Title
Applications of Near-Term Quantum Computers

Author
Huggins, William James

Publication Date
2020

Peer reviewed|Thesis/dissertation

Applications of Near-Term Quantum Computers

by

William Huggins

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor K. Birgitta Whaley, Chair
Professor Martin Head-Gordon
Professor Joel Moore

Summer 2020

Applications of Near-Term Quantum Computers

Abstract

Applications of Near-Term Quantum Computers

by

William Huggins

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor K. Birgitta Whaley, Chair

Quantum computers exist today that are capable of performing calculations that challenge the largest classical supercomputers. Now the great challenge is to make use of these devices, or their successors, to perform calculations of independent interest. This thesis focuses on a family of algorithms that aim to accomplish this goal, known as variational quantum algorithms, and the challenges that they face. We frame these challenges in terms of two kinds of resources, the number of operations that we can afford to perform in a single quantum circuit, and the overall number of circuit repetitions required. Before turning to our specific contributions, we provide a review of the electronic structure problem, which serves as a central example for our work. We also provide a self-contained introduction to the field of quantum computing and the notion of a variational quantum algorithm.

We begin the main body of the thesis by studying a generalization of the standard unitary coupled cluster ansatz. We present a sparse version of unitary coupled cluster and show that it can accurately represent the ground and excited states of small molecular systems using fewer quantum gates than the full unitary coupled cluster. We then introduce a strategy for representing molecular ground states as linear combinations of parameterized wavefunctions, allowing for a tradeoff between the number of operations required for each circuit and the number of circuit repetitions. We provide circuit primitives that allow for the efficient measurement of the required matrix elements between these wavefunctions. Subsequently, we show how the cost of estimating the energy of a quantum chemical wavefunction on a near-term quantum computer can be dramatically reduced by using a factorization of the two-electron integral tensor. Furthermore, we explain how this measurement strategy helps mitigate against errors during state preparation and measurement. We then present a Monte Carlo version of a classical algorithm for calculating the partition function of two-dimensional lattice models that shows a similar kind of tradeoff between two resources, albeit in a different context. Finally, we show how ideas based on tensor networks can inform the design of quantum circuits for machine learning tasks. We argue that this application is especially

tolerant to noise and present numerical data substantiating this argument. We conclude by explaining explicitly how our work addresses the problems of the limited resources available to near-term quantum computers and offering some optimism about the future of the field.

Dedicated to my parents, John and Diane, my brother Bobby, and my partner Hilary

Your support and encouragement has made this dissertation, and so much else in my life, possible.

Contents

List of Figures

List of Tables

Acknowledgments

This thesis is the culmination of my time at Berkeley, and I owe a debt of gratitude to the many people who have provided mentorship, support, and camaraderie on the journey.

I would like to begin by thanking my advisor, Professor K. Birgitta Whaley. I'm grateful for her willingness to take a chance on a student without much formal chemistry training, and her guidance over the years. She pushed me to become a better scientist and communicator, while also allowing me the freedom to explore and to make my own mistakes.

I've found myself the recipient of incredible generosity from a series of other guides along the way. Dr. Norm Tubman welcomed me into Birgitta's group and helped me channel my early excitement with his infinite patience. Dr. Miles Stoudenmire taught me much of what I know about tensor networks and helped me understand the importance of intuition in science. Dr. Jonathan DuBois and Professor Berni Alder graciously welcomed me for a summer and opened my eyes to the power of Monte Carlo methods. Professor Dave Bacon introduced me to the importance of beer in quantum computing, testing in software, and group theory in everything. Dr. Ryan Babbush inspired me with his enthusiasm for science and for life, while helping me to become more confident in my communication and creative in my problem solving.

This thesis would not exist without my collaborators, who contributed so much directly through their work and indirectly through what they have taught me. Much of what I know about quantum chemistry I owe to Dr. Joonho Lee, who also helped me understand the value of constructive criticism and the joy of fruitful collaboration. Bryan O'Gorman showed me how precise abstractions can make it possible to wave away irrelevant details and focus on the essence of a problem. Dr. Daniel Freeman and Unpil Baek bounced ideas back and forth with me, putting up with more poorly-drawn tensor network diagrams than anyone should have to. Piyush Patil and Bradley Mitchell dove fearlessly with me into work that was new for all of us in different ways. Dr. Jarrod McClean, Dr. Nicholas Rubin, Dr. Zhang Jiang, and Professor Nathan Wiebe welcomed me into their research team and I hope that I've absorbed something of each of their unique perspectives on how to approach the unknown.

I would like to thank Professor Martin Head-Gordon, who, besides being a valuable collaborator directly for a part of this thesis, has served as a member of my qualifying exam committee and now my thesis committee. I would also like to thank Professor Joel Moore for his service on both my qualifying exam committee and my thesis committee. I am grateful for Professor Eric Neuscamman's I am grateful for Professor Eric Neuscamman's service on my qualifying exam committee, and for his time and advice over the years. I'd like to acknowledge Professor Eran Rabani for his service on my qualifying exam committee, and for the course he co-taught with Professor Naomi Ginsberg; I am grateful to both of them for teaching me to use Taylor expansions like a real physicist.

I'd like to thank the others at Berkeley who have taught me, challenged me, and inspired me. Professor Littlejohn for his extraordinary dedication to physics education. Dr. Mohan Sarovar, Professor Lin Lin, and Dr. Gabriel Durkin for interesting discussions over the years. The other members of the Whaley Group, especially those I've had the pleasure of interacting

with the most frequently but haven't yet mentioned, Carlos Mejuto Zaera, Song Zhang, Ian Convy, and Haoran Liao. We may not (yet) have published work together, but I've learned much from you and cherish your friendship. I'm grateful too to the friends and companions I've made throughout my time here at Berkeley, and the ones that I brought with me. You lunch buddies, study buddies, party members, rabble-rousers, knights of the square table, and fellow basement dwellers, I wouldn't have made it without you.

I'd especially like to thank those that shaped me into the person ready (and crazy enough) to pursue this degree. Without a chain of teachers, Marty Caplan, Joel and Eric, Mrs. Wickham and Mr. Calhoun, Professor Casalaina-Martin, Professor Pflaum, and Professor Mozer, I would not have been able to begin this thesis. Without friends further along the path, Dr. Caleb Phillips, Esteban Rios, Peter Van Blerkom, William Treuren, Dr. Evan Starr, I would not have been inspired to continue. Without the loving support of my partner, Hilary, and the sweetness we share, I would have not have had the courage to finish it. I owe a special thanks to my brother Bobby, for his love and for endless late nights talking about math. Finally, I'd like to thank the giants whose shoulders I stand upon. I mean, of course, those scientists who came before and built the foundation for this thesis, and also my parents John Huggins and Diane Skufca, and grandparents, who came before and built the foundation for me.

Chapter 1

Introduction

> Nature isn't classical, dammit, and if you want to make a simulation of nature, you'd better make it quantum mechanical, and by golly it's a wonderful problem, because it doesn't look so easy.
>
> R. P.Feynman [1]

Quantum computing promises to revolutionize computational physics and chemistry. A quantum computer stores and manipulates information in a way that is fundamentally quantum mechanical. This would allow a sufficiently powerful quantum computer to sidestep the challenges associated with a computational treatment of quantum phenomena on classical computers [2]. Furthermore, quantum algorithms for specific computational tasks removed from the simulation of quantum mechanics itself have also been developed, most famously Shor's algorithm for integer factorization [3] and Grover's algorithm for unstructured search [4, 5]. From the point of view of computational complexity theory, quantum computing provides the first credible challenge to the extended Church-Turing thesis [6]. Informally stated, quantum computing provides the first reasonable model of computation that might be capable of performing certain problems exponentially faster than a classical computer.

Google's recent results on (the unfortunately named) "quantum supremacy" show that this revolution is drawing nearer [7]. Essentially, they carried out a computation with a programmable quantum computer orders of magnitude faster than would be possible using even the largest classical supercomputers. Importantly, however, this calculation was designed specifically to be challenging for a classical computer. We shall discuss this experiment in more detail in Section 1.1, explaining the computation and highlighting some of the caveats of their experiment. Notwithstanding these caveats, their experimental achievement is a milestone that heralds a new era in the field of quantum computing. Modern quantum

computers are now large and performant enough that they are no longer trivially simulable by classical computers.

The threshold theorem tells us that a scalable, fault-tolerant, quantum computer should be feasible in practice [8]. Using the appropriate generalizations of classical error-correcting codes, it will be possible to leverage precise but imperfect control over a collection of quantum systems in order to achieve an exponentially more accurate control over a smaller collection of logical degrees of freedom. These techniques are expected to eventually allow us to perform computations on quantum information with arbitrarily small error rates, incurring an overhead that grows polylogarithmically with the size of the computation [9]. Unfortunately, the constant factors in this overhead using the best known techniques put the use of quantum error correction far out of reach for today's quantum computers [10, 11].

It is the tremendous chasm between the quantum computer Google used in their supremacy experiment and a hypothetical quantum computer capable of fault-tolerant computation that motivates this thesis. John Preskill coined the term Noisy Intermediate-Scale Quantum (NISQ) to describe technologies that fall into this gap [10]. Small enough quantum computers are trivially simulable, and so can not answer calculations out of reach of classical computation. On the other hand, we have compelling applications for devices that are sufficiently large and powerful enough that they can perform fault-tolerant computation [2–4]. The central question that motivates this thesis is, how might we bridge that gap? How might we extract utility from quantum computers available in the NISQ era? We shall present several contributions that we have made in this direction, but first we lay the groundwork for their description.

1.1 Noisy Intermediate-Scale Quantum Computing

The term Noisy Intermediate-Scale Quantum refers to devices that are beyond the reach of classical simulation but not powerful enough to support fully fault-tolerant quantum computation. As Preskill notes in Ref. 10, this classification is a matter both of quantity and of quality. At the bottom end, a quantum computer composed of few enough qubits (two-level systems) can be simulated at a cost that is exponential in the number of qubits. Similarly, a quantum computer with sufficiently high levels of noise (imperfect control and measurement) is also efficiently simulable, regardless of its size [12]. These observations lower bound the necessary quantity and quality (of the qubits). At the upper end, it is known that tradeoffs are possible between the number of qubits and error rates required to achieve fault-tolerant quantum computation. Typical numbers suggest an overhead of 1000-fold in the number of physical qubits per error-corrected logical qubit will be required [10].

By this definition, Google's Sycamore chip is perhaps the first NISQ quantum computer [7]. With 53 qubits, exact simulation in the 2^{53} dimensional Hilbert space would require storing and manipulating $2^{54} - 1$ real numbers. Ref. 13 points out that such a simulation could be carried out to a reasonable accuracy using the majority of the disk space of the world's largest supercomputer for storage, but the simulation time would still be orders of magnitude slower than the operation of the device. Some advantages in simulation time can be achieved

by taking advantage of the fact that the non-zero error rates of the device, but the best algorithms to date are not capable of quickly simulating the quantum supremacy experiment of Ref. 7.

In this thesis, we shall focus on quantum computation, rather than analog quantum simulation. We make this distinction here because the reader could fairly object that the achievement of Ref. 7 is not particularly special. Other experiments have been carried out that seem to perform computations about quantum mechanics at the boundary classical simulation [14]. We define a quantum computer to be a device that is capable of universal computation, a notion that we shall specify more concretely in Section 3.1. An analog quantum simulator is a programmable device capable of emulating a certain family of quantum systems, but one that does not support arbitrary computations. In the near-term, this distinction may be fuzzy and only partly helpful. Eventually, however, there is a clear path towards fault-tolerance for quantum computation but not for analog quantum simulation [10]. For this reason, and for the flexibility that a universal set of operations affords even on a near-term device, we shall focus on quantum computers rather than special-purpose quantum simulators throughout this thesis.

The specific task of the quantum computer in the "quantum supremacy" experiment of Ref. 7 was to sample from the output distribution of a family of random quantum circuits. As we shall explain in Section 3.1, the output of a quantum computation can always be regarded as a string of bits. Arute et al. repeatedly executed particular randomly chosen quantum circuits and compared the bit strings they observed to the distributions that would be expected from an ideal quantum computer. There is strong evidence that performing these computations is challenging for a classical computer, even when accounting for the fact that the quantum device is noisy [7, 15, 16]. Although the circuits executed in Ref. 7 create highly entangled states that defy simulation by existing classical methods, it isn't clear that a device with similar characteristics is capable of treating computational problems of interest beyond this random circuit sampling task.

There have been a number of proposals to make use of NISQ-era quantum computers for meaningful applications. Heuristic approaches have been proposed for use in quantum chemistry [17], many-body quantum physics [18], combinatorial optimization [19], machine learning [20], and other fields. The level of theoretical justification for these different approaches varies, but, when executed on imperfect hardware, they all share a lack of provable advantage over existing classical techniques. Ref. 7 makes the case that a noisy 53-qubit quantum computer is capable of performing a task that is hard for existing classical computers. It remains an open question if such a device, or the larger and less noisy NISQ devices to come, will be able to provide a useful quantum advantage for a task of independent interest. This thesis presents several contributions that are part of vast body of work attempting to provide a constructive, affirmative, answer to this question.

1.2 Outline

We shall present a brief outline of the thesis in this section. Following this introductory chapter, we shall present background material necessary to provide context for the main body of the thesis in in Chapter 2 and Chapter 3. We shall also provide brief introductions specific to the contents of Chapter 7 and Chapter 8 in Section 7.2 and Section 8.2. The remaining chapters, except for the conclusion in Chapter 9, are derived from a series of published and unpublished papers and constitute the main contributions of the thesis.

In Chapter 2, we shall present some background material on the electronic structure problem. The challenges associated with the simulation of quantum phenomena are one of the core motivations for the field of quantum computing and for this thesis in particular. Chapter 4, Chapter 5, and Chapter 6 all deal with the application of NISQ quantum computing to quantum chemistry, and we shall introduce concepts in Chapter 2 that are later mentioned throughout these chapters. The field of quantum chemistry is large and we shall not attempt to present a thorough review. Instead, we shall aim for the more modest goal of discussing those threads that are directly relevant to the later chapters. For the most part, this means that we shall focus on introducing wavefunction based methods and discussing the drawbacks of existing approaches rather than their strengths.

In Chapter 3, we shall delve deeper into the formalism of quantum computing and present and overview of NISQ quantum computing in particular. The primary aim of this chapter shall be to introduce the notion of variational quantum algorithms and to set the stage for our concrete contributions by framing the challenges facing this subfield of quantum computation. In order to do so, we shall find it necessary to present a general introduction to quantum computing in Section 3.1 and a high-level review of current quantum computing hardware platforms in Section 3.2. In Section 3.6, we shall frame the challenges facing NISQ quantum computing in terms of two kinds of resources, the coherent time complexity (number of two-qubit gates) and the total time complexity (number of circuit repetitions). Most of the later chapters of this thesis can be understood in terms of attempts to make good use of these two limited resources, or to enable tradeoffs between them.

In Chapter 4, we shall explore a class of wavefunction ansatze for quantum chemistry on a quantum computer. We shall compare the performance of the standard unitary coupled cluster (UCC) approach with a generalized UCC that includes additional excitation operators. We shall also introduce a family of sparse doubles operators that allow for compact quantum circuit implementations. We shall variationally optimize these wavefunctions and compare their performance to classical benchmark calculations for the ground states and first excited states of H_4, H_2O, and N_2. We shall also show that the use of a specialized multi-determinental reference state is helpful for improving the excited state calculations.

In Chapter 5, we shall introduce a method for approximating the ground state of a system using a linear combination of parameterized quantum wavefunctions. We shall develop techniques for efficiently measuring the matrix elements between these wavefunctions and show how performing these pairwise measurements and solving a small generalized eigenvalue problem enables an extension to the usual variational quantum eigensolver formalism that

we shall introduce in Section 3.5. We shall explain how this extension enables the systematic improvement of a trial wavefunction without requiring additional coherent quantum resources (two-qubit gates) and present numerical data on the utility of this approach applied to wavefunctions developed in Chapter 4. We shall show how this approach results in a tradeoff, increasing the overall time required (number of circuit repetitions) for a fixed accuracy and we shall develop a technique to partially ameliorate this increased cost.

In Chapter 6, we shall introduce a new technique for measuring the expectation value of the quantum chemical Hamiltonian in the context of the variational quantum eigensolver. We shall explain how this technique, based on a factorization of the two-electron integral tensor, greatly reduces the overall time (number of circuit repetitions) required for measuring the energy of a wavefunction to within a fixed precision. In order to accomplish this, our approach requires the execution of a particular circuit before measurement, increasing the coherent quantum resources (number of two-qubit gates) required. We shall show how our approach enables an enhanced form of error mitigation and make the case that this benefit makes up for the additional complexity of the measurement operation and the potential for errors that the larger quantum circuit incurs.

In Chapter 7, we shall transition to a different topic and briefly introduce the idea of a tensor network before exploring a Monte Carlo strategy for parallelizing a particular tensor network algorithm. This chapter is set aside from the main body of the thesis in that it does not deal with quantum computing, but it is also connected in two basic ways. First of all, we have found the language of tensor networks helpful for reasoning about quantum circuits, and the inclusion of this chapter shall provide a good opportunity to discuss them. Secondly, like the rest of the thesis, this chapter shall focus on a tension between two different kind of resources. Echoing our work on making tradeoffs between coherent time complexity and total time complexity, this chapter shall explore a tradeoff between a serial mode of operation and a parallelizable one for a particular family of tensor network algorithms.

In Chapter 8, we shall apply the language of tensor networks introduced in Section 7.2 to the design of parameterized quantum circuits for machine learning. We shall briefly introduce some background on machine learning in Section 8.2 before describing our proposal. We shall show how a class of quantum circuits inspired by classical tensor networks are well suited to two different kinds of machine learning tasks. The structure of these circuits also enables us to perform simulations of our approach applied to a collection 64 qubit states representing the images of handwritten digits. We shall show that our approach is highly resilient to the kind of noise present on near-term devices.

In Chapter 9, we shall conclude by summarizing the work of the thesis and offering our perspective on the future on NISQ-era quantum computing. We shall focus on putting the previous chapters into the framework of the two resources, coherent time complexity (number of two qubit gates) and total time complexity (number of measurement repetitions), that we shall lay out in detail in Section 3.6. The quest to reduce the cost of near-term algorithms in terms of these two resources and to enable tradeoffs between them has been a theme running through our work, and we shall take this opportunity to make that through line explicit. Finally, we shall offer a few thoughts on the future of NISQ-era quantum computing.

Chapter 2

The Electronic Structure Problem

> The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.
>
> P. A. M. Dirac [21]

In this chapter, we shall review the aspects of the field of quantum chemistry necessary to understand the work of this thesis. As Dirac pointed out, the principle challenge facing this discipline is that we are generally unable to solve the necessary equations. The traditional approach taken is to make a series of approximations, eventually arriving at a problem that is numerically or analytically tractable. We review these approxiations, as well as some of the computational methods they enable, As we explain in later chapter, the tools of quantum computing may allow us to relax some of these approximations. The language and concepts introduced in our discussion of classical techniques in electronic structure will provide a foundation for our exploration of the quantum computational alternatives.

We shall first discuss the standard formulations of the electronic structure problem along with the approximations implicitly included in these formulations. Along the way, we shall explain why we usually focus on determining the ground state of a collection of electrons in a fixed external potential. We shall discuss two strategies for representing the Hamiltonian and explain how the second arises out of a desire to automatically incorporate the anti-symmetry of fermionic wavefunctions. This allows us to review a standard strategy for discretizing the electronic structure problem in order to obtain an eigenvalue equation involving finite-dimensional matrices.

Then, we shall review a selection of classical computational techniques related to the quantum computational approaches we shall present in later chapters. These techniques start with the Hartree-Fock approach, which is a 'mean-field' technique in the sense that we assume each electron experiences a potential derived by averaging over the positions of the other electrons. Hartree-Fock serves as a starting point for most of the other classical and quantum computational methods for the electronic structure problem discussed in this thesis. Following the discussion of Hartree-Fock, we shall briefly introduce some of the classical techniques that build upon it, with a focus on the configuration interaction and coupled cluster methods. Throughout this review, we shall attempt to point out where classical electronic structure techniques experience challenges.

Finally, we shall give a brief summary of how quantum computing might someday impact the field of electronic structure. Much of this thesis is concerned with the progress of near-term algorithms for quantum chemistry. However, the basic reason for the excitement around applying quantum computing to quantum chemistry has more to do with the techniques that will be available using more powerful quantum computers than we have available today. We find it helpful and motivating to review some of these techniques in this chapter, even if the focus of this thesis is on the more limited family of algorithms suitable for use on a near-term quantum computer.

2.1 Useful Approximations

We begin by making a series of standard approximations. The first of these is to treat the atomic nuclei as point particles. In reality, it would be more correct to treat nuclei as being composed of protons and neutrons. Even more fundamentally, we could consider the protons and neutrons themselves as composite particles. Making this approximation is rarely problematic when treating everyday matter at everyday temperatures and its relaxation is the domain of the field of nuclear chemistry [22].

The next approximation we make is to neglect relativistic effects. In situations where the kinetic energy of the electrons is small enough that their expected velocity is much smaller than the speed of light then this approximation is a good one [23]. Systems composed of lighter atoms satisfy this condition, but it can start to break down for heavier nuclei, depending on the accuracy required. When relativistic effects are small but non-negligible, it is possible to correct for them using perturbative approaches. Besides mentioning the fact that we are making an approximation by treating the non-relativistic case, we do not address relativistic electronic structure in this thesis.

We also make the Born-Oppenheimer approximation, neglecting the coupling between the nuclear and electronic degrees of freedom [24]. We qualitatively justify this approximation by appealing to the vastly different masses of electrons and nuclei. Because the nuclei are so much heavier than the electrons, they move much more slowly, and thus we can treat the electrons as if they exist in an external potential defined by a set of nuclei pinned in space. We therefore usually consider the electronic Hamiltonian and its eigenstates parametrically,

as functions of the nuclear coordinates. The resulting Hamiltonian is

$$\hat{H}(\vec{R}) = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{a=1}^{M}\frac{Z_a}{r_{ia}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}}, \tag{2.1}$$

where ∇^n is the Laplacian operator, N indicates the number of electrons, M indicates the number of nuclei, Z_a denotes the charge of the ath nuclei, r_{ij} is the distance between the ith and jth electron, r_{ia} is the distance between the ith electron and the ath nuclei, and we work in atomic units [24]. This approximation is known to perform extremely well in most situations, but also to break down for certain configurations of nuclei, such as in the vicinity of a conical intersection [25]. For the remainder of this thesis, we shall focus on situations where the Born-Oppenheimer approximation holds.

Typically, we are interested in the ground state or the low-lying eigenstates of the electronic structure Hamiltonian, i.e., the lowest eigenstates of the time independent Schrödinger equation,

$$\hat{H}\left|\psi\right\rangle = E\left|\psi\right\rangle. \tag{2.2}$$

This is because we expect a system in thermal equilibrium at some temperature T to be in the state

$$\rho = \frac{e^{-\beta\hat{H}}}{tr(e^{-\beta\hat{H}})}, \tag{2.3}$$

where $\beta = \frac{1}{k_b T}$. At ordinary temperatures, the gap between the ground state and the lowest-lying excited state of a typical electronic system is large compared to β, and therefore the occupation of the higher-lying states is exponentially suppressed. For example, consider the water molecule. Experimental photo-absorption spectroscopy indicated an energy gap between the ground state and first excited state of ≈ 7.447 eV (electron volts) [26]. At 70°F, Eq. 2.3 indicates that the population in the first excited state is suppressed by a factor of $\approx 3 \times 10^{-128}$ compared to the population in the ground state. Owing to this exponential suppression, we can often neglect the higher-lying electronic states entirely and understand many of the properties of a molecular system in terms of its electronic ground state. In particular, the ground state potential energy surface (PES), the ground state energy as a function of the nuclear coordinates, is a central object of study. Other properties of interest, such as the polarizability of a molecule, may formally depend upon characteristics of all of the eigenstates but be amenable to approximation using information derived from just the low-lying eigenstates [27].

We can calculate a wide range of useful quantities once we understand the potential energy surface. For example, if we neglect the quantum mechanical nature of the nuclei, then the equilibrium geometry for a molecule corresponds to a minima of the potential energy surface. Furthermore, we can understand certain types of chemical reactions by considering the combined ground state potential energy surface of the reactants. Through this lens, a chemical reaction is the traversal from one minima of the potential energy surface to another and is enabled when the barriers between the minima are sufficently small [28, 29]. Other

types of reactions might require an electronic excitation in order to be feasible [27], and understanding the properties of low-lying excited states can be useful in analyzing such reactions.

Unfortunately, even with these approximations and a focus on the ground state, exact solutions to the electronic structure problem are rarely available. Except for a handful of model systems, such as the Hydrogen atom, we do not have analytical solutions to Eq. 2.2. Therefore, we must turn instead to numerical methods. Before we are prepared to discuss exact or approximate numerical methods though, we shall discuss some subtleties relating to the representation of the electronic structure Hamiltonian of Eq. 2.1 and fermionic wavefunctions. Along the way, we shall see how we may arrive at a suitably discretized form of Eq. 2.2.

2.2 Fermionic Wavefunctions

2.2.1 First Quantization

For a one-electron system, like the Hydrogen atom, the Hamiltonian of Eq. 2.1 is the end of the story. In fact, for a system with only a single electron and a single nuclei, there are standard techniques for obtaining an analytical solution [27]. The Hamiltonian, as expressed in Eq. 2.1, is referred to as the first quantized electronic structure Hamiltonian. For a single electron system, first quantized wavefunctions are described by functions of the electron coordinates, $\vec{r}$. For a multi-electron system, a first quantized wavefunction is similarly a function of the coordinates of each of the electrons, $\{\vec{r}_i\}$. With more than one electron, it also becomes important to account for the spin (intrinsic angular momentum) of the electrons. We shall discuss spin further in Section 2.2.1.2. It is important to note, however, that not all properly normalized functions are valid wavefunctions in a many-electron system.

Because electrons are identical particles, any observable property of the system should remain the same when two electrons are exchanged or relabeled. For electrons, and other fermionic particles, this symmetry in the observables is enforced by an anti-symmetry in the wavefunction. For our purposes, we take it as an axiom that exchanging two electrons should cause the wavefunction to change in sign. The fact that fermions, particles with half-integer value of spin, have this anti-symmetry under exchange while bosons, particles with integer values of spin, are symmetric under exchange can be understood by appealing to the spin-statistics theorem [30], but we do not review this topic more deeply here.

The requirement that fermionic wavefunctions are anti-symmetric under exchange has profound consequences. Because the sign of the wavefunction must change when two identical particles are exchanged, it is impossible for two electrons to occupy the same state. This fact is famously known as the Pauli exclusion principle. This simple requirement gives rise to the shell structure of atoms, which in turn leads to the organization of the periodic table [31]. In our quest to analytically or numerically solve the time-independent Schrödinger equation we need to make sure that we only consider solutions that are properly anti-symmetric.

Fortunately, there is a natural way to build the anti-symmetry into our mathematical formalism. Let the states $|\phi_i\rangle$ be a (possibly infinite) set of single-particle wavefunctions that span the single-particle Hilbert space. As a specific example, we consider a three-electron wavefunction. Then states of the form $\phi_i(\vec{r_1}) \otimes \phi_j(\vec{r_2}) \otimes \phi_k(\vec{r_3})$ form a basis for states in the three-electron Hilbert space. In order to consider only those states which are properly anti-symmetric, we must take linear combinations of such wavefunctions. We can express the appropriate linear combinations as the determinant of a matrix,

$$\begin{vmatrix} \phi_i(\vec{r_1}) & \phi_j(\vec{r_1}) & \phi_k(\vec{r_1}) \\ \phi_i(\vec{r_2}) & \phi_j(\vec{r_2}) & \phi_k(\vec{r_2}) \\ \phi_i(\vec{r_3}) & \phi_j(\vec{r_3}) & \phi_k(\vec{r_3}), \end{vmatrix} \tag{2.4}$$

called a Slater determinant [32]. Exchanging two electrons is then equivalent to interchanging rows of this matrix, which induces the appropriate sign change in the determinant.

While the use of Slater determinants formally solves this problem, it does not eliminate the basic redundancy of first quantization. We are still representing many-particle wavefunctions explicitly using the coordinates of the separate electrons as if they were identical particles.

2.2.1.1 Second Quantization

The formalism of second quantization avoids this redundancy by not labeling the individual electrons. Rather than keeping track of the state of each electron, a second-quantized description keeps track of how many electrons are in each possible state. This is sometimes called the occupation number formalism and we frequently refer to these possible states as orbitals. When treating a Hilbert space with a finite number, N, of single-particle basis states, the Slater determinants that form a basis for the many-body Hilbert space can be represented by bit-strings of length N, e.g., $|0001010\rangle$. The full many-body Hilbert space is called Fock space.

Second quantization encodes the anti-symmetry of the wavefunction algebraically, and the primary objects of this formalism are the creation and annihilation operators. The creation operator $\hat{a}_i^\dagger$ is defined as the operator which adds an electron to the ith orbital. Its Hermitian conjugate, the annihilation operator $\hat{a}_i$, removes an electron from the same orbital. If there is no electron to remove, or if an operator would add an electron to an already occupied orbital, the result is instead zero. We define the vacuum state, $|vac\rangle$, as the normalized state with zero electrons. From this state, we build up arbitrary Slater determinants through the action of a product of creation operators.

We require that the creation and annihilation operators obey a set of anti-commutation relations,

$$a_i^\dagger, a_j^\dagger = 0 \qquad\qquad a_i^\dagger, a_j = \delta_{i,j}. \tag{2.5}$$

These anti-commutation relations enforce the anti-symmetry of the wavefunction. Because the creation operators commute with each other, the order in which we apply them to a state matters. Typically, we use the convention that a collection of creation operators acting

on the vacuum state should be put in increasing order to correspond with an occupation number string. Therefore, we have $|0001010\rangle = a_4^\dagger a_6^\dagger |vac\rangle$. We refer the reader to Ref. 24 for an extended discussion of the properties of creation and annihilation operators.

The electronic structure Hamiltonian expressed in second quantization looks somewhat different than the first quantized form of Eq. 2.1. In one standard notation, sometimes called the physicist's notation, we write it as

$$\hat{H} = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} g_{pqrs} a_p^\dagger a_r^\dagger a_s a_q + h_{nuclear}. \tag{2.6}$$

In this expression, we have absorbed all of the physics of the electrons and the nuclei into the definition of the coefficients h_{pq}, g_{pqrs}, and $h_{nuclear}$. Following Ref. [33], we define these coefficients through the following equations,

$$h_{pq} = \int \phi_p^*(\vec{x}) \Big(-\frac{1}{2}\nabla^2 - \sum_l \frac{Z_l}{r_l} \Big) \phi_q(\vec{x}) d\vec{x}, \tag{2.7}$$

$$g_{pqrs} = \int \int \frac{\phi_p^*(\vec{x}_1)\phi_r^*(\vec{x}_2)\phi_q(\vec{x}_1)\phi_s(\vec{x}_2)}{r_{12}} d\vec{x}_1 d\vec{x}_2, \tag{2.8}$$

$$h_{nuclear} = \frac{1}{2} \sum_{i \neq j} \frac{Z_i Z_j}{r_{ij}}. \tag{2.9}$$

In order to arrive at a numerically tractable Hamiltonian, we make a further approximation by discretizing space. Rather than working with an infinite set of single-particle basis functions which form a basis for three dimensional space, we instead choose a finite number of basis functions. Depending upon the type of problem, the appropriate choice can vary widely. For example, for solid state problems, we might take a set of plane waves, the eigenfunctions of the momentum operator, with quantum numbers regularly spaced up to some cutoff value of the momentum [34]. For atomic and molecular systems, a more typical choice is to use single-particle basis functions inspired by the atomic orbitals of the (exactly solvable) hydrogen atom, known as Slater-type orbitals [33].

Over the decades, much work has gone into constructing basis sets that yield physically meaningful results for chemical systems using a limited number of basis functions. There are a range of options available for the practitioner of numerical methods. Throughout this thesis, we tend to use relatively small basis sets in order to minimize the overhead for our numerical simulations. Our minimal basis set of choice is STO-3G, which approximates each Slater-type atomic orbital using a linear combinations of 3 Gaussian functions. We also employ the larger 6-31G and cc-PVDZ basis sets [35], which, besides using more basis functions per atom, attempt to improve upon smaller basis sets by performing optimization with respect to the accuracy of atomic Hartree-Fock calculations and more sophisticated molecular calculations respectively [33]. Regardless of the basis sets used for the specific calculations, we are ultimately interested in removing the impact of our finite basis set approximation by extrapolating to the infinite basis set limit.

2.2.1.2 Properties of Quantum Chemical Hamiltonians and Wavefunctions

Before proceeding, we find it useful to highlight some properties of the quantum chemical Hamiltonian and its eigenstates that we shall reference later in this thesis. In order to understand the details of the calculations we present in later chapters, we briefly review the concept of spin. We also touch on some of the symmetries of the Hamiltonian and its eigenfunctions that we shall use later to speed up our calculations and for other purposes.

We briefly mentioned the notion of spin when discussing fermionic anti-symmetry, but it deserves a more extended presentation. We take it as an axiom that electrons possess an intrinsic angular momentum, known as spin. This angular momentum is quantized with the quantum numbers $s = \frac{1}{2}$ and $m_s = +\frac{1}{2}$ or $m_s = -\frac{1}{2}$. Therefore, in addition to describing the spatial coordinates of electrons, we also need to include the spin-component of their wavefunctions. Starting with a given set of spatial orbitals (basis functions), $\{\phi_i\}$, we can obtain a set of basis functions which include the spin component of the single-particle wavefunctions, $\{\phi_i \otimes \alpha\} \cup \{\phi_i \otimes \beta\}$, where α denotes the spin state $\left|s = \frac{1}{2}, m_s = +\frac{1}{2}\right\rangle$ and β denotes $\left|s = \frac{1}{2}, m_s = -\frac{1}{2}\right\rangle$ [27]. Sometimes we use the notation $\uparrow$ and $\downarrow$ rather than α and β.

The quantum chemical Hamiltonian, defined in Eq. 2.6, commutes with the total spin operator, $\hat{S}^2 = \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2$ and individual components of the spin vector, $\hat{S}_x$, $\hat{S}_y$, and $\hat{S}_z$. In second quantization, we can express the components of the spin vector as

$$\hat{S}_x = \frac{1}{2} \sum_p \left(\hat{a}_{p\alpha}^\dagger \hat{a}_{p\beta} + \hat{a}_{p\beta}^\dagger \hat{a}_{p\alpha} \right), \tag{2.10}$$

$$\hat{S}_y = \frac{1}{2} \sum_p \left(\hat{a}_{p\alpha}^\dagger \hat{a}_{p\beta} - \hat{a}_{p\beta}^\dagger \hat{a}_{p\alpha} \right), \tag{2.11}$$

$$\hat{S}_z = \frac{1}{2} \sum_p \left(\hat{a}_{p\alpha}^\dagger \hat{a}_{p\alpha} - \hat{a}_{p\beta}^\dagger \hat{a}_{p\beta} \right), \tag{2.12}$$

where we use a standard convention of indicating the spatial component of the orbital indices with roman letters and the spin component with greek letters. The individual components of the spin vector don't commute with each other, so we typically choose to use the eigenvalues of the operators $\hat{S}^2$ and $\hat{S}_z$ as quantum numbers, which we denote with the symbols s and m_s. Because $\hat{S}^2$ and $\hat{S}_z$ commute with the Hamiltonian, we can simultaneously diagonalize all three operators. This allows us to choose an eigenbasis for the Hamiltonian where the states have definite values for s and m_s, as well as for energy.

The Hamiltonian also commutes with the total particle number operator,

$$\hat{N} = \sum_{p,\alpha} \hat{a}_{p\alpha}^\dagger \hat{a}_{p\alpha}. \tag{2.13}$$

The eigenstates of this operator are the states with a definite number of electrons and its eigenvalues correspond to the number of electrons in each eigenstate. Because the Hamiltonian also commutes with this operator, its eigenvalues can be used to label the eigenstates of the

Hamiltonian. Furthermore, as with $\hat{S}^2$ and $\hat{S}_z$, the Hamiltonian can be expressed as a direct sum over the different symmetry sectors. This can allow us to speed up computations by only considering the action of the Hamiltonian on a particular symmetry subspace. We shall make use of this fact later.

Another notable property of the quantum chemical Hamiltonian is that its matrix elements can be taken to be real. This is often true by construction for the single-particle basis sets used in practice. One can also show, in general, that there always exists a change of basis that allows for the matrix elements of the quantum Hamiltonian to be real. This fact is related to the time-reversal symmetry of the Hamiltonian, and is explained well in Chapter 4 of Ref. 36. When the Hamiltonian is represented using real matrix elements, it is always possible to choose an eigenbasis formed by taking linear combinations of the basis states with real coefficients. One can see that this must be true for non-degenerate eigenstates by noting that, with a real Hamiltonian and a real eigenvalue, the real and imaginary parts of any eigenvector must both satisfy the eigenvalue equations on their own. Because of the assumption of non-degeneracy, these real and imaginary solutions must represent the same state. The degenerate case can be reduced to the non-degenerate one by adding a commuting perturbation to the Hamiltonian that breaks the degeneracies.

2.3 Classical Techniques and Their Limitations

Despite the approximations we have described so far, determining the ground state of Eq. 2.6 remains challenging. The most straightforward approach is the full-configuration interaction (FCI) method, also known as exact diagonalization. Under this approach, we use standard numerical linear algebra tools to determine the lowest eigenstate of the Hamiltonian matrix, usually in a subspace corresponding to a particular set of quantum numbers for the available symmetries. Unfortunately, the number of basis states scales exponentially with the system size, quickly rendering this method unafforable for even modestly-sized molecules.

Furthermore, we often need more than a single solution in a single basis set. We may want to understand the whole potential energy surface, which requires finding the ground state energy across a range of nuclear coordinates. Besides this, the FCI energy within a fixed basis set is only exact within that model chemistry. If we are interested in comparison to experiment we need to converge the properties of interest to the infinite basis set limit, and this can require a series of calculations in larger basis sets. All of these desiderata pose a challenge when a single calculation might already be prohibitively expensive.

Therefore, we often use approximate methods to solve Eq. 2.6. It can sometimes be helpful to think about a hierarchy of methods, organized by the way the cost of the method scales with system size. As we shall review below, the Hartree-Fock method can enable very good approximate solutions at a cost that scales cubically with system size. At the upper end of the hierarchy, we have the exponentially scaling FCI method. Between these two extremes are approaches whose cost scales polynomially with system size but more rapidly than Hartree-Fock. We sometimes refer to the solutions generated by a particular method

as a 'model chemistry.' The extent to which a model chemistry corresponds to reality can vary, but it is a desirable property of an approximate method that it behaves consistently for a variety of problems. This enables the accurate calculation of relative energy differences between the ground states of different Hamiltonians.

The work in this thesis builds mainly upon a specific branch of electronic structure techniques which we shall review here. This family of approaches starts with the Hartree-Fock method and includes a range of post-Hartree-Fock approaches that aim to recover most of the difference between the approximate Hartree-Fock solution and the exact (within the chosen basis set) FCI solution. These methods are a particular example of so-called "wavefunction methods," which focus on approximating the ground state wavefunction.

2.3.0.1 Hartree-Fock and Beyond

We explained in Section 2.2.1 how any anti-symmetric wavefunction is naturally represented as a linear combination of Slater determinants. The essence of the Hartree-Fock method is to find the best single determinant approximation to the ground state. One perspective, which we shall explain in more detail below, is that Hartree-Fock is essentially a mean-field approach that accounts for anti-symmetry. We refer the reader to Refs. 24 and 33 for a more thorough treatment of the Hartree-Fock approach and focus here on providing the necessary intuition for the work of the body of the thesis.

Having fixed the functional form of the wavefunction to be a single Slater determinant, the remaining degrees of flexibility lie in the choice of the single-particle basis functions. If we assume that we start with an orthonormal set of N single-particle basis functions $\{\phi_i\}$, then we can rotate to any other orthonormal basis through the action of an $N \times N$ unitary matrix U,

$$\tilde{\phi}_i = \sum_i U_{ij}\phi_j. \tag{2.14}$$

We can think about an arbitrary Slater determinant with η electrons on N spin-orbitals as arising from the rotation of some fixed reference determinant into a new basis. Actually, rotations between the occupied orbitals and the virtual orbitals are redundant, so U contains more parameters than we need, but the intuition is still appropriate. With this understanding, we can define the Hartree-Fock wavefunction as the determinant with the lowest energy across all possible choices for the parameters.

This process can be performed efficiently in most cases by repeatedly solving an eigenvalue problem involving matrices whose size scales linearly with the system size. This implies an time complexity of $O(N^3)$ (per iteration) for a system with N orbitals. Specifically, the Hartree-Fock orbitals may be obtained by self-consistently diagonalizing the Fock operator,

$$\hat{f} = \hat{h} + \hat{V}. \tag{2.15}$$

In the above equation, $\hat{h}$ is the usual one-body term from the second-quantized Hamiltonian

of Eq. 2.6 and

$$\hat{V} = \sum_{pq} \Big(\sum_{i} g_{pqii} - g_{piiq} \Big) a_p^\dagger a_q, \tag{2.16}$$

where g_{pqii} and g_{piiq} are defined in Eq. 2.8, with the understanding that both operators are expressed in terms of the Hartree-Fock orbitals. The Hartree-Fock orbitals are themselves defined by the eigenvalues of the Fock matrix in the one-electron Hilbert space, and so the solution is usually obtained by an iterative procedure which is repeated until a self-consistent solution is obtained.

The $\hat{V}$ term in Eq. 2.15 represents the average potential that one electron experiences as a result of the others. This is the sense in which Hartree-Fock is a mean-field approach. Rather than accounting for the electron-electron interactions completely, the Hartree-Fock approximation replaces the two-electron term in the Hamiltonian by a term that accounts merely for the average influence of the electrons on each other. This point of view explains why the Hartree-Fock equations must be solved self-consistently. The average impact of the other electrons depends upon the wavefunction, which itself depends upon the average impact of the other electrons.

Once a solution has been obtained, we obtain the Hartree-Fock wavefunction by filling the η lowest energy single-particle basis functions resulting from diagonalizing Eq. 2.15 with electrons. The Hartree-Fock energy is the sum of the eigenvalues corresponding to those orbitals. The variational principle guarantees that this energy is an upper bound to the true ground state energy [24], i.e., that

$$\left\langle \hat{H} \middle| \phi \middle| \hat{H} \right\rangle \geq E_{FCI}. \tag{2.17}$$

The Hartree-Fock (HF) wavefunction and energy are so central to traditional electronic structure theory because they often form very good approximations to the true ground state wavefunction and energy. We define the correlation energy as the gap between the FCI energy and the HF energy. For the equilibrium geometries of molecules made up of light atoms, the correlation energy is often 1% or less of the total energy [33].

The Hartree-Fock approximation serves as a starting point for a collection of "post-Hartree-Fock" methods that attempt to approximate the correlation energy. Because Hartree-Fock is a mean-field theory, it fails to capture the effects of electronic correlation beyond those implied by anti-symmetry. When these effects are primarily small contributions from a number of other determinants, we often use the phrase "dynamic correlation." One natural approach to account for these small corrections to the wavefunction, and hence, the energy, is perturbation theory. In many cases, second order Møller Plesset perturbation theory provides a straightforward and affordable way to recover most of the correlation energy. In other cases though, we instead have "static correlation," where the ground state requires a superposition of several determinants with significant weight, no matter what single-particle basis is used.

2.3.0.2 Configuration Interaction and Exact Diagonalization

We have already discussed full configuration interaction (FCI), also known as exact diagonalization, as a solution to the electronic structure problem. When this method is too expensive to afford, but the Hartree-Fock approximation does not provide sufficient accuracy, a truncated configuration interaction (CI) calculation can provide a natural middle ground. In a truncated CI calculation, we pick a reference state (most often the Hartree-Fock determinant) and generate a collection of additional determinants by considering excitations from that state. For example, we might take all of the single and doubly excited determinants, i.e., the determinants formed by removing one or two electrons from the "occupied orbitals" (the orbitals where the reference determinant has electrons) and placing them in the "unoccupied orbitals" (the orbitals where the reference determinant does not have electrons). We then take this collection of states and diagonalize the Hamiltonian projected onto that subspace, or at least, we solve for it's ground state and ground state energy. This approach, where we take the singly and doubly excited determinants, is referred to as configuration interaction singles and doubles (CISD). As higher and higher orders of excitation are included in the calculation, the ground state and ground state energy converge to the FCI value, as does the cost of performing the calculation.

Another route to reducing the cost of a quantum chemical calculation is to solve the problem exactly in a small active space. In this approach, we begin with the Hartree-Fock state (or some other reference determinant) and designate some fraction of the orbitals as the active space. We refer to the occupied orbitals outside of the active space as the core orbitals and the unoccupied orbitals outside of the active space as the virtual orbitals. For those orbitals within the active space, we consider the collection of determinants that can be formed by exciting electrons from the occupied to the unoccupied orbitals. We then perform an FCI calculation using this smaller set of determinants. The resulting wavefunction is known as a complete active space (CAS) wavefunction [37]. This approach, of focusing on solving the electronic structure within an active space, can also be followed when using a non-exact method.

Frequently, quasi-exact solutions in an active space are combined with other tools to improve the overall accuracy of the calculation. For example, the complete active space self-consistent field theory (CASSCF) model involves the simultaneous optimization of the coefficients within the active space as well as the underlying single-particle basis [33, 37]. This offers a potential improvement over simply using the Hartree-Fock orbitals, which are only optimized to give the best single-determinant solution. Perturbative approaches, such as complete active space second-order perturbation theory (CASPT2), which attempt to recover dynamical correlation outside of the active space also have an important role to play in the treatment of certain problems [38]. In this thesis we shall mostly focus on using quantum algorithms to solve active space problems for small model systems but we expect the integration of these techniques with more sophisticated tools from classical electronic structure to become important as the field matures.

Truncated configuration interaction methods lack a desirable property known as size

consistency [39]. Imagine two Hydrogen molecules, infinitely separated so that there are no interactions between them. Because these molecules have only two electrons, CISD can find the exact ground state of either molecule individually, even as their bonds are stretched and there is some contribution from doubly excited determinants to the ground state. However, if we try to apply the CISD method to find the ground state of both molecules together, it is no longer an exact method because we neglect contributions from configurations with simultaneous double excitations on both molecules. Despite the seemingly reasonable scheme for truncating the CI space, CISD fails to return an answer for calculations involving two non-interacting systems consistent with the answers it would return on the systems considered individually. This failure is known as a lack of size consistency, and it is a problem that all CI methods besides FCI face [33, 39]. This can make it challenging to use calculations to correctly predict physical properties, which might naturally involve energy differences between calculations performed on a composite system and calculations performed on smaller molecular fragments.

2.3.0.3 Coupled Cluster

Coupled cluster methods can be seen as a natural solution to the problem of size consistency in truncated CI methods. Let i, j denote the indices of occupied orbitals, a, b denote unoccupied orbitals, and $|\phi\rangle$ denote the Hartree-Fock state. Then a CISD wavefunction takes the form

$$|\psi\rangle = \hat{T}\,|\phi\rangle\,, \qquad \hat{T} = \sum_{ij} t_i^a a_a^\dagger a_i + \sum_{ijab} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_j a_i, \tag{2.18}$$

where the t_i^a and t_{ij}^{ab} are the free parameters. The basic idea of coupled cluster is to instead use an exponential ansatz with the form

$$|\psi\rangle = e^{\hat{T}}\,|\phi\rangle\,. \tag{2.19}$$

The coupled cluster singles and double wavefunction (CCSD) uses the same set of excitation operators as the CISD wavefunction ansatz, but it uses them in a much different way, as the components of the "cluster operator," $\hat{T}$. We do not prove it here, but we note that coupled cluster with a truncated cluster operator theory yields size-consistent energies, unlike configuration interaction [40]. One can see how this might be true by considering the Taylor series expansion of the exponential; exponentiating the cluster operator naturally includes products of lower order excitations.

Typically, coupled cluster wavefunctions are not used as variational ansatze. If the energy of a coupled cluster wavefunction were evaluated by the usual expression,

$$\frac{\left\langle \hat{H}\middle|\psi\middle|\hat{H}\right\rangle}{\langle\psi|\psi\rangle}, \tag{2.20}$$

the resulting energy would be an upper bound to the true ground state energy. An efficient scheme for evaluating the energy in this way would be desirable, but in general is not

available [41]. Instead, classical approaches to coupled cluster theory follow a different approach that does not explicitly yield a normalized wavefunction or a variational estimate of the energy. Instead of using Eq. 2.20, they take the eigenvalue equation,

$$\hat{H}e^{\hat{T}}\phi = Ee^{\hat{T}}\phi, \tag{2.21}$$

as a starting point. Then, they multiplying from the left by $e^{-\hat{T}}$ and project onto a determinant $|\phi_{proj}\rangle$, yielding

$$\langle\phi_{proj}|\, e^{-\hat{T}}\hat{H}e^{\hat{T}}\,|\phi\rangle = E\,\langle\phi_{proj}|\, e^{-\hat{T}}e^{\hat{T}}\,|\phi\rangle\,, \tag{2.22}$$

$$\langle\phi_{proj}|\, e^{-\hat{T}}\hat{H}e^{\hat{T}}\,|\phi\rangle = E\delta_{\phi,\phi_{proj}}. \tag{2.23}$$

By varying the determinant $|\phi_{proj}\rangle$, a collection of equations can be obtained, which are simplified by writing $e^{-\hat{T}}\hat{H}e^{\hat{T}}$ as a series of nested commutators.

