

Improving the Qubit-Efficiency of Quantum Algorithms for the Electronic Structure
Problem Using Orbital Optimization

by

Joel Bierman

Department of Physics
Duke University

Defense Date: March 21, 2024

Approved:

Jianfeng Lu, Supervisor

Thomas Barthel

Shailesh Chandrasekharan

Kenneth Brown

Joshua Socolar

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Physics
in the Graduate School of Duke University
2024

ABSTRACT

Improving the Qubit-Efficiency of Quantum Algorithms for the Electronic Structure
Problem Using Orbital Optimization

by

Joel Bierman

Department of Physics
Duke University

Defense Date: March 21, 2024

Approved:

Jianfeng Lu, Supervisor

Thomas Barthel

Shailesh Chandrasekharan

Kenneth Brown

Joshua Socolar

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Physics
in the Graduate School of Duke University
2024

Copyright © 2024 by
Joel Bierman

All rights reserved except the rights granted by the Creative Commons
Attribution-Noncommercial Licence

Abstract

Solving the time-independent Schrödinger equation for the electronic structure Hamiltonian in quantum chemistry is hoped to be one problem where quantum computers may provide an early advantage over classical computers. The basic intuition behind this is that quantum computers are able to prepare states with an exponentially large number of probability amplitudes with a linear number of qubits, whereas classical methods must introduce approximations and heuristics to avoid the need to store and perform operations on such exponentially large states. We address two challenges that occur within the context of developing algorithms for solving the electronic structure problem on quantum computers:

1. developing methods which not only find the ground state, but also excited states and;
2. contending with the basis set truncation error which requires the use of large numbers of qubits using conventional methods.

We first develop the quantum Orbital Minimization Method (qOMM) and show through numerical simulations using Qiskit that it is able to converge much more quickly to a set of low-lying excited states than another method, the Subspace Search Variational Quantum Eigensolver (SSVQE) which has appeared in the literature in recent years. We then develop the optimal orbital variational quantum eigensolver (OptOrbVQE) algorithm and numerically simulate it using Qiskit to show that it can often achieve lower basis set truncation error in the ground state energy than methods using larger, conventional basis sets. We then generalize this method to find excited states in optimized basis sets and demonstrate analogous results to the ground state case through numerical simulations in Qiskit.

Dedication

To my family.

Contents

Abstract	iv
List of Tables	ix
List of Figures	x
Acknowledgements	xii
1 Introduction	1
1.1 The Electronic Structure Problem	1
1.2 A Brief Summary of Classical Methods	9
1.2.1 Hartree-Fock Methods	9
1.2.2 Configuration Interaction	12
1.2.3 Coupled-Cluster Methods	14
1.2.4 Orbital Optimization Methods	17
1.3 The Basics of Quantum Computing	24
1.4 The Electronic Structure Problem on Quantum Computers	33
2 Quantum Phase Estimation Background	39
2.1 Multi-Ancilla Phase Estimation	39
2.2 Iterative Phase Estimation	43
2.3 Bayesian Phase Estimation	46
2.4 Resource Estimates for Quantum Phase Estimation	47
3 Variational Eigensolvers Background	51
3.1 The Variational Framework	51
3.2 The Variational Quantum Eigensolver	53
3.2.1 VQE Measurements	54
3.2.2 The Ansatz Circuit	57
3.2.3 Classic Optimization Subroutine	61
3.3 Excited States Solvers	63

3.3.1	Quantum Subspace Expansion	64
3.3.2	Folded Spectrum VQE	65
3.3.3	Variational Quantum Deflation	66
3.3.4	State-Averaged Excited States Solvers	68
3.4	VQE Resource Estimates	71
4	The Quantum Orbital Minimization Method	74
4.1	qOMM Algorithm	74
4.2	Numerical Results	78
4.2.1	H ₂ and H ₄	81
4.2.2	Noisy H ₂ simulations	83
4.2.3	LiH	84
4.3	Discussion	86
4.3.1	Convergence Rate Plateaus	86
4.3.2	Initialization	88
4.3.3	Ansatz Circuit Depth	90
4.3.4	Inner Product Circuit Depth	91
4.3.5	Conclusion	93
5	OptOrbVQE Algorithm	95
5.1	Motivation and Background	95
5.2	Optimal Orbital VQE	97
5.3	Numerical Results	103
5.3.1	Minimal Qubit Usage	103
5.3.2	Increasing Qubit Resources	109
5.3.3	Robustness to Noise	113
5.4	Discussion	116
6	Excited States Orbital Optimization	118

6.1	Introduction	118
6.2	State-Averaged Quantum Eigensolvers	121
6.3	State-Averaged Orbital Optimization	122
6.4	Numerical Results	125
6.4.1	H ₂	127
6.4.2	H ₄	129
6.4.3	LiH	130
6.4.4	BeH ₂	132
6.4.5	H ₄ Noisy Binding Curve	132
6.5	Discussion and Conclusions	135
7	The Effect of Orbital Optimization on Hamiltonian Sparsity and Circuit Depth	138
7.0.1	Introduction	139
7.1	Numerical Results	141
7.1.1	H ₄	142
7.1.2	LiH	146
7.2	Discussion and Conclusions	150
8	Conclusions	152
	Appendix A Additional qOMM and SSVQE Convergence Plots	157
	Appendix B Excited States Initializations and Ansatz Expressiveness	161
	Appendix C CIS State Preparation	163
C.0.1	Example 1: 3 particles, 6 spin-orbitals	165
C.0.2	Example 2: Sparse 2 Particles, 6 Spin-Orbitals	168
C.0.3	General Procedure	170
	Bibliography	174
	Biography	186

List of Tables

4.1	The convergence success rates of qOMM and SSVQE for H_2 , H_4 , and LiH using random parameter initializations.	80
B.1	The final accuracy of the average energy for H_4 for given choices of eigensolver, initialization, and UCCSD ansatz expressiveness.	162

List of Figures

4.1	Convergence (noise-free) of the relative error of qOMM and SSVQE for 3 H ₂ states at an interatomic distance of 0.735 Å.	82
4.2	Convergence (noise-free) of the relative error of qOMM and SSVQE for H ₄ at an interatomic distance of 1.23 Å using a random parameter initialization.	82
4.3	Base circuit block pattern used for noisy H ₂ simulations.	83
4.4	Convergence of the relative error of qOMM and SSVQE for H ₂ at an interatomic distance of 0.735 Å in the presence of depolarizing noise.	84
4.5	Convergence of the relative error of qOMM and SSVQE for LiH.	85
4.6	Relative accuracy of the convergence of the first three energy levels of LiH for SSVQE using a random parameter initialization.	87
4.7	Convergence of the relative error of VQE for the ground state problem of LiH using a random parameter initialization.	89
5.1	Convergence of OptOrbVQE as a function of the outer loop iteration number for H ₄ at the near-equilibrium H-H distance of 1.23 Å.	105
5.2	Convergence of OptOrbVQE as a function of the outer loop iteration number for LiH at the near-equilibrium bond distance of 1.595 Å.	106
5.3	Dissociation curve of LiH calculated using OptOrbVQE.	106
5.4	Convergence of OptOrbVQE as a function of the outer loop iteration number for BeH ₂ at the near-equilibrium Be-H bond distance of 1.3264 Å.	108
5.5	Convergence of OptOrbVQE for H ₂ O at the near-equilibrium O-H bond distance of 0.9578 Å and bond angle 104.4776 degrees.	110
5.6	Converged energy of OptOrbVQE as a function of the number of spin orbitals for H ₂ at the near-equilibrium bond distance of 0.735 Å.	112
5.7	Binding curve of H ₂ using 10 ⁶ circuit samples.	114
5.8	Ansatz used for H ₂ simulations with depolarizing noise.	115
5.9	Binding curve of H ₂ using 10 ⁶ circuit samples with $p_{error} = 10^{-3}$	115
6.1	Convergence of orbital optimization methods for H ₂ using increasing numbers of spin-orbitals.	128
6.2	Convergence of excited states orbital optimization methods for H ₄ using 8 optimized spin-orbitals.	129

6.3	Convergence of excited states orbital optimization methods for LiH.	131
6.4	Convergence of excited states orbital optimization methods for BeH ₂	133
6.5	The potential energy surface of the first two energy levels of H ₄ calculated using OptOrbMCVQE.	134
7.1	Binding curve of the ground state of H ₄ calculated using OptOrbVQE and OptOrbAdaptVQE.	142
7.2	Measures of the Hamiltonian sparsity as a function of the outer loop iteration for OptOrbVQE for various interatomic distances for H ₄	143
7.3	The measurement overhead factor as a function of the outer loop iteration for OptOrbVQE for various interatomic distances for H ₄	144
7.4	Measures of the circuit depth for the optimal circuit found by ADAPT-VQE as a function of the outer loop iteration for various interatomic distances.	145
7.5	Convergence of the ground state energy of LiH for OptOrbVQE and OptOrbAdaptVQE as a function of the outer loop iteration.	147
7.6	The circuit depth of the optimal circuit found by OptOrbAdaptVQE at each outer loop iteration for LiH.	148
7.7	Convergence of the ground state energy of LiH for OptOrbVQE and OptOrbAdaptVQE as a function of the outer loop iteration.	149
7.8	The number of Pauli strings in the Hamiltonian for OptOrbVQE and OptOrbAdaptVQE for each outer loop iteration for LiH.	149
7.9	The measurement overhead factor for OptOrbVQE and OptOrbAdaptVQE for each outer loop iteration.	150
A.1	Convergence of the relative error of qOMM and SSVQE for H ₂ using a random parameter initialization.	157
A.2	Convergence of the relative error of qOMM and SSVQE for H ₂ at an interatomic distance of 0.735 Å with depolarizing noise.	158
A.3	Convergence of the relative error of qOMM and SSVQE for H ₄ at an interatomic distance of 1.23 Å using a random parameter initialization.	158
A.4	Convergence of the relative error of qOMM and SSVQE for H ₄ at an interatomic distance of 2.46 Å using a random parameter initialization.	159
A.5	Convergence of the relative error of qOMM and SSVQE for LiH using randomly initialized parameters.	160
B.1	Convergence of the state-averaged energy accuracy of MCVQE for H ₄	162

Acknowledgements

I would first like to thank my advisor, Professor Jianfeng Lu for his continued guidance and support throughout my graduate career. Having graduated with degrees in both mathematics and physics but being forced to choose one field for graduate school, I came to Duke looking for a research area that would provide the right balance between these two fields. I was unsure at the time what the right balance for me would be or which physics subfield would best provide this. He suggested that we start working on developing quantum algorithms for solving quantum many-body problems. Quantum information and computing lies at the fortuitous intersection of many different fields including mathematics, physics, chemistry, and computer science. The benefit this gives a researcher in this field is that there is opportunity to shift one's focus more in the direction of one of these fields or their respective subfields in response to shifting research interests. I thank Professor Lu for his suggestion that set me on a career path that would provide the optimal balance between my various interests. I would also like to thank our collaborator, Professor Yingzhou Li. His continued guidance throughout my graduate career and insights during our weekly discussions have been invaluable.

I would also like to thank my committee for their guidance. When doing research, it is easy to get bogged down in the details which are relevant to making progress on a day-to-day basis at the expense of forgetting about the big picture which motivates the work in the first place. Their questions during my preliminary exam helped to remind me not to forget about the bigger picture behind the motivation of why we do what we do. Additionally, Professor Barthel's class on open quantum systems has been invaluable in understanding the models which are used to understand the error-prone nature of quantum computers. Professor Brown's QEC class has also been an immense help in de-mystifying quantum error correction. While our research on near-term algorithms does not necessarily require us to think about error correction codes on a daily basis, having a baseline understanding of this subfield helps to understand the principles which guide the progress of the field of quantum computing as a whole and how our work fits into that long-term big picture.

Finally, I would like to thank my family. Their continued support is what has made all of this possible.

1. Introduction

The structure of this introduction is as follows. In Sec. 1.1 we introduce the electronic structure problem. Sec. 1.2 briefly summarizes classical methods in the literature for solving the electronic structure problem which are most relevant to later discussions on quantum algorithms and our work. Sec. 1.3 summarizes the basics of quantum computing and introduces notation. Sec. 1.4 introduces the steps involved in setting up the electronic structure problem on a quantum computer.

1.1 The Electronic Structure Problem

In the electronic structure problem,^{1,2} we are interested in finding the eigenvalues E_k and eigenstates $|\psi_k\rangle$ which comprise solutions to the time-independent Schrödinger equation:

$$\hat{H}|\psi_k\rangle = E_k|\psi_k\rangle \quad (1.1)$$

for a non-relativistic Hamiltonian \hat{H} describing the Coulomb interactions between a set of electrons and nuclei. We can write down this Hamiltonian in atomic units as:

$$\hat{H} = -\sum_i \frac{1}{2} \nabla_i^2 - \sum_I \frac{1}{2} \nabla_I^2 + \sum_{i,j>i} \frac{1}{r_{ij}} + \sum_{i,I} \frac{Z_I}{r_{iI}} + \sum_{I,J>I} \frac{Z_I Z_J}{R_{IJ}} \quad (1.2)$$

where lower case letters index the electrons and upper case letters index the nuclei. ∇_i^2 and ∇_I^2 are the coordinate space Laplacians representing the momentum operators of the i th electron and I th nuclei, respectively. $r_{ij} = |\vec{r}_i - \vec{r}_j|$ is the distance between the i th and j th electrons in position space. $r_{iI} = |\vec{r}_i - \vec{r}_I|$ is the distance between the i th electron and I th nucleus. $R_{IJ} = |\vec{r}_I - \vec{r}_J|$ is the distance between the I th and J th nuclei. Z_I is the atomic number of the I th nucleus.

The eigenvalues and eigenvectors found by solving the time-independent Schrödinger equation for this Hamiltonian often serve as the starting point for a much broader study of the properties of molecular and solid state systems. For example, in molecular chemistry, one is often interested in predicting which reactions can occur, the rate at which they

occur at a given temperature and pressure, and the mechanism by which they occurred. A chemist may want to design a catalyst which allows a particular reaction to occur quickly or a pharmaceutical company may want to know the reaction pathway by which the synthesis of a particular drug is both possible and efficient. A materials chemist may want to design new materials for solar cells, in which case its energy gaps will govern its interactions with light. A biological chemist may want to characterize the geometry of an active site of a protein to determine its biological function. Even for molecular systems much simpler than proteins, subtle differences in the geometry of molecules with the same stoichiometric composition and atomic connectivity (*i.e.* stereoisomers) can give rise to drastically different properties. A well-known example of this phenomenon comes from food chemistry. The carvone molecule, which has two possible stereoisomers that are mirror images of each other, will have distinctly different scents depending on which stereoisomer is present. One is responsible for the smell of spearmint and the other for the smell of dill and caraway seeds.³ Clearly, these two mirror image molecules interact quite differently with our smell receptors. In all of these examples, one could gain insight from a description of the potential energy surface of the system in question and its wavefunction at various points. That is, we solve the time-independent Schrödinger equation for various numerical values of the spatial degrees of freedom of the molecule and map out the energy as a function of these degrees of freedom. Local minima of this potential energy surface correspond to potentially stable (or metastable) geometries that could be observed. A molecule with many degrees of freedom could have several such local minima, each with potentially different properties. These local minima may be connected by a path on the potential energy surface with an intermediate geometry that is higher in energy than either of them, *i.e.* an energy barrier. The difference in energies between these two local minima will determine the ratios of these geometric configurations that would exist in equilibrium and the height of this barrier will influence how quickly a transition from one configuration to another will occur.⁴ A chemical reaction may occur for one geometry but not the other, in which case a chemist would be interested in this potential energy surface path and the relative

energies of the various possible configurations. Similarly, a chemical reaction may involve an unstable or metastable transition state which is higher in energy than both the reactants and products. The difference in Gibbs free energy between the reactants and transition state appears in the Eyring equation for the rate of this reaction. In order to estimate this reaction rate to within an accuracy of one order of magnitude (at room temperature), one needs to estimate the energies of the molecules involved to an accuracy of at least 1.6×10^{-3} Hartree.⁵ This accuracy is referred to as chemical accuracy in the literature. This is not a strict requirement and could either be excessive or insufficient depending on the setting, but it provides a chemically-motivated rule of thumb that is often used as a target to assess the quality of various computational methods. Usually when we refer to the potential energy surface of a system, we implicitly mean that of the electronic ground state. However, there are situations in which we may also want the potential energy surface of one or more of the low-lying excited states. An obvious example would be the study of photochemical reactions that are mediated by the absorption of a photon by one of the reactants.⁶ Transitions between the ground and excited states can however, occur without the absorption or emission of a photon. This can occur at points in the potential energy surface where the ground and first excited state intersect, known as conical intersections. The ability to predict the presence and location of these conical intersections is known to be important for predicting many reaction pathways which are mediated by their presence.⁷

To-date, all useful calculations of the properties of molecules have been done on classical computers using classical algorithms. The downside to using a classical computer is that the amount of computational resources involved in solving the time-independent Schrödinger equation numerically exactly scales exponentially with the size of the system, limiting this treatment to small system sizes such as diatomic and triatomic molecules with light nuclei. This is due to the dimensionality of the Hamiltonian and states in its Hilbert space growing exponentially with the size of the system. For most applications, classical calculations are forced to use approximate methods which use various heuristics to reduce the dimensionality of the problem. In using such heuristic methods we not only lose guar-

antees that any given method will offer sufficient accuracy for any given system, but we also lack rigorous bounds for the errors of these methods. This can make it difficult to assess the validity of their results unless we have accurate experimental results to which we can compare them. Quantum computers have the potential to avoid this curse of dimensionality without the use of heuristics because they store and process information on collections of 2-level quantum systems called qubits. A register of N qubits can store a quantum state that is a linear combination of 2^N basis states. This simple observation by itself does not automatically imply that a quantum computer can efficiently solve the electronic structure problem, but we do have good reasons to be optimistic. For example, we do have quantum algorithms (quantum phase estimation methods⁸) which can solve the time-independent Schrödinger equation efficiently to arbitrary accuracy with rigorous error bounds, provided that certain assumptions are met. The drawback to these quantum algorithms is that the amount of computational resources involved are beyond that which quantum computers will have in the near-term. We also have variational quantum algorithms such as VQE⁹ which use fewer resources and which retain the efficient preparation of quantum states, but which also re-introduce some of the heuristics and lack of rigorous error bounds from classical methods that we sought to avoid. The primary goal for the quantum computing community is to refine these methods to reduce the amount of resources required to use them to solve useful problems. Another goal is to find relevant chemical systems whose size puts them beyond the reach of the exact classical treatment and whose complexity causes the assumptions involved in classical approximate methods to break down. Finding the smallest possible systems for which this occurs is not necessarily trivial and is currently an active area of research. For example the Cr₂ molecule (Chromium dimer) has been identified as a deceptively simple (stoichiometrically and geometrically speaking) molecule which classical methods struggle to treat accurately.¹⁰ Chromium dimer is the smallest system for which demonstrating a plausible quantum advantage has been speculated, but it is primarily an academic curiosity with no practical relevance. A more industrially relevant system which highlights a deficiency in the capabilities of classical methods and for which

plausible quantum advantage has been speculated is the active site complex of the nitrogenase enzyme, FeMoco.¹¹ This enzyme acts as a catalyst in the biological conversion of atmospheric nitrogen to a form which plants can metabolize. The mechanism by which this occurs has eluded description thus far. It is hoped that an accurate understanding of this mechanism may allow for catalysts to be designed which improve upon the Haber-Bosch process currently used to convert nitrogen to fertilizer on an industrial scale.¹² Thus, the benefits of elucidating this mechanism are potentially immense as it could allow for the energy-efficient production of the amount of fertilizer needed to feed a global population of eight billion people. It has been proposed that modeling FeMoco may be a promising target problem for quantum computers,¹¹ although this early proposal has been criticized for being too optimistic in terms of the computation resources that would be required for this problem.¹³ Iron-Sulfur clusters such as 4Fe-4S and 2Fe-2S have also drawn attention in the quantum computing community for similar reasons as FeMoco.^{14,15} They too are present on the nitrogenase enzyme¹² and could play a role in the conversion of atmospheric nitrogen to ammonia. We should take note of what all these candidate systems have in common: they all contain multiple transition metals. It is therefore reasonable to conjecture that other transition metal complexes (especially those with multiple transition metals) may be a promising class of molecules to investigate the potential utility of quantum computers. FeMoco (and to a lesser extent, iron-sulfur clusters) has drawn much attention due to the high-profile nature of the problem it is involved in. If a quantum computer enabled the elucidation of the nitrogen fixation mechanism by nitrogenase which then in turn enabled the discovery of a more energy-efficient means to produce fertilizer, it would be difficult to understate the impact that innovation would have. However, we should not become so fixated on nitrogenase that we forget that there is an entire field of transition metal catalysis chemistry for which quantum computers may also find practical application in lower-profile problems that have to-date not received as much attention. In this area of study, it is difficult to pinpoint exactly one target system of interest because a research group may have a large number of candidate catalysts that they want to simulate in order

to narrow down the field. For example, one may want to predict which candidate catalyst produces the fastest reaction rate or the highest yield. Certainly in the case of predicting reaction rates, chemical accuracy could be seen as the bare minimum accuracy for which to aim. Given that chemical accuracy only ensures accuracy of reaction rates to within one order of magnitude, it may be necessary to achieve even higher accuracy for such problems. It is these types of problems which necessitate high accuracy solutions where the potential advantage of quantum computers becomes most promising, given that there are already a plethora of classical approximate methods which can readily produce low accuracy results. This use of quantum computers to design better transition metal catalysts to carry out various reactions more efficiently has been suggested by Elfving et. al¹³ and investigated by von Burg et. al.¹⁶ In the work by von Burg et. al, they estimate the computational resources needed to estimate the energies of molecular species involved in the catalytic conversion of carbon dioxide to methanol on a quantum computer. That is, they are looking at using quantum computers to solve problems that are directly relevant to carbon capture technologies. They note that for the ruthenium-based catalysts considered in their work, the mechanism by which this conversion occurs has already been elucidated, but that such ruthenium-based catalysts suffer from low turnover number. Thus, finding more practical catalysts for problems such as carbon capture is an ongoing computational challenge for which quantum computers may find utility.

We begin our solution of the electronic structure problem by applying the Born-Oppenheimer approximation,² where we assume we can consider the positions of the nuclei to be fixed and that the entire wavefunction can be written as a tensor product of the nuclear and electronic subsystem wavefunctions. The intuition behind this is that because the masses of the nuclei are several orders of magnitude greater than those of the electrons, any Coulomb interaction between a nucleus and an electron will have a far greater effect on the position of the electron than that of the nucleus. With the second term in Eq. 1.2 now contributing nothing to the total energy and the fifth now being taken to be a constant nuclear repulsion energy E_{nuc} , we can write down the electronic structure Hamiltonian as:

$$\hat{H} = -\sum_i \frac{1}{2} \nabla_i^2 + \sum_{i,j>i} \frac{1}{r_{ij}} + \sum_{i,I} \frac{Z_I}{r_{iI}} \quad (1.3)$$

keeping in mind that we will have to add E_{nuc} to its eigenvalues if we want the total energy. Given that this Hamiltonian describes a fermionic many-body system, we must enforce the anti-symmetry of its wavefunction. In the second-quantization formulation this anti-symmetry is enforced at the level of operators. The wavefunctions themselves are defined in terms of fermionic creation and annihilation operators acting on a vacuum state $|0\rangle$ with no particles. These creation and annihilation operators can be functions of either continuous variables (such as position space coordinates), or discrete variables (such as spin), or both. Here we adopt the shorthand notation $\vec{x} = (\vec{r}, m)$ where $m = \pm\frac{1}{2}$ is the spin-magnetization quantum number and \vec{r} is the position-space coordinate. The creation and annihilation operators $\hat{\psi}^\dagger(\vec{x})$ and $\hat{\psi}(\vec{x})$ obey the anticommutation relations:¹⁷

$$\{\hat{\psi}^\dagger(\vec{x}), \hat{\psi}^\dagger(\vec{x}')\} = \{\hat{\psi}(\vec{x}), \hat{\psi}(\vec{x}')\} = 0 \quad (1.4a)$$

$$\{\hat{\psi}(\vec{x}), \hat{\psi}^\dagger(\vec{x}')\} = \delta^3(\vec{r} - \vec{r}') \delta_{m,m'}. \quad (1.4b)$$

The Hamiltonian in Eq. 1.3 can be written in terms of these creation and annihilation field operators as:^{17,18}

$$\begin{aligned} \hat{H} = & \sum_m \int d^3\vec{r} \hat{\psi}^\dagger(\vec{x}) \left(-\frac{1}{2} \nabla^2 + \sum_I \frac{Z_I}{|\vec{r} - \vec{r}_I|} \right) \hat{\psi}(\vec{x}) \\ & + \frac{1}{2} \sum_{m,m'} \int d^3\vec{r} d^3\vec{r}' \hat{\psi}^\dagger(\vec{x}) \hat{\psi}^\dagger(\vec{x}') \frac{1}{|\vec{r} - \vec{r}'|} \hat{\psi}(\vec{x}') \hat{\psi}(\vec{x}). \end{aligned} \quad (1.5)$$

We now expand the field operators in terms of a complete set of orthonormal single-particle wavefunctions $\{\phi_n(\vec{x})\}$ called orbitals:

$$\hat{\psi}(\vec{x}) = \sum_n \phi_n(\vec{x}) \hat{a}_n \quad (1.6a)$$

$$\hat{\psi}^\dagger(\vec{x}) = \sum_n \phi_n^*(\vec{x}) \hat{a}_n^\dagger \quad (1.6b)$$

where \hat{a}_n^\dagger and \hat{a}_n create and annihilate an electron in orbital $\phi_n(\vec{x})$, respectively. These fermionic creation and annihilation operators obey the anticommutation relations:¹

$$\{\hat{a}_i, \hat{a}_j\} = \{\hat{a}_i^\dagger, \hat{a}_j^\dagger\} = 0 \quad (1.7a)$$

$$\{\hat{a}_i, \hat{a}_j^\dagger\} = \delta_{i,j}. \quad (1.7b)$$

When an orbital is a function of both position and spin, it is referred to as a spin-orbital. The position-space component of a spin-orbital (or an orbital which is a function of only position) is referred to as a spatial orbital. In general this expansion will need to be infinite for the wavefunctions to constitute a complete basis. In practice any computer (classical or quantum) will have a finite amount of resources, thus we are only capable of using a finite number of orbitals, inducing a basis set truncation error in the eigenvalues of \hat{H} . We must choose $\{\phi_n(\vec{x})\}$ carefully so as to minimize the computational resources involved while also keeping this truncation error to an acceptable level. Substituting the expansion in Eq. 1.6 into Eq. 1.5, we get:

$$\hat{H} = \sum_{p,q} h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{pqrs} v_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_s \hat{a}_r \quad (1.8)$$

where:

$$h_{pq} = \int d^3\vec{r} \phi_p^*(\vec{r}) \left(-\frac{1}{2} \nabla^2 + \sum_I \frac{Z_I}{|\vec{r} - \vec{r}_I|} \right) \phi_q(\vec{r}) \quad (1.9a)$$

$$v_{pqrs} = \int d^3\vec{r} d^3\vec{r}' \phi_p^*(\vec{r}) \phi_q^*(\vec{r}') \frac{1}{|\vec{r} - \vec{r}'|} \phi_s(\vec{r}') \phi_r(\vec{r}) \quad (1.9b)$$

are the one and two body integrals h_{pq} and v_{pqrs} , respectively.¹ Note that we have absorbed the summation over the spin indices m and m' into the summation over the indices p, q, r, s for the spin-orbitals to keep the notation compact. With \hat{H} now discretized and

finite-dimensional, we can begin to solve the problem at hand. We can construct many-body wavefunctions from a set of N single-particle wavefunctions $\phi_p(\vec{x})$ by considering basis states where each orbital has a well-defined occupation number n_p of either 0 or 1. These occupation number states are defined in terms of the action of the fermionic creation operators on the state with no electrons:^{1,17}

$$\prod_{p=0}^{N-1} (\hat{a}_p^\dagger)^{n_p} |0\rangle = |n_0, n_1, \dots, n_{N-1}\rangle. \quad (1.10)$$

These states are often referred to as Slater determinants. Under the assumption that the set of spin-orbitals that we have chosen is orthonormal, these Slater determinants are also orthonormal. The set of all possible Slater determinants spans the Hilbert space of \hat{H} and can therefore be used as a basis.

1.2 A Brief Summary of Classical Methods

The study of classical algorithms for solving the electronic structure problem has been developed over many decades, resulting in an enormous number of methods, each of which has several variants. A proper discussion on all of them could span several literature reviews. Here we will content ourselves with a brief discussion on the most basic versions of the methods which are most often mentioned in works on solving the electronic structure problem on quantum computers and which will be most relevant when discussing our own work.

1.2.1 Hartree-Fock Methods

Hartree-Fock refers to a class of approximate methods where the ground state is approximated as a single Slater determinant in some single-particle basis.^{1,2} We start with a set of atomic orbitals (AOs) centered around the nuclei that have a known analytical form. For molecular systems, these atomic orbitals typically take a form that is loosely analogous to solutions of the Schrödinger equation for the hydrogenic atom. That is, an atomic orbital is written as a product of an angular part and a radial part, where the radial part depends

on a principle quantum number n and an orbital angular momentum quantum number l . The angular part (usually taken to be a spherical harmonic function, a solid harmonic function, or a Cartesian coordinate polynomial of the form $x^a y^b z^c$) depends on l as well as the quantum number m_l representing the component of the orbital angular momentum along a particular axis. These quantum numbers follow the same rules as those for the solutions for the hydrogenic atom. That is, $n = 1, 2, \dots, \infty$, $l = 0, 1, \dots, n - 1$, and $-l \leq m_l \leq l$. The radial part is usually taken to be either a Gaussian or a linear combination of Gaussians. Single-particle basis sets can be constructed by selecting any number of orbitals with forms dictated by these quantum numbers according to various rules.¹ Popular basis sets include minimal basis sets such as STO-3G,¹⁹ split-valence basis sets such as 6-31G²⁰ and correlation-consistent basis sets such as cc-pVXZ (for X = D, T, Q, 5, ...).²¹ Much of the difficulty of computational chemistry comes from the fact that quite large basis sets (*e.g.* cc-pVQZ or cc-pV5Z) are almost always required to obtain quantitatively accurate and useful results. Modestly-sized basis sets such as STO-3G or 6-31G are typically used for demonstration purposes or to capture some qualitatively correct results.

The goal of Hartree-Fock methods is to output a single Slater determinant using an orbital set consisting of linear combinations of these AOs which best approximates the ground state energy of \hat{H} . These linear combinations of atomic orbitals are often called molecular orbitals (MOs). In general the true ground state will be a linear combination of such Slater determinants, thus Hartree-Fock is a crude approximation that is often a starting point for more sophisticated methods. The benefit of using Hartree-Fock in this way is threefold: **1.** it can be readily applied to large systems; **2.** the resulting MOs often form a more compact basis set representation of the Hamiltonian than the starting AOs and; **3.** the resulting Hartree-Fock wavefunction often serves as a reasonable initial guess for the wavefunction in more sophisticated methods. Hartree-Fock methods can be loosely grouped into two types: variational methods and self-consistent field methods.^{1,2} In the variational approach, we first start with the Hamiltonian \hat{H} and a Slater determinant $|\psi_{sd}\rangle$ in a basis of AOs. We consider unitary basis set transformations of the form:

$$\hat{U} = e^{-\hat{\kappa}} \quad (1.11a)$$

$$\hat{\kappa} = \sum_{p>q} \kappa_{pq} (\hat{a}_p^\dagger \hat{a}_q - \hat{a}_q^\dagger \hat{a}_p) \quad (1.11b)$$

and perform the minimization problem:

$$\min_{\{\kappa_{pq}\}} \langle \psi_{sd} | \hat{U}^\dagger \hat{H} \hat{U} | \psi_{sd} \rangle \quad (1.12)$$

where κ is an antisymmetric matrix with dimension equal to the number of orbitals. Note that the operator $\hat{\kappa}$ is not the same quantity as κ . $\hat{\kappa}$ is an operator acting on states in the Hilbert space of \hat{H} , whereas κ is a matrix storing the parameters of $\hat{\kappa}$. When an optimal κ , κ_{opt} is found, the Hartree-Fock wavefunction is given as:

$$|HF\rangle = e^{-\hat{\kappa}_{opt}} |\psi_{sd}\rangle \quad (1.13)$$

and the AOs are transformed to optimized MOs as:¹

$$\phi_p^{MO} = \sum_p [e^{-\kappa_{opt}}]_{qp} \phi_q^{AO}. \quad (1.14)$$

Self-consistent field methods take a mean-field approach where the two-body Coulomb interaction terms in the Hamiltonian in Eq. 1.8 are replaced with effective potentials where each electron is subject to a one-body potential that has a dependence on which orbitals are occupied in the Hartree-Fock state. This effective Hamiltonian is known as the Fock operator and is given by:¹

$$\hat{f} = \sum_{p,q} h_{pq} \hat{a}_p^\dagger \hat{a}_q + \sum_{p,q} \sum_i (2v_{pqii} - v_{pqi}) \hat{a}_p^\dagger \hat{a}_q \quad (1.15)$$

where the index i runs over orbitals which are occupied in the Hartree-Fock state. We note that because this operator contains only one-body terms, its eigenstates can be taken to be single-particle wavefunctions, *i.e.* orbitals. The state of the electrons are effectively treated

as independent particles, subject to the constraint that they are affected by an effective potential generated by all of the electrons in the Hartree-Fock state. If we could diagonalize this operator, we could generate a valid approximation to the ground state by filling the eigenstates of \hat{f} with the lowest eigenvalues with electrons. However, there is one subtlety to take note of: because the index i runs over orbitals which are occupied in the Hartree-Fock state, this operator is defined in terms of its own eigenvectors and eigenvalues. How can we begin to solve a problem which is formulated in terms of its own solution? The self-consistent field method takes an iterative approach. We start with an initial set of AOs and a guess for which orbitals are occupied. Call this operator \hat{f}_0 since this is the 0th iteration. We can diagonalize \hat{f}_0 to obtain its eigenvalues and eigenvectors. These updated eigenvectors will be linear combinations of the original AOs and we can recompute the one and two body integrals h and v in the basis of these new orbitals. We can now recompute an updated operator \hat{f}_1 by having the index i run over those of the new orbitals with the lowest eigenvalues up to the number of electrons in the system. Now we diagonalize \hat{f}_1 and repeat this process until the orbitals and their eigenvalues have converged. We formulate the solution Hartree-Fock state as a Slater determinant where the orbitals with the lowest energy are filled.¹

1.2.2 Configuration Interaction

Configuration Interaction (CI) methods are one step above Hartree-Fock in terms of sophistication and accuracy. In Hartree-Fock, the ground state was approximated as a single Slater determinant, which in general will only be a crude approximation. CI methods represent wavefunctions as linear combinations of Slater determinants.^{1,2} We are free to choose which Slater determinants appear in this expansion and are left to solve for their coefficients to obtain the wavefunction. The more determinants we include, the more accurate the results will be at the expense of greater computational cost. Not all determinants will contribute equally to the final energy accuracy, so we must be somewhat strategic in how we choose them. While any method that takes this general approach would fall un-

der the umbrella of CI methods, it is customary to start with a reference state (usually the Hartree-Fock state $|HF\rangle$) and generate the Slater determinants used in the calculation from fermionic excitation operators acting on this reference state. For example, a Slater determinant resulting from a single-particle excitation would be given by:

$$|\psi_{sd}\rangle = \hat{a}_p^\dagger \hat{a}_q |HF\rangle \quad (1.16)$$

where ϕ_p and ϕ_q would be unoccupied and occupied in $|HF\rangle$, respectively. In general we can generate determinants from exciting any number of electrons up to the amount in the system, categorizing them according to the degree of the excitation. Denote the set of integers which index orbitals occupied and occupied in $|HF\rangle$ by \mathcal{O} and \mathcal{U} , respectively. The wavefunction is expanded as:^{1,2}

$$|\psi_{CI}\rangle = C_{HF} |HF\rangle + \sum_{\substack{i \in \mathcal{O} \\ a \in \mathcal{U}}} C_i^a \hat{a}_a^\dagger \hat{a}_i |HF\rangle + \sum_{\substack{a > b \in \mathcal{U} \\ i > j \in \mathcal{O}}} C_{ij}^{ab} \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_a \hat{a}_b |HF\rangle + \dots \quad (1.17)$$

where it is implied that this expansion can continue beyond single and double-particle excitations. CI methods which include only single excitations are denoted as CIS, those that include single and double excitations are denoted as CISD, those that include up to triple excitations are denoted as CISDT, and so on. Methods which include all possible excitations are called Full Configuration Interaction (FCI) methods. FCI methods correspond to exact diagonalization of the Hamiltonian in the chosen finite basis set. In this sense FCI is numerically exact within the confines of the model (*i.e.* in a finite truncated basis and using the Born-Oppenheimer approximation, which will both introduce errors). However, FCI calculations are limited to small systems due to the fact that the number of determinants scales combinatorially with the number of orbitals and electrons. Thus, most of classical computational chemistry focuses on attempting to get as close to the FCI energy as possible using heuristic approximations and clever tricks. We can generate a truncated version of \hat{H} in the subspace spanned by the set of Slater determinants by using the Slater-Condon

rules to compute the matrix elements:^{1,2}

$$\langle \psi_{sd,1} | \hat{H} | \psi_{sd,2} \rangle \quad (1.18)$$

for pairs of Slater determinants $|\psi_{sd,1}\rangle$ and $|\psi_{sd,2}\rangle$. Using this contracted Hamiltonian and vectorizing the representation of the CI coefficients as \vec{c} , we can solve the eigenvalue problem:

$$\hat{H}\vec{c} = E_n\vec{c} \quad (1.19)$$

for the eigenvectors and their eigenvalues. This can be accomplished either through direct diagonalization or through variational methods which use \vec{c} as the set of parameters over which to minimize an objective function.¹ For example, in order to find the ground state E_0 , we can invoke the variational principle:²²

$$\frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \geq E_0 \quad (1.20)$$

which states that for any state $|\psi\rangle$ in the Hilbert space of \hat{H} , the expectation value of \hat{H} is lower-bounded by the ground state energy E_0 . Any CI wavefunction is in the Hilbert space of \hat{H} , thus techniques which iteratively minimize the expectation value of \hat{H} with respect to \vec{c} would be valid approaches to solving this problem.

1.2.3 Coupled-Cluster Methods

In computational chemistry, the term *ansatz* typically refers to a parameterized guess as to what form the solution wavefunction should take for a particular problem instance. We do not know what the wavefunction is *a priori*, but we conjecture that there should be some numerical values of the parameters for which the ansatz gives a reasonable approximation to the solution. In this sense the CI wavefunction in Eq. 1.17 is an example of ansatz. We guess that the wavefunction should be well-approximated by a linear combination of

fermionic excitation operators acting on a reference state and solve for their coefficients. Coupled-cluster methods^{1,5,23} adopt the following ansatz for the solution wavefunction:

$$|\psi_{cc}\rangle = e^{\hat{T}}|HF\rangle \quad (1.21a)$$

$$\hat{T} = \sum_{\substack{i \in \mathcal{O} \\ a \in \mathcal{U}}} t_i^a \hat{a}_a^\dagger \hat{a}_i + \sum_{\substack{i > j \in \mathcal{O} \\ a > b \in \mathcal{U}}} t_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_i \hat{a}_b^\dagger \hat{a}_j + \dots \quad (1.21b)$$

where it is implied that this summation can go up to the number of electrons in the system. The expansion of this exponential excitation operator is typically truncated at a specific excitation level in the same way as CI methods. Coupled-cluster singles and doubles is referred to as CCSD, coupled-cluster containing up to triple excitations is referred to as CCSDT, coupled-cluster containing up to quadruple excitations is referred to as CCSDTQ, and so on. Coupled-cluster where one of the degrees of excitations is handled approximately using perturbation theory is denoted by encapsulating the letter corresponding to that excitation level in parentheses. For example, CCSD(T) corresponds to standard coupled-cluster up to double excitations and the triple excitations are treated approximately using perturbation theory.¹ One advantage this ansatz has over its CI counterpart is that there is a sense in which it is more compact. For example, in the case of CISD it is not possible for the wavefunction to contain any Slater determinants expressible as any degree of excitation more than 2 above the reference state. This is explicitly baked into its construction. On the other hand, the exponential nature of the CCSD ansatz allows for the solution state to contain Slater determinants expressible as excitations of degree higher than 2. This is because one typically approximates the exponential ansatz as a sum of products of individual exponential operators (*e.g.* with a BCH or Trotter expansion). Products of lower degree excitations can give rise to higher degree excitation terms appearing in the wavefunction. Because of this, coupled-cluster often attains a higher degree of accuracy in the solution energy compared to configuration interaction methods of the same excitation degree. Furthermore, we can intuit how this property may lead to the accuracy of CC being less severely impacted

than that of CI when the number of electrons is increased.¹ These properties have allowed for coupled-cluster methods to be highly successful in molecular chemistry; it can be applied to larger systems than FCI due to its polynomially number of parameters while still achieving a reasonable degree of accuracy for many chemical systems. For this reason CCSD(T) is often referred to as the "gold standard" in classical computational chemistry. One known limitation of couple-cluster methods is that because the set of excitation operators in the expansion of \hat{T} is defined in terms of the occupied orbitals in the reference state $|HF\rangle$ upon which it acts, the accuracy of coupled-cluster methods may break down for systems where the reference state is a particularly poor approximation of the FCI wavefunction.¹

The task at hand is then to solve the Schrödinger equation for this ansatz state:

$$\hat{H}e^{\hat{T}}|HF\rangle = E_{cc}e^{\hat{T}}|HF\rangle. \quad (1.22)$$

or equivalently:

$$e^{-\hat{T}}\hat{H}e^{\hat{T}}|HF\rangle = E_{cc}e^{-\hat{T}}\hat{H}e^{\hat{T}}|HF\rangle. \quad (1.23)$$

This is solved by projecting Eq 1.23 onto the set of Slater determinants expressible as a fermionic excitation on the reference state with degree up to that included in \hat{T} . For example, for CCSD we would project onto all Slater determinants included in a CISD expansion. Let us denote such states by $|\psi_{sd,i}\rangle$. This leads to the set of equations:

$$\langle\psi_{sd,i}|e^{-\hat{T}}\hat{H}e^{\hat{T}}|HF\rangle = 0 \quad (1.24a)$$

$$\langle HF|e^{-\hat{T}}\hat{H}e^{\hat{T}}|HF\rangle = E_{cc} \quad (1.24b)$$

parameterized by the set of coefficients $\{t_i^a, t_{ij}^{ab}, \dots\}$ in Eq. 1.21b. This can be achieved through iterative schemes such as Newton or quasi-Newton methods.¹

1.2.4 Orbital Optimization Methods

In the previous methods discussed we started with a set of AOs, rotated these orbitals to obtain a set of MOs as linear combinations of the AOs, then performed an eigenvalue problem using a set of Slater determinants constructed from these MOs as a basis. Truncated CI and CC methods approximated the wavefunction by taking into account only a subset of excitation operators acting on the Hartree-Fock state. There is another way we can control the dimensionality of the wavefunction: we can perform the eigenvalue problem using only a subset of these MOs as a basis. In quantum chemistry this is called choosing an active space or making an active space approximation. The intuition behind this is simple. For many chemical systems we expect that for a subset of the spatial MOs, the expectation value of their occupation number will be approximately 2 (one electron in the spin-up orbital and one in the spin-down orbital). These are denoted as core orbitals which form the inactive space. Similarly, we expect that a subset of the orbitals will have expected occupation numbers to be approximately 0. These are called the virtual orbitals and form the virtual space. The orbitals for which neither of these assumptions is likely to be approximately true are called the active orbitals and form the active space. We then make the approximation that the total wavefunction can be written as a tensor product of states occupying these spaces separately:

$$|\psi\rangle = |\psi_c\rangle \otimes |\psi_a\rangle \otimes |\psi_v\rangle \quad (1.25)$$

and choose an eigensolver which solves for the active space component $|\psi_a\rangle$. The simplest way to choose these spaces would be to partition the MOs according to their Hartree-Fock energies. According to this heuristic, the lowest energy orbitals are the ones most likely to be doubly occupied and the highest energy orbitals are the ones most likely to have an occupation number of zero. However, we should keep in mind that the Hartree-Fock MOs and their energies were found by diagonalizing an effective Hamiltonian that treated the two-body interaction terms in an approximate mean-field manner. Thus, there is no

reason to believe that choosing an active space in this way is optimal. Ideally, we want a set of orbitals such that we can choose the active space to be as small as possible while also minimizing the error associated with this approximation. Here we discuss two methods which attempt to find such sets of orbitals.

1.2.4.1 CASSCF

The complete active space self-consistent field (CASSCF) method^{1,24-26} is an orbital optimization method wherein we consider unitary transformations of the form:

$$\hat{U} = e^{-\hat{K}} \quad (1.26a)$$

$$\hat{K} = \sum_{p>q} \kappa_{pq} (\hat{a}_p^\dagger \hat{a}_q - \hat{a}_q^\dagger \hat{a}_p) \quad (1.26b)$$

where κ_{pq} form a vector $\boldsymbol{\kappa}$ parameterizing this transformation. In a basis of M orbitals, this transformation will have dimension $M \times M$. If we have a set of orbitals $\{\phi_1, \dots, \phi_q, \dots, \phi_M\}$, then \hat{U} transforms the orbitals according to:

$$\tilde{\phi}_{p'} = \sum_q^M \hat{U}_{qp} \phi_q \quad (1.27)$$

to obtain a new set of orbitals $\{\tilde{\phi}_1, \dots, \tilde{\phi}_{q'}, \dots, \tilde{\phi}_M\}$. This is equivalent to transforming the creation and annihilation operators as:¹

$$\tilde{a}_{p'}^\dagger = \sum_q^M \hat{U}_{qp} \hat{a}_q^\dagger \quad (1.28a)$$

$$\tilde{a}_{p'} = \sum_q^M \hat{U}_{qp}^* \hat{a}_q \quad (1.28b)$$

Given a parameterized normalized wavefunction $|\psi(\mathbf{x})\rangle$, we want to minimize the function:

$$f(\boldsymbol{\kappa}, \mathbf{x}) = \langle \psi(\mathbf{x}) | e^{\hat{K}} \hat{H} e^{-\hat{K}} | \psi(\mathbf{x}) \rangle \quad (1.29)$$

over the combined spaces of parameters for $\boldsymbol{\kappa}$ and \boldsymbol{x} . Because \hat{H} and the basis states spanning its Hilbert space are both defined in terms of the creation and annihilation operators in the second-quantization formulation, we can equivalently view Eq. 1.29 as either a similarity transformation of \hat{H} or a rotation of the states $|\psi(\boldsymbol{x})\rangle$. One well-known way to solve this minimization problem is to use Newton's method, which is referred to as the Newton-Raphson method in the CASSCF literature when used in the context of minimizing Eq. 1.29. Using the parameter $\boldsymbol{\lambda} = (\boldsymbol{\kappa}, \boldsymbol{x})$ to denote a joint vector of the two types of parameters, the CASSCF Newton-Raphson method proceeds by approximating $f(\boldsymbol{\lambda} + \boldsymbol{\delta})$ by expanding f in its second-order Taylor series around the point $\boldsymbol{\lambda}$:

$$f(\boldsymbol{\lambda} + \boldsymbol{\delta}) \approx f(\boldsymbol{\lambda}) + \boldsymbol{\delta}^T \nabla f(\boldsymbol{\lambda}) + \frac{1}{2} \boldsymbol{\delta}^T \mathbf{H}_f(\boldsymbol{\lambda}) \boldsymbol{\delta} \quad (1.30)$$

where $\nabla f(\boldsymbol{\lambda})$ and $\mathbf{H}_f(\boldsymbol{\lambda})$ denote the gradient and Hessian of f at $\boldsymbol{\lambda}$, respectively. By taking the gradient of this expression with respect to $\boldsymbol{\delta}$ and setting it to zero, we get the following linear equation that must be solved:¹

$$\mathbf{H}_f(\boldsymbol{\lambda}) \boldsymbol{\delta} = -\nabla f(\boldsymbol{\lambda}) \quad (1.31)$$

The solution vector $\boldsymbol{\delta}$ is then used to define an update on an iterative procedure indexed by n :

$$\boldsymbol{\lambda}_{n+1} = \boldsymbol{\lambda}_n + \boldsymbol{\delta}_n. \quad (1.32)$$

That is, at iteration n , we Taylor expand the expression for the energy around $\boldsymbol{\lambda}_n$, find the point $\boldsymbol{\lambda}_n + \boldsymbol{\delta}_n$ for which the gradient of this expansion with respect to $\boldsymbol{\delta}_n$ vanishes, then repeat this procedure for the new point $\boldsymbol{\lambda}_{n+1}$. This procedure is repeated until some stopping condition for the algorithm is met, which will hopefully occur for a point reasonably close to the global minimum of f . It should be noted that there are many possible variants of this generic procedure which could be implemented. For example, one has the option to carry out the optimization with respect to the parameters $\boldsymbol{\kappa}$ and \boldsymbol{x} either concurrently or separately.

These two options are called the one and two-step Newton-Raphson methods, respectively.²⁶ Regardless of the method by which Eq. 1.29 is minimized, we should emphasize that the point of orbital optimization methods such as CASSCF is to avoid having to carry out an exponentially costly FCI problem in the full space of M orbitals by confining the FCI step to an active space of $N < M$ transformed orbitals. Thus we need a way of choosing which N rotated orbitals to include in the FCI step. One way to do this would be to select the N orbitals with the lowest transformed Hartree-Fock energies, although we note that this is a heuristic which is not guaranteed to be optimal. This is because the Hartree-Fock orbital energies are the eigenvalues of the mean-field Fock operator¹ in Eq. 1.15 which is a separate entity from the exact Hamiltonian \hat{H} . Choosing a subset of these transformed orbitals in this manner therefore does not guarantee that the expectation value of \hat{H} with respect to a wavefunction expanded in this orbital set is lower in energy than a wavefunction expanded using all other possible orbital sets.

