Efficient Parametrization and Solution of the Electronic Wavefunction on a Quantum Processor: A Review

Harper Grimsley

July 8, 2019

> With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.

John Von Neumann

Contents

1 Classical Chemistry Methods

1.1 Hartree-Fock

The Hartree-Fock (HF) approximation is a self-consistent field (SCF) approximation to the electronic structure of a molecule, based on approximately solving equation 1 for the various orbitals, $\{\chi\}$, via linear variation.[3] (This equation is essentially a restatement of the time-dependent Schrödinger equation in the syntax of Slater determinants.)

$$E_0 = \sum_i \left\langle \chi_i \left| \hat{H} \right| \chi_i \right\rangle + \sum_{i<j} \langle \chi_i \chi_j | \chi_i \chi_j \rangle - \langle \chi_i \chi_j | \chi_j \chi_i \rangle \tag{1}$$

While variational and size-consistent, HF fails to capture dynamical correlation, and therefore overestimates chemical energies.[3][4] In spite of its lack of chemical accuracy, The SCF method is formally only $O\left(N^4\right)$ where N is the number of molecular orbitals. (This comes from the number of 2-electron integrals.)[3][4] Furthermore, it provides a single-configuration wavefunction which can function as a reference for more powerful iterative methods.[3] An alternative view of the HF state is that it exactly solves the Hamiltonian in 2.

$$\hat{H}_0 = \sum_i \hat{f}(i) \tag{2}$$

The term $\hat{f}(i)$ is the Fock operator associated with the position and spin of electron i, defined in equation 3.[3]

$$\hat{f}(i) = \hat{h}(i) + \sum_j \hat{\mathcal{J}}_j(i) + \hat{\mathcal{K}}_j(i) \tag{3}$$

In equation 3, $\hat{h}$ accounts for nuclear attraction of electron i to the nucleus, $\hat{\mathcal{J}}$ accounts for the mean Coulombic interaction between electrons i and j, and $\hat{\mathcal{K}}$ accounts for the mean "exchange" interaction between electrons i and j.[3] Because electron j is assumed to be fixed and described by its average contributions to each term, this is a single-electron operator.[3] This definition of the HF state is helpful for describing perturbation theory.[3]

1.2 Configuration Interaction

While HF theory considers the single best electronic configuration of a system, configuration interaction (CI) theory allows for linear combinations of all configurations, seeking the ideal solution of the form in equation 4.[3]

$$c_0 \left| \Psi^{(0)} \right\rangle + \sum_{i,a} c_{ij} \left| \Psi_i^a \right\rangle + \sum_{i,j,a,b} c_{ijab} \left| \Psi_{ij}^{ab} \right\rangle + ... \tag{4}$$

In full CI, every level of excitation is included to yield the exact basis set limit energy of the system.[3] Unfortunately, there are $O(N!)$ coefficients in equation 4, making full CI intractable "for all but the very smallest systems."[3,4] Alternatively, it is possible to truncate at a fixed number of excitations to reduce computational complexity.[4] Only including single excitations does nothing by Brillouin's theorem, so at least double excitations need to be included.[3] CI with single and double excitations (CISD) is $O\left(N^6\right)$; the addition of triple excitations (CISDT) gives $O\left(N^8\right)$, the addition of quadruple excitations (CISDTQ) gives $O\left(N^{10}\right)$, and so on.[4] This scaling originates from the number of determinants in each excitation scheme.[4] All CI schemes are constructed to be variational; however, only full CI is size-consistent, making truncated CI progressively worse for increasingly large molecules.[5] For this reason, truncated CI is considered inappropriate for large systems, with methods like perturbation theory and coupled cluster used in its place.[3,5] Therefore, no form of CI is applicable to most systems of interest.

1.3 Perturbation Theory

One approach to improve on a reference state is to consider the exact Hamiltonian $\hat{H}$ as the Hartree-Fock Hamiltonian, $\hat{H}_0$, with a small perturbation $\hat{\mathcal{V}}$.[3]

$$\hat{H} = \hat{H}_0 + \hat{\mathcal{V}} \tag{5}$$

Similarly, we can expand the exact wavefunction $|\Psi\rangle$ as a sum of corrections to the HF wavefunction, $|\Psi^{(0)}\rangle$, as shown in equation 6, and expand the ground-state energy E as a sum of corrections to the HF energy, $E^{(0)}$, as shown in equation 7.[3][5]

$$|\Psi\rangle = |\Psi^{(0)}\rangle + \lambda|\Psi^{(0)}\rangle + \lambda^2|\Psi^{(0)}\rangle + ... \tag{6}$$

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + ... \tag{7}$$

Breaking up these corrections by the order of expansion parameter λ, which can hereafter be ignored, gives an arbitrarily large number of equations, as given below.[5] (Note that the perturbative correction to the Hamiltonian is of first order.)

$$\hat{H}_0|\Psi^{(0)}\rangle = E_0|\Psi^{(0)}\rangle$$
$$\hat{H}_0|\Psi^{(0)}\rangle + \hat{\mathcal{V}}|\Psi^{(0)}\rangle = E_0|\Psi^{(1)}\rangle + E_1|\Psi^{(0)}\rangle$$
$$\hat{H}_0|\Psi^{(2)}\rangle + \hat{\mathcal{V}}|\Psi^{(1)}\rangle = E_0|\Psi^{(2)}\rangle + E_1|\Psi^{(1)}\rangle + E_2|\Psi^{(0)}\rangle$$
$$...$$