Ultimately, a set of non-linear coupled equations must be solved to obtain the correct amplitudes and the coupled cluster energy. This process is non-variational and can sometimes fail dramatically, especially when used to study bond dissociation or other strongly correlated phenomena [42]. In practice, coupled cluster works extremely well, especially for "single-reference" systems whose ground states are dominated by a single determinant with corrections to account for dynamical correlation. Coupled cluster singles and doubles with perturbative corrections to account for triple excitations, CCSD(T), is often called the "gold-standard" of classical electronic structure methods. CCSD scales with the sixth power of system size, while CCSD(T) scales with the seventh power, limiting the applicability of these methods to larger systems [42]. A substantial amount of research effort has gone into developing more affordable approximations to these basic methods, or to otherwise tailoring them for specific applications [43–50].

2.4 The Promise of Quantum Computing for Quantum Chemistry

Despite many decades of work, the electronic structure problem remains intractable, or at least, challenging, for a range of interesting systems [70–75]. Understanding the low-energy landscape of strongly-correlated materials and molecules well enough to predict their properties can be extremely challenging [70–73]. This is particularly true, and also industrially relevant, when heavier atoms are involved in catalytic processes [72, 73]. In other situations, strong light-matter interactions can make the use of traditional computational tools challenging [74, 75].

Quantum computing promises to one day add powerful tools to the electronic structure toolbox. For example, efficient quantum algorithms for simulating time evolution by the electronic structure Hamiltonian are known [2, 76–79]. This is in contrast with the situation for classical computational techniques for time evolution, which are limited to small system

Method	Abbreviation	Original Reference(s)	Additional References
Hartree-Fock	HF	[51–53]	[24, 33]
Møller-Plesset Perturbation Theory	MP	[54]	[33]
Configuration Interaction	CI	[32]	[33]
Complete Active Space Self-Consistent Field theory	CASSCF	[37]	[33]
Coupled Cluster	CC	[55–57]	[33]
Density Functional Theory	DFT	[58, 59]	[60]
Diffusion Monte Carlo	DMC	[61, 62]	[63]
Auxiliary Field Quantum Monte Carlo	AFQMC	[64, 65]	[66]
Density Matrix Renormalization Group	DMRG	[67, 68]	[69]

Table 2.1: A collection of some important classical computational methods for the electronic structure problem. We do not present an exhaustive list, but merely attempt to mention some of the important approaches in the field. In some cases, there is a subjectivity in the choice of the original reference papers. Any exclusion of method or reference on our part is likely to reflect a lack of perspective rather than an intentional omission.

sizes or short times even with sophisticated algorithmic approaches [80–83]. Combined with the quantum phase estimation algorithm, these quantum algorithms for time evolution allow for direct measurement of the Hamiltonian operator, projecting a trial wavefunction into the eigenbasis of the Hamiltonian [76, 84]. More generally, quantum computers appear to be capable of efficiently performing a larger set of tasks than classical computers, and it is natural to ask how quantum chemical calculations might benefit from this fact [85].

The phase estimation algorithm has been the focus of much of the early excitement about the possibility of quantum computing for quantum chemistry. We shall explain it briefly here and review the algorithm more formally in Section 3.1.3. In Ref. 2, Lloyd showed how a quantum computer can be used to approximately simulate the real-time evolution of an arbitrary system. The quantum phase estimation algorithm is essentially the combination of real-time evolution and the quantum Fourier transform in order to perform interferometric measurement [2, 76, 86]. The quantum Fourier transform of Ref. 84 allows for the measurement of the relative phase accrued by applying a unitary operator to a superposition of states. This primitive can be used in conjunction with the ability of a quantum computer to simulate time evolution to measure the phase caused by the application of the time-evolution operator, e^{-iHt}. When the phase is resolved with enough precision, the possible outcomes of this measurement yield the eigenvalues of the Hamiltonian and the post-measurement states correspond to the eigenvectors. In Ref. 76, Aspuru-Guzik et al. showed that the use of this approach for quantum chemistry yields a polynomially scaling approach to determining the ground state energy of a molecule, provided that a state with sufficient overlap with the true ground state can be prepared. Fortunately, it is often straightforward to use existing classical approximation techniques to prepare such a reference state even without knowing how to fully charectarize the true ground state [87]. A quantum computer capable of executing this algorithm would therefore open up new chemistry and physics to computational study [88].

More generally, the ability to efficiently simulate real-time evolution, sometimes referred to as Hamiltonian simulation, may someday offer enormous promise. As we described above, the state of the art for real-time evolution using classical algorithms is limited to small systems or severe approximations [80–83]. Circumventing these limitations could allow for the understanding, and eventual control, of non-equilibrium quantum phenomena. For example, there has been experimental interest for decades in using coherent light to control chemical reactions, but a lack of computational techniques to complement the capabilities of experiment has handicapped this line of research [74]. As another example, consider the recent experimental evidence for transient, light-induced, superconducting behavior presented in Ref. 89. Accurate quantum-mechanical calculations of the real-time interaction between the light field and the material in question could help us understand such an effect and eventually leverage it to create new materials that can be reliably driven to a superconducting state at higher temperatures. Hamiltonian simulation also has implications for fields beyond quantum chemistry. There are fundamental questions about quantum many-body phenomena in condensed-matter physics, high-energy physics, and quantum field theory that might be addressable if real-time evolution could be performed tractably [90]. Algorithmically, real-time evolution has other uses besides the role that it plays in the quantum phase estimation

algorithm. For instance, real-time evolution can be used for state preparation by simulating adiabatic evolution under a slowly changing time-dependent Hamiltonian [91].

At the present day, however, the kind of quantum computers that can reliably perform these tasks are not available. Even when restricted to the simulation of model Hamiltonians where the simplicity of the model allows for a reduced cost, the most recent estimates require $\approx 10^9$ operations or more to perform calculations that might be challenging for classical computers [78]. As discussed briefly in Section 1.1, today's devices struggle to implement far fewer operations without error. Therefore, we turn our focus towards algorithms that are more suitable for today's and tomorrow's noisy devices.

Chapter 3

Noisy Intermediate-Scale Quantum Computing

> It is important to realize that we will need significant advances — in basic science as well as in systems engineering — to attain fully scalable fault-tolerant quantum computers.
>
> John Preskill [10]

The road between current hardware platforms and a fully error-corrected quantum computer is a long one. The clearest roadmap for constructing such a fault-tolerant device involves applying quantum error correction to systems with hundreds of thousands of physical qubits [92]. However, the most advanced machines that currently exist have fewer than 10^2 qubits [7, 93, 94]. Furthermore, the individual components of a hypothetical error-corrected device will have to be made at least somewhat more reliable than those of these existing smaller hardware platforms [10]. The term Noisy Intermediate-Scale Quantum (NISQ) was coined by in Ref. 10 to describe quantum computational devices large enough to do something that isn't trivially simulable but not powerful enough to support fault-tolerant error correction.

Before turning to the main topic of this thesis, the use of these pre-fault-tolerant machines, we shall begin by reviewing the formalism of quantum computing. We shall focus on the gate-based model of quantum computing, defining qubits and quantum gates. We shall also introduce some notation that will be useful throughout the rest of the thesis, including the commonly used diagrammatic notation for quantum circuits. We shall explain how larger quantum programs are constructed from a small set of operations, and introduce some of the strategies for quantum circuit compilation that the later chapters will leverage. We shall include a discussion of measurement and the estimation of observables. Finally, we shall

present an overview of the quantum phase estimation algorithm so that we will be able to compare it later with approaches more suitable for near-term quantum computers.

We shall provide a summary of the different leading hardware platforms for quantum computing before transitioning to a more detailed overview of the noise present on these devices. As of today, there are various competing technologies all capable of supporting universal quantum computing. We shall focus on devices that are roughly equivalent to the idealized quantum circuit model, which we shall introduce in Section 3.1, but we shall see that different hardware platforms offer different tradeoffs and challenges. All of them, however, face some degree of noise. This noise can come from imperfections in the construction or the control of the hardware, or it can come from unwanted environmental interference. Regardless of the ultimate sources of this noise, we shall mainly focus on qualitatively modeling it using phenomenological approximations.

Variational quantum algorithms represent a promising approach to ameliorating the challenges posed by this noise, and are the principle focus of this thesis. The phrase "hybrid quantum-classical" is sometimes used to describe these algorithms (as well as a handful of other proposals), but we eschew this term because all quantum computation is likely to involve interaction with classical computational resources. We shall introduce variational quantum algorithms, focusing first on the work of Peruzzo et al. in Ref. 17 which introduced the concept. Although this thesis focuses primarily on applications in quantum chemistry, with a small detour into machine learning in Chapter 8, variational quantum algorithms have been proposed for a wide range of different uses. We shall take the opportunity to highlight some of the interesting developments in the field, even if they only appear later in the body of the thesis in passing.

We shall then review the variational quantum eigensolver of Ref. 17 in detail in order to lay the groundwork for the main body of the thesis. The variational quantum eigensolver proposes to approximate the ground state wavefunction of a given Hamiltonian using a parameterized quantum circuit as an ansatz. We shall explain how wavefunctions and operators for a collection of indistinguishable fermions may be mapped to the Hilbert space of a collection of qubits. We shall discuss the design of these parameterized circuits and some of the concerns that arise therein. We also spend some time addressing the estimation of observables in this context, building on our treatment of measurement in Section 3.1. As we shall explain, the variational quantum eigensolver approach, and variational quantum algorithms in general, typically involve the optimization of parameterized circuits. We shall spend some time reviewing this optimization aspect of the variational quantum eigensolver and explaining where it faces challenges.

Finally, we shall segue into a broader discussion of our perspective on the challenges faced by variational quantum algorithms and NISQ computing in general. We shall explain how we think of these challenges primarily in terms of the constraints on two types of resources. The first resource being the number of two-qubit gates that can be ran before too many errors accumulate and the second resource being the overall number of circuit repetitions required for the algorithm. We shall review some of the efforts that have been made to address these challenges, setting the stage for the later chapters where we present our contributions in this

direction.

3.1 The Formalism of Quantum Computing

3.1.1 Quantum Gates in the Circuit Model

In this thesis, we shall focus exclusively on the circuit model of quantum computation [95]. As is the case with classical computing, other models of quantum computation are known. These alternatives, including the quantum Turing machine [85], adiabatic quantum computation [96], and measurement based quantum computing [97], are all polynomially equivalent to the circuit model. That is to say, it is known that a circuit model quantum computer can simulate any of these alternative models of quantum computation, and vice-versa. We focus on the quantum circuit model both for ease of understanding, and because most near-term hardware platforms are naturally describable in this language.

The fundamental objects of a circuit model quantum computer are qubits, the quantum analog of the classic bit. A qubit is a two-level quantum system whose basis states are typically labeled $|0\rangle$ and $|1\rangle$. The Pauli operators X, Y, and Z, combined with the two-qubit identity matrix, I, form a useful basis for the operators acting on such a two-level system. Expressed as matrices in the computational basis, they are

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.1}$$

We form the Hilbert space for a collection of quantum bits by taking the tensor product of their individual Hilbert spaces. One natural basis for this larger space is the "computational basis", defined to be the states $\bigotimes_i |q_i\rangle$, where $q_i \in \{0, 1\}$. Similarly, a natural basis for operators acting on multiple qubits is obtained by taking tensor products of the single-qubit Pauli operators and identity matrix.

The state of the quantum computer is manipulated by the application of quantum gates, unitary operations acting on the qubits. A quantum program on a quantum computer is ultimately specified as a series of quantum gates and measurement operations, which we shall discuss below. Typically, it is assumed that the device starts in a simple state, such as the zero state, $\bigotimes_i |0\rangle$. The usual approach is to define a fixed set of one and two-qubit gates and to construct the desired operation from these building blocks. Sometimes our construction of these operations involves the temporary use of extra qubits, which we call ancilla qubits. It turns out that such a small set of one and two-qubit gates can be sufficient to approximate any unitary operation on the whole collection of qubits to an arbitrary precision [98].

When a given collection of gates suffices to (approximately) generate any unitary, we call it a universal set of gates [86, 98]. One commonly referenced set of gates that have this property are the controlled NOT (CNOT), Hadamard (H), Phase (S), and T gates, defined

below [99],

$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad S = \begin{bmatrix} 1 & 0 \\ 0 & e^{\frac{i\pi}{2}} \end{bmatrix}, \quad T = \begin{bmatrix} 1 & 0 \\ 0 & e^{\frac{i\pi}{4}} \end{bmatrix}. \tag{3.2}$$

Note that we use the standard ordering of the basis states in the specification of the above matrices, $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ for the two-qubit gates and $|0\rangle, |1\rangle$ for the single-qubit gates. Notice that the controlled NOT gate, named because it acts to flip the state of the second qubit if the first qubit is in the $|1\rangle$ state, is the only two-qubit gate required for this set to be universal. In the context of NISQ computing, we often consider the CNOT gate together with arbitrary single qubit gates as our fundamental gate set, which is therefore also universal as it includes the discrete set mentioned above. Alternatively, we may instead focus on a universal set of gates that are natively implemented by a specific hardware platform.

Of course, even though a small set of one and two-qubit gates may suffice to generate any desired unitary, finding the sequence of gates that carries out a specific multi-qubit unitary may be challenging and there is no guarantee that the sequence must be short. The Solovay-Kitaev theorem guarantees that we can approximate any unitary on n qubits to within a precision ϵ (measured by operator norm) using at most $O(4^n log^c(\frac{1}{\epsilon}))$ gates from our universal set [100]. In the original proof, $c = 3.97$, but it has since been shown that specific families of gates exist with the property that $c = 1$ [101]. It is possible to see by a counting argument that the dependence on the number of qubits must be exponential [86]. Nevertheless, it is useful to be assured that we can efficiently approximate arbitrary unitary operators on a fixed number of qubits. Importantly, we note that the Solovay-Kitaev theorem can also be used constructively to find such decompositions [102].

Other strategies exist for constructing certain kinds of multi-qubit unitaries from smaller constituent parts, a task we sometimes refer to as compilation. One approach that shall come up throughout this thesis is the use of product formulas, specifically, the Lie-Suzuki-Trotter formula [103],

$$e^{t\sum_i A_i} = \prod_i e^{tA_i} + O(t^2). \tag{3.3}$$

Sometimes referred to colloquially by the shorter name "Trotter Formula," or the moniker "Trotterization," approaches based on product formulas date back to the earliest concrete proposals for simulating Hamiltonian evolution using a universal quantum computer [2]. Higher-order formulas that are correct up to arbitrary powers of t exist, but we shall only require the first order formula given above for the work in this thesis [104]. In the recent years, a variety of more sophisticated approaches have been developed for implementing complex multi-qubit unitary operations, with a special focus on the problem of performing time evolution for a variety of Hamiltonians [105–109]. Because this thesis focuses on algorithms suitable for near-term, noisy, quantum hardware, we shall not review these other approaches here.

3.1.2 Measurement

The other essential operation of the gate-based quantum computer is measurement. In quantum mechanics, one way of defining measurement is to use a collection of measurement operators, $\{M_i\}$, such that $\sum_i M_i^\dagger M_i = I$. A measurement of the state ρ yields outcome i with a probability

$$p_i = tr(M_i \rho_i M_i^\dagger), \tag{3.4}$$

and the subsequent post-measurement state is

$$\frac{M_i \rho M_i^\dagger}{tr(M_i \rho M_i^\dagger)}. \tag{3.5}$$

Frequently, we talk about the measurement of some Hermitian observable, O. We can write O in its eigenbasis as

$$O = \sum_i \lambda_i P_i = \sum_i \lambda_i P_i^\dagger P_i, \tag{3.6}$$

where the λ_is are the eigenvalues of O and the P_is are the projectors onto the associated eigenspaces. Because O is Hermitian, we have $\sum_i P_i^\dagger P_i = I$. Therefore, the projectors $\{P_i\}$ can take the role of the measurement operators $\{M_i\}$ and we see that the measurement of O is just a special case of the more general measurement. We sometimes refer to this special case as projective measurement. Ref. 86 contains a more detailed discussion of the formalism for both the general case and the special case of projective measurement.

In the circuit model of quantum computation, we can reduce a measurement of either type to the more specific operation of measuring one or more qubits in the computational basis [86]. Formally, we can consider measuring the qubits indexed $1, 2, \cdots N$ by taking the M_i operators described above to be

$$\{M_i\} = \{|i_1 i_2 \cdots i_N\rangle\langle i_1 i_2 \cdots i_N|\}, \qquad i_j \in \{0, 1\}, \tag{3.7}$$

e.g., the projectors onto the computational basis states of the qubits labelled 1 to N. Any general measurement on a circuit model quantum computer can be performed by applying the appropriate unitary and performing such a computational basis measurement [86]. This way of looking at measurement also lines up naturally with the capabilities of many hardware platforms, where single-qubit measurement in a particular fixed basis is the most natural type of measurement to perform.

When the objective of measurement is to evaluate an expectation value, statistical noise necessarily arises. If we have many copies of the state ρ and wish to estimate $\langle O \rangle = tr(\rho O)$, we can apply the measurement formalism from the paragraphs above. For each copy, when we perform a measurement of O we obtain one of the eigenvalues of O, λ_i, with probability p_i. Therefore, we have a classical random variable whose mean is equal to $\langle O \rangle$. Furthermore, the variance of this random variable is the natural generalization of the classical definition, $\sigma^2 = \langle O^2 \rangle - \langle O \rangle^2$. When we use samples from this random variable to estimate its mean, i.e.,

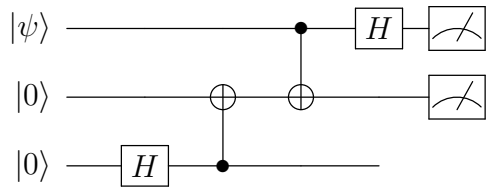

Figure 3.1: An example of a quantum circuit diagram. Each qubit is indicated by a horizontal line. Time progresses from left to right. The initial states of the three qubits are indicated by the symbols on the left, $|\psi\rangle$, $|0\rangle$, and $|0\rangle$. The boxes with letters inside indicate quantum gates, with H being an abbreviation for the Hadamard gate of Eq. 3.2. The small filled circles and larger open circles joined by lines indicate CNOT gates (also defined in Eq. 3.2), with the control qubit denoted by the small filled circle and the target qubit denoted by the larger open circle with the + sign inside. The two symbols at the right-hand side of the circuit, after the wires for the first two qubits terminate, denote measurement in the computational basis.

when we use measurements of O to estimation $\langle O \rangle$, we average the individual measurement outcomes and obtain an estimator of $\langle O \rangle$ with a variance σ^2/M, where M is the number of individual measurements we perform. Actually, it turns out that one is able to obtain an estimator with a variance that scales as σ^2/M^2 by performing the measurements coherently using techniques based on quantum amplitude estimation [110]. This approach is mostly unsuitable for NISQ devices but we mention it here for completeness.

3.1.3 The Quantum Phase Estimation Algorithm

In Section 2.4, we mentioned that the quantum phase estimation algorithm allows one to perform a measurement in the eigenbasis of the Hamiltonian. In this section, we shall explain this algorithm in deeper detail, although we refer the reader to Ref. 86 for a thorough review. We shall focus our explanation on the simplest case, where we have access to an exact eigenstate $|\psi\rangle$ of some unitary operator U. Furthermore, we shall assume that $U|\psi\rangle = e^{2\pi i\theta}$, where $2^n\theta$ is an integer for some value $n \in \{0, 1, 2 \ldots\}$. After we review the functioning of the algorithm with these assumptions, we shall state how the results generalize when the assumptions are removed. With or without these assumptions, we shall require access to a quantum circuit that implements a controlled version of the unitary U, the operation $|0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes U$, where I is the identity.

The quantum phase estimation algorithm begins by initializing n ancilla qubits in the state $|0\rangle^{\otimes n}$ and acting on each of them with the Hadamard gate to produce an even superposition over all possible bit strings. Let us consider the action of the first gate after the dotted line labeled 1 in Figure 3.2. This controlled U^{2^n-1} gate, shorthand for $2^n - 1$ applications of the controlled U gate, has the following effect (neglecting the other ancilla qubits and the

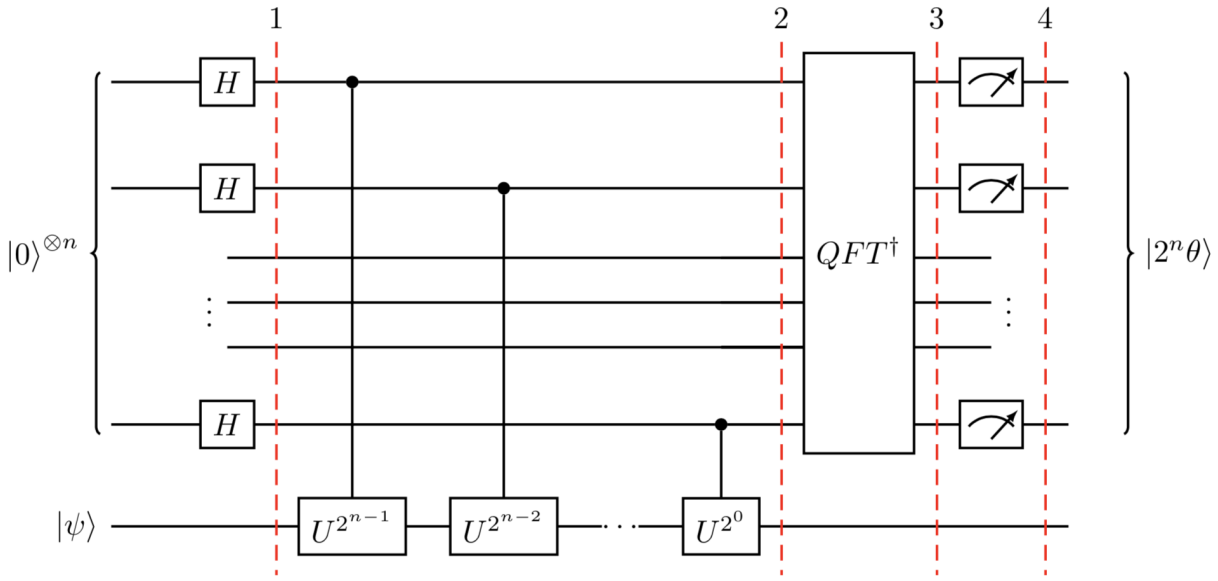

Figure 3.2: A circuit diagram of the quantum phase estimation algorithm. The simplest case is where $|\psi\rangle$ is an eigenstate of the unitary operator U with eigenvalue $e^{2\pi i\theta}$, where $2^n\theta$ is an integer. In this case, the measurement outcome will be $|2^n\theta\rangle \otimes |\psi\rangle$ with probability 1. Image taken with permission from Ref. 111.

normalization factor for now):

$$|0\rangle \otimes |\psi\rangle + |1\rangle \otimes |\psi\rangle \rightarrow |0\rangle \otimes |\psi\rangle + e^{2^n \pi i\theta} |1\rangle \otimes |\psi\rangle \,. \tag{3.8}$$

By repeatedly applying U to the state $|\psi\rangle$ conditioned on the ancilla qubit being in the $|1\rangle$ state, the algorithm accumulates a relative phase between the $|0\rangle$ and $|1\rangle$ states of the ancilla. The remaining gates between the first two dotted lines in Figure 3.2 work similarly, "kicking back" phases equal to θ multiplied by different powers of 2 onto the other ancillae.

At this point in the algorithm, the system is in the state

$$\frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} e^{2\pi ik\theta} |k\rangle \otimes |\psi\rangle, \tag{3.9}$$

where $|k\rangle$ is the computational basis state encoding the integer k in binary. The quantum Fourier transform can be defined by it action on a computational basis state [86, 112],

$$|j\rangle \rightarrow \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} e^{2\pi ijk/2^n} |k\rangle \,. \tag{3.10}$$

The inverse quantum Fourier transform, naturally, inverts this operation. By identifying j with $2^n\theta$, we can see that the application of the inverse quantum Fourier transform in

Figure 3.2 results in exactly the desired state, $|2^n\theta\rangle \otimes |\psi\rangle$. Although we do not prove it here, the quantum Fourier transform (or its inverse) can be implemented in a simple fashion using $O(n^2)$ two-qubit gates [86, 112]. More modern implementations reduce this cost to $O(n\,log(n))$ [113].

Of course, in the more general case, $|\psi\rangle$ may not be an eigenstate of U and it may have support on eigenvalues that are not precisely integers divided by powers of two. We can always expand $|\psi\rangle$ in the eigenbasis of U,

$$|\psi\rangle = \sum_i c_i \, |\phi_i\rangle \, . \tag{3.11}$$

Colloquially, the action of the phase estimation algorithm on $|\psi\rangle$ is to produce an n bit approximation to the eigenvalue of $|\phi_i\rangle$ with probability $|c_i|^2$. The post-measurement state of the system register (the qubits originally containing $|\phi\rangle$) in this case will be a state consistent with the measured approximation to the eigenvalue. If n is large enough to resolve the eigenvalue of a non-degenerate eigenstate, then the post-measurement state of the system register contains exactly the corresponding eigenvector. We refer the interested reader to the canonical Ref. 86 and also to Ref. 114 for a more sophisticated treatment of this case.

For our purposes, the most important fact can be summarized by the statement that phase estimation allows us to measure in the eigenbasis of U to within a precision ϵ with a cost (number of calls to the controlled U operator) that scales as $\tilde{O}(\frac{1}{\epsilon})$, where the tilde denotes that we have dropped the logarithmic factors. Ref. 76 showed how the controlled time-evolution operator for the quantum chemical Hamiltonian could be approximately implemented for a cost that scales polynomially with the system size. This thereby enabled Aspuru-Guzik et al. to efficiently calculate the ground state energy of the quantum chemical Hamiltonian, provided that a reference state with sufficient overlap on the true ground state could be obtained. More modern techniques have dramatically reduced the costs of such approaches, but the scaling with $\frac{1}{\epsilon}$ is at the fundamental Heisenberg limit [115]. Even with these improvements, it is widely expected that the application of these algorithms will not be practical for large systems on near-term devices [10]. It is useful, however, to keep in mind that the $\frac{1}{\epsilon}$ scaling with precision is possible with a sufficiently powerful quantum computer.

3.2 Near-Term Hardware

Today's, and tomorrow's, quantum computers are significantly more constrained than the idealized circuit model discussed in the section above. In reality, their fundamental components are not isolated two-level systems, and they are not perfectly controllable. Furthermore, they are limited in size, especially because the problem of engineering and controlling the individual qubits can become more challenging as the number of qubits is increased. Eventually, it is expected that fault tolerant quantum computers will more closely mimic the abstract circuit model quantum computer, but until that day arrives it is helpful to understand the non-ideal behavior of the various hardware platforms currently available [10].

The noise afflicting near-term devices has a variety of sources. The individual qubits may be imperfect, in the sense that it is only approximately true to describe them as isolated two-level systems. In order to enact quantum gates or to perform measurement it is necessary to couple to the qubits in a controlled way. Small errors or imperfections in the signals used to control and read out the state of the qubits can be the source of noise. These errors can result from an inability to perfectly send the desired signal, and also from a difficulty in knowing exactly what signal should be sent. Furthermore, unintended environmental coupling may occur at any time, both due to imperfections in the qubits and control signals, and due to other fluctuations in the environment beyond experimental control. We shall highlight some examples of these different types of noise below when we discuss particular hardware platforms.

Besides being noisy, existing devices are also limited in their size and connectivity. For example, superconducting quantum computing platforms typically arrange the qubits on a two-dimensional surface, with neighboring qubits connected to form a regular lattice [116]. In such a setup, two-qubit gates can be directly performed only between adjacent qubits. Ion trap quantum computers, by contrast, can be capable of executing gates between any pair of ions in an electromagnetic trap, but there are limits to how many ions may be put into a single trap before they become difficult to control individually [117]. The factors that limit the size of existing quantum computers are numerous, but challenge that is common to many of the specific implementations is the sheer amount of classical electronic infrastructure required to perform the control and readout. For example, Google's 53 qubit Sycamore chip was connected to 277 digital-to-analog converters each sending independent microwave signals in order to carry out the quantum supremacy experiment described in Ref. 7.

Devices based on superconducting qubits are some of the most advanced at the present day. Modern superconducting quantum computers use the lowest two levels of a weakly anharmonic oscillator as their qubits. By driving the qubits themselves, or other connected circuit elements, with microwave tones, they are able to achieve impressive performance. In the quantum supremacy experiment of Ref. 7, Google reported two-qubit gate error rates averaging .62% and single-qubit gate error rates averaging .16%. Furthermore, superconducting quantum computers are able to perform gates in 10s of nanoseconds, allowing for quick operation and repetition of experiments [116]. Superconducting quantum computers benefit from building on top of well-established semiconductor and telecommunications technologies, allowing for the use of off-the-shelf components for some aspects of the system.

However, certain aspects of quantum computing with superconducting circuits are especially difficult. Superconducting circuits are engineered on the surface of a material and defects within the substrate can interact with the qubits and control fields in unpredictable ways, leading to noise [118]. Performing readout and control via microwave signals offers advantages, but small errors in the shaping of the control fields or design of the hardware can lead to unintended interactions [7]. Leakage into higher levels of the anharmonic oscillators that make up the qubits can be particularly pernicious and hard to suppress [119], and can break assumptions that used by schemes for error mitigation and error correction [120].

Ion trap quantum computing offers several advantages [121]. By using atomic energy

levels as the basis for the two-level systems that make up a quantum computer, a trapped ion device is guaranteed to have identical qubits. These ions are suspended in a vacuum using electromagnetic fields, and so there are no defects in the substrate to cause unwanted interactions. Owing to these, and other, factors, ion trap quantum computers are extremely good at storing quantum information. Lifetimes for coherent quantum information of seconds or minutes have been reported [121], far larger than the microseconds typical of superconducting qubits [116]. Furthermore, state-of-the-art ion trap devices possess one-qubit gates, two-qubit gates, and measurement capabilities, with error gates below superconducting qubits (or, perhaps, competetive for the two-qubit gates) [121].

On the other hand, the engineering challenges of scaling up a trapped ion system are severe. The best experimental results reported for ion trap devices involve only a small number of ions in a linear trap. As the number of ions is increased, it becomes more challenging to perform the two-qubit gates accurately [121]. In order to circumvent this challenge, modular architectures that combine multiple ion traps have been proposed [117, 122]. Some proposals involve physically shuttling ions around the surface of a two-dimensional chip [117], while others propose to use entangled photons to teleport quantum information between separate traps [122]. These efforts promise to maintain the advantages small-scale ion trap computing while enabling larger devices to be constructed in a modular fashion, but experimental demonstrations remains challenging [121]. Ion trap quantum computing also struggles with gate times that are much longer than those of superconducting platforms, with typical two-qubit gate times of around 100 µs.

Photonic quantum computing is another route that promises to make good use of existing technologies, and it has the substantial advantage that information preserved in photonic degrees of freedom is naturally very stable. Linear optical quantum computing, as proposed in Ref 123, has been the focus of much activity over the past two decades. Linear optical quantum computing aims to construct circuit-model quantum computers out of well-established tools like beam-splitters and single-photon sources. The greatest disadvantage of this approach is that it requires the probabilistic implementation of certain two-qubit gates [123]. Dealing with qubits that are constantly in motion poses challenges as well. More recently, there has been substantial progress in photonic implementations of the one-way measurement-based model of quantum computation [97, 124, 125]. Under this approach, probabilistic generation of entanglement is used to build up a highly entangled state that is subsequently used for measurement-based quantum computing [125]. Proponents of this approach believe that, if challenges associated with single-photon generation and detection can be dealt with, it would be possible to quickly scale this approach to be competetive with superconducting and ion trap based approaches [125].

Other approaches, based on a variety of different physical platforms, are less well-developed at present. Devices based on Nitrogen-vacancy centers in diamond are already used as state-of-the-art quantum sensors, but the best demonstrations of multi-qubit systems to date have relied upon finding suitably coupled defects in samples, an approach that is not easily scalable [126]. Neutral atom qubits share some of the same desirable features of ionic qubits and they have shown great promise as analog simulation devices already [127]. Precisely

engineering the necessary multi-qubit gates using neutral atoms has proven challenging, but, if successful, could allow for the direct implementation of a useful family of gates involving three or more qubits [127]. Someday, devices may be constructed using exotic quasi-particles called non-Abelian anyons. It is believed that universal quantum computation can be performed by brading these quasi-particles around each other on a two-dimensional surface, and that such an architecture would display natural resilience towards noise [128]. Quantum computing based on Nuclear magnetic resonance was in vogue before its scalability was called into question [129]. Silicon-based spin qubits [130] and continuous variable quantum computing are also active areas of research [131]. The list of possible architectures goes on, and we do not attempt to exhaustively list them here.

It remains to be seen which approach, or approaches, will ultimately prove the most successful. It took decades for classical computer technology to settle on the metal-oxide-semiconductor field-effect transistor (MOSFET) technology that underlies modern electronics [132]. It may be that decades are required for quantum computing technology to likewise mature. The distinctions we have drawn in this section between different technologies are likely to blur and evolve over time as well. There are hints of this, for example, in the idea of using photonic interconnects to mediate interactions between components of an ion trap device [117, 122]. Technologies that seem well-positioned today may run into unexpected engineering roadblocks in the future. The main body of this thesis is concerned with near-term applications and we shall mainly focus on the two most mature technologies, superconducting circuits and ion traps, when we discuss implementation concerns.

3.3 The Noise in "Noisy"

We briefly mentioned the nuances behind the term "noise" in the last section. For the most part, we will neglect the exact sources of this noise and focus on a phenomenological description of noise, defined to be the deviation between the actual operation of the quantum computer and the desired operation. Of course, even from the point of view of applications and algorithm development, some understanding of the details of the incorrect operation is useful. For example, if attempts at the physical implementation of a particular two-qubit gate U deterministically fail by instead performing a different operation, $\tilde{U}$, this is a different kind of problem than a failure which occurs with some probability. We call the first kind of error a coherent error, while we call the second kind a stochastic error [133]. We shall discuss both kinds of errors as they occur during the execution of quantum gates and measurement. We shall also spend some time discussing slowly-varying noise that has particular implications for the NISQ algorithms that are the focus of this thesis.

Error rates during the execution of single-qubit gates, or while the qubits are idling, can be extremely low. We discuss these two errors together because the best experimental results for qubits not undergoing any operation are frequently obtained by using quantum control techniques, like dynamical decoupling, to zero out unintended interactions [134, 135]. The challenges in performing these idling sequences are comparable to those that occur in

the execution of single-qubit gates, and the error rates are therefore similar. Generally, the control problems become more challenging when more degrees of freedom are involved. This contributes to the fact that single-qubit gates (including the idling, or "identity" gate) are not a dominant source of errors in most near-term devices. Frequently, the single-qubit gates have error rates that are an order of magnitude or more smaller than the error rates for two-qubit gates [7, 93, 94, 116, 121].

Therefore, in the context of near-term quantum computing, we sometimes focus on the two-qubit error rate as the most important metric. We note here that there are a variety of different ways to quantify the performance of a quantum gate. Many of the notions of error and gate fidelity are related, Refs. 136 and 7 provide a good discussion of some of these relationships. In this section, we shall focus on the gate error rate defined as

$$\nu = \frac{1}{2}||C_{actual} - U_{ideal}||_{\diamond}, \tag{3.12}$$

where C_{actual} refers to the quantum channel actually implemented, U_{ideal} refers to the unitary quantum channel that was intended, and the $\diamond$ subscript indicates the use of the diamond norm [136]. Our discussion of error rates will be mostly qualitative rather than quantiative in any case. Because two-qubit gate errors are one of the dominant sources of error during a quantum computation, the number of two-qubit gates, sometimes referred to as the gate count, or two-qubit gate count, is a natural figure of merit for evaluating the feasibility of a near-term quantum algorithm. A related quantity, the two-qubit gate depth, (sometimes abbreviated as the "circuit depth" or "depth") counts any number of two-qubit gates performed in parallel as a single time step and focuses on the number of such time steps.

One and two-qubit gate noise is often modeled through the use of Kraus operators. The most general valid transformation of a quantum state is a completely positive trace-preserving (CPTP) map. A transformation $\Phi(\rho)$ is a CPTP map if and only if it can be expressed in the form

$$\Phi(\rho) = \sum_a M_a \rho M_a^\dagger, \tag{3.13}$$

for a collection of operators $\{M_a\}$ that satisfy

$$\sum_a M_a M_a^\dagger = I. \tag{3.14}$$

Ref. 137 provides an in-depth treatment of the theory. Modelling one and two-qubit gate noise as a quantum channel that acts on one or two-qubits only neglects correlations between errors, such as those arising from crosstalk in the control signals. However, such an approximation has been shown to provide qualitatively accurate results on real hardware [7], at least for certain types of circuits.

For example, one commonly used model is the single-qubit depolarizing channel. Kraus operator representations are not unique [10], but for the single-qubit depolarizing channel, the standard formulation provides an illuminating perspective,

$$M_1 = (1-p)I, \qquad M_2 = \frac{p}{3}X, \qquad M_3 = \frac{p}{3}Y, \qquad M_4 = \frac{p}{3}Z. \tag{3.15}$$

One interpretation of this channel is that, with probability $1-p$ there is no error and with probability p a randomly chosen single-qubit unitary from the set $\{X, Y, Z\}$ is applied. This probabilistic point of view is sometimes used to perform noisy simulations of quantum circuits using a quantum trajectories approach. By averaging over an ensemble of calculations where the action of the Kraus operators is replaced by a randomly chosen unitary M_a, it is possible to avoid full density matrix simulations and still capture the effects of noise [10, 138]. The single-qubit depolarizing channel is often taken as a starting point when calculating the properties of error-mitigation or error-correction techniques [139], and we shall use it for that purpose in Chapter 6. Single-qubit amplitude damping and dephasing channels, which we define in Section 8.6.3 of Chapter 8, are also commonly used for this purpose [10, 139].

The single-qubit depolarizing channel is an example of a stochastic error, but coherent errors are also widespread. One easy to understand example is provided by overrotation. Consider the $R_z(\theta)$ gate, which we define to be the unitary

$$R_z(\theta) = e^{-i\theta Z} = \begin{bmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{bmatrix}. \tag{3.16}$$

The $R_z(\theta)$ gate, along with other arbitrary single-qubit rotations, can be a natural operation on a near-term quantum computer. It might be the case that, for some fixed δ, the gate $R_z(\theta+\delta)$ is applied instead of $R_z(\theta)$ whenever the $R_z(\theta)$ gate is called for in a quantum circuit. We would call such an error an overrotation, and it is clear that this is a coherent error because it deterministically replaces the unitary that we wished to implement with a different, fixed, operation. We shall see later how coherent errors may be less problematic than stochastic ones in the context of NISQ computing [140]. In the context of quantum error correction, where stochastic errors may be easier to treat, there are randomized compiling techniques that can turn coherent errors into stochastic ones [141].

Errors during the measurement process are also a signficant challenge. The landmark experimental achievement of Ref. 7 was accomplished on a quantum processor with an average single-qubit readout rate at $\approx 3.8\%$ for simultaneous measurement. This translates to a probability of only $\approx \frac{1}{8}$ of correctly measuring all 53 qubits. Due to particular engineering flaws [7], that particular device has errors much worse than those reported for other state-of-the-art systems [116, 121], but the numbers serve to underscore the severity of the challenge. As with circuit-level noise, errors during the measurement process can be approximated by assuming a tensor-product structure, with an error channel acting independently on each qubit [142]. Noise during measurement can be particularly challenging to engineer against because of the fundamental need to couple the information in the qubit to the outside world. For example, dispersive readout of superconducting qubits is accomplished by engineering a shift in frequency that depends on the qubit's state [116]. However, the same physics can lead to the spontaneous emission of energy from the qubit, changing it's state in the process of reading it out [116].

Another challenging form of noise is the low-frequency noise that manifests itself as a slow drift in the performance and optimal control parameters of an experimental device [143, 144].

This kind of noise makes results averaged over time unreliable and can be catastrophic for certain kinds of near-term algorithms that rely on performing a large number of repetitions of a short circuit. For example, a particular application might rely on repeatedly preparing the same state and estimating the expectation values of different observables with respect to that state. If the actual state being prepared and measured by the device is slowly drifting, then it would no longer be possible to interpret all of the measurement statistics as belonging to a single (possibly noisy) state. Ref 145 discusses an example of this in the context of the experiment performed in Ref. 146. If it is necessary for a particular application to avoid this effect, slowly varying noise could set an upper limit on the duration of an experiment, and hence, the number of circuit repetitions that can be fruitfully used.

As larger quantum computers are built, understanding their noise processes better will be an important task. The results in the so-called "quantum supremacy" paper, Ref. 7, provide a good example of how experiments on near-term devices can contribute to this goal. The authors found that the error rate for their entire circuit could be predicted by a simple model that depended only on the error rates of the individual one and two-qubit gates, thereby providing support for the idealized error models that are used in the study of quantum error correction [147, 148]. The success of proposals for mitigating errors in near-term computations, which we shall discuss in Section 3.6, as well as fault tolerant quantum computation, will require a deep understanding of the actual characteristics of the noise on real hardware.

3.4 Variational Quantum Algorithms

Variational quantum algorithms arose as a response to the challenge of dealing with noise on near-term devices. The basic idea of a variational quantum algorithm is to cast the task being performed as a minimization problem over the parameters of a quantum circuit, $\vec{\theta}$. We define a function, $F(\vec{\theta})$, such that the minimum value of the function,

$$F(\vec{\theta}^*), \qquad \vec{\theta}^* = argmin_{\vec{\theta}}\big(F(\vec{\theta})\big), \tag{3.17}$$

or the circuit evaluated with the minimizing parameters, $U(\vec{\theta}^*)$, corresponds to the solution to some computational problem. The function $F(\vec{\theta})$ can depend on the parameterized circuit in a variety of ways, but the simplest example is one might be where $F(\vec{\theta})$ is the expectation value of some Hamiltonian, $\left\langle \psi(\vec{\theta}) \middle| H \middle| \psi(\vec{\theta}) \right\rangle$, with $\left| \psi(\vec{\theta}) \right\rangle$ defined to be the state prepared by parameterized quantum circuit acting on a fixed initial state. Then $F(\vec{\theta})$ may be evaluated by repeatedly preparing $\left| \psi(\vec{\theta}) \right\rangle$ and measuring H. We shall review this kind of variational quantum algorithm, called the variational quantum eigensolver (VQE), in detail in Section 3.5 and explain how such an approach helps to ameliorate the challenges posed by noise.

Variational quantum computing began with the variational quantum eigensolver of Ref. 17. Interested in studying the ground state of He-H^+ with a small photonic quantum computer,

Peruzzo et al. proposed a quantum computational analog to classical variational ansatz-based approaches in quantum chemistry. As we shall see, their proposal avoided the long coherence times required to perform quantum phase estimation by instead repeatedly preparing and measuring a parameterized quantum wavefunction. It wasn't initially expected that this approach would prove resilient to noise [17], but followup experimental work demonstrated a clear robustness against certain kinds of errors [149, 150].

The simplest example of an error that can be easily corrected by a variational quantum algorithm is an overrotation (or underrotation) error. Imagine that our problem would be exactly solved by a circuit with the parameters $\vec{\theta}^*$ in the absence of noise. Consider the case where some miscalibration in our device causes it to instead implement the circuit $U(\vec{\theta}+\vec{\delta})$ instead of $U(\vec{\theta})$ when given the parameters $\vec{\theta}$. By optimizing the parameters to minimize the observed value of $F(\vec{\theta})$ on our noisy device, we would find that we attain the minimal value of F using the parameters $\vec{\theta}^* - \delta$. The flexibility in the parameterized circuit would thus allow us to automatically compensate for this type of noise. This type of noise resilience, which was later described by the term "Optimal Value Resilience" in Ref. 151, can potentially correct for more general errors than a simple overrotation as well [140, 149].

Shortly after the initial experiment of Peruzzo et al., the Quantum Approximate Optimization Algorithm (QAOA) was proposed for combinatorial optimization [19]. Combinatorial optimization can be framed as the minimization of a function

$$-C(\vec{z}) = -\sum_{\alpha} C_\alpha(\vec{z}), \tag{3.18}$$

where $\vec{z}$ is a string of classical bits and the $C_\alpha(\vec{z})$s are functions that output a value in $\{0,1\}$. Farhi et al. proposed a variational quantum algorithm to solve this problem, promoting the function $-C(\vec{z})$ to a quantum Hamiltonian and describing a natural family of parameterized quantum circuits to use as ansatze for $U(\vec{\theta})$. The circuits they proposed have the appealing property that, given enough circuit depth, they can emulate the quantum adiabatic algorithm, an approach with a clear advantage over classical algorithms [152, 153].

The field of quantum machine learning has spawned a suite of proposals for variational quantum algorithms [20, 154–156], partly inspired by the success of similar techniques in the deep learning community [157, 158]. We shall expand on this connection when we introduce some deep learning background in Section 8.2, as a prelude to Chapter 8, which is adapted from our contribution to this application area, originally published as Ref. 159. Many of the variational quantum algorithms for machine learning build on the original proposal of Ref. 20 by implementing a transformation of some input data as a parameterized quantum circuit. Most commonly, the parameterized circuit $U(\vec{\theta})$ is optimized to act on a collection of input states $\{|\phi_i\rangle\}$ which represent some data. The function $F(\vec{\theta})$ is designed such that its minimizing parameters, $\vec{\theta}^*$, yield a circuit that performs some desired transformation on the data states.

Variational approaches have been applied in interesting ways to quantum simulation beyond the ground state problem. For example, in Ref. 160, Li and Benjamin showed how real-

time evolution under the Schrödinger equation can be approximated by a variational algorithm. Given an approximation of a wavefunction at time t by a parametrized wavefunction,

$$|\phi(t)\rangle \approx \left|\psi(\vec{\theta})\right\rangle, \tag{3.19}$$

one follow their approach to update parameters of the wavefunction to approximate the state at time $t + \delta t$. Real-time evolution in a different sense can be accomplished using the techniques of Ref. 161. By proposing a strategy to efficiently calculate energy derivatives with respect to the Hamiltonian parameters, O'Brien et al. provide the necessary tools to calculate the forces experienced by classical nuclei under the Born-Oppenheimer approximation. This, in turn, allows for classical simulation of the nuclear coordinates driven by variational quantum calculations of the electronic degrees of freedom.

Variational quantum algorithms have even been developed for the purpose of compiling other quantum algorithms. Refs. 162 and 151 explore strategies for approximating a given unitary or state preparation procedure using a parameterized quantum circuit. They focus on the tasks of optimizing a circuit $U(\vec{\theta})$ to approximate a target unitary,

$$U(\vec{\theta}^*) \approx V, \tag{3.20}$$

or to approximately prepare a target state,

$$U(\vec{\theta}^*)\,|0\rangle \approx V\,|0\rangle\,, \tag{3.21}$$

where $|0\rangle$ is some fixed reference state. The tools they develop allow for a longer circuit to be approximated by a shorter one, provided that the shorter parameterized circuit is flexible enough to enact the desired operation. This can be useful in the context of minimizing the impact of noise, which we shall revisit in Section 3.6

3.5 The Variational Quantum Eigensolver

Besides being one of the central foci of this thesis, the variational quantum eigensolver (VQE) offers an excellent case study of a variational quantum algorithm. We shall provide a self-contained review of the VQE formalism as it applies to quantum chemistry here and highlight some of its important features. Besides the more specific citations which we shall include along the way, we also refer the interested reader to Refs. 140, 163, and 164 for general discussions and literature reviews. We shall break our review into four pieces, beginning with an explanation of how fermionic wavefunctions may be represented using a circuit model quantum computer. Following this, we explain some of the issues that arise during ansatz design and review the most promising approaches to designing the parameterized quantum circuits for VQE. We then discuss the issue of energy measurement, explaining the state-of-the-art in the field prior to our contributions in this direction contained in Chapter 6. We conclude our discussion by reviewing the classical optimization strategies used to search the parameter landscape for minima.

Variational methods for the ground state problem have a long history in the classical electronic structure world. Variational algorithms that involve a parameterized quantum wavefunction offer new possibilities compared to existing classical approaches but it isn't clear that these new possibilities will ultimately unlock new capabilities. In general, determining the ground state of a given Hamiltonian is believed to be difficult even with a quantum computer [165]. More specific to the electronic structure problem is the observation that existing classical tools, variational and not, work extremely well, and that variational quantum algorithms will face steep challenges in surpassing them [166].