1.2.4.2 OptOrbFCI

The optimal orbital FCI (OptOrbFCI) method²⁷ shares surface-level similarities with CASSCF, however we note that there are important distinctions that will become apparent as the method is reviewed. Recall that CASSCF rotated a set of M orbitals by an $M \times M$ unitary parameterized by an anti-hermitian operator \hat{K} . This choice of parameterization allowed the parameters κ_{pq} constituting the vector $\boldsymbol{\kappa}$ to be independent from one another, thus unconstrained optimization schemes such as Newton's method could be used to find a set of optimal orbitals. This would not have not been the case had the method opted to use a unitary matrix not parameterized as an exponential operator of this form. Any optimization procedure would have needed to ensure that the matrix elements of such a unitary remain mutually orthonormal. The OptOrbFCI algorithm takes the orbital space transformation to be a partial unitary matrix \hat{U} of size $M \times N$ with real-valued elements. That is, it takes a set of M orthonormal orbitals $\{\phi_1, \dots, \phi_p, \dots, \phi_M\}$ and maps it to another set of $N < M$ orthonormal orbitals $\{\phi_1, \dots, \tilde{\phi}_{p'}, \dots, \tilde{\phi}_N\}$ via the transformation:

$$\tilde{\phi}_{p'} = \sum_p^M \hat{U}_{qp'} \hat{\phi}_p \quad (1.33)$$

which is equivalent to mapping the fermionic operators as:

$$\tilde{a}_{p'}^\dagger = \sum_p^M \hat{U}_{pp'} \hat{a}_p^\dagger \quad (1.34a)$$

$$\tilde{a}_{p'} = \sum_p^M \hat{U}_{pp'} \hat{a}_p. \quad (1.34b)$$

We want to diagonalize the transformed Hamiltonian $\tilde{H}(\hat{U})$:

$$\tilde{H}(\hat{U}) = \sum_{p',q'}^N \tilde{h}_{p'q'} \tilde{a}_{p'}^\dagger \tilde{a}_{q'} + \frac{1}{2} \sum_{p'q'r's'}^N \tilde{v}_{p'q'r's'} \tilde{a}_{p'}^\dagger \tilde{a}_{q'}^\dagger \tilde{a}_{s'} \tilde{a}_{r'}. \quad (1.35)$$

The transformed one and two-body integrals $\tilde{h}_{p'q'}$ and $\tilde{v}_{p'q'r's'}$ can be computed from their non-transformed counterparts v_{pqrs} and h_{pq} as:

$$\tilde{h}_{p'q'} = \int d^3\vec{r} \phi_{p'}^*(\vec{r}) \left(-\frac{1}{2} \nabla^2 + \sum_I \frac{Z_I}{|\vec{r} - \vec{r}_I|} \right) \phi_{q'}(\vec{r}) \quad (1.36)$$

$$= \sum_{p,q}^M \hat{U}_{pp'} \hat{U}_{qq'} \int d^3\vec{r} \phi_p^*(\vec{r}) \left(-\frac{1}{2} \nabla^2 + \sum_I \frac{Z_I}{|\vec{r} - \vec{r}_I|} \right) \phi_q(\vec{r}) \quad (1.37)$$

$$= \sum_{p,q}^M \hat{U}_{pp'} \hat{U}_{qq'} h_{pq} \quad (1.38)$$

and:

$$\begin{aligned}
\tilde{v}_{p'q'r's'} &= \int d^3\vec{r}d^3\vec{r}'\phi_{p'}^*(\vec{r})\phi_{q'}^*(\vec{r}')\frac{1}{|\vec{r}-\vec{r}'|}\phi_{s'}(\vec{r}')\phi_r(\vec{r}) \\
&= \sum_{p,q,r,s}^M \hat{U}_{pp'}\hat{U}_{qq'}\hat{U}_{ss'}\hat{U}_{rr'} \int d^3\vec{r}d^3\vec{r}'\phi_p^*(\vec{r})\phi_q^*(\vec{r}')\frac{1}{|\vec{r}-\vec{r}'|}\phi_s(\vec{r}')\phi_r(\vec{r}) \\
&= \sum_{p,q,r,s}^M \hat{U}_{pp'}\hat{U}_{qq'}\hat{U}_{ss'}\hat{U}_{rr'}v_{pqrs}.
\end{aligned} \tag{1.39}$$

We can therefore write the transformed Hamiltonian in terms of the matrix elements of \hat{U} as:

$$\begin{aligned}
\tilde{H}(\hat{U}) &= \sum_{p',q'}^N \sum_{p,q}^M h_{pq}\hat{U}_{pp'}\hat{U}_{qq'}\tilde{a}_p^\dagger\tilde{a}_{q'} \\
&+ \frac{1}{2} \sum_{p'q'r's'}^N \sum_{p,q,r,s}^M \hat{U}_{pp'}\hat{U}_{qq'}\hat{U}_{rr'}\hat{U}_{ss'}v_{pqrs}\tilde{a}_p^\dagger\tilde{a}_{q'}^\dagger\tilde{a}_{s'}\tilde{a}_{r'}.
\end{aligned} \tag{1.40}$$

Suppose we have a parameterized state $|\psi(\mathbf{x})\rangle$ in an active space of N orbitals. Define the sets $\mathcal{X} = \{\mathbf{x} \mid \langle\psi(\mathbf{x})|\psi(\mathbf{x})\rangle = 1\}$ and $\mathcal{U}(M \times N) = \{\hat{U} \in \mathbb{R}^{M \times N} \mid \hat{U}^\dagger\hat{U} = I_N\}$. *i.e.* \mathbf{x} is a set of parameters of fixed length such that the parameterized state is normalized and \hat{U} is any $M \times N$ partial unitary with real-valued elements. We want to perform the minimization problem:

$$\min_{\substack{\mathbf{x} \in \mathcal{X} \\ \hat{U} \in \mathcal{U}(M \times N)}} \langle\psi(\mathbf{x})|\tilde{H}(\hat{U})|\psi(\mathbf{x})\rangle \tag{1.41}$$

to find the ground state in an optimized basis set. This minimization problem is carried out in an alternating fashion between the two types of parameters \mathbf{x} and \hat{U} . We start with an initial guess for the partial unitary, transform the Hamiltonian according to Eq. 1.40 using this guess, then minimize Eq. 1.41 with respect to \mathbf{x} . This will return a state $|\psi(\mathbf{x})\rangle$ which we use to compute the quantities $\langle\psi(\mathbf{x})|\tilde{a}_p^\dagger\tilde{a}_{q'}|\psi(\mathbf{x})\rangle$ and $\langle\psi(\mathbf{x})|\tilde{a}_p^\dagger\tilde{a}_{q'}^\dagger\tilde{a}_{s'}\tilde{a}_{r'}|\psi(\mathbf{x})\rangle$. These

quantities are the elements of the 1 and 2 reduced density matrices (1 and 2-RDM) with respect to the state $|\psi(\mathbf{x})\rangle$. We then minimize Eq. 1.41 with respect to \hat{U} keeping the 1 and 2-RDM tensors fixed. This will return a new \hat{U} which we use to re-compute the transformed Hamiltonian and repeat the minimization with respect to \mathbf{x} . This alternating optimization procedure is repeated until the energy difference between successive FCI solutions is below some predetermined threshold.

The minimization with respect to \mathbf{x} is simply a diagonalization problem where any FCI solver would be suitable. In the OptOrbFCI proposal paper, the Coordinate-Descent FCI (CDFCI) solver is used.²⁸ The minimization problem with respect to \hat{U} must be handled more carefully as the matrix elements \hat{U}_{pq} are not independent parameters. They must be constrained such that \hat{U} is a partial unitary matrix. Several optimization algorithms which adhere to these constraints exist in the literature,^{29–32} however the authors of the OptOrbFCI proposal paper use a gradient projection method.³³ In the absence of orthogonality constraints, one could iteratively update the parameter vector \mathbf{x} of a function $f(\mathbf{x})$ in the opposite direction of its gradient:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \tau_k \nabla f(\mathbf{x}_k) \quad (1.42)$$

for some small step size τ_k . This is commonly known as gradient descent and is ubiquitous in optimization problems. In the projected gradient method, an analogous parameter update is applied, followed by projecting the resulting updated matrix back onto the manifold of partial unitaries. Denote the expectation value of the transformed Hamiltonian as a function of \hat{U} with fixed \mathbf{x} by $f(\hat{U})$. The update in the parameter \hat{U} is defined in the gradient projection method as:

$$\hat{U}_{k+1} = \text{orth} \left(\hat{U}_k - \tau_k \nabla f(\hat{U}_k) \right) \quad (1.43)$$

where the orthogonal projection function $\text{orth}(A)$ is defined by:³³

$$\text{orth}(A) = AQA^{-1/2}Q^T \quad (1.44)$$

where Q is a matrix whose columns consist of the eigenvectors of $A^T A$ and Λ is a diagonal matrix whose entries are the eigenvalues of $A^T A$ with the same ordering as the eigenvector columns of Q . This projection ensures that the parameter \hat{U} remains a partial unitary throughout the optimization. τ_k is the gradient descent step size, which is chosen adaptively as a Barzilai-Borwein step size according to the rule:^{27,33}

$$\tau_k^{odd} = \frac{\text{Tr} \left[(\hat{U}_k - \hat{U}_{k-1})^T (\hat{U}_k - \hat{U}_{k-1}) \right]}{\left| \text{Tr} \left[(\hat{U}_k - \hat{U}_{k-1})^T (\nabla f(\hat{U}_k) - \nabla f(\hat{U}_{k-1})) \right] \right|} \quad (1.45a)$$

$$\tau_k^{even} = \frac{\text{Tr} \left[(\hat{U}_k - \hat{U}_{k-1})^T (\nabla f(\hat{U}_k) - \nabla f(\hat{U}_{k-1})) \right]}{\left| \text{Tr} \left[(\nabla f(\hat{U}_k) - \nabla f(\hat{U}_{k-1}))^T (\nabla f(\hat{U}_k) - \nabla f(\hat{U}_{k-1})) \right] \right|} \quad (1.45b)$$

for odd and even values of the iteration number k , respectively. This method was shown through numerical simulations to be more adept at avoiding local minima than CASSCF, achieving more accurate ground state energies.²⁷ Thus, there is motivation to investigate its utility in other settings such as solving the ground state problem on quantum computers.

1.3 The Basics of Quantum Computing

In the circuit model of classical computing, the basic building blocks of algorithms are classical logic gates (*e.g.* NOT, AND, XOR, ect...) which map strings of bits to other strings of bits. By constructing sequences of these gates which act on registers of bits, we can carry out basic functionality such as adding two numbers together.³⁴ Sequences of these functionality units can be combined to perform more complex functions such as CPU instructions. Algorithms designed to carry out specific tasks can be implemented in a computer program, compiled into a sequence of such instructions, then executed by the CPU. The circuit model of quantum computing works in an analogous manner. Instead of having logical operations act on registers of bits (which can only have the value 0 or 1), we

store and process information on two-level quantum systems called qubits whose state can be any normalized linear combination of the "zero" state $|0\rangle$ and the "one" state $|1\rangle$:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{1.46}$$

where α and β are any two complex numbers such that $|\alpha|^2 + |\beta|^2 = 1$. This state occupies a two-dimensional Hilbert space and is denoted in matrix notation as:

$$|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{1.47}$$

One qubit is not enough to do anything interesting or useful, so in general we assume access to a register of N qubits. The collective state of these N qubits can be any state of the form:

$$|\psi\rangle = \sum_{n=0}^{2^N-1} c_n |n\rangle \tag{1.48}$$

where c_n are any complex numbers such that $\sum_n |c_n|^2 = 1$. The states $|n\rangle$ are tensor products of individual qubits whose states are either $|0\rangle$ or $|1\rangle$. n is the base-10 integer representation of the bitstring of length N represented by this tensor product state. For example, for a register of 3 qubits, a few examples of such states would be:

$$|0\rangle = |0\rangle \otimes |0\rangle \otimes |0\rangle = |000\rangle$$

$$|1\rangle = |0\rangle \otimes |0\rangle \otimes |1\rangle = |001\rangle$$

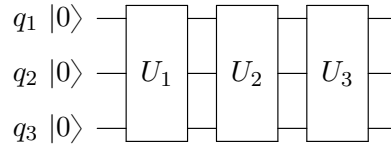
$$|5\rangle = |1\rangle \otimes |0\rangle \otimes |1\rangle = |101\rangle.$$

Such states are known as the computational basis states. Because there are 2^N possible bitstrings of length N , there are 2^N computational basis states which span the 2^N -dimensional Hilbert space of the register of N qubits. Quantum logic gates are represented by either unitary operators or projective measurement operators acting on qubit states. Quantum

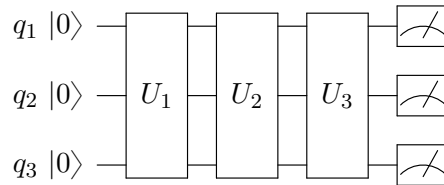
circuits are comprised of sequences of such operations acting on some initial state which is typically taken (without loss of generality) to be the "all-zero" state $|0\rangle = |0\rangle^{\otimes N}$. In matrix notation, the quantum circuit comprising the application of the unitaries U_1 , U_2 , and U_3 in that order is given by:

$$U_3 U_2 U_1 |0\rangle.$$

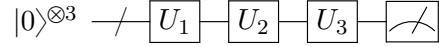
Quantum circuits are often denoted as a sequence of gates connected by wires. Each wire is associated with a qubit or register of qubits. These wires are represented by horizontal lines which denote the flow of time from left to right. Gates (denoted as boxes, dots, or circles) act on a subset of qubits if the symbols for these gates are connected by the wires associated with these qubits. For example, if we suppose that the circuit just described above acts on a register of 3 qubits $\{q_1, q_2, q_3\}$ initialized in the state $|0\rangle$, its circuit representation would be given by:



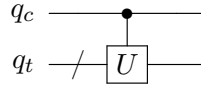
where each of these operations can in general act trivially or non-trivially on any subset of these qubits. Projective measurements in the computational basis are denoted by a meter symbol. If we performed this circuit with such a measurement at the end, the circuit would be denoted by:



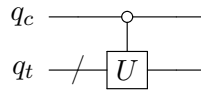
Multi-qubit registers are often compactly denoted by wires with a slash through them. Using this convention, this circuit would be denoted as:



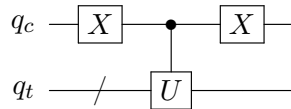
It is often the case that we want to apply an operation U on some subset of qubits conditioned on the state of some other register of qubits. The qubit register to which the operation U is (possibly) being applied is called the target qubit register. The qubits whose states this operation is conditioned on are called the control qubits. Such controlled gates are denoted by a box overlaying the target qubit wires connected by vertical wires to dots overlaying the wires of the control qubits. A closed dot denotes that the qubit is controlled on being in the state $|1\rangle$ and an open dot denotes that the qubit is controlled on being in the state $|0\rangle$. A gate controlled on the state of a qubit being $|1\rangle$ is denoted in the circuit notation as:



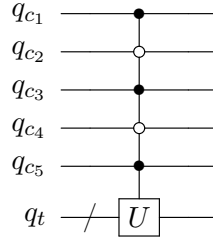
where q_c and q_t denote the control and target qubits, respectively. The analogous circuit where the application of U is conditioned on the control qubit being in the state $|0\rangle$ is given by:



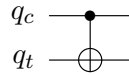
This is equivalent to:



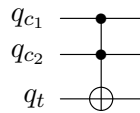
where the gate X is the gate given by the Pauli-X matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, which is analogous to the classical NOT gate. In general we can construct gates conditioned on the states of several qubits. An example of such a gate is given by:



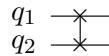
For any state in the joint Hilbert space $\mathcal{H}_c \otimes \mathcal{H}_t$ of the control and target qubits, this multi-qubit controlled gate applies the operator \hat{U} to components of the joint wavefunction where the control qubits are described by $|10101\rangle$. The gate acts as the identity on all other components of the wavefunction. A simple example of a controlled gate with one target qubit and one control qubit is the controlled-NOT (CNOT) gate, which applies a Pauli-X gate to the target qubit for components of the joint wavefunction where the control qubit is described by $|1\rangle$. The circuit notation for this gate is given by:



Another interesting gate is the Toffoli gate, which controls the application of a NOT gate on a target qubit q_t on two qubits q_{c_1} and q_{c_2} both being in the state $|1\rangle$:



One more gate worth adding to our library is the SWAP gate:



which does exactly what its name suggests: it swaps the states of two qubits. *i.e.* it performs the operation:

$$SWAP_{1,2}|a\rangle_1 \otimes |b\rangle_2 = |b\rangle_1 \otimes |a\rangle_2 \quad (1.50)$$

Although many quantum circuits are constructed and represented as multi-qubit gates such as these, in general we do not need to make the assumption that a quantum computer can directly implement gates acting on more than 2 qubits. This is because any unitary acting on any number of qubits can be approximated to arbitrarily small error by a sequence of gates which act on at most 2 qubits. Sets of gates which can be used to construct such sequences are referred to as universal sets. In particular, it has been proven that this can be accomplished using CNOT and arbitrary single-qubit gates.³⁴ Furthermore, it has been proven through the Solovay-Kitaev theorem that one can use a finite, discrete set of single-qubit gates to approximate any single-qubit gate to arbitrarily small error using a sequence of such gates whose length scales logarithmically with the inverse error.^{34,35} Thus, the CNOT together with a finite, discrete set of single-qubit gates can form a universal set. A well-known example of such a universal set is $\{\text{CNOT}, H, S, T\}$,³⁴ where H denotes the Hadamard gate. The matrix representation of H , S , and T are given by:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{1.51a}$$

$$S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \tag{1.51b}$$

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} \tag{1.51c}$$

The S gate is sometimes referred to as the phase gate since it applies a complex phase to the $|1\rangle$ component of any single-qubit state. Similarly, the T gate is sometimes referred to as a $\pi/8$ rotation gate.³⁴

The previous discussions included only operations that were either unitary or projective measurements. That is, we assumed that the qubits constituted a closed system that evolved unitarily with time, followed by measurements that we perform. However, it is worth taking a moment to emphasize that this is an idealization that is never truly the case in quantum computing. In general, the state of a register of qubits is not a closed

system represented by a vector of amplitudes, but is rather an open system represented by a density matrix that can interact with its environment and as a result will not evolve with time in a perfectly unitary way. Interactions such as these induce errors in the calculation that can corrupt the results. The development of quantum algorithms is usually formulated in the language of closed, error-free systems. The implicit assumption here is that we can separately develop methods which protect the quantum system from error and that can be seamlessly integrated into the physical implementation of these algorithms. Here we briefly review the notation that is used to model errors in the quantum circuit model and the types of methods that have been developed to protect these circuits from error or mitigate their impact.

The evolution of an open quantum system can be modeled by quantum channels \mathcal{E} which map valid density operators to other valid density operators as $\rho \rightarrow \mathcal{E}(\rho)$. In order for the output density operator $\mathcal{E}(\rho)$ to be a valid density operator, we require that \mathcal{E} be a linear map that preserves the trace of density operators and maps positive operators to positive operators. That is, \mathcal{E} is a linear, completely positive trace-preserving map. The action of \mathcal{E} on an arbitrary density operator ρ can be represented in terms of its Kraus operators $\{\hat{E}_i\}$ as:

$$\mathcal{E}(\rho) = \sum_i \hat{E}_i \rho \hat{E}_i^\dagger \quad (1.52)$$

where:

$$\sum_i \hat{E}_i^\dagger \hat{E}_i = \hat{I}. \quad (1.53)$$

The physical interpretation of this is that $\mathcal{E}(\rho)$ is a probabilistic mixture of states:

$$\mathcal{E}(\rho) = \sum_i p_i \frac{\hat{E}_i \rho \hat{E}_i^\dagger}{\text{Tr}(\hat{E}_i \rho \hat{E}_i^\dagger)} \quad (1.54)$$

with probabilities $p_i = \text{Tr}(\hat{E}_i \rho \hat{E}_i^\dagger)$.³⁴ We represent the composition of N quantum channels as:

$$\mathcal{E}_N \circ \mathcal{E}_{N-1} \circ \cdots \circ \mathcal{E}_1(\rho). \quad (1.55)$$

This is useful for representing a sequence of quantum gates which are each prone to a discrete set of errors. We can consider a noisy gate to be the composition of the exact gate which we intend to implement and a map representing one element of a set of errors that can occur with certain probabilities. By composing several of these noisy gate channels together, we can model a noisy quantum circuit. Some well-known examples of single-qubit error channels are the bit-flip channel:

$$\mathcal{E}(\rho) = (1 - p)\rho + p\hat{X}\rho\hat{X}, \quad (1.56)$$

the phase-flip channel:

$$\mathcal{E}(\rho) = (1 - p)\rho + p\hat{Z}\rho\hat{Z}, \quad (1.57)$$

and the depolarizing channel:³⁴

$$\begin{aligned} \mathcal{E}(\rho) &= (1 - p)\rho + \frac{p}{3}(\hat{X}\rho\hat{X} + \hat{Y}\rho\hat{Y} + \hat{Z}\rho\hat{Z}) \\ &= (1 - p)\rho + p\frac{\hat{I}}{2}. \end{aligned} \quad (1.58)$$

In each of these channels, the state is unchanged with probability $(1 - p)$ and incurs an error with probability p . We note that for each of these noise channels, successive applications of the channel to the initial state ρ has the effect that the state tends towards a steady state ρ_∞ of the channel such that $\mathcal{E}(\rho_\infty) = \rho_\infty$ in the limit of infinite applications. This steady state has no dependence on the initial state ρ , thus we intuitively expect that the effect of allowing errors to accumulate is that all information initially contained in ρ is lost.

The goal of quantum error-correction^{34,36–41} is to be able to detect errors through mid-circuit measurements and apply operations which undo these errors, conditioned on the outcome of these measurements. The main idea is that logical qubits are encoded into the collective state of many physical qubits, which will allow a certain number of local errors on physical qubits to be tolerated and corrected if done properly. We likewise encode logical Pauli operators as Pauli strings acting on the collective Hilbert space of these physical qubits and encode logical gates as sequences of physical gates acting on these physical qubits. When this encoding is done in such a way that errors are suppressed rather than propagated further, we say that we have achieved fault-tolerant quantum computation. By the threshold theorem of quantum error-correction,³⁴ in order to carry out scalable fault-tolerant error-correction, the error rates of the physical gates used in our encoding must be below some threshold error rate p_{th} characteristic of the code being used. When this condition is met, we can increase the code distance d (the minimum number of physical errors that can produce a logical error) by increasing the number of qubits to make the logical error arbitrarily low. The logical error p_L as a function of d , p_{th} , and the physical error rate p is approximated as:⁴⁰

$$p_L \cong C \left(\frac{p}{p_{th}} \right)^{(d+1)/2} \quad (1.59)$$

where C is a constant that is characteristic of the particular code being used. Thus, if the physical error rates are below a certain threshold, the logical error rates can be exponentially suppressed with increasing code distance d . For many quantum algorithms which require circuits that are quite deep, we can expect upwards of 10^3 – 10^4 physical qubits to be required to encode each logical qubit to bring the logical error rate down sufficiently.^{40,42} This is well beyond what current quantum computers are capable of, thus motivating quantum error mitigation techniques.

Error mitigation techniques^{43–51} refer to a class of methods wherein the effects of noise on the expectation values of observables are suppressed through additional measurements

and classical post-processing rather than the encoding of logical states into many physical qubits. Whereas error correction performs mid-circuit measurements to periodically detect and correct errors as they occur, error mitigation allows errors to propagate throughout the circuit undetected and uncorrected. As a result there is no threshold theorem for error mitigation techniques and these methods are only practical when the average number of errors per circuit is small enough that the amount of post-processing needed to recover the information lost through errors is not horrendous. The primary reason these techniques have been researched is that they can be implemented with either low or zero qubit count overhead, which makes them practical to implement on near-term devices for demonstration purposes. For example, zero noise extrapolation^{43,46} (ZNE) is a technique where we assume that we have the ability to intentionally make the error rate worse. That is, the noise level is modelled by some parameter λ and we assume that we can make λ larger than what it is optimally. If we have some observable E , we can measure E for various values of λ and extrapolate to the case where $\lambda = 0$. In probabilistic error cancellation (PEC),^{43,44,48} we assume that we can characterize the noise channel \mathcal{N} of a circuit sufficiently well. We can then estimate the inverse noise channel \mathcal{N}^{-1} and randomly sample from the Kraus operators which constitute it, then add these operations to the circuit when selected from a random sample. This has the effect that the bias induced by \mathcal{N} on the estimation of the observable is partially negated, but at the cost of an increased variance which results in a sampling overhead which increases exponentially with the circuit depth.⁵⁰ Thus, quantum error mitigation, in contrast to quantum error-correction, is not widely expected to be a scalable solution to suppressing the effects of noise in quantum algorithms. Error mitigation, however, can be a useful tool for experiments in the near-term.

1.4 The Electronic Structure Problem on Quantum Computers

The crux of the challenge in solving the electronic structure on classical computers is that in the worst case, solving for the exact (FCI) wavefunction of a quantum many-body system with N basis functions requires the specification of 2^N complex amplitudes as

shown in Eq. 1.48. This exponential scaling may be somewhat mitigated when it is known in advance that the system in question has a definite particle number and spin-magnetization. *i.e.* the wavefunction is an eigenstate of both the particle number operator \hat{N} and the spin-magnetization operator \hat{S}_z . This is often the case in non-relativistic molecular chemistry where the Hamiltonian commutes with both of these operators. In the case where our basis set consists of spin-orbitals with an equal number of spin-up and spin-down orbitals and we have n_\uparrow and n_\downarrow spin-up and spin-down electrons respectively, the number of possible Slater determinants N_{det} involved in the FCI wavefunction is given by:

$$N_{det} = \binom{N/2}{n_\uparrow} \binom{N/2}{n_\downarrow} \quad (1.60)$$

Under the assumption that the number of electrons $n_\uparrow + n_\downarrow$ and the number of spin-orbitals N needed to describe them accurately differ by some constant of proportionality, this combinatorial scaling becomes approximately exponential in the limit of large numbers of electrons or orbitals.¹ For this reason FCI calculations are limited to small molecular systems. Larger systems can be treated approximately using methods such as Hartree-Fock, truncated CI, and Coupled-Cluster. However, all of these approximate methods sidestep the exponential scaling of FCI by heuristically reducing the number of Slater determinants included in the solution. As a result we do not expect these methods to accurately treat systems for which their heuristics do not capture a sufficient number of determinants that make a non-negligible contribution to the total energy.

Because the main difficulty in accurately treating the electronic structure problem on classical computers stems from their inefficiency in representing quantum many-body wavefunctions which involve a large number of Slater determinants, it is natural to ask the question of whether a computer which is itself a quantum many-body system can efficiently solve problems relevant to quantum many-body physics. This line of reasoning is often attributed to Feynman.⁵² Thus, the potential ability for quantum computers to efficiently represent physically relevant wavefunctions with exponentially-many amplitudes may eliminate or re-

duce the reliance on the heuristics that are ubiquitous in classical computational chemistry methods, allowing us to accurately treat a more universal set of molecular systems. While this intuition is reasonable, the reality of solving the electronic structure problem efficiently turns out to be more involved than simply using a quantum computer to store many-body wavefunctions. For example, it is known that in the worst case, approximating an arbitrary n -qubit unitary operator with a sequence of 1 and 2-qubit gates scales exponentially with n .³⁴ Thus, there are some wavefunctions which cannot be efficiently prepared with a quantum computer. This should not deter us as we are not necessarily interested in the worst case or preparing arbitrary wavefunctions. We seek particular wavefunctions: eigenstates of the time-independent Schrödinger equation for physically-motivated systems.

The first step in solving this problem is to represent the second-quantized fermionic Hamiltonian in Eq. 1.8 in a form that involves operators that can be implemented on a quantum computer while simultaneously preserving the fermionic anticommutation relations in Eq. 1.7. This can be accomplished with one of several schemes which map the fermionic creation and annihilation operators to qubit operators in such a way that the resulting qubit operators also obey the fermionic anticommutation relations. Well-known mapping schemes include Jordan-Wigner, Bravyi-Kitaev, Bravyi-Kitaev Superfast, and Parity.^{53–58} As an example, the Jordan-Wigner mapping is defined by the transformation:⁵³

$$\hat{a}_i^\dagger = \sigma_i^+ \bigotimes_{j<i} Z_j \tag{1.61a}$$

$$\hat{a}_i = \sigma_i^- \bigotimes_{j<i} Z_j \tag{1.61b}$$

where the subscript indices i and j specify that the operator acts non-trivially only on the i th or j th qubit and trivially on all other qubits. Here, Z denotes the Pauli-Z operator and σ^\pm are the qubit raising and lowering operators given by:

$$\sigma_i^\pm = \frac{1}{2} (X_i \mp iY_i) \tag{1.62}$$

where X and Y denote the Pauli- X and Pauli- Y operators, respectively. The matrix representations of all the Pauli operators are given by:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1.63a}$$

$$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{1.63b}$$

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.63c}$$

The qubit raising and lowering operators by themselves have the same raising and lowering behavior as their fermionic counterparts, however they do not obey the fermionic anticommutation relations, necessitating the tensor product of Pauli Z operators in Eq. 1.61. This trail of Pauli- Z operators implies that the number of qubits on which the encoded operators act non-trivially scales with the number of qubits N as $\mathcal{O}(N)$.⁵³ For example, for N spin-orbitals the encoded operators for the N th spin-orbital act non-trivially on all the qubits. There are several contexts for which this is not necessarily a desirable property. For example, ultimately we want to encode the entire Hamiltonian \hat{H} as a qubit operator in the form of a linear combination of tensor products of Pauli operators. If we want to evolve some state in time according to the time evolution operator $e^{-i\hat{H}t}$, then we want each of the Pauli terms in \hat{H} to act non-trivially on as few of the qubits as possible to minimize the complexity of compiling this operator into a sequence of 1 and 2-qubit gates. Nonetheless, the Jordan-Wigner mapping has proven to be popular in the quantum computational chemistry literature for the simple reason that in this encoding the computational basis states have the same clear physical interpretation as Slater determinants in the classical computational chemistry literature: each qubit encodes the occupation number of the spin-orbital with the same index. This property can be useful for physically-motivated circuit design because in this encoding it is easy to construct unitaries which conserve particle-number and spin-magnetization and which act non-trivially on as few qubits as

possible. For example, the assumption of the use of the Jordan-Wigner encoding was used to demonstrate that a particular class of particle-conserving unitaries are universal for describing chemically-motivated wavefunctions with a definite number of electrons.⁵⁹ Other qubit mapping schemes do not have the property that the encoded fermionic operators act non-trivially on a number of qubits which scales as $\mathcal{O}(N)$. In the Bravyi-Kitaev mapping for instance, the encoded fermionic operators act non-trivially on a number of qubits which scales as $\mathcal{O}(\log N)$,⁵⁷ but the computational basis states do not have the same clear physical interpretation as they did the Jordan-Wigner encoding.

It is also worth noting that if the spin-orbitals are indexed such that the first $N/2$ spin-orbitals are spin-up (down) and the second half are spin-down (up), then for certain qubit mapping schemes the values of a subset of the qubits in the computational basis states may encode the values of observables which are conserved quantities. *i.e.* the observables are represented by operators which commute with the Hamiltonian \hat{H} . In the case where we want to restrict the simulation of various systems to a sector of the Hilbert space corresponding to particular eigenvalues of these operators, the values of the qubits encoding these conserved quantities ideally would be fixed. It has been shown that these symmetries can be exploited to reduce the number of qubits in the Hamiltonian, reducing the computational cost of any algorithm which uses it.⁵⁶ For example, the value of the i th qubit in a computational basis state in the Parity mapping encodes the sum (modulo 2) of the occupation numbers of the spin-orbitals up to the i th spin-orbital.⁵⁷ Thus, the value of the N th qubit encodes the number of particles in the system modulo 2. Similarly, the $\frac{N}{2}$ th qubit encodes the number of spin-up (or spin-down) particles modulo 2, which determines the spin-magnetization of the system modulo 2. Both of these quantities are conserved for the non-relativistic electronic structure Hamiltonian. By omitting these qubits and replacing the Pauli operators acting on them with their eigenvalues of the Pauli-Z operator, the number of qubits involved in calculations using these operators can be reduced by 2.⁵⁶

Regardless of the particular mapping scheme chosen, we can now map the second-quantized fermionic Hamiltonian to a qubit Hamiltonian that is some linear combination

of tensor products of Pauli operators (including the identity) acting on the Hilbert spaces of each qubit individually:

$$\hat{H} = \sum_i h_i \hat{P}_i \tag{1.64}$$

where h_i are complex coefficients and the operators \hat{P}_i are often referred to as Pauli strings or Pauli words. An example of a possible \hat{P}_i would be $X_1 \otimes Y_2 \otimes X_3 \otimes Z_4 \otimes I_5$. With the electronic structure Hamiltonian now represented as an operator acting on the Hilbert space of qubits, we can now discuss quantum algorithms for finding its eigenvalues and eigenvectors.

2. Quantum Phase Estimation Background

In this chapter we review some of the literature on quantum phase estimation. In Sec. 2.1 we discuss the version of phase estimation which uses multiple ancilla qubits. This is sometimes referred to as "textbook phase estimation" as it is the version which appears in the well-known textbook by Nielsen and Chuang³⁴ through which many students are introduced to this method for the first time. In Sec. 2.2 we discuss the iterative version of phase estimation which uses only one ancilla qubit. In Sec. 2.3 we review the version of phase estimation which uses Bayesian inference to iteratively update a guess for the phase. In Sec. 2.4 we review some of the works which have attempted to estimate the resources needed to use phase estimation to solve useful problems currently inaccessible by classical computers and the implications these findings have for its implementation on a physical device.

2.1 Multi-Ancilla Phase Estimation

The first quantum algorithm developed for finding the eigenvalues of the electronic structure Hamiltonian was quantum phase estimation (QPE).^{8,34,60} There are many variants of phase estimation,^{8,34,61-69} but at their core, all phase estimation variants address the following problem: given a unitary operator \hat{U} and one of its eigenstates $|\phi\rangle$ with eigenvalue $e^{2\pi i\phi}$ such that $0 \leq \phi < 1$, solve for ϕ . *i.e.* it is assumed that we can prepare $|\phi\rangle$ and implement \hat{U} on a quantum computer. The requirement that we need to prepare $|\phi\rangle$ exactly can be relaxed to the requirement that we can prepare a state which has non-vanishing overlap with $|\phi\rangle$.^{8,34,60} For now, we assume that we can prepare the exact $|\phi\rangle$ as doing so simplifies the presentation of these methods. The reason why this method is of particular interest to the electronic structure problem is that we can take \hat{U} to be the time evolution operator $e^{-i\hat{H}t}$ for some t such that $0 \leq -E_0t < 2\pi$, where E_0 is the ground state energy of the Hamiltonian \hat{H} in Eq. 1.64.^{8,60}

We will first consider the version of phase estimation that uses multiple ancilla.^{8,34,60} In order to further simplify the presentation of the method, we make the assumption that

ϕ can be represented as a binary fraction $0.\phi_1\phi_2\dots\phi_m$ with exactly m bits of accuracy. We start with a state consisting of a tensor product of a register of m ancilla qubits (all initialized in the zero state) and the state $|\phi\rangle$:

$$|0\rangle^{\otimes m} \otimes |\phi\rangle. \quad (2.1)$$

Applying a Hadamard gate to all m ancilla qubits, we get:

$$\frac{1}{2^{m/2}} (|0\rangle + |1\rangle)^{\otimes m} \otimes |\phi\rangle. \quad (2.2)$$

In the next step we impart partial information about the value of ϕ as a complex relative phase to the $|1\rangle$ component of the state of each of the ancilla qubits. This is accomplished by applying successive powers of 2 (2^k for integers k from 0 to $m-1$) of \hat{U} to $|\psi\rangle$ controlled on the state of each of the ancillas in sequence. Denoting the operator corresponding to \hat{U}^{2^k} being controlled on the state of the k th ancilla qubit as $C_k\hat{U}^{2^k}$, the evolution of the joint wavefunction as this sequence of operators is applied is given by:³⁴

$$\frac{1}{2^{m/2}} \prod_{k=m-1}^0 C_k\hat{U}^{2^k} (|0\rangle + |1\rangle)^{\otimes m} \otimes |\phi\rangle \quad (2.3)$$

$$= \frac{1}{2^{m/2}} \prod_{k=m-1}^1 C_k\hat{U}^{2^k} (|0\rangle + |1\rangle)^{\otimes m-1} \left(|0\rangle + e^{2\pi i(2^0\phi)}|1\rangle \right) \otimes |\phi\rangle \quad (2.4)$$

$$= \frac{1}{2^{m/2}} \prod_{k=m-1}^2 C_k\hat{U}^{2^k} (|0\rangle + |1\rangle)^{\otimes m-2} \left(|0\rangle + e^{2\pi i(2^1\phi)}|1\rangle \right) \left(|0\rangle + e^{2\pi i(2^0\phi)}|1\rangle \right) \otimes |\phi\rangle \quad (2.5)$$

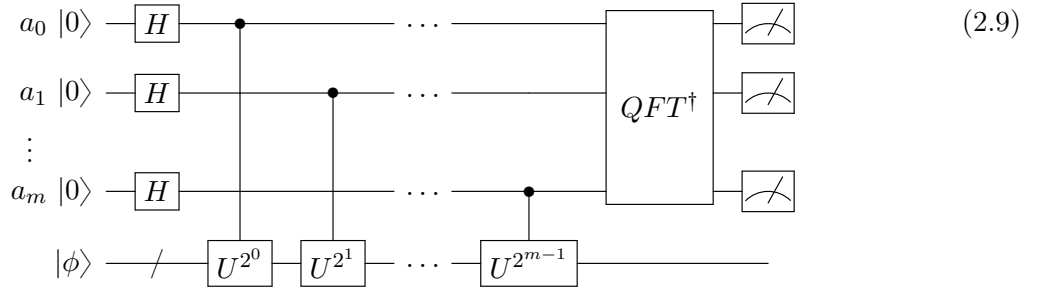
$$\vdots \quad (2.6)$$

$$= \frac{1}{2^{m/2}} \bigotimes_{k=m-1}^0 \left(|0\rangle + e^{2\pi i(2^k\phi)}|1\rangle \right) \otimes |\phi\rangle. \quad (2.7)$$

Next, we apply the inverse of the Quantum Fourier Transform (QFT), which can be defined by its action on the computational basis states $|\phi_1\phi_2\dots\phi_m\rangle$ as:³⁴

$$QFT|\phi_1\phi_2\dots\phi_m\rangle = \frac{1}{2^{m/2}} \bigotimes_{k=m-1}^0 \left(|0\rangle + e^{2\pi i(2^k\phi)}|1\rangle \right). \quad (2.8)$$

Thus if we apply the inverse Quantum Fourier Transform to the ancilla qubits following the sequence of controlled \hat{U} gates, the ancilla qubits state will be mapped to the computational basis state $|\phi_1\phi_2\dots\phi_m\rangle = |2^m\phi\rangle$. The total wavefunction is given by $|2^m\phi\rangle \otimes |\phi\rangle$. Measurement of the ancilla qubit then allows for the value of ϕ to be obtained simply by reading out the bits obtained by a single measurement. The circuit for this sequence of operations is given by:^{8,34}



This presentation of QPE works out nicely in that its success probability is 100%, however this is only because we made two rather unreasonable assumptions: **1.** that we know and can prepare the exact eigenstate $|\phi\rangle$ and **2.** that ϕ can be represented using exactly the same number of bits m as the number of ancilla qubits that we happened to use. In practice, neither of these is likely to ever be true. After all, if we already had a perfect preparation of $|\phi\rangle$, we could simply measure the expectation value of the Hamiltonian using statistical sampling, eliminating the need for phase estimation in the first place. Regarding the latter assumption, it was only because the controlled \hat{U} gates imparted phases into Eq. 2.7 which exactly matched those imparted by the Fourier transformation of $|\phi_1\phi_2\dots\phi_m\rangle$ that we were able to map Eq. 2.7 back to precisely the computational basis state $|\phi_1\phi_2\dots\phi_m\rangle$ with the inverse QFT. In practice, if ϕ cannot be represented with m bits (or if we do not know how many bits are required for its exact value), then the phases imparted by the controlled \hat{U} gates and those imparted by the Fourier transform will be mismatched. This implies

that the application of the inverse QFT gate will not map the state back to precisely the correct computational basis state, but rather a linear combination of computational basis states with bitstring values clustered around the bitstring encoding ϕ , giving rise to a measurement probability distribution which can be systematically improved to estimate ϕ with arbitrarily high accuracy. It can be shown that if one wants to upper bound the probability of measuring the value of ϕ to no fewer than m bits of accuracy by ϵ , then one needs to use at least N_a ancilla qubits given by:³⁴

$$N_a = m + \left\lceil \log_2 \left(2 + \frac{1}{2\epsilon} \right) \right\rceil \quad (2.10)$$

where $f(x) = \lceil x \rceil$ is the ceiling function denoting the smallest integer greater than x . We note that so far there are two sources of error in the estimation of ϕ : the measurement error ϵ and the truncation of ϕ to m bits of accuracy. Given that the total number of applications of the controlled \hat{U} gates that need to be applied scales as $\mathcal{O}(2^{N_a})$, the total runtime scales with ϵ as $\mathcal{O}(\frac{1}{\epsilon})$. The error associated with the truncation (which we denote by ϵ_t) to m bits is given by $\epsilon_t = 2^{-m}$, thus the runtime (as a function of m) scales as $\mathcal{O}(\frac{1}{\epsilon_t})$. We will see in future discussions that the fact that the runtime of QPE scales with the inverse error is important when comparing QPE to other quantum algorithms for solving the electronic structure problem.

We can also consider what the effect of relaxing the requirement that we must prepare the exact eigenstate $|\phi\rangle$ would be. We can take the eigenvalues of \hat{U} to be $\{\phi_0, \phi_1, \dots, \phi_n, \dots\}$ and their respective eigenstates to be labelled as $\{|\phi_0\rangle, |\phi_1\rangle, \dots, |\phi_n\rangle, \dots\}$. Suppose we want to find ϕ_0 and that we can prepare a state $|\tilde{\phi}_0\rangle$ which has non-zero overlap with the exact eigenstate $|\phi_0\rangle$. We can expand $|\tilde{\phi}_0\rangle$ in the eigenbasis of \hat{U} as:

$$|\tilde{\phi}_0\rangle = \sum_{n=0}^{n_{max}} \langle \phi_n | \tilde{\phi}_0 \rangle |\phi_n\rangle. \quad (2.11)$$

Using the fact that each of the gates in the QPE circuit is a linear map, the analog of the state right before the QFT gate is applied in Eq. 2.7 is now given by:

$$\frac{1}{2^{m/2}} \sum_{n=0}^{n_{max}} \langle \phi_n | \tilde{\phi}_0 \rangle \bigotimes_{k=m-1}^0 \left(|0\rangle + e^{2\pi i(2^k \phi_n)} |1\rangle \right) \otimes |\phi_n\rangle. \quad (2.12)$$

Under the assumption that all of the eigenvalues of \hat{U} can be represented as an m -bit binary fraction, application of the inverse QFT gives:

$$\sum_{n=0}^{n_{max}} \langle \phi_n | \tilde{\phi}_0 \rangle |2^m \phi_n\rangle \otimes |\phi_n\rangle. \quad (2.13)$$

There are two aspects of this wavefunction that deserve discussion. The first is that the probability of measuring the particular eigenvalue ϕ_0 that we wish to obtain now picks up a multiplicative factor $|\langle \phi_n | \tilde{\phi}_0 \rangle|^2$. This is true even in the case where ϕ_0 cannot be represented with exactly m bits. In this case the probability of success $1 - \epsilon$ becomes $|\langle \phi_n | \tilde{\phi}_0 \rangle|^2 (1 - \epsilon)$.³⁴ The second aspect to note is that when we measure the ancilla qubit register to measure ϕ_0 , the qubit register storing the state $|\tilde{\phi}_0\rangle$ collapses to the exact eigenstate $|\phi_0\rangle$.⁶⁰ Thus, the QPE procedure will output the exact eigenstate $|\phi_0\rangle$ when the desired eigenvalue is successfully measured.

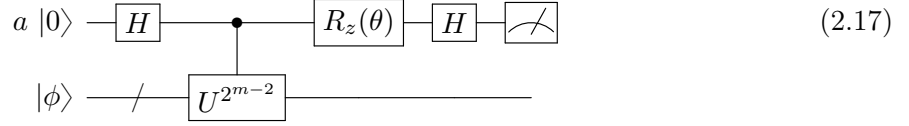
2.2 Iterative Phase Estimation

Real quantum computers currently are prone to error and have limited quantities of qubits. This will continue to be the case in the near term. The implication here is that in order for quantum algorithms to find practical application sooner, they should be designed to use as few qubits as possible and to minimize the number of gates that must be coherently implemented sequentially (*i.e.* the circuit depth) so as to minimize the impact of errors. For this reason, much of the research on quantum algorithms has focused on developing new methods which use fewer qubits and are more resilient to error, even if the asymptotic scaling of the resource requirements remain unchanged or possibly worsen. We can see this

that under the assumptions that we have prepared $|\phi\rangle$ exactly and that the circuit is noise-free, the measurement outcome will be $|\phi_m\rangle$ with 100% probability.⁶⁷ This is analogous to the measurement outcome of the idealized N_a -ancilla QPE. Next, we prepare a circuit such that the output wavefunction is analogous to Eq. 2.15, but with ϕ_{m-1} in the arguments of the sinusoidal amplitudes. This can be done by applying the controlled $\hat{U}^{2^{m-2}}$ gate to encode the phase $0.\phi_{m-1}\phi_m$. If we could then apply a gate that shifts this phase by $2^{-2}\phi_m = 0.0\phi_m$, then the encoded phase would be $0.\phi_{m-1}$. The gate which accomplishes this is the $\hat{R}_z(\theta)$ gate given by:

$$\hat{R}_z(\theta) = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix}. \quad (2.16)$$

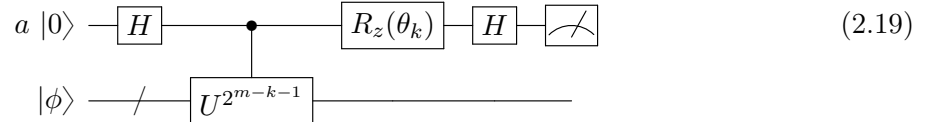
The modified circuit with the phase shift:



yields the state:

$$e^{\pi i 0.\phi_{m-1}\phi_m} [\cos(\pi 0.\phi_{m-1}\phi_m + \theta) |0\rangle - i \sin(\pi 0.\phi_{m-1}\phi_m + \theta) |1\rangle] \otimes |\phi\rangle. \quad (2.18)$$

We can see by inspection that the proper choice of θ is given by $-\pi 0.0\phi_m = 2^{-2}\phi_m$. Running this circuit with this phase shift allows us to measure θ_{m-1} . We repeat this procedure again using the controlled $\hat{U}^{2^{m-2}}$ gate followed by the phase shift $\theta = -\pi 0.0\phi_{m-1}\phi_m$. We can define the recursion rule: for iteration $0 \leq k \leq m-1$, apply the circuit:



to find ϕ_{m-k} , where $\theta_{k+1} = \theta_k/2 - 2^{-2}\pi\phi_{m-k+1}$.^{67,68} For convenience, we define $\phi_{m+1} = 0$ and $\theta_{-1} = 0$ so that $\theta_0 = 0$ as expected. Note that in the case where ϕ cannot be

represented using exactly m bits, ϕ_{m+1} in general will not be zero. This gives rise to a probability of measuring the wrong bit at each iteration. Various schemes can be devised to suppress this error through repeated measurements of each bit. For instance, it can be shown that carrying out an odd number of measurements for each bit and determining the most likely value through majority vote, the probability of error is exponentially suppressed with the number of measurements.⁵⁸ Given that the circuit depth scales as $\mathcal{O}(2^m)$ to obtain a maximum accuracy $\epsilon = 2^{-m}$, we can see that the total runtime of IQPE scales as $\mathcal{O}(\frac{1}{\epsilon})$, just as was the case for the N_a -ancilla version of QPE.

2.3 Bayesian Phase Estimation

Iterative Phase Estimation avoids having to apply several controlled unitaries in sequence in the same circuit by using prior information (the previous bit obtained) to update θ_k and find the next bit. One interesting QPE variant which generalizes this idea is Bayesian Phase Estimation.^{64–66} Instead of solving for the value of ϕ in a bit-by-bit fashion, Bayesian QPE models the phase as a probability distribution with mean ϕ representing the current best guess for the phase and standard deviation σ representing the error in this estimation. Experimental demonstrations of small instances of Bayesian QPE on silicon photonic qubits have suggested that this method may be more error-resilient than Iterative Phase Estimation.⁶⁵ Bayesian Phase Estimation uses a circuit similar to IQPE:

$$\begin{array}{c}
 a \ |0\rangle \text{ --- } [H] \text{ --- } [R_z(-M\theta)] \text{ --- } \bullet \text{ --- } [H] \text{ --- } \text{meter} \\
 | \phi \rangle \text{ --- } \text{meter} \text{ --- } [U^M] \text{ --- } \text{meter}
 \end{array}
 \tag{2.20}$$

where M and θ are chosen adaptively according to the current best guess for ϕ . The conditional probabilities on the measurement outcome x of the ancilla qubit, given an unknown ϕ and known M and θ are given by:⁶⁶

$$P(x = 0 | \phi, \theta, M) = \frac{1 + \cos(M[\theta + \phi])}{2}
 \tag{2.21a}$$

$$P(x = 1|\phi, \theta, M) = \frac{1 - \cos(M[\theta + \phi])}{2}. \quad (2.21b)$$

Note that in this construction, a factor of 2π has been absorbed into ϕ and θ as the lack of a need to represent ϕ as a binary fraction implies that we are no longer reliant on modulo 2 arithmetic. Bayes' theorem⁷⁰ gives a straightforward method for updating the probability distribution of a random variable upon obtaining partial information. Our guess for the distribution of this variable before receiving this information is called the prior distribution and our guess for the distribution after receiving this information and updating it according to Bayes' theorem is called the posterior distribution. In this context, the random variable whose distribution we wish to model is ϕ and the partial information we receive about it is the ancilla measurement outcome x . Bayes' rule in this context is given by:⁶⁶

$$P(\phi|x, \theta, M) = \frac{P(x|\phi, \theta, M)P(\phi)}{\int P(x|\phi, \theta, M)P(\phi)d\phi} \quad (2.22)$$

where $P(\phi)$ is the prior distribution of ϕ and $P(\phi|x, \theta, M)$ is the updated posterior distribution of ϕ given the measurement outcome x for our chosen θ and M . We can then define an iterative procedure (indexed by $k \geq 0$) where individual measurements are taken and the probability distribution $P_k(\phi)$ is updated according to:

$$P_{k+1}(\phi) = \frac{P(x_k|\phi, \theta_k, M_k)P_k(\phi)}{\int P(x_k|\phi, \theta_k, M_k)P_k(\phi)d\phi}. \quad (2.23)$$

The value of M_k chosen as $\left\lceil \frac{1.25}{\sigma_k} \right\rceil$ and θ_k is chosen to be proportional to μ_k , where σ_k and μ_k are the standard deviation and mean of $P_k(\phi)$, respectively.⁶⁶ Note that this choice of M_k implies that the runtime of Bayesian Phase Estimation scales linearly with the inverse error as was the case for previous phase estimation variants.