Sequentially solving these equations gives Møller-Plesset (MP) perturbation theory, a non-iterative, size-consistent method for finding any arbitrary order of correction to the Hartree-Fock wavefunction and energy.[5] Perturbation theory can be done to any order n, using MPn. MP1, however, does nothing by Brillouin's theorem.[4] The most common form, MP2 is $O(N^5)$, due to the four-index orbital transformation to molecular orbitals from atomic orbitals.[4] MP3 is $O(N^6)$, MP4 is $O(N^7)$, and so on due to increasing numbers of energy terms to evaluate.[4][6][7] Furthermore, the MP procedure is "not particularly accurate", and is non-variational.[7][8]

1.4 Coupled Cluster

Coupled cluster (CC) theory, like MP theory, sacrifices the variational property for size consistency.[5] CC theory is generally truncated to single and double excitations (CCSD), single, double, and triple excitations (CCSDT), and so on.[9] The coupled cluster ansatz for a given truncation level is given in equation 8.[9]

$$e^{\sum_{ia} t_i^a a_a^\dagger a_i + \frac{1}{4}\sum_{ijab} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_j a_i + ...}|\Psi^{(0)}\rangle \equiv e^{\hat{\mathcal{T}}_1 + \hat{\mathcal{T}}_2 + ...}|\Psi^{(0)}\rangle \equiv e^{\hat{\mathcal{T}}} \tag{8}$$

This form of the wavefunction allows one to separate the wavefunction of two isolated systems into a single product wavefunction, which gives the method its size consistency.[3][9] We can non-variationally solve for the t-amplitudes in equation 8 by similarity transfoming the Hamiltonian using the cluster operator, then considering the projections by the reference state and by each excited state, giving equations 9 and 10.[9]

$$\left\langle \Psi^{(0)} \left| e^{-\hat{\mathcal{T}}} \hat{H} e^{\hat{\mathcal{T}}} \right| \Psi^{(0)} \right\rangle = E \tag{9}$$

$$\left\langle \Psi_{ij...}^{ab...} \left| e^{-\hat{\mathcal{T}}} \hat{H} e^{\hat{\mathcal{T}}} \right| \Psi^{(0)} \right\rangle = 0 \tag{10}$$

The Baker-Campbell-Hausdorff (BCH) expansion enables the similarity-transformed Hamiltonian to be evaluated as a truncating sum of nested commutators.[9] The exact details of solving the t-amplitudes from equations 9 and 10 are beyond the scope of this work. It suffices to note that CCSD scales as $O(N^6)$, CCSDT scales as $O(N^8)$, and so on.[4]

1.5 Unitary Coupled Cluster

In ordinary coupled cluster theory, the BCH expansion gives a finite number of terms to evaluate; the tradeoff is that the cluster operator is not generally unitary, meaning coupled cluster is not variational.[9] An alternative approach is to replace $\hat{\mathcal{T}}$ with $\hat{\mathcal{T}} - \hat{\mathcal{T}}^\dagger$.[10] The parametrization of this new operator $\mathcal{T}'$ is as shown in equation 11.[9][10]

$$\hat{\mathcal{T}}' = \sum_{ia} t'_{ia} \left(a_a^\dagger a_i - a_i^\dagger a_a \right) + \sum_{ijab} t'_{ijab} \left(a_a^\dagger a_b^\dagger a_j a_i - a_i^\dagger a_j^\dagger a_b a_a \right) + ... \tag{11}$$

This generator is, by construction, anti-Hermitian, thus making $e^{\hat{\mathcal{T}}}$ unitary.[10,11] The unitary character of

$$e^{\hat{\mathcal{T}}} = e^{-\hat{\mathcal{T}}^\dagger}$$

makes equations 9 and 10 variational when $\hat{\mathcal{T}}$ is replaced by $\hat{\mathcal{T}}'$[9] Unfortunately, the BCH expansion no longer truncates in unitary coupled cluster, demanding a finite number of terms be included.[9,10,12,13] Unitary coupled cluster theory is not considered practical for real systems due to the exponentially increasing cost of solving equations 9 and 10.[12–14] There are other variational coupled cluster schemes with similar limitations, but we restrict our discussion to Hoffman's formulation.[9,10,12] Of the classical chemistry methods described, only unitary coupled cluster methods can be truncated to a fixed level of excitation while preserving the variational principle and maintaining size consistency. Its only downside is classical intractability; this serves as sufficient motivation to explore the application of quantum computers to quantum chemistry.

2 Quantum Information Background

2.1 Qubits

In classical computer science, all information is compressed to bits, binary units of information with value of either 0 or 1.[15] The quantum analogue of a bit is a qubit, which is assigned any unit vector in $\mathbf{C}^2$, as shown in equation 12.[15]

$$|\Psi\rangle = \alpha |0\rangle + \beta |1\rangle \equiv \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \tag{12}$$

Furthermore, a qubit obeys all of the postulates of quantum mechanics, making it intuitively useful for simulating quantum mechanical problems.[15][16] A system can exist with more than one qubit; for example, the state in equation 13 is a pure 2-qubit state, while the state in equation 14 is an entangled 2-qubit state.

$$|\Psi\rangle = |00\rangle \tag{13}$$

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{14}$$

Note that both states maintain the unit normalization condition. A pure 2-qubit state is one in which the outcomes are uncorrelated or *unentangled*. Entanglement describes how much the state of one qubit affects the other; for example, the qubit in equation 14 is *maximally entangled* because one qubit cannot be measured without collapsing the state of the other to match.