The approach taken with a quantum computer echoes that of these classical variational methods, some of which we reviewed in Section 2.3. When calculated in the usual way, $E(\vec{\theta}) = \left\langle \psi(\vec{\theta}) \middle| H \middle| \psi(\vec{\theta}) \right\rangle$, the energy of a parameterized wavefunction is guaranteed to be greater than or equal to the ground state energy of the Hamiltonian. Therefore, the minimum energy found by optimizing $E(\vec{\theta})$ as a function of $\vec{\theta}$ is an upper bound to the true ground state energy. By using a sufficiently flexible parameterized wavefunction and a powerful enough optimization routine, the ground state energy can be approximated extremely well.

Determining the energy of a parameterized wavefunction is the primary task of the quantum computer in the VQE formalism. By acting on a fixed reference state with a parameterized quantum circuit, the quantum computer prepares a qubit wavefunction that corresponds to a wavefunction of the fermionic system. The energy is then estimated by repeated state preparation and measurement. The optimization of the parameters is left, wholly or mostly, to a classical algorithm that interprets the measurement results and suggests new parameter values to the quantum co-processor [140]. Before we discuss the wavefunction, measurement, or optimization aspects though, we shall review how a system of distinguishable qubits can be used to represent a collection of indistinguishable fermionic particles.

3.5.1 Fermionic Wavefunctions on Qubits

Recall the second quantized form of the electronic structure Hamiltonian given in Eq. 2.6,

$$\hat{H} = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} g_{pqrs} a_p^\dagger a_r^\dagger a_s a_q + h_{nuclear}. \tag{3.22}$$

By fixing a set of single-particle basis functions and calculated the resulting coefficients, h_{pq} and g_{pqrs}, we already have a discretized version of the Hamiltonian that encodes all of the physics that we're interested in. For a system with N spin-orbitals, the Hilbert space constructed from all possible electronic states is a Fock space of dimension 2^N. We can construct exact isomorphisms between this space and the 2^N dimensional Hilbert space that describes the possible states of N qubits.

The first such isomorphism we consider, the Jordan-Wigner transformation, gives us a natural mapping between fermionic systems and qubits [167, 168]. The Jordan-Wigner

transformation is defined by the following equation,

$$a_p^\dagger \to \frac{X_p - iY_p}{2} \bigotimes_{j=1}^{p=1} Z_j. \tag{3.23}$$

Product of fermionic creation and annihilation operators form a basis for the operators in Fock space and products of Pauli operators likewise form a basis for operators in the qubit Hilbert space, so we have a way to map arbitrary operators between the two spaces. To complete the mapping, we define the fermionic vacuum state, $|vac\rangle$, to be the all 0 state of the qubits, $|0\rangle$. Under this mapping, the computational basis states of the qubit Hilbert space are associated with the determinants of the fermionic Hilbert space and the particular bit strings of the qubit basis states are the occupation number vectors of the determinants.

The downside to using the Jordan-Wigner transformation is that k-local fermionic operators (operators acting on exactly k fermionic modes) can be translated to qubit operators that act on any number of qubits. For example, assuming that $p > q$, we have

$$a_p^\dagger a_q + a_q^\dagger a_p = \frac{1}{2} X_p \Big(\bigotimes_{j=q+1}^{p-1} Z_j \Big) X_q + \frac{1}{2} Y_p \Big(\bigotimes_{j=q+1}^{p-1} Z_j \Big) Y_q. \tag{3.24}$$

The Jordan-Wigner transformation maps this fermionic operator with support on two modes into a qubit operator with support on $p-q+1$ qubits. This expansion of the support can cause overheads and challenges with various aspects of a quantum algorithm. For example, several of the proposals for parameterized quantum circuits that we shall discuss in the next section require real-time evolution by operators similar to the one in Eq. 3.24. A straightforward implementation of the required circuits on a quantum computer with restricted connectivity (like a grid of qubits) involves $O(N)$ two-qubit gates [169]. The expansion of support from the Jordan-Wigner transformation can also cause challenges during measurement, a problem that we discuss in more detail in Chapter 6.

Fortunately, these disadvantages can be mitigated or avoided in a variety of cases. In Ref. 170, Kivlichan et al. introduced two useful circuit primitives that enable this mitigation. They made use of the fermionic swap gate, originally developed in the context of classical tensor network simulations of quantum systems [171], to dynamically change the Jordan-Wigner ordering. The fermionic swap gate between modes i and $i+1$, described by the fermionic operators

$$1 + a_i^\dagger a_{i-1} + a_{i-1}^\dagger a_i - a_i^\dagger a_i - a^\dagger i - 1 a_{i-1}, \tag{3.25}$$

or the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \tag{3.26}$$

is like the normal swap gate for qubits except that it assigns a -1 phase is both qubits are in the 1 state in order to properly account for the anti-symmetry of a fermionic wavefunction.

Ref. 172 later showed how fermionic swap gates could be used to efficiently compile a number of different circuits for quantum simulation in a way that removes the Jordan-Wigner overhead.

Kivlichan et al. also introduced another important circuit primitive for Jordan-Wigner transformed fermionic wavefunctions, the Givens rotation network [170]. Consider the single-particle change of basis, defined in second quantization by the equation

$$c_p^\dagger = \sum_q a_q^\dagger u_{pq}, \tag{3.27}$$

where u is a unitary matrix in the single-particle Hilbert space. The Thouless theorem tells us that this change of basis can be implemented in the many-body Hilbert space through the action of the exponential of a related one-body operator [173],

$$exp\Big(\sum_{pq} t_{pq}(a_p^\dagger a_q - a_q^\dagger a_p)\Big), \tag{3.28}$$

where t is the matrix logarithm of u. Using the techniques of Ref. 170, the Jordan-Wigner transformed version of this unitary operator can be implemented exactly using $\binom{N}{2}$ two-qubit gates. This primitive has been instrumental to a number of algorithmic advances [77, 174, 175], including the work we present in Chapter 6.

Alternatives to the Jordan-Wigner encoding have been proposed, most notably the Bravyi-Kitaev transformation [176, 177]. Like the Jordan-Wigner transformation, the Bravyi-Kitaev transformation expands the support of a k-local fermionic operator, but the resulting qubit operator has support on at most $O(log(N))$ qubits [177, 178]. The exact form of the transformed operators is complicated enough and the details are not relevant to the rest of the thesis so we refer the interested reader to Ref. 177 for more details. We shall briefly summarize their qualitative explanation of the transformation here. Essentially, one can understand the Jordan-Wigner transformation as encoding the occupation of each spin-orbital locally, but in order to satisfy the fermionic anti-commutation relations, the action of a typical operator depends on the parity of the occupation of $O(N)$ other spin-orbitals. The Bravyi-Kitaev transformation provides and alternative fermion-to-qubit mapping that balances the storage of the occupation and parity information, using $O(log(N))$ qubits for each.

More exotic alternatives are potentially useful in specific cases. For example, a recent line of work has explored encodings based on Majorana fermions [179–182]. These mappings are capable of representing k-local fermionic operators using qubit operators with support on $O(1)$ qubits. This low overhead comes at the cost of restricting the allowable interactions between different fermionic modes, making these transformations most useful for simulating lattice models with restricted connectivity. These mappings also share some features with quantum error correcting codes, raising the possibility of error correction or error mitigation schemes specific to fermionic simulations using these encodings.

For the purposes of this thesis, we shall focus on the Jordan-Wigner transformation. This is mainly due to the existence of the efficient circuit primitives developed for the Jordan-Wigner transformation in Refs. 170, 166, and 172, but the simplicity of the Jordan-Wigner

transformation also makes it an appealing starting point. The algorithmic components in the aforementioned papers have the advantage that they are designed to work well using a linearly connected array of qubits, making them especially appealing for near-term implementation on superconducting quantum hardware where limited connectivities are typical.

3.5.2 Ansatz Design

When compared with classical variational techniques, the appeal of the VQE formalism is that the use of parameterized quantum circuits allows for new classes of ansatze to be explored. Historically, the first ansatz proposed for VQE was a Trotterized approximation to the unitary coupled cluster wavefunction [17]. As we discussed in Section 2.3.0.3, coupled cluster wavefunctions are parameterized by an exponential ansatz,

$$|\psi\rangle = e^T |\phi\rangle , \tag{3.29}$$

where T is a linear combination of excitation operators and $|\phi\rangle$ is some reference state, typically the Hartree-Fock state. Classical approaches to evaluating the coupled cluster are not variational, but by demanding that T be an anti-Hermitian operator, we ensure that $|\psi\rangle$ is a normalized quantum state at the cost of removing our ability to efficiently evaluate the energy using a classical computer [183–188]. However, we can efficiently implement an approximation to the unitary operator e^T using a Trotter expansion of the exponential and use the VQE framework to evaluate the energy variationally. This leads to a chemically motivated and well-understood ansatz for VQE that improves an already powerful classical method. We shall explore and discuss unitary coupled cluster further in Chapter 4.

A structurally similar ansatz that is also physically well-motivated is the Trotterized adiabatic state preparation of Ref. 18. The adiabatic theorem tells us that a system in the ground state of a time-dependent Hamiltonian $H(t)$ at time $t = 0$ will remain very close to the instantaneous ground state at later times, provided that the Hamiltonian is varying sufficiently slowly and that there is an energy gap between the ground state manifold and the higher excited states [153]. Wecker et al. proposed to construct a time-dependent Hamiltonian where the ground state of $H(0)$ is a straightforward to prepare reference state and the $H(1)$ is some other Hamiltonian of interest, such as the quantum chemical Hamiltonian of some molecule [18]. By using a Trotterized approximation to the time evolution of $H(t)$ as their VQE ansatz, they are guaranteed to obtain a circuit that would prepare the ground state given a sufficient number of Trotter steps so long as $H(t)$ has a non-zero gap in the region $0 \leq t \leq 1$.

Both Trotterized adiabatic state preparation and unitary coupled cluster are theoretically appealing, but they can require infeasibly deep circuits in practice. This has led to the development of other ansatz circuits, with a variety of tradeoffs being made to reduce the circuit depth. Our work in Chapter 4, along with several other papers [170, 189, 190], focuses on creating physically-motivated ansatz circuits whose depth scales linearly with the size of the system. Another family of approaches, first laid out in Ref. 191, iteratively constructs

ansatz circuits specific to the VQE instance being studied [192]. This adaptive approach is likely to be especially useful on near-term devices where minimizing the number of gates is crucial.

An alternative perspective is offered by the hardware-efficient ansatz [193]. Rather than taking an ansatz motivated by the physics of the Hamiltonian, a hardware-efficient ansatz is constructed from the naturally available interactions of the physical hardware. In Ref. 193, this means that the multi-qubit interactions were accomplished by allowing all of the qubits to evolve together under the native Hamiltonian of the device and the control was accomplished by interspersing parameterized single-qubit rotations. As noted in Ref. 194, these approaches can blur the line between analog quantum simulation and digital quantum computation. By using a hardware-efficient ansatz, the necessity of calibrating multi-qubit gates and compiling specific algorithmic components into these primitive gates is avoided, potentially making better use of noisy hardware.

Unfortunately, the potential advantages of a hardware efficient ansatz come at a cost. Using a hardware-efficient ansatz can make it challenging to take advantage of the natural symmetries of the Hamiltonian. For example, one is often interested in finding the ground state of the electronic structure Hamiltonian in a specific particle number sector. Using an ansatz built from particle-number conserving fermionic excitations, one can naturally focus on a specific particle number sector by choosing the appropriate reference. The hardware-efficient ansatz does not offer this capability. More seriously, finding the optimal parameters for a hardware-efficient ansatz can be prohibitively expensive. We shall revisit this issue below, in Section 3.5.4.

In general, a good ansatz should possess a variety of qualities that are sometimes in tension. It should have a low circuit depth and gate count, so as to make best use of the limited capabilities of a noisy, near-term device. Relatedly, it should not have too many parameters, in order to minimize the cost of optimization. At the same time, it should be flexible enough to represent the ground state of the Hamiltonian of interest. A good ansatz should either naturally fit the connectivity constraints of the hardware, or the overhead from compiling it to locally-connected hardware should, at least, be worth the cost. A good ansatz also needs to be easily optimizable. This oftentimes means that there is a reasonable way of initializing the parameters so that the optimization procedure doesn't have as much work to perform. Optimizability also requires that the energy landscape, as a function of the parameters, should be easily traversable and not too filled with bad local minima or other features that cause optimization algorithms to struggle.

3.5.3 Repeated State Preparation and Measurement

Now that we have established how fermionic operators and wavefunctions can be represented on a quantum computer, we are ready to discuss the core function of the quantum computer in the VQE framework, repeated state preparation and measurement. From the beginning, the variational quantum eigensolver was designed to replace the long-running circuits of quantum phase estimation with a large number of repetitions of shorter circuits [17]. On a

noisy device where circuits with a large number of gates are impossible to run, this can be a worthwhile tradeoff.

Recall the quantum chemical Hamiltonian,

$$\hat{H} = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} g_{pqrs} a_p^\dagger a_r^\dagger a_s a_q + h_{nuclear}, \tag{3.30}$$

where we use indices that represent spin-orbitals for simplicity. The general strategy for measuring an observable on a quantum computer is to diagonalize it, rotating the state into the eigenbasis of the observable and measuring in the computational basis. However, it is unclear how to construct such a circuit and determining one by brute force would be at least as challenging as solving the electronic structure problem classically. Quantum phase estimation offers one approach to performing a measurement of the entire Hamiltonian operator, but, as we discussed, VQE attempts to avoid the complicated circuits required to perform the measurement this way.

Because we are interested in the expectation value of the Hamiltonian rather than specific measurement outcomes, we have a simpler option available. We can estimate the expectation value of individual pieces of the Hamiltonian separately and sum them together. This approach is referred to as Hamiltonian averaging [140]. Typically, these pieces are chosen to allow for diagonalization using only single-qubit gates, but other choices are also possible and we explore one such alternative in Chapter 6. Regardless of how the Hamiltonian is decomposed into multiple pieces, measuring the Hamiltonian in this way has consequences. In particular, because the entire Hamiltonian is not being measured at once, applying Hamiltonian averaging to an eigenstate of the Hamiltonian won't result in an estimator with zero variance. This can be surprising to those with a background in the classical computational techniques referred to as varational Monte Carlo, where an analogous zero-variance behavior is typical [195].

The simplest form of Hamiltonian averaging can be arrived by applying the Jordan-Wigner transformation (or some other alternative) to the Hamiltonian of Eq. 3.30 and considering the resulting sum,

$$H = \sum_i h_i P_i. \tag{3.31}$$

Here, the P_i are products of single-qubit Pauli operators that are sometimes referred to as Pauli words or Pauli strings. Using only single-qubit rotations and measurements, our choice of measurement is limited to picking a measurement basis for each qubit. Under these restrictions, we can simultaneously measure two Pauli words P_i and P_j if and only if P_i and P_j contain the same Pauli operator for each qubit where they both act non-trivially [140]. This notion of simultaneous measurability is sometimes called qubit-wise commutativity [196]. A straightforward and experimentally accessible way of performing Hamiltonian averaging is to take the Hamiltonian of Eq. 3.31, break the collection of Pauli strings into qubit-wise commuting groups, and measure each group independently by repeatedly preparing the initial state and making the requisite single-qubit measurements.

The time required for measurement is an important factor in determining the overall time complexity of the variational quantum eigensolver. If we decompose the Hamiltonian into a

sum of operators $H = \sum_i O_i$, each of which is measured m_i times, then we can calculate the overall variance of the resulting energy estimator [140],

$$\sigma^2 = \sum_i \frac{\langle O^2 \rangle - \langle O \rangle^2}{m_i}. \tag{3.32}$$

To determine the expectation value of a particular term in this sum to within a fixed precision ϵ requires a number of measurements scaling with $\frac{1}{\epsilon^2}$. This quadratic scaling is unavoidable within the VQE framework but it can be mitigated by choosing a grouping which makes the numerators as small as possible [140]. We shall discuss some of the strategies for accomplishing this in Chapter 6 alongside our own work in this direction. It is important to note that the energies calculated using the VQE approach are only variational up to this statistical noise. If strict lower bounds on the ground state energy are desired the size of this uncertainty must be taken into account.

One can derive a simple upper bound to the number of measurements required for a fixed precision by determining upper bounds to the terms in the numerator of Eq. 3.32 and choosing the optimal strategy for distributing measurements between the terms [140, 197]. For example, consider a measurement strategy that measures each of the terms in the Hamiltonian of Eq. 3.31 separately. It can be shown that the overall number of measurements required (M) to determine the expectation value of the Hamiltonian to within an accuracy (ϵ) is upper bounded in the following way [197],

$$M \leq \left(\frac{\sum_i |h_i|}{\epsilon} \right)^2. \tag{3.33}$$

We discuss bounds of this type further and also calculate the actual variance of a variety of energy estimators with respect to the ground states of different model systems in Chapter 6.

3.5.4 Optimization

We have described the necessary ingredients required for the repeated state preparation and measurements aspects of VQE. Within the full VQE framework, these aspects are embedded in a classical "outer loop." Using noisy evaluations of the energy or its gradients, a classical optimization routine is responsible for suggesting new wavefunction parameters with the goal of minimizing the energy. Ultimately, the success of VQE relies on the ability of the classical optimizer to efficiently find a set of minimizing parameters. The stochastic error inherent in expectation value estimation can make this task particularly challenging [140]. Classical optimization routines are often not designed to be robust to noise and this can lead to either poor performance or exorbitant cost, depending on the number of circuit evaluations performed [169, 198].

Most early efforts used optimization routines that used only noisy estimates of the energy and did not attempt to make use of gradient information [146, 193], or avoided optimization

entirely [17]. In principle, energy evaluations can be used to approximate the derivatives with respect to parameter values using a finite difference approach. For small ϵ,

$$\frac{\partial \langle H \rangle (\vec{\theta})}{\partial \theta_i} \approx \frac{\langle H \rangle (\vec{\theta} + \epsilon \vec{\delta_i}) - \langle H \rangle (\vec{\theta} - \epsilon \vec{\delta_i})}{2\epsilon}. \tag{3.34}$$

In practice however, the statistical fluctuations in the estimation of the expectation values in the numerator are large compared to the difference at the two nearby finite difference points. As a result, optimization methods like COBYLA (constrained optimization by linear approximation) [199], BOBYQA (bound optimization by quadratic approximation) [200], the Nelder-Mead method [201], particle swarm optimization [202], and others that avoided the need for derivative information tended to be the methods of choice.

More recently, better tools for evaluating the gradients of hybrid quantum algorithms have been developed. The simplest approach, the parameter shift rule, doesn't work in every situation but is sufficient for a variety of cases [203]. Importantly, when it is applicable, it allows for the derivative with respect to a circuit parameter to be computed simply by querying the objective function at a handful of well-separated points. In other situations, it is possible to evaluate the derivatives using a slightly modified circuit that replaces one of the gates in the original circuit with its controlled form [204]. A generalization of the parameter shift rule has also been devices that claims to expand its applicability to cover the more general case [205]. Taken together, these tools remove the obstacles to using derivative-based optimization for the variational quantum eigensolver and variational quantum algorithms in general. It is worth noting that, even with these tools, variational quantum algorithms still require the derivatives with respect to each circuit parameter to be computed separately. Contrast this with the situation in classical computation, where automatic differentiation can enable the calculation of the entire gradient of a wide variety of functions with a cost similar to a single function evaluation [206].

There is some theoretical support for the use of gradient based optimization, even with the per-parameter overhead required. In Ref. 207, Harrow and Napp prove that there are certain cases where an optimization routine using the gradient is guaranteed to converge to the optimum faster than any derivative-free method. Their proof is specific to a family of artificially constructed problems, but is suggestive of a general principle. A recent numerical study focused on comparing the cost of actually performing various kinds of optimization on a near-term device, including the impact various kinds of communication latency [198]. The conclusions of Ref. 198 were less clear than Ref. 207, making the case that the relative performance of different kinds of optimization strategies depend on the specifics of the hardware and software used to control the quantum computing device itself.

3.6 Some Challenges for Variational Quantum Algorithms

We have already touched on some of the challenges facing variational quantum algorithms, particularly in our description of the variational quantum eigensolver. In this section, we shall review them more thoroughly and explain a unifying perspective on them that we have found helpful. We shall also highlight some of the attempted solutions to these challenges, focusing on those that have most impacted our own understanding of the field.

At a coarse-grained level, it can be helpful to think of these challenges in terms of two kinds of resources, the number of two-qubit gates and the number of measurement repetitions. Both of these quantities tell us something about the time required to execute a particular computation. The number of two-qubit gates in a quantum circuit is a rough measure of something that we shall call its coherent time complexity. As we shall explain, this notion is separate from the overall time required by the whole algorithm, the total time complexity, which is well approximated by the number of circuit repetitions. If a particular algorithm has a total time complexity (number of two-qubit gates) requirement that is beyond the capability of a near-term device, it might be impossible to run the algorithm without setting aside the quantum computer entirely and incurring the exponential overhead of simulating the whole procedure classically. Note that we don't focus on the number of qubits available as a resource, because this is unlikely to be a limiting factor for NISQ algorithms until sufficiently large devices are available to start serious experimentation with quantum error correction [10].

Any particular computation, such as a VQE calculation of the ground state energy, will have some required precision. As we discussed in Section 3.3, errors incurred during the implementation of two-qubit gates are frequently the dominant source of noise on near-term devices. Therefore, we could translate a particular precision requirement into an upper bound on the number of two-qubit gates. For example, one could calculate the expected number of gates after which a single error occurs by taking the inverse of the two-qubit gate error rate [10]. The particular precision requirement will depend heavily on the application. A VQE calculation aiming to estimate the ground state energy to within 1 milliHartree might struggle even with small error rates, as we see in Chapter 6. By contrast, combinatorial optimization with QAOA only requires that the probability of an error occurring throughout the computation is bounded significantly away from 1 [19].

Error mitigation techniques can dramatically affect the error rate that a particular application can tolerate, and hence, the two-qubit gate budget. For example, for applications that can be reduced to the estimation of expectation values, powerful techniques for extrapolation to the zero-noise limit have been developed [160, 208, 209]. By artificially increasing the error rates of a computation, this technique aims to characterize the expectation value of an observable O as a function of the error rate ϵ. Once an approximation to this function is constructed, the value at $\epsilon = 0$ can be inferred. Different proposals generate the required data points in a variety of ways, ranging from applying less intense control fields over a

longer time [208], to letting the natural variation of the hardware generate samples with different error rates [210], to artificially adding random gate operations to the circuits being executed [160]. Reductions in the error by an order of magnitude have been reported in state-of-the-art experimental implementations of these techniques [209]. Such reductions directly translate into the ability to use deeper circuits with larger numbers of two-qubit gates when targetting a fixed precision.

A wide variety of error mitigation strategies have been developed, including some that take advantage of the structure of specific problems. We do not exhaustively review them here, but instead merely attempt to mention a few important papers in the field. Quasi-probability strategies use properties analytically derived from an assumed error model [208], or measured by comparison with classically simulable circuits [211], to reconstruct the error-free expectation value from a family of circuits derived from the circuit of interest. When the quantity of interest is the ground state of a Hamiltonian, the quantum subspace expansion technique can correct errors by performing additional measurements and solving a small generalized eigenvalue problem [150, 212]. Symmetry verification, which we discuss in more detail in Chapter 6, can be used to enforce that a wavefunction satisfies some desired symmetry, such as fermionic parity [213–215].

Even with the existing suite of error-mitigation tools, applications like quantum chemistry that demand a high level of precision are challenging. As we discussed in Chapter 2, for many chemical applications we are interested in determining relative energies between different states to within "chemical accuracy," ≈ 1 milliHartree. Experiments on real devices struggle to meet this threshold. For example, the errors in Ref. 209 were typically an order of magnitude larger than this, despite the application of advanced error-mitigation techniques. In Ref. 216, the authors were able to achieve chemical accuracy only through the use of error-mitigation techniques that do not generalize to post-Hartree-Fock wavefunctions. It seems clear that improvements in error-mitigation will be required for variational quantum algorithms to make a significant impact on the field of quantum chemistry.

The other principle limitation on variational quantum algorithms is the overall time required to execute them. While a detailed model might account for the communication latency between the classical computer and the quantum hardware it controls, the number of repeated state preparation and measurement steps can be used as a basic proxy for the total time [198]. In state-of-the-art superconducting and ion trap devices, the time required to measure and reset the qubits is signficantly larger than the time required to execute the kinds of short circuits that are currently practical [116, 121]. The repetition rate is therefore not very sensitive to the details of the circuits, although it can vary significantly between different hardware platforms. The overall time required is an important figure of merit in the sense that an experimenter must be willing to spend the time required for a computation to finish, but it is also important for another reason. As we discussed in Section 3.3, slowly varying drift in the calibration of devices and other noise sources can lead to important algorithmic assumptions breaking down if the runtime of an experiment grows too large. This effect can set absolute upper limits to the total time beyond those implied by practical constraints.

Suppressing the stochastic noise inherent in the estimation of expectation values can

be extremely costly. Ref. 18 made some estimates of the total time required for a VQE calculation to attain chemical accuracy and found that $\approx 10^{19}$ circuit repetitions might be required to treat the molecule Fe_2S_2 in a minimal basis. Even with an experimental repetition rate of 10 kHz, a reasonable target for near-term superconducting devices, such a number is clearly infeasibly large [198]. Our work in Chapter 6 attempts to address this problem, but further improvements are likely necessary.

This challenge is magnified by the fact that variational quantum algorithms involve a classical optimization over circuit parameters. Ref. 217 points out that large families of random quantum circuits can be shown to have gradients with respect to an objective function of interest that are almost certain to be exponentially small in the number of qubits. In particular, a hardware-efficient ansatz initialized with random parameters is likely to be basically untrainable using an approach based on local search or gradient descent for large enough system sizes. Because the cost of evaluating the gradient to within a precision ϵ scales polynomially in $\frac{1}{\epsilon}$, an exponentially small gradient implies an exponentially large cost to resolve any meaningful information from a collection of noisy measurements [217]. This phenomenon suggests that great care must be taken in the design of parameterized quantum circuits and the strategies for initializing their parameters. One commonly used approach is to favor the use of ansatze that can be initialized with reasonable starting guesses for their parameters, but other solutions are possible as well [218, 219].

The number of measurements required to estimate the objective functions or its gradient to high precision must be multiplied by the number of optimization steps required to find a sufficiently good set of parameter values. Even if the gradient isn't exponentially small (in the number of qubits), the fact estimates of the gradient are subject to additive stochastic noise can make optimization difficult [217]. Further development of strategies and applications that minimize the overall time complexity of variational quantum algorithms will be necessary in order to scale them up to treat more interesting systems. Some hope is offered by recent results and strategies that suggest that benefits can be obtained without costly optimization strategies. For example, Ref. 220 shows that it may be possible to perform QAOA using a fixed set of parameters, rather than optimizing separately for different problem instances. As another example, Ref. 221 and Ref. 222 develop tools that use parameterized wavefunctions based on approximate real-time evolution rather than variational optimization to determine the ground state energy of a target system. It is also possible to trade additional coherent resources (an increased two-qubit gate count) for a lower overall time complexity, for example, by using the proposal of Ref. 223 to obtain some of the benefits of quantum amplitude estimation without the full cost.

Chapter 4

Generalized Unitary Coupled Cluster Wavefunctions for Quantum Chemistry on a Quantum Computer

4.1 Preface

This chapter is taken from the previously published Ref. 224, originally appearing in The Journal of Chemical Theory and Computation. It was co-authored by Dr. Joonho Lee, the author of this thesis, Professor Martin Head-Gordon, and Professor K. Birgitta Whaley. Dr. Lee and the author of this thesis contributed equally to the work, with Dr. Lee providing the quantum chemical expertise and the author of this thesis providing the quantum computing background. The code used to produce the numerical data presented was produced as a collaborative effort, except as noted in the text where the `Q-Chem` software was used to perform classical benchmark calculations [42]. The writing pertaining to the classical quantum chemical aspects and the analysis of the results was mostly completed by Dr. Lee, while the writing pertaining to the quantum computing aspects was mostly completed by the author of this thesis.

4.2 Introduction

Quantum computing promises to provide access to a new set of computational primitives that possess profoundly different limitations from those available classically. It was shown early on that quantum phase estimation (QPE) provides an exponential speed-up over the best "currently" known classical algorithms for determining the ground state of the molecular Hamiltonian [76]. However, the use of this approach is believed to require large, error-corrected, quantum computers to surpass what is possible classically [79, 88]. A more promising path to pursuing such "quantum supremacy" [15, 225] in the context of quantum chemistry on near-term quantum devices is a quantum-classical hybrid algorithm that is

referred to as the variational quantum eigensolver (VQE) [17]. Interested readers are referred to a more extensive review in Ref. 163.

Unlike phase estimation, VQE requires only a short coherence time. This hybrid approach uses a quantum computer to prepare and manipulate a parameterized wavefunction, and embeds this in a classical optimization algorithm to minimize the energy of the state as measured on the quantum computer, i.e.,

$$E = \min_\theta \langle \psi(\theta) | \hat{H} | \psi(\theta) \rangle, \tag{4.1}$$

where θ denotes the set of parameters specifying the quantum circuit required to prepare the state $|\psi\rangle$. From a quantum chemistry perspective, there are two key attractive aspects of the VQE framework:

1. The evaluation of the energy of a wide class of wavefunction ansätze which are exponentially costly classically (with currently known algorithms) requires only state preparation and measurement of Pauli operators, both of which can be carried out on a quantum processor in polynomial time. These wavefunction ansätze include unitary coupled-cluster (UCC) wavefunctions, [17, 226] the deep multi-scale entanglement renormalization ansatz (DMERA), [227] a Trotterized version of adiabatic state preparation (TASP), [18] the qubit coupled cluster approach (QCC),[228] and various low-depth quantum circuits inspired by the specific constraints of physical devices currently available. [193]

2. On a quantum processor, efficient evaluation of the magnitude of the overlap between two states is possible even when two states involve exponentially many determinants. Classically, this is a distinct feature only of tensor network [229] and variational Monte Carlo [230] approaches. However on a quantum computer, any states that can be efficiently prepared will also possess this advantage.

Given the recent progress and near-term prospects in quantum computing hardware, and the uniqueness of these capabilities, it is interesting to explore these two aspects from a quantum chemistry perspective and this constitutes the major motivation of this work.

The remainder of this paper is organized as follows. (1) We review existing UCC ansätze in the context of traditional coupled cluster theory, focusing in particular on unitary extensions of the generalized coupled-cluster ansatz of Nooijen [183]. We then present a new ansatz, referred to as k-UpCCGSD, that uses k products of the exponential of distinct pair coupled-cluster double excitation operators, together with generalized single excitation operators. We show that this ansatz is more powerful than previous unitary extensions of coupled-cluster, achieving a significant reduction in scaling of circuit depth relative to both straightforward unitary extensions of generalized UCC (UCCGSD) and conventional UCC with single and double excitations (UCCSD). (2) We analyze options for variational optimization of excited states that are subject to orthogonalization constraints with a previously variationally optimized ground state. [231] We explore several distinct options and make an analysis of the possible

errors encountered when using such a variational approach. We show that these excited state energies can be significantly improved by using a different reference state for the excited state variational calculation, specifically, by using single excitation reference states. (3) We undertake a systematic analysis of the resource requirements for realization of these UCC ansätze on a quantum computer, relevant to preparation of initial states of molecules for both QPE and VQE computations. Our resource analysis focuses on the scaling of gate count, circuit depth, and spatial resources with size of the quantum chemistry calculation. We find that the k-UpCCGSD ansatz exhibits a linear dependence of circuit depth (a measure of the computational time that we define explicitly below) on the number of spin-orbitals N, with higher order polynomial dependence obtained for both UCCGSD and UCCSD. (4) To assess the accuracy of the new ansatz, we undertake benchmarking calculations on a classical computer for ground and first excited states of three small molecular systems, namely H_4 (STO-3G, 6-31G), H_2O (STO-3G), and N_2 (STO-3G), making additional comparisons to conventional coupled cluster methods as relevant. Detailed analysis of potential energy curves for ground and excited states of all three species shows that k-UpCCGSD ansatz offers the best trade-off between low cost and accuracy. (5) We conclude with a summary and outlook for further development of unitary coupled cluster ansätze for efficient implementation of molecular electronic states in quantum computations.

4.3 Theory

We shall use $i, j, k, l, \cdots$ to index occupied orbitals, $a, b, c, d, \cdots$ to index unoccupied (or virtual) orbitals, and $p, q, r, s, \cdots$ to index either of these two types of orbitals. The indices will denote spin-orbitals unless mentioned otherwise. We use N to denote the number of spin-orbitals and η to denote the number of electrons.

4.3.1 Coupled-Cluster Theory

In this section, we first briefly review traditional coupled cluster (CC) theory and unitary CC (UCC). We shall then draw connections between an existing body of work on variants of coupled cluster theory and a recently described wavefunction ansatz for VQE,[18] before proposing a novel ansatz also motivated by previous work in quantum chemistry. We note that in the quantum information literature it is customary to use UCC to denote the unitary version of restricted CC, in contrast to the quantum chemistry literature where UCC generally refers to unrestricted CC. We follow the quantum information convention in this paper.

4.3.1.1 Traditional Coupled Cluster

Traditional CC is a successful wave function method used for treating correlated systems in quantum chemistry [56, 232, 233]. Coupled-cluster with singles and doubles (CCSD), i.e., where the excitations in the cluster operator $\hat{T}$ are restricted to singles and doubles, is suitable for treating most "weakly-correlated" chemical systems.

The CCSD wave function is usually written with an exponential generator acting on a reference state,

$$|\psi\rangle = e^{\hat{T}}|\phi_0\rangle, \tag{4.2}$$

where for CCSD we have a cluster operator

$$\hat{T} = \hat{T}_1 + \hat{T}_2, \tag{4.3}$$

with

$$\hat{T}_1 = \sum_{ia} t_i^a \hat{a}_a^\dagger \hat{a}_i \tag{4.4}$$

$$\hat{T}_2 = \frac{1}{4}\sum_{ijab} t_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i. \tag{4.5}$$

In traditional CCSD, we evaluate the energy by projection of the Schrödinger equation, $\hat{H}|\psi\rangle = E|\psi\rangle$ first with $\langle\phi_0|$:

$$E \equiv \langle\phi_0|\hat{H}|\psi\rangle. \tag{4.6}$$

We then project with $\langle\phi_\mu|$ where μ is any single ($\langle\phi_i^a|$) or double ($\langle\phi_{ij}^{ab}|$) substitution. The t-amplitudes are then obtained by solving a set of non-linear equations:

$$0 = \langle\phi_\mu|\hat{H}|\psi\rangle - Et_\mu, \tag{4.7}$$

with $|\phi_\mu\rangle = \hat{t}_\mu|\phi_0\rangle$. The cost of solving Eq. Eq. 4.7 scales as $\mathcal{O}(\eta^2(N-\eta)^4)$, where η is the number of electrons and N is the total number of spin-orbitals possessed by the system.

It is evident from Eq. Eq. 4.6 that the projective way of evaluating energy is not in general variational, except in some obvious limits where CCSD is exact (e.g., for non-interacting two-electron systems[56, 232, 233]). With spin-restricted orbitals, it is quite common to observe catastrophic non-variational failure of CCSD when breaking bonds or, more broadly, in the presence of strong correlation. This non-variational catastrophe is often attributed to the way in which traditional CCSD parametrizes quadruples (i.e., $\hat{T}_2^2/2!$)[48, 49, 234–237] and searching for solutions to this problem without increasing the computational cost is an active area of research [48, 49, 236, 237]. Unfortunately, attempting to avoid this breakdown by variationally evaluating the energy of a CC wave function leads to a cost that scales exponentially with system size.

4.3.1.2 Unitary CC

A simple approach to avoid the non-variational catastrophe on a quantum computer is to employ a unitary CC (UCC) wavefunction,[184–188]

$$|\psi\rangle = e^{\hat{T}-\hat{T}^\dagger}|\phi_0\rangle, \tag{4.8}$$

where for the case of UCCSD, $\hat{T}$ is defined as in Eqs. Eq. 4.3 - Eq. 4.5. We can then evaluate the energy in a variational manner,

$$E(\{t_i^a\},\{t_{ij}^{ab}\}) \equiv \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle}, \tag{4.9}$$

using the standard VQE approach[17, 140, 163] that is summarized later in this work. UCC has a long history in electronic structure for quantum chemistry, with a number of theoretical works dedicated to the approximate evaluation of Eq. Eq. 4.9 within a polynomial amount of time,[184–188] since the approach appears to scale exponentially if implemented exactly using a classical computer. UCC is more robust than traditional CC, due to the fact that the unitary cluster operator involves not only excitation operators ($\hat{T}$) but also de-excitation operators ($\hat{T}^\dagger$). Nevertheless, the single reference nature of Eq. Eq. 4.8 can still lead to difficulties when treating strongly correlated systems on classical computers. This was investigated in Ref. 238 for the Lipkin Hamiltonian.

Unlike a classical computer, a quantum computer can efficiently employ a UCC wavefunction, even with a complicated multi-determinantal reference state, since both preparation of the state and evaluation of its expectation values can be carried out using resources that scale polynomially with system size and number of electrons [17, 140]. For UCC with singles and doubles (UCCSD), one must implement a Trotterized version of the exponentiated cluster operator, with $\mathcal{O}((N-\eta)^2\eta^2)$ terms, where each term acts on a constant number of spin-orbitals.

4.3.1.3 Generalized CC

In the early 2000's, there was an active debate on the question of whether the exact ground state wavefunction of an electronic Hamiltonian can always be represented by a general two-body cluster expansion. Motivated by earlier work of Nakatsuji, [239] Nooijen conjectured [183] that it is possible to express an exact ground state of a two-body Hamiltonian as

$$|\psi\rangle = e^{\hat{T}}|\phi_0\rangle, \tag{4.10}$$

where

$$\hat{T} = \hat{T}_1 + \hat{T}_2 \tag{4.11}$$

$$= \frac{1}{2}\sum_{pq} t_p^q \hat{a}_q^\dagger \hat{a}_p + \frac{1}{4}\sum_{pqrs} t_{pq}^{rs} \hat{a}_r^\dagger \hat{a}_s^\dagger \hat{a}_q \hat{a}_p. \tag{4.12}$$

This yields an exponential ansatz with a number of free parameters, the t_p^q and t_{pq}^{rs} values, that is equal to the number of parameters in the Hamiltonian. Here the single and double "excitation" terms do not distinguish between occupied and unoccupied orbitals and they are therefore called "generalized" singles and doubles (GSD). Although early work showed that the numerical performance of the resulting wavefunction was promising, the conjecture of Ref.183 has been the subject of an active debate and was later disproved.[240–247]

4.3.2 Generalized Unitary CC

We explore here a generalized form of the UCC wavefunction introduced in the VQE literature.[17] Our approach uses the generalized excitations of of Nakatsuji and Nooijen described above in the ansatz

$$|\psi\rangle = e^{\hat{T}-\hat{T}^\dagger}|\phi_0\rangle, \tag{4.13}$$

with $\hat{T}$ the cluster operator from Eq. Eq. 4.11. We shall term this ansatz UCCGSD. A unitary version of coupled cluster with generalized singles and doubles was first mentioned in Nooijen's paper,[183] but has never been thoroughly studied classically without making an approximation to the energy evaluation.

We note that a similar approach to defining a UCC ansatz by relating the terms in the Hamiltonian to generalized singles and doubles operators has appeared recently in the quantum computing literature, [18] where the performance of a Trotterized version of such a UCCGSD on small hydrogen chains and equilibrium geometry molecular systems has been characterized. As we shall show explicitly later in this work, the UCCGSD wavefunction is far more robust and accurate than the simpler UCCSD wavefunctions for the chemical applications considered here.

4.3.2.1 Unitary Pair CC with Generalized Singles and Doubles Product Wavefunctions

The method of pair coupled-cluster double excitations (pCCD) [50], also known as AP1roG [248] extends a widely used quantum chemistry method known as generalized valence-bond perfect-pairing (GVB-PP) [249]. pCCD is less prone than spin-restricted CCSD (RCCSD) to a non-variational failure when breaking bonds, despite the fact that it is computationally much simpler than RCCSD. pCCD is a coupled cluster wavefunction with a very limited number of doubles amplitudes (containing only the two body excitations that move a pair of electrons from one spatial orbital to another),

$$\hat{T}_2 = \sum_{ia} t_{i_\alpha i_\beta}^{a_\alpha a_\beta} \hat{a}_{a_\alpha}^\dagger \hat{a}_{a_\beta}^\dagger \hat{a}_{i_\beta} \hat{a}_{i_\alpha}, \tag{4.14}$$

where the summation runs over occupied and unoccupied spatial orbitals. pCCD is capable of breaking a single-bond qualitatively correctly, but fails to break multiple bonds. Orbital optimization of pCCD wavefunctions includes the important effects of the single excitations in a UCC wavefunction. In exchange for its high computational efficiency and reduced incidence of non-variationality, pCCD has other disadvantages: it loses invariance to unitary transformation within the occupied-occupied and virtual-virtual subspaces present in CCD, and it does not recover the dynamic correlation that CCD has.

We define the unitary pCCSD (UpCCSD) wavefunction to have the full singles operator as in Eq. Eq. 4.4 together with the unitary doubles operator of Eq. Eq. 4.14. We show below in the analysis of the quantum resource requirements that the circuit depth (time complexity) of preparing a UpCCSD state on a quantum computer scales linearly with the system size as

quantified by the number of spin-orbitals. However, our initial exploration of UpCCSD yielded errors in the absolute energies that were generally larger than the threshold for chemical accuracy. We therefore improve this wavefunction by the following two modifications: (i) we use the generalized singles and doubles operators employed in Refs. 239, 183, and (ii) we take a product of a total of k unitary operators to increase the flexibility of the wavefunction. We shall refer to this model as k-UpCCGSD.

Formally, k-UpCCGSD is defined in the following manner. For a chosen integer k,

$$|\psi\rangle = \Pi_{\alpha=1}^{k} \left(e^{\hat{T}^{(\alpha)} - \hat{T}^{(\alpha)\dagger}} \right) |\phi_0\rangle, \tag{4.15}$$

where each $\hat{T}^{(k)}$ contains an independent set of variational parameters (i.e., the singles and paired doubles amplitudes, the t_p^q's and the $t_{p_\alpha p_\beta}^{q_\alpha q_\beta}$'s respectively). Since the doubles operator in UpCCGSD is very sparse, the circuit depth required to prepare a k-UpCCGSD state still scales linearly with the system size, with a prefactor that is increased by a factor of k. This is similar in spirit to other recently proposed low depth ansätze [190] and also to the repeated independent variational steps of the Trotterized adiabatic state preparation approach[18] but, to our knowledge, this form of wavefunction has never been explored in either classical or quantum computational electronic structure calculations for quantum chemistry.

4.3.3 Excited State Algorithms

4.3.3.1 Previous Work

Obtaining excited states under the variational quantum eigensolver (VQE) framework has attracted considerable interest recently due to the substantial progress made in experimental realization of ground state VQE simulations [17, 149, 193, 250–252]. Algorithms proposed to extend this hybrid approach to excited states include the quantum subspace expansion (QSE) algorithm [212], the folded spectrum (FS) method [17], the witnessing eigenstates (WAVES) strategy [250], and a method based on penalizing overlap with an approximate ground state [231, 253]. We shall refer to the last of these as orthogonally constrained VQE (OC-VQE).

The QSE method is motivated by a linear-response approach: it samples the Hamiltonian matrix elements in the linear response space of a ground state wave function and diagonalizes it to obtain an excitation spectrum. A major drawback of this method is an obvious steep increase in the number of measurements after the ground state VQE calculation, since every matrix element needs to be sampled. Furthermore, QSE suffers from the well-known problem of linear-response methods, that is, it can only describe excited states that are within a small perturbation of a given ground state. However, the proper description of chemically relevant excited states sometimes requires inclusion of a higher order of excitations. A classic example of this is the dark low-lying excited state of butadiene, which requires that the linear response space include quadruple excitations in order to obtain a converged result.[254]

The FS method is closely related to the variance minimization algorithm widely used in the quantum Monte Carlo community:[255]

$$E(\omega) = \min_\theta \langle\psi(\theta)|(\hat{H}-\omega)^2|\psi(\theta)\rangle. \tag{4.16}$$

One advantage of this algorithm over the WAVES and OC-VQE algorithms is its ability to target a state whose energy is the closest to a preset ω, as in Eq. Eq. 4.16. Although this ability to variationally target specific excited states is very desirable, the algorithm inherently involves the evaluation of a quadratic term in $\hat{H}$, which greatly increases the number of Hamiltonian terms. Due to its steep scaling, $\mathcal{O}(N^8)$ in a standard gaussian basis set, application of the FS method (if possible) is likely to be limited to very small systems.

The WAVES algorithm relies on the ability of a quantum computer to efficiently perform time evolution conditioned on the state of a control qubit.[250] The protocol applies single qubit tomography to the first qubit of the state $\frac{1}{\sqrt{2}}|0\rangle \otimes |\psi\rangle + \frac{1}{\sqrt{2}}|1\rangle \otimes e^{-i\hat{H}t}|\psi\rangle$, for a given input state $|\psi\rangle$ and time t. The reduced density matrix of the control qubit describes a pure state if and only if $|\psi\rangle$ is an eigenstate of the Hamiltonian, or a superposition of degenerate eigenstates. Using this idea, it is possible to variationally target excited states (although not specific energies as is possible with the FS method), by varying the parameters of the trial state to maximize the purity of the measured single qubit state. This advantage is offset by the requirement that the quantum computer must implement a controlled version of the time evolution operator, which imposes steep demands on the relatively noisy quantum computing devices currently available.

4.3.3.2 Orthogonally Constrained VQE

In this work we explore an alternative to the aforementioned three methods which has the advantage that it requires roughly the same number of measurements as the ground state VQE calculation and only a doubling of the necessary circuit depth. [231] This algorithm can be naturally used with the two generalized coupled cluster wavefunction ansätze described above, or with any other circuit suitable for ground state VQE. Furthermore, OC-VQE can describe excited states that lie beyond the linear-response regime of the ground state. The approach assumes that a circuit for the ground state wavefunction is already available from a standard VQE calculation. One then defines an effective Hamiltonian whose lowest eigenstate is the first excited state and whose lowest eigenvalue is the energy of said state. Under the assumption that the excited state of interest has an energy less than zero,

$$\hat{H}_{\text{OC-VQE}} = \hat{H} + \mu\,|\psi_0\rangle\,\langle\psi_0|\,, \tag{4.17}$$

where $|\psi_0\rangle$ is the ground state wavefunction and the second term constitutes a level shift operator. For the molecular systems studied here, both the ground and first excited states are bound states (i.e., the electronic energies of these states are negative). Under these assumptions, we can choose $\mu = -E_0 = -\langle\psi_0|\hat{H}|\psi_0\rangle$. [231] This level shift imposes an energy penalty of $\mu|\langle\psi_0|\psi_1\rangle|^2$ on any trial state $|\psi_1\rangle$ that overlaps with $|\psi_0\rangle$. Such an energy level

shift technique is commonly used in quantum chemistry to enforce constraints within a variational framework [256–259]. Similar techniques have also been used in density matrix renormalization group calculations. [229] Minimizing the expectation value of $\hat{H}_{\text{OC-VQE}}$ with respect to the parameters in $|\psi_1\rangle$ defines this first OC-VQE procedure.

The choice of effective Hamiltonian in Eq. Eq. 4.17 is not unique. We have also explored the form

$$\hat{H}'_{\text{OC-VQE}} = (1 - |\psi_0\rangle \langle\psi_0|) \hat{H} (1 - |\psi_0\rangle \langle\psi_0|) . \tag{4.18}$$

Eq. Eq. 4.17 and Eq. Eq. 4.18 are identical if and only if $|\psi_0\rangle$ is an eigenstate of $\hat{H}$ with an eigenvalue E_0. If we choose $\mu = \infty$, the two approaches yield the same first excited state for a given approximate ground state $|\psi_0\rangle$. Both Eqs. Eq. 4.17 and Eq. 4.18 minimize the trial energy in the orthogonal complement space of $|\psi_0\rangle$, and these two different effective Hamiltonians have been interchangeably utilized in various contexts in quantum chemistry. [257, 259] We choose to work with Eq. Eq. 4.17 here, since it has a clear implementation suitable for a near term quantum device without requiring costly controlled unitary implementations of the state preparation circuits.

Specifically, it is clear that OC-VQE can be effectively implemented using the Hamiltonian of Eq. Eq. 4.17 so long as an efficient algorithm for measuring the magnitude of the overlap between the ground state and a trial excited state is available. On a classical computer, measuring the overlap between, for instance, two UCC states scales exponentially while on a quantum device this task is only polynomial scaling [231]. We describe one implementation of the necessary overlap calculation between two parameterized quantum states in the Quantum Resource Requirements section below, and refer the reader to recent work by Higgott et al. [231] for additional discussion on minimizing the effect of errors on this measurement.