2.4 Resource Estimates for Quantum Phase Estimation

Algorithmic improvements of QPE over the years are encouraging, however it remains the case that there are significant challenges associated with scaling any of them up to solve

quantum many-body problem instances whose sizes deem them inaccessible by classical computers. Let us take a moment to develop an intuition as to why QPE is challenging to implement in the near-term for molecular chemistry. In the case of the electronic structure problem, the unitary operator \hat{U} is given by the time evolution operator $e^{i\hat{H}t}$, where \hat{H} is a sum of Pauli strings as in Eq. 1.64. A common method for implementing this time evolution operator is by using the Trotter-Suzuki formula:⁷¹

$$e^{i\hat{H}t} \approx \left(\prod_j e^{ih_j\hat{P}_j t/n} \right)^n + \sum_{i>j} h_i h_j [\hat{P}_i, \hat{P}_j] t^2/2n + \mathcal{O}(t^3) \quad (2.24)$$

where \hat{H} is a linear combination of Pauli strings as given in Eq. 1.64. The first term consists of a sequence of gates that we can further decompose into a sequence of one and two qubit gates on a quantum computer. In the limit of infinite n (where n is often referred to as the number of Trotter steps), this approximation becomes exact. Let us do a quick "back of the envelope" order of magnitude approximation of how many gates will be involved in this circuit. An often cited benchmark accuracy needed for ground state energy calculations to have useful, quantitative predictive power in some contexts is around 1.6×10^{-3} Hartree (or 1 kcal/mol),^{1,5,72,73} Thus, we need to implement a gate \hat{U}^M for an $M \approx 10^3$, given that M scales with the error ϵ as $\mathcal{O}(\frac{1}{\epsilon})$. Recall that the finite basis set truncation in the electronic structure problem induces an error that often limits the maximum attainable accuracy of a calculation. There is no universal agreement as to what the minimum number of spin-orbitals is that could be used in a calculation that both exceeds chemical accuracy and is inaccessible by classical supercomputers, however this number is commonly estimated to be on the order of hundreds of spin-orbitals or logical qubits.^{11,15,74} The $\mathcal{O}(N^4)$ scaling of the number of two-body interaction terms in the electronic structure Hamiltonian in Eq 1.8 with the number of spin-orbitals N implies that there can be as many as $\approx N^4$ interaction terms which need to be mapped to Pauli strings. Optimistically assuming that we can choose a system size with as few as $N = 100$ spin-orbitals, this would give approximately 10^8 Pauli

strings in the resulting qubit Hamiltonian. Optimistically assuming that a single Trotter step is sufficient to suppress the second order and higher terms in Eq. 2.24, this would imply that our QPE circuit would need to consist of 10^{11} multi-qubit gates of the form $e^{ih_j \hat{P}_j t}$. Now consider that what we actually need are the controlled versions of each of these 10^{11} gates compiled down to a series of 1 and 2-qubit gates. The number of the gates involved in the compilation of each of these gates will likely add at least another order of magnitude to the gate count and will depend on the compilation technique used. However, at this point we do not need to delve any deeper to see the challenge associated with implementing QPE on near-term devices, which have 1 and 2-qubit gate errors typically on the order of 10^{-3} to 10^{-2} for superconducting qubits.⁷⁵ We can assume that in the absence of error correction, a circuit with 10^{11} gates (even if they were 1 and 2-qubit gates, which these are not without further compilation) with these error rates would easily be dominated by errors. This would be true even if physical gate errors were improved by several orders of magnitude. Even under the assumption that we do not encounter the worst case scenario with N^4 Pauli terms in the Hamiltonian, but rather something optimistic such as N^2 , this would only bring the (non-compiled) gate count down to 10^7 . Thus, there is still a gap of several orders of magnitude between the number of gates that non-fault tolerant quantum computers will be able to coherently carry out and the number of gates required to do even the smallest (potentially) useful computations with quantum phase estimation. This is why phase estimation in the context of quantum many-body problems is often assumed to require fault-tolerant error correction, putting it firmly in the "long-term" category of quantum algorithms for solving the electronic structure problem.

There have been several works over the years which have done much more careful estimates of the resources required to use phase estimation to solve many-body problems that are currently inaccessible by classical computers.^{11,13,15,76} For reasons described above, these works assume the need for fault-tolerant error correction. The error correction code used is usually assumed to be the surface code^{36,39-41} because its ability to tolerate physical gate errors on the order of 10^{-2} and its assumption that only a 2D lattice of qubits with limited

connectivity are available makes its implementation on platforms such as superconducting qubits simpler than other codes. For this reason these resource estimates measure the cost associated with implementing QPE by the number of T (or possibly Toffoli) gates in the circuit, which are by far the most costly to implement in the surface code. Estimates for logical gate counts in these works are typically on the order of anywhere from 10^{10} to 10^{18} . The logical gate counts estimates in these works are subject to a host of assumptions chosen by the authors (*e.g.* the system size being considered, the target accuracy, and implementation details such as whether the time evolution operator is implemented using the Trotter formula⁷¹ or qubitization⁷⁷), however the picture they paint in the aggregate is clear: phase estimation will almost certainly never be used to solve useful problems in chemistry without fault-tolerant error correction. Further algorithmic improvements may bring these counts down, but they are unlikely to close the gap completely. Furthermore, the number of physical qubits needed to carry out circuits of these depths with sufficiently low logical error using the surface code is well beyond that which will be available on near-term quantum computers. For example, it has been estimated that calculating the ground state energy of the FeMoco molecule (54 electrons in a basis of 108 spin-orbitals) to an accuracy of 0.1 milliHartree would require at least 10^5 to 10^6 physical qubits.¹¹ However, it should be noted that qLDPC codes^{37,38} have the potential to reduce these qubit requirements considerably, should they ever be successfully implemented on a physical device. At time of writing, the number of physical qubits available on superconducting devices has only recently surpassed 10^3 with IBM's 1,121 superconducting qubit Condor chip. The infeasibility of producing physical error rates low enough to implement QPE non-fault-tolerantly and the near-term unavailability of machines with enough qubits to implement QPE fault-tolerantly has motivated the search for algorithms which can be implemented with lower circuit depth: variational eigensolvers.

3. Variational Eigensolvers Background

In this chapter we review some of the methods in the literature for solving the electronic structure problem using variational quantum-classical hybrid algorithms. In Sec. 3.1 we summarize the motivation behind variational methods and the generic framework by which variational algorithms are formulated. In Sec. 3.2 we review how the variational quantum eigensolver (VQE) algorithm attempts to solve the ground state problem as well as the various algorithmic subcomponent choices involved in its implementation. In Sec. 3.3 we review several methods in the literature which have extended VQE to find excited states as well. In Sec. 3.4 we review some of the works which have estimated the resources associated with implementing VQE and the implications these findings have for the feasibility of implementing VQE on near-term noisy hardware.

3.1 The Variational Framework

The deep circuits of phase estimation arose from the fact that a relatively large number of repetitions ($M \approx 10^3$) of a controlled unitary \hat{U} were required to reach chemical accuracy. That is, the circuit depth and maximum attainable accuracy were inextricably linked in its construction. This leads to a natural question: are there quantum algorithms for which the accuracy is less strongly dependent on the circuit depth? It is reasonable to conjecture that this could be the case for many chemical systems of interest. After all, phase estimation itself already required the preparation of a state which has high overlap with the true eigenstate. The caveat to this is that this state did not necessarily need to be sufficiently close to the exact eigenstate to reach chemical accuracy. The amount of additional circuit depth that would be needed to reach chemical accuracy is unclear. From this perspective, there are two main questions: **1.** are there quantum circuits with depth shallow enough to prepare chemically accurate states on quantum computers with limited or no error correction and **2.** can such circuits be found efficiently? Implicitly, we mean that we want to answer these questions for chemical systems of sizes where no classical methods are expected to attain chemical accuracy in the near-term. If we were able to do this, we could obtain the

eigenvalues through statistical sampling of the expectation value of the Hamiltonian with respect to this solution state. One popular approach to this problem in recent years has been to attempt to find such circuits variationally using a parameterized ansatz.

Variational algorithms re-frame problems as the minimization of some objective function $g(\mathbf{x})$, where \mathbf{x} is some vector of parameters. We do not know the solution of the problem, but we know enough to formulate a g such that at its global minimum, the parameters and the value of the function encode the solution. In the context of quantum computing, we can formulate functions of the parameters of ansatz circuits $\hat{U}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters $(\theta_1, \theta_2, \dots, \theta_n)$. We refer to the set of all possible parameters $\boldsymbol{\theta}$ as the parameter space. The set of all circuits $\hat{U}(\boldsymbol{\theta})$ such that $\boldsymbol{\theta}$ is in the parameter space is referred to as the ansatz space. We want to choose a function f and an ansatz $\hat{U}(\boldsymbol{\theta})$ such that:

$$\min_{\boldsymbol{\theta}} f(\hat{U}(\boldsymbol{\theta})) \approx \min_{\mathbf{x}} g(\mathbf{x}) \quad (3.1)$$

where $f(\hat{U}(\boldsymbol{\theta}))$ (which from this point forward we will abbreviate as $f(\boldsymbol{\theta})$) is comprised of quantities which can be measured on a quantum computer for any value of $\boldsymbol{\theta}$. Variational quantum algorithms (VQAs) find the global minimum iteratively through a series of circuit preparations and measurements followed by parameter updates. A quantum computer repeatedly prepares the state $\hat{U}(\boldsymbol{\theta})|\psi_i\rangle$ for some initial reference state $|\psi_i\rangle$ and performs measurements to calculate the value of $f(\boldsymbol{\theta})$, along with possibly its gradient and Hessian. A classical optimization algorithm receives this information and updates the value of $\boldsymbol{\theta}$ according to a rule which is used to prepare the updated circuit until the stopping condition of the optimizer is reached. The procedure for a generic VQA is as follows.

1. Choose an ansatz $\hat{U}(\boldsymbol{\theta})$, a reference state $|\psi_i\rangle$, and an initial set of parameters $\boldsymbol{\theta}_0$. Set $k = 0$.
2. Repeatedly prepare the circuit $\hat{U}(\boldsymbol{\theta}_k)|\psi_i\rangle$ and perform measurements sufficiently many times to calculate $f(\boldsymbol{\theta}_k)$ (along with possibly its gradient and Hessian) to some target accuracy.

3. Give the estimated value of $f(\boldsymbol{\theta}_k)$ and other necessary quantities to a classical optimization algorithm. If the stopping condition of classical optimizer is met, halt the algorithm. Else, return $\boldsymbol{\theta}_{k+1}$ according to optimizer rule.
4. Set $k = k + 1$.
5. Repeat steps 2-4 as necessary.

We note that this is not quite the most general case, as one could imagine many possible extensions for particular algorithms (*e.g.* allowing f to be a function of several different ansatz circuits), however this is the base procedure that many methods share.

3.2 The Variational Quantum Eigensolver

In the context of the ground state problem in classical computational chemistry, our goal would be to minimize the objective function:

$$g(\mathbf{x}) = \frac{\langle \psi(\mathbf{x}) | \hat{H} | \psi(\mathbf{x}) \rangle}{\langle \psi(\mathbf{x}) | \psi(\mathbf{x}) \rangle} \quad (3.2)$$

where $|\psi(\mathbf{x})\rangle$ denotes our parameterized state.¹ \mathbf{x} could be the coefficients of a CI expansion of a wavefunction, for instance. This is motivated by the well-known variational principle:

$$\frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \geq E_0 \quad (3.3)$$

which states that for any normalized state $|\psi\rangle$ in the Hilbert space \mathcal{H} of a hermitian operator \hat{H} , the expectation value of \hat{H} with respect to $|\psi\rangle$ must be greater than or equal to its lowest eigenvalue E_0 .^{1,22} In this context E_0 is the ground state energy of a Hamiltonian \hat{H} . For any \mathbf{x} parameterizing a wavefunction expanded in a set of basis states which span \mathcal{H} (or possibly a subspace of \mathcal{H}), the corresponding normalized, parameterized wavefunction will be in \mathcal{H} . Thus the ground state E_0 is a lower bound for any such parameterized state and we can minimize Eq. 3.2 to estimate E_0 . When we want to solve the ground state problem variationally on a quantum computer, the function we minimize is the expectation value of the Hamiltonian with respect to our ansatz state parameters:

$$f(\boldsymbol{\theta}) = \langle \psi_i | \hat{U}^\dagger(\boldsymbol{\theta}) \hat{H} \hat{U}(\boldsymbol{\theta}) | \psi_i \rangle \quad (3.4)$$

where $|\psi_i\rangle$ is some non-parameterized state that we use to initialize our circuit. This method is known as the variational quantum eigensolver (VQE).^{9,78–82}

3.2.1 VQE Measurements

We note that since \hat{H} is a linear combination of Pauli strings as in Eq. 1.64, the measurement of its expectation value is expressed as:

$$\langle \psi(\boldsymbol{\theta}) | \hat{H} | \psi(\boldsymbol{\theta}) \rangle = \sum_i h_i \langle \psi(\boldsymbol{\theta}) | \hat{P}_i | \psi(\boldsymbol{\theta}) \rangle \quad (3.5)$$

where $|\psi(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})|\psi_i\rangle$. Thus, we take measurements for each of the \hat{P}_i individually and compute their weighted sum classically. Suppose we take n_i independent measurements for each \hat{P}_i . By the central limit theorem, the variance ϵ_i^2 of the quantity $\langle \psi(\boldsymbol{\theta}) | \hat{P}_i | \psi(\boldsymbol{\theta}) \rangle$ approaches $\text{Var}(h_i \hat{P}_i)/n_i$ in the limit of large n_i , where $\text{Var}(h_i \hat{P}_i)$ is the variance of the probability distribution of measurement outcomes of the observable $h_i \hat{P}_i$.⁷⁰ The total number of measurements M needed to compute the expectation value of the Hamiltonian to error ϵ can be upper bounded by:^{5,14}

$$M \leq \frac{(\sum_i |h_i|)^2}{\epsilon^2}. \quad (3.6)$$

This implies that the spacetime cost of VQE scales with the error ϵ as $\mathcal{O}(\frac{1}{\epsilon^2})$, which is notable as this is asymptotically worse than QPE by a factor of $\frac{1}{\epsilon}$. In the worst case scenario, VQE also scales with the number of spin-orbitals N as $\mathcal{O}(N^4)$ as there are potentially $\mathcal{O}(N^4)$ Pauli strings to measure. Importantly, since these $\mathcal{O}(N^4/\epsilon^2)$ measurements are independent, they can be parallelized across many quantum processors, distributing at least a fraction of the cost over space rather than time. Note that because a general Pauli string can contain Pauli \hat{X} , \hat{Y} , and \hat{Z} operators, we might naively assume that we need a measurement device that is physically versatile enough to perform projective measurements

in the eigenbases of all three of these operators for each of the qubits. This turns out to not be the case for the simple reason that we can perform a change of basis on the circuit level rather than physically altering the measurement device. To see this, consider that the \hat{X} and \hat{Y} operators can be diagonalized according to:

$$\hat{X} = H\hat{Z}H \quad (3.7a)$$

$$\hat{Y} = R_x^\dagger\left(-\frac{\pi}{2}\right)\hat{Z}R_x\left(-\frac{\pi}{2}\right) \quad (3.7b)$$

where $R_x(\theta) = e^{-i\frac{\theta}{2}\hat{X}}$. This implies that if a Pauli string contains either an \hat{X} or \hat{Y} operator on the j th qubit, we would append either an H or $R_x\left(-\frac{\pi}{2}\right)$ gate to the end of the circuit on the j th qubit, respectively and perform a projective measurement in the Pauli- \hat{Z} basis.⁵ No gate is applied to a qubit if the Pauli string contains either a \hat{Z} or \hat{I} operator for that qubit index. This can be seen from the simple example of measuring a 2-qubit Pauli string $\hat{X}_1\hat{Y}_2$ with respect to the state $\hat{U}|\psi_i\rangle$:

$$\langle\psi_i|\hat{U}^\dagger\hat{X}_1\hat{Y}_2\hat{U}|\psi_i\rangle = \langle\psi_i|\hat{U}^\dagger H_1 R_x^\dagger\left(-\frac{\pi}{2}\right)_2 \hat{Z}_1 \hat{Z}_2 H_1 R_x\left(-\frac{\pi}{2}\right)_2 \hat{U}|\psi_i\rangle. \quad (3.8)$$

Equipped with a method for calculating $f(\boldsymbol{\theta})$, we may now also compute its gradient $\nabla f(\boldsymbol{\theta})$. The most intuitive method for doing so would be the finite difference method, which approximates the partial derivative of $f(\boldsymbol{\theta})$ with respect to each θ_i by perturbing each θ_i by a small quantity δ and computing the quantities:

$$\frac{\partial f}{\partial \theta_i} \approx \frac{f(\dots, \theta_i + \delta, \dots) - f(\dots, \theta_i - \delta, \dots)}{2\delta} \quad (3.9)$$

where all parameters θ_j with $i \neq j$ remain unperturbed. This becomes exact in the limit $\delta \rightarrow 0$. Note that while is not inherently problematic on classical computers (where machine epsilon is on the order of 10^{-15}), this is inherently more difficult to compute on a quantum computer than $f(\boldsymbol{\theta})$ by itself due to the need to resolve two values of f in parameter space which differ by only a tiny perturbation δ using statistical sampling. The more accurately

we want to approximate $\nabla f(\boldsymbol{\theta})$, the smaller we need to make δ and the more difficult it will be to resolve the two quantities in the numerator of Eq. 3.9. It can be shown that an upper bound for the number of measurements n_i needed to evaluate $\nabla f(\boldsymbol{\theta})$ to error ϵ resembles Eq. 3.6, but picks up a multiplicative factor of $\frac{1}{\delta^2}$ along with a multiplicative factor equal to the number of parameters.⁵ The finite difference formula is quite general in that it works in the limit of small δ for any function whose first partial derivatives exist, however we are not interested in the most general case. We only care about functions of parameterized quantum circuits. These quantum circuits are comprised of sequences of parameterized gates that have a structure that we can exploit, as demonstrated through the parameter-shift method.⁸³⁻⁸⁷ It has been shown that according to the parameter-shift rule, if all the parameterized gates in a quantum circuit (with no restriction on constant gates such as CNOT) can be written in the form:

$$\hat{U}_k(\theta_k) = e^{-i\theta_k \hat{H}_k/2} \quad (3.10)$$

where \hat{H}_k is a hermitian operator with two distinct eigenvalues, the sinusoidal nature of these gates (along with the product rule) leads an exact expression for the partial derivatives:

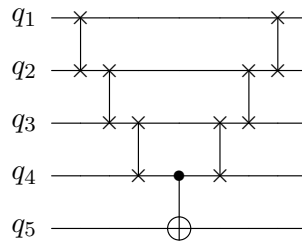
$$\frac{\partial f}{\partial \theta_k} = \frac{f(\dots, \theta_k + \delta, \dots) - f(\dots, \theta_k - \delta, \dots)}{2 \sin(\delta)} \quad (3.11)$$

for any δ such that $\sin(\delta) \neq 0$. Higher order derivatives can be found by successive re-applications of this rule. Eq. 3.11 has the direct implication that the $\frac{1}{\delta^2}$ dependence of the number of measurements needed for the finite difference approximation is replaced by $\frac{1}{\sin(\delta)^2}$ for the parameter-shift rule. Since this expression is exact, we can choose any δ that does not cause the denominator to vanish. Minimizing the number of measurements corresponds to choosing $\delta = \frac{\pi}{2}$.⁸⁴ Note that the requirement that each parameterized gate is an exponential of a hermitian operator with two distinct eigenvalues places an additional restriction on its application that was not present for the finite difference method. Not all

ansatz are automatically of this form, but those that are not can often be decomposed into a form that is.⁸⁵

3.2.2 The Ansatz Circuit

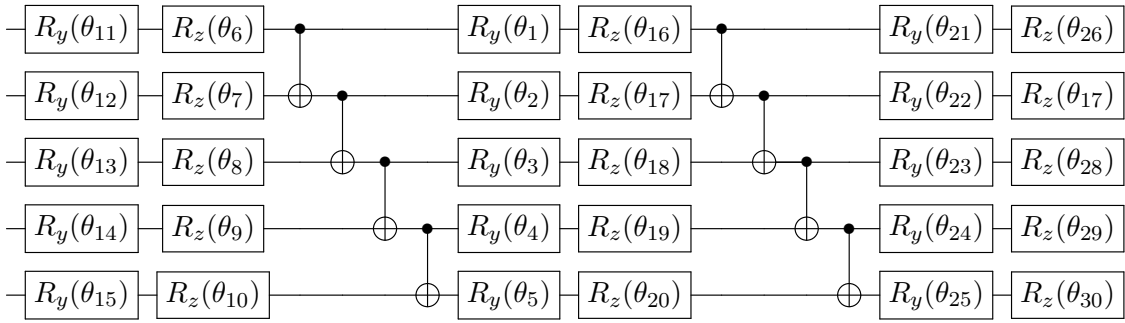
In the context of quantum many-body problems, ansatz circuits can be loosely grouped into two categories: hardware-efficient ansatzes and physically-motivated ansatzes. Hardware-efficient ansatzes^{88,89} attempt to minimize the circuit depth overhead associated with compilation by using gate sequences that are native to the basis gate set and qubit topology of a particular machine. That is, they are formulated only in terms of gates that do not require any further compilation into a longer sequence of gates. Both of these are important considerations, particularly for near-term devices which can only natively implement a small discrete set of basis gates and which often (at least for superconducting qubit devices) have limited nearest-neighbor connectivity. For example, if an ansatz contains a gate which entangles two qubits q_1 and q_2 which are not directly connected on the machine, then this gate must be decomposed via a sequence of entangling gates between a set of intermediary qubits which do connect them. An example of such a procedure would be a swap network.⁹⁰ These types of compilation procedures can be quite costly if we want to entangle two qubits which are quite far away on the connectivity graph of the machine. As an illustrative example, consider the problem of implementing a CNOT gate between two qubits q_1, q_5 that are connected by a sequence of qubits q_2, q_3, q_4 . One way to implement this operation would be to "move" q_1 to the spot occupied by q_4 via a sequence of SWAP gates, perform a CNOT with q_5 , then "move" q_1 back to its original spot. The circuit representation of this procedure is given by:



Using the fact that SWAP can be decomposed into three CNOT gates as:


(3.12)

we can see that implementing a non-native CNOT gate between q_1 and q_5 requires a total of 19 CNOT gates. In general (with the caveat that this is not a proof of optimality), implementing a non-native CNOT between two qubits separated by n qubits on the connectivity graph would require $3n + 1$ CNOT gates using this compilation method. Given this, the motivation for using hardware-efficient ansatzes on near-term noisy machines is clear. An example of a hardware-efficient ansatz for a machine with 1D nearest-neighbor connectivity and whose native gate set includes R_z , R_y , and CNOT would be:



The drawback to these types of ansatzes is that they are known to induce cost function landscapes whose gradient (on average over the parameter space) vanishes exponentially with the number of qubits. This is the so-called barren plateaus problem.⁸⁹ This has the implication that if one does not have a good initial guess for the parameters that is likely to be close to the solution, then the cost function landscape at the starting point is likely to be extremely flat, hindering convergence to the global minimum or preventing it altogether. Note that hardware-efficient ansatzes such as the one shown above are in a sense problem-agnostic. This is often a highly undesirable feature as there is no problem-specific structure that can be used to speed up convergence. For example, in quantum chemistry we often want to restrict the wavefunction to be in a particular subspace of a Hilbert space \mathcal{H}

corresponding to a specific number of particles and spin-magnetization. Hardware-efficient ansatzes in general do not enforce these symmetries unless we purposefully design them to do so⁹¹ or use a qubit mapping scheme which enforces them at the operator level.⁵⁶ Thus, in the absence of such symmetry enforcement our optimization procedure will need to search through a much larger space of states.

Physically-motivated ansatzes attempt to exploit problem-specific structure in order to increase the chances that the ansatz can well-approximate the solution and that the solution is found quickly. A well-known example of a physically-motivated ansatz is the Unitary Coupled-Cluster ansatz,^{5,86,92,93} motivated by the non-unitary coupled-cluster method in classical computational chemistry.¹ Recall from Eq. 1.21a that classical coupled-cluster adopts the wavefunction ansatz $|\psi_{cc}\rangle = e^{\hat{T}}|HF\rangle$, where $|HF\rangle$ is the Hartree-Fock state and \hat{T} is an anti-hermitian parameterized sum of fermionic excitation operators. The success of non-unitary coupled-cluster in classical computational chemistry tells us that states which take this form can often represent good approximations to chemical states. Thus, we may want to develop a similar version of it which can be prepared on a quantum computer. We can obtain a unitary analog of this ansatz by taking the exponential of $\hat{T} - \hat{T}^\dagger$:⁵

$$\hat{U}_{UCC} = e^{\hat{T} - \hat{T}^\dagger} \quad (3.13a)$$

$$\hat{T} - \hat{T}^\dagger = \sum_{i>j} \theta_j^i \left(\hat{a}_i^\dagger \hat{a}_j - \hat{a}_j^\dagger \hat{a}_i \right) + \sum_{\substack{i>j \\ k>l}} \theta_{kl}^{ij} \left(\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l - \hat{a}_l^\dagger \hat{a}_k^\dagger \hat{a}_j \hat{a}_i \right) + \dots \quad (3.13b)$$

where in general we can include up to any degree- N_e excitation terms for systems with N_e electrons. We can map this to a qubit operator just like we did for the system Hamiltonian and approximate it according to the first order Trotter-Suzuki formula in Eq. 2.24. The naming convention for UCC of varying excitation levels follows that of non-unitary classical CC. The most common method is to truncate the cluster operator at degree-2 excitations (which we refer to as UCCSD), then simply increase the number of Trotter steps if necessary to expand the ansatz space. Eq. 3.13 is quite flexible in that we can formulate a number

of different UCC variants based on the types of excitations we include. UCCSD typically assumes that half of the indices in each term run over a set of orbitals which are occupied in the Hartree-Fock state and the other half run over orbitals which are unoccupied. Generalized UCCSD (UCCGSD) includes excitations terms from all orbitals, regardless of whether or not they are occupied in the Hartree-Fock state. Pair UCCD (Up-CCD) includes only pairwise excitations from two electrons occupying the same spatial orbital with opposite spins. Up-UCCGD performs pairwise excitations with the generalized form of UCC.⁹² Adding the prefix n (*e.g.* n -UCCSD) specifies that n Trotter steps were used in approximating Eq. 3.13. Note that this ansatz automatically preserves the number of particles in the wavefunction as each excitation term contains an equal number of creation and annihilation operators. We can further enforce the preservation of the spin-magnetization quantum number by including only excitation terms which do not change the number of spin-up and spin-down electrons. These restrictions can significantly speed up the convergence to the solution state. Additionally, we note that because the UCC ansatz is equivalent to the identity when all parameters are zero, it is straightforward to choose a good initialization for the optimization by choosing a physically motivated reference state $|\psi_i\rangle$ such as Hartree-Fock or truncated CI. This is not necessarily the case for many hardware-efficient ansatzes unless one adds additional gates to the state initialization to cancel out the non-identity gates when all ansatz parameters are set to zero. The drawback to the UCCSD ansatz is that it is challenging to keep the circuit depth sufficiently shallow while also retaining the excitation operators which are important for achieving chemical accuracy. We can see this from a quick order of magnitude estimation of the number of cluster amplitudes, given that the number of these amplitudes scales as $\mathcal{O}(N^2 N_e^2)$ for a system of N spin-orbitals and N_e electrons.⁵ Consider the FeMoco molecule consisting of 108 spin-orbitals and 54 electrons that was previously used as a benchmark system for resource estimations of QPE.¹¹ Such a system would require a number of cluster amplitudes on the order of 10^7 under this scaling argument. This is many orders of magnitude below the $\approx 10^{14} - 10^{15}$ reported for QPE, however it is clearly still several orders of magnitude in excess of what non-fault tolerant

machines with physical error rates on the order of 10^{-3} are capable of implementing coherently. It should be noted that this 10^7 estimate does not take into account costs associated with the qubit mapping and gate compilation steps, which would likely add at least a few more orders of magnitude to this number, depending on the techniques used. More careful resource estimates of the number of 2-qubit gates (which are the more error-prone gates in the non-fault tolerant case) required for representing the H_2O molecule (10 electrons) in a basis size of $\approx 10^2$ spin-orbitals using the UCCSD ansatz in the Jordan-Wigner encoding put this number on the order of 10^5 , even when including circuit optimizations such as gate-cancellations and gate filtering based on pre-screening of amplitudes.⁷⁴ Notably, adaptive UCC-style ansatzes which attempt to lower the circuit depth by starting out with a circuit with no gates and iteratively adding excitation terms as needed to improve accuracy throughout the course of the optimization procedure have been proposed.⁹⁴⁻⁹⁶ The drawback to these adaptive methods is that they require additional measurements in order to determine which excitation operators should be appended to the ansatz.

3.2.3 Classic Optimization Subroutine

An optimizer is an algorithm which accepts a set of information (*e.g.* the value of a function, its gradient, its Hessian, or some subset of these quantities) that can be expressed as a function of some high-dimensional vector of parameters $\boldsymbol{\theta}_k$ and returns a new set of parameters $\boldsymbol{\theta}_{k+1}$ according to an update rule. By iteratively updating the parameters according to this rule, we hope to generate a sequence of parameters $\{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k, \dots\}$ such that the sequence $\{f(\boldsymbol{\theta}_0), f(\boldsymbol{\theta}_1), \dots, f(\boldsymbol{\theta}_k), \dots\}$ converges as close as possible to the global minimum of f with as few iterations as possible. This is of course not guaranteed in general. The quality and speed of this convergence will depend on what optimizer is used and the nature of the cost function landscape of f . For example, the optimizer may be unable to distinguish the global minima from local minima and saddle points, leading it to converge to a point in the parameter space which does not correspond to the solution. The convergence may also be too slow to be useful. The presence of noise on near-term

quantum computers is expected to exacerbate the difficulty of these challenges.

The prototypical example of an optimizer is gradient descent, which calculates the gradient of a function at a point and updates the parameter in the opposite direction of the gradient:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \tau_k \nabla f(\boldsymbol{\theta}_k) \quad (3.14)$$

for some small step size τ_k . Variants of gradient descent have been used in the quantum computing literature. The quantum natural gradient,^{97,98} for instance, attempts to remedy the fact that traditional gradient descent does not take into consideration the geometry that the ansatz induces on the cost function landscape. Without knowledge of this geometry, gradient descent may choose a sub-optimal parameter update that may, for instance, cause it to overshoot the global minimum or calculate a $|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k|$ that is smaller than necessary, resulting in slow convergence. In this method, the quantum natural gradient update method is formulated by replacing $\nabla f(\boldsymbol{\theta}_k)$ in Eq. 3.14 with $F_k^{-1} \nabla f(\boldsymbol{\theta}_k)$, where F_k is the Fubini-Study metric of the ansatz at the point $\boldsymbol{\theta}_k$. This method was shown using numerical simulations to converge more quickly than traditional gradient descent.⁹⁷ Methods which stochastically approximate both the gradient⁸⁸ and quantum natural gradient⁹⁹ have also been studied. Instead of analytically solving for these quantities at each iteration, these methods choose a uniformly random direction in parameter space, calculate the directional derivative in this direction, then carry out gradient descent-like parameter updates using this quantity. These methods go by the names Simultaneous Perturbation Stochastic Approximation (SPSA) and Quantum Nature SPSA (QNPSA). Whereas the analytic counterparts of these methods always update the parameter in the direction of steepest descent within a given geometry, these methods move in the direction of steepest descent only in an average sense. The benefit to this is that each iteration only requires a constant number of function evaluations. Furthermore, SPSA has been used for experimental demonstrations of VQE on quantum hardware,⁸⁸ demonstrating a certain robustness to

noise. Examples of other well-known optimizers include L-BFGS-B¹⁰⁰ and COBYLA.¹⁰¹ L-BFGS-B is a quasi-Newton method, wherein the Hessian is not calculated directly, but rather is approximated using a history of gradient evaluations from previous iterations. COBYLA is an example of a gradient-free simplex method. It keeps track of a collection of points, performs a linear fit to a hyperplane in parameter space using these points, then minimizes the objective function on this hyperplane subject to some bounds on the parameters. The point corresponding to this minimum then takes the place of one of the other points previously used. This process repeats until the region encompassed by this collection of points is sufficiently small.

3.3 Excited States Solvers

VQE found the ground state of a Hamiltonian by minimizing its expectation value with respect to a parameterized ansatz state $|\psi(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})|\psi_i\rangle$. This is often sufficient for electronic structure chemistry at room temperature, where the average thermal energy is often insufficient for exciting a molecule from its ground state. There are many other instances, however, where a molecule may become excited through some means other than thermal energy. For instance, we may wish to study a reaction mechanism wherein a molecule must be excited by a photon before it can react with another chemical species, *i.e.* photochemistry.⁶ Furthermore, at points in the potential energy surface where energy levels cross (*e.g.* conical intersections) or almost cross, it is possible for a molecule to transition between electronic states without the absorption or emission of a photon.⁷ For this reason, much work in recent years has been dedicated to extending VQE to find low-lying excited states of the electronic structure Hamiltonian as well. Here we will discuss a few of them. As we will see, this introduces additional challenges beyond those already present in the ground state case, much of which stems from the need to enforce the mutual orthogonality of the set of solution states.

3.3.1 Quantum Subspace Expansion

The principle behind quantum subspace expansion (QSE)^{102–105} is simple. Suppose we have carried out the VQE algorithm, which has returned some (hopefully accurate) approximate solution for the ground state $|\tilde{E}_0\rangle$. We can construct a set of operators $\{\hat{O}_i\}$ (one of which must be the identity) and seek excited states solutions of the form:

$$|E_j\rangle = \sum_i c_{ij} \hat{O}_i |\tilde{E}_0\rangle \quad (3.15)$$

for some coefficients c_{ij} for which we solve. The set of states $\{\hat{O}_i |\tilde{E}_0\rangle\}$ will span some subspace of the Hilbert space of the system Hamiltonian \hat{H} . We can generate a contracted Hamiltonian matrix H^c in this subspace using these states as a basis. The matrix elements of H^c are given by:

$$H_{ij}^c = \langle \tilde{E}_0 | \hat{O}_i^\dagger \hat{H} \hat{O}_j | \tilde{E}_0 \rangle. \quad (3.16)$$

The solution eigenstates can be found by solving the generalized eigenvalue equation:

$$H^c C = S C E \quad (3.17)$$

where S is the overlap matrix whose elements are given by:

$$S_{ij} = \langle \tilde{E}_0 | \hat{O}_i^\dagger \hat{O}_j | \tilde{E}_0 \rangle, \quad (3.18)$$

C is the coefficients matrix whose elements are given by c_{ij} , and E is a diagonal matrix whose i th entry is the i th eigenvalue.¹⁰² Thus by running VQE, performing additional measurements to compute these matrix elements, then solving this generalized eigenvalue problem, we can obtain estimates for low-lying excited states. There are three main points to briefly discuss here. The first is that in general the number of measurements that must be taken will be much higher than that required by VQE. The electronic structure Hamiltonian \hat{H} has $\mathcal{O}(N^4)$ terms. This implies that if use n_o operators in our set $\{\hat{O}_i\}$, then

the number of matrix elements H_{ij}^c that we need to measure scales as $\mathcal{O}(N^4 n_o^2)$. This can amount to quite a substantial number of additional measurements unless we are extremely restrictive regarding the number of operators we include in our expansion. For example, if our operators consisted of all possible single-particle fermionic excitation operators $\{\hat{a}_i^\dagger \hat{a}_j\}$, then the number of measurements needed scales as $\mathcal{O}(N^6)$. Including all double excitation operators would raise this cost to $\mathcal{O}(N^8)$. The second point to note is the surface-level similarity this method has with truncated configuration interaction methods in the classical computing setting. Instead of expanding using the Hartree-Fock state as a reference state in our wavefunction expansion, we use an estimate for the ground state $|\tilde{E}_0\rangle$. Thus we can conjecture that this method might suffer from some of the same accuracy issues that CI does when scaled up to systems with large number of electrons where higher order excitation terms become non-negligible. The counter-argument to this of course is that $|\tilde{E}_0\rangle$ may already contain some of these higher order excitation terms and that a lower number of further excitation terms may be necessary than in the case where the Hartree-Fock state was used as a reference. The third point is that we can note that there is a sense in which QSE doubles as an error mitigation technique for the quality of the state $|\tilde{E}_0\rangle$.¹⁰² In fact, QSE has been proposed as an error mitigation technique in a much broader context than just quantum chemistry.¹⁰⁵ To see this, we can simply note that in the case where $|\tilde{E}_0\rangle$ is not the exact ground state, a better estimate for the exact ground state may be contained in the span of $\{\hat{O}_i|\tilde{E}_0\rangle\}$, which would be found when solving Eq. 3.17. Thus QSE may partially correct certain types of errors such as incomplete convergence in the VQE problem.

3.3.2 Folded Spectrum VQE

In the folded spectrum method, one performs a "shift and square" operation on the Hamiltonian \hat{H} :

$$\hat{H}_i^{fs} = (\hat{H} - \lambda_i \hat{I})^2 \quad (3.19)$$

for each excited state $|E_i\rangle$. $\hat{H}_i^{fs}|E_i\rangle = (E_i - \lambda_i)^2|E_i\rangle$, thus minimizing the expectation value of \hat{H}_i^{fs} with respect to an ansatz will return the eigenstate of \hat{H} whose eigenvalue is closest to λ_i .¹⁰⁶ By providing such a λ_i for each excited state and running this optimization problem for each excited state, we can solve for each excited state. The drawbacks to this method are twofold. The first is that the number of Pauli terms in \hat{H}_i^{fs} will scale as the number of terms in \hat{H} squared. That is, if \hat{H} contains $\mathcal{O}(N^4)$ terms, then we must now measure $\mathcal{O}(N^8)$ Pauli terms, which can be prohibitively expensive. The second drawback is that we have to properly choose the hyperparameters λ_i in order to find the correct states. This estimation does not have to be particularly precise, with the only requirement being that it must be closer to the target eigenvalue than any of the others. In instances where the gaps between eigenvalues become smaller, we expect that this task will become more difficult.

3.3.3 Variational Quantum Deflation

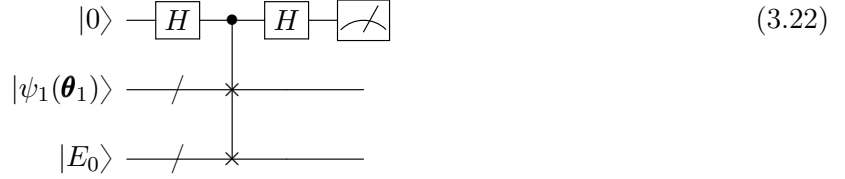
In variational quantum deflation (VQD),^{107,108} one assumes that we have already run VQE to find a ground state $|E_0\rangle = \hat{U}_0(\boldsymbol{\theta}_0^{opt})|\psi_0\rangle$, where \hat{U}_0 is the ansatz used to find the ground state, $\boldsymbol{\theta}_0^{opt}$ are the optimal ansatz parameters found in VQE, and $|\psi_0\rangle$ is the initial reference state used by VQE. To find the first excited state $|E_1\rangle$, VQD minimizes the expectation value of the effective Hamiltonian:

$$\hat{H}_1 = \hat{H} + \beta|E_0\rangle\langle E_0| \quad (3.20)$$

for some real-valued $\beta > 0$. $|E_0\rangle$ is an eigenstate of \hat{H}_1 with eigenvalue $E_0 + \beta$. If we choose β such that $E_0 + \beta > E_1$, then $|E_0\rangle$ is the first excited state of \hat{H}_1 and $|E_1\rangle$ is its ground state. Assuming we have chosen a β that meets this condition, we can minimize the expectation value of \hat{H}_1 with respect to the parameters of an ansatz state $|\psi_1(\boldsymbol{\theta}_1)\rangle = \hat{U}_1(\boldsymbol{\theta}_1)|\psi_1\rangle$. In practice we do not decompose the operator $|E_0\rangle\langle E_0|$ as this would potentially involve expanding it in an operator basis of exponentially-many Pauli strings, but instead minimize the objective function:

$$\langle \psi_1(\boldsymbol{\theta}_1) | \hat{H}_1 | \psi_1(\boldsymbol{\theta}_1) \rangle = \langle \psi_1(\boldsymbol{\theta}_1) | \hat{H} | \psi_1(\boldsymbol{\theta}_1) \rangle + \beta |\langle \psi_1(\boldsymbol{\theta}_1) | E_0 \rangle|^2. \quad (3.21)$$

The overlap term $|\langle \psi_1(\boldsymbol{\theta}_1) | E_0 \rangle|^2$ can be computed in several different ways. The most intuitive method is to prepare the state $\hat{U}_{ref,0}^\dagger \hat{U}^\dagger(\boldsymbol{\theta}_0^{opt}) \hat{U}(\boldsymbol{\theta}_1) |\psi_1\rangle$, where $\hat{U}_{ref,1} |0\rangle = |\psi_0\rangle$, and perform repeated projective measurements in the computational basis. The overlap term will be given by the probability that this measurement outcome yields the "all-zero" computational basis state $|0\rangle$. Note that the circuit depth required to compute this quantity is the sum of the ansatz circuits used for each state. A second option would be to use the SWAP test.¹⁰⁹



where the multi-qubit controlled SWAP gate means that we apply controlled SWAP gates for each pair of qubits between the states $|E_0\rangle$ and $|\psi_1(\boldsymbol{\theta}_1)\rangle$. This implies that the SWAP test required circuit depth scaling with the number of qubits N as $\mathcal{O}(N)$ on top of the circuit depth required to prepare $|E_0\rangle$ and $|\psi_1(\boldsymbol{\theta}_1)\rangle$. Additionally, we require $2N + 1$ total qubits to measure the overlap between two N -qubit states. The output of this circuit (pre-measurement) is given by:

$$\frac{1}{2} \left(|E_0\rangle |\psi_1(\boldsymbol{\theta}_1)\rangle + |\psi_1(\boldsymbol{\theta}_1)\rangle |0\rangle \right) + \frac{1}{2} \left(|E_0\rangle |\psi_1(\boldsymbol{\theta}_1)\rangle - |\psi_1(\boldsymbol{\theta}_1)\rangle |1\rangle \right). \quad (3.23)$$

Measurement of the ancilla qubit yields either 0 or 1 with probabilities:

$$P(0) = \frac{1}{2} + |\langle \psi_1(\boldsymbol{\theta}_1) | E_0 \rangle|^2 \quad (3.24a)$$

$$P(1) = \frac{1}{2} - |\langle \psi_1(\boldsymbol{\theta}_1) | E_0 \rangle|^2. \quad (3.24b)$$

Through repeated circuit preparation we can estimate the overlap to error ϵ with a number of measurements that scales as $\mathcal{O}(\frac{1}{\epsilon^2})$. It is also worth noting that it has been shown that one can compute this overlap with $2N$ qubits, a subcircuit which uses N CNOT and Hadamard gates to entangle the two states, and classical post-processing which scales linearly with N .¹¹⁰ Regardless of the method used to calculate the overlap, it is clear that doing so comes with additional costs beyond that required to prepare the two states individually. It is up to us whether the additional cost we want to incur is to double the number of qubits or to increase the circuit depth.

Once we have used this method to find the first excited state, we can repeat this process for successively higher excited states, adding an overlap penalty term after each state that we find. The objective function that we minimize for finding the n th excited state is given by:

$$\langle \psi_n(\boldsymbol{\theta}_n) | \hat{H}_n | \psi_n(\boldsymbol{\theta}_n) \rangle = \langle \psi_n(\boldsymbol{\theta}_n) | \hat{H} | \psi_n(\boldsymbol{\theta}_n) \rangle + \sum_{i=0}^{n-1} \beta_i |\langle \psi_n(\boldsymbol{\theta}_n) | E_i \rangle|^2. \quad (3.25)$$

The main drawback (besides the increased resource costs) to this method is that finding n eigenstates requires n optimization procedures which cannot be parallelized. This was not the case for the folded spectrum VQE and QSE methods where the additional steps required for finding the excited states could be parallelized to a large extent. Furthermore, because each iteration uses the parameters found for the previous excited state in the previous iteration, errors incurred in one iteration will cascade into all further iterations.¹⁰⁷

It remains to be seen how problematic this is in practice.

3.3.4 State-Averaged Excited States Solvers

Here we discuss two excited states solvers: the subspace-search variational quantum eigensolver¹¹¹ (SSVQE) and the multistate, contracted variational quantum eigensolver¹¹² (MCVQE). These two methods are quite similar, yet appear to have been developed independently by different authors and contain subtle differences that are not necessarily

unimportant. Both enforce the mutual orthogonality of the solution states by exploiting the fact that unitary transformations preserve the inner products of pairs of states. That is, if we have two states $|\psi_1\rangle$ and $|\psi_2\rangle$ to which we apply a unitary transformation \hat{U} to yield the states $|\tilde{\psi}_1\rangle$ and $|\tilde{\psi}_2\rangle$, then:

$$\langle \tilde{\psi}_1 | \tilde{\psi}_2 \rangle = \langle \psi_1 | \hat{U}^\dagger \hat{U} | \psi_2 \rangle = \langle \psi_1 | \psi_2 \rangle. \quad (3.26)$$

If we have a set of n reference states $\{|\psi_0\rangle, |\psi_1\rangle, \dots, |\psi_{n-1}\rangle\}$ that are chosen to be mutually orthogonal, then they will remain orthogonal if we apply the same ansatz circuit $\hat{U}(\boldsymbol{\theta})$ with the same parameter values to all of them. MCVQE and SSVQE both minimize an objective function of the form:

$$f(\boldsymbol{\theta}) = \sum_i^{n-1} w_i \langle \psi_i | \hat{U}^\dagger(\boldsymbol{\theta}) \hat{H} \hat{U}(\boldsymbol{\theta}) | \psi_i \rangle \quad (3.27)$$

where the weights w_i are positive real numbers. Where the two methods begin to diverge is in how they choose the weights w_i and reference states $\{|\psi_i\rangle\}$. SSVQE chooses the weights w_i to be distinct positive numbers that decrease with increasing energy level. Using this choice and under the assumption that the ansatz is sufficiently expressive and that the optimization converges to the global minimum, the ansatz \hat{U} maps the reference states to the eigenstates $\{|E_i\rangle\}$. The authors suggest choosing the reference states to be computational basis states as they are trivial to prepare. MCVQE on the other hand chooses the weights w_i to be equal positive-valued numbers. Under this choice of weights, the ansatz no longer maps the reference states to the exact eigenstates, but rather a set of states which span the low-lying eigenspace. The choice of equal weights introduces a degeneracy in the global minimum of $f(\boldsymbol{\theta})$ that allows the ansatz to map each of the reference states to many different possible linear combinations of the low-lying eigenstates. MCVQE solves for the exact eigenstates through a diagonalization post-processing step following the minimization of $f(\boldsymbol{\theta})$. For optimal parameters $\boldsymbol{\theta}^{opt}$, the contracted Hamiltonian \hat{H}^c with matrix elements given by:

$$\hat{H}_{ij}^c = \langle \psi_i | \hat{U}^\dagger(\boldsymbol{\theta}^{opt}) \hat{H} \hat{U}(\boldsymbol{\theta}^{opt}) | \psi_j \rangle \quad (3.28)$$

is computed and subsequently diagonalized. The authors show that if one can prepare states of the form $\frac{1}{\sqrt{2}}(|\psi_i\rangle \pm |\psi_j\rangle)$, then measurement of the expectation value of \hat{H} with respect to the states $\frac{1}{\sqrt{2}}\hat{U}(\boldsymbol{\theta}^{opt})(|\psi_i\rangle \pm |\psi_j\rangle)$ allows for the off-diagonal matrix elements to be solved for algebraically. If one can prepare CI states for instance, then linear combinations of CI states can also be prepared by adding their coefficients together and renormalizing the state.¹¹² Note the similarity between this diagonalization procedure and the one used by QSE. The advantage MCVQE has over QSE in this regard is that the lack of a need to expand some reference state with a potentially large number of operators means that the dimensionality of the contracted Hamiltonian is much smaller and can be constructed using far fewer measurements.

One appealing feature of these state-averaged methods is that the mutual orthogonality of the solution states is handled automatically, which eliminates the need for the types of overlap circuits used by VQD which explicitly added resource costs relative to those required to prepare the ansatz. In principle this is advantageous. Whether or not this is advantageous in practice is another question entirely and would need to be investigated numerically. The authors of the SSVQE paper themselves note that in their own numerical testing, SSVQE required a deeper ansatz circuit than VQE to properly represent the excited states.¹¹¹ Intuitively, we expect that the task of finding a single ansatz which can simultaneously map a set of reference states to a set of eigenstates is more difficult than the task of finding multiple ansatz circuits which can separately map a set of reference states to a set of eigenstates. To what extent this is a problem in practice has yet to be seen, however we note that if SSVQE or MCVQE requires roughly twice the circuit depth as an ansatz used by VQD, then there is no advantage in terms of circuit depth. Furthermore, with VQD we had the option of using twice the number of qubits rather than (roughly) twice the circuit depth, which is not an option that we have with these state-averaged approaches.