Measuring the state of a qubit is more complicated than measuring the state of a bit. Consider the arbitrary qubit in equation 12; quantum mechanics tells us that if we measure in the computational basis, we can get only outcomes $|0\rangle$ or $|1\rangle$ with probabilities given by equations 15 and 16.

$$P\left(|0\rangle\right) = |\langle 0|\Psi\rangle|^2 = |\alpha|^2 \tag{15}$$

$$P\left(|1\rangle\right) = |\langle 1|\Psi\rangle|^2 = |\beta|^2 \tag{16}$$

The physics literature tends to refer to this technique as projective measurement, and define the measurement operator M for state $|\phi\rangle$ by equation 17.[15]

$$M = |\phi\rangle \langle\phi| \tag{17}$$

This convention can be helpful when one is interested in the state of a system post-measurement, since it enacts the state collapse. (The measurement operator does not renormalize the state; this must be done separately.) The projective measurement equations 15 and 16 in this notation are given by equations 18 and 19 respectively.[15]

$$P(|0\rangle) = \langle\Psi|0\rangle \langle 0|0\rangle \langle 0|\Psi\rangle = |\alpha|^2 \tag{18}$$

$$P(|1\rangle) = \langle\Psi|1\rangle \langle 1|1\rangle \langle 1|\Psi\rangle = |\beta|^2 \tag{19}$$

In general, the probability of measuring $|\Psi\rangle$ in the state described by M is given in equation 20.[15]

$$P = \langle\Psi| M^\dagger M |\Psi\rangle \tag{20}$$

We now consider the case of measuring an arbitrary 2-qubit state, as given in equation 21.

$$|\Psi\rangle = \alpha |00\rangle + \beta |01\rangle + \gamma |10\rangle + \delta |11\rangle \tag{21}$$

Measuring both qubits simultaneously works the same way as measuring a single qubit; the only difference is that there are four orthogonal states the two qubits can be in. Measuring one qubit does not generally collapse the other qubit, so we need to apply the measurement operator to see what happens to the state. Consider the case where we measure the first qubit to be in state $|0\rangle$. Then the measurement operator is $|0\rangle_1 \langle 0|_1$, and we partially collapse our state according to equation 22.

$$|0\rangle_1 \langle 0|_1 |\Psi\rangle = |0\rangle_1 \otimes (\alpha \langle 0|0\rangle_1 |0\rangle_2 + \beta \langle 0|0\rangle_1 |1\rangle_2 + \gamma \langle 1|0\rangle_1 |0\rangle_2 + \delta \langle 1|0\rangle_1 |0\rangle_2) = \alpha |00\rangle + \beta |01\rangle \tag{22}$$

Evaluating the probabilities of measuring $|\Psi\rangle$ to be in each basis state is performed according to equation 20; the tensor bookkeeping works similarly to equation 22. Finally, it is worth noting that the global phase of a qubit is computationally irrelevant; many algorithms take advantage of this fact.[15]

2.2   Quantum Gates

Where the bits of classical computing are replaced with quantum states, the gates of classical computers are replaced with unitary linear operators, or *quantum gates*.[15] The first gates we consider are the Pauli operators, $\hat{X}$, $\hat{Y}$, and $\hat{Z}$. In the computational basis, these take the forms given in 23, along with the common Hadamard, phase, and $\frac{\pi}{4}$ gates.[15]

$$\hat{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad\qquad \hat{Y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad\qquad \hat{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{23}$$

$$\hat{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad\qquad \hat{S} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \qquad\qquad \hat{T} = \begin{pmatrix} 1 & 0 \\ 0 & e^{\frac{i\pi}{4}} \end{pmatrix}$$

We now introduce the CNOT gate, which can change the entanglement between two qubits.[15] Consider two qubits, the control and the target of the CNOT. If the control qubit is in state $|0\rangle$, nothing happens. If it is in state $|1\rangle$, the target qubit has the X gate applied to it, which is the quantum NOT operator. In conjunction with the single-qubit gates listed, the CNOT gate forms a universal quantum gate set; that is, any quantum circuit can be approximated to arbitrary precision with only this set. In fact, the Pauli gates can be excluded and this will still be a universal set.[15]

As a final note on gates, we introduce some basic circuit diagram notation. Each line in a circuit diagram represents a single qubit and is read left to right. Single qubit operators are represented as labeled boxes. CNOT gates are labeled with a dot on the control qubit, connected by a line to an open circle on the target qubit. Other controlled unitary operators instead have a labeled box on the target qubit. Measurements in the computational basis are shown by an electrical meter. By default, the initial state of each qubit is assumed to be $|0\rangle$. In figure 1, we give a simple example, a circuit which is used to prepare the maximally entangled spin singlet state given in equation 24. [15]

$$|\Psi^-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}} \quad (24)$$

Figure 1: Spin Singlet Preparation Gate

2.3 Contemporary Quantum Hardware

2.3.1 Metrics for Quantum Utility

We now have all of the theoretical tools to begin constructing useful algorithms. Unfortunately, we still require some experimental ones. A variety of experimental setups are areas of active research, but several issues persist, including state preparation, gate implementation, coherence of quantum states, and measurements. [15] Physicists have quantified all of these issues and use them to assess various implemenation schemes. We introduce a few common, related quantities in table 1. [15]

Table 1: Various Numerical Metrics

Metric	Definition
T_1	Classical relaxation time from $|1\rangle$ to $|0\rangle$
T_2	Coherence time of arbitrary superpositions of $|1\rangle$ and $|0\rangle$
$\mathcal{F}$	Fidelity of a gate implementation
t_{op}	Time to complete a gate implementation

These metrics will be described for each of several promising quantum computing technologies, with a focus on superconducting qubits.