4.3.3.3 Energy Error Analysis of OC-VQE

When an exact ground state $|\psi_0\rangle$ of $\hat{H}$ is used to construct the effective Hamiltonian $\hat{H}_{\text{OC-VQE}}$ in Eq. Eq. 4.17, the exact ground state of $\hat{H}_{\text{OC-VQE}}$ yields the exact excited state of the original Hamiltonian $\hat{H}$. We now show that use of an approximate ground state, $|\tilde{\psi_0}\rangle$, in the construction of $\hat{H}_{\text{OC-VQE}}$ will cause the excited state energy to incur an error that is similar in size to the error in the ground state energy, i.e. $E_0 - \langle\tilde{\psi_0}|\hat{H}|\tilde{\psi_0}\rangle$. We define the relevant excited state Hamiltonian,

$$\hat{\tilde{H}}_{\text{exc}} = \hat{H} - \tilde{E_0}|\tilde{\psi_0}\rangle\langle\tilde{\psi_0}|, \tag{4.19}$$

and consider the difference in energy between the ground states of $\hat{\tilde{H}}_{\text{exc}}$ and of $\hat{H}_{exc}$ in Eq. Eq. 4.17.

Writing the approximate ground state as $|\tilde{\psi_0}\rangle = \sqrt{1-\epsilon^2}\,|\psi_0\rangle + \epsilon\,|\psi_\perp\rangle$, where $\langle\psi_0|\psi_\perp\rangle = 0$, we can rewrite Eq. 4.19 as

$$\hat{\tilde{H}}_{\text{exc}} = \hat{H}_{\text{exc}} + \hat{V}, \qquad \hat{V} = -\epsilon E_0\,|\psi_\perp\rangle\,\langle\psi_0| - \epsilon E_0\,|\psi_0\rangle\,\langle\psi_\perp| + \mathcal{O}(\epsilon^2). \tag{4.20}$$

The first excited state of $\hat{H}$, which we denote $|\psi_1\rangle$, is by definition an approximation to the ground state of $\hat{\tilde{H}}_{\text{exc}}$. Assuming that ϵ is small, we compute the first order correction to the energy using Eq. Eq. 4.20. Because $|\psi_0\rangle$ and $|\psi_1\rangle$ are orthogonal, it is immediately clear that $\langle\psi_1|V|\psi_1\rangle$ is zero to first order in ϵ. Therefore, the difference between the true excited state energy, E_1, and the energy given by finding the ground state of the approximate excited state Hamiltonian, $\hat{\tilde{H}}_{\text{exc}}$, is $\mathcal{O}(\epsilon^2)$, which is on the same scale as the error in the ground state energy, $\epsilon^2(\langle\psi_\perp|\hat{H}|\psi_\perp\rangle - E_0)$.

Of course, in practice, we also do not find the exact ground state energy of $\hat{\tilde{H}}_{\text{exc}}$, instead incurring an additional error in our determination of the excited state energy from the second round of approximate minimization. However, if we make the assumption that the VQE procedure on $\hat{\tilde{H}}_{\text{exc}}$ is carried out well enough (and the ansatz is flexible enough) to yield an approximate ground state which is ϵ_1 away from the true ground state of $\hat{\tilde{H}}_{\text{exc}}$, then our overall error in the energy will be $\mathcal{O}(\epsilon^2 + \epsilon_1^2)$.

4.4 Quantum Resource Requirements

To assess the benefits of unitary coupled cluster theory for quantum computation it is important to quantify the cost of both state preparation and measurement needed to use these states on quantum processors. Our presentation here addresses the resources required for state preparation for a general quantum computation - we refer the reader to prior work for additional details specific to measurement in the VQE hybrid implementation [140]. This resource analysis requires an accounting of the number of quantum gates ("gate count" or "circuit size"), the time required to implement them, and the number of qubits on which they act. We shall take the total gate count to be determined by the number of two-qubit gates. In general, the relationship between the gate count and the number of sequential time steps required to implement them when parallelization is taken into account, the "circuit depth," will depend on the architectural details of the quantum processor. For many applications in quantum chemistry optimal results can nevertheless be obtained with minimal assumptions [170, 175].

We now present the implementation details necessary for evaluating the scaling of our proposed ansätze with respect to the numbers of spin-orbitals and electrons represented by the state. Our presentation here addresses the resources required for a general quantum computation - we refer the reader to prior work for additional details specific to the VQE hybrid implementation [140].

In order to treat the UCC ansatz on a quantum computer, it is necessary to map [168, 177, 260] the reference state and the exponentiated cluster operator from a Hilbert space of N fermionic spin-orbitals to a collection of quantum gates acting on N qubits. Therefore, the qubit resource requirement is linear in the number of spin-orbitals. For a UCC ansatz, the total gate count would be naïvely expected to be lower bounded by the number of cluster amplitudes t_p^q and t_{ps}^{rs}, possibly with additional overhead deriving from the mapping to

fermionic modes and the limited connectivity of a real device. Regarding the former, while the Jordan-Wigner transformation allows the representation of fermionic creation and annihilation operators in terms of products of single qubit Pauli operators in a way that properly encodes the canonical commutation relations,[168] direct application of this transformation maps the fermionic operators acting on individual spin-orbitals to qubit operators that act non-locally on $\mathcal{O}(N)$ qubits, leading to a corresponding overhead for the circuit depth. However, recent work in Refs. 175 and 172 describes procedures for implementing a Trotter step of unitary coupled cluster in a manner that not only entirely eliminates this Jordan-Wigner overhead, but also allows for the parallel implementation of individual exponentiated terms from the cluster operator on a linearly connected array of qubits. We note that a practical implementation of UCC relies on approximating $e^{\hat{T}-\hat{T}^\dagger}$ by a small number of Trotter steps, which leads to ansätze that are not exactly equivalent to the ones considered in our numerical calculations. Nevertheless, it has been demonstrated that the variational optimization of as few as one Trotter steps of UCC can yield highly accurate quantum chemical calculations [261].

Energy measurement and wavefunction optimization in the VQE framework both require repeated state preparation to overcome the statistical nature of the measurement process.[17, 140] Therefore, in analyzing the asymptotic time complexity for quantum computation of the approaches considered here, we focus on the cost of state preparation as quantified by the gate count and the circuit depth required for a fixed number of Trotter steps. Generally, we expect a practical benefit from minimizing both the number of free parameters that must be optimized (i.e., the cluster amplitudes) and the circuit depth.

The scaling of the circuit depth was derived here by assuming the maximum possible parallelization of terms in the cluster operator that act on distinct spin-orbitals and neglecting the Jordan-Wigner overhead. [172] Within this approach it is then clear that the k-UpCCGSD ansatz allows reduction of the circuit depth from the gate count by a factor of N, since the doubles pairs may be grouped into $\mathcal{O}(N)$ sets of $\mathcal{O}(N)$ terms, each of which acts on distinct spin-orbitals and can the $\mathcal{O}(N)$ sets can therefore be executed in parallel. We note that the results can also be obtained by using the procedure in Ref. 175 without additional numerical truncation. The resulting asymptotic scaling of gate count and circuit depth with respect to both the number of spin-orbitals N and electrons η is shown in Table 1 for all three unitary ansätze. Specific values for the numbers of cluster amplitudes used for the individual molecules for which benchmarking studies are performed will be shown in Table 4.9 in the results section.

4.4.1 Quantum implementation of Overlap Measurements

In order to implement the excited state algorithm used this work, Eq. Eq. 4.17, it is necessary to estimate not only the expectation value of the energy, but also $|\langle\psi_0|\psi_1(\theta)\rangle|^2$, where $|\psi_0\rangle$ is a parameterized guess for the ground state wavefunction and $|\psi_1(\theta)\rangle$ is the excited state ansatz. Allow $\hat{U}_1$ to be the quantum circuit that generates $|\psi_1(\theta)\rangle$ from the $|0\rangle$ state of the qubit register, i.e., $|\psi_1(\theta)\rangle = \hat{U}_1|0\rangle$. Let $\hat{U}_0$ be the unitary which prepares $|\psi_0\rangle$. The circuit that applies $\hat{U}_0^\dagger$ can be constructed simply by inverting each of the gates that compose $\hat{U}_0$.

Method	Gate Count	Circuit Depth
UCCSD	$\mathcal{O}((N-\eta)^2\eta^2)$	$\mathcal{O}((N-\eta)^2\eta)$
UCCGSD	$\mathcal{O}(N^4)$	$\mathcal{O}(N^3)$
k-UpCCGSD	$\mathcal{O}(kN^2)$	$\mathcal{O}(kN)$

Table 4.1: Resources required for preparing the three classes of UCC wavefunctions UCCSD, UCCGSD, and k-UpCCGSD, on a quantum device using a fixed number of Trotter steps. The gate count refers to the total number of quantum gates. The circuit depth is the number of sequential steps allowing for quantum gates acting on neighboring qubits to be executed in parallel (see text for details). η denotes the number of electrons and N the number of spin-orbitals in the active space for a given molecule. k denotes the number of products in the k-UpCCGSD wavefunction.

The quantity $|\langle\psi_0|\psi_1(\theta)\rangle|^2$ can therefore be rewritten as $|\langle 0|\hat{U}_0^\dagger\hat{U}_1|0\rangle|^2$. This is exactly equal to the probability that the zero state will be observed when the state $\hat{U}_0^\dagger\hat{U}_1|0\rangle$ is measured in the computational basis. Consequently, the magnitude of the overlap may estimated by repeated state preparation and measurement. Because of the necessity to apply both $\hat{U}_1$ and $\hat{U}_0^\dagger$, these measurements require a doubling of the circuit depth compared to the other observables. However, the overall cost of the measurements required for the OC-VQE approach for quantum chemistry in a molecular orbital basis will still be dominated by the measurement of the $\mathcal{O}(N^4)$ terms in the original Hamiltonian.

4.5 Benchmark implementations on a Classical Computer

4.5.1 Computational Details

All the full configuration interaction (FCI) calculations needed to benchmark the demonstration examples in this work are performed through `Psi4` [262] along with its `OpenFermion` [263] interface. All UCCSD calculations are performed with an in-house code that uses `OpenFermion` [263] together with `TensorFlow` [264] for efficient gradient evaluations. The energy as a function of the cluster amplitudes is computed variationally as in Eq. 4.9 and the gradient of this function is used in conjunction with `SciPy`'s implementation of the BFGS algorithm [265], a quasi-Newton method for optimization which does not require explicit calculation of the Hessian. The limit of our code is about 16 spin-orbitals, which allowed us to examine various model systems presented below. A production level code may follow the implementation of Evangelista [226], which may facilitate prototyping VQE ansätze.

All other calculations required for the demonstrations presented in this work are done with the development version of `Q-Chem`.[42] All calculations were performed with the frozen

core approximation applied to oxygen and nitrogen.

There are several possible strategies for optimizing the amplitudes of a k-UpCCGSD wavefunction. One attractive approach is to optimize only one set of amplitudes in $\hat{T}^{(k)}$, while fixing all the amplitudes associated with a $(k-1)$-UpCCGSD wavefunction. This has the potential benefit of reducing the extra computational cost for optimization of more amplitudes as the index k is increased. However, we found that in practice, this optimization generally requires a larger k value to achieve chemical accuracy then simultaneous optimization of all k sets of amplitudes in k-UpCCGSD. Therefore, for the results presented below, we optimized all k sets of amplitudes simultaneously.

In general, with UCC methods it is not clear whether one obtains global minima of the energy for a given class of wavefunctions. Efficiently obtaining a global minimum in a non-linear optimization problem is an open problem in applied mathematics.[266] In order to approximate the true minimum, each gradient-based optimization was therefore carried out between thirty and two hundred times (depending on the cost) starting from randomly chosen initial points.

We note that the BFGS optimization as we have performed it here on a classical computer is unsuitable for use on a quantum device due to the stochastic error associated with the measurement of observables in the VQE framework. Given this, it will be necessary to find better ways to handle optimization for large scale VQE experiments.

4.5.2 Applications to Chemical Systems

We now describe application of the three UCC ansätze, UCCSD, UCCGSD, and k-UpCCGSD, to three molecular systems possessing different geometries, namely H_4, H_2O, and N_2.

4.5.2.1 H_4(in D_{4h} and D_{2h} symmetry)

H_4 is an interesting model system for testing CC methods with singles and doubles. We study here the potential energy curve of H_4 for deviations from the square geometry with fixed bond distance, $R_{\text{H-H}} = 1.23$ Å. Then we vary R in the following coordinate system (values are given in Å),

$$\begin{aligned}
&\text{H1} : (0, 0, 0) \\
&\text{H2} : (0, 0, 1.23) \\
&\text{H3} : (R, 0, 0) \\
&\text{H4} : (R, 0, 1.23).
\end{aligned}$$

This particular geometry setup has been used by others in Refs. 236, 267–271. At $R = 1.23$ Å (the D_{4h} geometry), we have two quasidegenerate RHF determinants, which poses a great challenge to single-reference CC methods with only singles and doubles.

We assess the ground state UCC methods including those developed in this work and compare them against RCCSD and coupled-cluster valence bond with singles and doubles

(CCVB-SD) within the minimal basis, STO-3G.[272, 273] CCVB-SD corrects for ill-behaving quadruples in RCCSD and is able to break any number of bonds exactly within the valence active space. In this sense, it is one of the most powerful classical CC methods with singles and doubles within the valence active space. There are two solutions for RCCSD and CCVB-SD, each one being obtained with one of the two low-lying RHF determinants. The two RHF solutions cross at $R = 1.23$ Å. We present the results obtained with the lowest RHF reference for a given R.

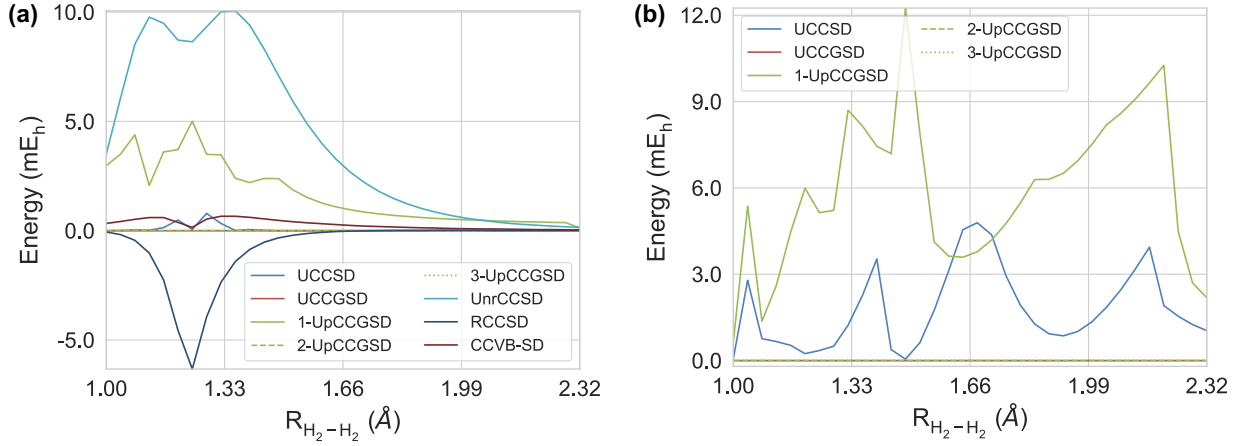

Figure 4.1: The error in the absolute energy of the various CC methods examined in this work for (a) the ground state and (b) the first excited state of H_4 as a function of the distance between two H_2's. The basis set used here is STO-3G ($N = 8$, $\eta = 4$). For both plots, UCCGSD, 2-UpCCGSD, and 3-UpCCGSD are overlapping near zero error in the absolute energy.

In Figure 4.1 (a), we present the absolute energy error in ground state of the aforementioned CC methods as a function of R. We first point out that unrestricted CCSD (UnrCCSD) performs worst in an absolute sense among the methods examined here. This is because the H-H distance in each H_2 is stretched enough to get spin-contamination on each H_2. This makes the entire potential energy curve of H_4 heavily spin-contaminated within the range of R examined. RCCSD has clearly gone non-variational while CCVB-SD remains above the exact ground state energy at all distances. Except 1-UpCCGSD and UCCSD, all the UCC variants are numerically exact. 1-UpCCGSD is much worse than all the rest of UCC methods and adding one more product (i.e. 2-UpCCGSD) makes the energy numerically exact.

Unlike full doubles CC models, the energy of k-UpCCGSD is generally not invariant under unitary rotations among orbitals. This is likely a primary cause of the multiple unphysical local minima observed for 1-UpCCGSD. This problem can be ameliorated by increasing the value of k, as shown in Figure 4.1 (a). The difficulty of optimizing pair wavefunctions has been discussed in some earlier works. Interested readers are referred to Ref. 274.

In Figure 4.1 (b), the performance of UCC methods on the first excited state of H_4 was assessed within the OC-VQE framework. It is clear that UCCSD and 1-UpCCGSD exhibit larger errors than those of the ground state. This illustrates a potential drawback of OC-VQE in terms of accuracy when we do not have a high quality ground state. However, with better ansätze this drawback can be made insignificant. The excited states from UCCGSD, 2-UpCCGSD, and 3-UpCCGSD are numerically exact, illustrating the power of these novel wavefunction ansätze which go beyond the capability of UCCSD while also offering a lower asymptotic scaling.

(a)

	UCCSD	UCCGSD	1-UpCCGSD	2-UpCCGSD	3-UpCCGSD	UnrCCSD	RCCSD	CCVB-SD
NPE	0.79	0.00	4.88	0.00	0.00	9.90	6.35	0.62

(b)

	UCCSD	UCCGSD	1-UpCCGSD	2-UpCCGSD	3-UpCCGSD
NPE	4.75	0.00	11.62	0.00	0.00

Table 4.2: The non-parallelity error (NPE) (mE_h) in (a) the ground state and (b) the first excited state of H_4 within the STO-3G basis set ($N = 8$, $\eta = 4$).

In Table 4.2, we present the non-parallelity error (NPE) in the ground state and the first excited state for each CC method. NPE is defined as the difference between the maximum and minimum error and is a useful measure of performance, since we are interested in relative energetics in most chemical applications. In the ground state, UnrCCSD is the worst in terms of NPE. CCVB-SD is comparable to UCCSD and RCCSD and 1-UpCCGSD are comparable. UCCGSD, 2-UpCCGSD, and 3-UpCCGSD all have zero NPEs as they are numerically exact everywhere. In the case of the first excited state, UCCSD and 1-UpCCGSD performs worse than their ground state performance as observed before. All the other UCC methods are numerically exact.

We repeat the same calculations within the 6-31G basis. There are a total of 16 spin-orbitals in this case: in terms of resource on a quantum device this corresponds to the most expensive calculation reported in this work. This test is interesting because some dynamic correlation effects can be captured in 6-31G, in contrast to STO-3G, and these pose a greater challenge to pair CC methods.

In Table 4.3, the error in the ground state is presented as a function of R. In terms of NPE, UCCGSD is again numerically exact and thus best. 2-UpCCGSD and 3-UpCCGSD are within 1 mE_h of UCCGSD and exhibit larger errors than the corresponding results in the STO-3G basis. RCCSD performs better with the 6-31G basis set and it is better than UCCSD. As it clearly becomes non-variational at $R = 1.23$ Å, we suspect that this is a fortuitous outcome for RCCSD. Moreover, UnrCCSD is the worst amongst the traditional CC methods considered in this work, which emphasizes the importance of spin-purity.

R	UCCSD	UCCGSD	1-UpCCGSD	2-UpCCGSD	3-UpCCGSD	UnrCCSD	RCCSD	CCVB-SD
1.01	0.25	0.00	3.56	0.33	0.05	3.14	0.24	0.55
1.12	0.84	0.00	7.49	0.54	0.05	5.78	0.40	1.34
1.23	3.83	0.00	18.20	0.38	0.05	3.24	-0.79	3.21
1.34	0.86	0.00	14.99	0.46	0.03	6.00	0.01	1.41
1.45	0.27	0.00	6.49	0.24	0.02	5.56	0.10	0.74
1.56	0.13	0.00	2.48	0.17	0.02	3.47	0.12	0.48
1.67	0.08	0.00	1.86	0.11	0.01	1.99	0.11	0.35
1.78	0.06	0.00	1.76	0.10	0.01	1.14	0.10	0.26
1.89	0.04	0.00	1.51	0.04	0.01	0.67	0.09	0.21
2.00	0.03	0.00	1.14	0.03	0.01	0.40	0.08	0.17
2.11	0.03	0.00	2.28	0.04	0.00	0.25	0.07	0.13
2.22	0.02	0.00	2.20	0.03	0.01	0.16	0.06	0.11
NPE	3.81	0.00	17.06	0.51	0.05	5.84	1.19	3.10

Table 4.3: The error in absolute energy (mE_h) and non-parallelity error (NPE) (mE_h) in the ground state of H_4 within the 6-31G basis ($N = 16$, $\eta = 4$) as a function of the distance (R) between two H_2's (Å).

R	UCCSD	UCCGSD	1-UpCCGSD	2-UpCCGSD	3-UpCCGSD
1.01	2.55	0.00	9.49	0.92	0.12
1.12	12.17	0.00	4.66	1.02	0.19
1.23	6.99	0.00	14.27	1.74	0.15
1.34	10.46	0.00	13.30	1.02	0.15
1.45	14.07	0.00	12.16	1.31	0.18
1.56	11.79	0.00	9.87	0.91	0.18
1.67	8.94	0.00	9.17	0.80	0.17
1.78	7.71	0.00	10.76	0.63	0.15
1.89	9.38	0.00	13.44	0.58	0.16
2.00	9.74	0.00	17.26	0.49	0.14
2.11	9.46	0.00	17.53	0.46	0.15
2.22	8.44	0.00	21.14	0.42	0.18
NPE	11.52	0.00	16.47	1.31	0.07

Table 4.4: The error in absolute energy (mE_h) and non-parallelity error (NPE) (mE_h) in the first excited state of H_4 within the 6-31G basis ($N = 16$, $\eta = 4$) as a function of the distance (R) between two H_2's (Å).

Lastly, we discuss the quality of the first excited state from UCC methods on H_4 within the 6-31G basis set [275] as presented in Figure 4.4. It is immediately obvious that the degraded ground state performance of UCCSD is amplified in the excited state calculation and that 1-UpCCGSD continue to perform poorly. This is consistent with the STO-3G results. However, it should be emphasized that UCCGSD is still numerically exact and the 3-UpCCGSD error is still less than 0.1 mE_h. UCCSD's poor performance strongly validates our development of better wavefunction ansätze beyond UCCSD, particularly for obtaining good excited states within the OC-VQE framework.

4.5.2.2 Double Dissociation of H_2O (C_{2v})

The double dissociation of H_2O is another classic test platform for various wavefunction methods.[276–279] As we stretch two single bonds, we have total 4 electrons that are strongly entangled. The traditional RCCSD method can easily become non-variational, as will be demonstrated below. At a fixed angle $\theta_{HOH} = 104.5°$ and within the C_{2v} symmetry, we varied the bond distance between H and O and obtained potential energy curves for various CC methods within the STO-3G basis set.[272, 273]

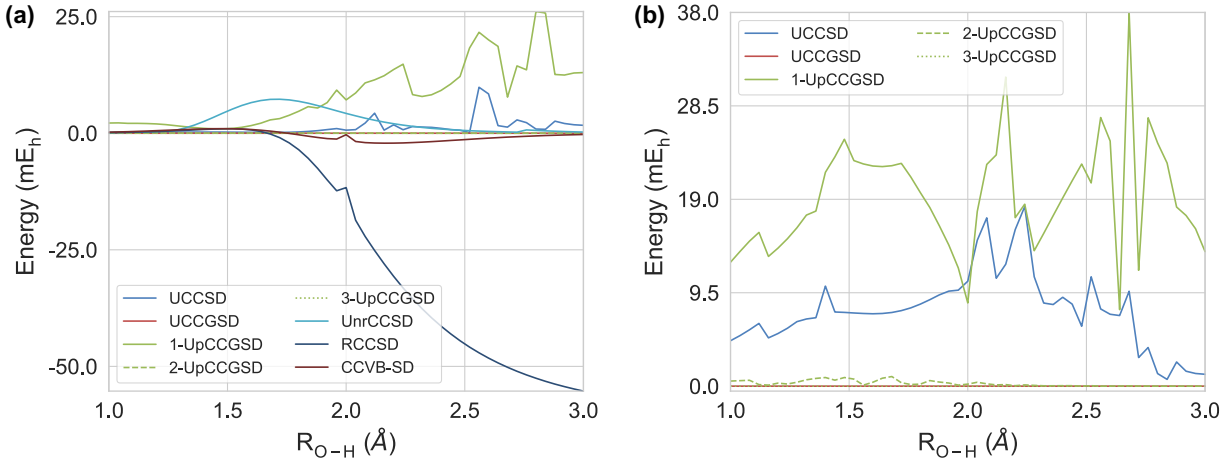

Figure 4.2: The error in the absolute energy of the various CC methods examined in this work for (a) the ground state and (b) the first excited state of H_2O as a function of the distance between O and H. The basis set used here is STO-3G ($N = 12$, $\eta = 8$). For the ground state (a), UCCGSD, 2-UpCCGSD, and 3-UpCCGSD are overlapping near zero error in the absolute energy. For the excited state (b), UCCGSD and 3-UpCCGSD are overlapping near zero error in the absolute energy.

In Figure 4.2, the error in the absolute energy of the ground state and the first excited state of H_2O is presented as a function of the $R_{O\text{-}H}$ distance. In Figure 4.2 (a), RCCSD performs much worse than CCVB-SD and UnrCCSD especially after 1.75 Å and exhibits a

very significant non-variationality upon increasing the O-H distance. There is a kink between 2.02 Å and 2.04 Å in both RCCSD and CCVB-SD, that is due to a change in the character of the converged amplitudes. The RHF solutions for these CC calculations are delocalized and obey spatial symmetry. We also note that there is another spatially-symmetric RHF solution that is lower in energy than the orbitals we found. This solution starts to appear from 2.02 Å and is more stable than the other for longer bond distances. This solution has orbitals either localized on O or two H's. This reference yields much higher CCVB-SD and RCCSD energies at 2.04 Å. These two low-lying RHF solutions might cause multiple amplitudes solutions close in energy. We found that the largest T_1 amplitude of CCVB-SD is 0.28 at 2.02 Å and 0.07 at 2.04 Å. This discontinuity does not appear with a larger basis set such as cc-pVDZ so it is likely an artifact of using a minimal basis. With the delocalized RHF solution, CCVB-SD performs best among the classical CC methods examined here.

UCCSD and 1-UpCCGSD perform much worse than the other UCC methods, as also observed above in H_4. Other UCC methods are more or less numerically exact on the scale of the plot. The performance of the first excited state as presented in Figure 4.2 (b) is consistent with the ground state performance. UCCGSD and 3-UpCCGSD are numerically exact and 2-UpCCGSD is within 1 mE_h for all $R_{\text{O-H}}$ values. UCCSD and 1-UpCCGSD do not deliver reliable excited state energies.

(a)

	UCCSD	UCCGSD	1-UpCCGSD	2-UpCCGSD	3-UpCCGSD	UnrCCSD	RCCSD	CCVB-SD
NPE	9.75	0.00	25.16	0.07	0.00	7.15	56.26	3.11

(b)

	UCCSD	UCCGSD	1-UpCCGSD	2-UpCCGSD	3-UpCCGSD
NPE	17.57	0.00	30.22	0.98	0.01

Table 4.5: The non-parallelity error (NPE) (mE_h) in (a) the ground state and (b) the first excited state of H_2O within the STO-3G basis set ($N = 12$, $\eta = 8$).

In Table 4.5, we present the NPE of both the ground state in (a) and the first excited state in (b) of H_2O. UCCGSD, 2-UpCCGSD, and 3-UpCCGSD all yield reliable potential energy curves, while curves from the other methods are not as reliable. It should be noted that UCCSD performs worse than the best classical method considered here, UnrCCSD, but improved wavefunctions such as UCCGSD and 3-UpCCGSD are more or less exact for both states.

4.5.2.3 Dissociation of N_2

The dissociation of N_2 is very challenging for CC methods with only singles and doubles.[279, 280] At a stretched geometry, there are a total of 6 electrons that are strongly entangled. RCCSD exhibits severe non-variationality and UnrCCSD has a non-negligible non-parallelity error due to poor performance in the intermediate bond length (spin-recoupling) regime.

To obtain a qualitatively correct answer within the traditional CC framework with a RHF reference, one would need RCCSD with the addition of triples, quadruples, pentuples and hextuples which contains far more excitations than RCCSD. Alternatively, one could employ CCVB-SD as it is able to break N_2 exactly within the STO-3G basis.[272, 273]

R	UCCSD	UCCGSD	1-UpCCGSD	2-UpCCGSD	3-UpCCGSD	4-UpCCGSD	5-UpCCGSD	6-UpCCGSD
1.0	1.63	0.44	37.69	9.46	1.55	0.23	0.07	0.03
1.4	4.04	1.50	27.36	10.08	4.00	1.58	0.67	0.27
1.8	3.56	0.77	23.19	10.00	3.05	1.30	0.78	0.38
2.2	11.80	0.42	45.90	15.66	2.70	0.32	0.25	0.19
2.6	12.97	0.16	60.54	15.46	1.86	0.07	0.07	0.02
3.0	14.64	0.43	64.28	18.31	3.73	0.08	0.01	0.03
NPE	13.01	1.33	41.09	8.84	2.45	1.51	0.77	0.37

R	RCCSD	UnrCCSD	CCVB-SD
1.0	2.58	2.58	3.98
1.4	11.53	7.47	14.46
1.8	-19.55	9.08	3.82
2.2	-159.58	2.09	-4.94
2.6	-199.20	0.48	-2.21
3.0	-208.21	0.11	-0.59
NPE	219.74	8.98	19.40

Table 4.6: The error in absolute energy (mE_h) and non-parallelity error (NPE) (mE_h) in the ground state of N_2 within the STO-3G basis ($N = 16$, $\eta = 10$) as a function of the distance (R) between two N's (Å).

In Table 4.6, we present the NPEs for ground state N_2 for the various CC methods examined in this work. In terms of the number of electrons that are strongly correlated, this system is the most challenging problem investigated in this work. RCCSD is highly non-variational and not acceptably reliable for any distance considered except for 1.0 Å. CCVB-SD exhibits non-variationality but eventually dissociates properly. However, in terms of NPE CCVB-SD is not reliable. UnrCCSD has a NPE of 8.98 mE_h due to poor performance at intermediate bond lengths. UCC methods also struggle to properly dissociate. UCCSD is worse than UnrCCSD in terms of NPE. Furthermore, UCCGSD is now not numerically exact, with a NPE of 1.33 mE_h. In order to achieve a NPE less than 1 mE_h, k needs to be greater than 4. The fact that k-UpCCGSD is systemetically improvable and can achieve very accurate results with a lower cost than UCCSD is very encouraging.

Lastly, we discuss the performance of the UCC methods in the first excited state $^1\Pi_g$ which is presented in Table 4.7. Obtaining an accurate description for the first excited state of N_2 within the OC-VQE framework is extremely challenging. The best performing UCC method is 6-UpCCGSD with a NPE of 1.61 mE_h. UCCGSD exhibits a NPE of 7.79 mE_h, which, while certainly better than that of UCCSD (31.94 mE_h), is not close to the threshold for chemical accuracy. These results highlight the challenge of constructing wavefunction

R	UCCSD	UCCGSD	1-UpCCGSD	2-UpCCGSD	3-UpCCGSD	4-UpCCGSD	5-UpCCGSD	6-UpCCGSD
1.0	25.77	1.00	45.13	17.56	6.46	2.85	1.33	0.70
1.4	28.35	3.21	62.22	24.20	9.41	4.72	2.62	1.71
1.8	40.46	7.89	42.53	14.68	2.97	1.77	0.78	0.94
2.2	31.57	6.55	8.57	7.32	4.16	2.44	1.78	0.64
2.6	15.09	1.13	-0.32	1.01	0.43	0.37	0.35	0.11
3.0	8.52	0.10	-5.05	-2.83	0.12	0.14	0.05	0.14
NPE	31.94	7.79	67.27	27.02	9.29	4.59	2.57	1.61

Table 4.7: The error in absolute energy (mE_h) and non-parallelity error (NPE) (mE_h) in the first excited state of N_2 within the STO-3G basis ($N = 16$, $\eta = 10$) as a function of the distance (R) between two N's (Å).

ansätze capable of accurately representing the excited states of strongly correlated systems.

4.5.2.4 Discussion of Excited State Energies

We analyze here the error of UCCGSD for the first excited state of N_2 at 1.8 Å , which is significant, at 7.89 mE_h. For the purpose of demonstration, we ran another set of calculations with an exact orthogonality constraint constructed from the exact ground state. The results obtained with this exact constraint are presented in Table 4.8.

Determinants	Error	Reference
1	10.23	Ground State RHF
2	3.18	Singly Excited Configuration ($\pi_x \rightarrow \pi_x^*$)
4	0.45	Two Singly Excited Configurations ($\pi_x \rightarrow \pi_x^*$ and $\pi_y \rightarrow \pi_y^*$)

Table 4.8: The error in absolute energy (mE_h) for the first excited state of N_2 at 1.8 Å when using the exact ground state for the OC-VQE penalty term together with the UCCGSD ansatz and multiple reference states. Here $\eta = 10$ electrons in $N = 8$ spin-orbitals.

The ground state RHF determinant is likely to be a poor reference state for excited states. This is clearly demonstrated in Table 4.8 with an error of 10.23 mE_h in the case of the ground state RHF reference. The first excited state of N_2 is a rather simple electronic state in the sense that it is mainly dominated by single excitations from the ground state wave function. At 1.8 Å, these single excitations are mainly $\pi \rightarrow \pi^*$ and there are a total of two excitations like this along x and y cartesian components assuming that the molecular axis is the z-axis. Therefore, a more sensible starting point for OC-VQE would be to use these singly excited configurations. This leads to an error of 3.18 mE_h with two determinants of the $\pi_x \rightarrow \pi_x^*$ type and to an error of 0.45 mE_h with additional two determinants of the $\pi_y \rightarrow \pi_y^*$ type. A total of 4 determinants (or 2 spin-adapted singlet configurations) were enough to reach the

chemical accuracy. In general, a much more sensible reference state for excited states like this can be cheaply obtained via regular linear response methods such as configuration interaction singles.[281] Furthermore, the natural transition orbital basis [281] can be used to generate a minimal multi-determinantal reference which will be usually of two determinants.

4.5.2.5 Summary of Chemical Applications

Molecule	Basis	Electrons	Spin-Orbitals	Ansatz/Method	Amplitudes	NPE (ground)	NPE (excited)
H_4	STO-3G	4	8	UCCSD	20	0.79	4.75
				UCCGSD	27	0.00	0.00
				1-UpCCGSD	12	4.88	11.62
				2-UpCCGSD	24	0.00	0.00
				3-UpCCGSD	36	0.00	0.00
H_4	6-31G	4	16	UCCSD	156	3.81	11.52
				UCCGSD	434	0.00	0.00
				1-UpCCGSD	56	17.06	16.47
				2-UpCCGSD	112	0.51	1.31
				3-UpCCGSD	168	0.05	0.07
H_2O	STO-3G	8	12	UCCSD	72	9.75	17.57
				UCCGSD	135	0.00	0.00
				1-UpCCGSD	30	25.16	30.22
				2-UpCCGSD	60	0.07	0.98
				3-UpCCGSD	90	0.00	0.01
N_2	STO-3G	10	16	UCCSD	240	13.01	31.94
				UCCGSD	434	1.33	7.79
				1-UpCCGSD	56	41.09	67.27
				2-UpCCGSD	112	8.84	27.02
				3-UpCCGSD	168	2.45	9.29
				4-UpCCGSD	224	1.51	4.59
				5-UpCCGSD	280	0.77	2.57
				6-UpCCGSD	336	0.37	1.61

Table 4.9: A summary of the results of this work: the number of amplitudes and the non-parallelity error (NPE) (mE_h) for each method applied to each molecule and basis. The excited NPEs are obtained with restricted Hartree-Fock references.

In Table 4.9, we present a summary of the results in this section. In particular, we focus on the tradeoff between the number of amplitudes and the accuracy (i.e. NPE). UCCSD does not perform very well given the number of amplitudes. k-UpCCGSD with a similar number of amplitudes always performs better than UCCSD which demonstrates the compactness of k-UpCCGSD. UCCGSD offers very accurate energies at the expense of requiring a significant number of amplitudes. In all cases we considered it was possible to achieve chemical accuracy using k-UpCCGSD with less amplitudes than UCCGSD. We also note that excited states are in general more challenging than ground state calculations. Furthermore, there is no fortuitous error cancellation in excitation gaps in this approach. Therefore, it is important to obtain

near-exact energies for both ground and excited states in order to achieve chemical accuracy for excitation gaps. As noted above, using multi-determinantal reference wavefunction can improve the accuracy significantly. Considering the tradeoff between the cost and the accuracy, we recommend k-UpCCGSD for general applications. However, it should be noted that for k-UpCCGSD to be effective, it is essential to choose k large enough to obtain sub-chemical accuracy. Otherwise the lack of smoothness associated with this novel ansatz will inhibit application goals such as exploring potential energy surfaces.

4.6 Summary and Outlook

In this work, we have presented a new unitary coupled cluster ansatz suitable for preparation, manipulation, and measurement of quantum states describing molecular electronic states, k-UpCCGSD, and compared its performance to that of both a generalized UCC ansatz UCCGSD, and the conventional UCCSD. A resource analysis of implementation of these new wavefunctions on a quantum device showed that k-UpCCGSD offers the best asymptotic scaling with respect to both circuit depth and amplitude count. Specifically, the circuit depth for k-UpCCGSD scales as $\mathcal{O}(kN)$ while that for UCCGSD scales as $\mathcal{O}(N^3)$ and that for UCCSD with $\mathcal{O}((N-\eta)^2\eta)$.

We performed classical benchmark calculations with these ansätze for the ground state and first excited state of three molecules with very different symmetries, H_4 (STO-3G, 6-31G), H_2O (STO-3G), and N_2 (STO-3G), to analyze the relative accuracy obtainable from these ansätze. Comparison was also with results from conventional coupled cluster wavefunctions where relevant. The benchmarking calculations show that the new ansatz of unitary pair coupled-cluster with generalized singles and doubles (k-UpCCGSD) offers a favorable tradeoff between accuracy and time complexity.

We also made excited state calculations, using a variant of the recently proposed orthogonally constrained variational quantum eigensolver (OC-VQE) framework [231]. Our implementation of this takes advantage of the close relation of this approach to some excited state methods in quantum chemistry.[231, 257, 282] OC-VQE works as a variational algorithm where there a constraint in imposed on the energy minimization in order to ensure the orthogonality of an excited state to a ground state wavefunction that has been previously obtained from a ground state VQE hybrid quantum-classical calculation. This approach requires only a modest increase in resources to implement on a quantum device compared to the resources required for ground state VQE, and is furthermore capable of targeting states outside of a small linear response subspace defined from the VQE ground state.

Assessing the classically computed potential energy curves of these three molecules, we found that the error associated with excited states obtained by the OC-VQE approach in conjunction with the standard UCCSD reference, is considerably larger than the error of the ground state calculation. The excited states of UCC singles and doubles are never of high quality, except for simple two-electron systems where UCCSD is exact.[231] We found that energies of both ground and excited states can be greatly improved by employing

either UCCGSD, i.e., UCC with generalized singles and doubles, or the k-fold products of k-UpCCGSD. Furthermore, we demonstrate that the quality of excited state calculations in the OC-VQE framework can be dramatically improved by choosing a chemically motivated reference wavefunction.

UCCGSD was found to be numerically exact for H_4 (STO-3G, 6-31G) and H_2O (STO-3G) for both ground and excited states. However, its non-parallelity error (NPE) is 1.33 mE_h for the ground state of N_2 and 7.79 mE_h for the first excited state of N_2. k-UpCCGSD was found to be numerically exact for a large enough k, where the required value of k increases with the difficulty of the problem. It would be interesting to study the required value of k for fixed accuracy on a broader class of problems in the future.

In summary, this work demonstrates the advantages of wavefunction ansätze that go beyond UCCSD and indicates the desirability of further refinement of such ansätze to forms that are accurate for both ground and excited states. The performance of k-UpCCGSD is particularly encouraging, showing a tradeoff between accuracy and resource cost that allows chemical accuracy to be achieved with resources scaling only linearly in the number of spin-orbitals. Our analysis of excited states indicates that these pose significant challenges and there is a need for focus on these. In particular, we anticipate that further development of novel algorithms not within the variational framework may be necessary to obtain high quality excited state energetics, particularly when working with an approximate ground state.

Finally we note that the wavefunctions we have investigated in this work can be fruitfully combined with existing classical approximations to UCC based on the truncation of the Baker-Campbell-Hausforff expansion of $\langle\phi_0|e^{T^\dagger - T} H e^{T - T^\dagger}|\phi_0\rangle$. [283–285] This would allow for the efficient initialization of the cluster amplitudes, making it possible to further optimize them using the VQE hybrid approach to quantum computation, and also avoiding the difficulties posed by a random initialization.[217] In future work, it would be interesting to further explore the balance between the cost and accuracy of unitary coupled cluster ansätze obtained here by building on chemically motivated approximations. Two especially promising directions that we believe could yield a further reduction of the number of amplitudes and the gate depth required for a fixed accuracy, are i) the adaption the recently proposed full coupled-cluster reduction [286] method for use on a quantum computer, and ii) the elimination of singles amplitudes through the use of approximate Brückner orbitals [287–291] obtained by classical pre-processing. Ultimately, the resulting wavefunctions could themselves serve as inputs to a fully quantum computation of more accurate ground and excited state energies, e.g., with the quantum phase estimation algorithm, or to a quantum simulation of quantum dynamics.

4.7 Additional Computational Details

In this section, we shall describe some of the operational details of the calculations we performed in order to make their reproduction easier. We shall not spend much time discussing the traditional quantum chemical calculations we performed, except to slightly

expand on their description in the main body of this chapter. We shall mainly focus on describing in more detail the unitary coupled cluster calculations we performed, and on highlighting the software tools that we used to strike a balance between ease of implementation and numerical performance.

In addition to the standard Hartree-Fock (HF) and full configuration interaction (FCI) calculations, we performed three other kinds of classical electronic structure calculations for the purpose of comparison against our unitary coupled cluster results. We applied both the spin-restricted and spin-unrestricted flavors of coupled cluster with singles and doubles, abbreviating these methods as RCCSD and UnrCCSD respectively. We already gave a very high-level summary of the coupled cluster approach in Section 2.3.0.3 and we refer the reader to Ref. 33 for a comprehensive reference. For pedagogical purposes, we note here that the difference between these two versions of coupled cluster lies with the choice of reference state and the symmetries of the cluster operator. Throughout this work, restricted coupled cluster refers to the spin-adapted singlet coupled cluster which is standard for closed-shell molecules [33]. The use of the restricted Hartree-Fock reference state combined with the spin-adapted cluster operator guarantees that RCCSD produces a singlet state. Unrestricted coupled cluster, by contrast, uses an unrestricted Hartree-Fock reference state and a cluster operator that does not conserve S^2.

The third method we use as a point of comparison, coupled cluster valence bond with singles and doubles (CCVB-SD) [49, 236], is a somewhat more niche approach that RCCSD or UnrCCSD. CCVB-SD can be understood as a modified version of RCCSD [292]. Specifically, CCVB-SD involves a technical change to the collection of coupled equations that make up RCCSD. We do not review the details of this method here, but we note that it results in a cubically scaling approach that has been shown to be effective at treating strongly-correlated closed shell systems [49, 236]. Like the RCCSD and UnrCCSD calculations, the CCVB-SD data was obtained using the `Q-Chem` software package [42].

Except for their application to the same systems, these benchmark calculations were performed independently of the unitary coupled cluster calculations that constitute most of the data presented in this chapter. As we mentioned in Section 4.5.1, the unitary coupled cluster calculations were performed with an in-house code that evaluated the unitary coupled cluster energy (Eq. 4.9). The coefficients of the second-quantized electronic structure Hamiltonian (Eq. 2.6) were one of the essential inputs to this energy evaluation. In order to determine them, we used the electronic structure package `Psi4` to perform Hartree-Fock calculations in the various basis sets we mentioned above [262]. We then loaded the resulting integrals for the molecular orbital basis, h_{pq} and g_{pqrs}, into the software package `OpenFermion` [263]. In order to perform our numerics as efficiently as possible, we used the `get_number_preserving_sparse_operator` function from `OpenFermion` to obtain a matrix representation of the Hamiltonian projected into the subspace corresponding to $S_z = 0$ and the correct total number of electrons for each of the systems we considered.

In Table 4.10, we present the coefficients of the molecular orbitals for H_4 in terms of the atomic STO-3G atomic orbitals. For a Hydrogen atom centered at the origin, the STO-3G basis set prescribes a single spatial orbital, designed to approximate the $1s$ orbital of the

Orbital	MO 1	MO 2	MO 3	MO 4
AO 1	0.360429	-0.552191	-0.552218	-0.757377
AO 2	0.360424	0.552218	-0.552191	0.75738
AO 3	0.360429	0.552191	0.552218	-0.757377
AO 4	0.360424	-0.552218	0.552191	0.75738

Table 4.10: The coefficients of the molecular orbitals (MO) for our square ($R = 1.23$ Å) H_4 system, in terms of the STO-3G atomic orbital (AO) basis set that we used for this system. This minimal basis set contains one spatial orbital for each Hydrogen atom. The molecular orbitals are numbered in order of increasing energy. The atomic orbitals are numbered in a counterclockwise fashion. The symmetry of the square leads to a degeneracy between the second and third molecular orbitals, which is clearly visible in the coefficients.

analytic solution by a sum of three Gaussian functions,

$$\left(\frac{2 \cdot 0.154328}{\pi}\right)^{\frac{3}{4}} e^{-3.42525 r^2} + \left(\frac{2 \cdot 0.535328}{\pi}\right)^{\frac{3}{4}} e^{-.623913 r^2} + \left(\frac{2 \cdot 0.444634}{\pi}\right)^{\frac{3}{4}} e^{-.168855 r^2}. \quad (4.21)$$

In general, atomic orbitals centered on different atoms will not be orthonormal. As part of the Hartree-Fock procedure, the solution to a generalized eigenvalue problem allows for the construction of a set of orthonormal molecular orbitals.

Before we expand upon our approach to the energy evaluation, we shall explain the exact form of the cluster operator that we used for our various UCC calculations. Because we were focusing on singlet states throughout this work, we used spin-restricted versions of the cluster operator designed to prepare singlet states from the closed-shell RHF determinants that we used as our reference states. In order to preserve the singlet nature of the reference states, we followed Ref. 33 and demanded that our cluster operators satisfy the following commutation relations,

$$[T, S_z] = 0, \qquad [T, S_+] = 0, \qquad [T, S_-] = 0, \quad (4.22)$$

with S_z, S_+, and S_- written below below for convenience,

$$S_z = \frac{1}{2} \sum_p \left(a^\dagger_{p\alpha} a_{p\alpha} - a^\dagger_{p\beta} a_{p\beta} \right) \quad (4.23)$$

$$S_+ = \sum_p a^\dagger_{p\alpha} a_{p\beta} \quad (4.24)$$

$$S_- = \sum_p a^\dagger_{p\beta} a_{p\alpha}. \quad (4.25)$$

Ref. 33 explains that one can satisfy these conditions by making the following choice of

cluster operators,

$$\begin{aligned} T &= T_1 + T_2 \\ T_1 &= \sum_{ai} t_i^a E_{ai} \\ T_2 &= \frac{1}{2} \sum_{abij} t_{ij}^{ab} E_{ai} E_{bj}, \qquad t_{ij}^{ab} = t_{ji}^{ba}. \end{aligned} \tag{4.26}$$

Here E_{ai} is the singlet one-electron excitation operator,

$$E_{pq} = a_{p\alpha}^\dagger a_{q\alpha} + a_{p\beta}^\dagger a_{q\beta}. \tag{4.27}$$

Note that we use the letters a and b to denote unoccupied orbitals, i and j to denote occupied orbitals, and p and q to denote an orbital which could be either, as explained above in Section 4.3. Our actual implementation enforced these symmetries in an equivalent way that yielded some extra (redundant) free parameters, but the most straightforward way would be to use Eq. 4.26 directly. Our cluster operator for UCCGSD took a similar approach, except that we used summations over all of the orbitals with no distinction between occupied and virtual orbitals.

$$\begin{aligned} T_1 &= \sum_{pq} t_p^q E_{qp} \\ T_2 &= \frac{1}{2} \sum_{pqrs} t_{pq}^{rs} E_{rp} E_{sq}, \qquad t_{pq}^{rs} = t_{qp}^{sr}. \end{aligned} \tag{4.28}$$

Note that because we use these operators to form the antihermitian argument to the exponential, $T - T^\dagger = T_1 - T_1^\dagger + T_2 - T_2^\dagger$, there are some additional redundancies between the parameters. We accounted for these in our calculations of the number of amplitudes listed in the tables throughout this work. For the k-UpCCGSD ansatz, the choice of cluster operator is described in detail in the text above.