3.4 VQE Resource Estimates

The primary motivation behind the variational quantum eigensolver (and by extension its excited states variants) was that it could serve as an alternative to phase estimation in instances where we do not yet have machines with the qubit counts or logical gate fidelities needed to run the deep circuits that phase estimation requires. The principle behind VQE was that it is reasonable to conjecture that there could exist circuits that represent chemically-accurate eigenstates with gate counts much lower than those used in QPE. We rid ourselves of circuits which are known to be deep, but we also traded the rigorous runtime bounds of QPE for the non-rigorous heuristics of VQE. We can choose an ansatz which is shallow, but we do not have any guarantee that it can represent the eigenstate we want to the degree of accuracy that we want. We can choose an objective function and a classical optimizer subroutine, but we have no guarantee on quick and reliable convergence to the global minimum, particularly in the presence of noise, which was found to induce a "barren-plateau"-like effect.¹¹³ Furthermore, we traded the $\mathcal{O}(\frac{1}{\epsilon})$ runtime scaling of QPE for the $\mathcal{O}(\frac{1}{\epsilon^2})$ spacetime scaling required to measure a single observable to error ϵ in VQE. Thus, in order to counter the sub-optimal scaling of VQE measurements, we would need roughly 10^3 quantum processors (assuming we are aiming for chemical accuracy) working in parallel to match the runtime scaling of QPE. The heuristic nature of VQE makes it difficult to estimate what the total wall clock runtime of a hypothetical simulation would be or what the minimum ansatz circuit depth needed would be for any given system. As a result, works attempting to estimate such quantities often make assumptions that require a considerable degree of speculation. Some only estimate the resources needed to perform a single iteration of VQE since the number of iterations needed to reach the ground state are unknown. Gonthier et al. estimated the wall clock runtime to perform one energy evaluation for several molecules to accuracy 5×10^{-4} Hartree under the (rather optimistic) assumption that we can prepare a noiseless hardware-efficient ansatz scaling as $5N - 3$ with the number of spin-orbitals N .⁷³ They estimated that this runtime would be roughly 1.9 days for the

case of the methane molecule (8 electrons in 104 spin-orbitals) and 71 days for the case of the ethanol molecule (20 electrons in 260 spin-orbitals). Bearing in mind that this energy estimation subroutine would likely have to be performed on the order of at least thousands of times, this demonstrates that even under the assumptions that an exceptionally shallow ansatz can be found and that neither error mitigation nor error correction is necessary, the number of measurements that VQE requires presents itself as a major bottleneck in the computation that cannot be ignored. Given that Kühn et al.⁷⁴ estimated that circuits with 2-qubit gate counts on the order of 10^5 using the UCCSD ansatz are needed to simulate a molecule with ≈ 100 spin-orbitals to an accuracy bounded above by CCSD(T), these wall clock estimates are likely overly optimistic by at least an order of magnitude. Wecker et al.¹⁴ estimated that using their proposed Hamiltonian Variational Ansatz (HVA), whose gate count scales as $\mathcal{O}(N^4)$, simulating the Fe_2S_2 molecule in a basis of 112 spin-orbitals would require about 10^{13} measurements per energy evaluation. If we take these numbers and assume that we can lower bound the circuit depth as $\approx (10^2)^4/10^2 = 10^6$ by parallelizing as many of these $\mathcal{O}(N^4)$ gates across the $\approx 10^2$ qubits as we can and if we use their assumption that each gate takes $10^{-6}s$ to implement, then each energy evaluation would take roughly on the order of $10^{13}s \approx 3 \times 10^5$ years. This runtime is clearly infeasible. We should also keep in mind that 10^2 spin-orbitals is roughly what is speculated to be the *smallest* problem size for which quantum computers may demonstrate an advantage. In order to go beyond the demonstration of a few niche chemical systems to the point where quantum computers have widespread application to a large number of systems, we would likely have to consider numbers of spin-orbitals well beyond this minimum. Furthermore, it could be that the estimate of 10^2 spin-orbitals is overly optimistic. It is difficult to say in advance whether an active space of any given size will be sufficient for approaching the infinite basis set limit accurately enough to have useful predictive power. The prediction of any given number of qubits being sufficient to have useful application therefore has a speculative component to it. For example, the FeMoco molecule with an active space of 108 spin-orbitals with 54 electrons that was previously proposed to be one target system of interest⁷⁶ has been criticized for

likely being overly-optimistic in the size of the active space size that would be required.¹³ Furthermore, the authors of this critique note that FeMoco contains molybdenum, an atom which could be heavy enough to potentially merit relativistic treatments which would be beyond the capabilities of the non-relativistic electronic structure Hamiltonian. What all of these resource estimates show in the aggregate is that VQE in its original unaltered formulation is unlikely to provide a viable path towards demonstrating quantum advantage for calculating eigenstates of molecular Hamiltonians. Algorithmic improvements of the subcomponents involved will be required. For this reason, much of the work (including our own) centers around proposing various methods for improving the efficiency of these subcomponents.

4. The Quantum Orbital Minimization Method

This chapter contains material that we have published in the Journal of Chemical Theory and Computation on the development of the quantum Orbital Minimization Method (qOMM).¹¹⁴ We begin by motivating and introducing the method in Sec. 4.1. In Sec. 4.2 we present the results for our numerical simulation of the qOMM algorithm and its comparison to the SSVQE algorithm. In Sec. 4.3 we discuss the results and potential future directions of research.

4.1 qOMM Algorithm

Much research attention in recent years has been devoted to developing variational quantum algorithms for solving the electronic structure problem in chemistry. This has been motivated by the fact that the quantum phase estimation algorithm^{8,60} requires circuits which are far too deep to feasibly implement it on near-term hardware. By contrast, variational algorithms such as VQE⁹ only require circuits that can prepare eigenstates of the Hamiltonian to chemical accuracy. The solution states are found by iteratively preparing a parameterized ansatz state, measuring a set of observables, and using a classical optimization subroutine to minimize an objective function that depends on these observables. VQE is an eigensolver which solves for the ground state, however in recent years there has been much work that extends VQE to find excited states as well. The quantum subspace expansion algorithm¹⁰³ first finds a ground state wavefunction by running VQE, generates a subspace spanned by a set of operators which act on the VQE ground state, then carries out a generalized eigenvalue problem in this subspace to find estimates for the low-lying excited states. (VQD)^{102,107} finds the low-lying eigenstates through a sequence of minimization problems where each iteratively adds terms which penalize the overlap with previously found eigenstates. The folded spectrum method¹⁰⁶ can find the eigenstate whose eigenvalue is closest to a hyperparameter λ by minimizing the expectation value of the squared Hamiltonian $(\hat{H} - \lambda\hat{I})^2$ in the same fashion as VQE. MCVQE¹¹² and SSVQE¹¹¹ both take a state-averaged approach where the weighted sum of the expectation values of \hat{H}

with respect to a set of mutually-orthogonal states is minimized. In order to ensure mutual orthogonality is maintained, the same ansatz with the same parameter values is applied to a set of mutually orthogonal reference states throughout the optimization procedure. SSVQE chooses the weights in this summation to be distinct, allowing its global minimum to correspond to the low-lying eigenstates. MCVQE chooses these weights to be equal, thus its global minimum consists of a set of states which span the low-lying eigenspace. A classical diagonalization post-processing step is conducted in this eigenspace in order to obtain the eigenvectors themselves.

One way to find novel methods for developing methods for finding excited states on quantum computers is to take inspiration from objective function methods which exist in the classical computing literature. Here we extend the Orbital Minimization Method¹¹⁵ to the quantum computing case and test its validity and performance using classical numerical simulations. The numerical linear algebra formulation of the OMM objective function in the classical literature is given by:

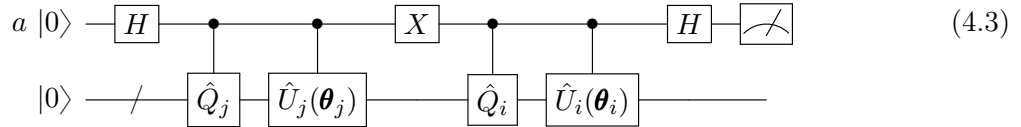
$$f(X) = \text{tr} \left((2I - X^\dagger X) X^\dagger H X \right) \quad (4.1)$$

where H is the Hamiltonian in matrix form and X is a $N \times K$ dimensional matrix whose K columns consist of N -dimensional statevectors representing our guesses for the K lowest eigenstates of H . This objective function has the interesting property that all local minima are equal in value to the global minima and represent valid solutions to the low-lying eigenspace.¹¹⁵ This is a highly desirable property since it means that if we minimize this objective function with respect to X , we know that it cannot converge to spurious local minima that do not represent valid solutions. The caveat to this is that it can still converge to saddle points or points in parameter space that are locally flat enough for the stopping condition of the optimizer to be activated. Nonetheless, there is sufficient motivation to investigate its usage in the quantum computing setting. If we want to represent f as a function of a set of parameterized ansatz states $\{\hat{U}_i(\boldsymbol{\theta}_i)|\psi_i\rangle\}$, we would replace each column

of X with one of the states in this set. This gives us the objective function:

$$\begin{aligned}
f(\boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_{K-1}) &= \sum_{i=0}^{K-1} \langle \psi_i | \hat{U}_i^\dagger(\boldsymbol{\theta}_i) \hat{H} \hat{U}_i(\boldsymbol{\theta}_i) | \psi_i \rangle \\
&- \sum_{\substack{i \neq j \\ i, j=0}}^{K-1} \langle \psi_i | \hat{U}_i^\dagger(\boldsymbol{\theta}_i) \hat{U}_j(\boldsymbol{\theta}_j) | \psi_i \rangle \langle \psi_j | \hat{U}_j^\dagger(\boldsymbol{\theta}_j) \hat{H} \hat{U}_i(\boldsymbol{\theta}_i) | \psi_i \rangle.
\end{aligned} \tag{4.2}$$

Here we have added subscripts i and j to both the parameters and their respective ansatz states. This is to emphasize the flexibility that we can choose different ansatz forms for each state and that in general each ansatz will have parameters with different values. Moreover, unlike SSVQE and MCVQE, we have the flexibility to choose reference states $|\psi_i\rangle$ that are not mutually orthogonal. The overlap terms in the double summation determine that the mutual orthogonality is enforced at the global minimum rather than at every point in the cost function landscape. Once this objective function is minimized, we get a set of mutually orthogonal states which span the low-lying eigenspace of \hat{H} . In order to obtain the eigenstates themselves, we perform a diagonalization post-processing step in this eigenspace in an analogous manner to those used in QSE and MCVQE. The expectation value terms in the first summation are calculated in the same standard way as in VQE. In order to measure quantities of the form $\langle \psi_i | \hat{U}_i^\dagger(\boldsymbol{\theta}_i) \hat{U}_j(\boldsymbol{\theta}_j) | \psi_i \rangle$, we carry out Hadamard-test¹¹⁶ style circuits of the form:



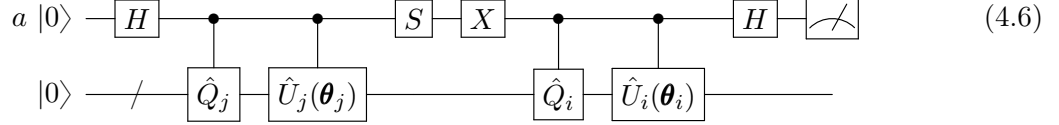
where $\hat{Q}_i|0\rangle = |\psi_i\rangle$. This circuit outputs the state:

$$\frac{1}{2} \left(|0\rangle \left(\hat{U}_j(\boldsymbol{\theta}_j) |\psi_j\rangle + \hat{U}_i(\boldsymbol{\theta}_i) |\psi_i\rangle \right) + |1\rangle \left(\hat{U}_j(\boldsymbol{\theta}_j) |\psi_j\rangle - \hat{U}_i(\boldsymbol{\theta}_i) |\psi_i\rangle \right) \right). \tag{4.4}$$

Measurement of the ancilla qubit then yields outcomes $|x\rangle$ with probabilities:

$$P(x) = \frac{1}{2} + \frac{(-1)^x}{2} \text{Re} \langle \psi_i | \hat{U}_i^\dagger(\boldsymbol{\theta}_i) \hat{U}_j(\boldsymbol{\theta}_j) | \psi_j \rangle. \quad (4.5)$$

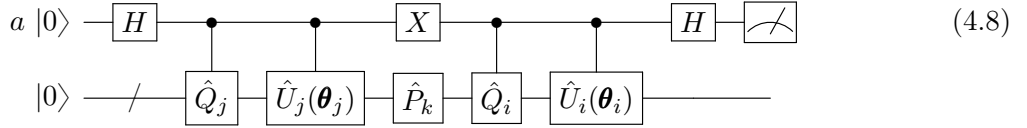
By carrying out a similar circuit, but with an S gate inserted between the first controlled \hat{U} gate and the X gate on the ancilla qubit:



we impart a relative phase $e^{\pi i/2}$ into the ancilla qubit, which allows us to analogously compute the imaginary part of the overlap from the probabilities:

$$P(x) = \frac{1}{2} + \frac{(-1)^x}{2} \text{Im} \langle \psi_i | \hat{U}_i^\dagger(\boldsymbol{\theta}_i) \hat{U}_j(\boldsymbol{\theta}_j) | \psi_j \rangle. \quad (4.7)$$

Note that if we choose an ansatz circuit such that the wavefunctions have purely real amplitudes, we do not have to compute the imaginary part. We can then estimate these overlap terms with a number of measurements that scales with the target error ϵ as $\mathcal{O}(\frac{1}{\epsilon^2})$. The overlap terms $\langle \psi_i | \hat{U}_i^\dagger(\boldsymbol{\theta}_i) \hat{H} \hat{U}_j(\boldsymbol{\theta}_j) | \psi_j \rangle$ can be computed in a directly analogous way. Recalling that the Hamiltonian \hat{H} is expressed as a linear combination of Pauli string operators \hat{P}_k with weights h_k , we can run circuits of the form:



to compute the real part. To obtain the imaginary part, we simply run the circuit directly analogous to that in Eq. 4.6. We then repeat this for all Pauli strings in the expansion of \hat{H} and compute the weighted sum classically. We make one final important note about this objective function before presenting the numerical results. The classical OMM objective function has no spurious local minima, however the quantum OMM (qOMM) objective

function in Eq. 4.2 may very well have spurious local minima that could impede its convergence. This is largely due to the fact that we represent each state as a parameterized quantum circuit, whereas in the classical case each state was parameterized in a linear way as a statevector. The quantum circuit parameterization will cause the states to be parameterized in a highly non-linear way, which may induce spurious local minima that do not represent valid solutions. Furthermore, there are of course no guarantees that our choice of ansatz circuits will be sufficiently expressive to represent the solution states. Our numerical simulations serve in part to investigate how much of an issue both of these factors are in practice.

4.2 Numerical Results

We present the numerical results of our proposed method obtained from a classically simulated quantum computer. We apply it to the problem of finding low-lying eigenvalues of the electronic structure Hamiltonian for near-equilibrium configurations of two different molecules as well as H_4 , a toy model consisting of 4 hydrogen atoms arranged in a square lattice. The numerical tests in this section serve to explore what advantages or disadvantages may arise for the qOMM and weighted SSVQE approaches (using weight vectors of the form $[n, n - 1, \dots, 1]$ for finding n states) in different situations. Section 4.2.1 presents the results for H_2 and H_4 obtained from noiseless simulations. Section 4.2.2 presents the results obtained from simulating H_2 with a depolarizing noise model. Section 4.2.3 presents the results for LiH obtained from noiseless simulations.

All code for these simulations is implemented using the Qiskit¹¹⁷ library, with SSVQE being implemented as a modification of Qiskit’s VQE implementation. The electronic structure Hamiltonians were generated using molecular data from PySCF¹¹⁸ in the STO-3G basis and mapped to qubit Hamiltonians using the Jordan-Wigner mapping for the noiseless simulations and the Parity mapping for the noisy simulations. The relative accuracies of the results for both methods are calculated by comparing them to their numerically exact counterparts obtained by numerically exact diagonalization of the Hamiltonians. Circuits

involved in noiseless qOMM simulations were conducted using Qiskit’s *StateVector* simulator. Circuits involved in noiseless SSVQE simulations were conducted using Qiskit’s *QasmSimulator* in conjunction with the *AerPauliExpectation* method for computing expectation values in order to improve runtime performance. Both of these methods produce ideal, noiseless results. Simulations for finding the energy levels of H₂ were performed at an interatomic distance of 0.735 Angstroms, those of LiH were performed at an interatomic distance of 1.595 Angstroms, and those of H₄ were performed at an interatomic distance of 1.23 Angstroms. Simulation results for the hydrogen square model and H₂ at stretched bond distances can also be found in Appendix A. For the LiH simulations, we freeze the two core electrons to reduce the problem size from 12 qubits to 10 qubits.

The set of initial states for both algorithms and all molecular Hamiltonians were chosen to be the Hartree-Fock state and low-lying single-particle excitations above it. For example, if we wish to find three energy levels for 4-qubit H₂, in the Jordan-Wigner mapping this would be the set: $\{|01\rangle_\alpha|01\rangle_\beta, |01\rangle_\alpha|10\rangle_\beta, |10\rangle_\alpha|01\rangle_\beta\}$. For LiH, this would be the set: $\{|00001\rangle_\alpha|00001\rangle_\beta, |00001\rangle_\alpha|00010\rangle_\beta, |00010\rangle_\alpha|00001\rangle_\beta\}$, where the subscript α and β denote the \hat{S}_z quantum number of the orbitals.¹ Given that the UCCSD⁵ ansatz preserves the numbers of spin-up and spin-down electrons, the above choice of initial states constrains these variational quantum algorithms to search only the $\hat{S}_z = 0$ subspace, lessening the computational difficulty and producing physically meaningful results at the same time. For this reason, UCCSD is our ansatz of choice for these simulations. If we wish to find only two energy levels, we would omit the highest-energy state from the set. This choice of the set of states allows us to study two different types of algorithm initializations: one in which all of the ansatz parameters are randomly sampled according to a uniform distribution on $[-2\pi, 2\pi)$ and another one in which they are all set to zero. With the random parameter initialization, we can study the robustness of the algorithms with respect to their

¹ In the Qiskit implementation of n -qubit basis states in the Jordan-Wigner representation, the first $\frac{n}{2}$ qubits encode spin-up (α) and the second $\frac{n}{2}$ qubits encode spin-down (β). Thus we have chosen these three basis states because they are elements of the 2-particle, spin magnetization $S_z = 0$ subspace of the full 2^n -dimensional Fock space. Orbitals of a given spin are ordered from right to left in ascending Hartree-Fock energies.

Table 4.1: The success rate of given problem instances converging to within a relative accuracy of 10^{-5} for a given number of UCCSD repetitions using a randomized ansatz parameter initialization.

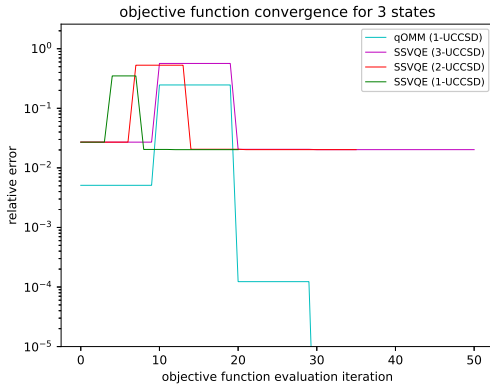
Molecule	Algorithm	UCCSD Repetitions						
		1-rep	2-rep	3-rep	4-rep	5-rep	6-rep	7-rep
H ₂ (0.735 Å)	qOMM (3 states)	100%	—	—	—	—	—	—
	SSVQE (3 states)	0%	70%	100%	—	—	—	—
H ₂ (1.47 Å)	qOMM (3 states)	100%	—	—	—	—	—	—
	SSVQE (3 states)	0%	90%	100%	—	—	—	—
H ₄ (1.23 Å)	qOMM (2 states)	0%	100%	—	—	—	—	—
	qOMM (3 states)	0%	100%	—	—	—	—	—
	SSVQE (2 states)	—	0%	100%	—	—	—	—
	SSVQE (3 states)	—	—	0%	100%	—	—	—
H ₄ (2.46 Å)	qOMM (2 states)	0%	90%	—	—	—	—	—
	qOMM (3 states)	0%	100%	—	—	—	—	—
	SSVQE (2 states)	—	0%	100%	—	—	—	—
	SSVQE (3 states)	—	—	0%	100%	—	—	—
LiH (1.595 Å)	qOMM (2 states)	0%	100%	—	—	—	—	—
	qOMM (3 states)	0%	100%	—	—	—	—	—
	qOMM (4 states)	—	100%	—	—	—	—	—
	qOMM (5 states)	—	100%	—	—	—	—	—
	qOMM (6 states)	—	100%	—	—	—	—	—
	qOMM (7 states)	—	100%	—	—	—	—	—
	SSVQE (2 states)	0%	80%	—	—	—	—	—
	SSVQE (3 states)	0%	0%	70%	—	—	—	—
	SSVQE (4 states)	—	—	0%	100%	—	—	—
	SSVQE (5 states)	—	—	—	10%	100%	—	—
	SSVQE (6 states)	—	—	—	—	10%	100%	—
SSVQE (7 states)	—	—	—	—	—	60%	100%	

starting point in the parameter space. With the “zero vector” initialization, the UCCSD ansatz circuit is initialized to the identity and the states after the initial ansatz circuit application remain the Hartree-Fock state and low-lying single-particle excitations above the Hartree-Fock states. In Qiskit’s implementation of ansatz circuits such as UCCSD, one can increase the expressiveness of the ansatz by repeating the corresponding quantum circuit block pattern r times, which comes at the cost of both the circuit depth and the number of variational parameters increasing by a factor of r . In this work, we refer to an ansatz circuit that consists of the UCCSD circuit block pattern repeated r times as r -repetition UCCSD

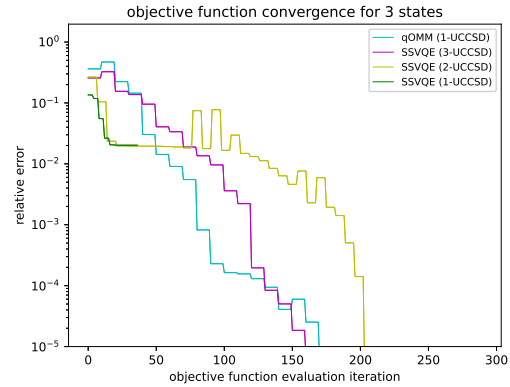
(or simply r -UCCSD). We run both algorithms for several different numbers of repetitions in order to account for the fact that the necessary ansatz expressiveness for convergence is not known *a priori*. Such a study allows us to explore the dependence of convergence success on the circuit depth for each algorithm. When using a random initialization, we run each algorithm ten times for each setting. In Sections 4.2.1 and 4.2.3, for each setting we plot one run roughly representative of the average convergence. In the noisy results in Section 4.2.2, we plot the run which obtained the lowest objective function convergence. The complete set of all ten runs for all of these simulations are plotted in Appendix A and their convergence success rates are summarized in Table 4.1 for completeness. When using the “zero vector” initialization, we run each algorithm only once. Running the algorithm multiple times for this initialization is neither necessary nor useful in the absence of noise because the outcome is deterministic for a given initial point.

4.2.1 H₂ and H₄

We begin by presenting the results for the 4 and 8 qubit systems we consider here: H₂ and the hydrogen square model H₄. Figure 4.1 illustrates the convergence of the objective function for each algorithm when attempting to calculate the first three energy levels of the H₂ molecule at the equilibrium bond distance. Figure 4.1b demonstrates how qOMM and SSVQE both converge to the global minimum, but notably, SSVQE cannot do so with just 1-UCCSD, requiring two repetitions in order to converge to within a relative accuracy of 10^{-5} . The success rate is further reported in Table 4.1. Figure 4.1a illustrates an attempt to solve the same problem, except the ansatz parameters are all initialized to zero. In this instance, we see that qOMM converges more quickly by a factor of roughly 5 compared to its random initialization counterpart, while SSVQE gets stuck in a local minimum and does not converge regardless of the ansatz circuit depth used. In both figures and all later convergence figures in this chapter, we depict the convergence curve as the relative error (defined as $\frac{|f_{exact} - f_{measured}|}{|f_{exact}|}$) against the number of the objective function evaluations. The L-BFGS-B implementation used in this work uses a 2-point finite difference method

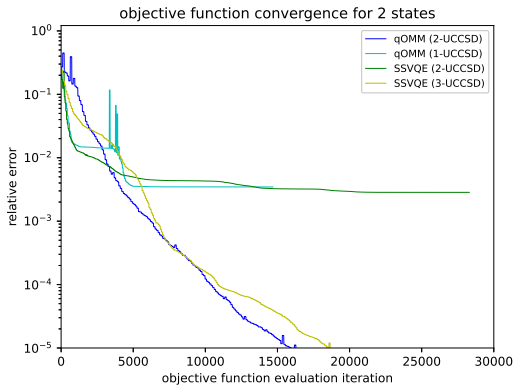


(a) Zero vector initialization

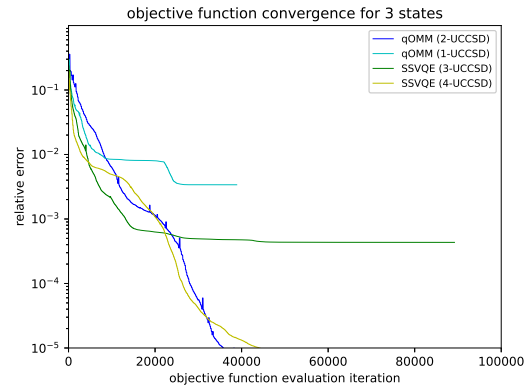


(b) Random initialization

FIGURE 4.1: Convergence (noise-free) of the relative error $\frac{|f_i - f_{exact}|}{|f_{exact}|}$ of qOMM and SSVQE for 3 H_2 states at an interatomic distance of 0.735 \AA .



(a) Two states



(b) Three states

FIGURE 4.2: Convergence (noise-free) of the relative error $\frac{|f_i - f_{exact}|}{|f_{exact}|}$ of qOMM and SSVQE for the hydrogen square model at an interatomic distance of 1.23 \AA using a random parameter initialization.

such that each parameter is perturbed slightly from a given reference point common to all of them. Thus for an n -parameter objective function, $n + 1$ function evaluations are required for each parameter update step in order to estimate the gradient. Thus, in all convergence figures, we observe stair-like curves of width $n + 1$. Figure 4.2 illustrates the convergence results for H_4 at an interatomic distance of 1.23 \AA . We can see from

these figures that similarly to H_2 , the difference between SSVQE and qOMM is primarily in the number of UCCSD repetitions needed to converge. When calculating 2 states, qOMM requires 2-UCCSD whereas SSVQE requires 3-UCCSD. When calculating 3 states, qOMM still requires only 2-UCCSD, whereas SSVQE requires 4-UCCSD. These observations are further tabulated in Table 4.1.

4.2.2 Noisy H_2 simulations

We now run H_2 simulations in the presence of a classically simulated noise model on Qiskit’s *AerSimulator*. Each one-qubit gate in the circuit is modelled as being accompanied by a local depolarizing channel with some probability of error p_{error} . Two-qubit gates are accompanied by a tensor product of two local depolarizing channels acting on each qubit. No error mitigation strategies are employed. We use the Parity mapping in order to use symmetry considerations to reduce the H_2 Hamiltonian to a 2 qubit representation.⁵⁶ We construct ansatzes for each algorithm using the circuit block pattern shown in Figure 4.3. This pattern can be repeated an arbitrary number of times to construct increasingly ex-

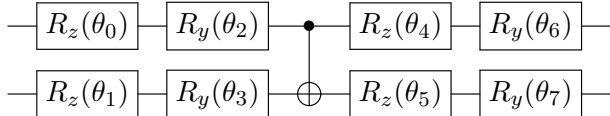
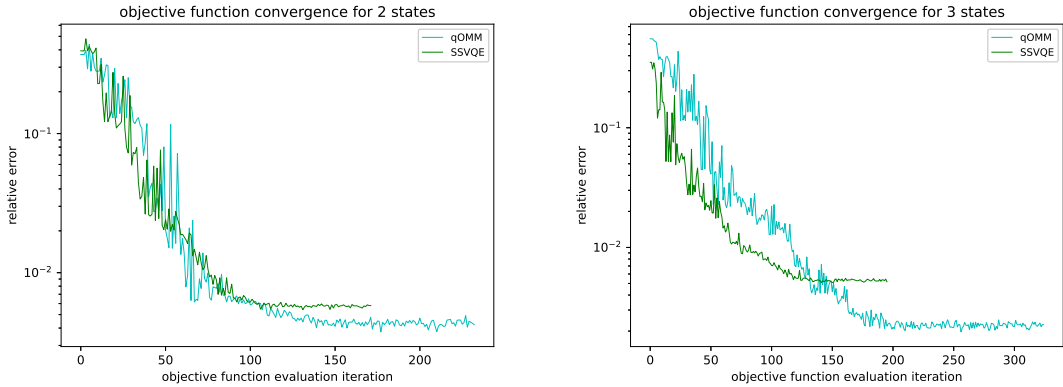


FIGURE 4.3: Base circuit block pattern used for noisy H_2 simulations.

pressive ansatzes in the same way that the number of UCCSD repetitions was varied in the noiseless simulations. The Qiskit compiler is used to compile all circuits to the set of basis gates consisting of R_z , R_y , CNOT, \sqrt{X} , and the identity. For all problem instances, we use the minimum number of repetitions necessary to converge in the absence of noise in order to ensure that we are not simply measuring the ability of the ansatz to represent the global minimum. This corresponds to one repetition for qOMM and two repetitions for SSVQE. We use COBYLA¹⁰¹ as the classical optimization subroutine. We have found this optimizer to be more robust to noise than the L-BFGS-B optimizer used in the noiseless tests. 10^6 circuit samples are used to evaluate all inner product terms and expectation values. p_{error}



(a) Two states

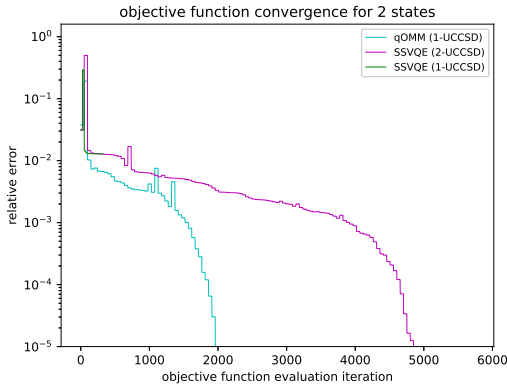
(b) Three states

FIGURE 4.4: Convergence of the relative error $\frac{|f_i - f_{exact}|}{|f_{exact}|}$ of qOMM and SSVQE for H_2 at an interatomic distance of 0.735 \AA , where each circuit gate is modeled as having a probability of local depolarizing error $p_{error} = 0.001$.

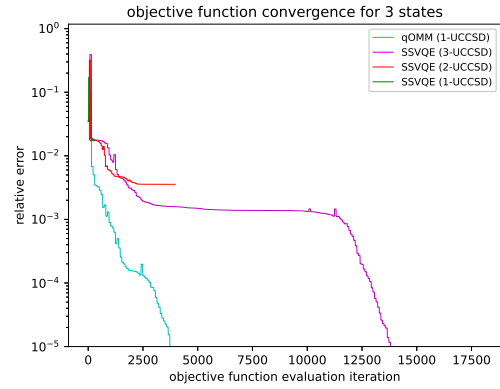
is set to 10^{-3} . Both SSVQE and qOMM are run ten times with randomly initialized parameters. The convergence of the runs which achieved the lowest objective function values are given in Figure 4.4. The results for all ten runs are given in Appendix A. We can see from these figures that both SSVQE and qOMM demonstrate a robustness to noise, although notably they do not achieve chemical accuracy. This is to be expected since the 10^6 shots lower bound the error by 3 digits of accuracy before even taking into consideration the effects of the depolarizing noise. Nonetheless, these tests serve to demonstrate that these two methods are not so fragile that cannot attain anything qualitatively close to their global minima in the presence of modest amounts of noise.

4.2.3 LiH

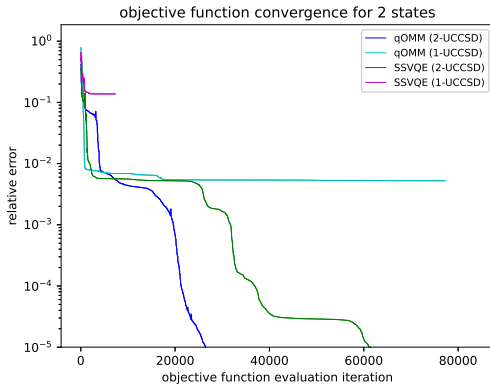
We now present the results for the 10-qubit system we consider here: LiH with two core electrons frozen. Figure 4.5 illustrates the convergence of the objective function for both algorithms with various UCCSD repetitions when calculating up to the first three energy levels. We can see from Figure 4.5c that when the ansatz parameters are randomly initialized, both algorithms demonstrate the ability to converge within an accuracy



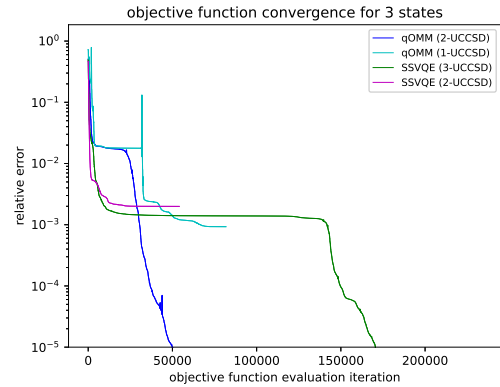
(a) Two states (zero vector initialization)



(b) Three states (zero vector initialization)



(c) Two states (random initialization)



(d) Three states (random initialization)

FIGURE 4.5: Convergence of the relative error $\frac{|f_i - f_{exact}|}{|f_{exact}|}$ of qOMM and SSVQE for LiH.

of 10^{-5} , but require 2-repetition UCCSD to do so. From Figure 4.5a we can see that when all of the parameters are initialized to zero, both algorithms converge to within a relative accuracy of 10^{-5} much more quickly than their randomly initialized counterparts. Notably, qOMM requires only 1-UCCSD to achieve this convergence, while SSVQE requires 2-UCCSD. When three states are calculated, qOMM can achieve a convergence below 10^{-5} with 2-UCCSD, whereas SSVQE requires three repetitions. From Figure 4.5b we can see that when the ansatz parameters are all initialized to zero, both algorithms quickly converge to the global minimum. Notably, qOMM requires only 1-UCCSD repetition with this initialization. When attempting to find larger numbers of states, we can see from Table 4.1

that qOMM can converge with 2-UCCSD for all numbers of states considered, whereas SSVQE requires an increasing number of ansatz repetitions to find increasing numbers of states.

4.3 Discussion

In this section, we discuss three notable aspects of the numerical results for LiH: plateaus in the convergence rate, the effect of initialization choice, and circuit depths required to run each algorithm.

4.3.1 Convergence Rate Plateaus

In many of the LiH convergence plots we show here, both qOMM and SSVQE initially converge at a rapid rate but then hit a plateau and stall the convergence until they escape from these regions of the energy landscape. As long as the algorithm escapes from these regions, the convergence rate resumes being rapid. These regions of the energy landscape (regardless of what their nature may be) seem to be the main obstacle for the convergence of both algorithms, which we observe to be qualitatively more severe for SSVQE. We can further investigate the convergence of each individual eigenvalue as a function of the number of function evaluations in an attempt to gain additional insight and provide a plausible explanation for this behavior. Since the issue is more severe for SSVQE, we focus on the SSVQE rather than qOMM. This may allow us to gain some intuition as to why the SSVQE objective function convergence is observed to plateau in many of the LiH runs before steeply converging below a relative accuracy of 10^{-5} . In Figure 4.6 we plot the eigenvalues and their relative accuracies, respectively, for one of the ten randomly-initialized SSVQE runs. We can see that while E_0 converges rapidly, the convergence of E_1 and E_2 plateau and do not converge for many more function evaluations. E_1 does not escape from its plateau until the accuracy of E_0 is sufficiently decreased. This is followed by the accuracies of E_0 and E_1 increasing together. E_2 does not escape from its plateau until the accuracies of E_0 and E_1 are both sufficiently decreased, after which the accuracies of all three energy levels collectively increase, allowing the objective function to converge to its global minimum.

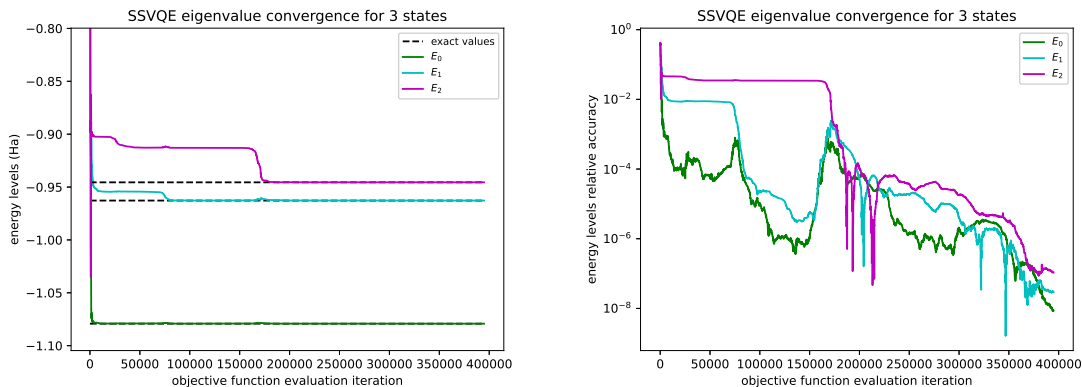


FIGURE 4.6: Relative accuracy $\frac{|E_{n,i} - E_{n,exact}|}{|E_{n,exact}|}$ of the convergence of the first three energy levels of LiH for SSVQE using a random parameter initialization, represented using both linear and logarithmic scaling on the vertical axis.

While we only illustrate this for one of the ten runs, we observe that this qualitative behavior is typical for the other nine runs. We can offer qualitative, speculative explanations for the observation of these plateaus. First, we can observe that these plateaus appear to coincide with points where a subset of the eigenvalues are reasonably converged, but the others are not. The optimizer tends to gravitate to these points (perhaps optimizing a subset of eigenvalues initially offers the steepest gradient) and gets stuck for many iterations. This suggests that there could be some difficulty in simultaneously optimizing multiple eigenvalues using the same ansatz that are not obvious from the construction of SSVQE. From Fig. A.5 in Appendix A we can see that SSVQE requires an increasing number of UCCSD repetitions as the number of excited states is increased, however the severity of these plateaus is diminished with increasing repetitions. In particular, we note that for 6 states, SSVQE with 5-UCCSD can converge, albeit at a considerably slow rate. When the ansatz is increased to 6-UCCSD, the convergence rate is similar to that of qOMM with 2-UCCSD. This would suggest that these observations are largely related to the flexibility of the objective function for each method. Since SSVQE is constrained to use the same ansatz for all states, its only means of being sufficiently flexible to avoid getting stuck in plateaus is to "over-parameterize" the ansatz at the cost of increased circuit depth. qOMM

on the other hand acquires this flexibility through its encoding of each state with separate parameters and enforcing their mutual orthogonality through overlap terms. From Fig. 4.5b and Fig 4.5a we see that starting with a good initial guess can reduce this effect, resulting in significantly improved convergence.

We also briefly note that although we have chosen to measure convergence speed in terms of the number of calls to the objective function, another valid measure of convergence speed would be the total number of optimization iterations if it can be assumed that one can parallelize multiple calls to the objective function. This is because one optimization iteration can require multiple calls to the objective function. In the noiseless results we have presented here, each optimization iteration of the L-BFGS-B optimizer requires $n+1$ calls to the objective function for n total parameters. For problem instances we have presented here where qOMM has more parameters than SSVQE, the speedup of qOMM over SSVQE would be even greater by this measure. For instance, in Fig. 4.5d, qOMM with 2-UCCSD has 144 total parameters and SSVQE with 3-UCCSD has 72.

4.3.2 Initialization

Initialization plays an important role in the convergence for both qOMM and SSVQE. Even for the corresponding VQE ground state problem with 1-repetition UCCSD given in Fig. 4.7, random parameter initialization requires about 100 times more objective function evaluations than the zero vector parameter initialization on average. Both initializations in Figure 4.7 converge to a relative error of 10^{-5} . In the excited states case, the two different initialization strategies for both qOMM and SSVQE can differ not only in the number of iterations but also in whether or not the method converges to the global minimum or a local minimum. Comparing Figs. 4.5a and 4.5c, for SSVQE with 2-repetition UCCSD, random parameter initialization converges more than ten times slower than that of zero vector parameter initialization. Furthermore, we see that qOMM can converge with 1-UCCSD in the zero vector initialization, but requires 2-UCCSD with a random initialization. A similar result can be observed in Figs. 4.5b and 4.5d. Overall, in all results included in this

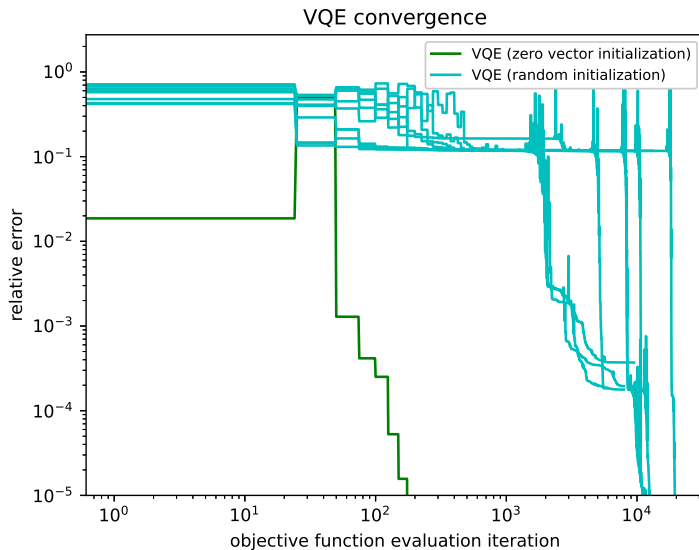


FIGURE 4.7: Convergence of the relative error $\frac{|E_{0,i} - E_{0,exact}|}{|E_{0,exact}|}$ of VQE for the ground state problem of LiH using a random parameter initialization.

work where both types of initializations are tested, zero vector parameter initializations outperform their random parameter initialization counterparts for both algorithms. However, we still include numerical results for random parameter initializations to demonstrate that the ability of these methods to converge is not dependent on this particular choice of initialization. We expect that since this strategy initializes each excited state as a single Slater determinant, its performance would scale poorly with the size of the system and the number of excited states. This "zero vector" initialization is demonstrated primarily not to advocate for its use as a scalable initialization strategy, but to illustrate what the improved convergence could look like if it can be assumed that one has a reasonable initialization strategy. We note that in addition to the overall flexibility of the cost function, qOMM also is more flexible with respect to its initialization. Whereas SSVQE must choose a set of mutually orthogonal reference states with the same set of initial parameters, qOMM is subject to neither of these constraints. For example, one could use a chemically-motivated initialization for low-lying states and a random initialization for higher excited states where one might expect that initialization to break down. Such a mixed initialization strategy

could help avoid being trapped in spurious local minima and benefit from the fast convergence of those states with zero vector parameter initialization. This motivates further investigation into developing and benchmarking initialization strategies more sophisticated than the ones we consider in this work. For example, the MCVQE paper¹¹² proposes an efficient quantum circuit implementation of CIS states for excited state initializations. This initialization would likely improve the convergence results of the simulations for both qOMM and SSVQE presented in this paper. Furthermore, because the optimization stage of MCVQE can be seen as a special case of SSVQE that uses a CIS initialization and an equal weighting of the states, performing analogous simulations using a CIS initialization would offer insight into how the convergence of qOMM compares to that of MCVQE. The approach of using chemically-motivated initializations inspired from classical computational chemistry is likely to further improve the convergence of algorithms such as qOMM and SSVQE. The main challenge in this approach is that chemically-inspired initializations require mapping a classical representation of the state to an efficient quantum circuit, which may be non-trivial for many cases.

4.3.3 Ansatz Circuit Depth

Based on our numerical results, increasing the ansatz circuit depth, *i.e.*, increasing the repetition in UCCSD, has two main effects: **1.** it can increase the ability of the ansatz to represent multiple excited states, and **2.** it can accelerate convergence or help avoid local minima, even in instances where fewer repetitions are demonstrated to be sufficiently expressive. We begin by noting that for ground state of LiH, as in Fig. 4.7, 1-UCCSD is sufficient for the ansatz to approximate the solution state to chemical accuracy. When solving for excited states, this is often not the case and additional circuit depth is needed to represent the solution. When two states of LiH are needed, 1-UCCSD, as shown by the SSVQE results in Fig. 4.5, is not able to simultaneously approximate the ground state and the first excited states using the same parameter values, whereas 2-repetition UCCSD is able to. From the qOMM results in Fig. 4.5a, we know that 1-UCCSD is able to approximate the

ground state and the first excited state separately using different parameter values. When three states are needed, as shown by the SSVQE results in Fig. 4.5b, we need 3-UCCSD to simultaneously approximate three states with the same parameter values. Again, 1-UCCSD is still able to approximate any of these three states separately using different parameter values, as shown by the qOMM results. More generally, 2-UCCSD is sufficient for all LiH cases studied when using qOMM, whereas SSVQE requires n -UCCSD in order to reliably find n states. The results for H_4 in Fig. 4.2 demonstrates a similar pattern, except the situation is more severe for SSVQE, which requires more than n -UCCSD for finding n states. We further report the success rate (out of 10 randomly-initialized runs) for each method to converge to a relative accuracy of at least 10^{-5} using various numbers of UCCSD repetitions in Table 4.1. We see that the number of repetitions needed for qOMM is largely independent of the number of excited states, whereas SSVQE requires an increasingly deep ansatz to converge reliably. Furthermore, we observe instances where increasing the circuit depth may accelerate convergence or help avoid getting stuck in local minima. For instance, in Fig. 4.5c qOMM gets stuck in a local minima with 1-UCCSD, but converges with 2-UCCSD. Again, we know from Fig. 4.5a that this lack of convergence is not a consequence of insufficient expressability. Regarding the rate of convergence, we see from Fig. A.5e in Appendix A that even though 5-UCCSD is sufficiently expressive for SSVQE to converge for 6 states, increasing the circuit depth to 6-UCCSD greatly speeds up convergence. Thus, in certain situations "over-parameterizing" the ansatz may provide some benefit. This must of course be balanced with the fact that increasing the circuit depth will increase the runtime for each circuit run and make the method more susceptible to noise.

4.3.4 Inner Product Circuit Depth

We conclude this section by discussing the circuit depth involved in calculating the inner product terms in the qOMM objective function. Currently, the only method we are aware of that computes these terms involves running the circuit in Eq. 4.3. At a minimum, the

inner product circuit will incur a circuit depth twice that of the original ansatz. There will be additional circuit depth due to the need to compile controlled versions of the ansatz to the finite basis gate set of the machine. The extent of this additional depth compared to the original ansatz will depend on a number of factors including the basis gate set and qubit connectivity of the hardware, the choice of ansatz, and the efficiency of the compilation algorithm used. We can give some intuition for the circuit depths involved for the particular systems we studied here for a particular basis gate set. Using the Qiskit compiler, the gate depth for the 1-UCCSD ansatz used in our LiH simulations is approximately 1400 gates when compiled to the basis gate set consisting of R_z , \sqrt{X} , X , CNOT, and the identity. The corresponding circuit depth for computing the inner product terms is approximately 19600 gates. These counts would be scaled roughly by a factor of n for n -UCCSD. This underscores the importance of future work that addresses how to reduce these gate counts. In particular, finding (if they exist) more efficient implementations of the inner product computations will be important for the applicability of qOMM in the NISQ era. We emphasize that while there are several known subroutines for computing the overlap of two states for other applications,¹¹⁰ these generally compute real quantities of the form $|\langle\psi_1|\psi_2\rangle|$, whereas for qOMM we need the full complex quantity $\langle\psi_1|\psi_2\rangle$. The observation that SSVQE appears to require an increasingly expressive ansatz for increasingly larger systems and an increasing number of states is also a concern. Future work that addresses how to mitigate this effect will be important for the applicability of SSVQE in the NISQ era. In particular, it has been shown that certain conditions can induce cost function landscapes with gradient magnitudes that vanish exponentially (*i.e.* barren plateaus) with the number of qubits, rendering convergence infeasible. This has been shown to occur in instances where one uses a hardware-efficient ansatz in combination with a non-local cost function⁸⁹ and in the presence of noise when one uses an ansatz with circuit depth that scales linearly or faster with the number of qubits.¹¹³ Because both of these types of barren plateaus worsen with increasing circuit depth, the extent to which the two methods studied in this paper are susceptible to this effect will likely depend on how the circuit depth required for each algorithm to converge

scales with increasing problem size. Such an investigation would require the simulation of larger chemical systems and would be an interesting direction of future research. We also note that there are various strategies that could be employed to facilitate the study of larger systems by representing the active space with as few qubits as possible, effectively delaying the onset of barren plateaus. A simple, well-known example of this idea was used in this work when we froze the two core orbitals in LiH to reduce the problem size from 12 qubits to 10. Another strategy would be to optimize the basis set under a fixed qubit budget, as is done in OptOrbFCI in the context of classical computational chemistry.²⁷ Generalizing this work to quantum variational algorithms such as VQE, SSVQE, and qOMM will be the topic of future work.

4.3.5 Conclusion

In this work we proposed the qOMM algorithm and compared it to a similar method in the literature, SSVQE. We have shown that in general, qOMM is more adept at avoiding local minima and can often converge much more quickly using a less expressive ansatz. This comes at the cost of the need to run circuits to calculate the inner product which come with a considerable circuit depth cost compared to that of the ansatz. We note the difficulty that seems to be inherent to all variational methods for calculating excited states on quantum computers. SSVQE, while it has no strict circuit depth overhead requirements, often requires more expressive ansatz circuits which will lead to deeper circuits in practice. qOMM often requires less expressive ansatz circuits, but requires additional circuit depth when computing overlap terms. Variational Quantum Deflation requires either doubling the circuit depth or doubling the number of qubits to compute its overlap terms. Furthermore, we note from a comparison of Fig. 4.7 and Fig. 4.5a that for LiH, computing two states requires about an order of magnitude more function evaluations than the ground state when using a chemically-motivated initialization. This could pose a significant barrier for practical implementation in the near-term, where it has been optimistically estimated that even a single energy evaluation in VQE would require several days or weeks of runtime

for systems where demonstrating quantum advantage would be plausible.⁷³ Algorithmic advances which improve both the circuit depth and convergence rate will likely be needed for any excited states method to find practical relevance in the near-term.

5. OptOrbVQE Algorithm

This chapter contains material that we have published in the Journal of Chemical Theory and Computation on our work incorporating orbital optimization into quantum algorithms.¹¹⁹ Excerpts from our work on excited states orbital optimization¹²⁰ are also used here for the sake of the cohesiveness of this thesis.

5.1 Motivation and Background

One of the main areas of research being conducted in quantum computing today is exploring the extent to which near-term quantum computers can be useful for solving practical problems. Any algorithm developed for this purpose must fulfill three primary criteria: **1.** use as few qubits as possible, **2.** minimize circuit depth, and **3.** be robust to noise. One of the most promising problems for demonstrating quantum advantage on near term quantum hardware is the electronic structure problem.¹ The canonical approach to this problem in quantum computing has been to use the second quantization formulation, wherein we take the spatial coordinate representation of the electronic structure Hamiltonian and project it onto a finite set of basis functions. The choice of which basis to use ultimately determines how closely the obtained energy levels using this truncated Hamiltonian will match those of laboratory experimental results. Experimental results for demonstrating quantum algorithms have so far been limited to representing small molecules with minimal basis sets.^{88,121,122} Such basis sets are useful for proof-of-concept demonstrations and for benchmarking progress, but they do not represent results that would match laboratory results well enough to be useful to a chemist. The ability to move beyond these minimal basis sets will be an important step towards demonstrating quantum advantage in computational chemistry. Doing so, however, presents an obvious obstacle: Using larger basis sets increases the qubit requirements for the simulation. Furthermore, many near-term quantum algorithms developed for the electronic structure problem involve the use of ansatz circuits with depth scaling polynomially with the size of the spin orbital basis set. Thus, increasing the size of the basis set results in increased circuit depth as well.