2.3.2 Superconducting Qubits

One approach to introducing controllable quantum excitation is by controlling fluxons with a magnetic field or controlling Cooper pairs with an electric field. [17] The electrical case is referred to as a charge qubit or Cooper pair box, diagrammed in figure 2. [17][18] By convention, the elements of the circuit are, from left to right, a capacitor, a Josephson junction, and a controlled voltage source. The purpose of this circuit is to trap variable numbers of excess Cooper electron pairs between the Josephson junction and the capacitor on a "superconducting island." [18] (A Josephson junction is simply an inductor sandwiched between superconducting electrodes, such that quantum tunneling is the only mechanism of current flow through the junction.) [17] The charge gradient on the capacitor can be controlled by the voltage applied, enabling the manual control of the number of electrons on the island. [17,18] Because the Josepson junction is a non-linear inductor, the island can be modeled as an anharmonic oscillator, with its lowest two eigenstates as $|0\rangle$ and $|1\rangle$. [17] The anharmonicity is important to avoid leakage into excited states. [17] In practice, the high charge densities of these systems have caused high sensitivity to environmental effects, motivating the addition of a charge shunt to the design;

Figure 2: Cooper Pair Box
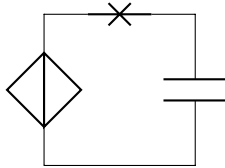
shunted charge qubits, or *transmons*, are the primary tools of our experimental collaborators.[17] Transmons have reasonably good coherence times on the order of 100 μs, with the unusual feature of T_2 and T_1 times being roughly equal so that T_2 is not the only bottleneck.[15,17]

Single-qubit gates are fairly easy to implement on transmons, with fidelities above 99% typical.[17] These are typically accomplished by driving individual qubits with microwave pulses.[17] Only some two-qubit gates have had implementation demonstrated, with only a few achieving $\mathcal{F} > .99$.[17] Coupling schemes between qubits are more nuanced; conceptually, the simplest is the microwave driving of one qubit at the frequency of a second qubit. Other setups allow for magnetic tuning of the qubits into resonance with one another or driving some single-qubit gate times take only a few nanoseconds, while two-qubit gate times on transmons vary from a few nanoseconds to several microseconds depending on implementation.[17][19] The fastest gates are associated with normal, 2D transmons, while the longest coherence times are associated with transmons embedded in a 3D cavity.[19]

2.3.3 Other Physical Qubit Realizations

While we are primarily interested in transmons, it is worth comparing them to other experimental setups to understand the advantages of transmons. In the table below, we give approximate figures of merit for both transmons and other common qubit systems.[19–21]

Table 2: Comparison of Various Experimental Setups

	T_1	T_2	t_{op} (single-qubit)	t_{op} (two-qubit)	$\mathcal{F}$ (single-qubit)	$\mathcal{F}$ (two-qubit)
2D Transmon	$\approx 40\ \mu$s	$\approx 40\ \mu$s	10-20 ns	10-40 ns	$\approx .999$	$>.99$
3D Transmon	100 μs	$>140\ \mu$s	30-40 ns	≈ 450 ns	$>.999$	.96-.98
Ion Trap (^{43}Ca$^+$)	≈ 1s	≈ 1s	1-3000 μs	$\approx 10\mu$s	$>.99$	$>.9999$
...						

In reality, trapped ions represent an extremely diverse class of qubit; however, it suffices for our discussion to note that they have longer coherence times than transmons, but longer gate times.

3 Quantum Algorithms for Quantum Chemistry

We have, until now, ignored the mathematical issue of translating spin-orbitals to qubits. This problem has several existing solutions, including the Jordan-Wigner, Bravyi-Kitaev, and parity mappings. We offer an overview of Jordan-Wigner in appendix A. Our usage of the other mappings is limited, and they are less chemically intuitive, despite in certain cases being more resource-efficient.

3.1 Phase Estimation Algorithm

The phase estimation algorithm, PEA, was developed as a general way to quickly compute the spectra of large Hermitian matrices; the value of such an algorithm was quickly recognized by physicists and chemists. [1,2] One possible circuit for PEA is given in figure 3. [1,2]

Figure 3: Phase Estimation Algorithm Circuit
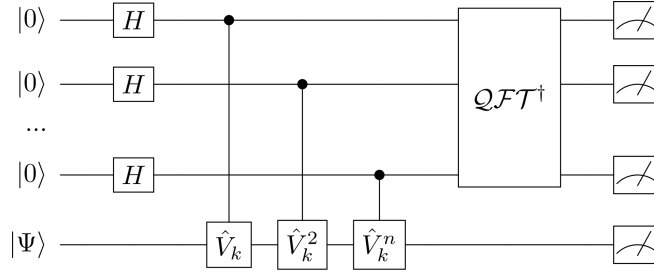
In traditional phase estimation of a Hamiltonian, this circuit is run repeatedly with $\hat{V}_0$ defined in equation 25. [1]

$$\hat{V}_0 = e^{-\frac{i}{\hbar}\hat{H}t} \tag{25}$$

$|\Psi\rangle$ is an n-qubit reference state which is not orthogonal to the Hamiltonian eigenstates. [1] A more detailed explanation of the algorithm is provided in appendix C, but the end result is a way to compute every eigenstate and eigenvalue of the Hamiltonian in polynomial time, provided the orthogonality requirement on the reference is met. [1] Aspuru-Guzik and co. introduced a recursive variant of this scheme to require only a constant number of qubits in the readout register regardless of system size. [2] Unfortunately, the coherent evolution of $|\Psi\rangle$ required by PEA is beyond the reach of existing quantum hardware, motivating the development of the variational quantum eigensolver, or VQE. [22]