In order to conveniently evaluate the derivative of the energy with respect to the cluster amplitudes ($\{t_i^a\}, \{t_{ij}^{ab}\}$ for the standard UCC), we took a roundabout route to evaluate the energy of Eq. 4.9. For each of the terms in the cluster operator (e.g. $\hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i$), we obtained a matrix representation using `get_number_preserving_sparse_operator`, in the same way that we did for the Hamiltonian. We then used the `TensorFlow` software library of Ref. 264 to write a function that would multiply the terms by the appropriate coefficients, apply the matrix exponential of the cluster operator to the Hartree-Fock reference state, and evaluate the energy. The advantage of using `TensorFlow`, or some other software package that supports automatic differentiation, was that we were able to automatically obtain the gradient of the energy function with respect to our parameters. Furthermore, despite the fact that the gradient involves a separate derivative for each parameter, automatic differentiation allowed us to evaluate it using an amount of time comparable to $O(1)$ evaluations of the function which computes the energy.

The remaining aspects of the numerical calculations were performed in a standard way. We performed the optimization using `SciPy`'s implementation of the BFGS algorithm [265], as discussed above. BFGS is a standard quasi-Newton method, i.e., it builds an approximation to the Hessian matrix from the repeated gradient queries and uses this approximate Hessian along with the gradient information to approximate search for a local minimum [293–297]. As we mentioned in Section 4.5.1, we performed the optimization multiple times from different randomly chosen initial points in order to find high-quality minima. Throughout the paper we took the resulting minimum energy values and plotted their difference with the energies obtained from FCI.

Chapter 5

A Non-Orthogonal Variational Quantum Eigensolver

5.1 Preface

This chapter is taken from the previously published Ref. 298, originally appearing in New Journal of Physics. It was co-authored by the author of this thesis, Dr. Joonho Lee, Unpil Baek, Bryan O'Gorman, and Professor K. Birgitta Whaley. The project evolved from early discussions between Dr. Lee, the author of this thesis, and Professor Whaley, and matured through further discussions between all the co-authors. Unpil Baek contributed significantly to the collection and analysis of the numerical data. Bryan O'Gorman developed the circuit implementations of the ansatz used throughout this work and contributed significantly to the development of the improved Hadamard test. The majority of the writing was completed by the author of this thesis, with substantial input from all of the co-authors.

5.2 Introduction

Large, error-corrected quantum computers are expected to provide powerful new tools for understanding quantum many-body physics. For example, such devices will be able to efficiently simulate long-time dynamics [2], and through phase estimation, measure the energy of a trial wavefunction while projecting it into the eigenbasis of the Hamiltonian [76]. Prior to the availability of such devices, it is natural to ask how today's noisy, intermediate-scale quantum (NISQ) platforms may be used for similar ends. One appealing strategy, the variational quantum eigensolver (VQE) [17, 140], uses a potentially noisy quantum computer as a black box to prepare parametrized wavefunctions and measure their energy. By optimizing over the wavefunction parameters in a classical outer loop, one obtains a variational upper bound on the true ground state energy.

While it is believed that even a noisy, modestly-sized quantum computer can prepare and measure states that are out of reach for a classical computer [15], it will still likely be difficult

to take advantage of this fact to surpass the capabilities of classical variational methods [18, 146, 193, 217]. One serious challenge is that noise is particularly damaging for quantum chemical calculations that demand a high degree of precision [146, 193, 299]. Recent works have presented a variety of approaches to overcoming this difficulty, including combining error detection schemes with postselection [213, 214, 299], extrapolating to the zero-noise limit [208, 209, 300], and using additional measurements and post-processing to construct better energy estimators [150, 212, 213, 215]. A complementary body of research has focused on developing new variational ansätze that use fewer gates and thus offer less opportunity for errors to occur [170, 191, 224]. We shall present a new approach in this latter direction that allows for a systematic increase in wavefunction complexity without a growing circuit depth.

The standard VQE approach uses a quantum computer to measure the expectation value of the Hamiltonian for some parametrized wavefunction, $|\psi(\boldsymbol{\theta})\rangle$, in conjunction with a classical coprocessor that interprets the measurement outcomes and suggests new values for the $\boldsymbol{\theta}$ parameters in order to minimize the energy [140]. In our approach, we define instead a logical ansatz

$$\left|\psi(\mathbf{c}, \boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)})\right\rangle = \sum_{i=1}^{M} c_i \left|\phi_i(\boldsymbol{\theta}^{(i)})\right\rangle, \tag{5.1}$$

where each $\left|\phi_i(\boldsymbol{\theta}^{(i)})\right\rangle$ is an independently parametrized wavefunction with a compact quantum circuit description. For brevity, we shall sometimes omit the parameters and refer to these wavefunctions more compactly as $|\psi\rangle$ and $|\phi_i\rangle$. Rather than preparing the state $|\psi\rangle$ directly on our device and measuring its energy, we use our quantum computer to prepare simpler pairwise superpositions of the states $\{|\phi_i\rangle\}$. We then measure the matrix elements of the Hamiltonian and overlap matrices,

$$\begin{aligned} H_{ij} &= \langle\phi_i|\hat{H}|\phi_j\rangle, \\ S_{ij} &= \langle\phi_i|\phi_j\rangle. \end{aligned} \tag{5.2}$$

This allows us to classically solve a generalized eigenvalue problem,

$$H\mathbf{c} = ES\mathbf{c}, \tag{5.3}$$

thereby finding the optimal $\mathbf{c}$ parameters and minimizing the energy in the subspace spanned by the set of states $\{|\phi_i\rangle\}$. The $\boldsymbol{\theta}^{(i)}$ values that parametrize each basis function $\left|\phi_i(\boldsymbol{\theta}^{(i)})\right\rangle$ can then be optimized by a classical outer loop to lower the energy further, solving a new generalized eigenvalue problem at each step.

Our approach shares certain features with a variety of recent proposals for quantum algorithms that involve solving generalized eigenvalue problems [212, 221, 222, 301–303]. However, our approach also differs from these works in some key respects. Most importantly, we make no assumptions about the form of the component wavefunctions $|\phi_i\rangle$, other than that they have efficient quantum circuit implementations. In the context of quantum algorithms, prior work has assumed that these wavefunctions are generated by excitations from a fixed reference state [212], by real or imaginary time evolution [221, 222, 301, 302], or by the

simultaneous rotation of a set of orthogonal reference wavefunctions [303]. Two of these works in particular, Refs. 221 and 222, were released contemporaneously with our own and provide an interesting contrast to our approach. Specifically, they require the same off-diagonal matrix element measurements used in this work but construct the non-orthogonal basis function by real-time propagation of trial wavefunctions rather than the variational approach we take here.

In the context of classical simulations, multireference methods which make use of a superposition of configurations have a long and storied history [304–313]. Most directly similar to this work are those which demand each of the $|\phi_i\rangle$ wavefunctions to be a Slater determinant (not necessarily in the same single particle basis) [307]. This basic direction has been elaborated upon under a variety of names, including the non-orthogonal configuration interation (NOCI) method [308, 309, 312], the non-orthogonal Multicomponent Adaptive Greedy Iterative Compression (NOMAGIC) algorithm [311], and the non-orthogonal multi-Slater determinant (NOMSD) expansion approach [310, 313], among others. The restriction to Slater determinants allows for the efficient classical evaluation of the required matrix elements, while the relaxation of the requirement that the determinants be orthogonal to one another allows for more flexible and accurate wavefunctions when compared to orthogonal CI expansions with the same number of determinants.

The difference between these various approaches mainly lies in the way in which they obtain a set of non-orthogonal determinants. For example, NOCI separately optimizes individual determinants by finding a collection of different solutions to the Hartree-Fock equations before performing a single diagonalization of the Hamiltonian matrix [308, 309, 312]. Other approaches more closely parallel the one we take here, iteratively adding new states and variationally optimizing their parameters [307, 310, 313]. We do not exhaustively review the classical literature here, but note that the variational approach has been found to be prone to optimization challenges and that a number of the methods we cite arise out of attempts to ameliorate this difficulty [310, 311, 313].

By taking the basis functions $|\phi_i\rangle$ to be independently parametrized quantum circuits rather than single Slater determinants, we obtain an extremely flexible form for our logical ansatz, $|\psi\rangle = \sum c_i |\phi_i\rangle$. For a wide variety of ansatz circuits, we shall show that the required matrix element measurements between any $|\phi_i\rangle$ and $|\phi_j\rangle$ pair can be implemented efficiently using a number of quantum gates that is equal to the sum of the gates required to prepare $|\phi_i\rangle$ and $|\phi_j\rangle$, plus a small factor that scales linearly with the system size. Notably, the circuit size required is independent of the number of wavefunctions in the logical ansatz, making it possible to systematically add flexibility to $|\psi\rangle$ without increasing the required gate fidelity or coherence times of the quantum hardware.

This flexibility, however, comes at the cost of demanding more matrix element measurements. To ameliorate this cost we propose using a Monte Carlo technique to estimate the uncertainty in the ground state energy and to adaptively allocate our measurements of the matrix elements. Essentially, this scheme involves sampling from the distributions representing the uncertainty in the estimates of the Hamiltonian and overlap matrices, and solving a small generalized eigenvalue problem for each sampled matrix pair. We then characterize

the resulting distribution of ground state energy values by a sample variance. We suggest a heuristic that repeatedly determines which measurement to perform by calculating the sensitivity of this sample variance to additional measurements of each of the matrix elements.

We apply these ideas to two model chemical systems, a square configuration of H_4 and the π-system of hexatriene (C_6H_8), both of which exhibit mixed strong correlation and dynamical correlation effects. In terms of strong correlation, we shall focus on a pair of strongly entangled electrons. Specifically, entanglement between a pair of electrons can lead to two exactly degenerate determinants for certain geometries of these systems, with the rest of the electrons contributing to dynamical correlation. We present two types of numerical experiments. In the first, we explore how well the ground state of these systems can be represented by an NOVQE logical ansatz, varying both the complexity of the constituent basis functions and the size of the subspace. In the second, we take a fixed set of basis wavefunctions and compare our adaptive protocol for scheduling measurements with a simpler alternative.

5.3 Theory

5.3.1 Matrix Element Measurement

The off-diagonal matrix elements of the Hamiltonian, $H_{ij} = \langle\phi_i|\hat{H}|\phi_j\rangle$, do not correspond to physical observables and therefore cannot be measured directly in the usual manner. Nevertheless, it is possible to construct circuits that allow us to estimate them, for example, by using the Hadamard test [84]. In this section we present a simple strategy for measuring these matrix elements. We combine ideas from recent proposals for measuring off-diagonal matrix elements that appear in other contexts [156, 302] with a trick inspired by the literature on the impossibility of black box protocols for adding controls to arbitrary unitaries [314]. Our strategy offers several benefits over a naive application of the Hadamard test. Namely, it doesn't require implementing controlled versions of the ansatz preparation circuits, and it enables the simultaneous measurement of matrix elements of multiple commuting observables while also yielding information about the overlap matrix elements, $S_{ij} = \langle\phi_i|\phi_j\rangle$.

For simplicity, we will describe below the case where $\hat{H}$ is a sum of commuting operators that can easily be simultaneously measured. In the more general case, the usual Hamiltonian averaging approach of grouping the terms into multiple sets that are each simultaneously measurable and measuring the sets separately can be applied without modification [140, 193, 197, 315].

We begin by preparing the state

$$|+_{ij}\rangle := \frac{1}{\sqrt{2}}(|\phi_i\rangle\,|0\rangle + |\phi_j\rangle\,|1\rangle), \tag{5.4}$$

where the second register is an ancilla qubit. This task can be accomplished by using controlled versions of the unitaries $\hat{U}_i$ and $\hat{U}_j$ that prepare $|\phi_i\rangle$ and $|\phi_j\rangle$ from a fixed reference

state. Given some quantum circuit that implements the unitaries $\hat{U}_i$ and $\hat{U}_j$, it is possible to construct circuits that implement the controlled version of $\hat{U}_i$ and $\hat{U}_j$, by replacing each gate in the original circuits with its controlled form. Even setting aside the difficulty of compiling such a circuit on a physical device with limited connectivity, the cost of implementing such a circuit on a near-term device (quantified by counting the number of two-qubit gates) will be substantially increased. For example, it is known that the decomposition of the Toffoli gate (the controlled-controlled-NOT gate) into a collection of single qubit and CNOT gates requires the use of six CNOT gates [316]. Given the limited coherence times and two-qubit gate fidelities of near-term hardware, we must ask if there are alternatives for implementing controlled versions of $\hat{U}_i$ and $\hat{U}_j$.

An ideal protocol might allow us to implement a controlled version of an arbitrary $\hat{U}$ using a single execution of the original, unmodified circuit that implements $\hat{U}$. Unfortunately, a single use of oracle (blackbox) access to a general $\hat{U}$ is insufficient for implementing a controlled version of $\hat{U}$ in the quantum circuit model [314]. However, if $\hat{U}_i$ and $\hat{U}_j$ preserve fermionic (or bosonic) excitation number and act trivially on the vacuum state, then we can circumvent this no-go result. We now show how this can be accomplished in the construction of a controlled unitary operator,

$$\hat{U}_i, \hat{U}_j \rightarrow \hat{U}_i \otimes |0\rangle\langle 0| + \hat{U}_j \otimes |1\rangle\langle 1| \,. \tag{5.5}$$

We begin with a generic input state $|\psi_0\rangle\,|0\rangle + |\psi_1\rangle\,|1\rangle$, subject to the restriction that $|\psi_0\rangle$ and $|\psi_1\rangle$ are both states that are orthogonal to the state with zero particles, $|\text{vac}\rangle$.

1. First, we adjoin an ancilla system register in the vacuum state to obtain

   $|\psi_0\rangle \otimes |\text{vac}\rangle \otimes |0\rangle + |\psi_1\rangle \otimes |\text{vac}\rangle \otimes |1\rangle$.

2. Treating the final qubit as the control, we apply a controlled-SWAP operation between the two system registers, resulting in

   $|\psi_0\rangle \otimes |\text{vac}\rangle \otimes |0\rangle + |\text{vac}\rangle \otimes |\psi_1\rangle \otimes |1\rangle$.

3. Next, we execute the unmodified circuit for $\hat{U}_i$ on the first system register, while doing the same with $\hat{U}_j$ on the second system register, yielding

   $\hat{U}_i\,|\psi_0\rangle \otimes |\text{vac}\rangle \otimes |0\rangle + |\text{vac}\rangle \otimes \hat{U}_j\,|\psi_1\rangle \otimes |1\rangle$.

4. We follow this with a second controlled-SWAP operation to produce the state,

   $\hat{U}_i\,|\psi_0\rangle \otimes |\text{vac}\rangle \otimes |0\rangle + \hat{U}_j\,|\psi_1\rangle \otimes |\text{vac}\rangle \otimes |1\rangle$.

5. Finally, we discard the now unentangled second system register to show completion of the action of the controlled unitary gate and obtain the desired result,

   $\hat{U}_i\,|\psi_0\rangle \otimes |0\rangle + \hat{U}_j\,|\psi_1\rangle \otimes |1\rangle$.

For our purposes, we can take $|\psi_0\rangle$ and $|\psi_1\rangle$ to be the same fixed reference state, usually a Hartree-Fock state $|\psi_{\text{HF}}\rangle$. Then $|\phi_i\rangle = \hat{U}_i |\psi_{\text{HF}}\rangle$ and $|\phi_j\rangle = \hat{U}_j |\psi_{\text{HF}}\rangle$ and we see that with the last step we have successfully prepared the desired state, $|+_{ij}\rangle := \frac{1}{\sqrt{2}}(|\phi_i\rangle |0\rangle + |\phi_j\rangle |1\rangle)$. We then apply a Hadamard gate on the ancilla qubit and perform a $\hat{Z}$ measurement. It is easy to see that the expectation value of $\hat{Z}$ for the ancilla qubit will be $\langle \hat{Z}_{\text{anc}} \rangle = \text{Re}\, \langle \phi_i | \phi_j \rangle$. Furthermore, the post-measurement state of the system register is either

$$\frac{|\phi_i\rangle + |\phi_j\rangle}{\sqrt{2 + 2\text{Re}\, \langle \phi_i | \phi_j \rangle}}, \tag{5.6}$$

if the ancilla qubit was found to be in the $+1$ eigenstate, or

$$\frac{|\phi_i\rangle - |\phi_j\rangle}{\sqrt{2 - 2\text{Re}\, \langle \phi_i | \phi_j \rangle}}, \tag{5.7}$$

if the measurement outcome was -1. These outcomes occur with probabilities $\frac{1+\text{Re}\langle \phi_i | \phi_j \rangle}{2}$ and $\frac{1-\text{Re}\langle \phi_i | \phi_j \rangle}{2}$ respectively.

In both cases, we proceed to measure the Hamiltonian $\hat{H}$ on the system register. Depending on the result of the ancilla qubit measurement, the resulting expectation values will be either

$$\langle \hat{H} \rangle^{(+1)} = \frac{\langle \hat{H} \rangle_i + \langle \hat{H} \rangle_j + 2\text{Re}\, \langle \phi_i | \hat{H} | \phi_j \rangle}{2 + 2\text{Re}\, \langle \phi_i | \phi_j \rangle}, \tag{5.8}$$

or

$$\langle \hat{H} \rangle^{(-1)} = \frac{\langle \hat{H} \rangle_i + \langle \hat{H} \rangle_j - 2\text{Re}\, \langle \phi_i | \hat{H} | \phi_j \rangle}{2 - 2\text{Re}\, \langle \phi_i | \phi_j \rangle}. \tag{5.9}$$

Now we consider the expectation value of the operator $\hat{H}\hat{Z}_{\text{anc}}$. By multiplying each of the conditional expectation values of $\hat{H}$ by the corresponding eigenvalue of $\hat{Z}_{\text{anc}}$ and taking the appropriate weighted average, we find that

$$\langle \hat{H}\hat{Z}_{\text{anc}} \rangle = \text{Re}\, \langle \phi_i | \hat{H} | \phi_j \rangle \,. \tag{5.10}$$

Furthermore, if $\hat{H}$ is a sum of Pauli operators, then the usual Hamiltonian averaging approach and upper bounds on the variance of a VQE observable apply to Eq. 5.10 [197]. Therefore, by repeated measurement we can estimate $\text{Re}\, \langle \phi_i | \hat{H} | \phi_j \rangle$ to a fixed precision ϵ using approximately the same number of measurements that we would need to measure a diagonal matrix element to the same accuracy. A similar approach allows us to estimate $\text{Im}\, \langle \phi_i | \phi_j \rangle$ and $\text{Im}\, \langle \phi_i | \hat{H} | \phi_j \rangle$ by starting with the state $\frac{1}{\sqrt{2}}(|0\rangle |\phi_i\rangle + i\, |1\rangle |\phi_j\rangle)$.

Consider an ansatz $|\theta\rangle = U(\theta) |\psi_0\rangle$ for N spin-orbitals, represented by N qubits after a Jordan-Wigner transformation, such that the size and depth of the circuit for U is independent of θ; this is typical of VQE ansätze, but the following can be easily generalized when it is not the case. Suppose also that we have a protocol for measuring the Hamiltonian H on

the N-qubit register. What are the additional resources required to implement NOVQE? First, we require $2N$ qubits and at least one ancilla. The variational unitaries U_i and U_j can be applied in parallel, doubling the size of the circuit but not the depth. The measurement protocol for H can be applied without modification to the first register. For the two controlled swaps, there is a space-time tradeoff. First, consider the case without geometric constraints. Each controlled swap of the registers can be implemented using the single ancilla and N 3-qubit CSWAP gates in series on pairs of the corresponding qubits from the two registers, adding $2N\tau_{\text{CSWAP}}$ to the depth, where τ_{CSWAP} is the effective depth of the CSWAP gate. Alternatively, we can use N ancillas and in $\lceil \log_2 N \rceil$ depth produce a logical ancilla in the form of a GHZ state. Then the N CSWAPS can be done in parallel, adding only $2\tau_{\text{CSWAP}}$ to the depth.

Suppose now that we are restricted, e.g., to some subgraph of a 2D square grid, and that $U(\theta)$ can be implemented only using gates on linearly adjacent qubits. Then we can place the computational registers on adjacent rows and the ancilla at the end of one. Now, in addition to the CSWAP gates, we must use N 2-qubit SWAP gates to move the ancilla through the line, so that the contribution to the depth is now $2N(\tau_{\text{CSWAP}} + \tau_{\text{SWAP}})$. Alternatively, we can use a whole row of ancillas between the two computational rows, and in $\lceil N/2 \rceil \tau_{\text{CNOT}}$ prepare the ancilla GHZ state as we did without geometric constraints, and again the CSWAP gates can be done in parallel.[1]

5.3.2 Diagonalization With Uncertainty

Given a collection of states $\{|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_n\rangle\}$, we are interested in determining the minimum energy state in the subspace that they span. To do this, we use our protocol described above to measure the matrix elements of the Hamiltonian and overlap matrices (Eq. 5.2), and solve the generalized eigenvalue problem (Eq. 5.3). However, because we perform only a finite number of measurements of each of matrix element, we have some level of statistical uncertainty. In this section, we shall lay out a simple Monte Carlo strategy to estimate the resulting uncertainty in the minimum eigenvalue of Eq. 5.3. We shall aim to provide a self-contained presentation for convenience, but we note that this approach is related to a long tradition of applying Monte Carlo methods to statistical problems, including the diagonalization of noisy matrices [317–319].

We model the experimentally determined values of each matrix element using a normal distribution. In practice, the experimental measurements of the matrix elements are individually described by draws from Bernoulli random variables, but variational quantum algorithms typically work in the regime where the average of such measurements are well-approximated by a normal distribution [140]. In the context of an actual experiment, one could approximately determine the parameters of these distributions from the experimental measurement record of the Hamiltonian and overlap matrix elements.

[1] Note that technically we should distinguish between different values for τ_{CSWAP} depending on geometric constraints on the 2-qubit gates into which the CSWAP is decomposed, e.g., between when the control qubit is in the middle of the three on a line and when it is at one of the ends.

For the purposes of the numerical experiments in this work, we determine the variance of the Hamiltonian matrix element measurements using the upper bounds described in Refs. 140 and 197. Similarly, we observe that our scheme for measuring the overlap matrix elements will have a variance that is at most $\frac{1}{m}$, where m is the number of measurements performed, and we use this upper bound as an approximation to the true variance. We use these approximations both in our simulation of the experimental measurement record and in our subsequent protocol to determine the uncertainty in the ground state energy. Throughout this section, we use a notation which separates the intrinsic component of the variance, which we denote by σ^2, from the scaling with the number of measurements, m.

Experimentally, we only have access to estimates of $\langle\phi_i|\hat{H}|\phi_j\rangle$ and $\langle\phi_i|\phi_j\rangle$ from our measurement record, which we denote by $\tilde{h}_{ij}$ and $\tilde{s}_{ij}$. Taken together with our estimates of the variances, $\tilde{\sigma}^2_{H_{ij}}$ and $\tilde{\sigma}^2_{S_{ij}}$, we can define the random variables

$$\tilde{H}'_{ij} = \tilde{h}_{ij} + \frac{\tilde{\sigma}_{H_{ij}}}{\sqrt{m_{H_{ij}}}}\mathcal{N}(0,1), \tag{5.11}$$

$$\tilde{S}'_{ij} = \tilde{s}_{ij} + \frac{\tilde{\sigma}_{S_{ij}}}{\sqrt{m_{S_{ij}}}}\mathcal{N}(0,1). \tag{5.12}$$

These distributions represent our uncertainty about the true value of the matrix elements given the limited information provided by our experimental data.

To quantify the corresponding uncertainty in the ground state energy in the NOVQE subspace, we use a Monte Carlo sampling procedure. We accomplish this by repeatedly drawing from the distributions $\tilde{H}'_{ij}$ and $\tilde{S}'_{ij}$, and solving the resulting generalized eigenvalue problems. However, it is possible that the noise in our matrix element measurements and subsequent sampling destroys the positive semi-definite character of the overlap matrix. To deal with this, we follow the canonical orthogonalization procedure described in Ref. 24, discarding the eigenvalues of the sampled overlap matrices that are less than some numerical cutoff (and their associated eigenvectors). Each sampled pair of matrices yields a sample from the unknown distribution over possible NOVQE ground state energies. We then quantify our uncertainty in our estimate of this lowest eigenvalue by calculating the sample variance of this distribution of possible energies, σ^2_{MC}.

It is important to note that this distribution is not Gaussian and that its mean is not an unbiased estimate of the ground state energy in the NOVQE subspace [317]. This is true for a number of reasons, but it can be seen, for example, by considering the fact that the usual second order correction to the energy is quadratic in the off-diagonal matrix elements. Therefore, even unbiased and normally distributed noise in the matrix elements leads to a bias in the estimated eigenvalues. Furthermore, the rate at which our estimate of the mean and variance of the distribution over possible NOVQE ground state energies converges (with respect to the number of Monte Carlo samples) will vary according to the underlying distribution. The most meaningful consequences of this for our purposes are that convergence with respect to the number of Monte Carlo samples should be checked before being relied upon and that one should be cautious in using the standard error to generate error bars. As

the number of measurements made increases and the amount of uncertainty diminishes, these effects are naturally suppressed.

5.3.3 Experiment Design Heuristic

In the previous section, we proposed a Monte Carlo scheme for estimating the uncertainty in the NOVQE ground state energy caused by a finite number of measurements of the individual matrix elements. By repeatedly sampling from $\tilde{H}'_{ij}$ and $\tilde{S}'_{ij}$ and solving the resulting generalized eigenvalue problems, we obtained a distribution over NOVQE ground state energies with some mean μ_{MC} and standard deviation σ_{MC}. Here we build on this proposal to determine the relative impact of performing additional measurements. Ultimately, our goal is to create a reasonable heuristic for adaptively scheduling measurements to most efficiently use a limited amount of device time.

We determine the impact of additional measurements of the matrix elements on the uncertainty in the ground state energy by calculating the derivatives of the sample standard deviation, σ_{MC}, with respect to the number of measurements performed, $m_{H_{ij}}$ and $m_{S_{ij}}$. The resultant quantities, $\frac{d\sigma_{\text{MC}}}{dm_{H_{ij}}}$ and $\frac{d\sigma_{\text{MC}}}{dm_{S_{ij}}}$, estimate how much we expect the sample deviation to shrink if we perform additional measurements of H_{ij} or S_{ij}. Note that we take these derivatives only with respect to $m_{H_{ij}}$ and $m_{S_{ij}}$ in the Monte Carlo sampling procedure of Eq. 5.11 and Eq. 5.12, not in the original measurements on the device. Therefore, no additional quantum resources are required. We use the TensorFlow software package to perform the Monte Carlo sampling of $\tilde{H}'_{ij}$ and $\tilde{S}'_{ij}$, to calculate of the ground state energies, and to estimate σ_{MC} [264]. This enables us to evaluate the analytical expressions for each of $\frac{d\sigma_{\text{MC}}}{dm_{H_{ij}}}$ and $\frac{d\sigma_{\text{MC}}}{dm_{S_{ij}}}$ (for a fixed set of samples drawn from $\tilde{H}'_{ij}$ and $\tilde{S}'_{ij}$) without explicitly deriving the equations.

To optimally allocate our experimental measurements, we begin by performing a small number of measurements of each matrix element. We then estimate the derivatives $\frac{d\sigma_{\text{MC}}}{dm_{H_{ij}}}$ and $\frac{d\sigma_{\text{MC}}}{dm_{S_{ij}}}$. Using these estimates, we simply choose to perform additional measurements on the matrix element whose corresponding derivative is the most negative. In practice, we perform these measurements in small batches so that the time taken by the classical processing of the measurement results is small compared to the time performing the measurements. By repeating this process for many steps, until we either achieve the desired accuracy or exhaust a pre-defined measurement budget, we aim to approximately optimize allocation of measurements between the different terms.

5.3.4 Implementation

The tools presented above are applicable for use with a variety of different ansätze, and are subject only to the constraint that the circuits act on a common reference state and conserve fermionic excitation number in order to benefit from the efficient implementation of the matrix element measurements. For our numerical experiments, we shall focus on a particular

class of wavefunctions known as ***k***-fold products of **u**nitary **p**aired **c**oupled **c**luster with **g**eneralized **s**ingle and **d**ouble excitations [224] (k-UpCCGSD). These wavefunctions have the appealing properties that 1) the required circuit depth scales only linearly in the size of the system, and 2) they can be systematically improved by increasing the refinement parameter k. We briefly review this ansatz below and then describe in more detail the implementation details of our numerical experiments.

5.3.4.1 The k-UpCCGSD Ansatz

The essential idea behind the k-UpCCGSD ansatz is to act on a reference state, Hartree-Fock in the case of this paper, with a product of k elementary blocks. Each block is an independently parametrized approximation to a unitary coupled cluster circuit generated by a sparse cluster operator containing only single and paired double excitations [50, 320]. To this end, the wavefunction (before the Trotter approximation involved in compiling the circuits) is defined as follows.

$$|\psi\rangle = \prod_{x=1}^{k} \left(e^{\hat{T}^{(x)} - \hat{T}^{(x)\dagger}} \right) |\phi_0\rangle, \tag{5.13}$$

where each cluster operator

$$\hat{T} = \sum_{ia} t_{ii}^{aa} \hat{a}_{a\alpha}^{\dagger} \hat{a}_{a\beta}^{\dagger} \hat{a}_{i\beta} \hat{a}_{i\alpha} + t_i^a (\hat{a}_{a\alpha}^{\dagger} \hat{a}_{i\alpha} + \hat{a}_{a\beta}^{\dagger} \hat{a}_{i\beta}). \tag{5.14}$$

possesses an independent collection of variational parameters. (We omit the (x) superscript for simplicity and use Latin and Greek letters for spatial and spin indices respectively.)

In contrast with the standard unitary coupled cluster single and doubles (UCCSD), k-UpCCGSD only includes doubles excitations which collectively move a pair of electrons from one spatial orbital to another. The resulting loss of flexibility is ameliorated by the use of generalized excitations that do not distinguish between occupied and unoccupied orbitals [183, 240], and by the k-fold repetition of the elementary circuit block. As a result, the number of free parameters in the ansatz scales as $O(kN^2)$. We make use of the generalized swap networks of Ref. 172 to implement a single Trotter step approximation to the k-UpCCGSD ansatz with the open source Cirq and OpenFermion-Cirq libraries [263, 321].

The circuits consist of the following elementary gates:

- $\mathrm{FSIM}_2(w_0, w_1) = \exp(iH)$ for
  $H = (w_0 \,|10\rangle \langle 01| + \mathrm{h.c.}) + w_1 \,|11\rangle \langle 11|$,
- $\mathrm{FSWAP} = \mathrm{SWAP} \cdot \mathrm{CZ}$, and
- $\mathrm{FSIM}_4(w) = \exp(iH)$ for
  $H = w \,|0011\rangle \langle 1100| + \mathrm{h.c.}$

Because each FSWAP immediately follows an $FSIM_2$, we can compile them together to get an effective duration τ_2. Let τ_4 be the effective duration of $FSIM_4$. The overall depth then is $kN(\tau_2 + \tau_4/2)$. There are $\binom{N}{2}$ pairs of $FSIM_2$ and FSWAP gates, and $\binom{N/2}{2}$ $FSIM_4$ gates. This is simply an upper bound; the depth may be compressed further by combining the compilation of each $FSIM_4$ with the immediately following 2-qubit gates.

5.3.4.2 Computational Details

The quantum chemical calculations of the full configuration interaction (FCI) ground states and Hartree-Fock (HF) reference wavefunctions were performed using the open source packages Psi4 and OpenFermion [262, 263]. We optimized the ground state energy in the NOVQE subspace by varying the parameters of the most recently added ansatz wavefunction, diagonalizing the Hamiltonian and overlap matrices at each step. Inspired by recent proposals for adaptive ansatz construction [191, 192, 322], each k-UpCCGSD wavefunction was grown iteratively by adding a single UpCCGSD block at a time, as described in more detail below. We performed this optimization using the Scipy implementation of the quasi-Newton limited-memory BFGS (L-BFGS-B) method [265, 323], treating the ground state energy in the NOVQE subspace as the objective function. We calculated the gradient at each step using a finite difference method with a step size of $\delta = 10^{-6}$. Each circuit was optimized using up to 2000 gradient evaluations, terminating early if the magnitude of all components of the gradient fell below 10^{-5}.

In order to escape local minima, we repeatedly applied random kicks to the variational parameters. After each 500 gradient evaluations we compared the current value of the objective function to the best observed value and reset the parameters if appropriate. Subsequently, we added random values drawn from the a normal distribution with zero mean and variance $\sigma^2 = 1$ (after 500 steps), $\sigma^2 = 10^{-1}$ (after 1000 steps), or $\sigma^2 = 10^{-2}$ (after 1500 steps). The best observed value of the energy across this whole procedure is the one we ultimately report. We randomly initialized the parameters of the $k = 1$ circuits by drawing from a normal distribution with mean 0 and variance $\sigma^2 = 10^{-6}$. Parameters for circuits with higher values of k were initialized by taking the parameters from an optimized circuit with $k-1$ UpCCGSD blocks and appending a new block with random variational parameters drawn from the same distribution, $\mathcal{N}(0, 10^{-6})$.

5.4 Results

H_4 is often used as a small testbed for single-reference coupled-cluster methods [267–271]. We shall focus on the square (D_{4h}) geometry here. The system exhibits two exactly degenerate determinants at the D_{4h} geometry, leading to a mix of strong and weak correlation effects. Another important class of chemical systems to investigate is hydrocarbons. In this work, we shall study a simple hydrocarbon, hexatriene (C_6H_8). The interesting aspect of this molecule is that the torsional PES of a double bond leads to a strong correlation problem. At $\theta = 90°$,

it exhibits two exactly degenerate determinants and therefore it is strongly correlated. To form the active space, we include the entire set of π electrons in the system along with both Π and Π^* orbitals. The resulting active space is then (6e, 6o), and this also possesses a good mixture of weak and strong correlation.

In the following subsections, we present the results of two types of experiments related to our proposed NOVQE approach on these chemical systems and discuss the potential utility of our approach for more general chemical problems. With the first class of experiments, we focus on understanding how effectively the ground state can be represented by a linear combination of parametrized wavefunctions, optimized using the gradient-based approach we described above. We vary both the complexity of the individual ansatz wavefunctions by adjusting the number of circuit blocks (k) in the k-UpCCGSD ansatz and the number of states (M) in the NOVQE subspace. For these calculations, we neglect the challenges posed by a finite number of measurements and the impact of circuit noise. In our second set of numerical experiments, we explore the extent to which our proposal for an adaptive measurement scheme is successful in reducing the number of circuit repetitions required to resolve the NOVQE ground state energy to a fixed precision. In both cases we compare the NOVQE energies with the numerically exact FCI energies.

5.4.1 NOVQE Ground State Energies

5.4.1.1 A Hydrogen Complex, H_4

Figure 5.1 presents data on the application of NOVQE to the square geometry of H_4 with fixed bond distance $R_{\text{H-H}} = 1.23$ Å in a minimal STO-3G basis set, an $N = 8$ qubit problem. We consider the performance of the k-UpCCGSD ansatz for $k = 1$ and $k = 2$ with $M = 1$ up to $M = 6$ states in the NOVQE subspace, noting that $M = 1$ is equivalent to the regular VQE procedure. For each value of k and M we perform five independent calculations and consider the error in the median ground state energy found by the optimization procedure as a proxy for the ansatz's ability to describe the ground state. Note that Figure 5.1 switches from a logarithmic scale to a linear one below 10^{-5}, in order to include points that are zero to numerical precision and to reflect the fact that our numerical optimization may behave inconsistently below this threshold due to its convergence threshold.

Focusing first on understanding the behavior of the wavefunctions in the context of the standard VQE approach ($M = 1$), we note that for $k \geq 2$ the k-UpCCGSD ansatz is essentially exact for this problem. Looking more closely at the data for $k = 2$, $M = 1$ in Figure 5.1, one can see that two of the five calculations failed to find the global optimum (the pale orange points). In general, we found that the optimization of this ansatz was challenging. We expect these challenges to become more severe with increasing system size, and when the stochastic nature of the quantum measurements are taken into account.

In the case of $k = 1$ we observe that we can systematically improve the accuracy of the estimated ground state energy by increasing the number of states included in the NOVQE subspace (M). Given $M = 3$ independent copies, even this relatively simple ansatz is able

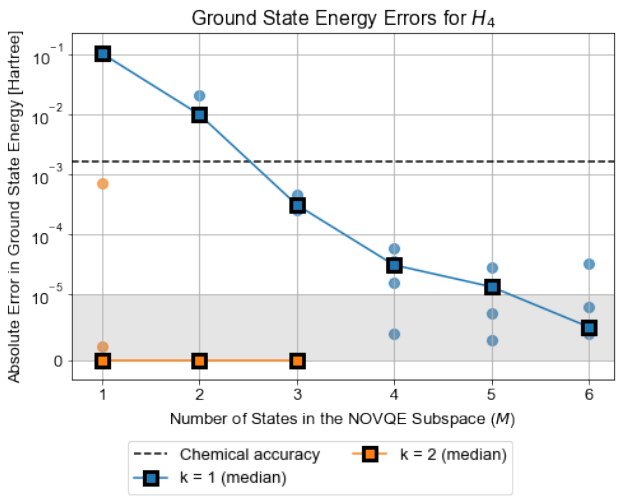

Figure 5.1: Difference between NOVQE energies and FCI energies for the ground state of H_4 for a variety of k-UpCCGSD ansätze and sizes of the NOVQE subspace (M). The NOVQE energy is optimized by varying the parameters of the most recently added state to minimize the ground state energy in the subspace. For each value of M and k we plot five independent calculations as separate points and show the median values as squares connected by lines. The scale of the plot switches from logarithmic to linear below 10^{-5} in order to include points which are zero to numerical precision and to reflect the fact that our numerical optimization may behave inconsistently below this threshold due to its convergence threshold. The dotted horizontal line indicates 1 kcal/mol $\approx$ 1.59 millihartree, a commonly accepted value for "chemical accuracy". As more states are added to the NOVQE subspace, the error in the ground state energy declines substantially for the $k = 1$ version of k-UpCCGSD. For larger values of k, a single state (equivalent to a regular VQE procedure) is sufficient to capture the ground state to a high precision.

to represent that ground state almost exactly. This supports our thesis that a collection of ansatz states which are individually not capable of targetting a desired state may be fruitfully combined to yield a sufficiently powerful logical ansatz. However, the measurements of the off-diagonal matrix elements for NOVQE require slightly more than twice the gate count necessary for the measurements of individual ansatz states in the regular VQE formalism. For this particular system, it may therefore be more effective to use a single $k = 2$ ansatz than multiple $k = 1$ circuits.

5.4.1.2 Hexatriene

Here we present our results for the ground state energy of two molecular configurations of Hexatriene (C_6H_8) in an STO-3G basis with an active space of 6 electrons in 6 π orbitals

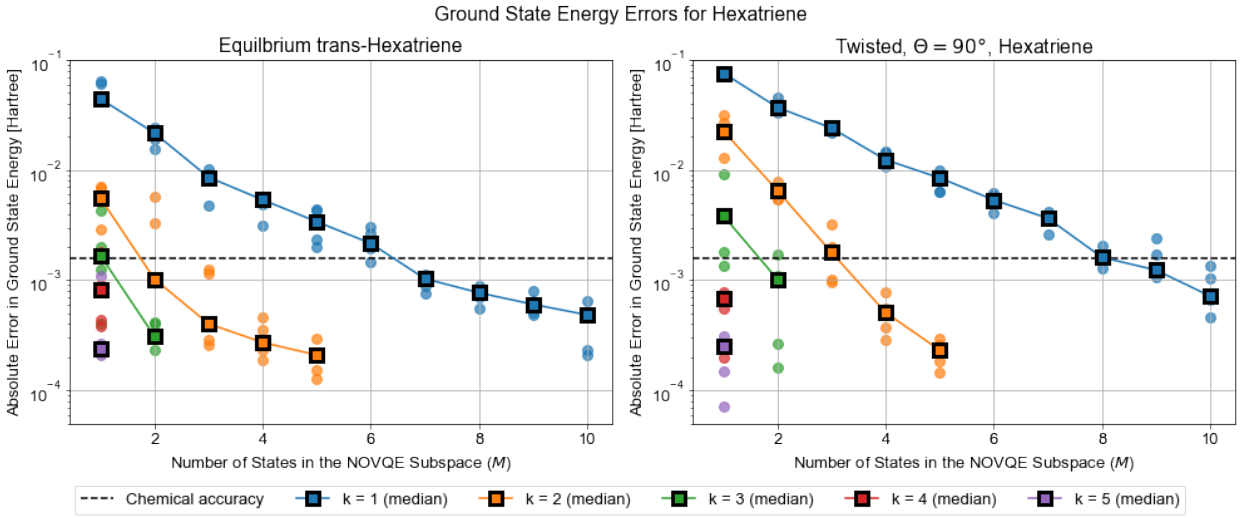

Figure 5.2: Difference between NOVQE energies and FCI energies for the ground states of the equilibrium configuration of trans-Hexatriene and a 90° twisted configuration for a variety of k-UpCCGSD ansätze and sizes of the NOVQE subspace (M). The NOVQE energy is optimized by varying the parameters of the most recently added state to minimize the ground state energy in the subspace. For each value of M and k we plot five independent calculations as separate points and show the median values as squares connected by lines. The dotted horizontal line indicates 1 kcal/mol $\approx$ 1.59 millihartree, a commonly accepted value for "chemical accuracy". The flexibility of the NOVQE wavefunction may be increased both by adding more states to the NOVQE subspace (M), or more parametrized blocks to each individual circuit (k). In either case, the error is driven below the threshold for chemical accuracy.

($N = 12$ qubits). Here, due to the system's increased complexity, we consider circuits with up to $k = 5$ UpCCGSD blocks and subspace sizes as large as $M = 10$. In Figure 5.2 we show the calculations for an equilibrium geometry (the trans isomer, obtained by performing geometry optimization using density functional theory) and a configuration with a 90° twist on the central Carbon-Carbon double bond respectively. We provide the geometries for these two configurations in Appendix 5.7, Table 5.1 and Table 5.2.

Once again we notice that increasing the circuit complexity by taking larger values of k provides a substantial benefit, driving the estimated ground state energy well below the threshold for chemical accuracy without resorting to the multiple states of the NOVQE formalism. Likewise, as the number of NOVQE states increases, the NOVQE ground state energy reaches chemical accuracy even with the most limited ansatz. For Hexatriene we see that multiple $k = 1$ states are able to achieve a performance on par with a single $k = 4$ state. The NOVQE procedure for the $k = 1$ states requires almost a factor of four less circuit depth

and half as many quantum gates as performing a single VQE calculation with the $k = 4$ state.

Interestingly, for the $k = 1$ case in both configurations, and the $k = 2$ case in the twisted configuration, Figure 5.2 shows regimes where the error in the ground state energy decreases exponentially as a function of M. We compare this with the observation in classical non-orthogonal electronic structure calculations, where a small number of determinants are often sufficient to capture most of the wavefunction, but a long tail of dynamic correlation can result in a slow convergence to the true ground state as determinants are added to the variational space [307–313, 324]. The classical intractability of calculating matrix elements between different coupled cluster wavefunctions means that relatively little work has been done on the representational power of wavefunctions like those used in NOVQE. This is in contrast with another class of quantum non-orthogonal methods which, by virtue of building their basis states by time-evolving a set of reference wavefunctions, demonstrate exponential convergence by construction [221, 222]. In the future, it would be interesting to determine whether the rapid convergence with respect to M we observe here gives way to a regime of slow convergence, like that observed in classical non-orthogonal methods, when more challenging systems are treated.

5.4.2 NOVQE Matrix Element Measurements

In the previous subsection we presented data on the performance of NOVQE in the absence of noise during the circuit execution and measurement process. Now we consider the effects of statistical noise during measurement. Specifically, we determine how many circuit repetitions are necessary to evaluate the ground state energy within a target precision, for a subspace defined by a fixed set of NOVQE states. For simplicity, we do not combine this analysis with an investigation of the optimization procedure. Instead, we take the optimized circuit parameters for a collection of M NOVQE states and compare the effectiveness of the adaptive protocol we described in Section 5.3.3 to a simpler alternative for determining the ground state energy in the subspace spanned by the optimized states, which we shall explain below.

The simpler protocol, which we shall refer to as non-adaptive, consists of measuring each matrix element of the Hamiltonian and overlap matrices the same number of times. For the adaptive protocol, we repeatedly use the procedure described in Section 5.3.3 to select a particular matrix element and perform measurements in batches of $\approx 10^5$ circuit repetitions. For the purpose of this comparison, we treat a 'measurement' of a particular Hamiltonian or overlap matrix element as a draw from a Gaussian random variable whose mean is the true value of the matrix element and whose variance is set by the upper bound described in Ref. 140, scaled by the number of measurements performed. Note that in a real experiment, or a finer-grained simulation, the Hamiltonian has to be decomposed into groups of terms that can be simultaneously measured. One could apply an adaptive scheme like the one we propose to schedule measurements between these groups as well. For both kinds of numerical experiments we calculate a 2σ error bar using a bootstrapping sample size of 200 with the techniques of Section 5.3.2.

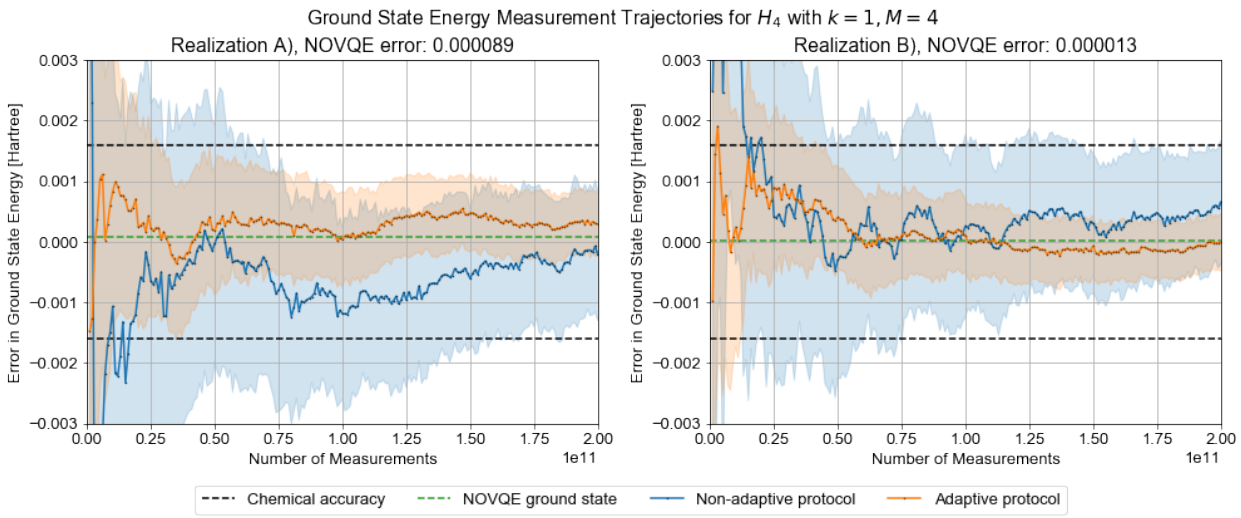

Figure 5.3: Comparison of the ability of the adaptive and non-adaptive schemes for scheduling measurements to resolve the ground state energy of H_4 in two different NOVQE subspaces of $M = 4$ optimized $k = 1$ k-UpCCGSD states. The evolution of the estimated ground state energies is plotted in solid lines together with 2σ error bars indicated by the shaded regions. The actual energies of the ground states in the NOVQE subspaces are indicated with dashed green lines. Panels A and B show two different typical realizations of the measurement record as the total number of measurements increases. In both cases, the adaptive protocol converges significantly more quickly than the non-adaptive one. Note that the variance of the experimental measurements are approximated using upper bounds and that the true numbers required for both the adaptive and non-adaptive schemes are likely to be lower [140, 299].