Several methods have been proposed in recent years to make the representation of the electronic structure Hamiltonian as compact and resource-efficient on quantum computers as possible. These methods can be roughly grouped into three categories: **1.** Classical pre-processing of compact effective Hamiltonians, **2.** Orbital optimizations interleaved between successive quantum eigensolver problems, and **3.** post-processing to partially correct the basis set error. Downfolded effective Hamiltonian techniques^{123–125} use a unitary coupled-cluster ansatz operator to rotate the Hamiltonian in the full orbital space, where the coupled-cluster amplitudes are solved for classically. The transformed Hamiltonian is approximated according to a second-order Baker-Campbell-Hausdorff expansion and projected onto a chosen active space. Transcorrelated and explicitly correlated Hamiltonian methods^{126,127} are conceptually similar to downfolded methods, with the main difference being that the similarity transformed applied to the Hamiltonian has an explicit dependence on the coordinate space positions of the electrons. The purpose of this is to efficiently capture the anti-correlation effects arising from the Coulomb repulsion between electrons that would traditionally require large basis set expansions. Orbital optimization methods share some similarities to effective Hamiltonian methods in that they also apply a similarity transformation to the Hamiltonian, but differ in how the transformation parameters are found. Whereas effective Hamiltonian methods solve for the transformation parameters in a pre-processing step, orbital optimization methods^{128–130} apply a parameterized unitary transformation to the Hamiltonian, projecting the resulting parameterized Hamiltonian onto a chosen active space and minimizing an objective function. Furthermore, downfolded Hamiltonians, unlike orbital optimization techniques, are not limited to single orbital rotations, necessitating the use of BCH expansions. The use of post-processing to partially correct the error arising from the truncated basis set has also been proposed. Virtual Quantum Subspace Expansion (VQSE)¹³¹ is a method where the ground state problem is first solved within a chosen active space using an algorithm such as VQE. An improved estimate for the ground state is then obtained by classically solving a generalized eigenvalue problem over a contracted subspace spanned by single and double fermionic excitation operators

acting on the solution to the previous active space problem. These excitation operators are allowed to include excitations to the virtual space and thus contribute to a correction to the energy from the limited active space solution.

In this work we generalize the OptOrbFCI²⁷ algorithm (developed in the context of classical computing for settings in which classical computational resources are limited) to the quantum computing setting in which qubit counts and coherent circuit depth are limited resources. OptOrbFCI is an orbital optimization method that applies a partial unitary transformation to the set of basis functions, collapsing it to one of a smaller size and introducing the elements of the matrix representation of this transformation as additional parameters to be optimized in the overall ground state search problem. An FCI solver is used to find the ground state energy in a reduced basis. Extending OptOrbFCI to the quantum computing setting corresponds to replacing the FCI solver subroutine with one of several quantum eigensolvers such as the Variational Quantum Eigensolver (VQE),⁹ Quantum Imaginary Time Evolution (QITE),^{132–135} or Quantum Monte Carlo.¹³⁶ In this work, we pair the orbital optimization subroutine with VQE, calling the resulting overall method OptOrbVQE. We find that OptOrbVQE consistently achieves lower ground state energy compared to standard VQE methods when using the same number of qubits. Higher accuracy results are also achieved while simultaneously using fewer qubits than these methods in several instances.

5.2 Optimal Orbital VQE

Let us now introduce the orbital optimization in the VQE setting, motivated by a similar scheme in the classical setting as the OptOrbFCI algorithm proposed by Li et. al.²⁷ If our set of basis functions has size M , then this will require the use of M qubits if no techniques to reduce this count are employed. Suppose we have access to a quantum computer with only $N < M$ qubits or that we are using an ansatz circuit that scales with the number of qubits in such a way that we are limited to calculations using N qubits. We thus have to restrict to a Hamiltonian with only N spin orbitals by applying a partial

unitary transformation for the basis change, which we represent using a $M \times N$ real partial unitary matrix \hat{V} . The basis functions will transform according to

$$\tilde{\psi}_i = \sum_j^M \hat{V}_{ji} \psi_j \quad (5.1)$$

This corresponds to the one and two body integrals transforming according to Eq. 5.2 and Eq. 5.3.

$$\tilde{h}_{p'q'} = \sum_{p,q=1}^M h_{pq} \hat{V}_{pp'} \hat{V}_{qq'} \quad (5.2)$$

$$\tilde{v}_{p'q'r's'} = \sum_{p,q,r,s=1}^M v_{pqrs} \hat{V}_{pp'} \hat{V}_{qq'} \hat{V}_{ss'} \hat{V}_{rr'} \quad (5.3)$$

The ground state energy is now a function of not only the ansatz parameters $\boldsymbol{\theta}$, but the partial unitary matrix \hat{V} as well. The ground state search problem is now a minimization problem over both the space of ansatz parameters and the space of all real partial unitary matrices of dimension $M \times N$:

$$\min_{\substack{\boldsymbol{\theta} \\ \hat{V} \in \mathcal{U}(M,N)}} \langle \psi_{\text{ref}} | \hat{U}^\dagger(\boldsymbol{\theta}) \tilde{H}(\hat{V}) \hat{U}(\boldsymbol{\theta}) | \psi_{\text{ref}} \rangle \quad (5.4)$$

where

$$\mathcal{U}(M, N) = \{ \hat{V} \in \mathbb{R}^{M \times N} | \hat{V}^T \hat{V} = I_N \} \quad (5.5)$$

The transformed Hamiltonian as a function of \hat{V} is given by:

$$\begin{aligned} \tilde{H}(\hat{V}) &= \sum_{p',q'=1}^N \sum_{p,q=1}^M h_{pq} \hat{V}_{pp'} \hat{V}_{qq'} \tilde{a}_{p'}^\dagger \tilde{a}_{q'} \\ &+ \frac{1}{2} \sum_{p',q',r',s'=1}^N \sum_{p,q,r,s=1}^M v_{pqrs} \hat{V}_{pp'} \hat{V}_{qq'} \hat{V}_{ss'} \hat{V}_{rr'} \tilde{a}_{p'}^\dagger \tilde{a}_{q'}^\dagger \tilde{a}_{s'} \tilde{a}_{r'} \end{aligned} \quad (5.6)$$

where the primed and unprimed indices index the transformed and original basis wavefunctions, respectively. (*e.g.* \hat{a}_p^\dagger is the fermionic creation operator corresponding to spin-orbital

ψ_p and $\tilde{a}_{p'}^\dagger$ is the fermionic creation operator corresponding to the transformed spin-orbital $\tilde{\psi}_{p'}$.) This fermionic Hamiltonian can then be mapped to a weighted sum of Pauli string operators acting on qubits. We leave the Hamiltonian expressed in terms of fermionic operators to emphasize that the method is independent of the mapping chosen. The expectation values $\langle \psi_{\text{ref}} | \hat{U}^\dagger(\boldsymbol{\theta}) \tilde{a}_{p'}^\dagger \tilde{a}_{q'} \hat{U}(\boldsymbol{\theta}) | \psi_{\text{ref}} \rangle$ and $\langle \psi_{\text{ref}} | \hat{U}^\dagger(\boldsymbol{\theta}) \tilde{a}_{p'}^\dagger \tilde{a}_{q'}^\dagger \tilde{a}_{s'} \tilde{a}_{r'} \hat{U}(\boldsymbol{\theta}) | \psi_{\text{ref}} \rangle$ are the 1-RDM and 2-RDM elements ${}^1D_{q'}^{p'}$ and ${}^2D_{r',s'}^{p',q'}$, respectively. These quantities are (after being mapped to qubit operators) measured on a quantum computer with respect to the ansatz state $|\psi(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})|\psi_{\text{ref}}\rangle$ in the same fashion as conventional VQE.

It is important to note that the optimization problem in Eq. 5.4 consists of two distinctly different types of parameters subject to different types of constraints: the partial unitary \hat{V} and the vector $\boldsymbol{\theta}$ (which typically consists of real numbers subject to some bounds). Thus, it is natural to treat the two sets of variables separately. In this work we adopt the procedure originally proposed by OptOrbFCI in the classical setting. The minimization problem in Eq. 5.4 is divided into two subproblems: minimizing the energy with respect to \hat{V} (keeping $\boldsymbol{\theta}$ fixed) and minimizing the energy with respect to $\boldsymbol{\theta}$ (keeping \hat{V} fixed). We alternate between these two subproblems until some stopping criterion is reached. Because this algorithm involves two minimization subproblems (each with their own iteration number counter) that are both repeated multiple times (where this number of times is associated with an additional “outer loop” iteration number counter), we specify which indices are used to denote which type of iteration counter throughout this paper in order to reduce any ambiguity:

- l will be used to denote the iteration number within the minimization with respect to \hat{V} ;
- m will be used to denote the iteration number within the minimization with respect to $\boldsymbol{\theta}$ (the same as what is typically referred to as the iteration number within the context of VQE without orbital optimization);
- n will be used to denote the outer loop iteration number (*i.e.* the number of times

the minimization subproblem with respect to \hat{V} has been conducted so far).

The superscript *opt* will be used to denote the optimal point for each of the minimization subproblems within a given outer loop iteration. The OptOrbVQE algorithm can be summarized as follows:

1. Set the outer loop iteration number $n = 0$ and choose an initial partial unitary transformation $\hat{V}_{n=0,l=0}$ and initial VQE parameters $\boldsymbol{\theta}_{n=0,m=0}$. Choose an outer loop stopping tolerance ϵ_{outer} .
2. On a classical computer, calculate the transformed Hamiltonian $\tilde{H}(\hat{V}_n)$ and use one of several known mappings to generate the corresponding transformed qubit Hamiltonian.
3. Initialize the ansatz state as $\hat{U}(\boldsymbol{\theta}_{n,m=0})|\psi_{\text{ref}}\rangle$ and perform VQE on a quantum computer to obtain $\boldsymbol{\theta}_n^{opt}$ and the estimated ground state energy $E(\hat{V}_{n=0,l=0}, \boldsymbol{\theta}_n^{opt})$.
4. If $|E(\hat{V}_{n-1,l=0}, \boldsymbol{\theta}_{n-1}^{opt}) - E(\hat{V}_{n,l=0}, \boldsymbol{\theta}_n^{opt})| < \epsilon_{outer}$, halt the algorithm and return $E(\hat{V}_{n=0,l=0}, \hat{U}(\boldsymbol{\theta}_n^{opt})|\psi_{\text{ref}}\rangle)$, and $\hat{V}_{n,l=0}$ as the optimal quantities of interest. Else, continue to next step.
5. On a quantum computer, measure the 1-RDM and 2-RDM elements with respect to the state $\hat{U}(\boldsymbol{\theta}_n^{opt})|\psi_{\text{ref}}\rangle$.
6. Initialize the partial unitary as $\hat{V}_{n,l=0}$ and perform the minimization subproblem in Eq. 5.4 with respect to \hat{V} (using the 1- and 2-RDM tensors from the previous step) to obtain \hat{V}_n^{opt} .
7. Set $\hat{V}_{n+1,l=0} = \hat{V}_n^{opt}$ and $\boldsymbol{\theta}_{n+1,m=0} = \boldsymbol{\theta}_n^{opt}$. Optionally, a small random perturbation can be added to $\hat{V}_{n+1,l=0}$ to avoid shallow local minima.
8. Set $n = n + 1$ and repeat steps 2-8.

There are a few clear initializations $\hat{V}_{n=0,l=0}$ and $\hat{V}_{n,l=0}$ that can be used in this algorithm. Throughout this work, we choose $\hat{V}_{n=0,l=0}$ to be the permutation matrix that selects N spin orbitals from the starting basis with the lowest Hartree-Fock energy ordered by ascending energy. This is equivalent to starting with a large basis, but restricting the

active space to these N spin orbitals. This is not the only initialization that could be used, but it is an intuitive one. In general, we can take any $M \times N$ real matrix A and project it onto one which is a partial unitary through the orthonormalization function:

$$\text{orth}(A) = AQA^{-\frac{1}{2}}Q^\dagger \quad (5.7)$$

where Q and Λ together are a solution of the diagonalization equation $A^\dagger A = Q\Lambda Q^\dagger$. We could, for instance, orthonormalize a matrix whose elements are sampled from a random distribution of our choice. The normal distribution or the uniform distribution over some interval would be natural choices. If \hat{P} is the permutation matrix used in this work, then one alternative choice for $\hat{V}_{n=0,l=0}$ would be $\text{orth}(\hat{P} + \text{Rand}(M, N))$, where $\text{Rand}(M, N)$ is a random $M \times N$ matrix. Throughout this paper, the partial unitary $\hat{V}_{n+1,l=0}$ in step 7 of the algorithm is chosen to be $\text{orth}(\hat{V}_n^{\text{opt}} + \text{Rand}(M, N))$, with the elements of $\text{Rand}(M, N)$ in this instance being sampled from the normal distribution centered about mean 0 with a standard deviation 0.01. The random perturbation matrix is added to help the method avoid getting trapped in shallow local minima.

We end this section by noting the differences between this proposed method and specific examples of methods in categories mentioned in the introduction. **1.** In contrast to explicitly correlated and downfolded Hamiltonian parameters, where the similarity transformation parameters are found as a pre-processing step according to a pre-defined set of equations or chemical intuition, OptOrbVQE (like other orbital optimization methods) finds the optimal parameters by minimizing an objective function. **2.** several of the techniques referenced in the introduction such as the DUCC¹²³ and CT-F12 Hamiltonians,¹²⁶ and QDSRG¹²⁴ use a similarity transformation which takes the form of a chemically-motivated ansatz. They further approximate the transformed Hamiltonian according to a second-order BCH expansion in a pre-processing step. In OptOrbVQE, the basis set representation of the Hamiltonian is iteratively updated throughout the optimization procedure through single-particle transformations of the one and two-body integrals. In the other orbital optimization techniques mentioned in the introduction, such as OO-UCC,¹²⁸ SA-

OO-VQE,¹³⁰ and quantum CASSCF,¹²⁹ a CASSCF-like approach is taken, representing the orbital rotation as an exponential of an $M \times M$ anti-hermitian matrix, approximating this exponential operator by a second-order Taylor expansion, performing successive Newton steps using this approximation until convergence of the energy is achieved, then choosing an active space of $N < M$ orbitals in which to solve for the eigenvalues and eigenstates. Instead of parameterizing the orbital rotation operator as an $M \times M$ exponential operator and carrying out successive Newton steps, we parameterize it directly as an $M \times N$ partial unitary matrix. This allows us to take advantage of modern optimization techniques that have been developed in recent years which have orthogonality constraints built in.^{29–33} Such optimization methods render the conventional exponential parameterization as one choice of parameterization rather than a strict requirement. We note that aside from access to a wider range of optimization techniques, this approach comes with other potential advantages. The first is that the partial unitary transformation has NM constrained parameters, rather than the $(M - 1)(M - 2)/2$ independent parameters in the anti-hermitian matrix that appears in CASSCF. This results in a net reduction of the size of the parameter space involved for many problems, particularly when $M \gg N$. This may ease the convergence procedure. Additionally, the partial unitary nature of the parameterization implies that unlike CASSCF-like methods, we do not need to resort to the use of heuristics or chemical intuition when choosing the post-optimization active space. The dimensionality of the partial unitary matrix handles this automatically during the optimization procedure. We also note that there is a sense in which the dimensionality of this partial unitary operator is a degree of freedom which we can control. For example, we still retain the option to optimize a full $M \times M$ unitary matrix using orthogonally-constrained optimizers and use heuristics to choose the active space of N orbitals. Alternatively, we could optimize over the set of $M \times m$ partial unitaries for $N < m < M$ and use heuristics to choose an active space of m orbitals from the space of N orbitals. This flexibility can be seen as an advantage in and of itself. Furthermore, the use of an orthogonally-constrained projected gradient descent method³³ in conjunction with this orbital optimization procedure has been

numerically demonstrated to be more adept at avoiding local minima and achieving better accuracy than CASSCF in a previous work by Li et. al in a classical computing context.²⁷ This offers a clear motivation for the continued study of this method in a broader range of contexts such as incorporation into various quantum eigensolvers.

5.3 Numerical Results

Our implementation of the OptOrbVQE algorithm is a combination of in-house code and code from the open source packages Qiskit¹¹⁷ (Qiskit Nature 0.3.2, Qiskit Aer 0.10.4, and Qiskit Terra 0.20.0) and PyTorch¹³⁷ 1.11.0. The method of finding the optimal \hat{V} with fixed θ is the same as that used in the OptOrbFCI proposal paper: a projection method with alternating Barzilai-Borwein stepsize.³³ The code for this optimizer was developed in-house using several tensor functionalities of PyTorch. We choose to use PyTorch for several reasons: **1.** We find that it has an efficient *einsum* implementation which greatly speeds up the computation of Eq. 5.4. **2.** It has support for automatic differentiation, which enables efficient computation of the gradient of Eq. 5.4 with respect to \hat{V} in the projection method. **3.** It offers support for GPU acceleration, which can speed up the calculation significantly, especially for larger starting basis sets. The subproblem of minimizing the energy with respect to θ uses Qiskit’s VQE implementation.

5.3.1 Minimal Qubit Usage

In this section we investigate the ground state accuracy achievable by OptOrbVQE when using the same number of spin orbitals as the minimal STO-3G basis. We then compare the results to VQE and FCI simulations using basis sets of the same size or larger. Ideally, we would only compare OptOrbVQE to VQE as this is a more appropriate comparison than classical FCI methods. However, we find that simulating VQE in Qiskit is much more computationally expensive than carrying out an FCI problem of the same size using PySCF. Thus, FCI results are a convenient stand-in for VQE results that would be computationally infeasible. The assumption here is that the FCI ground state energy serves as a lower bound for what is achievable by VQE. In the best-case scenario where a sufficiently powerful ansatz

is used and VQE achieves convergence to the global minimum, these values would closely match.

The classical optimizer used in VQE subproblem instances in this section is L-BFGS-B.¹⁰⁰ We use Qiskit’s *AerSimulator* in combination Qiskit’s *AerPauliExpectation* algorithm to compute expectation values of both the molecular Hamiltonian and the observables involved in computing the 1 and 2-RDM. This combination yields ideal, noiseless results. Thus, these simulations serve to test the ability of the OptOrbVQE algorithm to converge under ideal conditions, but not its robustness to noise. We defer a study of the robustness to noise of the method to Sec. 5.3.3. The stopping tolerances for both the orbital rotation subproblem and the OptOrbVQE algorithm as a whole are set to 10^{-5} . The maximum outer loop iteration number is set to 19 so that the VQE subproblem is run at most 20 times.

5.3.1.1 H₄

We begin by presenting classically-simulated results for H₄, a toy model which consists of 4 hydrogen atoms arranged in a square with an H-H distance of 1.23 Å. The ansatz used is 2-UCCSD.⁵ In Qiskit, one has the ability to repeat a base ansatz circuit n times to produce a more expressive ansatz. When we refer to n -UCCSD, we mean an ansatz which consists of the base UCCSD ansatz repeated n times in this fashion. Using n -UCCSD has the effect of increasing both the circuit depth and the number of independent parameters by a factor of n over UCCSD. We find that two repetitions are necessary for VQE in the STO-3G basis to converge to within the chemical accuracy of the FCI value (calculated using PySCF¹¹⁸ 2.0.1) in the same basis for the H₄ toy model.

We set the number of spin orbitals to be 8 for H₄, the number of spin orbitals for this system in the minimal STO-3G basis set. Fig. 5.1 illustrates the convergence of the OptOrbVQE algorithm as a function of the outer loop iteration number for various starting basis sets. We compare the results to that obtained from VQE in the 6-31G basis (16 spin orbitals) using 2-UCCSD as the ansatz. Despite the fact that OptOrbVQE is using half

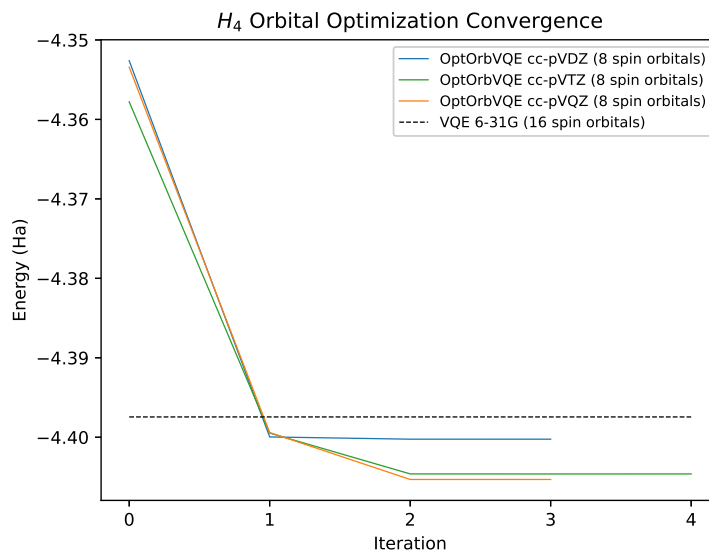


FIGURE 5.1: Convergence of OptOrbVQE as a function of the outer loop iteration number for H_4 at the near-equilibrium H-H distance of 1.23 Å.

the number of qubits as VQE, we find that it achieves a lower ground state energy for all the starting basis sets used. This lower energy is achieved after just the $n = 1$ outer loop iteration, which corresponds to carrying out the orbital rotation subroutine once and the VQE subroutine twice. The energy is lowered further when cc-pVTZ and cc-pVQZ are used as starting basis sets with further iterations. We note that the ground state energy achieved in the optimized basis set is approximately -4.404 Hartree, whereas the FCI energy for this system in the cc-pVDZ basis (40 spin-orbitals) is approximately -4.430 Hartree and the FCI energy in the cc-pVTZ basis (112 spin-orbitals) is approximately -4.443 Hartree, indicating clearly that larger optimized basis sets would be needed to approach the infinite basis set limit to chemical accuracy. Nonetheless, the fact that orbital optimization can exceed the accuracy of larger basis sets while simultaneously using half the number of qubits is encouraging.

5.3.1.2 LiH

For LiH, we use 1-UCCSD as the ansatz. We set the number of spin orbitals for OptOrbVQE to be 12, the number of spin orbitals for this system in the minimal STO-3G

basis set. We compute the ground state energy at the near-equilibrium Li-H distance of 1.595 Å as well as the binding curve of LiH.

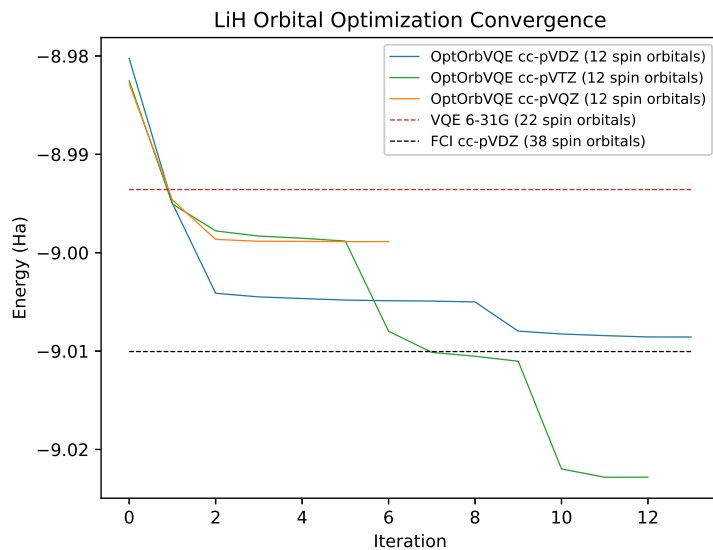


FIGURE 5.2: Convergence of OptOrbVQE as a function of the outer loop iteration number for LiH at the near-equilibrium bond distance of 1.595 Å.

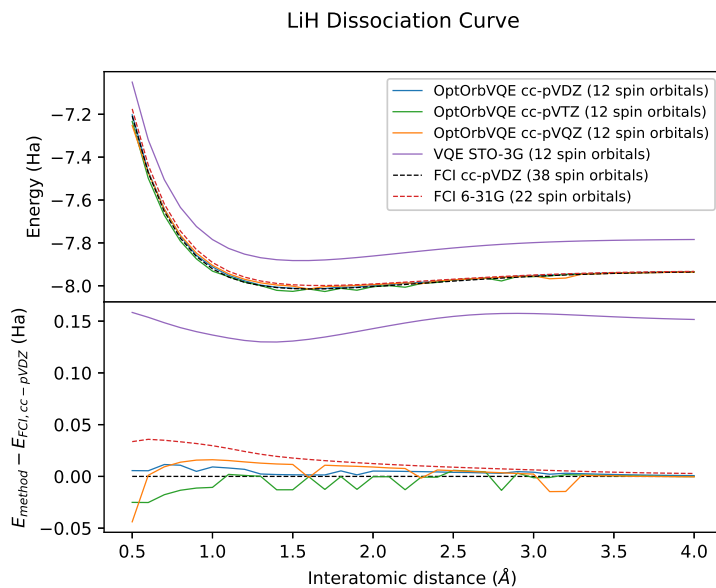


FIGURE 5.3: Top: Dissociation curve of LiH. Bottom: Difference of energy relative to FCI (cc-pVDZ).

Fig. 5.2 illustrates the convergence of OptOrbVQE as a function of the outer loop iteration number. We find that OptOrbVQE achieves a lower energy than VQE in the 6-31G basis (22 spin-orbitals) after the $n = 1$ iteration. The energy is further improved with additional iterations. In particular, OptOrbVQE using cc-pVTZ as the starting basis surpasses the FCI energy in the cc-pVDZ basis at the $n = 7$ iteration. OptOrbVQE starting from the cc-pVDZ basis also approaches, but does not surpass this value. This is to be expected as the FCI value in the cc-pVDZ basis puts a lower limit on what is achievable by OptOrbVQE starting from the cc-pVDZ basis for any active space size. We also note that starting from a larger basis does not always result in a more accurate value, as can be seen from OptOrbVQE (cc-pVQZ starting basis) not achieving the same accuracy as the other two starting basis sets. Given that the cc-pVQZ basis is comprised of far more orbitals than either cc-pVDZ or cc-pVTZ, this could indicate that OptOrbVQE starting from the cc-pVQZ basis converged to a local minimum in this instance.

Fig. 5.3 illustrates the results obtained for the binding curve of LiH. We can see that OptOrbVQE easily outperforms VQE using the same number of qubits. OptOrbVQE consistently achieves an energy lower than the FCI energy in the 6-31G basis. OptOrbVQE also often achieves an energy lower than the FCI energy in the cc-pVDZ basis, although this is not guaranteed and sometimes fails to do so. We also note that the jagged nature of the orbital optimized curve may indicate that the method is converging to local minima for some of the geometries considered here.

5.3.1.3 BeH₂

In this section we test OptOrbVQE on BeH₂, a linear molecule with a near-equilibrium Be-H bond distance of 1.3264 Å. 1-UCCSD is the ansatz used. The number of spin orbitals used by OptOrbVQE is set to 14, the number of spin orbitals for this system in the minimal STO-3G basis.

Fig. 5.4 illustrates the convergence of OptOrbVQE at the equilibrium configuration. We find that starting from either the cc-pVTZ or cc-pVQZ basis set results in OptOrbVQE

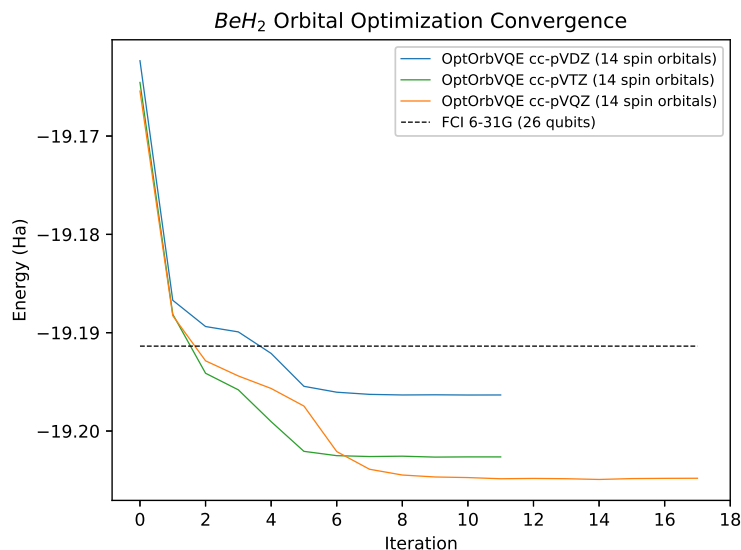


FIGURE 5.4: Convergence of OptOrbVQE as a function of the outer loop iteration number for BeH_2 at the near-equilibrium Be-H bond distance of 1.3264 Å.

surpassing the FCI energy in the 6-31G basis at the $n = 2$ iteration. Further iterations result in improved energy. Starting from the cc-pVDZ also surpasses the FCI (6-31G basis), but requires more iterations to do so. We note that the orbital optimized energy obtained here is roughly -19.204 Hartree, whereas the FCI energy in the cc-pVDZ basis (48 spin-orbitals) is roughly -19.227 Hartree, indicating that larger numbers of optimized orbitals would be needed to approach the infinite basis limit to chemical accuracy.

5.3.1.4 H_2O

In this section we test OptOrbVQE on the H_2O molecule. The ansatz used is 1-UCCSD. The number of spin orbitals used by OptOrbVQE is set to 14, the number of spin orbitals for this molecule in the minimal STO-3G basis. Fig. 5.5 plots the difference of the OptOrbVQE energy from the FCI energy in the 6-31G basis (26 spin orbitals) for H_2O at the near-equilibrium configuration of O-H distance 0.9578 Å and H-O-H bond angle of 104.4778 degrees. These results are different from the other systems presented in that while the method still easily outperforms VQE using the same number of spin orbitals, we do not observe OptOrbVQE using a minimal number of spin orbitals to surpass the FCI energy

in the larger 6-31G basis. OptOrbVQE can however be observed to approach the FCI (6-31G basis) energy at the milli-hartree level, with the energy difference converging to approximately 2.5×10^{-3} Hartree when using cc-pVQZ as the starting basis. One notable feature about this convergence curve is that the rate of convergence is most rapid up until the $n = 3$ iteration, then hits a plateau. The energy then fluctuates until the maximum number of iterations is reached, indicating the possible presence of multiple local minima which differ in energy at the milli-Hartree level. A similar trend is observed when starting from the cc-pVTZ basis, although the converged energy accuracy is worse and the fluctuations are less pronounced in this case. It is also worth noting that the 0th iteration of OptOrbVQE outperforms VQE in Fig. 5.5. Because the initial partial unitary for OptOrbVQE is set to be the matrix which selects the N lowest energy spin orbitals, the 0th iteration corresponds to starting with a large basis, but reducing the active space to one the same size as the STO-3G basis. Thus, using orbital optimization would not necessary to outperform VQE in the STO-3G basis in this instance. The main benefit of orbital optimization is further accuracy improvements at the milli-Hartree level.

5.3.2 Increasing Qubit Resources

One important feature of OptOrbVQE is that the number of spin orbitals used is a tunable parameter that can be set to any positive integer up to the number used by the starting basis set. The previous sections examined the performance of OptOrbVQE for various systems when using a number of spin orbitals equal to the minimal STO-3G basis. This choice of active space size was largely due to **1.** the observation that these small active space sizes are what are most commonly used in current experimental demonstrations of VQE on quantum hardware and **2.** the formidable computational runtime for running VQE using Qiskit compared to running FCI using PySCF. We showed that even such small active spaces were sufficient for optimized basis sets to outperform FCI using larger basis sets, however they are not sufficiently large to approach the infinite basis set limit to chemical accuracy. In order to achieve this, we would have to run OptOrbVQE with active space

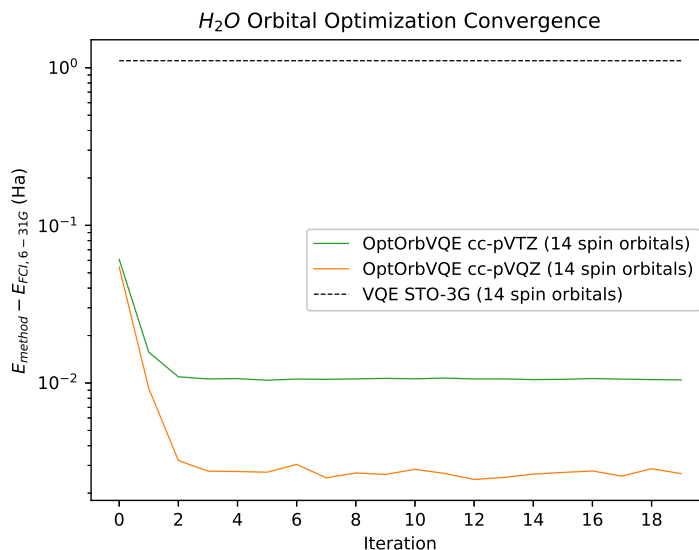


FIGURE 5.5: Convergence of OptOrbVQE as a function of the outer loop iteration number for H₂O at the near-equilibrium O-H bond distance of 0.9578 Å and bond angle 104.4776 degrees.

sizes larger than that of the minimal STO-3G basis. In this section we increase the number of spin orbitals used by OptOrbVQE in order to examine the potential for the method to further improve energy accuracies as the capabilities of quantum computers improve with time. We test OptOrbVQE on H₂ using even integer numbers of spin orbitals from 4 to 16. This system is simple enough that we can test the method for active spaces larger than the minimal basis. Qiskit’s *AerSimulator* and *AerPauliExpectation* are used to obtain ideal noiseless results as in Sec. 5.3.1. The optimizer used is L-BFGS-B and the ansatz used is 1-UCCSD. Fig. 5.6 plots the difference of the OptOrbVQE energy at the near-equilibrium bond distance of 0.735 Å using OptOrbVQE and the FCI energy in the cc-pVTZ basis (56 spin orbitals). The FCI energy in the cc-pVDZ basis (20 spin orbitals) is also included for reference. The most significant (but expected) feature of this plot is that the energy accuracy obtained by OptOrbVQE can be improved by increasing the number of spin orbitals that it uses. This comes with the caveat that using more qubits (starting from a particular basis) does not always result in a lower converged energy. Several plateaus can be

seen over the interval considered. For example, increasing the number of spin orbitals from 6 to 8 does not result in significantly improved energy when starting from either the cc-pVTZ or cc-pV5Z basis sets. Increasing the number of qubits from 10 to 16 also does not appear to result in improved energies when starting from the cc-pVQZ basis. Another notable feature of this plot is that for a given number of qubits, starting from a larger basis set does not always result in lower energy. This can be seen from OptOrbVQE starting from the cc-pVQZ basis (120 spin orbitals) achieves a lower energy than starting from the cc-pV5Z basis (220 spin orbitals) for 8 and 10 qubits. Finally, we note that in Fig. 5.6, the green curve compares the logarithmic difference between the energy obtained by OptOrbVQE starting from the cc-pVTZ basis and the FCI energy in the full 56 spin-orbital cc-pVTZ basis. The highest degree of accuracy obtained at 16 spin-orbitals is approximately 3 milliHartree. There are several factors contributing to this discrepancy: **1.** OptOrbVQE consists of two optimization subproblems, neither of which is guaranteed to converge to the global minimum. Each may converge to a spurious local minima or within a neighborhood of the global minimum. **2.** The VQE subproblem utilizes a wavefunction ansatz. This comes with an associated ansatz representation error that is not present in the classical FCI algorithm. **3.** It is well-known that large basis set expansions improve the ability of computational methods to capture energy contributions that arise from electron correlation effects, in particular correlation arising from Coulomb repulsion between electrons which is difficult to capture with small basis sets. Although orbital optimization helps the method capture some of this energy contribution, its smaller basis size precludes it from capturing all of it.

The first two points listed here may also help to explain some un-intuitive behavior exhibited by some of the tests in this paper. For example, in Fig. 5.2, the largest starting basis used by OptOrbVQE, cc-pVQZ, is the one which achieved the least accurate energy among the three starting basis sets considered. This is counter to what one would intuitively expect, where the more flexible variational space should give it the potential to achieve the highest quality accuracy. Furthermore, several plateaus are observed for all three starting

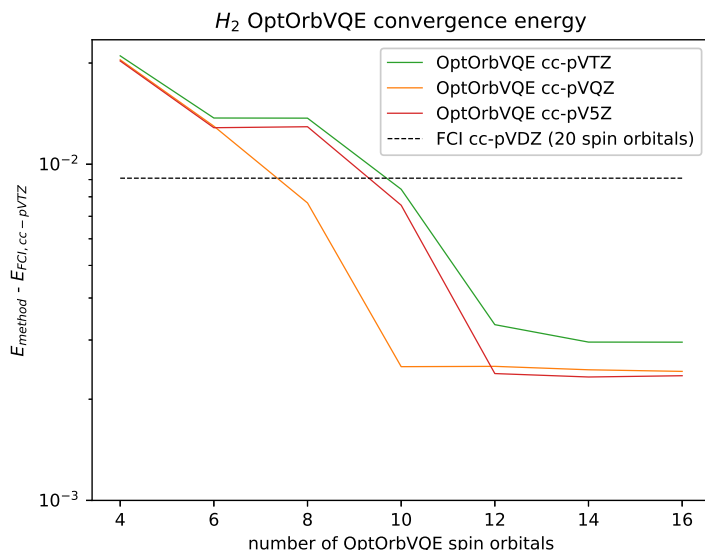


FIGURE 5.6: Converged energy of OptOrbVQE as a function of the number of spin orbitals for H_2 at the near-equilibrium bond distance of 0.735 \AA .

basis sets. Similarly, in Fig. 5.6 there are several instances where increasing the size of the variational space through an increase in the number of qubits does not strictly result in an increase in accuracy, but rather appears to occasionally result in a plateau. There are a few possible explanations for this behavior. We note that in order for the benefits of an increased variational space to be apparent in the final accuracy obtained, it is necessary for both the orbital optimization and VQE subproblems to converge sufficiently close to their global minima and for the VQE ansatz to have sufficient representation accuracy in the rotated basis sets determined by the orbital optimization subroutine at each iteration. If any of these conditions are not met, the final energy accuracy may not reach its full potential. We defer a more in-depth study on how to improve the convergence of OptOrbVQE to future work. For example, one could investigate incorporating adaptive ansatz strategies^{94,138} into the VQE subproblem. The intuition behind this approach is that an adaptive ansatz may be better suited for representing the ground state of a system than a fixed ansatz when the basis set representation itself is iteratively changing. A second possibility would be to add a random perturbation to the initial parameters of each VQE iteration. In these tests, a

random perturbation is added to the initial partial unitary to help the orbital optimization escape from shallow local minima, but the VQE subproblem may also benefit from a similar initialization.

5.3.3 Robustness to Noise

We now investigate the robustness of the OptOrbVQE algorithm to noise, which we carry out in two stages using the binding curve of the H_2 molecule as a test system. In Sec. 5.3.3.1 we incorporate statistical sampling as the only source of the noise. On quantum hardware, this type of noise arises from the repeated circuit preparation and observable measurement process. For example, to measure the quantity $\langle \psi_{\text{ref}} | \hat{U}^\dagger(\boldsymbol{\theta}) \hat{H} \hat{U}(\boldsymbol{\theta}) | \psi_{\text{ref}} \rangle$ we would prepare the circuit $\hat{U}(\boldsymbol{\theta})$ n times, measuring each of the Pauli terms \hat{P}_i in Eq. (1.64) n times and classically compute the weighted sum of their expectation values. Because this form of noise is independent of the ansatz circuit depth, starting with this form of noise allows us to compare OptOrbVQE using a smaller basis to VQE using a larger basis while keeping the effects that would arise from the difference in circuit depth between these two problem instances separate. In Sec. 5.3.3.2 we add a local depolarizing noise model to the statistical noise.

5.3.3.1 Statistical Sampling Noise

For the noisy simulations, we choose COBYLA¹⁰¹ as the classical optimizer, informed by our previous experience with these two optimizers that COBYLA is more resilient to noise than L-BFGS-B. The ansatz used is 1-UCCSD. The mapping used is Jordan-Wigner. 10^6 circuit samples are used for observable measurements. OptOrbVQE is set to use cc-pVQZ as the starting basis and uses an active space of 4 spin orbitals in the transformed basis. We compare it to VQE in the 6-31G basis (8 spin orbitals), using the FCI (6-31G basis) as a baseline. Fig. 5.7 illustrates the results obtained for these tests. The outer loop stopping tolerance is set to 10^{-3} . The error bars are calculated internally by Qiskit, which records the statistical variance σ associated with expectation values from n circuit samples and returns the error as $\sqrt{\frac{\sigma}{n}}$.

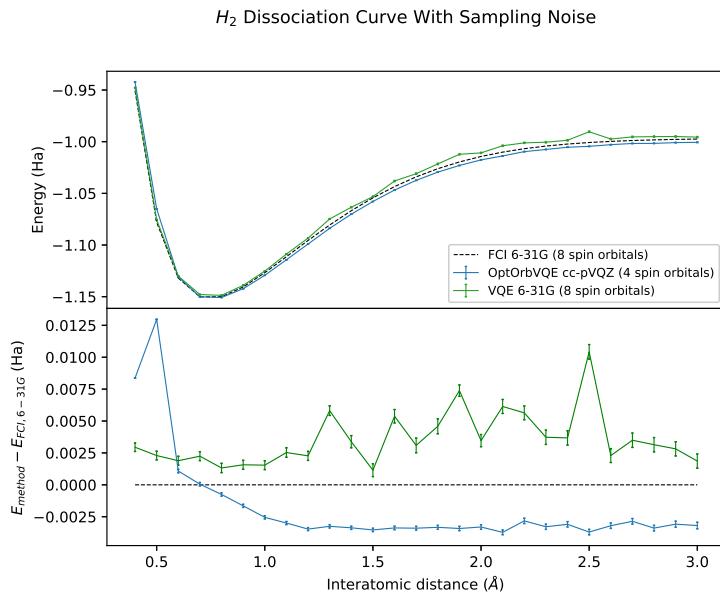


FIGURE 5.7: *Top*: binding curve of H_2 using 10^6 circuit samples. *Bottom*: difference in energy from the FCI (6-31G basis) energy.

We can see that in the presence of statistical sampling noise, OptOrbVQE retains its ability to achieve a lower ground state energy than VQE (6-31G, 16 spin orbitals) while only using half the number of qubits for interatomic distances 0.6 \AA and greater.

5.3.3.2 Depolarizing Noise

In order to model the effects of gate noise, we add a local depolarizing channel to each one-qubit gate and a tensor product of two local depolarizing channels to each two-qubit gate. This has the effect that every time a one-qubit gate is applied, one of the three Pauli operators (with equal likelihood) is also applied with probability p_{error} . For two-qubit gates, this probabilistic error event occurs independently for each qubit involved. In this work we set $p_{error} = 10^{-3}$. No error mitigation techniques are used. Aside from adding gate noise, the methodology remains the same as in Sec. 5.3.3.1, except that the ansatz is changed from 1-UCCSD to a hardware-efficient ansatz shown in Fig. 5.8.

In Qiskit, this corresponds to the *RealAmplitudes* circuit with the number of repetitions set to 2. The first layer of this circuit prepares the qubits in the Hartree Fock state. The

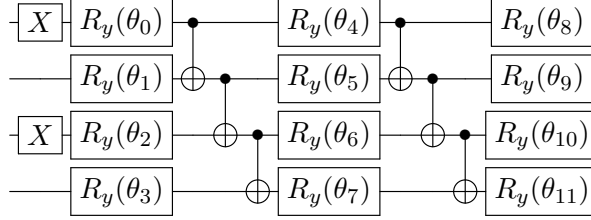


FIGURE 5.8: Ansatz used for H_2 simulations with depolarizing noise.

parameters are initialized to zero. We compare OptOrbVQE to VQE (STO-3G basis), using the FCI (6-31G basis) energy as a baseline. The results of these tests are shown in Fig. 5.9.

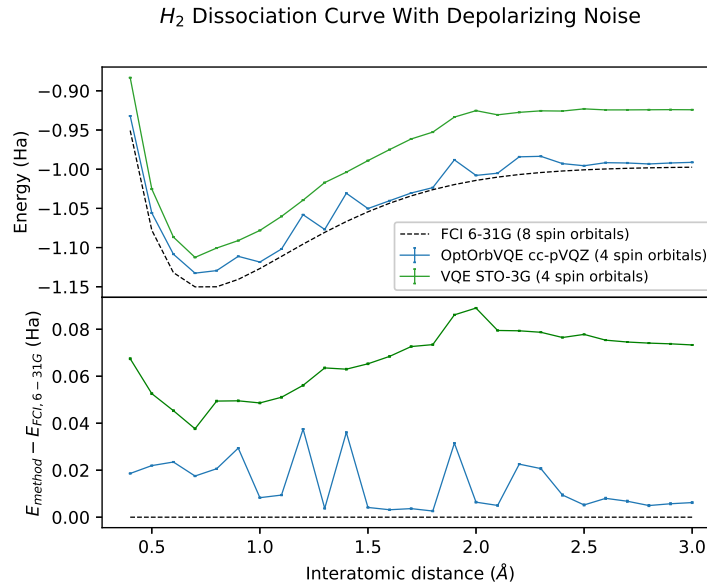


FIGURE 5.9: *Top*: binding curve of H_2 using 10^6 circuit samples with $p_{error} = 10^{-3}$. *Bottom*: difference in energy from the FCI (6-31G basis) energy.

We find that OptOrbVQE consistently achieves lower energy than VQE when using the same number of qubits. Unlike in Sec. 5.3.3.1 when only statistical sampling noise was used, OptOrbVQE no longer achieves energy lower than FCI in the 6-31G basis. It does, however, approach this reference energy at the milli-Hartree level for several interatomic distances. We note that the jagged nature of the orbital optimized binding curve possibly indicates a tendency to converge to local minima, likely exacerbated by the fact that the 1 and 2-RDM quantities used for the orbital optimization step are subject to noise in this instance. By

comparison, VQE in a fixed minimal basis has a much smoother binding curve, although it is still uniformly less accurate than the orbital optimized case. This demonstrates that even though some noise is injected into the start of the orbital optimization procedure through the 1 and 2-RDM measurements, the method is robust enough that the advantage of using an optimized basis set is not spoiled.

5.4 Discussion

One of the main challenges that exists today in quantum computing is demonstrating quantum advantage on a problem with practical utility. One such problem is calculating the ground state of electronic chemical systems to high accuracy when compared to laboratory results. In this work we have demonstrated that OptOrbVQE offers a clear path towards this goal in two ways: **1.** When using a number of qubits equal to that in a minimal basis, OptOrbVQE consistently achieves higher accuracy than VQE using a minimal basis set. In many cases it can even outperform VQE methods of larger basis sets using a fraction of the number of qubits. **2.** The number of qubits used by OptOrbVQE is a tunable parameter. Increasing the number of qubits typically has the effect of improving the energy accuracy, which provides a convenient method for systematically demonstrating improved results as the capabilities of quantum computers progress. This improved performance comes at the cost of running the orbital optimization and VQE subproblems multiple times. While we find that our classical simulations can in some instances utilize 10 or more iterations before the stopping condition is reached, the bulk of the convergence typically occurs during the first 2-5 iterations. These first few iterations are typically sufficient for the method to surpass VQE and FCI methods of larger basis sets. A user of this algorithm could simply choose to limit the number of iterations to 2-5 and still see most of the benefit of this method over using VQE with a basis set of the same size or larger.

One final point to note is that although we have used VQE to demonstrate this method, many other quantum eigensolvers could be used in its place. The main criterion is that the eigensolver returns an improved estimate for the eigenstate(s) over its input state(s),

which is trivially satisfied by many quantum eigensolvers. One could, for instance, explore using this orbital optimization method to improve the accuracy of Quantum Phase Estimation,⁸ α -VQE,¹³⁹ Quantum Imaginary Time Evolution (QITE),¹³² variational QITE,¹³³ Quantum Monte Carlo,¹³⁶ ADAPT-VQE,⁹⁴ and qubit-ADAPT-VQE.¹³⁸ Excited state eigensolvers could be explored as well. The three most obvious candidates would be Quantum Subspace Expansion (QSE),^{102,103} quantum Equation of Motion (qEoM),¹²² and EOM-VQE.¹⁴⁰ These methods operate by first performing the ground state search using an algorithm such as VQE, then performing a classical post-processing diagonalization step to find low-lying excited states of the Hamiltonian. Thus, OptOrbVQE could be used as a ground state solver for these methods. Two other excited states eigensolver for which it would be straightforward to incorporate this orbital optimization procedure would be multistate contracted VQE (MC-VQE)¹¹² and Subspace Search VQE (SSVQE).¹¹¹ These two methods both apply an ansatz circuit to a set of mutually orthogonal input states and minimize an objective function consisting of a weighted sum of expectation values of the Hamiltonian with respect to each of the resulting parameterized states. OptOrbVQE could easily be generalized to “OptOrbMC-VQE” or “OptOrbSSVQE” by modifying Eq. 5.4 to be a weighted sum of the transformed Hamiltonian with respect to mutually orthogonal parameterized states in the same manner as these methods. Orbital optimization could also be applied to the quantum Orbital Minimization Method (qOMM)¹¹⁴ by modifying Eq. 5.4 in an analogous way. These methods all find low-lying excited states simultaneously through the minimization of a single objective function. Variational Quantum Deflation (VQD)¹⁰⁷ is different from these other methods in that it finds the low-lying excited states sequentially through a series of minimization procedures. Thus, the application of orbital optimization to VQD would be more involved than simply modifying Eq. 5.4, but could still be investigated. We leave the investigation of the application of the orbital optimization procedure to these eigensolvers to future work.

6. Excited States Orbital Optimization

This chapter contains material that we have published in the Journal of Chemical Theory and Computation on our work on excited states orbital optimization.¹²⁰ In Sec. 6.1 we discuss the motivation for the current work. In Sec. 6.2 we briefly review quantum state-averaged eigensolvers. In Sec. 6.3 we formally outline the orbital optimization method presented. In Sec. 6.4 we present the numerical simulations of the proposed method for several small molecules. In Sec. 6.5 we discuss the results and potential directions of future research.