3.2 Variational Quantum Eigensolvers

VQE fundamentally trades the coherence requirement of PEA for more measurements. [22] In its original form (and the form of greatest interest to us,) Peruzzo and co. considered the UCC ansatz as a fixed unitary operator to replace the evolution unitary operator, limiting the parameter space to the dimension of the level of coupled cluster theory and guarenteeing polynomial scaling for any given level of truncation. [22] In general, the Suzuki-Trotter approximation is employed to ameliorate computational expense of the UCC ansatz, and in some cases, improves the accuracy of the VQE. [14,23,24] The approximation uses a truncated form of equation 26. [24]

$$e^{A+B} = \lim_{n \to \infty} e^A e^B \tag{26}$$

A rudimentary explanation of the necessity of the Trotter approximation is given in appendix B. The VQE algorithm variationally minimizes the expectation value of the parametrized ansatz, dividing the work between classical and quantum processors; the classical processor computes integrals, performs transforms, and runs a classical parameter optimization, outsourcing energy evaluations to the quantum processor. [22] While VQE is more feasible in the near future than PEA, methods which reduce coherence requirements even further are a subject of active investigation, leading to schemes like k-UPCCGSD and ADAPT. [14,22,23]

3.3 Toward NISQ Viability of VQE

The endgame of quantum computing is to construct true, fault-tolerant quantum computers; however, this is unlikely to be feasible for many years. [25] Of more interest is what Preskill dubbed the Noisy, Intermediate-

Scale Quantum (NISQ) era, which anticipates the construction of 50-100 qubit quantum processors which cannot be manipulated, maintained, or measured with perfect fidelity.[25] The NISQ regime is characterized by quantum processors becoming too large to simulate clasically; in the case of chemistry, this is expected to translate to running polynomial-time FCI calculations on larger systems than can be done with contemporary classical hardware.[2,25] Algorithm design in this regime generally involves solving problems with as few quantum resources as possible, even if it means a dramatic increase in classical resource requirements; VQE is one such example of a "hybrid" method, but it has been (and is still being) improved in a variety of schemes since its inception.[14,23,25,26]

Of particular promise are adaptive methods such as Adaptive Derivative-Assembled Pseudo-Trotterized (ADAPT) ansatz VQE.[23] While this type of scheme has been around less than a year, ADAPT has spawned interest in the physics community in developing similarly adaptive methods in other contexts, and continues to be a subject of research in our own group.[26,27] The ADAPT algorithm is presented graphically in figure 4.[23]

Figure 4: ADAPT Algorithm
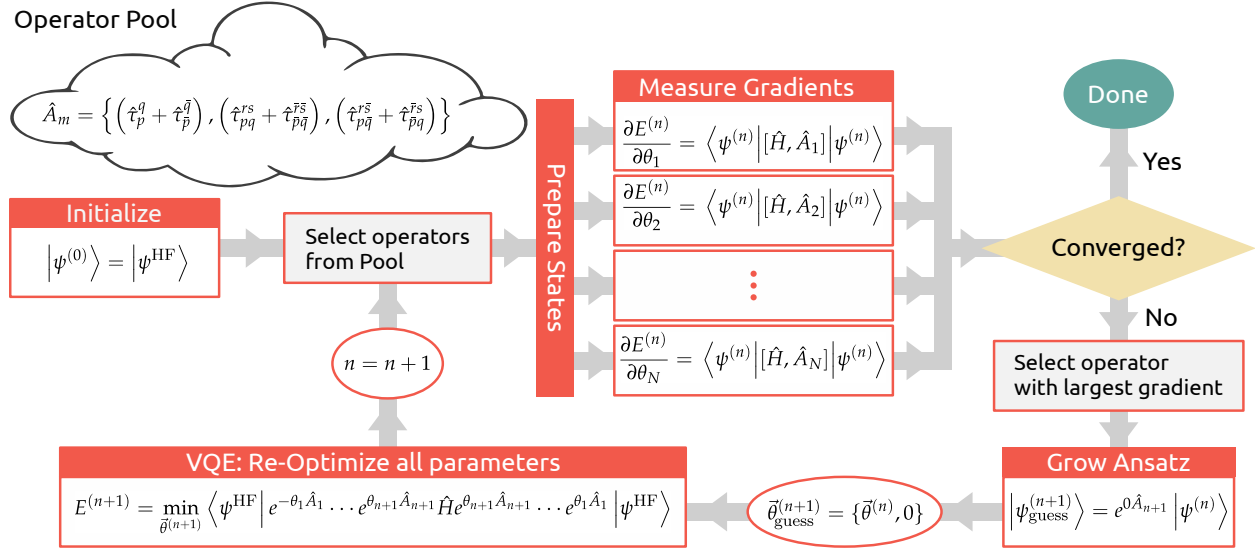
The algorithm begins with a reference state, generally the HF state. At this point, we consider each of the operators in a predetermined pool. The HF state can be thought of as the HF state being acted on by one of the operators with the parameter $\theta = 0$. The energy gradient with respect to this operator is facile to compute, especially using the gradient algorithm derived by our group, and is not generally zero.[23] Every gradient is measured according to the formulas pictured in figure 4. The operator giving the largest magnitude gradient is then permanently added to the ansatz. The parameter θ is then optimized by gradient descent. We repeat this process for the next gradient, reoptimizing all parameters with each addition to obtain a quasi-optimal ansatz. Stopping criteria are presently somewhat arbitrary, and remain a subject of interest to our group.