5.4.2.1 A Hydrogen Complex, H_4

In Figure 5.3 we plot the actual trajectories of the estimates for the ground state energies, together with their error bars for both the adaptive and non-adaptive approaches to measurement. We show two realizations of this numerical experiment applied to an NOVQE simulation of H_4 with $M = 4$ of the 1-UpCCGSD states. In both cases, we see that the adaptive protocol converges more quickly towards the NOVQE ground state energy than the non-adaptive one. We find that the data qualitatively supports the assumption that the variance in the ground state energy estimate settles into an asymptotic regime where its behavior is well described by the relationship

$$\sigma^2(N_m) \approx \frac{\kappa}{N_m}, \tag{5.15}$$

where N_m indicates the total number of measurements performed and κ is a constant, which we shall refer to as the 'intrinsic variance'. For these particular realizations, we find κ to be

approximately $5.3 \cdot 10^4$ E_h^2 and $5.5 \cdot 10^4$ E_h^2 for the non-adaptive scheme in panels A) and B), and approximately $1.4 \cdot 10^4$ E_h^2 and $9.7 \cdot 10^3$ E_h^2 for the adaptive scheme. Using the same upper bounds to calculate the variance for a regular VQE calculation performed on the same system would yield $\kappa \approx 28$ E_h^2.

For some fixed precision $\sigma^2(N)$, the relative increase (or decrease) in the number of measurements between the two approaches is can be determined by setting the ratio κ/N_m to be the same, i.e.,

$$N_m^1/N_m^2 \approx \kappa^1/\kappa^2$$

, where 1 refers to non-adaptive and 2 to adaptive. Therefore, for these applications to H_4, our scheme for iterative measurement achieves a modest reduction in variance. When targeting a fixed accuracy this would translate into a few-fold (≈ 3.7 for realization A and ≈ 5.7 for realization B) savings in the total number of measurements. Carrying out the same comparison between our adaptive measurement scheme for NOVQE and ordinary VQE, we see that, unfortunately, the overall measurement cost is still two orders of magnitude ($\sim 300-500$) larger than that required for energy measurement in an ordinary VQE approach. In order for NOVQE, or other forms of quantum non-orthogonal methods to be made practically useful, this increased measurement time will have to be accounted for and minimized.

5.4.2.2 Hexatriene

As in our analysis of H_4, we compare the proposed adaptive approach to distributing measurements between the elements of the Hamiltonian and overlap matrices with a non-adaptive one. We do so by choosing collections of optimized NOVQE states and applying both methods to determine the ground state energy in the resulting subspaces. In this case we choose to use $M = 8$ states, each of which is generated by a 1-UpCCGSD circuit, and focus on the equilibrium configuration of trans-Hexatriene. Examining the two realizations of this experiment plotted in Figure 5.4, we see immediately that the increased difficulty of this problem compared to H_4 is reflected in the much larger gaps between the FCI ground states and the ground states in the NOVQE subspaces, as well in the larger numbers of measurements required for convergence.

Figure 5.4 shows the same substantial difference between the performances of the adaptive and non-adaptive approaches that was observed for H_4. In panel B, we see that the true ground state of the subspace lies outside of the error bars for the non-adaptive scheme during small portions of the measurement procedure. This is a manifestation of the phenomenon mentioned in Section 5.3.2, where using an insufficient number of Monte Carlo samples may result in misestimating the magnitude of the uncertainty in the ground state energy. We note that the adaptive scheme moves quickly to a regime where the uncertainty estimates are reliable even with a small number of samples. We once again observe that the variance qualitatively converges with the expected long-time $\frac{1}{N}$ behavior of Eq. 5.15 for most of the numerical experiment. Therefore, we can determine κ, defined in Eq. 5.15, for each method and compare their statistical efficiencies.

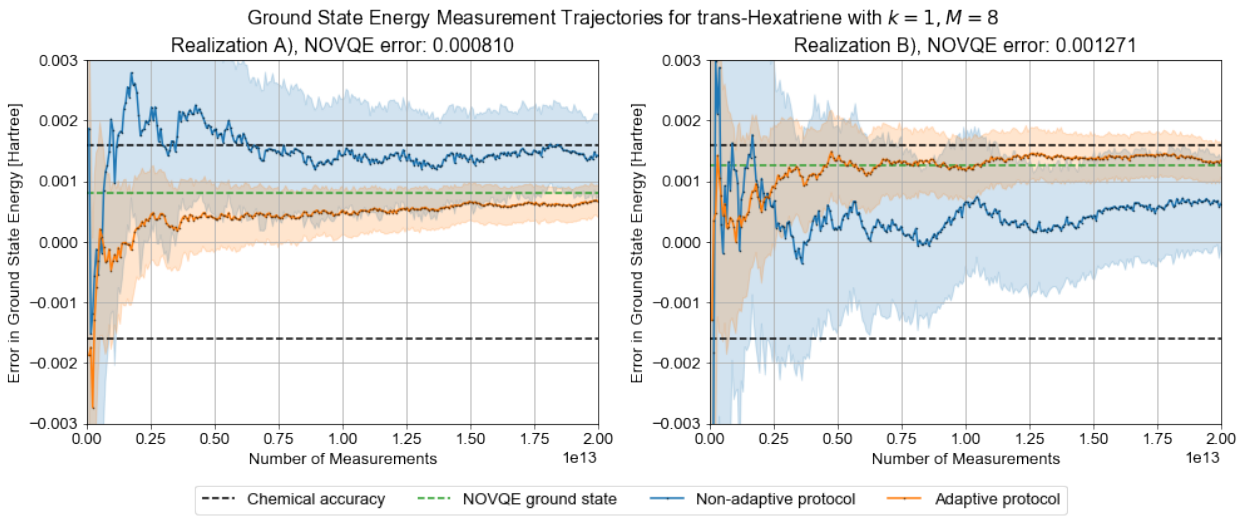

Figure 5.4: Comparison of the ability of the adaptive and non-adaptive schemes for scheduling measurements to resolve the ground state energy of trans-Hexatriene in two different NOVQE subspaces of $M = 8$ optimized $k = 1$ k-UpCCGSD states. The evolution of the estimated ground state energies is plotted in solid lines together with 2σ error bars indicated by the shaded regions. The actual energies of the ground states in the NOVQE subspaces are indicated with dashed green lines. Panels A and B show two different typical realizations of the measurement record as the total number of measurements increases. In both cases, the adaptive protocol converges significantly more quickly than the non-adaptive one. Note that the variance of the experimental measurements are approximated using upper bounds and that the true numbers required for both the adaptive and non-adaptive schemes are likely to be lower [140, 299].

For the non-adaptive scheme we observe $\kappa \approx 2.4 \cdot 10^6$ and $\kappa \approx 2.9 \cdot 10^6$ for panels A and B, while for the adaptive scheme we see $\kappa \approx 3.7 \cdot 10^5$ and $\kappa \approx 6.5 \cdot 10^5$. The reference value for a regular VQE calculation is $\kappa \approx 1.6 \cdot 10^2$, determined using the same bounds assumed throughout this comparison [140, 197]. Comparing with the simpler H_4 example, we see that the adaptive scheme for measuring the NOVQE ground state energy of Hexatriene results in a slightly larger gain when compared to the non-adaptive scheme, but still falls short of the goal of reducing the number of measurements to an experimentally plausible number. One promising avenue to further reducing this cost is the adaptation of recently proposed strategies for measurement in the context of regular VQE to NOVQE [299]. These strategies have been shown to reduce the number of circuit executions by orders of magnitude when compared with the bounds used to derive the number of measurements in this work. Further study is required in order to determine if one can alter the optimization process of the NOVQE states themselves or their coefficients, in order to achieve an additional reduction

in the measurement cost.

5.5 Discussion and Outlook

We have introduced an extension to the variational quantum eigensolver that calls for the ground state energy to be approximated by solving a generalized eigenvalue problem in a subspace that is spanned by a linear combination of M parametrized quantum wavefunctions. The resulting logical wavefunction ansatz is a linear combination of all M states in the subspace, but its properties can be determined by only pairwise measurements of the Hamiltonian and overlap matrices. In particular, construction of all off-diagonal matrix elements required for the generalized eigenvalue equation can be made using the unmodified state preparation circuits for each parameterized basis wavefunction, together with O(N) additional two-qubit gates, where N is the number of spin-orbitals. Therefore, it is possible to increase the flexibility of the ansatz without requiring additional coherent quantum resources. By analogy with the non-orthogonal configuration interaction method of classical quantum chemistry [308, 309, 312], we call our approach the non-orthogonal variational quantum eigensolver, NOVQE.

Our proposal necessitates off-diagonal measurements of the Hamiltonian and overlap matrices. We perform these using a modified Hadamard test. Naively, this would require us to implement controlled versions of the quantum circuits for state preparation. To avoid this cost, we demanded that the state preparation circuits all act on a common reference state and preserve fermionic excitation number. This allowed us to avoid the need to add controls to the ansatz circuits, by instead performing controlled swap operations between two copies of the system register, a cost that scales linearly and modestly with the system size.

To determine the ground state energy in the subspace, our approach requires that we measure all M^2 elements of the Hamiltonian and overlap matrices in the NOVQE subspace. We presented a statistical strategy for estimating the uncertainty in the resultant ground state energy estimate for a given uncertainty in the matrix elements. We also pointed out how the machinery that generates these estimates can be leveraged in a Monte Carlo sampling process to determine which matrix element should be chosen for additional measurements to optimally reduce the uncertainty. We proposed an iterative approach, in which small batches of measurements are repeatedly performed according to this Monte Carlo prescription, to minimize the overall number of circuit repetitions required by our NOVQE method.

We demonstrated an implementation of our approach using a collection of k-UpCCGSD wavefunctions to approximate the ground state of two model strongly-correlated systems, a square geometry of H_4 and the π-space of Hexatriene in two configurations. Growing the NOVQE subspace by adding and optimizing one state at a time, we showed how a collection of ansätze which individually struggle to represent the ground state can be fruitfully combined combined to form a more powerful logical ansatz. In our numerical experiments we observed that the marginal utility of adding additional states to the NOVQE subspace remained large, even as the size of the space increased. It is interesting to compare this with the

commonly noted behavior of classical non-orthogonal methods, which generate a collection of non-orthogonal Slater determinants and diagonalize in the resulting subspace [307–313]. These approaches eventually enter a regime where convergence slows down significantly as states are added to the subspace, sometimes before the desired accuracy is reached. This suggests that the there is a benefit in NOVQE's ability to make use of wavefunctions more sophisticated than the Slater determinants available to classical non-orthogonal methods, allowing for a balance between the number of distinct wavefunctions and their flexibility.

To characterize our proposal for adaptively scheduling measurements to minimize the number of circuit repetitions required by our approach, we focused on quantifying the number of measurements required to approximate the ground state energy in a fixed NOVQE subspace. For the purposes of this investigation we approximated the variance of the individual matrix element measurements using the bounds described in Refs. 140 and 197. For both our square H_4 and our equilibrium configuration of trans-Hexatriene, we optimized collections of NOVQE states and froze their parameters. We then applied our adaptive approach for scheduling measurements and compared it to a simpler non-adaptive scheme, in which each matrix element was measured the same number of times. We found that our adaptive approach used somewhat fewer measurements than a simpler non-adaptive strategy, but still dramatically more than it would take to measure the energy in the standard VQE formalism. It would be worthwhile to understand whether similar challenges appear for other proposed quantum non-orthogonal methods [221, 222, 301–303].

We can imagine several routes towards ameliorating this difficulty and developing NOVQE further. First, having states that are nearly linearly dependent in the NOVQE subspace can dramatically increase the cost of measurement. Developing an optimization strategy for the individual states, or their coefficients, that regularizes this behavior away would be useful. Related to this is the possibility of extending the tools for measuring analytical gradients of parametrized quantum circuits to work with the NOVQE formalism. Another avenue for future work would be the development of good initialization strategies for NOVQE, potentially using reference states derived from a classical NOCI calculation. Finally, recent work has shown that a measurement strategy based on factorizations of the two-electron integral tensor can dramatically reduce the cost of the standard VQE approach, lowering the number of separately measured terms from $O(N^4)$ to $O(N)$ [299]. The resulting cost reduction is especially large when compared to the type of bounds used throughout this paper [140, 197]. Adapting this approach for use with NOVQE is likely to offer a significant improvement.

Beyond these modifications to the NOVQE approach outlined in this paper, it is also conceivable that the tools we have presented might be usefully employed in other ways. For example, we have focused here on the variational optimization of a logical ansatz that is a superposition of individual parametrized wavefunctions. An alternative is to take inspiration from Ref. 301 and from the classical NOCI method [308, 309, 312], and optimize the individual wavefunctions separately, solving the generalized eigenvalue problem only once with the final collection of states. In this vein, there are several recent proposals which form a non-orthogonal basis using a collection of time-evolved reference states [221, 222]. Another possible direction

to pursue is the inclusion of one or more states in the NOVQE subspace that can be classically optimized, only turning to the use of more general parametrized quantum circuits to prepare small corrections to the classically tractable states. All of these ideas have the potential to benefit from the tools we have developed for efficiently performing the required matrix element measurements.

In summary, this work has presented a promising new extension to the VQE formalism and highlighted both its advantages and its drawbacks, some of which may be of general concern for developers of other quantum non-orthogonal methods. We have also presented a strategy for compiling off-diagonal matrix element measurements and promoted a general approach to Monte Carlo estimation of uncertainty. The circuit simulations of the k-UpCCGSD ansatz presented here add to the analyses of Refs. 224, 325, and 189. We believe that the ability of our NOVQE to trade off coherent quantum resources for additional measurements may prove to be a useful tool in making use of NISQ-era quantum hardware for studying challenging strongly correlated systems.

In the final stages of preparing this manuscript two works were posted which independently developed approaches using the matrix elements between collections of quantum states for other applications. One appears in the context of variational quantum algorithms for solving linear systems of equations [326], while the other proposes a strategy for approximating the low energy subspace of a Hamiltonian in terms of time-evolved trial wavefunctions [221].

5.6 Additional Computational Details

In this section, we shall provide some additional miscellaneous details related to the implementation of our numerical calculations. The main thing that we expand upon is our implementation of a Trotter step of the k-UpCCGSD ansatz. In Section 5.3.4.1 we recalled the k-UpCCGSD ansatz introduced in Chapter 4. We gave the form of the ansatz and the cluster operator in Eq. 5.13 and Eq. 5.14 respectively. We mentioned that we implemented a Trotterized approximation to this ansatz using the techniques of Ref. 172. In Figure 5.5 we present a circuit diagram for this ansatz, as applied to H_4 in the text above.

We already discussed some of the basic computational techniques for the electronic structure calculations important to this chapter at the end of Chapter 4 in Section 4.7. Most of the additional tools required for this chapter have already been described above. In Section 5.3.2 and Section 5.3.3 we gave a detailed overview of our numerical strategy for dealing with statistical noise during matrix element measurement. This statistical noise was accounted for using bounds of the type that we mentioned in Section 3.5.3. We shall review these bounds in more detail in Section 6.5. In order to calculate the matrix elements, we classically simulated the action of the state preparation circuit for the individual ansatz states. We then evaluated the matrix elements of Eq. 5.2 using these classically simulated wavefunctions. We did this, rather than simulating our Hadamard test circuit, in order to make the simulations more affordable by removing the requirement for a second system register.

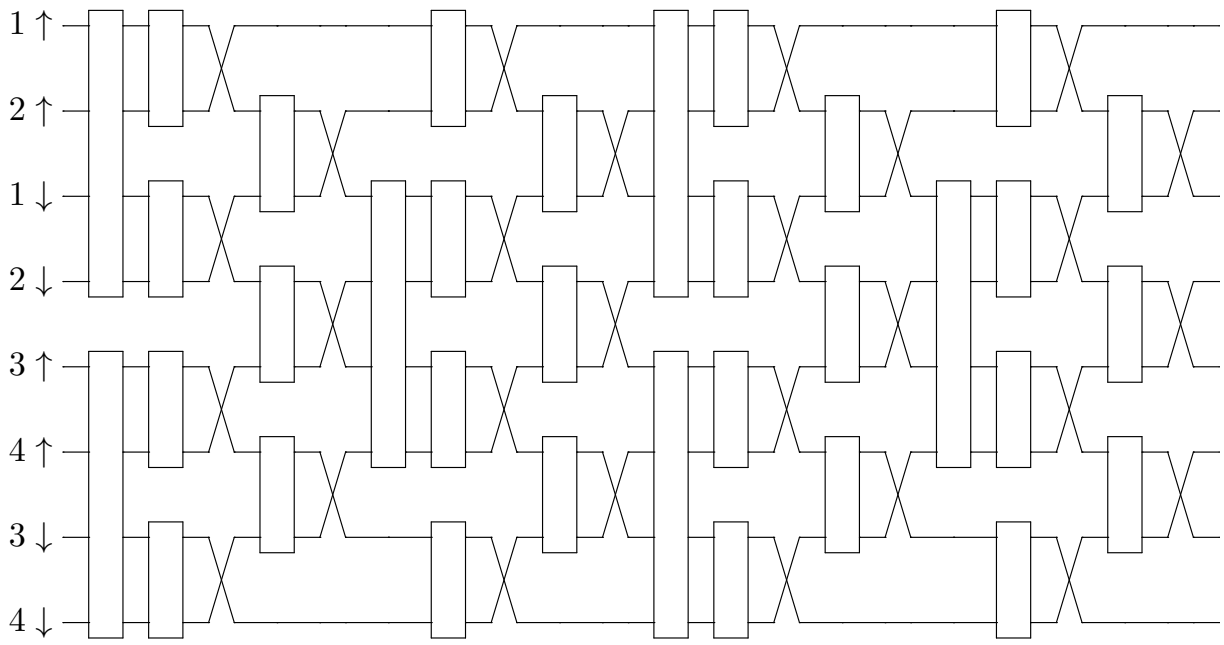

Figure 5.5: A circuit diagram for our implementation of one block of the k-UpCCGSD ansatz. Using k repetitions of this circuit we can construct a single Trotter step. Here the crossed wires should be understood as applications of the fermionic swap gate of Eq. 3.26, and the white squares should be understood as placeholders that contain the two and four-qubit interactions that implement the exponential of the individual terms in the cluster operator (as described in Section 5.3.4.1). Note that this construction implies a particular Trotter ordering. The symbols (e.g., $1\uparrow$) on the left-hand side of the diagram indicate the initial positioning of the fermionic modes. This figure is reproduced with permission from Ref. 172.

We emphasize here that we encountered severe challenges in performing the optimization. The scheme with random restarts and perturbations that we describe in Section 5.3.4.2 was necessary in order to obtain good quality solutions. It is highly likely that with the addition of statistical noise on the objective function queries we would have struggled to achieve the same quality of results. What portion of the blame for this difficulty lies with the choice of ansatz circuit and what portion lies with the NOVQE formalism is a question that deserves further investigation.

5.7 Hexatriene Geometries

The equilibrium geometry was obtained from the geometry optimization with ωB97X-D [327] and cc-pVTZ [35] using a development version of Q-Chem[42]. The 90° twisted configuration was obtained by rotating the middle C-C double bond out-of-plane. All distances are given in angstroms.

Atom	X	Y	Z
C	(0.5987833,	0.2969975,	0.0000000)
H	(0.6520887,	1.3822812,	0.0000000)
C	(-0.5987843,	-0.2970141,	0.0000000)
H	(-0.6520904,	-1.3822967,	0.0000000)
C	(-1.8607210,	0.4195548,	0.0000000)
H	(-1.8010551,	1.5036080,	0.0000000)
C	(-3.0531867,	-0.1693136,	0.0000000)
H	(-3.9685470,	0.4053361,	0.0000000)
H	(-3.1479810,	-1.2485605,	0.0000000)
C	(1.8607264,	-0.4195599,	0.0000000)
H	(1.8010777,	-1.5036141,	0.0000000)
C	(3.0531816,	0.1693296,	0.0000000)
H	(3.9685551,	-0.4052992,	0.0000000)
H	(3.1479561,	1.2485793,	0.0000000)

Table 5.1: The geometry of the equilibrium configuration of trans-Hexatriene.

Atom	X	Y	Z
C	(0.5987833,	0.2969975,	0.0000000)
H	(1.3716346,	-0.0683717,	0.6707370)
C	(-0.5987843,	-0.2970141,	0.0000000)
H	(-1.3716354,	0.0683544,	0.6707361)
C	(-0.9484080,	-1.4197297,	-0.8504282)
H	(-0.1721763,	-1.7803215,	-1.5183873)
C	(-2.1390983,	-2.0121775,	-0.8520831)
H	(-2.3554088,	-2.8468591,	-1.5037144)
H	(-2.9353514,	-1.6772360,	-0.1982062)
C	(0.9484189,	1.4197134,	-0.8504230)
H	(0.1721980,	1.7803171,	-1.5183881)
C	(2.1391167,	2.0121462,	-0.8520613)
H	(2.3554502,	2.8468291,	-1.5036834)
H	(2.9353585,	1.6771903,	-0.1981764)

Table 5.2: The geometry of the 90° twisted configuration of Hexatriene.

Chapter 6

Efficient and Noise Resilient Measurement for Quantum Chemistry on a Quantum Computer

6.1 Preface

This chapter is taken from the previously released Ref. 299, currently available on the arXiv and under review for publication. It was co-authored by the author of this thesis, Dr. Jarrod McClean, Dr. Nicholas Rubin, Dr. Zhang Jiang, Professor Nathan Wiebe, Professor K. Birgitta Whaley, and Dr. Ryan Babbush. The author of this thesis and Dr. Babbush conceived the idea and co-wrote the majority of the paper. The author of this thesis performed all numerical simulations except for the Bayesian analysis, which Professor Wiebe carried out. All of the co-authors participated in discussions which developed the theory and shaped the project.

6.2 Introduction

Given the recent progress in quantum computing hardware, it is natural to ask where the first demonstration of a quantum advantage for a practical problem will occur. Since the first experimental demonstration by Peruzzo et al. [17], the variational quantum eigensolver (VQE) framework has offered a promising path towards utilizing small and noisy quantum devices for simulating quantum chemistry. The essence of the VQE approach is the use of the quantum device as a co-processor which prepares a parameterized quantum wavefunction and measures the expectation value of observables. In conjunction with a classical optimization algorithm, it is possible to then minimize the expectation value of the Hamiltonian as a function of the parameters, arriving at approximations for the wavefunction, energy, and other properties of the ground state [17, 18, 140, 149, 193, 224, 303, 328]. A growing body of work attempting to understand and ameliorate the challenges associated with using VQE to target non-trivial

systems has emerged in recent years [150, 172, 196, 208, 212–215, 300, 315, 329–332]. In this article we address the challenge posed by the large number of circuit repetitions needed to perform accurate measurements and propose a new scheme that dramatically reduces this cost. Additionally, we explain how our approach to measurement has reduced sensitivity to readout errors and also enables a powerful form of error mitigation.

Within VQE, expectation values are typically estimated by Hamiltonian averaging. Under this approach, the Hamiltonian is decomposed into a sum of operators that are tensor products of single-qubit Pauli operators, commonly referred to as Pauli strings. The expectation values of the Pauli strings are determined independently by repeated measurement. When measurements are distributed optimally between the Pauli strings P_ℓ, the total number of measurements M is upper bounded by

$$M \leq \left(\frac{\sum_\ell |\omega_\ell|}{\epsilon} \right)^2, \quad \text{where} \quad H = \sum_\ell \omega_\ell P_\ell \tag{6.1}$$

is the Hamiltonian whose expectation value we estimate as $\sum_\ell \omega_\ell \langle P_\ell \rangle$, the ω_ℓ are scalars, and ϵ is the target precision [18, 197]. Prior work assessing the viability of VQE has used bounds of this form and concluded that chemistry applications require "a number of measurements which is astronomically large" (quoting from Ref. 18).

Several recent proposals attempt to address this obstacle by developing more sophisticated strategies for partitioning the Hamiltonian into sets of simultaneously measurable operators [172, 196, 315, 329–331]. We summarize their key findings in Table 6.2. This work has a similar aim, but we take an approach rooted in a decomposition of the two-electron integral tensor rather than focusing on properties of Pauli strings. We quantify the performance of our proposal by numerically simulating the variances of our term groupings to more accurately determine the number of circuit repetitions required for accurate measurement of the ground state energy. This contrasts with the analysis in other recent papers that have instead focused on using the number of separate terms which must be measured as a proxy for this quantity. By that metric, our approach requires a number of term groupings that is linear in the number of qubits - a quartic improvement over the naive strategy and a cubic improvement relative to these recent papers. However, we argue that the number of distinct term groupings alone is not generally predictive of the total number of circuit repetitions required, because it does not consider how the covariances of the different terms in these groupings can collude to either reduce or increase the overall variance. We will show below that our approach benefits from having these covariances conspire in our favor; for the systems considered here our approach gives up to three orders of magnitude reduction in the total number of measurements, while also providing an empirically observed asymptotic improvement.

Although there are a variety of approaches to simulating indistinguishable fermions with distinguishable qubits [176, 179, 181], the Jordan-Wigner transformation is the most widely used. This is due to its simplicity and to the fact that it allows for the explicit construction of a number of useful circuit primitives not available under more sophisticated encodings. These include the Givens rotation network that exactly implements a change of single-particle

Ref.	Partitioning Method	Circuit Description	# of Partitions	Gate Count	Depth	Connectivity	Diagonal
[140]	commuting Pauli heuristic	-	$O(N^4)$	-	-	-	-
[193]	compatible Pauli heuristic	single rotations	$O(N^4)$	N	1	any	no
[197]	n-representability constraints	single rotations	$O(N^4)$	N	1	any	no
[331]	mean-field partitioning	fast feed-forward	$O(N^4)$	$O(N)$	$O(N)$	full	no
[196]	compatible Pauli clique cov.	single rotations	$O(N^4)$	N	1	any	no
[172]	counting argument	swap networks	$O(N^3)$	$O(N^2)$	$O(N)$	linear	no
[329]	commuting Pauli graph color.	stabilizer formalism	$O(N^3)$	-	-	full	no
[315]	anticommuting Pauli clique cov.	Pauli evolutions	$O(N^3)$	$O(N^2 \log(N))$	-	full	no
[330]	commuting Pauli clique cover	symplectic subspaces	$O(N^3)$	$O(N^2/\log N)$	-	full	no
[332]	commuting Pauli clique cover	stabilizer formalism	$O(N^3)$	$O(N^2)$	-	full	no
here	integral tensor factorization	Givens rotations	$O(N)$	$N^2/4$	$N/2$	linear	yes

Table 6.1: A history of ideas reducing the measurements required for estimating the energy of arbitrary basis chemistry Hamiltonians with the variational quantum eigensolver. Here N represents the number of spin-orbitals in the basis. Gate counts and depths are given in terms of arbitrary 1- or 2-qubit gates restricted to the geometry of 2-qubit gates specified in the connectivity column. What we mean by "compatible" Pauli groupings is that the terms can be measured at the same time with only single qubit rotations prior to measurement. We report whether terms are measured in a diagonal representation as this is important for enabling strategies of error-mitigation by postselection. The number of partitions refers to the number of unique term groupings which can each be measured with a single circuit - thus, this reflects the number of unique circuits required to generate at least one sample of each term in the Hamiltonian. However, we caution that one cannot infer the total number of measurements required from the number of partitions, and often this metric is highly misleading. The overall number of measurements required is also critically determined by the variance of the estimator of the energy. As explained in the first entry of this table, when terms are measured simultaneously one must also consider the covariance of those terms. In some cases, a grouping strategy can decrease the number of partitions but increase the total number of measurements required by grouping terms with positive covariances. Alternatively, strategies such as the third entry in this table actually increase the number of partitions while reducing the number of measurements required overall by lowering the variance of the estimator.

basis [170, 172, 175, 333]. A disadvantage of using the Jordan-Wigner transformation is the fact that it maps operators acting on a constant number of fermionic modes to qubit operators with support on up to all N qubits. In the context of measurement, the impact of this non-locality can be seen by considering a simple model of readout error such as a symmetric bitflip channel. Under this model, a Pauli string with support on N qubits has N opportunities for an error that reverses the sign of the measured value, leading to estimates of expectation values that are exponentially suppressed in N (see Section 6.3.3). It has recently been shown that techniques based on fermionic swap networks can avoid the overheads and disadvantages imposed by the non-locality of the Jordan-Wigner encoding in a variety of contexts, including during measurement [170, 172, 175]. Our work will likewise avoid this challenge without leaving the Jordan-Wigner framework, allowing estimation of 1- and 2-particle fermionic operator expectation values by the measurement of only 1-local and 2-local qubit operators, respectively.

In addition to this reduction in the support of the operators that we measure, our work offers another opportunity for mitigating errors. It has been observed that when one is interested in states with a definite eigenvalue of a symmetry operator, such as the total particle number, η, or the z-component of spin, S_z, it can be useful to have a method which removes the components of some experimentally prepared state with support on the wrong symmetry manifold [150, 213, 214, 328]. Two basic strategies to accomplish this have been proposed. The first of these strategies is to directly and "non-destructively" measure the symmetry operator and discard those outcomes where the undesired eigenvalue is observed, projecting into the proper symmetry sector by postselection. In order to construct efficient measurement schemes, prior work in this direction has focused on measuring the parities of η and S_z, rather than the full symmetry operators, [213, 214]. These proposals involve non-local operations that usually require $O(N)$ depth, which may induce further errors during their implementation. The second class of strategies builds upon the foundation of Ref. 212 and uses additional measurements together with classical post-processing to calculate expectation values of the projected state without requiring additional circuit depth [150, 213, 328], a procedure which can be efficiently applied to the parity of the number operator in each spin sector. In this work, we show how our proposal for measurement naturally leads to the ability to postselect directly on the proper eigenvalues of the operators η and S_z, rather than on their parities.

6.3 Results

6.3.1 Using Hamiltonian Factorization for Measurements

The crux of our strategy for improving the efficiency and error resilience of Hamiltonian averaging is the application of tensor factorization techniques to the measurement problem. Using a representation discussed in the context of quantum computing in Refs. 77, 175, 334, we begin with the factorized form of the electronic structure Hamiltonian in second

quantization:

$$H = U_0 \left(\sum_p g_p n_p \right) U_0^\dagger + \sum_{\ell=1}^{L} U_\ell \left(\sum_{pq} g_{pq}^{(\ell)} n_p n_q \right) U_\ell^\dagger, \tag{6.2}$$

where the values g_p and $g_{pq}^{(\ell)}$ are scalars, $n_p = a_p^\dagger a_p$, and the U_ℓ are unitary operators which implement a single particle change of orbital basis. Specifically,

$$U = \exp\left(\sum_{pq} \kappa_{pq} a_p^\dagger a_q \right), \quad U a_p^\dagger U^\dagger = \sum_q \left[e^\kappa \right]_{pq} a_q^\dagger, \tag{6.3}$$

where $[e^\kappa]_{pq}$ is the p, q entry of the matrix exponential of the anti-Hermitian matrix κ that characterizes U.

Numerous approaches that accomplish this goal exist, including the density fitting approximation [335, 336], and a double factorization which begins with a Cholesky decomposition or eigendecomposition of the two-electron integral tensor [175, 336–342]. In this work, we use such an eigendecomposition and refer readers to Section 6.7 and to Refs. 175 and 77 for further details. The eigendecomposition step permits discarding small eigenvalues to yield a controllable approximation to the original Hamiltonian. While such low rank truncations are not central to our approach and would not significantly reduce the number of measurements, doing so would asymptotically reduce L (and thus ultimately, the number of distinct measurement term groupings). Such decompositions have been explored extensively in the context of electronic structure on classical computers on a far wider range of systems than those considered here [337, 339, 342–345]. It has been found that $L = O(N)$ is sufficient for the case of arbitrary basis quantum chemistry, both in the large system and large basis set limits [337]. Furthermore, specific basis sets exist where $L = 1$, such as the plane wave basis or dual basis of Ref. 333.

Our measurement strategy, which we shall refer to as "Basis Rotation Grouping," is to apply the U_ℓ circuit directly to the quantum state prior to measurement. This allows us to simultaneously sample all of the $\langle n_p \rangle$ and $\langle n_p n_q \rangle$ expectation values in the rotated basis. We can then estimate the energy as

$$\langle H \rangle = \sum_p g_p \langle n_p \rangle_0 + \sum_{\ell=1}^{L} \sum_{pq} g_{pq}^{(\ell)} \langle n_p n_q \rangle_\ell , \tag{6.4}$$

where the subscript ℓ on the expectation values denotes that they are sampled after applying the basis transformation U_ℓ. The reason that the $\langle n_p \rangle_\ell$ and $\langle n_p n_q \rangle_\ell$ expectation values can be sampled simultaneously is because under the Jordan-Wigner transformation, $n_p = (\mathbb{1} + Z_p)/2$, which is a diagonal qubit operator. In practice, we assume a standard measurement in the computational basis, giving us access to measurement outcomes for all diagonal qubit operators simultaneously. Thus, our approach is able to sample all terms in the Hamiltonian with only $L + 1 = O(N)$ distinct term groups.

Fortunately, the U_ℓ are exceptionally efficient to implement, even on hardware with minimal connectivity. Following the strategy described in Ref. 170, and assuming that the system is an eigenstate of the total spin operator, any change of single-particle basis can be performed using $N^2/4 - N/2$ two qubit gates and gate depth of exactly N, even with the connectivity of only a linear array of qubits [170]. This gate depth can actually be improved to $N/2$ by further parallelizing the approach of [170], making using ideas that are explained in the context of multiport interferometry in [346]. In fact, a further optimization is possible by performing the second matrix factorization discussed in Ref. 175. This would result in only $O(\log^2 N)$ distict values of the $g_{pq}^{(\ell)}$ and a gate complexity for implementing the U_ℓ which is reduced to $O(N \log N)$; however, we note that this scaling is only realized in fairly large systems when N is growing towards the thermodynamic (large system) rather than continuum (large basis) limit.

The primary objective of our measurement strategy is to reduce the time required to measure the energy to within a fixed accuracy. Because different hardware platforms have different repetition rates, we focus on quantifying the time required in terms of the number of circuit repetitions. We shall present data for electronic ground states that demonstrates the effectiveness of our Basis Rotation Grouping approach in comparison to three other measurement strategies and the upper bound of Eq. 6.1. All calculations were performed using the open source software packages OpenFermion and Psi4 [262, 263]. Specifically, we used exact calculations of the variance of expectation values with respect to the full configuration interaction (FCI) ground state to determine the number of circuit repetitions required. The calculations presented here are performed for symmetrically stretched Hydrogen chains with various bond lengths and numbers of atoms, for a symmetrically stretched water molecule, and for a stretched Nitrogen dimer, all in multiple basis sets. We justify our focus on the electronic ground states here by noting that most variational algorithms for chemistry attempt to optimize ansatze that are already initialized near the ground state.

In order to calculate the variance of the estimator of the expectation value of the energy, it is necessary to determine the distribution of measurements between the different term groupings. Refs. 18 and 197 provide a prescription for the optimal choice. They demand that (in the notation of Eq. 6.1) each term H_ℓ is measured a fraction of the time f_ℓ equal to

$$f_\ell = \frac{|\omega_\ell| \sqrt{1 - \langle H_\ell \rangle^2}}{\sum_j |\omega_j| \sqrt{1 - \langle H_j \rangle^2}}. \tag{6.5}$$

In practice, the expectation values in the above expression are not known ahead of time and so the optimal measurement fractions f_ℓ cannot be efficiently and exactly determined a priori. For the purposes of this paper, we approximated the ideal distribution of measurements by first performing a classically tractable configuration interaction singles and doubles (CISD) calculation of the quantities in Eq. 6.5. We shall show that this approximation introduces a negligible overhead in measurement time for all systems considered in this work. One could also envisage using an adaptive measurement scheme that makes additional measurements

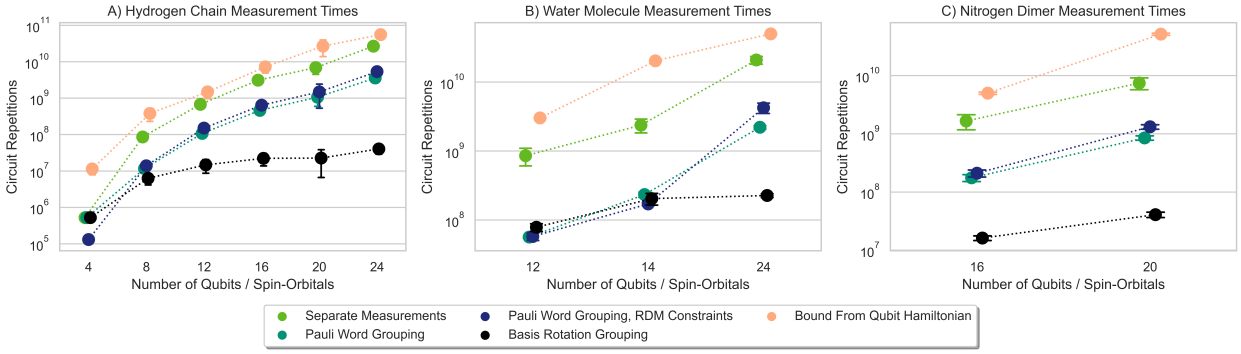

Figure 6.1: The number of circuit repetitions required to estimate the ground state energy of various Hydrogen chains, a water molecule, and a Nitrogen dimer with each of the five measurement strategies indicated in the legend. The specific systems considered are enumerated in Table 6.2. A target precision corresponding to a 2σ error bar of 1.0 millihartree is assumed. Calculations performed on systems which require the same number of qubits (spin-orbitals) are plotted together in columns. The cost of our proposed measurement strategy appears to have a lower asymptotic scaling than any other method we consider and obtains a speedup of more than an order of magnitude compared to the next best approach for a number of systems.

based on the observed sample variance, in order to approximate the ideal partitioning of measurement time, such as the one described in Ref 298.

6.3.2 Circuit Repetitions Required for Energy Measurement

In Figure 6.1 we plot the number of circuit repetitions for our proposed "Basis Rotation Grouping" measurement approach (black circles), together with three other measurement strategies and the upper bound based on Eq. 6.1 for the systems listed in Table 6.2. The first and most basic alternative strategy is simply to apply no term groupings and measure each Pauli string independently, a strategy we refer to as "Separate Measurements" (lime green circles). A more sophisticated approach, similar to the one described in Ref. 196, is to partition the Pauli strings into groups of terms that can be measured simultaneously. In the context of a near-term device we consider two Pauli strings P_j and P_k simultaneously measurable if and only if they act with the same Pauli operator on all qubits on which they both act non-trivially. Pauli Strings that satisfy this condition can be simultaneously measured using only single qubit rotations and measurement. In order to efficiently partition the Pauli strings into groups we choose to take all of the terms which only contain Z operators as one partition and then account for the remaining Pauli words heuristically by adding them at random to a group until no more valid choices remain before beginning a new group. We refer to this approach as "Pauli Word Grouping" (teal circles). The final strategy that we compare with

System	Interatomic Spacings (Å)	Basis Set	Frozen Orbitals	Number of Qubits
H_2	.6, .7, ... 1.3	STO-3G	None	4
H_2	.6, .7, ... 1.3	6-31G	None	8
H_4	.6, .7, ... 1.3	STO-3G	None	8
H_6	.6, .7, ... 1.3	STO-3G	None	12
H_4	.6, .7, ... 1.3	6-31G	None	16
H_8	.6, .7, ... 1.3	STO-3G	None	16
H_2	.6, .7, ... 1.3	cc-pVDZ	None	20
H_{10}	.6, .7, ... 1.3	STO-3G	None	20
H_6	.6, .7, ... 1.3	6-31G	None	24
H_2O	.8, .9, ... 1.5	STO-3G	1	12
H_2O	.8, .9, ... 1.5	STO-3G	None	14
H_2O	.8, .9, ... 1.5	6-31G	1	24
N_2	.9, 1.0, ... 1.6	STO-3G	2	16
N_2	.9, 1.0, ... 1.6	STO-3G	None	20

Table 6.2: List of the molecular systems considered in this work, displayed in order of increasing number of qubits, for each type of system. The hydrogen systems consist of a chain of atoms arranged in a line, with equal interatomic spacing. The interatomic spacing for the water molecules refers to the length of the symmetrically stretched bonds O-H bonds, which are separated by a fixed angle of 104.5 deg. The active space used for each system has one spatial orbital for every two qubits. A non-zero number of frozen orbitals indicates the number of molecular orbitals fixed in a totally occupied state.

preprocesses the Hamiltonian by applying the techniques based on the fermionic marginal (RDM) constraints described in Ref. 197, before applying the Jordan-Wigner transformation and using the same heuristic grouping strategy to group simultaneously measurable Pauli strings together [1]. We call this latter strategy "Pauli Word Grouping, RDM Constraints" (dark blue circles).

We refer to the bound of Eq. 6.1 as being based on the Hamiltonian coefficients and calculate it from the Jordan-Wigner transformed Hamiltonian, (meaning that the ω_ℓ in Eq. 6.1 are the coefficients of Pauli strings). This bound is indicated by salmon-colored circles in Figure 6.1. We note that attempting to calculate a similar bound directly from the fermionic Hamiltonian (meaning that the ω_ℓ in Eq. 6.1 would be the coefficients of the

[1] In the process of preparing this manuscript we have become aware of several recent works that employ more sophisticated strategies for grouping Pauli words together or employing a different family of unitary transformations than those we consider to enhance the measurement process [196, 315, 329, 330]. It would be an interesting subject of future work to calculate and compare the number of circuit repetitions required by these approaches.

Measurement Strategy	$\langle\log(a)\rangle$	$\Delta(a)$	$\langle\mathbf{b}\rangle$	$\sigma(b)$
Bound from Qubit Hamiltonian	-6.0	0.3	**4.90**	0.02
Separate Measurements	-9.3	0.4	**5.70**	0.06
Pauli Word Grouping	-8.9	0.4	**4.88**	0.06
RDM Constraints	-10.8	0.4	**5.63**	0.06
Basis Rotation Grouping	-6.0	0.3	**2.75**	0.01

Table 6.3: Bounds and uncertainties result from Bayesian inference using a Monte-Carlo approximation with 10^6 particles for all Hydrogen FCI data [347]. We assume $\log(N_{meas}) = \log(a) + \hat{x} + b\log(N)$ where $\hat{x} \sim \mathcal{N}(0, 0.1)$. The prior distributions are uniform for $\log(a)$ and b over $[-15, 1]$ and $[1, 20]$ respectively. Here $\sigma(b)$ is the posterior standard deviation for b and $\Delta(a)$ is the posterior standard deviation of $\log(a) + \hat{x}$. "RDM Constraints" refers to the Pauli Word Grouping approach with the RDM constraints applied, as in the text.

terms $a_p^\dagger a_q$ or $a_p^\dagger a_q^\dagger a_r a_s$) leads to different bounds. These are derived in Section 6.5, where they are shown to be substantially looser for the systems we consider in this work. While one would not measure the fermion operators directly, it is surprising that these bounds would be significantly different. We refer the interested reader to Section 6.5 for an analysis and discussion of this phenomenon.

Considering first the Hydrogen chain systems in Figure 6.1 (left panel, A), we note that our Basis Rotation Grouping approach consistently outperforms the other strategies for simulations with more than four fermionic modes, requiring significantly fewer measurements. Interestingly, while the bounds from the qubit Hamiltonian and other three methods appear to have relative performances that are stable across a variety of system sizes, the Basis Rotation Grouping method appears to have a different asymptotic scaling, at least for Hydrogen chains of increasing length and basis set size. This is likely due to large scale effects that only manifest when approaching a system's thermodynamic limit (which one approaches particularly quickly for Hydrogen chains) [70]. In Table II we quantify this asymptotic scaling by assuming that the dependence of the variance on the number of qubits N in the Hydrogen chain's Hamiltonian can be modeled by the functional form aN^b for some constants a and b which we fit using a Bayesian analysis described in the table caption [347]. By contrast, the data from the minimal basis water molecule (panel B in Figure 6.1) shows no benefit in measurement time from our method compared to the heuristic grouping strategies. However, the advantage of our approach becomes significant for that system in larger basis sets, a trend which is also apparent to a lesser extent for the Nitrogen dimer (panel C in Figure 6.1).

We find that applying the RDM constraints of Ref. 197 to our Pauli Word Grouping strategy (the combination is plotted with dark blue circles in Figure 6.1) does not significantly reduce the observed variance, despite the fact that the use of the RDM constraints have been previously shown to dramatically reduce the bounds on the number of circuit repetitions required [197]. In Section 6.6, we explore the possibility that this is due to the fact that these

constraints were applied to minimize a bound of the same form as Eq. 6.1 that is however formulated using the fermionic representation of the Hamiltonian. We present evidence in Section 6.5 that, in the context of such bounds, the use of the Jordan-Wigner transformed operators leads to surprisingly different results. However, as we show in Section 6.6, we find that the actual variance with respect to the ground state is not substantially changed by applying the same constraints and performing the minimization using the qubit representation of the Hamiltonian.

Earlier we explained that the data presented in Figure 6.1 was calculated by distributing the measurements between different term groupings according Eq. 6.5 using the variance of each term calculated with a classically efficient CISD approximation to the ground state. Any deviation from the ideal allocation of measurement cycles (obtained by evaluating Eq. 6.5 with respect to the true ground state) must increase the time required for measurement. In Figure 6.2 we present the ratio between the time required with the approximate distribution and the time required under the optimal one for each of the systems treated in the work. We find that impact from this approximation is negligible, with the largest observed increase in measurement time being below 3%.

Overall, Figure 6.1 speaks for itself in showing that in most cases there is a very significant reduction in the number of measurements required when using our strategy - sometimes by up to three orders of magnitude for even modestly sized systems. Furthermore, these improvements become more significant as system size grows.

6.3.3 Error Mitigation

Beyond the reduction in measurement time, our approach also provides two distinct forms of error mitigation. First, it reduces the susceptibility to readout errors by replacing the measurement of $O(N)$ qubit operators with 1 and 2 qubit operators. Second, it allows us to perform postselection based on the eigenvalues of the particle number operators in each spin sector. Both properties stem from measuring the Hamiltonian only in terms of density operators in different basis sets.

The first benefit, the reduction in readout errors, is a consequence of only needing to measure expectation values of operators that have support on one or two qubits. Direct measurement of the Jordan-Wigner transformed Hamiltonian using only single-qubit rotations and measurement involves measuring operators with support on $O(N)$ qubits. To demonstrate how reducing the support of the operators helps to mitigate errors, we consider a simple model of measurement error: the independent, single-qubit symmetric bitflip channel. When estimating the expectation value of a Pauli string P_ℓ acting on K qubits with a single-qubit bitflip error rate p, a simple Kraus operator analysis shows that P_ℓ is modified to

$$\langle P_\ell \rangle_{bitflip} = (1-2p)^K \langle P_\ell \rangle_{true}\,, \tag{6.6}$$

which means that the noise channel will bias the estimator of the expectation value towards zero by a factor exponential in K. Thus, the determination of expectation values is highly

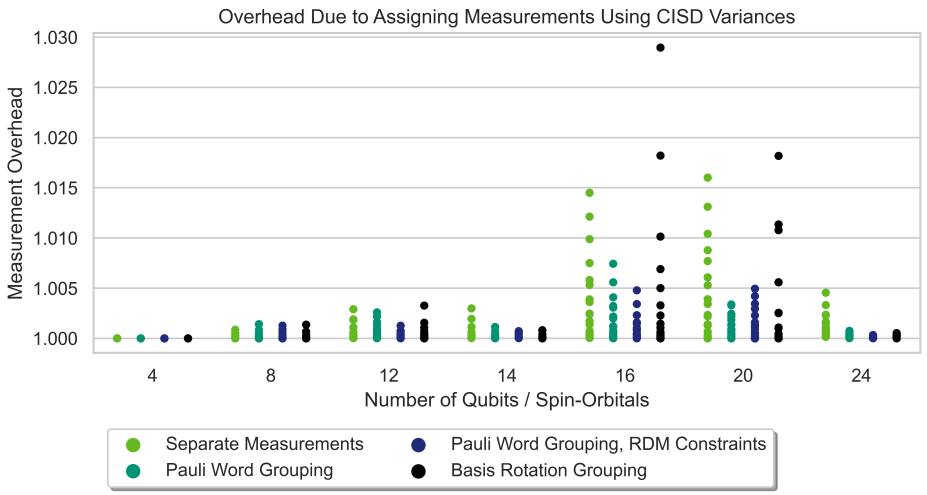

Figure 6.2: The increase in the time (or number of circuit repetitions) required to measure the ground state energy to a fixed precision when the measurements are distributed between groups using the variances calculated with the configuration interaction singles and doubles (CISD) approximation rather than the true ground state. For each of the systems and measurement techniques considered in this work, we present the ratio of the time required when using this approximate distribution of measurement repetitions compared with the time required using the optimal distribution, both calculated using Eq. 6.5 and then applied to the measurement of the actual ground state of the system. We find that using a classically tractable CISD calculation to determine the distribution of measurements between groups results in only a small increase in total measurement time.

sensitive to the extent of locality of the P_ℓ, a behavior that we expect to persist under more realistic models of readout errors.

