res.* 2020, 2, 033421.
- (15) Metcalf, M.; Bauman, N. P.; Kowalski, K.; de Jong, W. A. J. Chem. Theory Comput. **2020**, 16, 6165–6175.
- (16) Kandala, A.; Mezzacapo, A.; Temme, K.; Takita, M.; Brink, M.; Chow, J. M.; Gambetta, J. M. Nature 2017, 549, 242–246.
- (17) Ryabinkin, I. G.; Yen, T.-C.; Genin, S. N.; Izmaylov, A. F. J. Chem. Theory Comput.
  2018, 14, 6317–6326.
- Meitei, O. R.; Gard, B. T.; Barron, G. S.; Pappas, D. P.; Economou, S. E.; Barnes, E.;
  Mayhall, N. J. Npj Quantum Inf. 2021, 7, 1–11.
- (19) Kutzelnigg, W. In *Methods of Electronic Structure Theory*, Schaefer, H. F., Ed., 1st ed.;
   Modern Theoretical Chemistry, Vol. 3; Springer Science & Business Media: New York,
   NY, USA, 1977, pp 170–172.
- (20) Grimsley, H. R.; Claudino, D.; Economou, S. E.; Barnes, E.; Mayhall, N. J. J. Chem. Theory Comput. 2020, 16, Publisher: American Chemical Society, 1–6.
- Barkoutsos, P. K.; Gonthier, J. F.; Sokolov, I.; Moll, N.; Salis, G.; Fuhrer, A.; Ganzhorn,
  M.; Egger, D. J.; Troyer, M.; Mezzacapo, A.; Filipp, S.; Tavernelli, I. *Phys. Rev. A* **2018**, *98*, 022322.
- (22) Hatano, N.; Suzuki, M. 2005, DOI: 10.48550/arXiv.math-ph/0506007.
- (23) Bittel, L.; Kliesch, M. Phys. Rev. Lett. **2021**, 127, 120502.
- McClean, J. R.; Boixo, S.; Smelyanskiy, V. N.; Babbush, R.; Neven, H. Nat. Commun.
   2018, 9, 4812.

#### BIBLIOGRAPHY

- (25) Arrasmith, A.; Holmes, Z.; Cerezo, M.; Coles, P. J. Quantum Sci. Technol. 2022.
- (26) Grimsley, H. R.; Economou, S. E.; Barnes, E.; Mayhall, N. J. Nat. Commun. 2019, 10, 3007.
- (27) Grimsley, H. R.; Barron, G. S.; Barnes, E.; Economou, S. E.; Mayhall, N. J. 2022, DOI: 10.48550/arXiv.2204.07179.
- (28) Grimsley, H. R.; Mayhall, N. J. J. Chem. Theory Comput. 2022, DOI: 10.1021/acs. jctc.2c00751.
- (29) Bartlett, R. J.; Kucharski, S. A.; Noga, J. Chem. Phys. Lett. 1989, 155, 133–140.
- (30) Kutzelnigg, W. Theor. Chim. Acta **1991**, 80, 349–386.
- (31) Crawford, T. D.; Schaefer, H. F. In *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd: 2007, pp 33–136.
- (32) Chen, J.; Cheng, H.-P.; Freericks, J. K. J. Chem. Theory Comput. 2022, 18, 2193–2198.
- (33) Meyer, W. J. Chem. Phys 1973, 58, 1017–1035.
- (34) Zoboki, T.; Szabados, Á.; Surján, P. R. J. Chem. Theory Comput. 2013, 9, 2602–2608.
- (35) Ryabinkin, I. G.; Lang, R. A.; Genin, S. N.; Izmaylov, A. F. J. Chem. Theory Comput. **2020**, 16, 1055–1063.
- (36) Tang, H. L.; Shkolnikov, V.; Barron, G. S.; Grimsley, H. R.; Mayhall, N. J.; Barnes,
  E.; Economou, S. E. *PRX Quantum* 2021, *2*, 020310.
- (37) Yordanov, Y. S.; Armaos, V.; Barnes, C. H. W.; Arvidsson-Shukur, D. R. M. Commun.
   Phys. 2021, 4, 1–11.
- (38) Fan, Y.; Cao, C.; Xu, X.; Li, Z.; Lv, D.; Yung, M.-H. 2021, DOI: 10.48550/arXiv.
  2106.15210.

### BIBLIOGRAPHY

(39) Zhang, Z.-J.; Kyaw, T. H.; Kottmann, J. S.; Degroote, M.; Aspuru-Guzik, A. Quantum Sci. and Technol. 2021, 6, 035001.