The ADAPT ansatz is demonstrably more operator-compact than the large majority of possible random ansätze for chemical systems, as shown in figure 5.[23] A key caveat to ADAPT's performance is that it is generally more parameter-efficient than its corresponding unitary coupled cluster method, even for strongly correlated systems such as H_6. The ineffectiveness of untrotterized UCC is well-known in this regime, making it an area of great interest.[12]

plot

4 Active Research

4.1 ADAPT Development

4.2 Demystifying Trotterization

4.3 Design of the ADAPT Operator Pool

A Jordan-Wigner Transformation

We give here a brief overview of the Jordan-Wigner transfomation, the most intuitive mapping between fermionic and qubit operators. Our second-quantized operators are tools for the spin-orbital basis, not for the 2^n occupation states of n qubits. We seek then to define the annihilation operator σ^- in the qubit basis, such that the anti-commutation relations pictured in 27 hold. [3]

$$\{\sigma_j^+, \sigma_k^-\} = \delta_{jk} \tag{27}$$
$$\{\sigma_j^-, \sigma_k^-\} = 0$$

where $\sigma^+ = (\sigma^-)^\dagger$. Consider the action of the annihilation operator on an occupation product state $|\alpha\rangle = \otimes_{i=1}^n |\alpha_i\rangle$ given in 28, where each α_i is restricted to 0 or 1, and σ_j has the same properties as the fermionic annihilation operator such that it maps $|1\rangle$ to $|0\rangle$ and $|0\rangle$ to 0. [28]

$$\begin{aligned}
\sigma_j^- |\alpha\rangle &= \sigma_j^- |\alpha_1\rangle \otimes_{i=2}^n |\alpha_i\rangle \tag{28}\\
&= \sigma_j^- \left(\sigma_1^\dagger\right)^{\alpha_1} \otimes_{i=2}^n |\alpha_i\rangle\\
&= -1^{\alpha_1} |\alpha_1\rangle \sigma_j^- \otimes_{i=2}^n |\alpha_i\rangle\\
&= Z |\alpha_1\rangle \sigma_j^{-(2)} \otimes_{i=2}^n |\alpha_i\rangle\\
&\quad \text{Repeat j-2 additional times}\\
&= \otimes_1^{i=j-1} Z |\alpha_i\rangle \sigma_j^{-(j)} |\alpha_j\rangle \otimes_{i=j+1}^n |\alpha_i\rangle\\
&= \otimes_1^{i=j-1} Z |\alpha_i\rangle |0\rangle \langle 1|\alpha_j\rangle \otimes_{i=j+1}^n |\alpha_i\rangle
\end{aligned}$$

where $\sigma^{-(j)}$ acts only on qubits of index j or greater. Similarly, the creation operator acting on $|\alpha\rangle$ produces **??**

$$\otimes_1^{i=j-1} Z |\alpha_i\rangle |1\rangle \langle 0|\alpha_j\rangle \otimes_{i=j+1}^n |\alpha_i\rangle$$

It follows from the definitions of the Pauli operators that:

$$\sigma_j^- = -\frac{1}{2} Z \otimes_1^{j-1} (X - iY) I \otimes_{j+1}^n \quad \sigma_j^+ = -\frac{1}{2} Z \otimes_1^{j-1} (X + iY) I \otimes_{j+1}^n \tag{29}$$

It can be shown that both of the anti-commutation relations in 27 hold for these operators. Consequently, the Jordan-Wigner transformation allows the Pauli string expression of any operator or vector which we know how to express in terms of creation and annihilation operators. This includes initial states, Hamiltonians, and parametrized excitations. Naturally, this extends to the unitary coupled cluster ansatz.

B Suzuki-Trotter Decomposition Background

The Suzuki-Trotter decomposition is generally considered in the context of time evolution operators such as the propagator $e^{i\hat{H}}$; using a finite product of these "slices" is equivalent to using finite time steps to approximate the evolution.[24,29,30] It follows that in the limit of infinite slices, one obtains the true evolution.[24] However, finite truncations of the Trotter decomposition suffer not only from truncation error, but are not generally uniquely defined; in general, there is no reason to expect the unitary operators used to commute, and Trotter error is well-understood as a source of error in evaluating the Hamiltonian.[29] The effect of unitary coupled cluster ansatz decomposition ordering remains unexplored in the literature, and has become an active area of research for our group. In 30, we give examples of untrotterized, first-order trotterized, and second-order trotterized ansätze.

$$\text{Untrotterized:} \quad e^{A\theta_a + B\theta_b + \cdots} = e^{B\theta_b + A\theta_a + \cdots} \tag{30}$$

$$\text{First-order trotterized:} \quad e^{A\theta_a} e^{B\theta_b} \cdots \neq e^{B\theta_b} e^{A\theta_a} \cdots$$

$$\text{Second-order trotterized:} \quad \left(e^{\frac{A\theta_a}{2}} e^{\frac{B\theta_b}{2}} \cdots \right)^2 \neq \left(e^{\frac{B\theta_b}{2}} e^{\frac{A\theta_a}{2}} \cdots \right)^2$$

Note the lack of uniqueness in all finite Trotter decompositions.