One could also accomplish the reduction in the support of the operators that our method achieves by other means. For example, one could measure each of the $O(N^4)$ terms separately, localizing each one to a single qubit operator by applying $O(N)$ two-qubit gates. Other schemes have been proposed which simultaneously allow generic two-electron terms to be measured using $O(1)$ qubits each while simultaneously accomplishing the parallel measurement of $O(N)$ terms at a time, at the cost of using $O(N^2)$ or $O(N^2 log(N))$ two-qubit gates [172, 315, 332]. One unique advantage of our approach is that we achieve this reduction in operator support at the same time as the large reduction in the number of measurement repetitions presented in Sec. 6.3.2 above.

Our approach also enables a second form of error mitigation. Each measurement we prescribe is also simultaneously a measurement of the total particle number operator, η, and of the z component of spin, S_z. We can therefore reduce the impact of circuit and measurement errors by performing postselection conditioned on a desired combination of quantum numbers

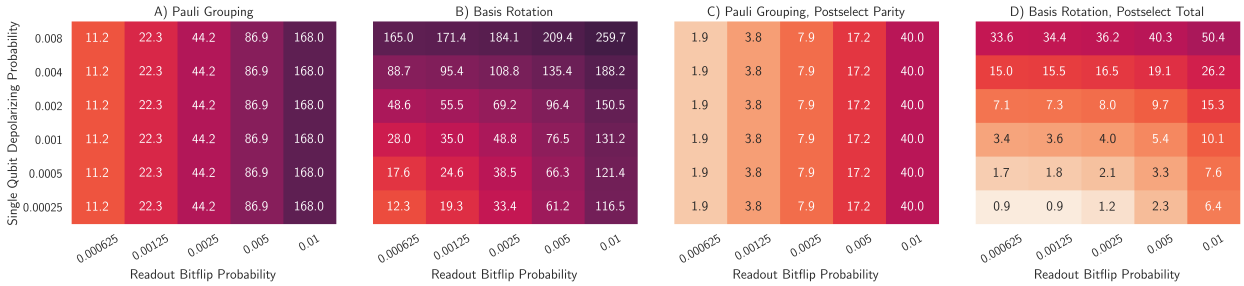

Figure 6.3: The absolute error in millihartrees of ground state measurements of a stretched chain of six Hydrogen atoms under an error model composed of single qubit dephasing noise applied after every two qubit gate together with a symmetric bitflip channel during readout. We consider single qubit depolarizing noise with probabilities ranging from 2.5×10^4 to 8×10^3, corresponding to two qubit gate error rates of $\approx 5 \times 10^4$ to $\approx 1.6 \times 10^2$. For the measurement noise, we take the single qubit bitflip error probabilities to be between 6.25×10^4 and 1×10^2. From left to right: A) The error incurred by a "Pauli Grouping" measurement strategy involving simultaneously measuring compatible Pauli words in the usual molecular orbital basis. B) The error when using our "Basis Rotation Grouping" scheme which performs a change of single-particle basis before measurement. C) The errors using the same Pauli word grouping strategy together with additional measurements and post-processing which effectively project the measured state onto a manifold with the correct parities of the total particle number and S_z operators. D) Those found when using our basis rotation strategy and postselecting on outcomes where the correct particle number and S_z were observed. In all panels we consider the measurement of the exact ground state without any error during state preparation.

for each of these operators. Let P denote the projector onto the corresponding subspace and let ρ denote the density matrix of our state. We obtain access to the projected expectation value,

$$\langle H \rangle_{proj} = \frac{\mathrm{Tr}\left(P\rho H\right)}{\mathrm{Tr}\left(P\rho\right)}, \tag{6.7}$$

directly from the experimental measurement record by discarding those data points which fall outside of the desired subspace. The remaining data points are used to evaluate the expectation values of the desired Pauli strings.

This postselection is efficient in the sense that it require no additional machinery beyond what we have already proposed. The only cost is a factor of $\approx 1/\mathrm{Tr}(P\rho)$ additional measurements. This factor is approximate because discarding measurements with the wrong particle number is likely to lead to a lower observed variance. Specifically, by removing measurements in the wrong particle number sector, we avoid having to average over large fluctuations caused by the energetic effects of adding or removing particles. This therefore

presents an additional route by which our Basis Rotation Grouping scheme will reduce the number of measurements in practice.

Several recent works have proposed error mitigation strategies which allow for the targeting of specific symmetry sectors. We make a brief comparative review of these here in order to place our work in context. One class of strategies focuses on non-destructively measuring one or more symmetry operators [213, 214]. After performing the measurements and conditioning on the desired eigenvalues, the post-measurement state becomes $P\rho P/\mathrm{Tr}(P\rho)$ and the usual Hamiltonian averaging can be performed. These approaches share some features with our strategy in that they also require an additional number of measurements that scale as $1/\mathrm{Tr}(P\rho)$ and an increased circuit depth. However, they also have some drawbacks that we avoid. Because they separate the measurement of the symmetry operator from the measurement of the Hamiltonian they require the implementation of relatively complicated non-destructive measurements. As a consequence, existing proposals focus on measuring only the parity of the η and S_z operators, leading to a strictly less powerful form of error mitigation than the approach we propose. Additionally, most errors that occur during or after the symmetry operator measurement are undetectable, including errors incurred during readout.

A different class of approaches avoids the need for additional circuit depth at the expense of requiring more measurements [150, 213, 328]. To understand this, let Π denote the fermionic parity operator and $P = (1 + \Pi)/2$ the projector onto the $+1$ parity subspace. Then,

$$\langle H \rangle_{proj} = \frac{\mathrm{Tr}\left(P\rho H\right)}{\mathrm{Tr}\left(P\rho\right)} = \frac{\mathrm{Tr}\left(\rho H\right) + \mathrm{Tr}\left(\rho\Pi H\right)}{1 + \mathrm{Tr}\left(\rho\Pi\right)}. \tag{6.8}$$

To construct the projected energy it then suffices to measure the expectation values of the Hamiltonian, the parity operator, and the product of the Hamiltonian and parity operators. A stochastic sampling scheme and a careful analysis of the cost of such an approach reveals that it is possible to use post-processing to estimate the projection onto the subspace with the correct particle number parity in each spin sector at a cost of roughly $1/\mathrm{Tr}(P_\uparrow P_\downarrow \rho)^2$ (where $P_\uparrow$ and $P_\downarrow$ are the parity projectors for the two spin sectors) [150]. Unlike our approach, this class of error mitigation techniques does not easily allow for the projection onto the correct eigenvalues of η and S_z, owing to the large number of terms required to construct these projection operators. Furthermore, the scaling in the number of additional measurements we described above, already more costly than our approach, is also too generous. This is because the product of the parity operators and the Hamiltonian will contain a larger number non-simultaneously measurable terms than the same Hamiltonian on its own. Maximum efficiency may require grouping schemes that consider this larger number of term groupings.

The most significant drawback of our method in the context of error mitigation is that the additional time and gates required for the basis transformation circuit lead to additional opportunities for errors. We believe the reduction in circuit repetitions we have shown makes our method the most attractive choice when it is feasible to use an additional $O(N^2)$ two-qubit gates during the measurement process. We therefore focus on comparing the performance of

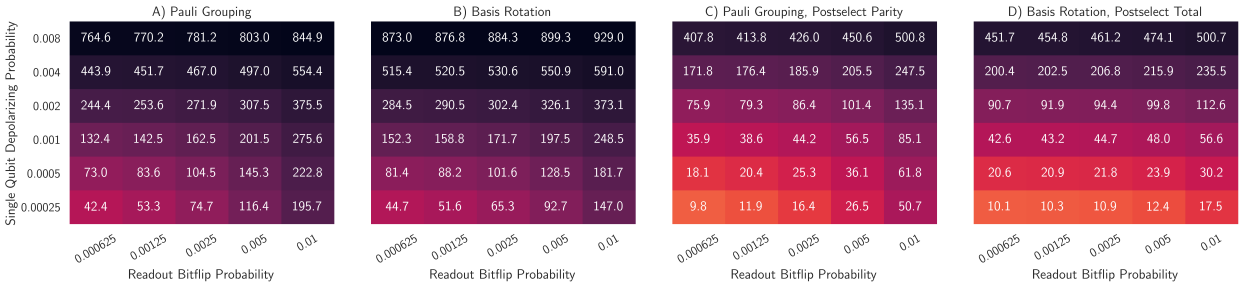

Figure 6.4: The absolute error in millihartrees of ground state measurements of a stretched chain of six Hydrogen atoms under an error model composed of single qubit dephasing noise applied after every two qubit gate together with a symmetric bitflip channel during readout. We consider single qubit depolarizing noise with probabilities ranging from 2.5×10^4 to 8×10^3, corresponding to two qubit gate error rates of $\approx 5 \times 10^4$ to $\approx 1.6 \times 10^2$. For the measurement noise, we take the single qubit bitflip error probabilities to be between 6.25×10^4 and 1×10^2. From left to right: A) The error incurred by a "Pauli Grouping" measurement strategy involving simultaneously measuring compatible Pauli words in the usual molecular orbital basis. B) The error when using our "Basis Rotation Grouping" scheme which performs a change of single-particle basis before measurement. C) The errors using the same Pauli word grouping strategy together with additional measurements and post-processing which effectively project the measured state onto a manifold with the correct parities of the total particle number and S_z operators. D) Those found when using our basis rotation strategy and postselecting on outcomes where the correct particle number and S_z were observed. In all panels, for the purpose of approximating a realistic ansatz circuit, three random Givens rotation networks which compose to the identity were simulated acting on the ground state prior to measurement.

our strategy with a strategy that requires no additional gates and uses a quantum subspace error mitigation approach that effectively projects onto the correct parity of the number operator on each spin sector [150, 213]. In order to do so, we use the open source software package Cirq [321] to simulate the performance of both strategies for measuring the ground state energy of a chain of six Hydrogen atoms symmetrically stretched to 1.3Å in an STO-3G basis. We take an error model consisting of i) applying a single qubit depolarizing channel with some probability to both qubits following each two qubit gate, and ii) applying a bitflip channel during the measurement process with some other probability. We report results for a wide range of gate and readout noise levels inspired by the capabilities of state of the art superconducting and ion trap quantum computers [10, 116, 121, 348]. Specfically, we consider single qubit depolarizing noise with probabilities ranging from 2.5×10^4 to 8×10^3 and single qubit bitflip error probabilities between 6.25×10^4 and 1×10^2. Here we do not consider the effect of a finite number of measurements and instead report the expectation values from the final density matrix.

Figure 6.3 shows the error in the measurement of the ground state energy for the error-mitigated Basis Rotation Grouping (far right panel) and Pauli Word Grouping (second panel from right) approaches together with the expectation values for both measurement strategies without error mitigation (two left panels). In these calculations we assumed that the ground state wavefunction under the Jordan-Wigner transformation is prepared without error. Circuit level noise is considered only during the execution of the Givens rotation required for our Basis Rotation Grouping approach. In order to include the impact of our proposed error mitigation strategy on state preparation as well as measurement, we have also carried out calculations including circuit noise during state preparation. The results of these calculations are presented in Figure 6.4. Here we have approximated a realistic state preparation circuit by applying three random basis rotations which compose to the identity to the ground state wavefunction. These state preparation circuits are simulated with the same gate noise as the measurement circuits. This choice is motivated by the assumption that low-depth circuits will be required for the successful application of VQE and the expectation that 90 two-qubit gates represents a reasonable lower bound to the size of circuit for a strongly-correlated problem on 12 qubits.

Figures 6.3 and 6.4 show that the Pauli Word Grouping and Basis Rotation Grouping approaches to measurement benefit significantly from their respective error mitigation strategies. Despite the fact that our proposed Basis Rotation Grouping technique requires 30 additional two-qubit gates compared to the Pauli Word Grouping approach, we see that the errors remaining after mitigation are comparable in some regimes and are lower for our strategy when noise during measurement is the dominant error channel (compare the bottom right corners of the two rightmost panels in both figures). Focusing first on Figure 6.3, we can see that this is true even when the errors during state preparation are not taken into account. Examining the left two panels of both figures, we can see that even without applying postselection,the locality of our Jordan-Wigner transformed operators leads to a considerable benefit in suppressing the impact of readout errors.

We note that the absolute errors we find when including noise during state preparation (Figure 6.4), even at the lowest noise levels considered here, are larger than the usual target of "chemical accuracy" ($\sim$ 1 mHa). In practice, an experimental implementation of VQE on non-trivial systems will require the combination of multiple forms of error mitigation. Prior work has shown that error mitigation by symmetry projection combines favorably with proposals to extrapolate expectation values to the zero noise limit [214]. We expect that such an extrapolation procedure could significantly improve the numbers we present here. Other avenues for potential improvements are also available. For example, one could rely on the error mitigation and efficiency provided by our measurement strategy during the outer loop optimization procedure, before utilizing a richer quantum subspace expansion in an attempt to reduce errors in the ground state energy after determining the optimal ansatz parameters.

6.4 Discussion

We have presented an improved strategy for measuring the expectation value of the quantum chemical Hamiltonian on near-term quantum computers. Our approach makes use of well studied factorizations of the two-electron integral tensor, in order to rewrite the Hamiltonian in a form which is especially convenient for measuring under the Jordan-Wigner transformation. By doing so, we obtain $O(N)$ distinct sets of terms which must be measured separately, instead of the $O(N^4)$ required by a naive counting of terms approach. Application to specific molecular systems show that in practice, we require a much smaller number of repetitions to measure the ground state energy to within a fixed accuracy target. For example, assuming an experimental repetition rate of 10 kHz (consistent with the capabilities of commercial superconducting qubit platforms), a commonly referenced bound based on the Hamiltonian coefficients suggests that approximately 55 days are required to estimate the ground state energy to of a symmetrically stretched chain of 6 Hydrogen atoms encoded as a wavefunction on 24 qubits to within chemical accuracy, while our approach requires only 44 minutes. Our proposed measurement approach also removes the susceptibility to readout error caused by long Jordan-Wigner strings and allows for postselection by simultaneously measuring the total particle number and S_z operators with each measurement shot.

The tensor factorization that we used to realize our measurement strategy is only one of a family of such factorizations. Future work might explore the use of different factorizations, or even tailor the choice of single particle bases for measurement to a particular system, by choosing them with some knowledge of the variances and covariances between terms in the Hamiltonian. As a more concrete direction for future work, the data we show in Section 6.5 regarding the difference between the bounds when calculated directly from the fermionic operators and the same approach applied to the Jordan-Wigner transformed operators, suggests that the cost estimates for error-corrected quantum algorithms should be recalculated using the qubit Hamiltonian.

For the largest systems we consider in this work, the 24 qubit Hydrogen chain and water simulations, and the 20 qubit Nitrogen calculations, our numerical results indicate that using our approach results in a speedup of more than an order of magnitude when compared to recent state-of-the-art measurement strategies. Furthermore, we observe a speedup of more than three orders of magnitude compared to the bounds commonly used to perform estimates in the literature. We also present strong evidence for an asymptotic improvement in our data on Hydrogen chains of various sizes. We performed detailed circuit simulations that show that reduction in readout errors combined with the error mitigation enabled by our work largely balances out the requirement for deeper circuits, even when compared against a moderately expensive error mitigation strategy based on the quantum subspace expansion [213]. We expect that the balance of reduced measurement time and efficient error mitigation provided by our approach will be useful in the application of variational quantum algorithms to more complex molecular systems.

Finally, we note that these techniques will generally be useful for quantum simulating any fermionic system, even those for which the tensor factorization cannot be truncated, such as

the Sachdev-Ye-Kitaev model of many-body chaotic dynamics [349, 350]. In that case, L will attain its maximal value of N^2, and our scheme will require $N^2 + 1$ partitions. Likewise, if the goal is to use the basis rotation grouping technique to estimate the fermionic 2-particle reduced density matrix rather than just the energy, one would need to measure in all $O(N^2)$ bases.

6.5 Variance Bounds

Earlier in this chapter, we reviewed the standard approach to upper bounding the number of measurements (M) required to measure the energy of a Hamiltonian (H) to within a desired target precision ϵ [18, 197]. This bound is calculated in a straightforward way from the qubit representation that is obtained by making the Jordan-Wigner transformation on a second quantized representation of the fermionic states, by expressing H as the sum of Pauli strings (products of single-qubit Pauli operators) P_ℓ acting on the qubit representation of the state. Then we have

$$M_q \leq \left(\frac{\sum_\ell |\omega_\ell|}{\epsilon} \right)^2, \quad \text{where} \quad H = \sum_\ell \omega_\ell P_\ell. \tag{6.9}$$

More generally, if the Hamiltonian is expressed as a linear combination $H = \sum_\ell \omega_\ell O_\ell$, one can work out the optimal way of distributing independent measurements between these terms and the overall number of measurements required for the resulting estimator to attain a target precision. We refer the reader to Ref. 197 for more details and simply recall the expression here,

$$M = \left(\frac{\sum_\ell |\omega_\ell| \, \sigma_\ell}{\epsilon} \right)^2. \tag{6.10}$$

The notation is the same as above, except that σ_ℓ is the positive square root of the variance of the operator O_ℓ. The upper bound of Eq. 6.9 is derived by noting that the variance of a Pauli operator measurement is at most one and by performing the appropriate substitutions. Our primary concern here is to show how the calculation of such a bound for fermionic Hamiltonians depends in a subtle manner on the representation of the Hamiltonian. We have denoted the bounds above M_q, to refer to their evaluation in the qubit representation of the Hamiltonian.

When calculating an upper bound of this type for a quantum chemical Hamiltonian,

$$H = \sum_{pq} h_{pq} a_p^\dagger a_q + \sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s, \tag{6.11}$$

it might seem natural to work directly with the coefficients h_{pq} and h_{pqrs} and the fermionic representation rather than performing the Jordan-Wigner transformation to the qubit representation. Provided that one is careful to count the coefficient for each term and its Hermitian conjugate only once, it is possible to obtain an upper bound to the number of measurements required directly from the coefficients h_{pq} and h_{pqrs}. Provided that the Hamiltonian is

normal-ordered first, this bound can be expressed as

$$M_f \leq \left(\frac{\sum_{p,q \leq p} |h_{pq}| + \sum_{pqr,s<q} |h_{pqrs}| + \sum_{pq,r \leq p,s=q} |h_{pqrs}|}{\epsilon} \right)^2 . \tag{6.12}$$

We denote this bound by M_f to indicate it was derived in the fermionic representation of the Hamiltonian. However, this bound is looser than necessary in multiple ways. In Table 6.4 we show a breakdown of the calculation of the sum of the absolute values of the coefficients for the Hamiltonian of a chain of eight equally spaced hydrogen atoms. We consider five different types of terms from Eq. 6.11 and calculate the sum of the absolute values of the coefficients for all terms of each type in both the qubit and fermionic representations. By comparing the two approaches in this way we show below that we can shed light on the difference in the resulting bounds.

In column I of Table 6.4 we begin with this analysis for the 'number operator' terms, $a_p^\dagger a_p$ of partition I. We see that the value reported in the qubit representation is exactly half of that reported in the fermionic one. This is because the Jordan-Wigner transformation applied to $a_p^\dagger a_p$ yields $\frac{1}{2}\mathcal{I} + \frac{1}{2}Z_p$. However, we may neglect the first contribution since this is a constant and so does not affect the variance. Another way of understanding this difference between the qubit and fermionic contributions is to realize that the bound of Eq. 6.9 is derived with the assumption that the maximum variance of each term is 1. The number operator, however, has eigenvalues 0 and 1 rather than -1 and 1 like a Pauli operator, and so its maximum variance is lower. We shall present an alternative to Eq. 6.12 below in Eq. 6.16 that accounts for this lower variance. The analysis of the other one-body term, $a_p^\dagger a_q + h.c.$, from partition II, is simpler. Column II shows that the part of the total magnitude of the coefficients of these terms is the same, regardless of which representation is used for the calculation.

The two-body terms in the Hamiltonian display more varied behavior. Taking a term from partition III and applying the Jordan-Wigner transformation results in exactly $\frac{1}{4}$ of the weight being assigned to a constant term, explaining the difference between the values for this partition. Because terms such as $a_p^\dagger a_q^\dagger a_p a_q$ can be rewritten as the product of two number operators, they must have eigenvalues 0 and 1 and thus, a maximum variance smaller than one. This improved bound would actually be lower than the one suggested by the analysis of the Jordan-Wigner transformed terms, because it would account for the covariances between the Z_p, Z_q, and Z_pZ_q terms that emerge. We shall incorporate this tighter bound on the variance of the individual terms in this class into the alternative to Eq. 6.12 presented below as Eq. 6.16.

The disparity for partitions IV and V has a different source. If one performs the Jordan-Wigner transformation of a term from either class individually, there is no difference between the total magnitudes of the coefficients for the fermionic operators and their qubit counterparts. However, when one sums all such terms together there is some cancellation between the qubit terms that reduces the overall total magnitude. Specifically, the terms in class V benefit from some cancellation due to the eight-fold symmetries of the two-electron integral tensor for real

Hamiltonian Partition	I	II	III	IV	V	Whole Hamiltonian
Example Term	$a_p^\dagger a_p$	$a_p^\dagger a_q + h.c.$	$a_p^\dagger a_q^\dagger a_p a_q$	$a_p^\dagger a_q^\dagger a_q a_r + h.c.$	$a_p^\dagger a_q^\dagger a_r a_s + h.c.$	-
Fermionic $\sum_\ell \lvert w_\ell \rvert$	32.288	2.852	34.214	7.436	35.579	112.368
Qubit $\sum_\ell \lvert w_\ell \rvert$	16.144	2.852	25.660	6.794	17.790	33.500

Table 6.4: Consider the normal ordered second quantized quantum chemistry Hamiltonian of Eq. 6.11, calculated for a chain of eight hydrogen atoms equally spaced 1.0Å apart in an STO-3G basis. We group the terms in this Hamiltonian into five partitions. Partitions I and II contain the one particle terms from the first summation. Partition I consists of those terms where $p = q$, while II consists of those where $p \neq q$. Partitions III, IV, and V contain the two particle terms from the second summation. Partition III consists of those where there are two unique values among p, q, r, and s, while IV consists of those with three unique values and V consists of those with four eigenvalues. For each partition, we report the sum of the absolute values of the coefficients of these terms in the fermionic representation of the Hamiltonian (counting the coefficient of a term and its Hermitian conjugate only once). We also report the same quantity calculated in the qubit representation after applying the Jordan-Wigner transformation. We drop any constant terms which appear as a result of the Jordan-Wigner transformation, since these do not contribute to the variance. Additionally, we report the sum of the absolute value of the coefficients for the entire Hamiltonian calculated in both ways in the final column.

orbitals [24],

$$h_{pqrs} = h_{qpsr} = h_{rspq} = h_{srqp} = h_{rqps} = h_{qrsp} = h_{psrq} = h_{spqr} = h. \tag{6.13}$$

We claim that the cancellations between symmetric terms result in a value for the sum of the magnitudes of the coefficients that is exactly half as large in the qubit representation as it is in the fermionic one. As an example, consider the case with $p = 4, q = 2, r = 3, s = 1$. After normal-ordering, the eight terms become four and we have

$$2ha_4^\dagger a_2^\dagger a_3 a_1 + h.c. + 2ha_3^\dagger a_2^\dagger a_4 a_1 + h.c., \tag{6.14}$$

where h is given by Eq. (6.13) and denotes the value of the coefficients before normal-ordering. The Jordan-Wigner transformation leads to terms from $a_4^\dagger a_2^\dagger a_3 a_1 + h.c.$ that cancel with terms from $a_3^\dagger a_2^\dagger a_4 a_1$ as shown below:

$$\begin{aligned}
&\tfrac{-h}{4}(X_1X_2X_3X_4 + X_1X_2Y_3Y_4 - X_1Y_2X_3Y_4 + X_1Y_2Y_3X_4 + Y_1X_2X_3Y_4 - Y_1X_2Y_3X_4 + Y_1Y_2X_3X_4 + Y_1Y_2Y_3Y_4) + \\
&\tfrac{-h}{4}(X_1X_2X_3X_4 + X_1X_2Y_3Y_4 + X_1Y_2X_3Y_4 - X_1Y_2Y_3X_4 - Y_1X_2X_3Y_4 + Y_1X_2Y_3X_4 + Y_1Y_2X_3X_4 + Y_1Y_2Y_3Y_4) = \\
&\tfrac{-h}{4}(X_1X_2X_3X_4 + X_1X_2Y_3Y_4 + Y_1Y_2X_3X_4 + Y_1Y_2Y_3Y_4 + X_1X_2X_3X_4 + X_1X_2Y_3Y_4 + Y_1Y_2X_3X_4 + Y_1Y_2Y_3Y_4).
\end{aligned} \tag{6.15}$$

It is straightforward, although tedious, to prove that the same cancellation occurs generically for terms in class V as a consequence of this eight-fold symmetry. As a result, for this class of terms the sum of the magnitudes of the coefficients is exactly half as large in the qubit representation as it is in the fermionic one. Analogous cancellations in the sum of the class IV terms do not show an obvious symmetry but they are also the source of the difference between the contributions from the two representations in partition IV.

Further cancellation is also apparent when one combines all five classes and calculates the sum of the absolute values of the coefficients (dropping the constant terms) for both representations of the Hamiltonian. The sum of magnitudes of the individual classes in the fermionic representation is the same as the magnitude of the sum. However, in the qubit representation, this value calculated for the entire Hamiltonian is roughly half the size of the sum of the individual partitions. One substantial reason for this difference is the fact that the terms in partitions I and III naturally give rise to terms proportional to products of single qubit Z operators having opposite signs. This is behavior that we should expect for any molecular Hamiltonian, where the single number operator terms arise from the Coulomb attraction between nuclei and individual electrons (negative sign), while the terms containing two number operators arise from the Coulomb repulsion between pairs of electrons (positive sign). Furthermore, unlike the tighter bounds achievable by accounting for the smaller variance of the terms in partitions I and III, this cancellation derives from the underlying form of the Hamiltonian and can not be accounted for in a straightforward way by using a better upper bound for the σ_ℓ values in Eq. 6.10 when deriving a fermionic bound like Eq. 6.12.

In the first three rows of Table 6.5, we now tabulate the bounds on the variance of the energy estimator, in units of 100 E_h^2, that arise from the sums of the absolute values of the coefficients. We perform these calculations for a chain of eight hydrogen atoms at various symmetric stretched interatomic spacings, including the 1.0Å distance explored in Table 6.4. In this table, 'Qubit Variance Bound' refers to the bound of Eq. 6.9 calculated using the qubit form of the Hamiltonian. 'Naive Fermionic Variance Bound' is calculated in a similar way, except using the sums of the fermionic coefficients, as in Eq. 6.12. As noted above, the terms in the fermionic Hamiltonian which consist of number operators (class I in Table 6.4) or products of number operators (class II in Table 6.4) actually have a variance which is upper-bounded by $\frac{1}{4}$ rather than one. One can substitute this tighter bound in Eq. 6.10 to yield the expression

$$M_f \leq \left(\frac{\frac{1}{2}\sum_p |h_p| + \sum_{p,q<p} |h_{pq}| + \sum_{pqr,s<q} |h_{pqrs}| + \sum_{pq,r<p,q} |h_{pqrq}| + \frac{1}{2}\sum_{pq} |h_{pqpq}|}{\epsilon} \right)^2, \tag{6.16}$$

where we have assumed that the Hamiltonian is normal-ordered to simplify the expression. We present calculation based on this improved bound in the row titled 'Fermionic Variance Bound.' However, it is clear that the bounds obtained directly from the qubit representation of the Hamiltonian are considerably tighter than either of the bounds obtained using the fermionic representation. This difference is explained by the cancellation effects that we have described above.

In addition to these bounds on the variance of the estimator for the Hamiltonian, we also consider two approximations to this variance that are not guaranteed to be upper bounds (rows 4 and 5 of Table 6.5). These approximations, which we refer to for brevity as FVA and QVA, are calculated using the methodology of Ref. 18 using the fermionic and qubit Hamiltonians respectively. In that work Wecker et al. reasoned that, in a typical quantum chemical calculation, the orbitals would have occupation numbers near 0 or 1. Therefore, the number and number-number terms in the Hamiltonian (partition I and partition III in Table 6.4) would have a variance that is close to zero. This assumption is satisfied exactly for the Hartree-Fock state when the appropriate single-particle basis is used, and should be approximately true when Hartree-Fock is qualitatively correct. Based on this assumption, Ref. 18 neglected these terms and then approximated the variance of the remaining terms in the Hamiltonian using the type of bounds we have already discussed.

Rows 4 and 5 of Table 6.5 show that that there is still a substantial difference between the variances calculated under this approximation using the two different representations of the Hamiltonian, i.e., between FVA and QVA. This is primarily due to the reduction caused by the cancellations among the double-excitation terms (class V in Table 6.4). Interestingly, the numbers presented for the 'Qubit Variance Approximation' (QVA) are nearly identical to those for the actual variance expected when measuring the Hartree-Fock state (row 6). In fact, any differences between these values are found to arise purely from numerical precision issues in the data. One can examine the Pauli operators that arise from performing the Jordan-Wigner transformation on the Hamiltonian after deleting the diagonal terms and

Interatomic Spacing (Angstrom)	.6	.7	.8	.9	1.0	1.1	1.2	1.3
Qubit Variance Bound	22.499	17.657	14.680	12.675	11.222	10.134	9.297	8.629
Fermionic Variance Bound	99.393	86.271	76.352	68.683	62.596	57.721	53.768	50.462
Naive Fermionic Variance Bound	205.251	177.997	156.704	139.847	126.267	115.234	106.168	98.545
Fermionic Variance Approximation	31.501	27.404	24.580	22.557	21.038	19.893	19.023	18.321
Qubit Variance Approximation	7.212	6.225	5.567	5.118	4.786	4.540	4.356	4.209
Hartree-Fock Variance	7.211	6.224	5.565	5.118	4.785	4.539	4.355	4.208
Ground State Variance	9.206	8.194	7.568	7.181	6.929	6.779	6.695	6.641

Table 6.5: Variances for a symmetrically stretched chain of 8 hydrogen atoms in an STO-3G basis. Rows 1-5 show values of the variance bounds and approximations to these that are described in the text. The variances are presented in units of 100 E_h^2. The bound in row 1 is calculated using Eq. 6.9, while row 2 uses Eq. 6.16, and row 3 uses Eq. 6.12. The approximations in rows 4-5 are calculated using the methodology of Ref. 18, which amounts to using Eq. 6.9 or Eq. 6.12 but neglecting some of the terms in the Hamiltonian as described in the text below. The last two rows, 6-7, present the actual variance of an estimator that measures each term in the Jordan-Wigner transformed Hamiltonian separately, for the Hartree-Fock state and for the ground state, respectively.

see that each of them has an expectation value of exactly zero on the Hartree-Fock state. Measurements of these terms therefore achieve the maximum possible variance of 1, while measurements of the deleted diagonal terms would have a variance that is exactly 0. Thus, the calculation of the actual variance (when measuring each term in the Hamiltonian separately) using Eq. 6.10 for the Hartree-Fock state then yields the same value as the calculation of the bound of Eq. 6.9 under the approximation proposed by Ref. 18.

6.6 Applying the fermionic RDM Constraints to the Qubit Hamiltonian

In the previous section we saw a substantial difference between the bounds calculated from the fermionic operators and those calculated from the qubit operators after applying the Jordan-Wigner transformation. In light of this, it is natural to ask how the reduced density matrix (RDM) approach of Ref. 197 might perform when formulated using the qubit representation of the Hamiltonian. In Ref, [197], Rubin et al. proposed that one could use known n-representability constraints on the expectation values of few-fermion operators, in order to construct estimators for the expectation value of the Hamiltonian that will have lower variance. They showed how one could take a collection of algebraic equalities from these fermionic n-representability constraints and use them to construct a new Hamiltonian $\tilde{H}$ from the original H. According to their approach, $\tilde{H}$ is constructed to have the same expectation value as H, but a lower maximum variance according to the bounds discussed above. They performed this minimization of the upper bound using standard linear programming techniques.

We are primarily focused here on the impact of these techniques for a real-world experiment. Therefore, we shall compare the impact of performing this minimization on the fermionic and qubit representations of the Hamiltonian, using the actual observed variance with respect to the ground state as the figure of merit, rather than employing the bounds or approximations discussed above. We take the same Pauli Word Grouping strategy described above and apply it to the Hamiltonians $\tilde{H}_{fermionic}$ and $\tilde{H}_{qubit}$. We define $\tilde{H}_{fermionic}$ and $\tilde{H}_{qubit}$ as the Hamiltonians that result from performing the upper bound minimization procedure of Ref. 197 in the fermionic and qubit representations respectively.

In Figure 6.5 we plot the ratio between the variances of $\tilde{H}_{qubit}$ and $\tilde{H}_{fermionic}$ for the ground state of each of the systems considered in this chapter. We list these systems in Table 6.2. Despite the substantial differences in the variance bounds formulated in the two representations, the impact of applying the RDM constraints to the qubit Hamiltonian rather than the fermionic one is found to be marginal, at best. For the majority of the systems it appears that the qubit-based bounds perform slightly better, but there are also a number of cases where this pattern is reversed.

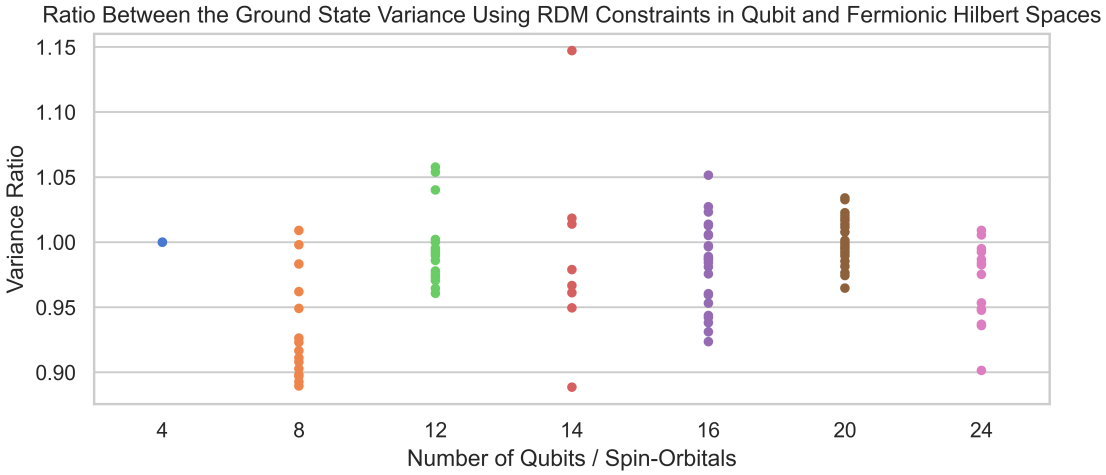

Figure 6.5: For each of the systems considered in this chapter we apply the techniques of Ref. 197 to the Hamiltonians in the fermionic and qubit Hibert spaces. We list these systems in Table 6.2. Using fermionic n-representability constraints, we construct the Hamiltonians $\tilde{H}_{fermionic}$ and $\tilde{H}_{qubit}$, that have the same expectation value but a lower maximum variance under bounds of the type described by Eq. 6.9 and Eq. 6.12. We then consider the variance of these Hamiltonians with respect to the ground state. We calculate these variances assuming measurement is performed using the Pauli Word Grouping strategy. Finally, we plot the ratio of the variance obtained for $\tilde{H}_{qubit}$ with the variance obtained for $\tilde{H}_{fermionic}$. The fact that all of these ratios are found to be near 1 shows that reformulating the work of Ref. 197 in the qubit representation does not offer a substantial improvement.

6.7 Low Rank Decomposition

In Section 6.3.1, we explained that our strategy for measurement is based on rewriting the standard quantum chemical Hamiltonian in the following form:

$$H = U_0 \left(\sum_p g_p n_p \right) U_0^\dagger + \sum_{\ell=1}^{L} U_\ell \left(\sum_{pq} g_{pq}^{(\ell)} n_p n_q \right) U_\ell^\dagger, \tag{6.17}$$

where the values g_p and $g_{pq}^{(\ell)}$ are scalars, $n_p = a_p^\dagger a_p$, and the U_ℓ are unitary operators which implement a single particle change of orbital basis. Here we shall explain how one obtains that factorization starting from a standard representation. We follow the presentation of Berry et al. with minor deviations and refer the reader to Ref. 77 for more details.

First, it is necessary to obtain the Hamiltonian in the chemist's standard form,

$$H = \sum_{\sigma \in \{\uparrow,\downarrow\}} \sum_{pq} T_{pq} a_{p,\sigma}^\dagger a_{q,\sigma} + \frac{1}{2} \sum_{\alpha,\beta \in \{\uparrow,\downarrow\}} \sum_{pqrs} V_{pqrs} a_{p,\alpha}^\dagger a_{q,\alpha} a_{r,\beta}^\dagger a_{s,\beta}. \tag{6.18}$$

This differs from the physicist's convention of Eq. 6.11, where the operators in the two-electron component of the Hamiltonian have both creation operators to the left of both annihilation operators. We assume the use of purely real spatial orbitals, and therefore the tensor V_{pqrs} inherits the eight-fold symmetry,

$$V_{pqrs} = V_{srqp} = V_{psqr} = V_{qprs} = V_{qpsr} = V_{rsqp} = V_{rspq} = V_{srpq}, \tag{6.19}$$

from the definition of the two-electron integrals [24].

Now we can perform the decomposition. We treat the tensor V as a matrix indexed by the collective indices pq and rs. We can eigendecompose this matrix to yield

$$V_{pqrs} = \sum_{\ell} w_\ell v_{pq}^{(\ell)} v_{rs}^{(\ell)} \tag{6.20}$$

In the above equation, w_ℓ are the eigenvalues of V, $v^{(\ell)}$ are the eigenvalues. We proceed by using this equality to rewrite the two-electron component of the Hamiltonian,

$$\frac{1}{2} \sum_{\alpha,\beta \in \{\uparrow,\downarrow\}} \sum_{pqrs} V_{pqrs} a_{p,\alpha}^\dagger a_{q,\alpha} a_{r,\beta}^\dagger a_{s,\beta} = \frac{1}{2} \sum_{\ell} w_\ell \Big(\sum_{\sigma \in \{\uparrow,\downarrow\}} \sum_{pq} v_{pq}^{(\ell)} a_{p,\sigma}^\dagger a_{q,\sigma} \Big)^2, \tag{6.21}$$

with $v_{pq}^{(\ell)}$ inheriting the symmetry between the p and q indices from V.

The final remaining step is to transform Eq. 6.21, as well as the one-electron component of the Hamiltonian by diagonalizing each of the one-body operators. It is straightforward to express each of the one-body operators as diagonal operators in a rotated single-particle basis. The appropriate change of basis matrices can be obtained from the eigenvalues of the coefficient tensors, T and the $g^{(\ell)}$s in our case. We can therefore express the Hamiltonian in the form of Eq. 6.17, dropping the spin indices for simplicity. The coefficients g_p come from rotating to a basis where T_{pq} is diagonal. The coefficients $g_{pq}^{(\ell)}$ likewise come from rotating to a series of bases where the tensors v^ℓ are diagonal between their p and q indices. The operators U_ℓ are the inverse of the operators that diagonalize the one-body operators $\sum_{\sigma \in \{\uparrow,\downarrow\}} \sum_{pq} T_{pq} a_{p,\sigma}^\dagger a_{q,\sigma}$ and $\sum_{\sigma \in \{\uparrow,\downarrow\}} \sum_{pq} v_{pq}^{(\ell)} a_{p,\sigma}^\dagger a_{q,\sigma}$. Note that the p and q indices of Eq. 6.17 represent new dummy indices and that the w_ℓ terms have been absorbed into $g_{pq}^{(\ell)}$, together with the contributions from the squares of the diagonalized v_{pq}^ℓ terms.

6.8 Additional Computational Details

In this section we shall provide some additional description of the calculations we presented throughout this chapter. We shall first step through the process of performing the variance calculations presented in Figure 6.1. We shall then briefly discuss the noisy circuit simulations that we performed to generate Figure 6.3 and Figure 6.4. Along the way, we shall also mention the specific functionality of the OpenFermion (Ref. 263) and Cirq (Ref. 321) software packages that we used.

The first step was to partition the Hamiltonian into a sum of terms, each of which could be measured directly,

$$H = \sum_i O_i. \tag{6.22}$$

For our new approach, this meant that we had to perform the decomposition described in Section 6.7. We used the OpenFermion utilities contained in `low_rank.py` to perform this decomposition. For the other approaches we considered, we began with the Jordan-Wigner transformed Hamiltonian and partitioned them as described near the beginning of Section 6.3.2. Having obtained the groupings, we then calculated the variance of each term using both the true ground state (which we denote by σ_i^2) and the CISD approximation to the ground state (which we denote by $\tilde{\sigma}_i^2$).

We gave an expression above (Eq. 6.5) for the fraction of time that each term should be measured in order to minimize the overall variance. Here we recall that expression but cast it in terms of the variance of each term,

$$f_i = \frac{\sqrt{\sigma_i^2}}{\sum_j \sqrt{\sigma_j^2}}. \tag{6.23}$$

If we performed M overall measurements of the Hamiltonian, then (ignoring the fact that fractional measurements don't make sense) we would perform $f_i M$ measurements of each term O_i. The overall variance of our energy estimator would therefore be

$$\sigma^2_{Hamiltonian} = \sum_i \frac{\sigma_i^2}{M f_i}. \tag{6.24}$$

In this chapter, we approximated the ratios f_i by using the variance with respect to the CISD state,

$$f_i \approx \frac{\sqrt{\tilde{\sigma}_i^2}}{\sum_j \sqrt{\tilde{\sigma}_j^2}}. \tag{6.25}$$

By using this approximation in Eq. 6.24, we obtained the data used to generate Figure 6.1. We recommend Ref. 197 for a thorough description of the mathematics, including a proof that Eq. 6.23 yields the optimal distribution of measurement effort.

The noisy circuit simulations are described already in Section 6.3.3, but we shall make some comments that might be of use to future practitioners. First of all, it is important to note that the code for generating the Cirq circuits for the basis rotation can be found in OpenFermion, in the file `optimal_givens_decomposition.py`. Cirq has a well-developed infrastructure for performing noisy simulations using a Kraus operator representation (see Section 3.3 for pedagogical background) contained in the `cirq.NoiseModel` abstraction. This allowed us to perform the noisy circuit simulations with just a few lines of additional code.

The only non-trivial component of our noisy circuit simulations was therefore the use of the two post-selection schemes. Because we were not focused on performing an experiment

or simulating the effects of a finite number of measurements, we were able to take a shortcut here. We explained in Section 6.3.3 how the ultimate effect of our error mitigation scheme, as well as the one we compared against, is to apply a projector to the final density matrix. Rather than simulating the measurements and post-selection procedure, we just performed this projection directly on the density matrices from the noisy simulations. Because these density matrices were already expressed in a basis where the relevant quantum numbers are diagonal, it was simple to remove the columns and rows corresponding to the wrong occupation numbers and then normalize the resulting projected density matrices.

Specifically, we determined the projected expectation value of Eq. 6.7 by applying the projectors to the density matrix. For the Pauli Word Grouping strategy, we used the projector P_{parity}, which removes the components of the state where the parity of the occupation number in either spin sector was incorrect. For our Basis Rotation Grouping strategy, we used the projector P_{total}, which removed the components of the state where the occupation number in either spin sector was incorrect. For example, assume that the spin-orbitals were $1_\alpha, 2_\alpha, \cdots 6_\alpha, 1_\beta, \cdots 6_\beta$ for our H_6 example. Then the basis state $|111000111100\rangle$ would be annihilated by both projectors, while $|111000111110\rangle$ would be annihilated only by P_{total}.

Chapter 7

Monte Carlo Approaches to the Tensor Renormalization Group

7.1 Preface

Except for Section 7.2, which was written specifically for this thesis, this chapter is taken from the previously released Ref. 351, currently available on the arXiv. It was co-authored by the author of this thesis, Dr. C. Daniel Freeman, Dr. Miles Stoudenmire, Dr. Norm Tubman, and Professor K. Birgitta Whaley. The author of this thesis wrote the majority of the manuscript and most of the code used to perform the numerical simulations. Dr. Freeman and Dr. Stoudenmire provided invaluable guidance on the tensor network formalism while Dr. Tubman did the same for the Monte Carlo aspects of the work. All of the co-authors contributed to the revision of the manuscript and provided necessary encouragement for the completion of the author of this thesis' first academic paper.

7.2 Tensor Network Background Material

This section will provide a brief introduction to the idea of tensor networks and the associated diagrammatic notation. Before we formally define a tensor network, we shall need to introduce some preliminary notions. Colloquially, though, we can say that tensor networks are tools for representing and manipulating tensors that would otherwise be unwieldy to treat directly. Many objects, including quantum mechanical wavefunctions, as well as the partition functions of both classical and quantum systems, have natural descriptions in terms of tensors. We shall provide a few simple examples throughout this section in order to highlight the utility of tensor networks and make our abstract definitions more concrete.

A tensor, for our purposes, is simply a multi-dimensional array of numbers. We do not need the more sophisticated algebraic or geometric definitions of a tensor, but we refer the interested reader to Ref. 352 for a deeper discussion. We shall refer to an N-dimensional array as a rank-N tensor. A vector is an example of a rank-1 tensor, and a matrix is an

example of a rank-2 tensor. We can often represent objects of interest using tensors in various ways. For example, a wavefunction on N two-level systems can be thought of as a rank-1 tensor if we flatten the indices into one collective index, or a rank-N tensor if we leave a separate index for each subsystem. To be concrete, let us specify a basis,

$$\{|i_1 i_2 i_3, \ldots, i_N\rangle\}. \tag{7.1}$$

We can represent a wavefunction in terms of its coefficients in this basis,

$$|\psi\rangle = \sum_{i_1,i_2,i_3,\ldots i_N} c_{i_1,i_2,i_3,\ldots i_N} \, |i_1, i_2, i_3, \ldots i_N\rangle \, . \tag{7.2}$$

In the above equation, the coefficients are elements of the rank-N tensor c, but we could also flatten the tensor c into a rank-1 tensor v, whose single index takes values in the range $\{1, 2, \ldots 2^N\}$. For example, we could set define the vector v through the following equation,

$$v_k = c_{i_1,i_2,i_3,\ldots i_N}, \qquad k = \sum_{j=1}^{N} i_j \cdot 2^j. \tag{7.3}$$

Given two tensors that have one or more pairs of indices of the same dimension, we can define a contraction over those indices, e.g.,

$$a_i = \sum_j M_{ij} v_j \quad \text{or} \quad \phi_{abc} = \sum_{ijk} U_{abcijk} \psi_{ijk}. \tag{7.4}$$

Formally, we could define contraction as a two-step process where we first take the tensor product between two tensors and then perform a joint summation over one or more pairs of indices [353]. These indices must have the same dimension in order for the summation to make sense and we typically focus on contraction over a pair of matching indices, rather than three or more. Less formally, we can point to a few familiar examples of tensor contraction, matrix multiplication, the dot product, the action of an operator on a wavefunction, taking the trace of a matrix, the equations of Eq. 7.4, and lean on our experience with these familiar operations.

We define a tensor network as a collection of tensors, together with a pattern for contracting over some or all of their indices. The collection of tensors,

$$\{A^{(1)}_{i_1,v_1}, A^{(2)}_{i_2,v_1,v_2}, A^{(3)}_{i_3,v_2,v_3}, A^{(4)}_{i_4,v_3,v_4}, A^{(5)}_{i_5,v_4}\}, \tag{7.5}$$

together with the understanding that we shall contract over repeated indices, is one example of a tensor network. With the contraction made explicit, we have

$$\sum_{v_1,v_2,v_3,v_4} A^{(1)}_{i_1,v_1}, A^{(2)}_{i_2,v_1,v_2}, A^{(3)}_{i_3,v_2,v_3}, A^{(4)}_{i_4,v_3,v_4}, A^{(5)}_{i_5,v_4}. \tag{7.6}$$

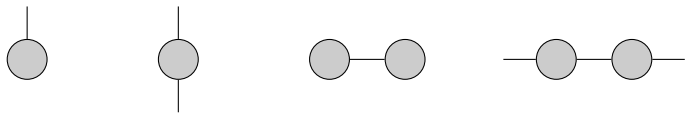

Figure 7.1: A few simple examples of tensor network diagrams. From left to right, we have: a vector, a matrix, the dot product between two vectors, the product of two matrices.

This tensor network might represent the wavefunction of a quantum mechanical system on 5 subsystems by encoding the coefficient tensor in a particular basis,

$$|\psi\rangle = \sum_{i_1,i_2,i_3,i_4,i_5} \left(\sum_{v_1,v_2,v_3,v_4} A^{(1)}_{i_1,v_1}, A^{(2)}_{i_2,v_1,v_2}, A^{(3)}_{i_3,v_2,v_3}, A^{(4)}_{i_4,v_3,v_4}, A^{(5)}_{i_5,v_4} \right) |i_1 i_2 i_3 i_4 i_5\rangle \,. \tag{7.7}$$

The order in which the indices are contracted does not change the result of a contraction, but it can greatly affect the computational cost.