6.1 Introduction

In the previous chapter we presented the OptOrbVQE algorithm and showed that it can often outperform fixed large basis VQE and FCI while using a fraction of the qubits. This offers strong motivation for generalizing this method to find orbital-optimized excited states as well. Here we present one such way to do so and present numerical simulations demonstrating results analogous to the ground state case. Conceptually, the most straightforward way to do this is to construct an orbital optimization objective function which takes the same form as objective functions used for fixed-basis excited states problems such as SSVQE,¹¹¹ MCVQE,¹¹² VQD,¹⁰⁷ or qOMM.¹¹⁴ The ground state orbital optimization method presented previously works because the orbital rotation parameters can be viewed as extending the variational space of the parameters used in the ansatz circuit. In essence what we are doing is taking a high-dimensional Hamiltonian with a corresponding high-dimensional Hilbert space and using orbital rotations to find an optimal low-dimensional subspace of that larger Hilbert space. The eigensolver then finds the optimal ground state within that subspace. Because we are at all times constrained to solving for states that are not outside the Hilbert space of the original large-dimensional Hamiltonian, the variational principle holds for the extended space as well, allowing us to minimize the expectation value of the energy with respect to both the ansatz and orbital rotation parameters. The same is true for excited states objective functions as well. We can minimize such objective

functions over the joint space of the ansatz parameters and the orbital rotation parameters to find low-lying excited states in an optimized basis. The simplest excited states method to generalize to the orbital-optimized case would be the state-averaged approach used by SSVQE and MCVQE. This method finds one set of ansatz parameters and one orbital rotation operator which minimize the weighted sum of the energies of an ensemble of low-lying excited states. This is known as state-averaged orbital optimization and has been previously explored in the literature for CASSCF-style orbital optimization.¹³⁰ An orbital optimization scheme inspired by the VQD and qOMM objective functions would allow us to perform what is known in the literature as state-specific orbital optimization. That is, separate basis sets are used to represent the individual excited states and the orthogonality of the excited states is enforced in the objective function. This has been explored in the classical computing literature for the case of a VQD-like orbital optimization objective function.¹⁴¹ The benefit to this approach is that optimizing a basis for each state individually may result in more accurate results. There is no reason to think that the optimal basis set representation for one excited state will be equally optimal for representing any of the other excited states. Thus using one basis set to represent multiple states as is done in state-averaged methods may result in sacrificing some accuracy in a democratic way among the ensemble of excited states. The main drawback to state-specific orbital optimization is that the overlap between pairs of excited states in different basis representations must be enforced so that the orbital optimization subroutine does not simply default to the basis which separately finds the optimal ground state basis for each excited state. The method which Yalouz et. al propose that does this in the classical computing setting assumes that we have access to a configuration interaction expansion representation of each of the excited states.¹⁴¹ In the quantum computing setting, we do not necessarily have easy access to such a representation as doing so would require estimating an exponential number of coefficients through tomography and storing them in classical memory. In VQE for example, we only know the parameter values of the ansatz circuit and the expectation values of various observables with respect to this circuit. The exponential scaling of this CI representation

negates the entire purpose of using quantum computers to evade the exponential scaling of classical FCI. In order to develop state-specific approaches for use on a quantum computer, we must find other methods for computing such overlaps or developing orbital optimization techniques which avoid their use altogether. In the present work we constrain ourselves to a state-averaged method and leave the search for state-specific methods to future work. Even without the potential intricacies involved in state-specific approaches, there is much that can be investigated for state-averaged approaches on quantum computers. For example, in our previous work on the development of the qOMM algorithm, we compared qOMM to the SSVQE algorithm.¹¹⁴ We found that both methods demonstrated a certain amount of sensitivity to the ansatz expressiveness and choice of initialization, with SSVQE being more severely affected. What this work did not show was whether this was due to the fact that SSVQE converges to the set of low-lying eigenstates (whereas qOMM converges to a low-lying eigenspace) or the fact that SSVQE rigidly imposes the constraint that the set of states must remain orthogonal throughout the optimization (whereas the states in qOMM are only guaranteed to be orthogonal at the global minimum). MCVQE is more similar to qOMM in that it also converges to a low-lying eigenspace, but applies the same rigid orthogonality constraint as SSVQE. In this work we combine both SSVQE and MCVQE with orbital optimization to see which converges more reliably in the orbital-optimized case. We attempt to gain some insight to what was not explored in the qOMM work by comparing the orbital optimized MCVQE and SSVQE in numerical simulations. Furthermore, we test both methods using various state initialization methods and levels of ansatz expressiveness. This is motivated by the fact that in the fixed-basis excited states work we published, we learned that the state-averaged SSVQE approach is sensitive to the choice of ansatz and initialization. This serves to explore the extent to which both of these factors affect the convergence of state-averaged orbital optimization. We note that even if a state-averaged orbital optimization objective function is used, the eigensolver subroutine is not required to be a state-averaged eigensolver. An overlap scheme such as VQD or qOMM could be used with state-averaged orbital optimization, however in this work we constrain ourselves

to use state-averaged eigensolvers as it simplifies the notation and the interpretation of the data.

6.2 State-Averaged Quantum Eigensolvers

Hybrid quantum-classical variational methods for finding eigenvalues of chemical Hamiltonians operate by classically minimizing an objective function constructed from quantities measured on a quantum computer. For example, to find the ground state of a Hamiltonian \hat{H} we would first prepare a parametrized state $|\psi(\boldsymbol{\theta})\rangle$ on the quantum computer, measure the expectation value of \hat{H} , and carry out the minimization problem:

$$\min_{\boldsymbol{\theta}} \langle \psi(\boldsymbol{\theta}) | \hat{H} | \psi(\boldsymbol{\theta}) \rangle \quad (6.1)$$

classically. This is the original formulation of the variational quantum eigensolver^{9,78} (VQE). In order to extend this method to low-lying excited states, the mutual orthogonality of these states must be accounted for. Several methods have been proposed that accomplish this. SSVQE¹¹¹ and MCVQE¹¹² are state-averaged approaches which apply a parameterized circuit $\hat{U}(\boldsymbol{\theta})$ to a set of mutually orthogonal initial states $\{|\psi_i\rangle\}$, then minimize an objective function of the form:

$$f(\boldsymbol{\theta}) = \sum_i w_i \langle \psi_i | \hat{U}^\dagger(\boldsymbol{\theta}) \hat{H} \hat{U}(\boldsymbol{\theta}) | \psi_i \rangle \quad (6.2)$$

where $\{w_i\}$ is a set of positive, real-valued weights. The main difference between MCVQE and SSVQE is that MCVQE chooses the weights $\{w_i\}$ to be equal, whereas SSVQE chooses them to be distinct. At first glance this difference seems trivial, however it should be noted that unequal weights corresponds to a global minimum comprised of the low-lying eigenvectors, whereas an equal weighting corresponds to a global minimum comprised of states which span the low-lying eigenspace. MCVQE adds a classical post-processing step which diagonalizes these states in this low-dimensional eigenspace to acquire the low-lying eigenvectors. It is unclear which of these approaches is advantageous or if their convergence is equivalent in practice.

6.3 State-Averaged Orbital Optimization

In OptOrbVQE we take the electronic structure Hamiltonian in its fermionic second-quantization representation:

$$\hat{H} = \sum_{p,q=1}^M h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{p,q,r,s=1}^M v_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_s \hat{a}_r, \quad (6.3)$$

and rotate the set of M orbitals $\{\psi_1, \psi_2, \dots, \psi_M\}$ according to the partial unitary transformation \hat{V} :

$$\tilde{\psi}_i = \sum_j^M \hat{V}_{ji} \psi_j \quad (6.4)$$

resulting in a new set of $N < M$ orbitals $\{\tilde{\psi}_1, \tilde{\psi}_2, \dots, \tilde{\psi}_N\}$. This is equivalent to transforming the Hamiltonian as:

$$\begin{aligned} \tilde{H}(\hat{V}) &= \sum_{p',q'=1}^N \sum_{p,q=1}^M h_{pq} \hat{V}_{pp'} \hat{V}_{qq'} \hat{a}_p^\dagger \hat{a}_{q'} \\ &+ \frac{1}{2} \sum_{p',q',r',s'=1}^N \sum_{p,q,r,s=1}^M v_{pqrs} \hat{V}_{pp'} \hat{V}_{qq'} \hat{V}_{ss'} \hat{V}_{rr'} \hat{a}_p^\dagger \hat{a}_{q'}^\dagger \hat{a}_{s'} \hat{a}_{r'}. \end{aligned} \quad (6.5)$$

The orbital optimization then corresponds to minimizing the expectation value of this Hamiltonian with respect to a fixed quantum state $\hat{U}(\boldsymbol{\theta})|\psi_{ref}\rangle$ provided by a quantum eigensolver. The total minimization problem is then given by:

$$\min_{\substack{\boldsymbol{\theta} \\ \hat{V} \in \mathcal{U}(M,N)}} \langle \psi_{ref} | \hat{U}^\dagger(\boldsymbol{\theta}) \tilde{H}(\hat{V}) \hat{U}(\boldsymbol{\theta}) | \psi_{ref} \rangle \quad (6.6)$$

where $\mathcal{U}(M,N)$ is the set of $M \times N$ real partial unitary matrices. The simplest way to generalize this problem is to consider Eq. 6.2 to be a function of both $\boldsymbol{\theta}$ and \hat{V} :

$$f(\boldsymbol{\theta}, \hat{V}) = \sum_i w_i \langle \psi_{ref,i} | \hat{U}^\dagger(\boldsymbol{\theta}) \tilde{H}(\hat{V}) \hat{U}(\boldsymbol{\theta}) | \psi_{ref,i} \rangle \quad (6.7)$$

and minimize the resulting state-averaged analog problem of Eq. 6.6:

$$\min_{\hat{V} \in \mathcal{U}(M,N)} f(\boldsymbol{\theta}, \hat{V}). \quad (6.8)$$

Such state-averaged analogs of CASSCF-like orbital optimization schemes^{130,142} have previously been explored in the literature. It is worth noting that an overlap-based orbital optimization objective function has been proposed in the classical literature,¹⁴¹ which allows for a separate optimal basis to be computed for each excited state. The authors claim that this allows for more accurate excitation energies to be computed. The method assumes the availability of the CI coefficients found by the eigensolver, which would require exponentially-expensive full state tomography to acquire in the quantum computing setting.

The total minimization problem Eq. 6.8 is divided into two subproblems: minimization with respect to the ansatz parameters θ and minimization with respect to \hat{V} . These two subproblems are solved in an alternating fashion, where one is fixed while the other is varied. The optimal parameters for one subproblem are then used for the initialization of the next run of the other until some global stopping criteria are met. For example, for a given optimal \hat{V} we can compute $\tilde{H}(\hat{V})$ and carry out a quantum excited states solver to find an optimal θ in the rotated basis. For a given θ found by a quantum excited states solver, we can compute the 1 and 2-RDMs with respect to each state in the set of computed excited states $\{\hat{U}(\theta)|\psi_{\text{ref},i}\rangle\}$, then vary Eq. 6.8 with respect to \hat{V} . The optimization with respect to θ is handled via one of several known quantum excited states solvers such as SSVQE,¹¹¹ MCVQE,¹¹² VQD,¹⁰⁷ or qOMM.¹¹⁴ The optimization with respect to \hat{V} (keeping θ fixed) is carried out using an orthogonally-constrained optimization procedure. In this work we use an orthogonally-constrained projected gradient method,³³ which has a parameter update step defined as:

$$\hat{V}_{n+1} = \text{orth}(\hat{V}_n - \eta \nabla f(\hat{V}_n)) \quad (6.9)$$

where $\nabla f(\hat{V}_n)$ is the gradient of Eq. 6.7 with respect to \hat{V} with fixed θ , η is a step size which is chosen adaptively in an alternating Barzilai-Borwein fashion, and the orth function is defined as:^{27,33}

$$\text{orth}(A) = AQA^{-\frac{1}{2}}Q^\dagger. \quad (6.10)$$

Here Q is a matrix whose columns are the eigenvectors of $A^\dagger A$ and Λ is a diagonal matrix whose entries are the eigenvalues of $A^\dagger A$. As was done for OptOrbVQE, we explicitly state the super and subscript notation used for the total problem to avoid confusion:

- The subscript l will index the iteration number in the minimization problem where \hat{V} is varied.
- The subscript m will index the iteration number in the minimization problem where $\boldsymbol{\theta}$ is varied.
- The subscript n will index a global “outer loop” iteration number that characterizes how many times both subproblems have been carried out.
- The superscript opt will denote the optimal parameter found in each subproblem for a given outer loop iteration number.

We now give an explicit step-by-step procedure for the total problem:

1. Set $n = 0$. Choose an initial partial unitary $\hat{V}_{n=0,l=0}$, an initial set of ansatz parameters $\boldsymbol{\theta}_{n=0,m=0}$, and a stopping threshold ϵ .
2. Calculate $\tilde{H}(\hat{V})$ on a classical computer and run a quantum eigensolver algorithm to obtain $\boldsymbol{\theta}_n^{opt}$.
3. If $|f(\boldsymbol{\theta}_n^{opt}, \hat{V}_{n-1}^{opt}) - f(\boldsymbol{\theta}_{n-1}^{opt}, \hat{V}_{n-2}^{opt})| < \epsilon$, halt the algorithm. Else, continue to the next step.
4. Measure the 1 and 2-RDMs with respect to the set of states $\{\hat{U}(\boldsymbol{\theta}_n^{opt})|\psi_{\text{ref},i}\rangle\}$ on a quantum computer.
5. Using the 1 and 2-RDMs from the previous step, minimize Eq. 6.8 with respect to \hat{V}_n to obtain \hat{V}_n^{opt} .
6. Set $n = n + 1$, $\hat{V}_{n+1,l=0} = \hat{V}_n^{opt}$, and $\boldsymbol{\theta}_{n+1,m=0} = \boldsymbol{\theta}_n^{opt}$. Optionally, a small random perturbation can be added to the latter two quantities. Repeat steps 2-6.

Step 5 requires the use of a classical optimizer which constrains \hat{V} to be a partial unitary.

Several methods which do this exist,^{29–32} but in our work we use an orthogonal projection method.³³ In general, $\boldsymbol{\theta}$ and $\hat{V}_{n+1,l=0}$ could be any real vector and real partial unitary, respectively, however it is intuitive to use information from the n th outer loop iteration to inform this choice. In our work we choose $\boldsymbol{\theta}_{n+1,l=0} = \boldsymbol{\theta}_n^{opt}$ and $\hat{V}_{n+1,l=0} = \text{orth}(\hat{V}_n^{opt} + \text{Rand}(M, N))$, where $\text{Rand}(M, N)$ is an $M \times N$ matrix whose elements are sampled from a normal distribution with average 0 and standard deviation 0.01. Additionally, although Eq. 6.8 is written as a state-averaged function of $\boldsymbol{\theta}$, step 2 does not necessarily need to be carried out using a state-average quantum eigensolver. The only requirement is that the solver returns solution states to be used for the calculation of 1 and 2-RDMs. Overlap-based methods such as qOMM¹¹⁴ and VQD¹⁰⁷ could be used, however for our work we test MCVQE¹¹² and SSVQE.¹¹¹

We further note that because the method derives an optimized basis of $N < M$ orbitals from an initial large basis of M orbitals, it cannot match or exceed the accuracy obtained by solving the eigenvalue problem in the full M orbital active space. This is intuitive from the perspective that the ansatz and orbital rotation parameters constitute a joint variational space. Restricting the variational space will in general restrict the maximum attainable accuracy. In particular, it is known that orbital optimization methods in general are not efficient at capturing correlation effects arising from Coulomb repulsion between electrons. For this, one could consider combining the orbital optimization scheme with other methods such as explicitly correlated methods.^{126,127,143}

6.4 Numerical Results

The code used for our numerical simulations is an extension of the functionality provided by the open source package Qiskit.¹¹⁷ Qiskit provides an implementation of VQE,^{9,78} which we have modified to produce implementations of SSVQE¹¹¹ and MCVQE.¹¹² The code for the state-averaged orbital optimization is a modification of the code used in our ground state orbital optimization work,¹¹⁹ with the main modification being the objective function to be minimized. The Qiskit package versions used are Qiskit-Aer 0.12.0, Qiskit-

Nature 0.4.5, and Qiskit-Terra 0.23.2. The 1 and 2-body integrals are obtained through the PySCF¹¹⁸ electronic structure driver in Qiskit, which uses PySCF to perform a restricted Hartree-Fock problem to obtain the un-optimized molecular integrals. Configuration interaction circuits are obtained in two steps. First, the truncated Hamiltonians are constructed from the 1 and 2-body integrals using the Slater-Condon rules,¹ which are then exactly diagonalized using NumPy.¹⁴⁴ The reasons why we do not use PySCF’s configuration interaction implementation are two-fold: **1.** PySCF does not have an CIS implementation and **2.** We have found that PySCF’s CISD implementation does not always produce orthogonal CI wavefunctions, with fidelity between two states being as large as on the order of 10^{-1} , even in the case where the corresponding eigenvalues are not degenerate. This is problematic for quantum algorithms such as SSVQE and MCVQE which require that the initial states be mutually orthogonal.

This statevector can then be used to initialize a circuit using Qiskit’s arbitrary statevector initialization implementation. We note that although this particular implementation requires the storage of an exponentially large statevector in classical memory, in principle configuration interaction state preparation on a quantum computer could be done in a completely sparse manner with resources scaling polynomially with the number of qubits. For example, it has been shown that Givens rotations are universal for preparing chemically-motivated states with the Jordan-Wigner mapping,⁵⁹ where the authors also give a general procedure for preparing an arbitrary statevector. In Appendix C we give an explicit example of how the particular case of CIS statevectors can be prepared on a quantum computer. Whether or not an efficient analogous procedure can be developed for CISD states is not discussed here, however in our simulations we include CISD initializations to investigate whether or not doing so would lead to further improvement. We also utilize an "excited Hartree-Fock" initialization that consists of the Hartree-Fock state and the lowest energy singly-excited states from it. In this paper, we will refer to this initialization as just "Hartree-Fock" or "HF". In Appendix B we test all three of these initializations for state-averaged quantum eigensolvers in the fixed minimal STO-3G basis.

In Qiskit, one can use any ansatz circuit as a base pattern to be repeated n times, increasing the circuit depth and number of parameters by a factor of n . In our simulations we use Qiskit’s implementation of the UCCSD ansatz⁵ as a circuit block pattern to be repeated for various values of n . We denote this as n -UCCSD. The classical optimizer used for all test instances is L-BFGS-B,¹⁰⁰ the FCI reference values are calculated using Coordinate Descent FCI (CDFCI),²⁸ and all orbital-optimized tests are run using Qiskit’s *AerSimulator* in noiseless statevector mode unless stated otherwise. In Sec. 6.4.5 we make slight adjustments to this methodology when calculating a potential energy surface of H_4 .

6.4.1 H_2

We begin with our results for the simplest model tested, H_2 at the near-equilibrium bond distance 0.735 Å, which are given in Fig. 6.1. We use cc-pVQZ (120 spin-orbitals) as the starting basis and reduce the active space for even numbers of spin-orbitals from 4 to 14 using the proposed orbital optimization scheme. The difference between the average orbital optimized energy and that of FCI (over the ensemble of ground and excited states) in the cc-pVTZ basis is plotted as a function of the outer loop iteration. Tests using both 2 and 3-UCCSD are included to investigate the effect of increasing the ansatz expressiveness in the algorithm. Both eigensolvers are initialized with configuration interaction singles (CIS) states. SSVQE is additionally tested using the Hartree-Fock initialization. It is evident that orbital optimization has the potential to achieve more accurate average energies than FCI in the cc-pVDZ basis (20 spin-orbitals) and can even approach cc-pVTZ quality values, but this is highly dependent on the choice of eigensolver, ansatz, and number of optimized spin-orbitals. A minimum of 8 spin-orbitals are needed to achieve a higher accuracy than cc-pVDZ. At this point, OptOrbMCVQE can do this for both 2 and 3-UCCSD, although OptOrbSSVQE cannot. Using 10 spin-orbitals, both eigensolvers surpass cc-pVDZ for both 2 and 3-UCCSD, although MCVQE offers roughly a 5 milli-Hartree improvement over SSVQE for 2-UCCSD. When 3-UCCSD is used, MCVQE offers a measurable but negligible improvement over SSVQE. At 14 spin-orbitals, cc-pVTZ-quality results are achievable.

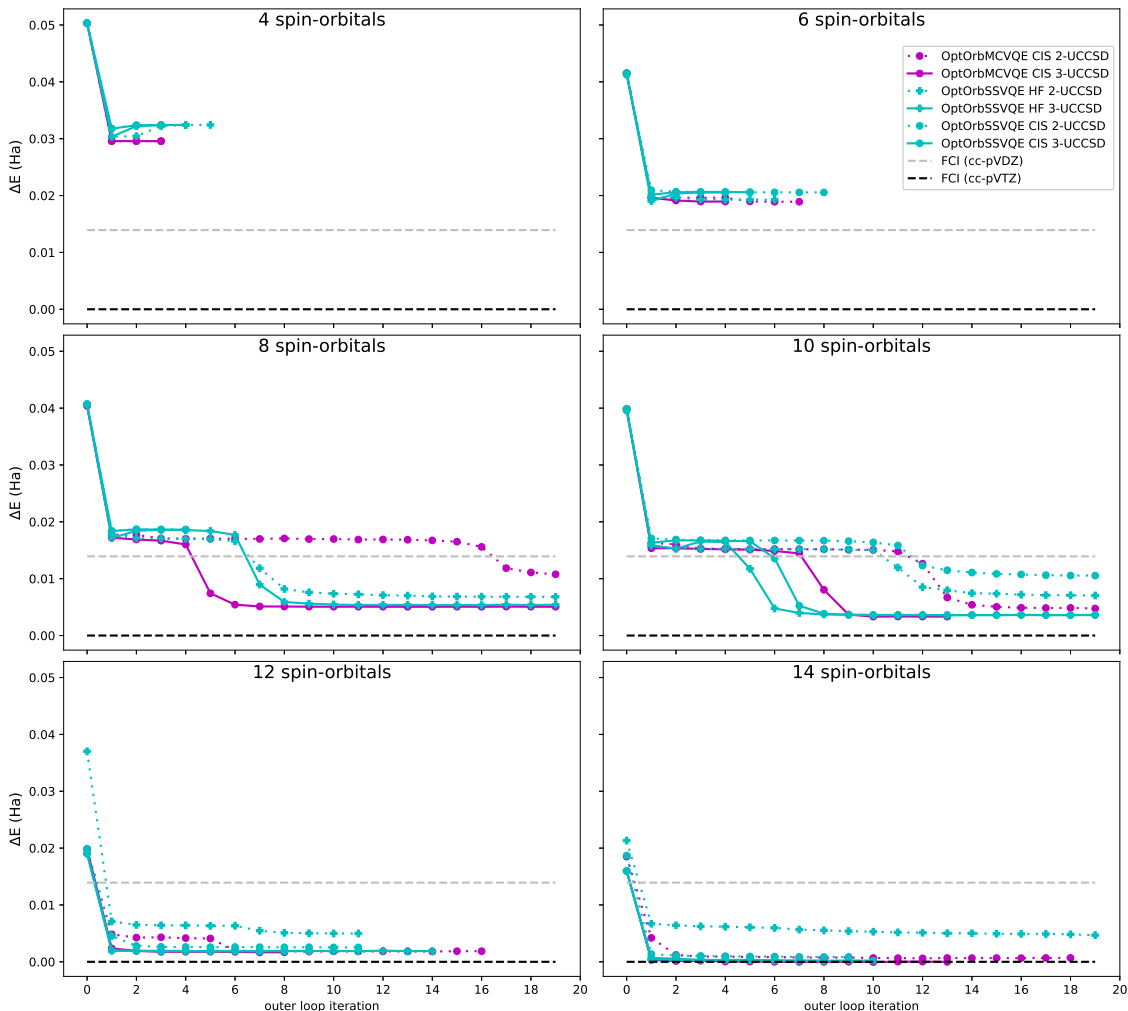


FIGURE 6.1: Convergence of orbital optimization methods for H_2 using various numbers of spin-orbitals (taken from the cc-pVQZ basis) as a function of the outer loop iteration. ΔE is the difference between the average energy and that of FCI in the cc-pVTZ basis (56 spin-orbitals).

Also notable is the effect that increasing the active space has on not only the quality of convergence, but its rate of convergence. Note that for 8 and 10 spin-orbitals, the convergence appears to plateau, hovering just above cc-pVDZ accuracy for several iterations before rapidly surpassing it. This behavior is not present at 12 and 14 spin-orbitals, with the energy quickly converging to near or at cc-pVTZ accuracy for the majority of tests run.¹

¹ As a sidenote, we note that the 14 spin-orbital tests using 3-UCCSD with a CIS initialization were stopped manually at iterations 10 and 13 for SSVQE and MCVQE, respectively as the runtime for these simulations proved to be the longest among these tests. However, we note that given that nearly all of the

We also note that for the 8 and 10 spin-orbital cases, using 3-UCCSD offers not only more accurate results than 2-UCCSD, but the convergence is faster as well. This is analogous to what we found in our work on qOMM,¹¹⁴ where it was found that an over-parameterization of the ansatz can result in faster convergences in some instances.

6.4.2 H₄

We now present the results for H₄, a toy system comprised of four hydrogen atoms arranged in a square with a nearest-neighbor distance of 1.23 Å. The starting basis set is cc-pVQZ (240 spin-orbitals) and an active space of 8 optimized spin-orbitals is used. Both 2 and 3-UCCSD are tested as ansatzes and both CIS and CISD are tested as initializations. SSVQE is additionally tested using the Hartree-Fock initialization. The results are given in Fig. 6.2.

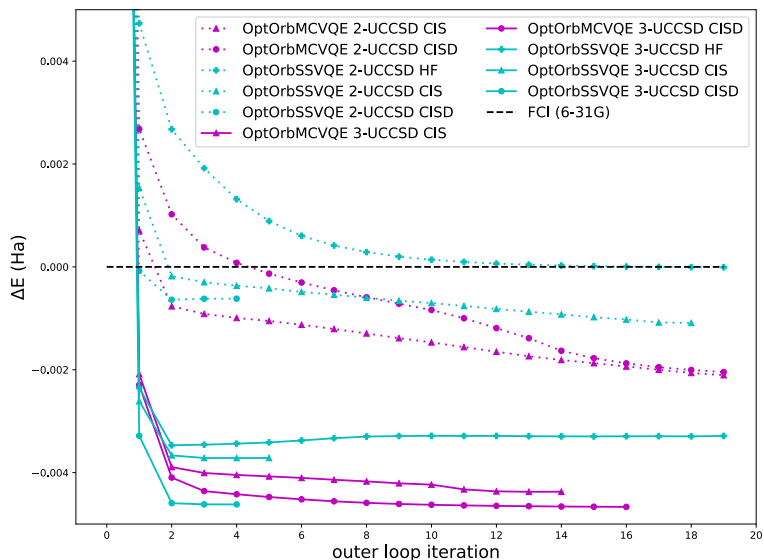


FIGURE 6.2: Convergence of orbital optimization methods for H₄ using 8 optimized spin-orbitals (taken from the cc-pVQZ basis) as a function of the outer loop iteration. ΔE is the difference between the average energy and that of FCI in the 6-31G basis (16 spin-orbitals).

We can see that for this system, orbital optimization can be used to achieve a more energy convergence occurred within the first 2 or 3 iterations, allowing the simulations to continue would likely not have resulted in further improvement.

accurate average energy than the 6-31G basis (16 spin-orbitals), despite the fact that it is utilizing half the number of spin-orbitals. Convergence approaching FCI cc-pVDZ (40 spin-orbitals) accuracy was not observed in our testing, indicating that as in the ground state case in our previous work, larger active spaces will be needed for approaching the infinite basis set limit to chemical accuracy. Between the three different algorithmic choices considered (the eigensolver, the initialization, and the ansatz), increasing the ansatz expressiveness from 2-UCCSD to 3-UCCSD had the most significant effect on the converged accuracy. Changing the initialization from CIS to CISD offered a clear improvement when used with the 3-UCCSD ansatz, however the same is not true for 2-UCCSD. With 2-UCCSD, OptOrbSSVQE using CISD converges quickly to a local minimum, whereas OptOrbSSVQE using CIS converges (albeit comparatively slowly) to a more accurate average energy. The final converged values for OptOrbMCVQE are similar between CIS and CISD when using 2-UCCSD. Note also that for instances using the same ansatz and initialization, using MCVQE as the eigensolver typically offers an improvement over SSVQE. The one exception to this is using CISD with 3-UCCSD, where the difference between these two converged values is negligible.

6.4.3 LiH

We now present the results for the first two energy levels of LiH at the near-equilibrium interatomic distance of 1.595 Å. The starting basis set is cc-pVTZ (88 spin-orbitals) and an active space of 12 optimized spin-orbitals is used. Both 1 and 2-UCCSD are used to assess the effect of ansatz expressiveness. CIS and CISD initializations are tested for MCVQE whereas SSVQE additionally tests the Hartree-Fock initialization. The results are shown in Fig. 6.3.

The most notable feature of this plot is that orbital-optimized solvers can achieve more accurate results than FCI using much larger basis sets. For example, most tests run for this system outperform FCI 6-31G (22 spin-orbitals) while using only 12 spin-orbitals. Depending on the choice of solver, ansatz, and initialization, some instances also outperform FCI

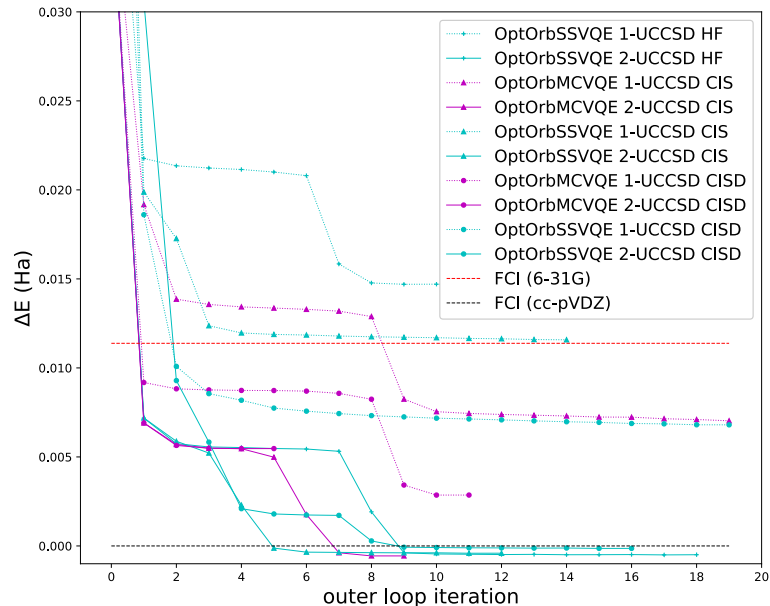


FIGURE 6.3: Convergence of orbital optimization methods for LiH using 12 optimized spin-orbitals as a function of the outer loop iteration. ΔE is the difference between the average energy and that of FCI in the cc-pVDZ basis (36 spin-orbitals).

cc-pVDZ (36 spin-orbitals). The second notable feature is that ansatz expressiveness (typically) has a greater influence on the final accuracy than the choice of initialization. The 1 and 2-UCCSD tests almost form two cleanly separated accuracy tiers, except for OptOrbMCVQE using 1-UCCSD with a CISD initialization, which achieves a higher accuracy than MCVQE using CISD with 2-UCCSD. (It is likely that the 2-UCCSD test here converged to a local minimum.) The choice of initialization has a greater impact when the less expressive 1-UCCSD ansatz is used and has little impact when the ansatz is sufficiently expressive to approximate the solution states well. The third notable feature is that OptOrbMCVQE typically outperforms OptOrbSSVQE when using the same ansatz and initialization, with the one exception to this being when CISD and 2-UCCSD are used. This effect is most noticeable when the less expressive 1-UCCSD ansatz is used.

6.4.4 BeH₂

We now present the results for the first two energy levels of BeH₂ with a linear geometry at the near-equilibrium Be-H distance of 1.3264 Å. We find the full system with 14 spin-orbitals and 6 electrons to be intractable for our computational budget, so we freeze two electrons in the Hartree-Fock orbitals with the lowest energy and compare the active space energy against that of FCI using the same frozen core approximation. Because we do not wish for the quality of the frozen core approximation across different basis sets to influence the comparison against FCI values, here we will only compare the orbital optimized results starting with the cc-pVQZ basis with an active space of 12 spin-orbitals against FCI in the cc-pVQZ basis using an active space of 228 spin-orbitals. Because of the large disparity in active space size, we do not expect the orbital optimized tests to approach chemical accuracy compared to FCI, however we may still gain some insight as to what portion of the full basis set energy is attainable using a small active space and what kind of improvement orbital optimization offers over a naive approach which chooses a fixed active space based on Hartree-Fock orbital energies. These results are given in Fig. 6.4.

We find that orbital optimization offers about a 0.6 Hartree improvement over naively choosing an active space based on the Hartree-Fock energy of the orbitals, however there remains a 0.4 Hartree energy gap between the orbital optimized case and the full active space FCI value. This again clearly indicates that larger active spaces will be required to approach the infinite basis set limit.

6.4.5 H₄ Noisy Binding Curve

In this section we use OptOrbMCVQE to compute the potential energy surface resulting from uniformly stretching the nearest-neighbor interatomic distance of the H₄ toy model. We find that this potential energy surface is difficult for state-averaged solvers such as MCVQE to accurately describe in the STO-3G basis, even when starting from CISD states. This is especially true for stretched geometries, where MCVQE often converges to local minima. Thus, this may serve as an interesting test bench for orbital-optimized state-

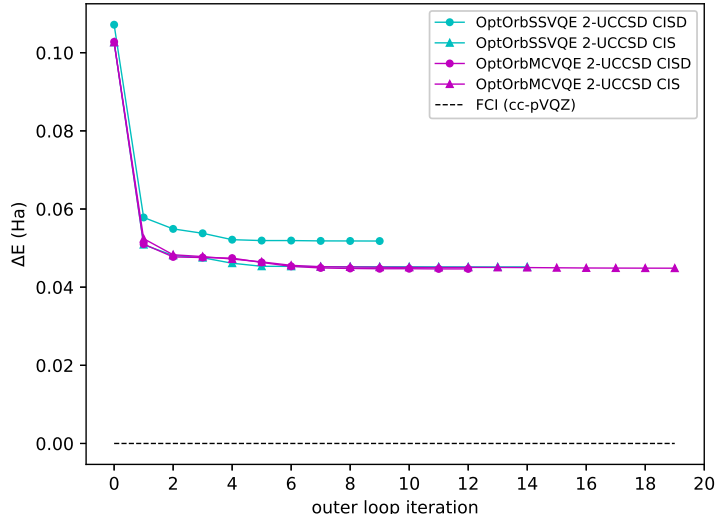


FIGURE 6.4: Convergence of orbital optimization methods for BeH_2 using 12 optimized spin-orbitals as a function of the outer loop iteration. ΔE is the difference between the average energy and that of FCI in the cc-pVTZ basis.

averaged eigensolvers such as OptOrbMCVQE. We study both MCVQE (STO-3G with 8 spin-orbitals) and OptOrbMCVQE (cc-pVTZ with 8 spin-orbitals) in the noiseless case as well as in the case where noise arises from statistical sampling. These results are compared against the exact FCI values in the STO-3G and cc-pVTZ basis sets for comparison. The methodology is largely the same as in previous sections with some notable exceptions. The first is that we now use the COBYLA¹⁰¹ optimizer for the eigensolver subroutine as we find that it is more robust to noise than L-BFGS-B. Furthermore, we use PySCF’s FCI implementation to generate exact comparison values. As opposed to previous sections where we studied various initializations and ansatz expressiveness, here we restrict ourselves to CISD initializations with the 3-UCCSD ansatz. As we will present the individual energy levels instead of the state-averaged energy, this simplifies presentation of the data. OptOrbMCVQE uses cc-pVTZ as the starting basis and uses an active space of 8 spin-orbitals. MCVQE uses the 8 spin-orbital STO-3G basis. For the tests which incorporate statistical sampling noise, we use 10^6 shots per observable evaluation using Qiskit’s approximate treatment of sampling

noise, which assumes a Gaussian distribution for observable measurements and returns the variance σ . We estimate the error associated with these measurements for a number of shots n as $\sqrt{\frac{\sigma}{n}}$. We find that the variance returned by Qiskit is typically on the order of 10^{-1} . Fig. 6.5a shows the absolute energies for the binding curve for the first two energy levels of H_4 for OptOrbMCVQE and MCVQE. The STO-3G (8 spin-orbitals) and cc-pVTZ (112 spin-orbitals) are included for comparison. Fig. 6.5b illustrates the relative accuracies of the energies as compared to the FCI value in the cc-pVTZ basis. Note that OptOrbMCVQE uses cc-pVTZ as its starting basis but uses an active space of 8 spin-orbitals. Thus, these (along with the other tests in this work) represent an aggressive reduction in active space size. We can see that while OptOrbMCVQE offers a significant increase in the absolute

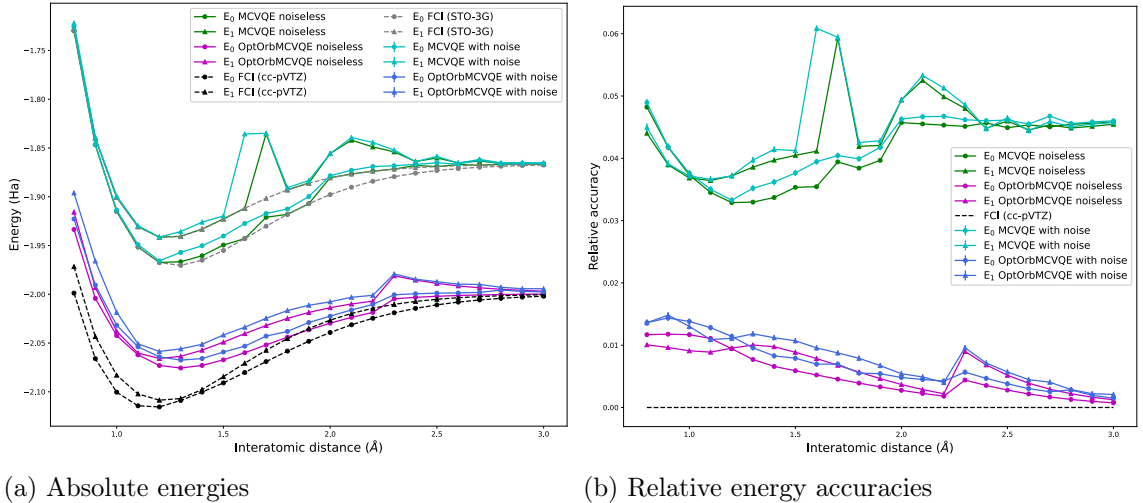


FIGURE 6.5: The potential energy surface of the first two energy levels of H_4 corresponding to uniformly stretching nearest-neighbor interatomic distances. a) Absolute energies for OptOrbMCVQE (cc-pVTZ, 8 spin-orbitals) and MCVQE (STO-3G, 8 spin-orbitals) along with the FCI values in the STO-3G and cc-pVTZ basis sets for comparison. b) Relative accuracies of E_0 and E_1 for the various methods compared to their exact FCI values in the full cc-pVTZ basis (112 spin-orbitals). Error bars are included but are small.

energies over MCVQE in the fixed STO-3G basis, its relative accuracy compared to the full cc-pVTZ value is still on the order of 10^{-2} for many geometries. This demonstrates that although orbital optimization offers a compact basis set representation, basis sets beyond

the minimal basis will still clearly be required in order to approach chemical accuracy in the infinite basis set limit. Interestingly, we also note that the potential energy surface of the orbital-optimized tests are much more smooth in comparison to the fixed-basis MCVQE tests. MCVQE in the STO-3G basis clearly struggles to reliably converge to the exact values, particularly for interatomic distances of 1.6 Å and larger. The jump in energies around 2.0 Å for OptOrbMCVQE similarly indicates that the method is likely converging to local minima for stretched bond distances, although this effect is qualitatively less severe than in the fixed basis case. We also note that as one would expect, the effect of noise arising from statistical sampling is to degrade the energy accuracies by an amount on the milli-Hartree level. Encouragingly, the majority of the additional accuracy attained through orbital optimization persists despite the presence of this noise.

6.5 Discussion and Conclusions

In this work we have proposed an orbital optimization scheme which uses a state-averaged approach to compute excited states of electronic structure Hamiltonians. We have shown that this method can achieve more accurate results than FCI using much larger fixed basis sets, however active space sizes much larger than the ones here will be required to reach chemical accuracy. We have also investigated the effects of the choice of quantum eigensolver, ansatz expressiveness, and state initialization. While exceptions to these trends can be found in our results, we can make the following general observations:

- Increasing the ansatz expressiveness offers the most significant effect among these factors.
- MCVQE often offers an improvement in accuracy over SSVQE for lower ansatz expressiveness. When higher expressiveness is used, the difference is often less significant.
- CIS initializations often offer an improvement over Hartree-Fock initializations, however the advantage of using CISD over CIS is unclear. There are several instances of CISD initialized tests achieving a lower accuracy than their CIS (and even Hartree-Fock) counterparts.

The first of these is not surprising. The ansatz expressiveness is what primarily determines the variational flexibility of the quantum eigensolver at each outer loop iteration. The second observation can be explained by noting that for a given initialization and ansatz, the solution space of SSVQE is more restricted than MCVQE. The solution space of SSVQE with unequal weights consists of the low-lying eigenvectors themselves, whereas the solution space of MCVQE consists of the subspace spanned by the low-lying eigenvectors. MCVQE utilizes a post-processing step involving a low-dimensional diagonalization problem in this subspace. This additional variational flexibility may ease the convergence process and allow it to partially compensate for an insufficiently expressive ansatz. The third point, while less easily explained than the other two, can be conjectured about. While the initialization does have an effect on ansatz expressiveness as it determines which excitation operators in the UCCSD ansatz act non-trivially, is not as variationally flexible as the parameterized ansatz itself is. Furthermore, this state is computed in the initial basis set guess, which is usually low-quality compared to the optimized basis set. Thus there is no guarantee that CISD computed in this initial basis will continue to be advantageous over CIS for successive basis set rotations. Furthermore, improving the initialization guess does not guarantee that the chances of converging to a local minima are decreased. It could be that certain chemically-motivated initializations happen to be close to local minima. For example, we note that in our work on qOMM, there were some instances for which starting with a chemically-motivated initialization with all ansatz parameters set to zero resulted in the method converging to a local minimum, whereas their randomly-initialized counterparts successfully converged to their global minima.¹¹⁴

One compelling and well-motivated extension of this work would be to take a state-specific orbital optimization approach rather than a state-averaged one. State-specific orbital optimization (as the name implies) optimizes a different basis set for each excited state individually rather than optimizing one basis set for an ensemble of excited states by minimizing its average energy. State-specific orbital optimization has been developed in the context of classical orbital optimization algorithms,¹⁴¹ however this particular method relies

on a full CI expansion of the wavefunction at every outer loop iteration. In the quantum computing setting, such an explicit wavefunction expansion (as opposed to the expectation value of observables used here) would involve exponentially costly tomography and classical storage. These CI wavefunction expansions are used to compute the overlap of two different excited states in two different basis sets and uses them to enforce their orthogonality. In the quantum computing setting, one would have to develop a method which can compute these overlaps without access to CI expansions or which does not require overlaps at all. This is an interesting problem and will be a direction of future investigation.

7. The Effect of Orbital Optimization on Hamiltonian Sparsity and Circuit Depth

The contents of this chapter are not published, but consist of last-minute simulations that I decided to run to satiate my own curiosity while writing this thesis. I would consider the observations therein more as "food for thought" that could motivate future (more thorough) investigations rather than as a complete work in their own right. I noticed during the completion of the simulations for the orbital optimization methods in previous chapters that the per-iteration runtime of VQE was roughly 2-3 times longer after the first orbital optimization run compared to that for the naive guess in the $n = 0$ iteration. The most likely explanation for this is that the orbital optimization procedure rotated the problem into a basis where the number of Pauli terms in the Hamiltonian was much greater than in the initial basis. Furthermore, a work by Kivlichan et. al¹⁴⁵ briefly mentioned a potential trade-off between the "compactness" of the basis and the sparsity of the Hamiltonian. Here, by "compact", we mean that a smaller number of orbitals is needed to achieve the same accuracy as a less compact basis. That is, if you design a basis to be efficient in terms of the energy accuracy as a function of the number of spin-orbitals, the price you pay is that the resulting Hamiltonian will have more Pauli terms compared to a less compact basis of the same size. This was used as motivation for formulating circuits in their work in the basis of atomic orbitals rather than molecular orbitals. This is an interesting comment that should be investigated in the context of our own work. Optimized orbital basis sets are found iteratively with the explicit intention of being as compact as possible. If such a trade-off exists for optimized orbitals such as the ones in our works, it should be straightforward to demonstrate through numerical simulations. Additionally, one may also ask whether a similar relation could be true for the basis set compactness and the minimum ansatz expressiveness needed for VQE to converge. That is, if using a more compact basis leads to a less sparse Hamiltonian, then perhaps its eigenstates are less sparse as well. This point is investigated here.

7.0.1 Introduction

The primary purpose of orbital optimization is to maximize the accuracy of the energy levels of a given system while minimizing the size of the basis needed to do so. For this reason the previous chapters focused on comparing the energy accuracy of orbital-optimized problem instances with fixed basis sets of the same size or larger. Encouragingly, we found that orbital optimization is often able to outperform larger basis sets while simultaneously using fewer spin-orbitals. There are, however, other practical considerations to take into account. In particular, the sparsity of the Hamiltonian (the number of Pauli strings in its expansion) and the minimum circuit depth needed to represent the solution in a given basis are both important considerations for both near-term algorithms such as VQE and long-term algorithms such as QPE. In QPE, the number of Pauli strings in the Hamiltonian directly affects the circuit depth required to implement controlled versions of the time evolution operator $e^{i\hat{H}t}$. In VQE, the number of Pauli strings has an effect on the number of measurements we need to take, since the total number of measurements needed to estimate the energy to accuracy ϵ can be upper bounded as:¹⁴⁶

$$M \leq \frac{(\sum_i |h_i|)^2}{\epsilon^2} \tag{7.1}$$

where $\{h_i\}$ are the coefficients of the Pauli strings in the Hamiltonian expansion. The intuition behind this is that the amount that a particular Pauli term contributes to the overall variance of the energy estimation depends on the magnitude of its coefficient. In order to minimize the error of the energy estimation as a whole while using as few measurements as possible, Pauli strings with larger coefficients should have more measurements devoted to them and Pauli strings with smaller coefficients should have fewer measurements devoted to them. In this chapter we call the numerator of Eq. 7.1 the measurement overhead factor. We can guess that there should be some relation between the number of non-zero coefficients and the choice of basis used. Similarly, we can intuitively guess that the basis set representation and the ansatz circuit used in variational methods should not be completely

independent of each other. The "expressiveness" of an ansatz is a term that is interchangeably used to describe the ability of a given ansatz to cover a portion of the Hilbert space of a Hamiltonian or its ability to describe the solution accurately. Generally, in order to accelerate the optimization process we want an ansatz that does not cover a large portion of the Hilbert space in which the solution does not lie while simultaneously being able to represent the solution to high accuracy. What we do not know is what the effect of changing the Hilbert space itself is with regards to the ability of the ansatz to represent states in that Hilbert space. It could be that the ansatz is sufficiently expressive for one basis, then the orbital optimization process changes the basis to one where the ansatz no longer is sufficiently expressive. Because the orbital optimization process keeps the state fixed when rotating the basis, it does not have access to this information and cannot take this into account when choosing an optimal basis. The only way to ensure that the ansatz is sufficiently expressive is to choose one that is likely to be simultaneously sufficiently expressive for every basis set considered.

In this chapter we use the dissociation curve of the H_4 square toy model and the equilibrium geometry of LiH to test the effects of orbital optimization on both the Hamiltonian sparsity and the circuit depth. That is, we take four Hydrogen atoms arranged in a square lattice and uniformly stretch the interatomic distance between nearest neighbor atoms. LiH results are included to corroborate the H_4 results, given that the H_4 ground state energy appears to converge to local minima in several instances. In order to simplify this study, we consider only the ground state, although one could imagine repeating this study for excited states and comparing the relative difficulty between these two cases. In order to study the Hamiltonian sparsity, we run orbital optimized VQE with a fixed ansatz. In order to study the circuit depth, we consider the adaptive ansatz counterpart of VQE, ADAPT-VQE.⁹⁴ In ADAPT-VQE, one starts out with an initial reference state (often taken to be the Hartree-Fock state) and a pool of gates given by parameterized fermionic excitation operators such as those found in the Trotterized version of UCCSD. The ansatz used for optimization starts out empty and the gradient of each of the gates in this pool is measured. The gate

with the largest gradient is added to the ansatz and the energy is minimized. After this minimization procedure, the gradient of each gate in this pool is again measured with respect to this state and the gate with the largest gradient (for a gate parameter of zero) is added to the end of the ansatz with initial parameter zero. This process is repeated until all gradients in this pool are below a specified threshold. The goal here is to iteratively build up an ansatz such that we use only the minimum number of gates required to represent the solution for a given system. Thus, incorporating orbital optimization in ADAPT-VQE is (in principle) a natural way to explore the relation between orbital optimization and the circuit depth needed to find the solution. Let us refer to orbital optimized ADAPT-VQE as OptOrbAdaptVQE. This is a much more fine-grained approach than was taken in Chapter 6 where we tested excited states orbital optimization with various numbers of UCCSD repetitions.

7.1 Numerical Results

As in previous numerical studies, we use Qiskit,¹¹⁷ PySCF,¹¹⁸ and PyTorch.¹³⁷ Specifically, we use Qiskit Aer 0.12.0, Qiskit Terra 0.23.2, Qiskit Nature 0.4.5, Pyscf 2.0.1, and PyTorch 1.12.0. Qiskit’s implementations of VQE and ADAPT-VQE are used. The classical optimizer used is L-BFGS-B.¹⁰⁰ The Jordan-Wigner encoding is used for all tests. In order to study the circuit depth, we use the Qiskit compiler to compile the optimal circuit found by ADAPT-VQE to the set of basis gates given by CZ , R_z , SX , X , and I for a qubit connectivity graph where all qubits can be entangled with all other qubits. Qiskit’s noiseless simulator is used for OptOrbVQE and VQE. We note that there is a quirk (likely a bug) of the Qiskit code that does not allow for completely noiseless simulations of ADAPT-VQE to occur. For this reason we set the number of shots for ADAPT-VQE to be 10^{30} so that the statistical error is roughly on the order of machine epsilon (10^{-15}). This should be sufficient for the simulations here to be considered effectively noiseless.