C Phase Estimation Algorithm Details

Here we provide more detailed explanations of the phase estimation algorithm and the recursive variant developed by Aspuru-Guzik and co.[1,2] Referring to figure 3 will be helpful for the latter.

C.1 Traditional Phase Estimation[1]

This scheme begins by using the n-qubit Hadamard gate to obtain a linear combination of each of the states which can be obtained in an n-qubit working register according to equation31.

$$H^{\otimes n} |0\rangle^{\otimes n} |\Psi\rangle = \frac{1}{\sqrt{N}} \sum_i |i\rangle |\Psi\rangle \tag{31}$$

An input register is initialized to some arbitrary phase $|\Psi\rangle$ which (hopefully) has non-zero overlap with the true eigenvectors of the system. At this point, a series of controlled unitary gates $\hat{V}^i$ is prepared; the specific gates of this preparation depend on hardware capabilities. This leads eventually to the preparation of the state in 32, using the spectral decomposition of $|\Psi\rangle$ with respect to $\hat{V}$.

$$\frac{1}{\sqrt{N}} \sum_i |i\rangle \hat{V}^i |\Psi\rangle \equiv \frac{1}{\sqrt{N}} \sum_{i\lambda} |i\rangle \hat{V}^i c_\lambda |\lambda\rangle \tag{32}$$

Taking λ to be the eigenvalue associated with $|\lambda\rangle$, this becomes represented by 33.

$$\frac{1}{\sqrt{N}} \sum_{i\lambda} c_\lambda \lambda^i |i\rangle |\lambda\rangle \tag{33}$$

Defining λ by equation 34, using $\hat{j}$ as the imaginary unit, we can perform a quantum Fourier transform to obtain the phase ω_λ and consquently the λ for whatever we measure the system in. Repeated evaluations allow the determination of all eigenstates and eigenvalues, obtaining accuracy which is theoretically identical to FCI.

$$\lambda = e^{i\omega_\lambda \hat{j}} \tag{34}$$

C.2 Recursive Phase Estimation[2]

The primary advantage of this algorithm is the ability to obtain high precision with a small number of working qubits. The recursive algorithm begins with a four-qubit PEA regardless of system size; giving the phase of the ground state ω (and consequently λ_0) with a precision of 1/16. The time evolution operator $\hat{V}_0$ is replaced now with 35, and another PEA is conducted. Squaring $\hat{V}_0$ and dividing out the approximate ground state energy causes a phase shift and ensures that the second PEA will only search the space of improvements on this eigenstate, giving approximate precision .004. This is repeated to an arbitrary degree of precision.

$$\hat{V}_1 = \frac{1}{\lambda_0}\hat{V}_0^2 \tag{35}$$

D   Efficient Classical Gradient Computation

While analytic formulae for the gradient of the trotterized UCC ansatz have existed for a few years now, in ref. 23 we derived the most tractable algorithm by far for this evaluation on a classical computer.[23,31–33] We outline here the proof and execution details of this recursive algorithm. We begin by expanding the analytic derivative and recognizing some basic identities in 36.

$$\frac{\partial E}{\partial \theta_k} = \frac{\partial}{\partial \theta_k} \langle HF| e^{-A\theta_a}...e^{-K\theta_k}...e^{-Z\theta_z}\hat{H}e^{Z\theta_z}...e^{K\theta_k}...e^{A\theta_a} |HF\rangle \tag{36}$$

$$= \langle HF|...e^{-K\theta_k}...\hat{H}...Ke^{K\theta_k}...|HF\rangle - \langle HF|...Ke^{-K\theta_k}...\hat{H}...e^{K\theta_k}...|HF\rangle$$

$$= 2\mathcal{R} \langle HF|...e^{-K\theta_k}...\hat{H}...Ke^{K\theta_k}|HF\rangle$$

Because we intend to evaluate a large number of gradients, we introduce a recursive scheme to obtain all of the gradients between A and Z. We define some useful quantities in 37.

$$\langle \psi_{\{A,...,N\}}^H| = \langle HF| e^{-A\theta_a}...e^{-Z\theta_z}\hat{H}e^{A\theta_a}...e^{N\theta_n} \tag{37}$$

$$|\psi_{\{A,...,N\}}\rangle = e^{N\theta_n}...e^{A\theta_a} |HF\rangle$$

We recognize that

$$\frac{\partial E}{\partial \theta_a} = 2\mathcal{R} \langle \psi_\emptyset^H| A |\psi_{\{A,...,Z\}}^H\rangle \tag{38}$$

Computing 38 requires evaluating N exponential matrix actions, where N is the number of unitary operators in the ansatz. (It is not 2N, because both terms in 38 are related by transposition and multiplication by the Hamiltonian.) Additionally

$$\frac{\partial E}{\partial \theta_b} = \langle \psi_{\{A\}}^H| B |\psi_{\{B,...Z\}}\rangle \tag{39}$$

Computing 39 only requires two exponential matrix actions, because

$$\langle \psi_{\{A\}}^H| = \langle \psi_\emptyset^H| e^{A\theta_a} \tag{40}$$

$$|\psi_{\{B,...Z\}}\rangle = e^{-A\theta_a} |\psi_{\{A,...,Z\}}\rangle \tag{41}$$

This procedure can be repeated N times for a total of 3N exponentiated matrices to compute the complete gradient.

References

[1] Abrams, D. S.; Lloyd, S. *Phys. Rev. Lett.* **1999**, *83*, 5162–5165.

[2] Aspuru-Guzik, A.; Dutoi, A. D.; Love, P. J.; Head-Gordon, M. *Science* **2005**, *309*, 1704–1707.

[3] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, revised ed. edition ed.; Dover Publications: Mineola, N.Y, 1996.

[4] Jensen, F. *Introduction to Computational Chemistry*, 3rd ed.; Wiley: Chichester, UK ; Hoboken, NJ, 2017.

[5] Atkins, P. W.; Friedman, R. S. *Molecular Quantum Mechanics*, 5th ed.; Oxford University Press: Oxford; New York, 2010.

[6] Diner, S.; Malrieu, J. P.; Claverie, P. *Theoret. Chim. Acta* **1967**, *8*, 390–403.

[7] Pulay, P.; Saebø, S. *Theoret. Chim. Acta* **1986**, *69*, 357–368.

[8] Saebø, S.; Pulay, P. *J. Chem. Phys.* **1987**, *86*, 914–922.

[9] Crawford, T. D.; Schaefer, H. F. *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd, 2007; pp 33–136.

[10] Hoffmann, M. R.; Simons, J. *J. Chem. Phys.* **1988**, *88*, 993–1002.

[11] Gallier, J. H. *Geometric Methods and Applications: For Computer Science and Engineering*; Springer Science & Business Media, 2001; Google-Books-ID: B4JtblR1lkMC.

[12] Evangelista, F. A. *J. Chem. Phys.* **2011**, *134*, 224102.

[13] Taube, A. G.; Bartlett, R. J. *International Journal of Quantum Chemistry* **2006**, *106*, 3393–3401.

[14] Lee, J.; Huggins, W. J.; Head-Gordon, M.; Whaley, K. B. *J. Chem. Theory Comput.* **2019**, *15*, 311–324, arXiv: 1810.02327.

[15] Nielsen, M. A.; Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition*, anniversary edition ed.; Cambridge University Press: Cambridge ; New York, 2011.

[16] Feynman, R. P. *Found Phys* **1986**, *16*, 507–531.

[17] Kjaergaard, M.; Schwartz, M. E.; Braumller, J.; Krantz, P.; Wang, J. I.-J.; Gustavsson, S.; Oliver, W. D. *arXiv:1905.13641 [cond-mat, physics:physics, physics:quant-ph]* **2019**, arXiv: 1905.13641.

[18] Wu, R. B.; Li, T. F.; Kofman, A. G.; Zhang, J.; Liu, Y.-X.; Pashkin, Y. A.; Tsai, J.-S.; Nori, F. *Phys. Rev. A* **2013**, *87*, 022324.

[19] Wendin, G. *Rep. Prog. Phys.* **2017**, *80*, 106001.

[20] Bruzewicz, C. D.; Chiaverini, J.; McConnell, R.; Sage, J. M. *Applied Physics Reviews* **2019**, *6*, 021314.

[21] Lucas, D. M.; Keitch, B. C.; Home, J. P.; Imreh, G.; McDonnell, M. J.; Stacey, D. N.; Szwer, D. J.; Steane, A. M. *arXiv:0710.4421 [quant-ph]* **2007**, arXiv: 0710.4421.

[22] Peruzzo, A.; McClean, J.; Shadbolt, P.; Yung, M.-H.; Zhou, X.-Q.; Love, P. J.; Aspuru-Guzik, A.; OBrien, J. L. *Nature Communications* **2014**, *5*, 4213.

[23] Grimsley, H. R.; Economou, S. E.; Barnes, E.; Mayhall, N. J. *arXiv:1812.11173 [cond-mat, physics:physics, physics:quant-ph]* **2018**, arXiv: 1812.11173.

[24] Hatano, N.; Suzuki, M. *arXiv:math-ph/0506007* **2005**, *679*, 37–68, arXiv: math-ph/0506007.

[25] Preskill, J. *Quantum* **2018**, *2*, 79.

[26] Ostaszewski, M.; Grant, E.; Benedetti, M. *arXiv:1905.09692 [quant-ph]* **2019**, arXiv: 1905.09692.

[27] Ryabinkin, I. G.; Genin, S. *https://arxiv.org/abs/1906.11192 [quant-ph]* **2019**,

[28] Nielsen, M. A. Complete notes on fermions and the Jordan-Wigner transform. 2005; `http://michaelnielsen.org/blog/complete-notes-on-fermions-and-the-jordan-wigner-transform/`.

[29] Hastings, M. B.; Wecker, D.; Bauer, B.; Troyer, M. *arXiv:1403.1539 [quant-ph]* **2014**, arXiv: 1403.1539.

[30] Babbush, R.; McClean, J.; Wecker, D.; Aspuru-Guzik, A.; Wiebe, N. *Phys. Rev. A* **2015**, *91*, 022311, arXiv: 1410.8159.

[31] Crooks, G. E. *arXiv:1905.13311 [quant-ph]* **2019**, arXiv: 1905.13311.

[32] Romero, J.; Babbush, R.; McClean, J. R.; Hempel, C.; Love, P.; Aspuru-Guzik, A. *arXiv:1701.02691 [quant-ph]* **2017**, arXiv: 1701.02691.

[33] Guerreschi, G. G.; Smelyanskiy, M. *arXiv:1701.01450 [quant-ph]* **2017**, arXiv: 1701.01450.