7.1.1 H₄

For H₄ we consider VQE and ADAPT-VQE in the fixed STO-3G basis (8 spin-orbitals) as well as orbital optimized simulations which start in the cc-pVTZ basis (112 spin-orbitals) and compress the active space to one with 8 spin-orbitals. We initialize the methods in a CISD state as we find that such an initialization offers much improved convergence reliability compared to a Hartree-Fock initialization. We observe a significant degree of instability when a Hartree-Fock initialization is used. For VQE and OptOrbVQE, adding a small amount of randomness to the initial parameters eliminates the instability, however ADAPT-VQE cannot benefit as greatly from such a random initialization since it starts out with an ansatz with only one parameter. For this reason we use a CISD initialization for all tests. VQE and OptOrbVQE use 2-UCCSD as the ansatz. We find that 1-UCCSD is not sufficient and 3-UCCSD offers no significant improvement. The gradient magnitude threshold for AdaptVQE and OptOrbAdaptVQE is set to 10^{-4} . The binding curve of H₄ is shown in Fig. 7.1. We can see that both the orbital-optimized and non-orbital-optimized

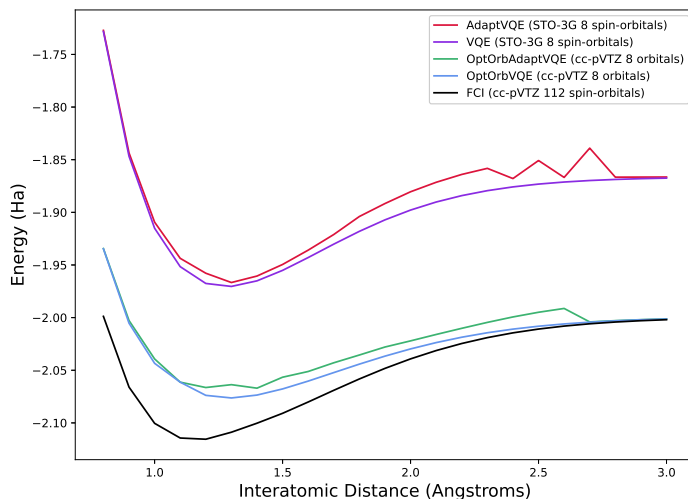


FIGURE 7.1: Binding curve of the ground state of H₄ for uniformly stretched nearest-neighbor interatomic distances using both orbital-optimized and non-orbital-optimized eigensolvers. The exact FCI energy in the cc-pVTZ basis is shown for comparison.

ADAPT-VQE instances converge to local minima for most points in the binding curve, often achieving a less accurate ground state energy than VQE and OptOrbVQE. This effect becomes more significant for moderately stretched bond distances. This is unfortunate as it implies that what information we can gain about the relation between the basis set and the ansatz is somewhat limited. Nonetheless we may still be able to gain some insights while keeping this in the back of our minds. Fig. 7.2 shows the number of Pauli strings in the Hamiltonian for OptOrbVQE as a function of the outer loop iteration. We show

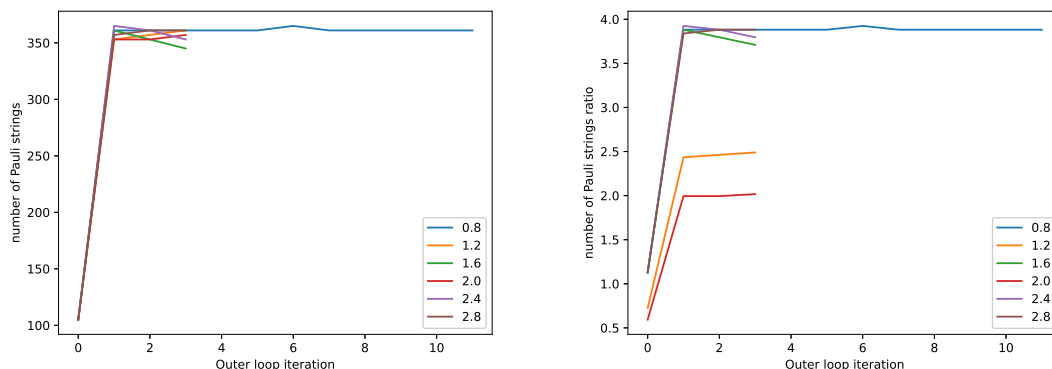


FIGURE 7.2: Measures of the Hamiltonian sparsity as a function of the outer loop iteration for OptOrbVQE (cc-pVTZ, 8 spin-orbitals) for various interatomic distances (in Angstroms). Left: The total number of Pauli strings in the Hamiltonian. Right: The ratio of the number of Pauli strings in OptOrbVQE to that in VQE (STO-3G, 8 spin-orbitals) as a function of the outer loop iteration.

the raw number of Pauli strings as well as the ratio of the number of Pauli strings in OptOrbVQE to the number of Pauli strings in VQE with the fixed STO-3G basis. Recall that the 0th iteration of OptOrbVQE corresponds to taking a large basis and selecting the spin-orbitals with the lowest Hartree-Fock energy. Thus, the 0th iteration represents a naive active space selection where one starts with a large basis and selects a small number of orbitals according to a heuristic that guesses which orbitals are most likely to be occupied in the ground state. All further outer loop iterations represent results obtained by VQE in a basis which has been rotated using orbital optimization. We can see that orbital optimization has the effect that the number of Pauli strings is increased by a multiplicative

factor of roughly 3.5 relative to the naive initial guess. The number of Pauli strings is quickly saturated at the $n = 1$ outer loop iteration and then plateaus and does not change significantly. Furthermore, this number of Pauli strings does not appear to have a strong dependence on the geometry of H_4 . This number of Pauli strings relative to the number for VQE in the STO-3G basis paints a slightly different picture, where we can see that the orbital-optimized basis has anywhere from roughly 2 to 4 times the number of Pauli strings depending on the geometry. It appears that whether one is comparing the orbital-optimized case to the naive initial guess or the non-orbital-optimized STO-3G case, orbital optimization has the effect that it tends towards less sparse Hamiltonians with more Pauli strings in its expansion. This does not, however, elucidate any information regarding the magnitude of the coefficients for this increased number of Pauli strings. In order to study this and to estimate how orbital optimization affects the number of measurements needed to estimate the energy, we plot the measurement overhead factor given in the numerator of Eq. 7.1. These results are shown in Fig. 7.3. We can see that using this measure of

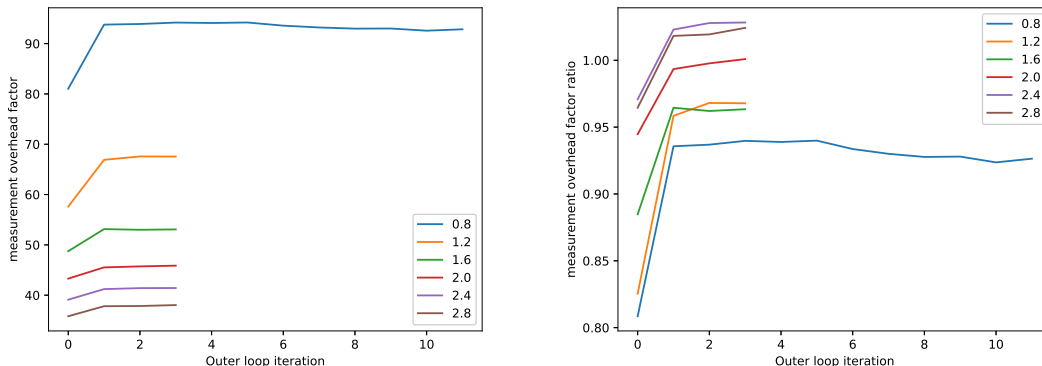


FIGURE 7.3: The measurement overhead factor as a function of the outer loop iteration for OptOrbVQE (cc-pVTZ, 8 spin-orbitals) for various interatomic distances (in Angstroms). Left: The total measurement overhead factor. Right: The ratio of the measurement overhead factor for OptOrbVQE to that for VQE (STO-3G, 8 spin-orbitals).

the Hamiltonian sparsity where their coefficients are taken into account paints a slightly different picture than when we considered the total number of Pauli strings. We can see

that while orbital optimization has some effect on this overhead factor compared to the naive initial guess for the $n = 0$ outer loop iteration, this effect is arguably modest, with the overhead increasing by approximately 34% at an interatomic distance of 1.2 Å and approximately 15% at an interatomic distance of 0.8 Å. The comparison to the fixed STO-3G basis case is even more muted, with orbital optimization requiring roughly 3-5% more measurements in the worse case and approximately 4% fewer measurements in the best case, depending on the geometry. We now consider the effect that orbital optimization has on the circuit depth of the optimal circuit found by ADAPT-VQE. These results are shown in Fig. 7.4. We begin by noting that from these circuit depth measurements, it does appear

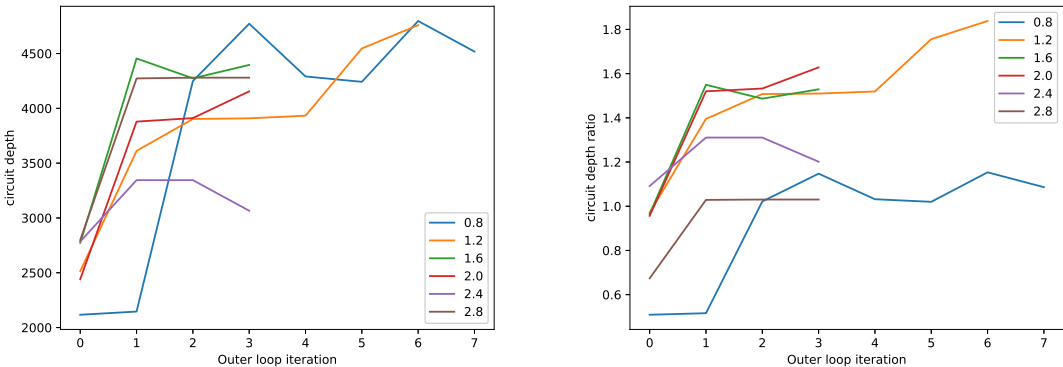


FIGURE 7.4: Measures of the circuit depth for the optimal circuit found by ADAPT-VQE as a function of the outer loop iteration for various interatomic distances (in Angstroms). Left: The total circuit depth as a function of the outer loop iteration. Right: The ratio of the circuit depth found by OptOrbAdaptVQE to that of ADAPT-VQE as a function of the outer loop iteration.

in this case that orbital-optimized basis sets result in increased circuit depth when using ADAPT-VQE as the eigensolver. This is true in many cases for both the comparison to the native $n = 0$ guess and to the fixed STO-3G basis case. For example, we note that the circuit depth for 1.2 Å is roughly 1.9 times that of both the naive guess and STO-3G. For other interatomic distances this difference is much less significant. For example, the circuit depth does not increase significantly beyond that of the naive guess for an interatomic distance of 2.4 Å. Similarly, the circuit depth for 0.8 and 2.8 Å are similar to that of the

STO-3G case. This large variation across geometries is to be expected as ADAPT-VQE is observed to have inconsistent convergence for both the orbital-optimized and fixed-basis cases for H_4 . Nonetheless, we do observe a trend that the circuit depths of the naive guess and fixed STO-3G basis sets appear to be a lower-bound for the circuit depth of the orbital-optimized basis sets. It appears that orbital-optimized basis sets do tend to require more expressive ansatzes than fixed basis sets. For reference, we note that the circuit depth of the 2-UCCSD ansatz used in the non-adaptive solvers is 7785 gates.

7.1.2 LiH

For LiH we consider VQE and ADAPT-VQE in the fixed STO-3G basis (8 spin-orbitals) and orbital-optimized versions of these eigensolvers starting in the cc-pVTZ basis (88 spin-orbitals) and compressed to an active space of 12 spin-orbitals. All methods are initialized in the Hartree-Fock state. The gradient magnitude threshold for AdaptVQE and OptOrbAdaptVQE is set to 10^{-3} . We consider LiH at its near-equilibrium interatomic distance of 1.595 Å. These tests are intended to corroborate the H_4 results, given that AdaptVQE and OptOrbAdaptVQE appeared to struggle to converge for H_4 . In particular, this put a limit on how much confidence we had in the results showing a relation between orbital optimization and an increased ansatz circuit depth. For LiH, we observe that both VQE (using 1-UCCSD) and AdaptVQE converge to an accuracy of roughly 10^{-5} Ha in the STO-3G basis compared to the exact FCI value, increasing our confidence that convergence to local minima is less likely for LiH with these methods. The convergence of OptOrbVQE and AdaptVQE as a function of the outer loop iteration is given in Fig. 7.5. The FCI (cc-pVDZ basis) value is given for reference. Note that the convergence for OptOrbAdaptVQE is much more rapid than OptOrbVQE in this instance. Whereas the adaptive ansatz case converges consistently and rapidly, the fixed ansatz case is often seen to either plateau or converge to a suboptimal value (compared to the adaptive ansatz case) at various iterations. This lends credence to the conjecture behind the motivation for investigating the relation between the basis and the ansatz in the first place: it may be the case that orbital optimization can

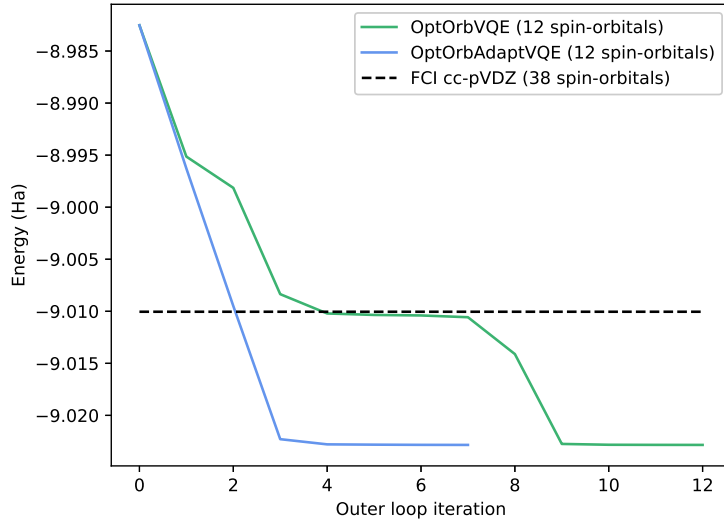


FIGURE 7.5: Convergence of the ground state energy of LiH for OptOrbVQE and OptOrbAdaptVQE as a function of the outer loop iteration.

rotate into a basis such that a particular fixed ansatz either has insufficient expressiveness or induces a cost function landscape that is difficult to navigate. If the basis itself is variable, then an optimal strategy may be to allow the structure of the ansatz to adapt to the changing basis. Note that utilizing an adaptive ansatz does not appear to improve the final accuracy obtained by orbital optimization in this instance, but simply allows for a more consistent and rapid convergence. We further plot the circuit depth of the ansatz found by OptOrbAdaptVQE as a function of the outer loop iteration in Fig. 7.6. These results illustrate that orbital optimization has the effect that the adaptive ansatz subroutine typically finds a deeper ansatz as the orbital optimization procedure progresses. For reference, the circuit depth of 1-UCCSD for this system is 21824 gates. Thus, OptOrbAdaptVQE converges more rapidly despite using a more shallow ansatz than its fixed-ansatz counterpart. Note that there is an alternate plausible explanation for the observed differences in speed of convergence that must be considered. Whereas OptOrbVQE initializes VQE in the optimal ansatz parameters found in the previous outer loop iteration, OptOrbAdaptVQE builds an ansatz from scratch on top of the Hartree-Fock state at every outer loop iteration. Thus,

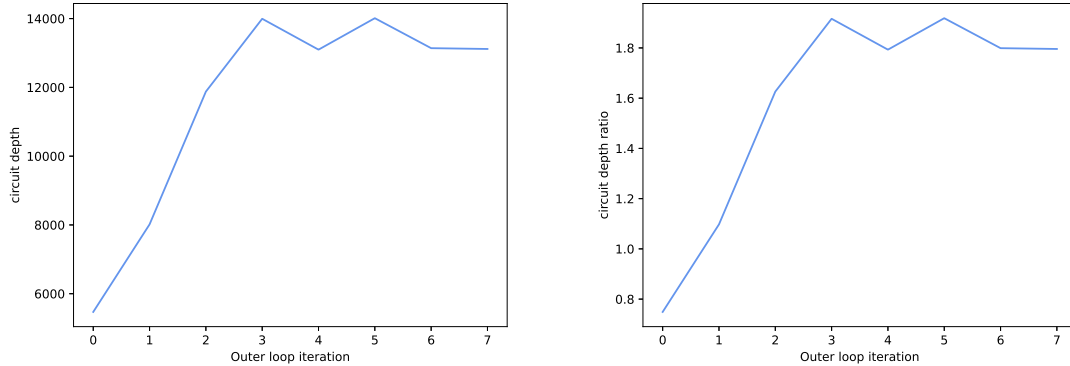


FIGURE 7.6: Left: The circuit depth of the optimal circuit found by OptOrbAdaptVQE at each outer loop iteration. Right: The ratio of the circuit depth of OptOrbAdaptVQE to that found by AdaptVQE in the fixed STO-3G basis as a function of the outer loop iteration.

the initialization strategies are different. It could be that re-using VQE parameters from previous iterations is suboptimal or happens to be close to local minima for some outer loop iterations. In order to increase our confidence that this is not the case, we re-run this comparison, but this time OptOrbVQE will set the VQE ansatz parameters to zero at each outer loop iteration instead of re-using parameters from the previous outer loop iteration. These results are shown in Fig. 7.7 The plateauing effect qualitatively appears to be less severe using this initialization, however convergence is clearly still slower than the adaptive ansatz case.

As we did for H_4 , we will also investigate the dependence of the Hamiltonian sparsity and the measurement overhead factor on orbital optimization. Fig. 7.8 plots the number of Pauli strings in the Hamiltonian as a function of the outer loop iteration for both OptOrbVQE and OptOrbAdaptVQE. We see results similar to what was found for H_4 . The number of Pauli terms starts out low and quickly saturates to a particular number. In this instance the number of Pauli terms in the optimized basis sets is roughly three times that of the fixed basis cases. We note that the number of Pauli terms in the STO-3G basis is the same as the number of Pauli terms in the 0th outer loop iteration: 631. To see how this increase

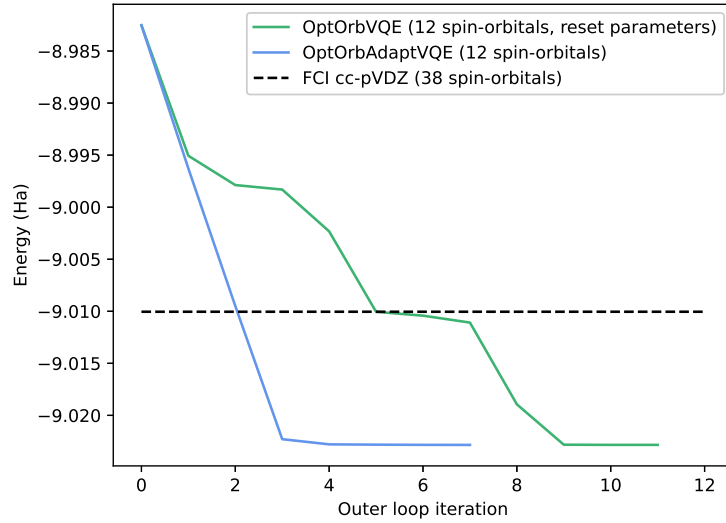


FIGURE 7.7: Convergence of the ground state energy of LiH for OptOrbVQE and OptOrbAdaptVQE as a function of the outer loop iteration. VQE ansatz parameters are reset to zero at all outer loop iterations.

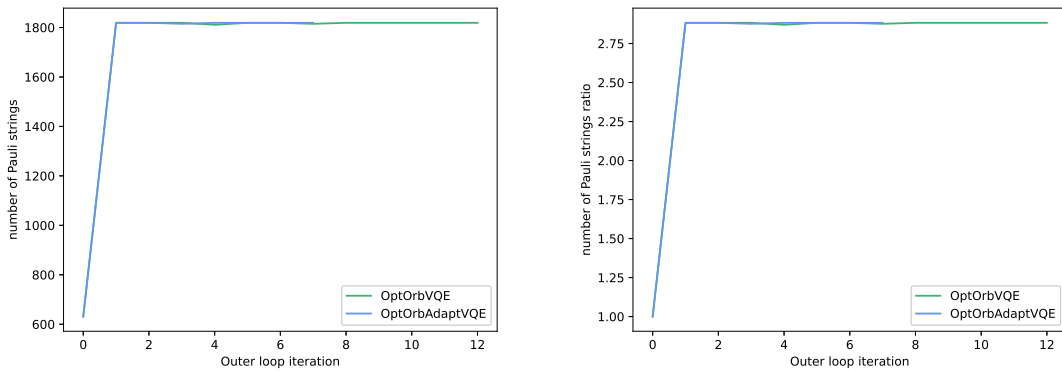


FIGURE 7.8: Left: The number of Pauli strings in the Hamiltonian for OptOrbVQE and OptOrbAdaptVQE for each outer loop iteration. Right: The ratio of the number of Pauli strings in the Hamiltonian for the orbital optimized methods to the number of Pauli strings in the Hamiltonian in the fixed STO-3G basis.

in number of Pauli terms affects the number of measurements required by VQE, we plot the measurement overhead factor in Fig. 7.9. We can see that for LiH, in contrast to where the additional measurement overhead for the orbital optimized case compared to the fixed

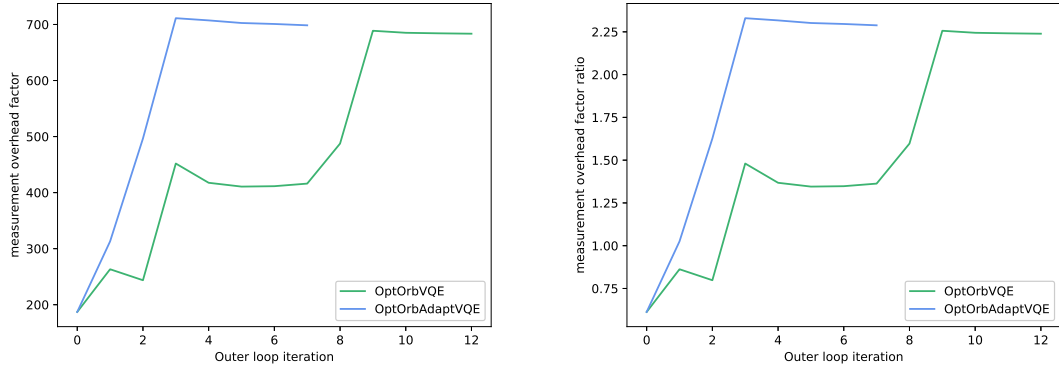


FIGURE 7.9: Left: The measurement overhead factor for OptOrbVQE and OptOrbAdaptVQE for each outer loop iteration. Right: The ratio of measurement overhead factor for the orbital optimized methods to the number of Pauli strings in the Hamiltonian in the fixed STO-3G basis.

STO-3G basis case was modest, here the measurement overhead factor is increased by a factor of roughly 2. This highlights the importance of testing multiple systems. If we had tested only H_4 , we might have been deceived into coming to the conclusion that orbital optimization does not significantly increase the number of measurements required. The LiH results hint that there might be some hidden scaling effect here, rather than some fixed multiplicative overhead, the exact nature of which cannot be elucidated through the two systems considered here. This is a possible effect which should be taken into consideration for future developments in the area of orbital optimized basis sets.

7.2 Discussion and Conclusions

In this chapter we have investigated the relation between orbital optimized basis sets and the sparsity of the Hamiltonian as well as the circuit depth of the ansatz. We have found that for one of our test systems, H_4 , Hamiltonians in orbital-optimized basis sets can contain as many as 4 times as many Pauli strings as the fixed STO-3G basis with the same number of spin-orbitals. LiH demonstrated a similar tendency towards reduced sparsity, with the number of Pauli terms increasing by a factor of roughly 3. This suggests that if one was to carry out phase estimation in an orbital-optimized basis set, the circuit

depth could increase by factors of 3 and 4 for LiH and H_4 , respectively. On the other hand, H_4 did not reveal a corresponding significant increase in the number of required measurements for VQE, implying that a large number of these additional Pauli strings are likely to have small coefficients. The measurement overhead factor did however, increase by a factor of roughly 2 for LiH, indicating that as much as twice as many measurements (compared to a fixed basis) would be required for this system. We have also attempted to draw some relation between orbital optimization and the level of ansatz expressiveness needed to represent the solution. Given the unreliable convergence of AdaptVQE for H_4 , only a weak relation between the orbital rotation iteration and the ansatz circuit depth could be qualitatively observed. The convergence of AdaptVQE is much more reliable for LiH, allowing us to gain more insight from this system. For this system, the circuit depth found by OptOrbAdaptVQE increased monotonically until the $n = 3$ iteration, then roughly stabilized at a depth 1.8 times that of AdaptVQE in the fixed STO-3G basis. We note that this stabilization coincides with a stabilization of the energy at the same outer loop iteration. Thus, when choosing an ansatz for orbital-optimized solvers, one should take into consideration the possibility that an ansatz which was sufficiently expressive for a particular fixed basis may still limit the accuracy of orbital-optimized eigensolvers in one or more of the basis set representations in successive outer loop iterations. One could try using adaptive ansatz strategies such as the one used in ADAPT-VQE to ameliorate this potential issue. Indeed, this did appear to greatly accelerate the convergence for LiH, however we note that the inability of ADAPT-VQE to converge for H_4 is a concern. This shows that while adaptive ansatz strategies are a promising way to approach the circuit depth issue for variational methods, they may also be more prone to instability for some molecular systems. Regarding the Hamiltonian sparsity, one potentially interesting direction of future research would be to try and modify the orbital optimization scheme to optimize the sparsity in some way. The optimal method of doing this is unclear at present, however if one could accomplish this, this would have positive consequences for the practical implementation of phase estimation where the sparsity has a direct impact on the circuit depth involved.

8. Conclusions

In the big picture, there is the question of whether or not there are useful problems for which quantum computers can outperform state of the art classical computers. One promising candidate problem is the electronic structure problem in molecular chemistry which we have focused on in this thesis. In the near-term, demonstrating such an advantage presents a significant challenge as classical computation has the benefit of over a century of hardware advancements, whereas quantum computing hardware is in its infancy. At time of writing, superconducting digital quantum computers have only recently passed the 1,000 physical qubit milestone with IBM'S 1,121 qubit Condor chip. When I started at Duke in 2019, IBM's largest chip was the 27 qubit Falcon chip and Google's largest chip was its 53 qubit Sycamore chip. Over the course of my five year graduate career, superconducting qubit counts have increased by a factor of roughly 20-40 (depending on whether or not you restrict the comparison to chips of the same company or not). While this rapid advancement in hardware is encouraging, physical gate error rates are currently on the order of 10^{-3} , which limits the number of gates that can be coherently carried out to the order of 10^3 or fewer if one assumes that error mitigation can be used when the average number of errors per circuit is on the order of unity or smaller. This is far below the number of gates required to implement phase estimation for molecular systems with size large enough ($\gtrsim 100$ spin-orbitals and dozens of electrons) that demonstrating quantum advantage is plausible. Furthermore, physical qubit counts are at present much lower than what would be required to implement 100 or more logical qubits using the surface code with logical fidelity high enough for the deep circuits involved in phase estimation. This has motivated the development of a class of variational algorithms which attempt to represent the solution state as a shallow (at least compared to phase estimation) parameterized ansatz circuit, offloading the problem of finding the optimal parameters to a classical computer. While these methods do not explicitly bake the requirement of deep circuits into their construction, there is still the practical matter of whether or not sufficiently shallow ansatz circuits can be found where the solution state can be quickly found. We note that the optimal circuits

found by ADAPT-VQE for the 8-qubit, 4 electron H_4 system simulated in Chapter 7 already have total gate depths on the order of thousands of gates. Although not included, the total gate count is likely even higher, putting this past what a quantum computer with an average physical error rate of 10^{-3} could coherently carry out. LiH, a slightly larger system with 4 electrons in 12 spin-orbitals, had gate depths on the order of several thousand gates. Now consider that these are quite simple systems where a modern laptop could calculate the FCI energy (in the minimal STO-3G basis) in a fraction of a second. Clearly, finding ansätze which are sufficiently expressive, do not induce barren plateaus, and are sufficiently shallow will be highly non-trivially. This requirement of the ansatz alone might suggest that variational algorithms are better thought of as early fault-tolerant methods rather than methods which can solve practical problems without error correction of any kind. This would, however, present additional challenges for variational methods. We recall that it was shown by Gonthier et. al.⁷³ that even under the assumption of optimistically shallow ansatz circuits and neglecting the likely overhead associated with error mitigation, a single energy estimation for VQE would take approximately 1.9 days for methane (104 spin-orbitals, 8 electrons) to 71 days for ethanol (260 spin-orbitals, 20 electrons). The cost of encoding these systems using error correction could exacerbate these estimates tremendously. Now we should consider that optimistically, the number of energy estimations needed would be on the order of hundreds (see Fig. 4.7 for instance) and more plausibly on the order of tens of thousands or more (see Fig. 4.2a for instance). It is not difficult to see that practical implementation of variational methods for useful problems will be far-fetched if additional algorithmic advances are not made.

In our work we have made contributions in an attempt to chip away at the obstacles presented for solving the electronic structure problem on quantum computers. In our work on developing the qOMM algorithm, we showed that qOMM can often converge more quickly and is more adept at avoiding local minima than SSVQE, a criteria which is critical in light of the energy evaluation runtime estimates given by Gonthier et. al. The drawback to the qOMM algorithm is that the overlap estimation circuits involved in its construction

are at least an order of magnitude deeper than the ansatz alone. One interesting direction of research would be investigating whether there are more efficient circuits for estimating such overlaps. In our work on developing the OptOrbVQE algorithm, we showed that more accurate ground state estimates can be achieved while simultaneously using fewer qubits compared to larger fixed-basis methods. We further extended this to excited states where we showed analogous results. These are encouraging results as it suggests that the energy estimates of Gonthier et. al could be brought down if one uses optimized basis sets using this method. We should emphasize that these pessimistic runtime estimates are largely due to three main factors: **1.** the $\mathcal{O}(\frac{1}{\epsilon^2})$ scaling of VQE energy estimations, **2.** the worst case scaling of $\mathcal{O}(N^4)$ in the number of Pauli strings in the Hamiltonian with respect to the number of spin-orbitals N , and **3.** the fact that the slow convergence to the infinite basis set limit often necessitates the use of a large N number of spin-orbitals to reach chemical accuracy relative to laboratory results. We note that orbital optimization could ameliorate the third of these issues by accelerating convergence to the infinite basis set limit. The extent to which this accelerated convergence would be possible should be investigated further.

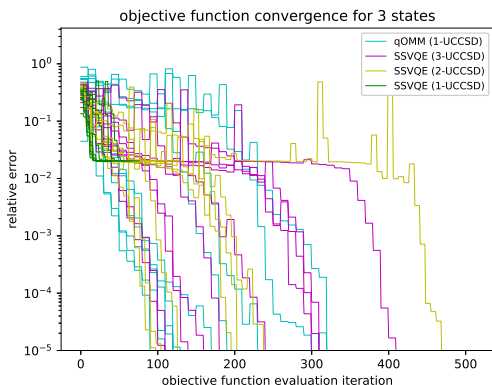
While in these works we considered the integration of orbital optimization into variational methods, one could consider also integrating it into long-term methods such as phase estimation. Naively replacing VQE with QPE while leaving the rest of the method unchanged is conceptually trivial, but not necessarily optimal. Determining the optimal method of including orbital optimization into phase estimation and demonstrating that it consistently improves upon fixed basis methods is not necessarily trivial and could be an interesting direction of research. To see this, consider a naive implementation where we replace VQE with QPE and the rest of the overall method is unchanged. Suppose we have prepared an approximation of the ground state $|\tilde{E}_0\rangle$ that has sufficiently high overlap with the exact ground state $|E_0\rangle$ such that with high probability we measure E_0 to accuracy ϵ . This will require a circuit with depth scaling as $\mathcal{O}(\frac{N^4}{\epsilon})$ for electronic structure Hamiltonians in the worst case. Now consider the fact that the orbital optimization step (in its current

formulation) will require the measurement of $\mathcal{O}(N^4)$ 2-RDM tensor elements. If one was to measure each of these elements using independent measurements, this would require $\mathcal{O}(\frac{N^4}{\epsilon^2})$ circuit preparations of $|E_0\rangle$, putting the total scaling at $\mathcal{O}(\frac{N^8}{\epsilon^3})$. In this naive formulation we would inherit the primary disadvantage of phase estimation (its deep circuits) as well as the primary disadvantage of VQE (the large number of measurements) and deal with the cumulative effects of both. Clearly, we must be more clever if we want phase estimation to benefit from optimized basis sets. Another approach would be to skip the RDM measurements altogether and just minimize the value of E_0 as a function of the rotation \hat{V} , using phase estimation to provide the value of $E_0(\hat{V})$ directly. Note that for a starting basis of M orbitals and an active space of size N orbitals, calculating the gradient of E_0 with respect to \hat{V} would require $2MN$ phase estimation circuit preparations using a finite difference approximation method (or perhaps $MN + 1$ circuit preparations if one uses a biased finite difference approximation). Thus the runtime for each orbital optimization microiteration (in this formulation there would be no outer loop iteration) would scale as $\mathcal{O}(\frac{MN^5}{\epsilon})$, which would be advantageous compared to the first formulation whenever $M < \frac{N^3}{\epsilon^2}$. In this formulation, one would also have to consider two subtleties that could prevent it from converging. The first is that for the approximate prepared state $|\tilde{E}_0\rangle$, there is a nonzero probability that measurement of the ancilla qubit(s) will result in the state being projected onto some eigenstate other than the ground state, introducing some error into the calculation of the orbital optimization gradient. Furthermore, the phase encoding the energy is approximated as a truncated binary fraction, introducing another source of error. It is unclear how much error of these types the orbital optimization procedure can tolerate and still converge reliably. The second subtlety is that the success probability of phase estimation is dependent on the approximate state $|\tilde{E}_0\rangle$ having high overlap with the exact eigenstate $|E_0\rangle$. One would have to develop a state preparation method which can prepare sufficiently accurate approximate eigenstates for not just one basis, but a wide range of basis sets since \hat{V} is variable. The alternative would be to prepare the same $|\tilde{E}_0\rangle$ for all values of \hat{V} and hope that the orbital

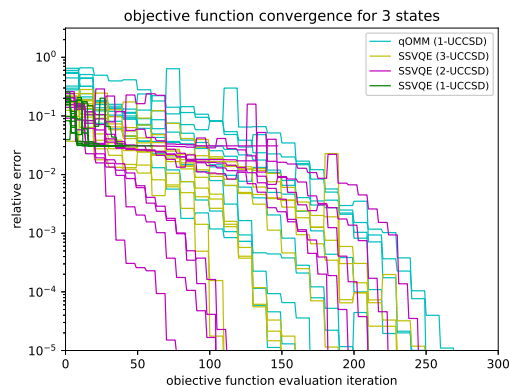
rotation does not spoil the high overlap. Additionally, one could consider the approximate initial state $|\tilde{E}_0\rangle$ to be a function of \hat{V} , $|\tilde{E}_0(\hat{V})\rangle$ and perform orbital optimization as a pre-processing step to phase estimation. In this scenario, one would formulate an efficient method for preparing this approximate state for various basis representations and interleave successive orbital optimization procedures in between in such approximate state preparations, similar to what was done for VQE in our work. This would avoid the cost of having to run phase estimation every time we want to update the basis set. However, it is unclear to what extent performing orbital optimization on an approximation of the solution (rather than the exact solution) would have on the maximum attainable accuracy or the speed of convergence. Evidently, there are a multitude of ways in which one could imagine incorporating orbital optimization into phase estimation, in contrast to the situation for VQE where its application is rather straightforward. It would be worth numerically investigating which of these approaches offers the best accuracy relative to its cost of implementation. Furthermore, we showed in Chapter 7 that orbital optimization tends to have the effect that the sparsity of the Hamiltonian is reduced, resulting in deeper circuits being required for phase estimation. This is a subtlety that should be addressed as well. Incorporating orbital optimization will be an important part of using quantum computers to solve useful problems in quantum chemistry in the long term. Addressing these potential challenges is not necessarily trivial and could be an interesting direction of future research.

Appendix A. Additional qOMM and SSVQE Convergence Plots

This appendix section contains work published in the Journal of Chemical Theory and Computation on our work on the qOMM algorithm.¹¹⁴ Here we provide all ten runs for each of the randomly initialized tests for which only one run was shown in Section 4.2. We also provide some additional tests such as H_2 and the hydrogen square model at stretched bond distances and up to 7 states of LiH. The success rates for the noise-free simulations are summarized in Table 4.1 in Section 4.2. Figure A.1 depicts all ten randomly initialized runs for the H_2 results discussed in Section 4.2.1. Figure A.2 depicts all of the ten randomly initialized runs for the noisy H_2 results discussed in Section 4.2.2. Figure A.3 depicts all of the ten randomly initialized runs for the hydrogen square toy model at equilibrium bond distance discussed in Section 4.2.1. Figure A.4 further depicts ten randomly initialized runs for this same system at a stretched bond distance. Figure A.5 is an extension of the LiH results discussed in Section 4.2.3, where all ten randomly-initialized runs for the results given in that section are depicted here as well as randomly-initialized results for up to seven states.

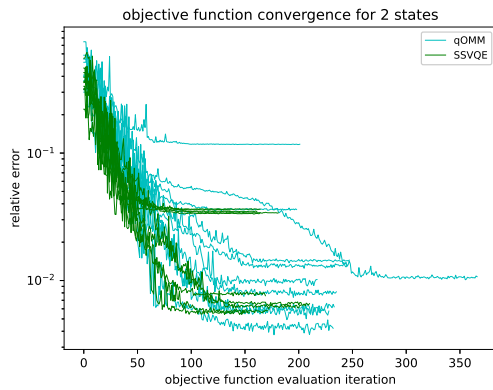


(a) 0.735 Å bond distance

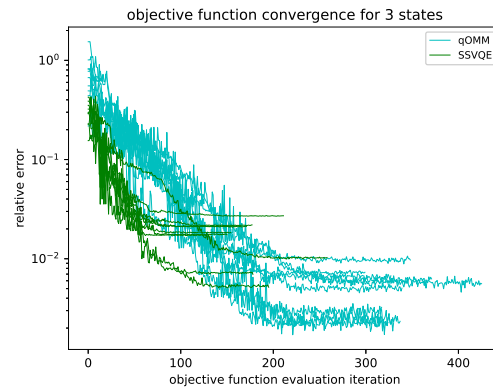


(b) 1.47 Å bond distance

FIGURE A.1: Convergence of the relative error $\frac{|f_i - f_{exact}|}{|f_{exact}|}$ of qOMM and SSVQE for H_2 using a random parameter initialization.

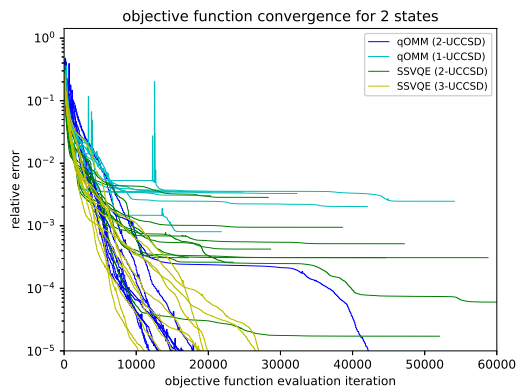


(a) Two states

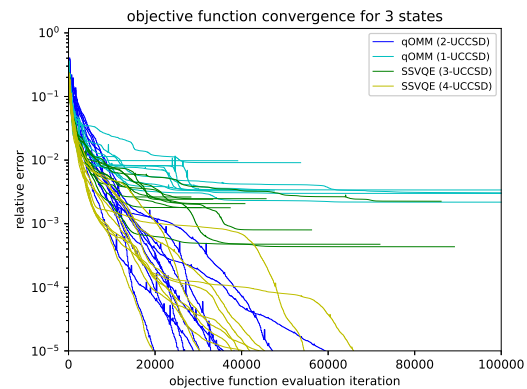


(b) Three states

FIGURE A.2: Convergence of the relative error $\frac{|f_i - f_{exact}|}{|f_{exact}|}$ of qOMM and SSVQE for H_2 at an interatomic distance of 0.735 \AA , where each circuit gate is modeled as having a probability of local depolarizing error $p_{error} = 0.001$.

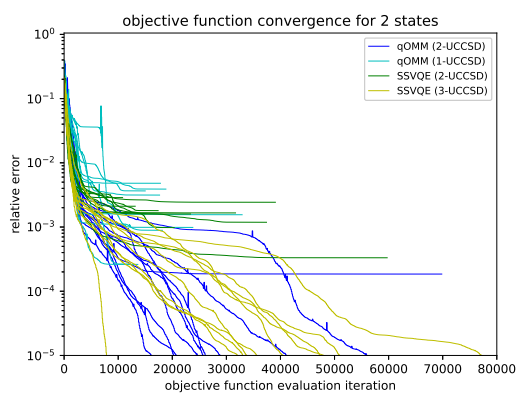


(a) Two states

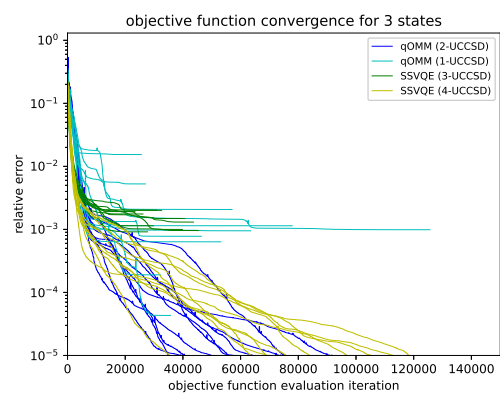


(b) Three states

FIGURE A.3: Convergence of the relative error $\frac{|f_i - f_{exact}|}{|f_{exact}|}$ of qOMM and SSVQE for the hydrogen square model at an interatomic distance of 1.23 \AA using a random parameter initialization.

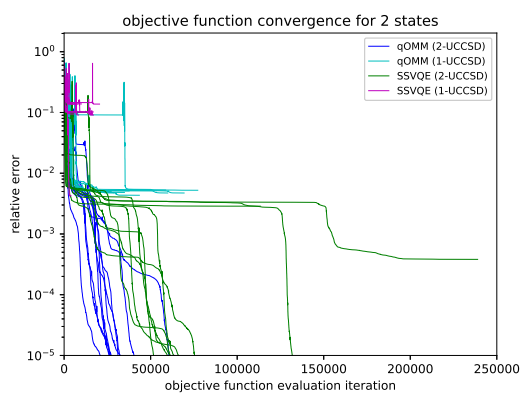


(a) Two states

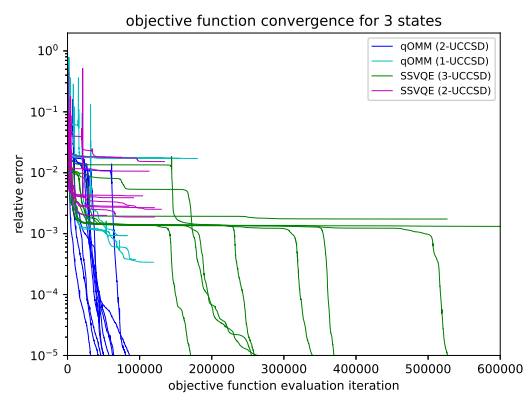


(b) Three states

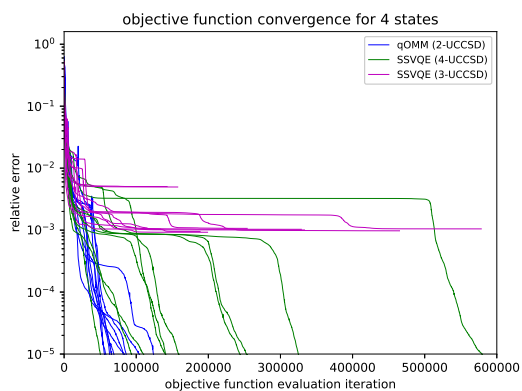
FIGURE A.4: Convergence of the relative error $\frac{|f_i - f_{exact}|}{|f_{exact}|}$ of qOMM and SSVQE for the hydrogen square model at an interatomic distance of 2.46 Å using a random parameter initialization.



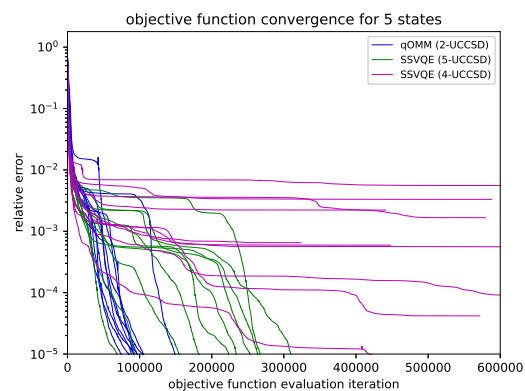
(a) Two states



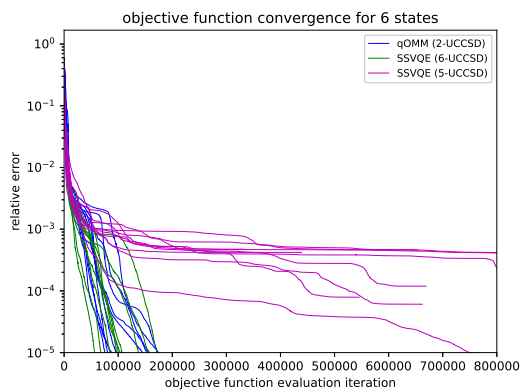
(b) Three states



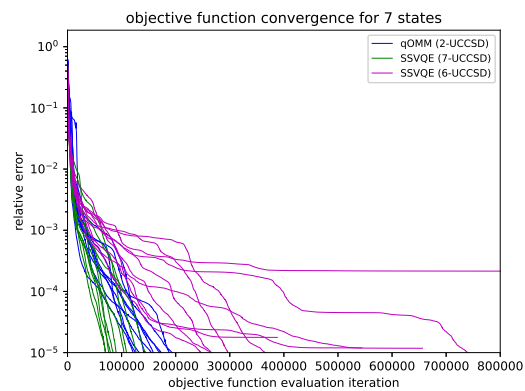
(c) Four states



(d) Five states



(e) Six states



(f) Seven states

FIGURE A.5: Convergence of the relative error of qOMM and SSVQE for LiH using randomly initialized parameters.

Appendix B. Excited States Initializations and Ansatz Expressiveness

This appendix contains material that we have published in the Journal of Chemical Theory and Computation on our work on excited states orbital optimization.¹²⁰ Here we test the effects of various initialization choices and levels of ansatz expressiveness on the convergence of MCVQE and SSVQE in a fixed minimal basis. These tests serve to illustrate our motivation for our particular choices in the orbital optimized tests in Section 6.4 of the main text. By “initialization” we mean the choice of non-parameterized subcircuit prepended to the parameterized ansatz. The ansatz parameters themselves are initialized to zero as this corresponds to the identity subcircuit. Thus, this allows us to explore various chemically motivated initializations. MCVQE is tested with configuration interaction singles (CIS) and configuration interaction singles and doubles (CISD) state initializations. SSVQE is tested with CIS and CISD as well as an “excited Hartree-Fock” initialization used in our work on qOMM.¹¹⁴ This initialization applies single-particle fermionic excitations to the Hartree-Fock state and chooses the resulting Slater determinants with the lowest energy to initialize the circuit. Such states are orthogonal and can thus be used with both MCVQE and SSVQE. The ansatz expressiveness is varied by varying the number of times the base UCCSD circuit pattern is repeated, where we denote the circuit consisting of n UCCSD repetitions as n -UCCSD.

Table B.1 shows the final state-averaged energy accuracy for the first three states of H_4 at a nearest neighbor distance of 1.23 Å for various choices of eigensolver, state initialization, and UCCSD expressiveness. We can see that Hartree-Fock and CIS initializations fail to produce an accuracy greater than 10^{-2} Hartree for any eigensolver or level of ansatz expressiveness. Furthermore, increasing the ansatz expressiveness offers no meaningful improvement for these initializations. On the other hand, the CISD initialization offers the ability to achieve greater than chemical accuracy. With 2-UCCSD, both eigensolvers fall just short of chemical accuracy, but increasing the ansatz to 3 and 4-UCCSD offers further

improvements. Thus, we can see that there is motivation for developing circuits which correspond to CISD states. We now compare the speed of convergence between MCVQE

Table B.1: The final accuracy of the average energy for H_4 for given choices of eigensolver, initialization, and UCCSD ansatz expressiveness.

Eigensolver	Initialization	UCCSD Repetitions			
		1-rep	2-rep	3-rep	4-rep
MCVQE	CIS	4.25×10^{-2}	3.85×10^{-2}	3.70×10^{-2}	3.70×10^{-2}
MCVQE	CISD	5.16×10^{-3}	2.16×10^{-3}	2.53×10^{-4}	3.73×10^{-8}
SSVQE	HF	5.78×10^{-2}	4.67×10^{-2}	3.70×10^{-2}	3.70×10^{-2}
SSVQE	CIS	4.26×10^{-2}	3.81×10^{-2}	3.70×10^{-2}	3.70×10^{-2}
SSVQE	CISD	5.41×10^{-3}	2.45×10^{-3}	1.56×10^{-4}	3.67×10^{-8}

and SSVQE for the four test instances in Fig. B.1 which were able to surpass chemical accuracy. Fig. B.1 plots the state-averaged energy accuracy as a function of the number of objective function evaluations. We can see that for all four instances, the state-averaged energy plateaus for many iterations before escaping and converging to (or closer to) the global minimum. This is consistent with our previous studies on excited states solvers.¹¹⁴ Notably, MCVQE is less prone to this issue.

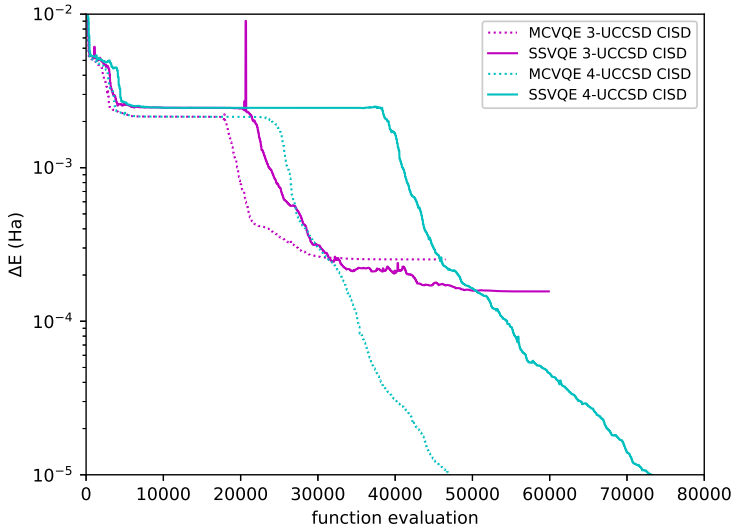


FIGURE B.1: Convergence of the state-averaged energy accuracy (ΔE).

Appendix C. CIS State Preparation

This appendix contains material that we have published in the Journal of Chemical Theory and Computation on our work on excited states orbital optimization.¹²⁰ Here we give an example of how configuration interaction singles (CIS) states can be prepared as a quantum circuit on a quantum computer using Givens rotations. It was proven that Givens rotations form a universal set of gates for chemically-motivated statevectors.⁵⁹ The authors accomplish this constructively by giving a procedure for preparing an arbitrary state using Givens rotations controlled on the states of multiple qubits. They comment that for particular classes of states the resources involved may be reduced by controlling the rotation only on certain qubit subsets. What remains to be done is to work out the details of how to apply this idea to specific classes of CI statevectors (CIS, CISD, CISDT, ect...) in a way that is as gate-efficient as possible. Here we give an example of how both dense and sparse CIS statevectors can be mapped to quantum circuits using Givens rotations.

We briefly note that the CIS state preparation circuit outlined in the MCVQE proposal paper¹¹² assumes a particular encoding where the reference state from which electrons are being excited is encoded as the "all-zero" state $|00\dots 0\rangle$ where the qubit registers encode the occupation number of orbitals unoccupied in the reference state, but not those occupied in the reference state. Thus, the singly-excited wavefunction components contain no information about the particular Hartree-Fock occupied orbital from which the electron was excited. Here we seek a CIS state preparation circuit in the Jordan-Wigner encoding where the reference state is the Hartree-Fock state and the occupation number of orbitals occupied in this state are included for all wavefunction components. Thus, each singly-excited wavefunction component does contain information about the occupied Hartree-Fock orbital from which the electron was excited.

The matrix representation of a Givens rotation involving qubits n and m with angle θ

is given by:⁵⁹

$$G_{nm}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta & 0 \\ 0 & \sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{C.1})$$

where the basis ordering is: $|0\rangle_m|0\rangle_n$, $|0\rangle_m|1\rangle_n$, $|1\rangle_m|0\rangle_n$, $|1\rangle_m|1\rangle_n$. or notational convenience, we will often omit the subscript n and m labels on qubit registers. We can also make use of Givens rotations controlled on the state of a target qubit t , which we denote by $C_t G_{nm}(\theta)$. This gate can be represented as:

$$C_t G_{nm}(\theta) = |0\rangle\langle 0|_t \otimes \hat{I}_{nm} \quad (\text{C.2})$$

$$+ |1\rangle\langle 1|_t \otimes G_{nm}(\theta). \quad (\text{C.3})$$

We also note that we adopt the convention of Qiskit where in the Jordan-Wigner encoding, the qubits are ordered according to spin and Hartree-Fock energy. Orbitals with the same spin are ordered from right to left in ascending Hartree-Fock energy. Thus, for our purposes the relevant action of a Givens rotation is:

$$G_{nm}(\theta)|01\rangle = \cos(\theta)|01\rangle + \sin(\theta)|10\rangle \quad (\text{C.4})$$

We do not have to consider the action of Givens rotations on the state $|10\rangle$ as we are only interested in exciting particles to orbitals of higher energies from lower ones. The circuit notation for the single-excitation Givens rotation is given by:⁵⁹

$$\begin{array}{c} |0\rangle \text{---} \boxed{G} \text{---} \\ | \\ |1\rangle \text{---} \boxed{G} \text{---} \end{array} \quad (\text{C.5})$$

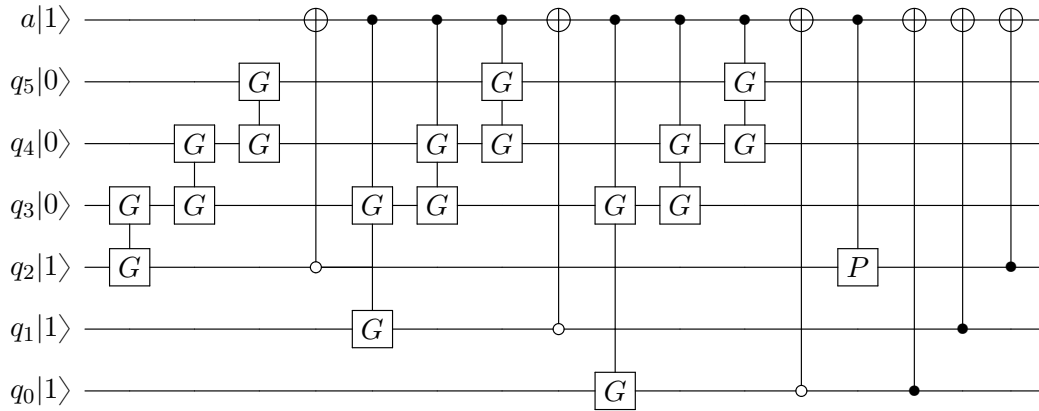
We want to construct a circuit from Givens rotations which produces the state:

$$|CIS\rangle = C_{HF}|HF\rangle + \sum_{\substack{p \in \mathcal{O}_{HF} \\ q \in \mathcal{U}_{HF}}} C_p^q |\phi_q \leftarrow \phi_p\rangle \quad (\text{C.6})$$

where $|\phi_q \leftarrow \phi_p\rangle$ means the computational basis state produced by exciting an electron from orbital ϕ_p to orbital ϕ_q from the Hartree-Fock ground state. \mathcal{O}_{HF} and \mathcal{U}_{HF} denote the set of orbitals occupied and unoccupied in the Hartree-Fock state, respectively. We can solve for the coefficients C_p^q classically then set them equal to the parameterized coefficients of the wavefunction expansion produced by a circuit comprised of Givens rotations. This produces a set of equations which can be solved to find the Givens angles which produce the circuit that prepares arbitrary CIS states.

C.0.1 Example 1: 3 particles, 6 spin-orbitals

We now give an example for the particular case where we want to generate the CIS wavefunction with 6 spin-orbitals and 3 particles, where all possible single-particle excitations are considered. The circuit for accomplishing this is given by:



where the register labelled as a is an ancilla qubit and those labelled as q_i are data qubits used to store the CIS state. CNOT gates with an open dot instead of the typical filled dot denote a CNOT gate where the NOT operation is controlled on the target qubit being in the state $|0\rangle$ instead of $|1\rangle$. Although it is not explicitly given in the circuit due to space constraints, each Givens rotation has its own parameter. The controlled phase gate P (implicitly $P(\lambda)$) is given in matrix form by:

$$P(\lambda) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\lambda} \end{pmatrix}$$

where the columns and rows are ordered as: $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. We will see later that we only need λ to be 0 or π . $\lambda = 0$ corresponds to a 2-qubit identity gate, in which case we could omit this gate entirely, whereas $\lambda = \pi$ corresponds to a controlled-Z gate. We denote this gate as P in order to keep full generality. The purpose of the final sequence of CNOT gates is to disentangle the ancilla qubit from the data qubits, putting the final state in the form $|CIS\rangle \otimes |0\rangle$. The final state of the data qubits is given by:

$$\begin{aligned}
|CIS\rangle = & e^{i\lambda} \cos \theta_2^3 \cos \theta_1^3 \cos \theta_0^3 |000111\rangle \\
& + \cos \theta_2^3 \cos \theta_1^3 \sin \theta_0^3 \cos \theta_0^4 |001110\rangle \\
& + \cos \theta_2^3 \cos \theta_1^3 \sin \theta_0^3 \sin \theta_0^4 \cos \theta_0^5 |010110\rangle \\
& + \cos \theta_2^3 \cos \theta_1^3 \sin \theta_0^3 \sin \theta_0^4 \sin \theta_0^5 |100110\rangle \\
& + \cos \theta_2^3 \sin \theta_1^3 \cos \theta_1^4 |001101\rangle \\
& + \cos \theta_2^3 \sin \theta_1^3 \sin \theta_1^4 \cos \theta_1^5 |010101\rangle \\
& + \cos \theta_2^3 \sin \theta_1^3 \sin \theta_1^4 \sin \theta_1^5 |100101\rangle \\
& + \sin \theta_2^3 \cos \theta_2^4 |001011\rangle \\
& + \sin \theta_2^3 \sin \theta_2^4 \cos \theta_2^5 |010011\rangle \\
& + \sin \theta_2^3 \sin \theta_2^4 \sin \theta_2^5 |100011\rangle.
\end{aligned} \tag{C.7}$$

We denote the angle which first adds the component $|\phi_q \leftarrow \phi_p\rangle$ to the overall wavefunction as θ_p^q . By setting these coefficients equal to those of the CI wavefunction expansion given in Eq. C.6, we arrive at the following set of equations:

$$\begin{aligned}
e^{i\lambda} \cos \theta_2^3 \cos \theta_1^3 \cos \theta_0^3 &= C_{HF} \\
\cos \theta_2^3 \cos \theta_1^3 \sin \theta_0^3 \cos \theta_0^4 &= C_0^3 \\
\cos \theta_2^3 \cos \theta_1^3 \sin \theta_0^3 \sin \theta_0^4 \cos \theta_0^5 &= C_0^4 \\
\cos \theta_2^3 \cos \theta_1^3 \sin \theta_0^3 \sin \theta_0^4 \sin \theta_0^5 &= C_0^5 \\
\\
\cos \theta_2^3 \sin \theta_1^3 \cos \theta_1^4 &= C_1^3 \\
\cos \theta_2^3 \sin \theta_1^3 \sin \theta_1^4 \cos \theta_1^5 &= C_1^4 \\
\cos \theta_2^3 \sin \theta_1^3 \sin \theta_1^4 \sin \theta_1^5 &= C_1^5 \\
\\
\sin \theta_2^3 \cos \theta_2^4 &= C_2^3 \\
\sin \theta_2^3 \sin \theta_2^4 \cos \theta_2^5 &= C_2^4 \\
\sin \theta_2^3 \sin \theta_2^4 \sin \theta_2^5 &= C_2^5.
\end{aligned} \tag{C.8}$$

The recursive structure of this circuit allows us to solve for all of these parameters analytically in a recursive way. We can partition these 10 equations into 3 blocks of 3 equations and one block with one equation according to the occupied Hartree-Fock orbital from which the excitations are generated. We start with $p = 2$ and solve for θ_2^5 , θ_2^4 , and θ_2^3 in that order. This has the solution:

$$\begin{aligned}
\theta_2^5 &= \arctan \left(\frac{C_2^5}{C_2^4} \right) \\
\theta_2^4 &= \arctan \left(\frac{1}{\cos \theta_2^5} \frac{C_2^4}{C_2^3} \right) \\
\theta_2^3 &= \arcsin \left(\frac{C_2^3}{\cos \theta_2^4} \right).
\end{aligned} \tag{C.9}$$

The equations corresponding to $p = 1$ are the same in structure to those of $p = 2$, except that the left hand side is multiplied by a constant factor of $\cos \theta_2^3$, a quantity that we solved

for in the $p = 2$ equations. We define $\alpha_2 = \cos \theta_2^3$ and divide both sides of these equations by α_2 . We arrive at a second set of solutions:

$$\begin{aligned}\theta_1^5 &= \arctan\left(\frac{C_1^5}{C_1^4}\right) \\ \theta_1^4 &= \arctan\left(\frac{1}{\cos(\theta_1^5)} \frac{C_1^4}{C_1^3}\right) \\ \theta_1^3 &= \arcsin\left(\frac{1}{\alpha_2} \frac{C_1^3}{\cos(\theta_1^4)}\right).\end{aligned}\tag{C.10}$$

The $p = 0$ block of equations also has the same form, but the left side is multiplied by a factor of $\alpha_1\alpha_2 = \cos \theta_1^3 \cos \theta_2^3$. We divide by sides of each equation in this block by $\alpha_1\alpha_2$ and arrive at the solution for this third block:

$$\begin{aligned}\theta_0^5 &= \arctan\left(\frac{C_0^5}{C_0^4}\right) \\ \theta_0^4 &= \arctan\left(\frac{1}{\cos \theta_0^5} \frac{C_0^4}{C_0^3}\right) \\ \theta_0^3 &= \arcsin\left(\frac{1}{\alpha_1\alpha_2} \frac{C_0^3}{\cos \theta_0^4}\right).\end{aligned}\tag{C.11}$$

This leaves only the parameter λ for which to solve. The magnitude of C_{HF} will match that of $\alpha_2\alpha_1\alpha_0$ due to the normalization condition, but the two may differ by a factor of either $+1$ or -1 . The parameter λ will determine this phase. If the phase of the two quantities match, then $\lambda = 0$ and the phase gate can be omitted entirely. If the two differ by a phase of -1 , then $\lambda = \pi$.

C.0.2 Example 2: Sparse 2 Particles, 6 Spin-Orbitals

The previous example dealt with the particular case of 3 particles and 6 spin-orbitals where every single-particle excitation from any occupied Hartree-Fock orbital is possible. We now give an example for a different number of particles and spin-orbitals for the case where the CIS wavefunction is sparse and some of the coefficients are zero. This demonstrates that we can also generate approximate CIS wavefunctions at lower circuit depth in

We solve for θ_1^5 , θ_1^2 , θ_0^4 , and λ recursively in that order. The solution is given by:

$$\begin{aligned}
\theta_1^5 &= \arctan\left(\frac{C_1^5}{C_1^2}\right) \\
\theta_1^2 &= \arcsin\left(\frac{C_1^2}{\cos\theta_1^5}\right) \\
\theta_0^4 &= \arcsin\left(\frac{C_0^4}{\alpha_1}\right) \\
e^{i\lambda} &= \frac{C_{HF}}{\alpha_1\alpha_0}
\end{aligned} \tag{C.14}$$

C.0.3 General Procedure

Based on the particular examples given, we can observe a general procedure for any number of particles and spin-orbitals. We first partition the spin-orbitals into two sets \mathcal{O}_{HF} and \mathcal{U}_{HF} , the set of spin-orbitals occupied and unoccupied in the Hartree-Fock reference state, respectively. For each $\phi_p \in \mathcal{O}_{HF}$, we generate an ordered set L_p of orbitals $\phi_q \in \mathcal{U}_{HF}$ for which the CI amplitude C_p^q is not zero or is not below a desired truncation threshold. These orbitals are ordered in ascending Hartree-Fock energy. For every spin-orbital in each L_p , we map the orbital indices q to new indices n_p^q . This is simply so that we may write down a general analytical expression for the gate sequence which reflects the fact we may not want or need the full, dense CI wavefunction. n_p^q is the index of the list L_p which was mapped from the original index of the spin-orbital ϕ_q . *e.g.* the original set of unoccupied orbitals from which a particular occupied orbital may be given by $\{\phi_3, \phi_6, \phi_8\}$, but we map this ordered set to the list indices $\{0, 1, 2\}$.

The general sequence of gates is given by:

$$\begin{aligned}
|CIS\rangle = & \left[\prod_{\phi_p \in \mathcal{O}_{HF}} CNOT(a, p) \right] C_a P(\lambda)_0 \\
& \times \prod_{\phi_p \in \mathcal{O}_{HF} \setminus \{\phi_0\}}^{q_{max}-1} \left[X_p CNOT(a, p) X_p \prod_{\phi_q \in L_p} C_a G_{n_p^q, n_p^q+1} \right] \\
& \times X_0 CNOT(a, 0) X_0 \prod_{\phi_q \in L_0}^{q_{max}-1} G_{n_0^q, n_0^q+1}(\theta_0^q) |HF\rangle |1\rangle_a
\end{aligned} \tag{C.15}$$

The rightmost terms denote the fact that for the set of excitations from the first occupied spin-orbital, we do not have to apply the Givens rotations conditioned on the state of the ancilla qubit. Without loss of generality, we may take ϕ_0 to be the orbital which has the longest list L_p of possible excited orbitals. This will reduce the circuit depth as compiling controlled Givens rotations into a sequence of 1 and 2-qubit basis gates will in general be more expensive than regular Givens rotations. After this, we apply a NOT gate to the ancilla qubit conditioned on the state of the qubit from which we just generated excitations being $|0\rangle$. This marks all data qubit wavefunction components that are not the Hartree-Fock component so that future Givens rotations will not apply excitations to these components. We then repeat this with Givens rotations controlled on the state of the ancilla for all the other Hartree-Fock occupied orbitals. If there is an orbital for which there are no possible excitations, we simply skip it. We then apply a phase gate $P(\lambda)$ to any of the Hartree-Fock occupied orbitals conditioned on the state of the ancilla. This applies the relative phase $e^{i\lambda}$ to the Hartree-Fock component of the wavefunction. Finally, for each of the Hartree-Fock occupied orbitals, we apply a NOT gate controlled on the state of the ancilla. This disentangles the data qubits from the ancilla qubit so that the final result is a product state of these registers.

Finally, we give a general procedure for mapping the CI coefficients C_p^q to the Givens rotation angles θ_p^q . In order to do this, we temporarily re-index the CI coefficient indices in the same way that we did for the general circuit expression. For each L_p , we map the

orbital indices $q \rightarrow n_p^q$. Here, $C_p^{n_p^q}$ is the re-indexed CI coefficient mapped from C_p^q for list L_p . The sequence of steps for this procedure can be given by:

1. For each $\phi_p \in \mathcal{O}_{HF}$ (In the corresponding order applied in the circuit):

(a) If $\text{length}(L_p) = 1$:

$$\theta_p^{n_p^q} = \arcsin \left(\frac{C_p^{n_p^q}}{\prod_{p' < p} \alpha_{p'}} \right)$$

(b) If $\text{length}(L_p) = 2$:

$$\theta_p^{n_p^q=1} = \arctan \left(\frac{C_p^{n_p^q=1}}{C_p^{n_p^q=0}} \right)$$

$$\theta_p^{n_p^q=0} = \arcsin \left(\frac{1}{\prod_{p' < p} \alpha_{p'}} \frac{C_p^{n_p^q=0}}{\cos(\theta_p^{n_p^q=0})} \right)$$

(c) If $\text{length}(L_p) > 2$:

$$\theta_p^{n_p^{q_{max}}} = \arctan \left(\frac{C_p^{n_p^{q_{max}}}}{C_p^{n_p^{q_{max}}-1}} \right)$$

$$\theta_p^{n_p^{q_{max}}-1} = \arctan \left(\frac{1}{\cos(\theta_p^{n_p^{q_{max}}})} \frac{C_p^{n_p^{q_{max}}-1}}{C_p^{n_p^{q_{max}}-2}} \right)$$

\vdots

$$\theta_p^{n_p^q} = \arctan \left(\frac{1}{\cos(\theta_p^{n_p^q+1})} \frac{C_p^{n_p^q}}{C_p^{n_p^q-1}} \right)$$

\vdots

$$\theta_p^{n_p^q=0} = \arcsin \left(\frac{1}{\prod_{p' < p} \alpha_{p'}} \frac{C_p^{n_p^q=0}}{\cos(\theta_p^{n_p^q=0})} \right)$$

2. Solve for λ :

$$\lambda = \begin{cases} 0, & \text{if } \frac{C_{HF}}{\prod_p \alpha_p} = 1 \\ \pi, & \text{if } \frac{C_{HF}}{\prod_p \alpha_p} = -1 \end{cases}$$

Here, for the sake of notational convenience we define $\prod_{p' < 0} \alpha_{p'} = 1$.

Bibliography

- (1) Helgaker, T.; Jorgensen, P.; Olsen, J., *Molecular Electronic-Structure Theory*; Wiley: 2014.
- (2) Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Courier Corporation: 2012.
- (3) de Carvalho, C. C. C. R.; da Fonseca, M. M. R. Carvone: Why and How Should One Bother to Produce This Terpene. *Food Chemistry* **2006**, *95*, 413–422.
- (4) Mortimer, R. G., *Physical Chemistry*; Academic Press: 2000.
- (5) 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 Science and Technology* **2018**, *4*, 014008.
- (6) Evans, R. C.; Douglas, P.; Burrow, H. D., *Applied Photochemistry*; Springer Science & Business Media: 2014.
- (7) Mai, S.; González, L. Molecular Photochemistry: Recent Developments in Theory. *Angewandte Chemie International Edition* **2020**, *59*, 16832–16846.
- (8) Aspuru-Guzik, A.; Dutoi, A. D.; Love, P. J.; Head-Gordon, M. Simulated Quantum Computation of Molecular Energies. *Science (New York, N.Y.)* **2005**, *309*, 1704–1707.
- (9) 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. *Nature Communications* **2014**, *5*, 4213.
- (10) Purwanto, W.; Zhang, S.; Krakauer, H. An Auxiliary-Field Quantum Monte Carlo Study of the Chromium Dimer. *The Journal of Chemical Physics* **2015**, *142*, 064302.
- (11) Reiher, M.; Wiebe, N.; Svore, K. M.; Wecker, D.; Troyer, M. Elucidating Reaction Mechanisms on Quantum Computers. *Proceedings of the National Academy of Sciences* **2017**, *114*, 7555–7560.
- (12) Hoffman, B. M.; Lukoyanov, D.; Yang, Z.-Y.; Dean, D. R.; Seefeldt, L. C. Mechanism of Nitrogen Fixation by Nitrogenase: The Next Stage. *Chemical Reviews* **2014**, *114*, 4041–4062.
- (13) Elfving, V. E.; Broer, B. W.; Webber, M.; Gavartin, J.; Halls, M. D.; Lorton, K. P.; Bochevarov, A. How Will Quantum Computers Provide an Industrially Relevant Computational Advantage in Quantum Chemistry?, 2020.

- (14) Wecker, D.; Hastings, M. B.; Troyer, M. Progress towards Practical Quantum Variational Algorithms. *Physical Review A* **2015**, *92*, 042303.
- (15) Wecker, D.; Bauer, B.; Clark, B. K.; Hastings, M. B.; Troyer, M. Gate-Count Estimates for Performing Quantum Chemistry on Small Quantum Computers. *Physical Review A* **2014**, *90*, 022305.
- (16) von Burg, V.; Low, G. H.; Häner, T.; Steiger, D. S.; Reiher, M.; Roetteler, M.; Troyer, M. Quantum Computing Enhanced Computational Catalysis. *Physical Review Research* **2021**, *3*, 033055.
- (17) Coleman, P., *Introduction to Many-Body Physics*; Cambridge University Press: 2015.
- (18) Kottmann, J. S.; Schleich, P.; Tamayo-Mendoza, T.; Aspuru-Guzik, A. Reducing Qubit Requirements While Maintaining Numerical Precision for the Variational Quantum Eigensolver: A Basis-Set-Free Approach. *The Journal of Physical Chemistry Letters* **2021**, *12*, 663–673.
- (19) Pietro, W. J.; Hehre, W. J. Molecular Orbital Theory of the Properties of Inorganic and Organometallic Compounds. 3. STO-3G Basis Sets for First- and Second-Row Transition Metals. *Journal of Computational Chemistry* **1983**, *4*, 241–251.
- (20) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent Molecular Orbital Methods 25. Supplementary Functions for Gaussian Basis Sets. *The Journal of Chemical Physics* **1984**, *80*, 3265–3269.
- (21) Peterson, K. A.; Dunning Jr., T. H. Accurate Correlation Consistent Basis Sets for Molecular Core–Valence Correlation Effects: The Second Row Atoms Al–Ar, and the First Row Atoms B–Ne Revisited. *The Journal of Chemical Physics* **2002**, *117*, 10548–10560.
- (22) Cohen-Tannoudji, C.; Diu, B.; Laloë, F., *Quantum Mechanics*; Wiley: 1977.
- (23) Bartlett, R. J.; Musiał, M. Coupled-Cluster Theory in Quantum Chemistry. *Reviews of Modern Physics* **2007**, *79*, 291–352.
- (24) Siegbahn, P. E. M.; Almlöf, J.; Heiberg, A.; Roos, B. O. The Complete Active Space SCF (CASSCF) Method in a Newton–Raphson Formulation with Application to the HNO Molecule. *The Journal of Chemical Physics* **1981**, *74*, 2384–2396.
- (25) Siegbahn, P.; Heiberg, A.; Roos, B.; Levy, B. A Comparison of the Super-CI and the Newton-Raphson Scheme in the Complete Active Space SCF Method. *Physica Scripta* **1980**, *21*, 323.

- (26) Sun, Q.; Yang, J.; Chan, G. K.-L. A General Second Order Complete Active Space Self-Consistent-Field Solver for Large-Scale Systems. *Chemical Physics Letters* **2017**, *683*, 291–299.
- (27) Li, Y.; Lu, J. Optimal Orbital Selection for Full Configuration Interaction (OptOrbFCI): Pursuing the Basis Set Limit under a Budget. *Journal of Chemical Theory and Computation* **2020**, *16*, 6207–6221.
- (28) Wang, Z.; Li, Y.; Lu, J. Coordinate Descent Full Configuration Interaction. *Journal of Chemical Theory and Computation* **2019**, *15*, 3558–3569.
- (29) Gao, B.; Liu, X.; Yuan, Y.-x. Parallelizable Algorithms for Optimization Problems with Orthogonality Constraints. *SIAM Journal on Scientific Computing* **2019**, *41*, A1949–A1983.
- (30) Wen, Z.; Yin, W. A Feasible Method for Optimization with Orthogonality Constraints. *Mathematical Programming* **2013**, *142*, 397–434.
- (31) Zhang, X.; Zhu, J.; Wen, Z.; Zhou, A. Gradient Type Optimization Methods For Electronic Structure Calculations. *SIAM Journal on Scientific Computing* **2014**, *36*, C265–C289.
- (32) Huang, W.; Gallivan, K. A.; Absil, P.-A. A Broyden Class of Quasi-Newton Methods for Riemannian Optimization. *SIAM Journal on Optimization* **2015**, *25*, 1660–1685.
- (33) Gao, B.; Liu, X.; Chen, X.; Yuan, Y.-x. A New First-Order Algorithmic Framework for Optimization Problems with Orthogonality Constraints. *SIAM Journal on Optimization* **2018**, *28*, 302–332.
- (34) Nielsen, M. A.; Chuang, I. L., *Quantum Computation and Quantum Information: 10th Anniversary Edition*; Cambridge University Press: 2010.
- (35) Dawson, C. M.; Nielsen, M. A. The solovay-kitaev algorithm. *arXiv preprint quant-ph/0505030* **2005**.
- (36) Litinski, D. A Game of Surface Codes: Large-Scale Quantum Computing with Lattice Surgery. *Quantum* **2019**, *3*, 128.
- (37) Xu, Q.; Ataiades, J. P. B.; Pattison, C. A.; Raveendran, N.; Bluvstein, D.; Wurtz, J.; Vasic, B.; Lukin, M. D.; Jiang, L.; Zhou, H. Constant-Overhead Fault-Tolerant Quantum Computation with Reconfigurable Atom Arrays, 2023.
- (38) Bravyi, S.; Cross, A. W.; Gambetta, J. M.; Maslov, D.; Rall, P.; Yoder, T. J. High-Threshold and Low-Overhead Fault-Tolerant Quantum Memory, 2023.

- (39) Fowler, A. G.; Stephens, A. M.; Groszkowski, P. High-Threshold Universal Quantum Computation on the Surface Code. *Physical Review A* **2009**, *80*, 052312.
- (40) Fowler, A. G.; Mariantoni, M.; Martinis, J. M.; Cleland, A. N. Surface Codes: Towards Practical Large-Scale Quantum Computation. *Physical Review A* **2012**, *86*, 032324.
- (41) Horsman, C.; Fowler, A. G.; Devitt, S.; Meter, R. V. Surface Code Quantum Computing by Lattice Surgery. *New Journal of Physics* **2012**, *14*, 123011.
- (42) Campbell, E. T.; Terhal, B. M.; Vuillot, C. Roads towards Fault-Tolerant Universal Quantum Computation. *Nature* **2017**, *549*, 172–179.
- (43) Temme, K.; Bravyi, S.; Gambetta, J. M. Error Mitigation for Short-Depth Quantum Circuits. *Physical Review Letters* **2017**, *119*, 180509.
- (44) Piveteau, C.; Sutter, D.; Bravyi, S.; Gambetta, J. M.; Temme, K. Error Mitigation for Universal Gates on Encoded Qubits. *Physical Review Letters* **2021**, *127*, 200505.
- (45) Koczor, B. Exponential Error Suppression for Near-Term Quantum Devices. *Physical Review X* **2021**, *11*, 031057.
- (46) Endo, S.; Zhao, Q.; Li, Y.; Benjamin, S.; Yuan, X. Mitigating Algorithmic Errors in a Hamiltonian Simulation. *Physical Review A* **2019**, *99*, 012334.
- (47) Endo, S.; Benjamin, S. C.; Li, Y. Practical Quantum Error Mitigation for Near-Future Applications. *Physical Review X* **2018**, *8*, 031027.
- (48) Suzuki, Y.; Endo, S.; Fujii, K.; Tokunaga, Y. Quantum Error Mitigation as a Universal Error Reduction Technique: Applications from the NISQ to the Fault-Tolerant Quantum Computing Eras. *PRX Quantum* **2022**, *3*, 010345.
- (49) Czarnik, P.; Arrasmith, A.; Cincio, L.; Coles, P. J. Qubit-Efficient Exponential Suppression of Errors, 2021.
- (50) Xiong, Y.; Chandra, D.; Ng, S. X.; Hanzo, L. Sampling Overhead Analysis of Quantum Error Mitigation: Uncoded vs. Coded Systems. *IEEE Access* **2020**, *8*, 228967–228991.
- (51) Huggins, W. J.; McArdle, S.; O’Brien, T. E.; Lee, J.; Rubin, N. C.; Boixo, S.; Whaley, K. B.; Babbush, R.; McClean, J. R. Virtual Distillation for Quantum Error Mitigation. *Physical Review X* **2021**, *11*, 041036.
- (52) Feynman, R. P. Simulating Physics with Computers. *International Journal of Theoretical Physics* **1982**, *21*, 467–488.

- (53) Tranter, A.; Love, P. J.; Mintert, F.; Coveney, P. V. A Comparison of the Bravyi–Kitaev and Jordan–Wigner Transformations for the Quantum Simulation of Quantum Chemistry. *Journal of Chemical Theory and Computation* **2018**, *14*, 5617–5630.
- (54) Chien, R. W.; Xue, S.; Hardikar, T. S.; Setia, K.; Whitfield, J. D. Analysis of Superfast Encoding Performance for Electronic Structure Simulations. *Physical Review A* **2019**, *100*, 032337.
- (55) Setia, K.; Whitfield, J. D. Bravyi-Kitaev Superfast Simulation of Electronic Structure on a Quantum Computer. *The Journal of Chemical Physics* **2018**, *148*, 164104.
- (56) Bravyi, S.; Gambetta, J. M.; Mezzacapo, A.; Temme, K. Tapering off qubits to simulate fermionic Hamiltonians. *arXiv preprint arXiv:1701.08213* **2017**.
- (57) Seeley, J. T.; Richard, M. J.; Love, P. J. The Bravyi-Kitaev Transformation for Quantum Computation of Electronic Structure. *The Journal of Chemical Physics* **2012**, *137*, 224109.
- (58) Havlíček, V.; Troyer, M.; Whitfield, J. D. Operator Locality in the Quantum Simulation of Fermionic Models. *Physical Review A* **2017**, *95*, 032332.
- (59) Arrazola, J. M.; Matteo, O. D.; Quesada, N.; Jahangiri, S.; Delgado, A.; Killoran, N. Universal Quantum Circuits for Quantum Chemistry. *Quantum* **2022**, *6*, 742.
- (60) Abrams, D. S.; Lloyd, S. Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors. *Physical Review Letters* **1999**, *83*, 5162–5165.
- (61) Whitfield, J. D.; Biamonte, J.; Aspuru-Guzik, A. Simulation of Electronic Structure Hamiltonians Using Quantum Computers. *Molecular Physics* **2011**, *109*, 735–750.
- (62) O’Malley, P. J. J. et al. Scalable Quantum Simulation of Molecular Energies. *Physical Review X* **2016**, *6*, 031007.
- (63) O’Brien, T. E.; Tarasinski, B.; Terhal, B. M. Quantum Phase Estimation of Multiple Eigenvalues for Small-Scale (Noisy) Experiments. *New Journal of Physics* **2019**, *21*, 023022.
- (64) Svore, K. M.; Hastings, M. B.; Freedman, M. Faster Phase Estimation. *Quantum Information & Computation* **2014**, *14*, 306–328.
- (65) Paesani, S.; Gentile, A. A.; Santagati, R.; Wang, J.; Wiebe, N.; Tew, D. P.; O’Brien, J. L.; Thompson, M. G. Experimental Bayesian Quantum Phase Estimation on a Silicon Photonic Chip. *Physical Review Letters* **2017**, *118*, 100503.

- (66) Wiebe, N.; Granade, C. Efficient Bayesian Phase Estimation. *Physical Review Letters* **2016**, *117*, 010503.
- (67) Dobšíček, M.; Johansson, G.; Shumeiko, V.; Wendin, G. Arbitrary Accuracy Iterative Quantum Phase Estimation Algorithm Using a Single Ancillary Qubit: A Two-Qubit Benchmark. *Physical Review A* **2007**, *76*, 030306.
- (68) Knill, E.; Ortiz, G.; Somma, R. D. Optimal Quantum Measurements of Expectation Values of Observables. *Physical Review A* **2007**, *75*, 012328.
- (69) Childs, A. M.; Preskill, J.; Renes, J. Quantum Information and Precision Measurement. *Journal of Modern Optics* **2000**, *47*, 155–176.
- (70) Pitman, J., *Probability*; Springer Science & Business Media: 1999.
- (71) Lloyd, S. Universal Quantum Simulators. *Science* **1996**, *273*, 1073.
- (72) Peterson, K. A.; Feller, D.; Dixon, D. A. Chemical Accuracy in Ab Initio Thermochemistry and Spectroscopy: Current Strategies and Future Challenges. *Theoretical Chemistry Accounts* **2012**, *131*, 1079.
- (73) Gonthier, J. F.; Radin, M. D.; Buda, C.; Duskocil, E. J.; Abuan, C. M.; Romero, J. Measurements as a Roadblock to Near-Term Practical Quantum Advantage in Chemistry: Resource Analysis. *Physical Review Research* **2022**, *4*, 033154.
- (74) Kühn, M.; Zanker, S.; Deglmann, P.; Marthaler, M.; Weiß, H. Accuracy and Resource Estimations for Quantum Chemistry on a Near-Term Quantum Computer. *Journal of Chemical Theory and Computation* **2019**, *15*, 4764–4780.
- (75) Kjaergaard, M.; Schwartz, M. E.; Braumüller, J.; Krantz, P.; Wang, J. I.-J.; Gustavsson, S.; Oliver, W. D. Superconducting Qubits: Current State of Play. *Annual Review of Condensed Matter Physics* **2020**, *11*, 369–395.
- (76) Berry, D. W.; Gidney, C.; Motta, M.; McClean, J. R.; Babbush, R. Qubitization of Arbitrary Basis Quantum Chemistry Leveraging Sparsity and Low Rank Factorization. *Quantum* **2019**, *3*, 208.
- (77) Low, G. H.; Chuang, I. L. Hamiltonian Simulation by Qubitization. *Quantum* **2019**, *3*, 163.
- (78) Tilly, J.; Chen, H.; Cao, S.; Picozzi, D.; Setia, K.; Li, Y.; Grant, E.; Wossnig, L.; Rungger, I.; Booth, G. H.; Tennyson, J. The Variational Quantum Eigensolver: A Review of Methods and Best Practices. *Physics Reports* **2022**, *986*, 1–128.

- (79) McClean, J. R.; Romero, J.; Babbush, R.; Aspuru-Guzik, A. The Theory of Variational Hybrid Quantum-Classical Algorithms. *New Journal of Physics* **2016**, *18*, 023023.
- (80) Cade, C.; Mineh, L.; Montanaro, A.; Staniscic, S. Strategies for Solving the Fermi-Hubbard Model on near-Term Quantum Computers. *Physical Review B* **2020**, *102*, 235122.
- (81) Moll, N. et al. Quantum Optimization Using Variational Algorithms on Near-Term Quantum Devices. *Quantum Science and Technology* **2018**, *3*, 030503.
- (82) Saib, W.; Wallden, P.; Akhalwaya, I. In *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, 2021, pp 42–53.
- (83) Schuld, M.; Bergholm, V.; Gogolin, C.; Izaac, J.; Killoran, N. Evaluating Analytic Gradients on Quantum Hardware. *Physical Review A* **2019**, *99*, 032331.
- (84) Mari, A.; Bromley, T. R.; Killoran, N. Estimating the Gradient and Higher-Order Derivatives on Quantum Hardware. *Physical Review A* **2021**, *103*, 012405.
- (85) Crooks, G. E. Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition. *arXiv preprint arXiv:1905.13311* **2019**.
- (86) Kottmann, J. S.; Anand, A.; Aspuru-Guzik, A. A Feasible Approach for Automatically Differentiable Unitary Coupled-Cluster on Quantum Computers. *Chemical Science* **2021**, *12*, 3497–3508.
- (87) Wierichs, D.; Izaac, J.; Wang, C.; Lin, C. Y.-Y. General Parameter-Shift Rules for Quantum Gradients. *Quantum* **2022**, *6*, 677.
- (88) 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.
- (89) Cerezo, M.; Sone, A.; Volkoff, T.; Cincio, L.; Coles, P. J. Cost Function Dependent Barren Plateaus in Shallow Parametrized Quantum Circuits. *Nature Communications* **2021**, *12*, 1791.
- (90) O’Gorman, B.; Huggins, W. J.; Rieffel, E. G.; Whaley, K. B. Generalized swap networks for near-term quantum computing. *arXiv preprint arXiv:1905.05118* **2019**.
- (91) Gard, B. T.; Zhu, L.; Barron, G. S.; Mayhall, N. J.; Economou, S. E.; Barnes, E. Efficient Symmetry-Preserving State Preparation Circuits for the Variational Quantum Eigensolver Algorithm. *npj Quantum Information* **2020**, *6*, 1–9.

- (92) Lee, J.; Huggins, W. J.; Head-Gordon, M.; Whaley, K. B. Generalized Unitary Coupled Cluster Wave Functions for Quantum Computation. *Journal of Chemical Theory and Computation* **2019**, *15*, 311–324.
- (93) 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. *Physical Review A* **2017**, *95*, 020501.
- (94) 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.
- (95) Claudino, D.; Wright, J.; McCaskey, A. J.; Humble, T. S. Benchmarking Adaptive Variational Quantum Eigensolvers. *Frontiers in Chemistry* **2020**, *8*.
- (96) Yordanov, Y. S.; Armaos, V.; Barnes, C. H. W.; Arvidsson-Shukur, D. R. M. Qubit-Excitation-Based Adaptive Variational Quantum Eigensolver. *Communications Physics* **2021**, *4*, 1–11.
- (97) Stokes, J.; Izaac, J.; Killoran, N.; Carleo, G. Quantum Natural Gradient. *Quantum* **2020**, *4*, 269.
- (98) Wierichs, D.; Gogolin, C.; Kastoryano, M. Avoiding Local Minima in Variational Quantum Eigensolvers with the Natural Gradient Optimizer. *Physical Review Research* **2020**, *2*, 043246.
- (99) Gacon, J.; Zoufal, C.; Carleo, G.; Woerner, S. Simultaneous Perturbation Stochastic Approximation of the Quantum Fisher Information. *Quantum* **2021**, *5*, 567.
- (100) Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing* **1995**, *16*, 1190–1208.
- (101) Powell, M. J. D. Direct Search Algorithms for Optimization Calculations. *Acta Numerica* **1998**, *7*, 287–336.
- (102) 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. *Physical Review A* **2017**, *95*, 042308.
- (103) 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. *Physical Review X* **2018**, *8*, 011021.

- (104) Yoshioka, N.; Hakoshima, H.; Matsuzaki, Y.; Tokunaga, Y.; Suzuki, Y.; Endo, S. Generalized Quantum Subspace Expansion. *Physical Review Letters* **2022**, *129*, 020502.
- (105) McClean, J. R.; Jiang, Z.; Rubin, N. C.; Babbush, R.; Neven, H. Decoding Quantum Errors with Subspace Expansions. *Nature Communications* **2020**, *11*, 636.
- (106) Tazi, L. C.; Thom, A. J. W. Folded Spectrum VQE : A Quantum Computing Method for the Calculation of Molecular Excited States, 2023.
- (107) Higgott, O.; Wang, D.; Brierley, S. Variational Quantum Computation of Excited States. *Quantum* **2019**, *3*, 156.
- (108) Ibe, Y.; Nakagawa, Y. O.; Earnest, N.; Yamamoto, T.; Mitarai, K.; Gao, Q.; Kobayashi, T. Calculating Transition Amplitudes by Variational Quantum Deflation. *Physical Review Research* **2022**, *4*, 013173.
- (109) Buhrman, H.; Cleve, R.; Watrous, J.; de Wolf, R. Quantum Fingerprinting. *Physical Review Letters* **2001**, *87*, 167902.
- (110) Cincio, L.; Subaşı, Y.; Sornborger, A. T.; Coles, P. J. Learning the Quantum Algorithm for State Overlap. *New Journal of Physics* **2018**, *20*, 113022.
- (111) Nakanishi, K. M.; Mitarai, K.; Fujii, K. Subspace-Search Variational Quantum Eigensolver for Excited States. *Physical Review Research* **2019**, *1*, 033062.
- (112) Parrish, R. M.; Hohenstein, E. G.; McMahon, P. L.; Martínez, T. J. Quantum Computation of Electronic Transitions Using a Variational Quantum Eigensolver. *Physical Review Letters* **2019**, *122*, 230401.
- (113) Wang, S.; Fontana, E.; Cerezo, M.; Sharma, K.; Sone, A.; Cincio, L.; Coles, P. J. Noise-Induced Barren Plateaus in Variational Quantum Algorithms. *Nature Communications* **2021**, *12*, 6961.
- (114) Bierman, J.; Li, Y.; Lu, J. Quantum Orbital Minimization Method for Excited States Calculation on a Quantum Computer. *Journal of Chemical Theory and Computation* **2022**, *18*, 4674–4689.
- (115) Lu, J.; Thicke, K. Orbital Minimization Method with ℓ_1 Regularization. *Journal of Computational Physics* **2017**, *336*, 87–103.
- (116) Wu, B.; Ray, M.; Zhao, L.; Sun, X.; Rebentrost, P. Quantum-Classical Algorithms for Skewed Linear Systems with an Optimized Hadamard Test. *Physical Review A* **2021**, *103*, 042422.

- (117) Qiskit contributors Qiskit: An Open-source Framework for Quantum Computing, 2023.
- (118) 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. *WIREs Computational Molecular Science* **2018**, *8*, e1340.
- (119) Bierman, J.; Li, Y.; Lu, J. Improving the Accuracy of Variational Quantum Eigensolvers with Fewer Qubits Using Orbital Optimization. *Journal of Chemical Theory and Computation* **2023**, *19*, 790–798.
- (120) Bierman, J.; Li, Y.; Lu, J. Qubit Count Reduction by Orthogonally Constrained Orbital Optimization for Variational Quantum Excited-State Solvers. *Journal of Chemical Theory and Computation* **2024**, Article ASAP. DOI: 10.1021/acs.jctc.3c01297 (accessed 2024-04-10).
- (121) McCaskey, A. J.; Parks, Z. P.; Jakowski, J.; Moore, S. V.; Morris, T. D.; Humble, T. S.; Pooser, R. C. Quantum Chemistry as a Benchmark for Near-Term Quantum Computers. *npj Quantum Information* **2019**, *5*, 1–8.
- (122) Ollitrault, P. J.; Kandala, A.; Chen, C.-F.; Barkoutsos, P. K.; Mezzacapo, A.; Pistoia, M.; Sheldon, S.; Woerner, S.; Gambetta, J. M.; Tavernelli, I. Quantum Equation of Motion for Computing Molecular Excitation Energies on a Noisy Quantum Processor. *Physical Review Research* **2020**, *2*, 043140.
- (123) Metcalf, M.; Bauman, N. P.; Kowalski, K.; de Jong, W. A. Resource-Efficient Chemistry on Quantum Computers with the Variational Quantum Eigensolver and the Double Unitary Coupled-Cluster Approach. *Journal of Chemical Theory and Computation* **2020**, *16*, 6165–6175.
- (124) Huang, R.; Li, C.; Evangelista, F. A. Leveraging Small-Scale Quantum Computers with Unitarily Downfolded Hamiltonians. *PRX Quantum* **2023**, *4*, 020313.
- (125) Claudino, D.; Peng, B.; Bauman, N. P.; Kowalski, K.; Humble, T. S. Improving the Accuracy and Efficiency of Quantum Connected Moments Expansions. *Quantum Science and Technology* **2021**, *6*, 034012.
- (126) Motta, M.; Gujarati, T. P.; Rice, J. E.; Kumar, A.; Masteran, C.; Latone, J. A.; Lee, E.; Valeev, E. F.; Takeshita, T. Y. Quantum Simulation of Electronic Structure with a Transcorrelated Hamiltonian: Improved Accuracy with a Smaller Footprint on the Quantum Computer. *Physical Chemistry Chemical Physics* **2020**, *22*, 24270–24281.
- (127) Sokolov, I. O.; Dobrutz, W.; Luo, H.; Alavi, A.; Tavernelli, I. Orders of Magnitude Reduction in the Computational Overhead for Quantum Many-Body

- Problems on Quantum Computers via an Exact Transcorrelated Method. **2023**, arXiv:2201.03049.
- (128) Mizukami, W.; Mitarai, K.; Nakagawa, Y. O.; Yamamoto, T.; Yan, T.; Ohnishi, Y.-y. Orbital Optimized Unitary Coupled Cluster Theory for Quantum Computer. *Physical Review Research* **2020**, *2*, 033421.
- (129) Tilly, J.; Sriluckshmy, P. V.; Patel, A.; Fontana, E.; Rungger, I.; Grant, E.; Anderson, R.; Tennyson, J.; Booth, G. H. Reduced Density Matrix Sampling: Self-consistent Embedding and Multiscale Electronic Structure on Current Generation Quantum Computers. *Physical Review Research* **2021**, *3*, 033230.
- (130) Yalouz, S.; Senjean, B.; Günther, J.; Buda, F.; O'Brien, T. E.; Visscher, L. A State-Averaged Orbital-Optimized Hybrid Quantum-Classical Algorithm for a Democratic Description of Ground and Excited States. *Quantum Science and Technology* **2021**, *6*, 024004.
- (131) Takeshita, T.; Rubin, N. C.; Jiang, Z.; Lee, E.; Babbush, R.; McClean, J. R. Increasing the Representation Accuracy of Quantum Simulations of Chemistry without Extra Quantum Resources. *Physical Review X* **2020**, *10*, 011004.
- (132) Motta, M.; Sun, C.; Tan, A. T. K.; O'Rourke, M. J.; Ye, E.; Minnich, A. J.; Brandão, F. G. S. L.; Chan, G. K.-L. Determining Eigenstates and Thermal States on a Quantum Computer Using Quantum Imaginary Time Evolution. *Nature Physics* **2020**, *16*, 205–210.
- (133) McArdle, S.; Jones, T.; Endo, S.; Li, Y.; Benjamin, S. C.; Yuan, X. Variational Ansatz-Based Quantum Simulation of Imaginary Time Evolution. *npj Quantum Information* **2019**, *5*, 1–6.
- (134) McClean, J. R.; Aspuru-Guzik, A. Compact Wavefunctions from Compressed Imaginary Time Evolution. *RSC Advances* **2015**, *5*, 102277–102283.
- (135) Tsuchimochi, T.; Ryo, Y.; Ten-no, S. L.; Sasasako, K. Improved Algorithms of Quantum Imaginary Time Evolution for Ground and Excited States of Molecular Systems. *Journal of Chemical Theory and Computation* **2023**, *19*, 503–513.
- (136) Huggins, W. J.; O'Gorman, B. A.; Rubin, N. C.; Reichman, D. R.; Babbush, R.; Lee, J. Unbiasing Fermionic Quantum Monte Carlo with a Quantum Computer. *Nature* **2022**, *603*, 416–420.
- (137) Paszke, A. et al. In *Advances in Neural Information Processing Systems*, ed. by Wallach, H.; Larochelle, H.; Beygelzimer, A.; dAlché-Buc, F.; Fox, E.; Garnett, R., Curran Associates, Inc.: 2019; Vol. 32.

- (138) 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.
- (139) Wang, D.; Higgott, O.; Brierley, S. Accelerated Variational Quantum Eigensolver. *Physical Review Letters* **2019**, *122*, 140504.
- (140) Asthana, A.; Kumar, A.; Abraham, V.; Grimsley, H.; Zhang, Y.; Cincio, L.; Tretiak, S.; Dub, P. A.; Economou, S. E.; Barnes, E.; Mayhall, N. J. Quantum Self-Consistent Equation-of-Motion Method for Computing Molecular Excitation Energies, Ionization Potentials, and Electron Affinities on a Quantum Computer. *Chemical Science* **2023**, *14*, 2405–2418.
- (141) Yalouz, S.; Robert, V. Orthogonally Constrained Orbital Optimization: Assessing Changes of Optimal Orbitals for Orthogonal Multireference States. *Journal of Chemical Theory and Computation* **2023**, *19*, 1388–1392.
- (142) Omiya, K.; Nakagawa, Y. O.; Koh, S.; Mizukami, W.; Gao, Q.; Kobayashi, T. Analytical Energy Gradient for State-Averaged Orbital-Optimized Variational Quantum Eigensolvers and Its Application to a Photochemical Reaction. *Journal of Chemical Theory and Computation* **2022**, *18*, 741–748.
- (143) McArdle, S.; Tew, D. P. Improving the accuracy of quantum computational chemistry using the transcorrelated method. *arXiv preprint arXiv:2006.11181* **2020**.
- (144) Harris, C. R. et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362.
- (145) Kivlichan, I. D.; McClean, J.; Wiebe, N.; Gidney, C.; Aspuru-Guzik, A.; Chan, G. K.-L.; Babbush, R. Quantum Simulation of Electronic Structure with Linear Depth and Connectivity. *Physical Review Letters* **2018**, *120*, 110501.
- (146) Huggins, W. J.; McClean, J. R.; Rubin, N. C.; Jiang, Z.; Wiebe, N.; Whaley, K. B.; Babbush, R. Efficient and Noise Resilient Measurements for Quantum Chemistry on Near-Term Quantum Computers. *npj Quantum Information* **2021**, *7*, 1–9.

Biography

Joel Bierman received B.S. degrees in physics and mathematics with a minor in chemistry from Indiana University Bloomington in spring 2019. He started attending Duke University in fall 2019 and started working as a graduate student researcher with Professor Jianfeng Lu in summer 2020. He will start working as a postdoctoral researcher with Professor Yuan Liu at North Carolina State University in summer 2024.