Tensor network notation is a convenient graphical language for expressing the pattern of tensors and their contractions. Modern tensor network notation is a simplified version of a graphical notation introduced in Ref. 354 by Roger Penrose. We refer readers to Ref. 353 for a more comprehensive introduction and simply summarize the basic rules here. Essentially, tensor network diagrams represent rank-N tensors as shapes with N legs, one for each index. A contraction over two indices is indicated by joining the legs from two different shapes. A tensor product between tensors is implicit in their inclusion in the same diagram. For example, if the two tensors from the left-hand side of Figure 7.1 were included in a single diagram, they would represent the separable rank-3 tensor, $T = A \otimes B$, where $T_{ijk} = A_i B_{jk}$. Note that the diagrams do not convey information about the entries of the tensors or the dimension of the indices (although the latter is sometimes added to the diagrams as an annotation). We present a few examples in Figure 7.1.

Tensor networks can be designed to approximately represent a particular family of tensors or computations. For example, the class of tensor networks known as matrix product states parameterize a variational manifold useful for describing the low-energy states of one-dimensional Hamiltonians. Consider the tensor network described in Eq. 7.5, Eq. 7.6, and Eq. 7.7, represented graphically by the diagram in Figure 7.2. Standard terminology

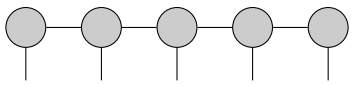

Figure 7.2: The tensor network diagram corresponding to the matrix product state tensor network of Eq. 7.5, Eq. 7.6, and Eq. 7.7. This tensor network represents a wavefunction over fives sites. The five indices corresponding to these sites are the dangling legs of each of the five grey circles.

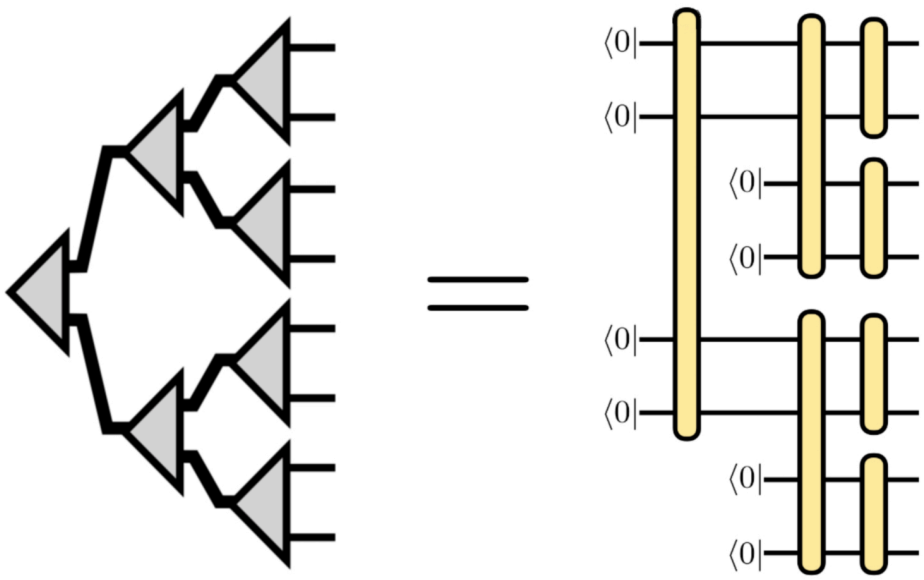

Figure 7.3: A tensor network representation of a wavefunction on eight qubits alongside a representation of the same wavefunction as a quantum circuit. Time goes from left to right in the right-hand diagram. On the left, we visually distinguish between indices with dimension four and dimension two by using bolder lines for the indices with dimension four.

labels the contracted indices ($\{v_1, v_2, \cdots v_4\}$) as virtual indices, and the remaining indices ($\{i_1, i_2, \cdots i_5\}$) as physical indices. For simplicity, we take virtual indices to all have the same dimension, which we refer to as the bond dimension. This tensor network, composed of a one-dimensional chain of tensors, each possessing one physical index and linked by virtual indices, is an example of a matrix product state. In Ref. 355, Hastings showed that the ground state of a gapped, local, one-dimensional quantum system could be well-approximated by a matrix product state with a fixed bond dimension. The story is more complicated for Hamiltonians of higher-dimensional systems [356], but tensor networks are routinely applied to study two-dimensional lattice systems and strongly correlated electronic Hamiltonians with great success as well [69, 357].

As another example, we present a tree tensor network representation of a wavefunction on eight qubits alongside a hybrid quantum circuit diagram / tensor network diagram for the same wavefunction in Figure 7.3. In fact, all quantum circuits are examples of tensor networks. In Chapter 8, we shall make us of this connection by drawing on intuition from the tensor network literature to design parameterized quantum circuits. Note that the requirement that the two and four qubit gates on the right-hand side of Figure 7.3 are unitary imposes a restriction on the tensors in the left-hand diagram if we wish to put them in one-to-one correspondence. Following Ref. 358, which first explored the connection between quantum

circuits and tensor networks, we refer to tensors that obey these restrictions as isometries and use triangular shapes to indicate them in Figure 7.3.

The calculation of classical partition functions can naturally be organized as a tensor network. For example, consider the partition function of the $2D$ Ising model,

$$Z = \sum_{\{\sigma\}} exp\Big(- \beta E(\{\sigma\})\Big) = \sum_{\{\sigma\}} exp\Big(- \beta J \sum_{<ij>} \sigma_i \sigma_j\Big), \tag{7.8}$$

where β is the inverse temperature, E is the energy, and the notation $\sum_{<ij>}$ indicates a sum over all i and j such that the spins σ_i and σ_j are adjacent in a two-dimensional lattice. By defining tensors of the form

$$A_{\sigma_1,\sigma_2,\sigma_3,\sigma_4} = exp\Big(- \beta J\big(\sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_3\sigma_4 + \sigma_4\sigma_1\big)\Big), \tag{7.9}$$

we can rewrite this partition function as a contraction over a collection of A tensors, as we show in Figure 7.2. In this figure, each spin corresponds to one of the indices (legs) in the tensor network diagram. Each of the A tensors encodes the contribution to the partition function from the interactions between the spins corresponding to its indices. Because of this unintuitive relationship between the tensor network diagram and the 2d spin model itself, we indicate the spins and their connectivity in the physical lattice explicitly in Figure 7.2. We recommend Ref. 359 as a reference for the interested reader, as well as Ref. 360.

7.3 Introduction

In the decades since the invention of the density matrix renormalization group [67] (DMRG) algorithm for determining the ground state of quantum systems, great strides have been made in understanding and generalizing its success. These developments includes many applications to 1D and 2D problems as well as small molecular chemistry Hamiltonians [68, 69, 229, 361–364]. The realization that DMRG could be seen as an efficient algorithm for variationally optimizing over a one dimensional matrix product state ansatz [365] led to the development of new tensor network wave functions [358, 366–369]. Other efforts (which we might call "calculational methods") have eschewed the variational approach entirely, instead directly representing the partition functions of classical or quantum systems as tensor networks [360, 370, 371].

Methods in this second class, as well as variational calculations with the projected entangled pair states (PEPS) ansatz, all face a common problem. Namely, that they involve tensor network contractions whose cost naively scales exponentially with system size. Fortunately, a host of algorithms, such as the tensor renormalization group [360] (TRG), tensor network renormalization [371], and the corner transfer matrix renormalization group [372], have been developed to overcome this exponential cost with well-controlled approximations. Despite these promising results, severe challenges still hamper the program of creating tensor network methods that match DMRG's power beyond the domain of one

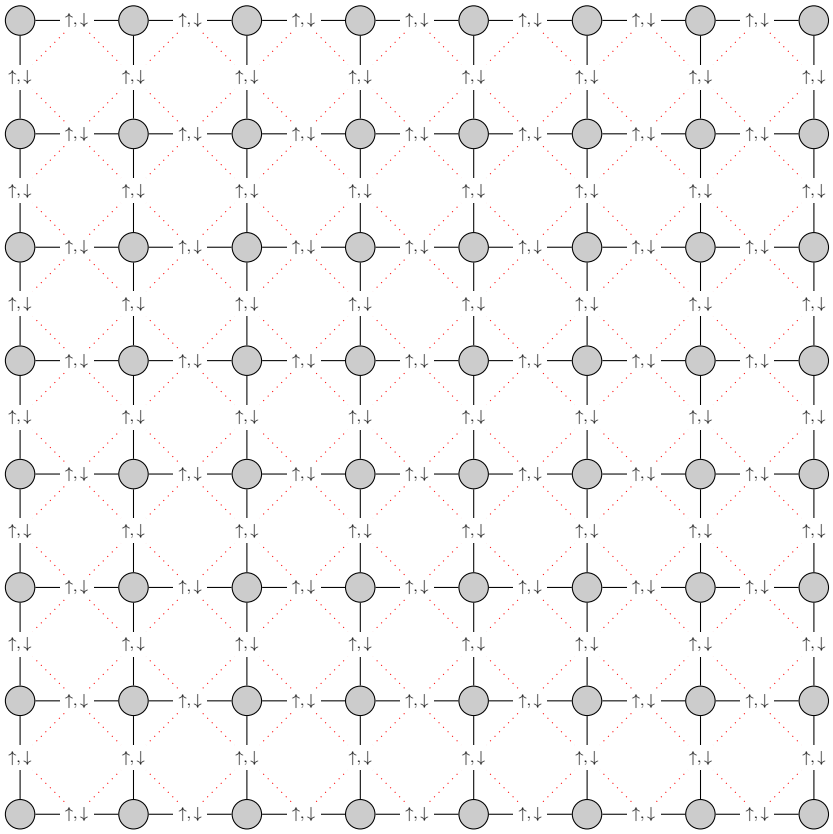

Figure 7.4: A tensor network diagram for the partition function a classical $2D$ Ising model. We have augmented the usual tensor network notation by adding the symbol $\uparrow, \downarrow$ over the legs of the tensors to indicate that each index represents a classical spin. We show the connectivity of the Ising model using dotted red lines that connect the symbols denoting the spins. These dotted red lines should not be confused with the legs of the tensor network, which are represented with solid black lines.

dimensional quantum systems [229, 373, 374]. One route forward it to make use of modern supercomputers by parallelizing the basic operations of tensor network calculations across hundreds or thousands of compute nodes, however, this promises to be a difficult feat to requiring an effort hand-tailored to each algorithm [375, 376]. In contrast, many Monte Carlo techniques used to simulate quantum systems use sampling techniques with which straightforward scaling to large numbers of nodes can be achieved easily [65, 377–380]. This observation motivates our exploration of stochastic tensor network contraction techniques.

The use of the singular value decomposition (SVD) to generate a low-rank tensor approximation is a key component of many tensor network algorithms. We consider a tensor T of order k: $T_{i_1,i_2...i_k}$ and divide the indices of T into two sets, $i_1, i_2...i_j$, and $i_{j+1}, i_{j+2}...i_k$. T can be treated as a matrix $T_{m,n}$ indexed by a tuple $m = (i_1, i_2, ..., i_j)$ of elements from the first set and a tuple $n = (i_{j+1}, i_{j+2}, ..., i_k)$ of elements from the second set. The best rank $\tilde{\chi}$ approximation (in the sense of minimizing the Frobenius norm of the difference between the exact T and the approximation $\tilde{T}$) is given by discarding all but the $\tilde{\chi}$ largest singular values, yielding:

$$T_{mn} \approx \tilde{T}_{mn} = \sum_{i}^{\tilde{\chi}} U_{mi} S_{ii} V_{in} \tag{7.10}$$

It was recently shown that one can sample from an ensemble of rank $\tilde{\chi}$ approximations of T (for any $\tilde{\chi} < \chi$) and exactly recover T as the average of an infinite number of samples [381]. Allow $\mathcal{E}$ to denote an ensemble of samples, with a particular element $\mathbf{e}$ of this ensemble defined by a subset of size $\tilde{\chi}$ of the nonzero singular values,

$$\mathbf{e} = \{s_1, ..., s_{\tilde{\chi}}\},\ s_i \in \{S_{11}, ..., S_{\chi\chi}\}. \tag{7.11}$$

Let the matrix $S^{(\mathbf{e})}$ be a diagonal matrix with the same shape as S but having $\tilde{\chi}$ nonzero entries, determined from $\mathbf{e}$ in a way which we will specify later. Then we define $\tilde{T}^{(\mathbf{e})} = US^{(\mathbf{e})}V$, a rank $\tilde{\chi}$ matrix which could also be equivalently but more compactly expressed terms of submatrices of $S^{(\mathbf{e})}$, U, and V.

In this approach we demand that the collection of rank $\tilde{\chi}$ matrices $\tilde{T}_j^{(\mathbf{e})}$ satisfy:

$$\lim_{N \Rightarrow \infty} \frac{1}{N} \sum_{j}^{N} \tilde{T}_j^{(\mathbf{e})} = T \tag{7.12}$$

By substituting the definitions of T and the $\tilde{T}^{(\mathbf{e})}$ into Eq. 7.12 we see that U and V can be canceled, yielding

$$\lim_{N \Rightarrow \infty} \frac{1}{N} \sum_{j}^{N} S_j^{(\mathbf{e})} = S. \tag{7.13}$$

Or, in other words, the matrices $S_j^{(\mathbf{e})}$ must average to S. Understanding this, we can formulate a constructive procedure for generating the matrices $S^{(\mathbf{e})}$. We begin with the original S and

randomly select a subset of the singular values $\mathbf{e}$ (with probability $p(\mathbf{e})$) to keep, setting the rest to zero. In order to satisfy Eq. 7.13 we then rescale the retained singular values. We do this by multiplying each of them by the inverse of the probability of including that particular value in an individual sample: $\frac{1}{r(S_{ii})}$. We note that this works for a general set of inclusion rates $r(S_{ii})$, determined from the subset selection probabilities $p(\mathbf{e})$ by the following expression,

$$r(S_{ii}) = \sum_{\mathbf{e}:\ S_{ii}\in\mathbf{e}} p(\mathbf{e}). \tag{7.14}$$

After the reweighting, any choice of scheme for selecting the subsets that has a finite probability of including each nonzero singular value S_{ii} will cause Eq. 7.13, and hence Eq. 7.12, to be satisfied and therefore lead to a valid ensemble. In order to minimize the expectation value of the error, $||T-\tilde{T}^{(e)}||^2$, we follow Ref. [381] and set the relative probability of each sample $\mathbf{e}$ to

$$w(\mathbf{e}) = \prod_{j}^{\tilde{\chi}} (s_j)^2 \tag{7.15}$$

and normalize these weights to form the probability distribution

$$p(\mathbf{e}) = \frac{w(\mathbf{e})}{\sum_{\mathbf{n}} w(\mathbf{n})}. \tag{7.16}$$

We refer the reader to Ref. [381] for details on a method to efficiently sample from this distribution and determine the probabilities of selection. With this sampling scheme in hand, we turn to its application as a component of an algorithm for tensor network renormalization.

Let $\mathcal{N}_I$ be a tensor network, composed of tensors $T_0, T_1, ..., T_k$, whose (tensor) trace represents some quantity of interest. We will specifically consider networks whose geometry is amenable to contraction with the tensor renormalization group (TRG) algorithm [360], an iterative procedure for approximately contracting a tensor network. However, as long as the truncation steps can be reduced to an application of the singular value decomposition, the following applies generally to other schemes for contraction by renormalization. We will consider here the case where all of the indices of $\mathcal{N}_I$ are summed over but the same arguments hold also when calculating a partial trace. Note that we use Roman numeral subscripts to refer to different levels of coarse-graining.

For each tensor $T_i \in \mathcal{N}_I$ which we must approximate by a truncated singular value decomposition, we can define an appropriate ensemble $\mathcal{E}_i$ as above such that $T_i = \langle \tilde{T}_i^{(\mathbf{e}_i)} \rangle_{\mathcal{E}_i}$ to yield the following equality,

$$\begin{aligned} Tr(\mathcal{N}_I) = Tr(T_1 T_2 ... T_k) = \\ Tr(\langle \tilde{T}_1^{(\mathbf{e}_1)} \rangle_{\mathcal{E}_1} \langle \tilde{T}_2^{(\mathbf{e}_2)} \rangle_{\mathcal{E}_2} ... \langle \tilde{T}_k^{(\mathbf{e}_k)} \rangle_{\mathcal{E}_k}). \end{aligned} \tag{7.17}$$

Allow the symbol $\mathcal{C}_I$ to denote the Cartesian product of the ensembles $\mathcal{E}_1$ through $\mathcal{E}_k$. Then, by linearity, and by the independence of the samples, we have

$$\begin{aligned} Tr(\langle \tilde{T}_1^{(\mathbf{e}_1)} \rangle_{\mathcal{E}_1} \langle \tilde{T}_2^{(\mathbf{e}_2)} \rangle_{\mathcal{E}_2} ... \langle \tilde{T}_k^{(\mathbf{e}_k)} \rangle_{\mathcal{E}_k}) = \\ \langle Tr(\tilde{T}_1^{(\mathbf{e}_1)} \tilde{T}_2^{(\mathbf{e}_2)} ... \tilde{T}_k^{(\mathbf{e}_k)}) \rangle_{\mathcal{E}_1 \mathcal{E}_2 ... \mathcal{E}_k} = \\ \langle Tr(\tilde{\mathcal{N}}_{II}^{(\mathbf{c}_I)}) \rangle_{\mathcal{C}_I} = \sum_{\mathbf{c}_I \in \mathcal{C}_I} p_{\mathbf{c}_I} Tr(\tilde{\mathcal{N}}_{II}^{(\mathbf{c}_I)}), \end{aligned} \tag{7.18}$$

where each $\tilde{\mathcal{N}}_{II}^{(\mathbf{c}_I)}$ is the coarse-grained tensor network associated with a particular set of samples $\mathbf{c}_I \in \mathcal{C}_I$ and the application of a single TRG step, and $p_{\mathbf{c}_I}$ denotes the probability of choosing the collection of samples $\mathbf{c}_I$. We continue recursively, being careful to note that both the coarse grained tensor networks and the ensembles that allow us to coarse grain them again in an unbiased fashion depend upon our choice of $\mathbf{c}_I$,

$$\begin{aligned} Tr(\tilde{\mathcal{N}}_{II}^{(\mathbf{c}_I)}) = \langle Tr(\tilde{\mathcal{N}}_{III}^{(\mathbf{c}_{II})}) \rangle_{\mathcal{C}_{II|\mathbf{c}_I}} = \\ \sum_{\mathbf{c}_{II} \in \mathcal{C}_{II|\mathbf{c}_I}} p_{\mathbf{c}_{II}} Tr(\tilde{\mathcal{N}}_{III}^{(\mathbf{c}_{II})}). \end{aligned} \tag{7.19}$$

Together then, we find that

$$\begin{aligned} Tr(\mathcal{N}_I) = \sum_{\mathbf{c}_I \in \mathcal{C}_I} p_{\mathbf{c}_I} \sum_{\mathbf{c}_{II} \in \mathcal{C}_{II|\mathbf{c}_I}} p_{\mathbf{c}_{II}} Tr(\tilde{\mathcal{N}}_{III}^{(\mathbf{c}_{II})}) = \\ \sum_{\mathbf{c}_I \in \mathcal{C}_I} \sum_{\mathbf{c}_{II} \in \mathcal{C}_{II|\mathbf{c}_I}} ... \sum_{\mathbf{c}_m \in \mathcal{C}_{m|\mathbf{c}_I \mathbf{c}_{II} ... \mathbf{c}_{m-1}}} p_{\mathbf{c}_I} p_{\mathbf{c}_{II}} ... p_{\mathbf{c}_m} Tr(\tilde{\mathcal{N}}_m^{(\mathbf{c}_m)}), \end{aligned} \tag{7.20}$$

where the $\tilde{\mathcal{N}}_m^{(\mathbf{c}_m)}$, because they contain only a small number of tensors whose bond dimensions have been controlled by the TRG truncations steps, are sufficiently simple that their trace can be computed explicitly.

We will approximate the sum from Eq. 7.20 by a Monte Carlo sampling. Beginning with $\mathcal{N}_I$, we perform the full singular value decompositions as usual and then choose a subset of singular values to keep according to the proscription described earlier. Each decomposition is sampled independently, and by completing the coarse graining step as in Ref. [360] we generate a coarse grained tensor network $\tilde{\mathcal{N}}_{II}^{(\mathbf{c}_I)}$ with the appropriate probability $p_{\mathbf{c}_I}$. By repeating the same stochastic coarse-graining steps several times we can efficiently sample from the distribution described by Eq. 7.20 and we find that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{i=N} Tr(\tilde{\mathcal{N}}_m^{(i)}) = Tr(\mathcal{N}_I). \tag{7.21}$$

We note that it is essential that the samples of different tensors be generated independently, regardless of the symmetries of the physical model. Therefore, the computational time required

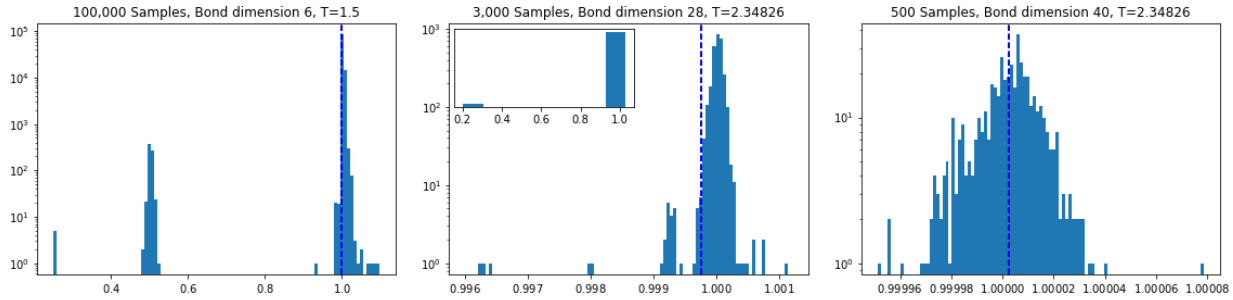

Figure 7.5: The distribution of results from our stochastic TRG calculations of the partition function at representative points. The dashed line represents the average over all samples and the x axis is scaled so that the exact value of the partition function is one. All data points are shown except in the middle plot, where there is a single sample only visible in the zoomed-out inset.

by this approach scales linearly with the number of tensors in the original network. This growth is comparable to the situation for the deterministic algorithm when the underlying tensor network is not translationally invariant and each tensor must be decomposed separately. It does, however, represent a substantial overhead versus the logarithmic scaling of the non-stochastic approach applied to the case where the underlying system is translationally invariant.

7.4 Numerical Results

To benchmark our algorithm we present calculations of the partition function of the 2d classical Ising model at zero field both near and far from the critical temperature. All calculations are performed for 128 spins on a periodic lattice using the ITensor library [1] and compared to an exact summation of the partition function carried out to machine precision. Six full renormalization steps are performed on the 8x8 tensor network, each reducing its by a factor of two, before the single remaining tensor is traced over.

We consider the behavior of our approach over a range of bond dimensions and sample sizes and emphasize that, in contrast with the deterministic application of TRG, the algorithm discussed above is completely unbiased. Regardless of the bond dimension, the estimate of the partition function and the error bars generated for that estimate have no systematic errors and the individual samples are totally uncorrelated. Therefore, we are focused on understanding its statistical efficiency in different regimes.

We can gain a qualitative insight into the efficiency of the algorithm by examining Fig. 7.5, where at small bond dimensions we observe multimodal distributions with one dominant

[1] Calculations performed using the ITensor C++ library (version 2.0.7), http://itensor.org/

mode and a series of much smaller peaks. This multimodal character implies that a relatively large number of samples would be required to characterize such a distribution and is the source of the large variances that we see at small bond dimensions. This is seen clearly in the leftmost plot in Fig. 7.5, which has more than ninety-nine percent of its probability mass concentrated in the largest mode but also has a non-trivial weight present in a smaller satellite peak.

Bond Dimension	Temperature	Standard Deviation
6	1.5	4.171e-02
28	2.34826	5.017e-04
40	2.34826	1.353e-05

Table 7.1: The per-sample standard deviation at the three temperature/bond dimension pairs highlighted in Fig. 7.5. The true value of the partition function has been normalized to one. The tightly peaked and unimodal distributions seen at higher bond dimension lead to better individual samples.

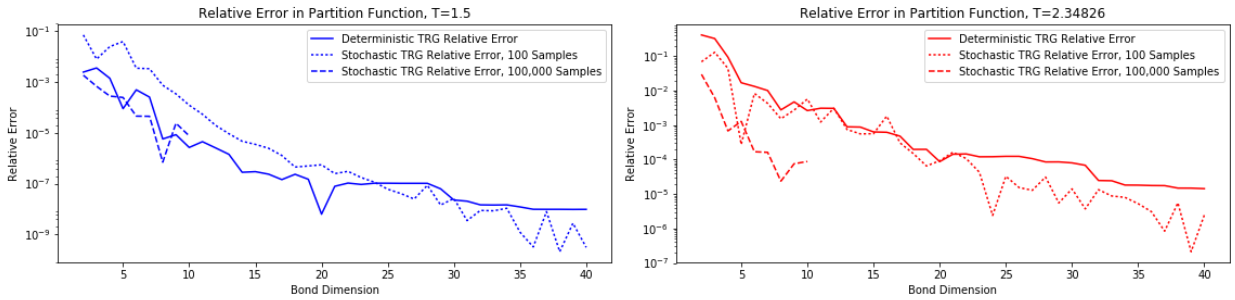

Figure 7.6: The relative error in the partition function for the deterministic version of TRG alongside the relative error for calculations performed with 100 and 100,000 samples using our stochastic TRG plotted at T=1.5 (left) and T=2.34 (right) over various bond dimensions. We see that the stochastic calculations performed with different numbers of samples follow roughly the same path, except that the curves with more samples are translated downward on a logarithmic scale.

The effect that the peculiarities of the underlying distributions have on the quality of the results has been studied extensively in the context of statistical mechanics and in the quantum Monte Carlo community [382, 383]. In some situations one must take care to ensure that the quantities of interest and their variances do not become ill defined. Tensor network Monte Carlo methods do not suffer from such obstacles because the summation being sampled

ranges over a finite (albeit intractably large) number of terms, all of which are themselves finite. However, it is still the case that the efficiency of these techniques depends on having well behaved distributions, as seen in the dramatic drop in per-sample standard deviation as the bond dimension is increased that we highlight in table 7.1.

When using a deterministic approach to tensor network renormalization one can improve the accuracy of the results only by increasing the bond dimension, whereas a sampling approach provides two ways to accomplish this goal. The first is to increase the number of samples at fixed bond dimension, and the second is to to increase the bond dimension used in each sample. In the limit, as the number of samples, or the bond dimension, goes to infinity, both of these approaches are guaranteed to drive the error to zero. However, while its possible to get the exact partition function or energy at a very small bond dimension using a stochastic approach, it still might be inefficient due to the number of samples needed. To give some perspective on the relative effectiveness of tuning these two parameters we plot a comparison between a deterministic TRG calculation and our MCTRG with a fixed number of samples across a range of bond dimensions. We can see that the deterministic and stochastic versions of TRG become dramatically more accurate with higher bond dimension. While not shown in Fig. 7.6, the effect of increasing the number of samples is straightforward to understand. Because our samples are calculated independently, we are guaranteed that the expected value of our error will be suppressed inversely proportional to the square root of the number of additional samples we generate.

Interestingly, we see that our stochastic TRG tends to benefit from an increase in bond dimension slightly more than the deterministic version. We speculate that in the low bond dimension regime the deterministic algorithm benefits from fortuitous cancellation of errors while the stochastic approach is beset by the difficulties of sampling from skewed multimodal distributions like those in the left panel of Fig. 7.5. As the bond dimension increases and both of these effects are attenuated the average error per sample in our algorithm drops and we see that we are able to significantly improve on the deterministic results at a given bond dimension by taking a modest number of samples. Also notable is that in neither case does the increase in accuracy appear monotonic. The jaggedness of the curve for the deterministic results is a well known phenomenon which is frequently encountered in attempts to extrapolate to the infinite bond dimension limit. For enough samples we would expect the stochastic nature of our Monte Carlo version of TRG to smooth out these effects and show a consistent decreases in error with increasing bond dimension but this behavior does not manifest strongly at the sample sizes we have considered.

7.5 Discussion

Renormalization group approaches have already proven extremely useful in the quest to take tensor network methods beyond DMRG, but substantial challenges remain. These algorithms can be prohibitively expensive in terms of both time and memory, especially for higher dimensional systems. Furthermore, if the renormalization scheme is not well-suited to the

entanglement structure of the tensor network of interest, the bond dimension required for a given accuracy may increase exponentially with the size of the system, or, equivalently, the accuracy may increase very slowly with bond dimension. Extrapolations to the infinite bond dimension limit and estimations of error bounds are challenging to perform [368], often rest on unproven heuristics, and can be hampered by a variational bias towards certain states [381]. Finally, these algorithms are labor-intensive to program and optimize even without planning for the parallelizability necessary to make good use of modern computing resources.

Our aim in this paper has been to investigate the feasibility of using the randomized truncation techniques presented in Ref. [381] to alleviate these difficulties. To that end, we implemented a stochastic version of Levin and Nave's tensor renormalization group [360], the simplest of a family of related algorithms, and dissected its performance on the well-studied 2d classical Ising model. We found that, in many cases, an average of a hundred independent samples could outperform a deterministic calculation at an equivalent bond dimension.

We argue that this result is near optimal for such a stochastic analogue of TRG. The standard approach of choosing to retain the top χ singular values is the best choice for a single sample. By randomly choosing a different subset of singular values at each step we expect to do worse with each sample. In exchange for paying this penalty in accuracy per sample, we gain, in addition to unbiased error bars, another parameter besides bond dimension that can be used to systematically improve the accuracy of our calculations. This means that by using parallel computing resources to take more samples, we can arbitrarily suppress the error, controllably approach the exact result, and get an unbiased estimate of the remaining errors. We expect this to be especially useful in regimes where tensor network algorithms struggle and the utility of increasing the bond dimension is limited. Our specific finding that a small number of stochastic samples can be competitive in accuracy with a deterministic calculation at the same bond dimension suggests that a wide swath of tensor network techniques could be made more accurate for a reasonable overhead in parallelizable computing resources.

It is also possible to think of many potential improvements to, and specialized applications of, the approach presented here. For example, even in cases where the underlying system is translationally invariant and one would be reluctant to pay the overhead necessary to do a full stochastic summation we suggest that significant advantages could be had by performing several deterministic coarse-graining steps before switching to a Monte Carlo approach as the number of non-trivial singular values starts to increase. We expect that larger bond dimensions could be made accessible without losing the unbiased nature of our approach by using the randomized singular value decomposition as in Ref. [384]. We are also excited about the potential for making use of the unbiased estimates provided by Monte Carlo tensor network contraction as a component of a new PEPS optimization procedure. Furthermore, we are generally hopeful that the unbiased estimate and error bars of a stochastic approach to tensor network renormalization will enhance the interpretability and reliability of calculations performed using these techniques.

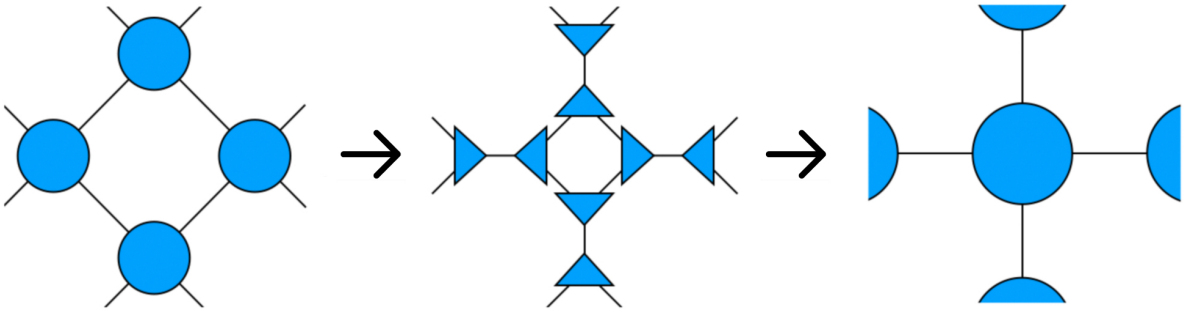

Figure 7.7: A schematic of the tensor renormalization group (TRG) algorithm applied to a zoomed-in view of a square lattice of tensors [360]. Between the left-hand diagram and the central diagram, a singular value decomposition is used to decompose each rank-4 tensor into two rank-3 tensors according to Eq. 7.22. Specifically, the circular rank-4 tensor corresponds to T in Eq. 7.22 and the triangular rank-3 tensors correspond to A and B. The use of triangles instead of circles here is merely a matter of convention; the tensors in the central panel do not represent isometries (which are sometimes indicated by triangles in such diagrams). Subsequently, the tensors are grouped into groups of four and contracted to yield the renormalized lattice of the right-hand diagram. The new lattice contains half as many tensors as the original lattice, which can be seen by observing that we first double the number of tensors by decomposing each circle into two tensors and then quarter it by contracting tensors in groups of four.

7.6 Additional Computational Details

In the main text of this chapter, we focused on our new contributions and did not present an overview of the tensor renormalization group (TRG) algorithm of Ref. 360 to which we applied the Monte Carlo approach of Ref. 381. Here we shall fill in these details with a short description of the algorithm, explaining how it essentially comes down to repeated applications of the decomposition described in Eq. 7.10. We shall focus on describing a single TRG step for a tensor network consisting of a square lattice of tensors with periodic boundary conditions. Figure 7.2 is an example of such a tensor network, except that it has open boundary conditions instead of periodic boundary conditions. As we shall see, the algorithm is most straightforwardly applied to a tensor network with 2^n tensors, for some natural number n. This is because each step of the TRG algorithm takes a lattice with 2^i tensors to one with (approximately) the same trace but with 2^{i-1} tensors.

The basic idea is to treat each tensor as a matrix between pairs of indices and to use a

truncated singular value decomposition (SVD) to decompose that tensor into two pieces,

$$T_{(ab)(cd)} = \sum_{i=1}^{\chi} U_{(ab)i} S_{ii} V_{i(cd)} \approx \sum_{i=1}^{\tilde{\chi}\leq\chi} \left(U_{(ab)i}\sqrt{S_{ii}}\right)\left(\sqrt{S_{ii}}V_{i(cd)}\right) = \sum_{i=1}^{\tilde{\chi}\leq\chi} A_{(ab)i} B_{i(cd)}. \tag{7.22}$$

In the above equation, we describe how a rank-4 tensor T is decomposed into a contraction of two tensors A and B. We indicate the grouping of the indices of T into two pairs with parentheses. The transition between the left-hand diagram in Figure 7.7 and the middle one illustrates the application of this decomposition to four tensors in a square lattice. Notice that the adjacent tensors are matricized in an alternating pattern, with half of them decomposed horizontally and half of them decomposed vertically. This pattern is repeated across the entire lattice. If the original tensor network is translationally invariant, the decomposition only needs to be performed twice, once for each grouping of the indices.

After this decomposition is performed, we finish the renormalization step by contracting the tensors together in groups of four to recover a new square lattice. In Figure 7.7, this can be seen in the transition from the middle diagram to the right-hand diagram. The four triangular tensors at the center of the middle diagram are contracted together to form the circular tensor at the center of the right-hand diagram. The new square lattice is rotated 45° with respect to the original lattice and contains exactly half as many tensors. If the singular value decomposition indicated in Eq. 7.22 is performed without truncation then the trace of the resulting tensor network would be the same as the trace of the original. Otherwise, there is some approximation. After repeated applications of this renormalization procedure, the tensor network will be reduced to a sufficiently small size that it can be contracted exactly. Alternatively, if one is interested in the convergence to the thermodynamic limit, one can contract the tensors after each TRG step. The resulting values of the trace correspond to a series of finite-sized periodic tensor networks whose sizes are growing exponentially. One can then halt once the value of the partition function converges appropriately.

Chapter 8

Towards Quantum Machine Learning with Tensor Networks

8.1 Preface

Except for Section 8.2, which was written specifically for this thesis, this chapter is taken from the previously published Ref. 159, originally appearing in Quantum Science and Technology. It was co-authored by the author of this thesis, Piyush Patil, Bradley Mitchell, Professor K. Birgitta Whaley, and Dr. Miles Stoudenmire. The author of this thesis developed the original ideas related to the discriminative machine learning techniques under the guidance of Dr. Stoudenmire, who also contributed the aspects related to generative machine learning and the qubit-efficient application of these models. The author of this thesis designed and performed the numerical experiments. Piyush Patil rewrote the original implementation to be faster and more suitable for further experimentation. Bradley Mitchell implemented the noisy circuit simulations and, together with Piyush Patil, contributed to discussions around the classical machine learning context. Dr. Stoudenmire wrote the majority of the paper, with the author of this thesis contributing to the sections describing the implementation. All of the co-authors helped to revise the manuscript into its final form.

8.2 Machine Learning Background Material

In this section, we shall briefly review some basic concepts from the field of machine learning. We shall focus on a few ideas relevant to laying the foundation for the rest of the chapter, mostly drawn from a family of modern approaches known as deep learning. We refer to interested readers to Ref. 157 for a comprehensive general introduction to the field. The field of deep learning primarily concerns itself with parameterized models known as neural networks. This chapter shall focus on a different kind of parameterized machine learning model, but in both cases the idea is to choose values for a set of parameters, $\vec{\theta}$, such that the model performs a particular task.

The first important notion we review is the idea of a discriminative model. A discriminative model aims to approximate a conditional probability distribution $P(y|x)$, where the possible ys are labels for the object x. For example, a discriminative model might assign a probability distribution over the labels $\{cat, dog\}$ to a particular image x. We are primarily interested here in supervised learning. In the context of discriminative models, supervised learning means that we train our model (choose its parameters) by fitting it to a series of labeled examples. We call the set of labeled examples, which are pairs $\{(x_i, y_i)\}$ of examples and their corresponding label, that we use to train our model the training set. By optimizing the model's parameters ($\vec{\theta}$) to correctly reproduce the desired behavior on the training set,

$$\delta(y, y_i) = P(y|x_i) = F(x_i, \vec{\theta}), \tag{8.1}$$

we hope to use our model, $F(x, \vec{\theta})$, to approximate the unknown function $P(y|x)$.

This process can fail because of overfitting, a name for the situation where $F(x, \vec{\theta})$ reproduces the conditional probability distribution on the training set but performs poorly on unseen examples. This failure to generalize is often guarded against by validating the model using a split between training and testing data. This entails holding back some labeled examples, which we call the test set, and using only a subset of the labeled examples as the training set. If both the training and test form a representative collection of the possible input values (x), then the models performance on the test set should approximate its performance on other unseen examples drawn from the same distribution.

Generative models aim to perform a different task, either approximating a probability distribution $P(x)$ or a joint probability distribution $P(x, y)$. Generative models are often considered to be examples of unsupervised learning, although there are nuances in the definition that we do not debate here. A generative machine learning algorithm is generally responsible for inferring the underlying structure in a probability distribution ($P(x)$) from a collection of examples $\{x_i\}$ sufficiently well to enable sampling from a distribution that approximates $P(x)$. Because of the relative ambiguity of this task compared to discriminative machine learning, generative machine learning is generally regarded as challenging. However, recent progress, particularly using an approach known as generative adversarial networks, is extremely promising [385].

The construction of both discriminative and generative machine learning models is often framed as an optimization problem, particularly in the context of deep learning. In this framing, the objective function is known as the loss function. Variational methods in physics and chemistry usually come with a single, natural, choice of objective function, such as the expectation value of energy. By contrast, in machine learning tasks, designing a loss function that leads to the desired behavior can be a substantial part of model design [386]. The fact that the measurable quantities (such as performance on the training examples) are only an approximation to the real desiderata (such as performance on new samples from the underlying distribution) leads to other subtle differences between machine learning tasks and more familiar variational algorithms. For example, it has been observed that stopping the optimization of a neural network before a minimum value of the loss function has been attained can improve its performance on unseen examples [387].

The choices made during the design of the model and the training process are commonly referred to as hyperparameters. This language serves to distinguish them from the parameters of the model itself. However, in order to achieve the best possible performance, these hyperparameters can also be optimized. A variety of approaches are taken here, with a combination of random search, grid search, and manual tuning (known in academic circles as "grad student descent") being particularly common [388]. More sophisticated methods are also employed. In recent years, the hyperparameter optimization has evolved into a full-fledged subfield of machine learning dedicated to the automated design of machine learning approaches [389].

8.3 Introduction

For decades, quantum computing has promised to revolutionize certain computational tasks. It now appears that we stand on the eve of the first experimental demonstration of a quantum advantage [15]. With noisy, intermediate scale quantum computers around the corner, it is natural to investigate the most promising applications of quantum computers and to determine how best to harness the limited, yet powerful resources they offer.

Machine learning is a very appealing application for quantum computers because the theories of learning and of quantum mechanics both involve statistics at a fundamental level, and machine learning techniques are inherently resilient to noise, which may allow realization by near-term quantum computers operating without error correction. But major obstacles include the limited number of qubits in near-term devices and the challenges of working with real data. Real data sets may contain millions of samples, and individual samples are typically vectors with hundreds or thousands of components. Therefore one would like to find quantum algorithms that can perform meaningful tasks for large sets of high-dimensional samples even with a small number of noisy qubits.

The quantum algorithms we propose in this work implement machine learning tasks—both discriminative and generative—using circuits equivalent to tensor networks [357, 366, 390], specifically tree tensor networks [391–394] and matrix product states [361, 390, 395]. Tensor networks have recently been proposed as a promising architecture for machine learning with classical computers [396–398], and provide good results for both discriminative [397–402] and generative learning tasks [403]. Tensor networks were also recently put forward as an appealing framework for using quantum computers to find ground states of many-body quantum systems, with benefits including the use of a small number of qubits and resilience to noise [227].

The circuits we will study contain many parameters which are not determined at the outset, in contrast to quantum algorithms such as Grover search or Shor factorization [4, 404]. Only the circuit geometry is fixed, while the parameters determining the unitary operations must be optimized for the specific machine learning task. Our approach is therefore conceptually related to the quantum variational eigensolver [17, 140] and to the quantum approximate

optimization algorithms [19], where quantum circuit parameters are discovered with the help of an auxiliary classical algorithm.

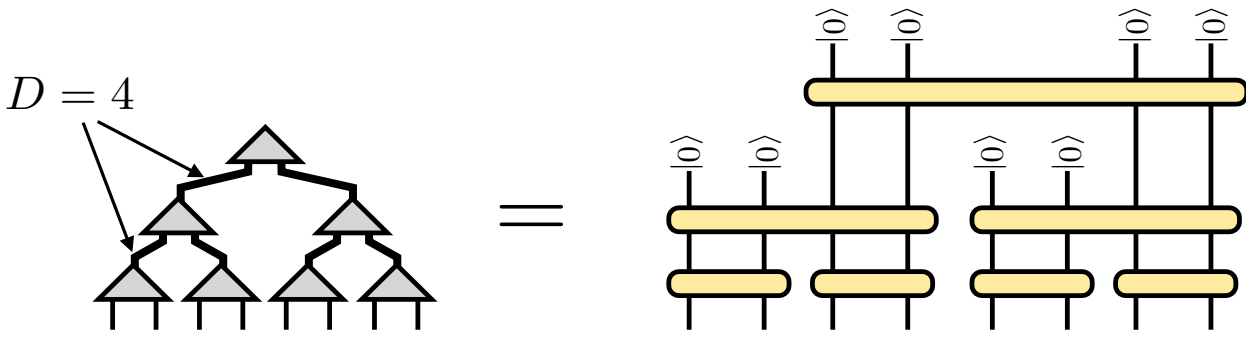

Figure 8.1: The quantum state of N qubits corresponding to a tree tensor network (left) can be realized as a quantum circuit acting on N qubits (right). The circuit is read from top to bottom, with the yellow bars representing unitary gates. The bond dimension D connecting two nodes of the tensor network is determined by number of qubits V connecting two sequential unitaries in the circuit, with $D = 2^V$.

The application of such hybrid quantum-classical algorithms to machine learning was recently investigated by several groups for labeling [20, 405] or generating data [154, 406, 407]. The proposals of Refs. 20, 154, 405, 407 are related to approaches we propose below, but consider very general classes of quantum circuits. This motivates the question: is there a subset of quantum circuits which are especially natural or advantageous for machine learning tasks? Tensor network circuits might provide a compelling answer, for three main reasons:

1. Tensor network models could be implemented on **small, near-term quantum devices** for input and output dimensions far exceeding the number of physical qubits. If the hardware permits the measurement of one of the qubits separately from the others, then the number of physical qubits needed can be made to scale either logarithmically with the size of the processed data, or independently of the data size depending on the particular tensor network architecture. Models based on tensor networks may also have an inherent resilience to noise. We explore both of these aspects in Section 8.6.

2. There is a **gradual crossover** from classically simulable tensor network circuits to circuits that require a quantum computer to evaluate, and exploring this boundary is an area of active research [227, 408]. With classical resources, tensor network models already give very good results for supervised [397, 398, 400, 402] and unsupervised [402, 403] learning tasks. The same models—with the same dataset size and data dimension—can be used to initialize more expressive models requiring quantum hardware, making the optimization of the quantum-based model faster and more likely to succeed. Algorithmic

improvements in the classical setting can be readily transferred to the quantum setting as well.

3. There is a rich **theoretical understanding** of the properties of tensor networks [355, 357, 361, 366, 390, 409], and their relative mathematical simplicity (involving only linear operations) will likely facilitate further conceptual developments in the machine learning context, such as interpretability and generalization. Properties of tensor networks, such as locality of correlations, may provide a favorable inductive bias for processing natural data [399]. One can prove rigorous bounds on the noise-resilience of quantum circuits based on tensor networks [227].

All of the experimental operations necessary to implement tensor network circuits are available for near-term quantum hardware. The capabilities required are preparation of product states; one- and two-qubit unitary operations; and measurement in the computational basis.

In what follows, we first describe our proposed frameworks for discriminative and generative learning tasks in Section 8.4. Then we present results of a numerical experiment which demonstrates the feasibility of the approach using operations that could be carried out with an actual quantum device in Section 8.5. We conclude by discussing how the learning approaches could be implemented with a small number of physical qubits and by addressing their resilience to noise in Section 8.6.

8.4 Learning with Tensor Network Quantum Circuits

The family of tensor networks we will consider—tree tensor networks and matrix product states—can always be realized precisely by a quantum circuit; see Fig. 8.1. Typically, the quantum circuits corresponding to tensor networks are carefully devised to make them efficient to prepare and manipulate with classical computers [358]. With increasing bond dimension, tree and matrix product state tensor gradually capture a wider range of states, and eventually the full Hilbert space.

In the context of machine learning, we use these states to represent the trainable "weight" parameters of a particular family of models, defined by a *feature map* Φ, with an example given below. For the types of feature maps we will choose, working in the space of all possible weights (full Hilbert space) corresponds to extremely powerful and expressive models which would likely overfit most data sets [396–398].

Using a finite-bond-dimension tensor network to represent the weights instead limits this expressivity, but the limitation can be rapidly lifted by increasing the bond dimension. Currently, the limit for practical bond dimensions is in the tens of thousands for matrix product state tensor networks, and many hundreds or perhaps thousands for more general tree tensor networks. By implementing tensor network circuits on quantum hardware instead, one could in principle go far beyond these classical limitations on model expressivity.

In this section, we first describe our tensor-network based proposal for performing discriminative tasks with quantum hardware. The goal of a discriminative model is to produce a specific output given a certain class of input; for example, assigning labels to images. Then we describe our proposal for generative tasks, where the goal is to generate samples from a probability distribution inferred from a data set. For more background on various types of machine learning tasks, see the recent review Ref. 410.

For clarity of presentation, we shall make use of multi-qubit unitary operations in this work. However we recognize that in practice such unitaries must be implemented using a more limited set of few-qubit operations, such as the universal gate sets of one- and two-qubit operators. Whether it is more productive to classically optimize over more general unitaries then "compile" these into few-qubit operations as a separate step, or to parameterize the models in terms of fewer operations from the outset remains an interesting and important practical question for further work. In either case, the observation that the models we consider in this work all demand $\mathcal{O}(N)$ parameterized multi-qubit unitary operations (where N is the number of input qubits for the discriminative models and the number of output qubits for the generative models) can guide more specific estimates of the resource counts for different implementation choices.

8.4.1 Discriminative Algorithm

To explain the discriminative tensor network framework that we propose here, assume that the input to the algorithm takes the form of a vector of N real numbers $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, with each component normalized such that $x_i \in [0, 1]$. For example, such an input could correspond to a grayscale image with N pixels, with individual entries encoding normalized grayscale values. We map this vector $\mathbf{x} \in \mathbb{R}^N$ to a product state on N qubits according to the feature map proposed in Ref. 398:

$$\begin{aligned} &\mathbf{x} \;\rightarrow \\ &|\Phi(\mathbf{x})\rangle = \begin{bmatrix} \cos\left(\frac{\pi}{2}x_1\right) \\ \sin\left(\frac{\pi}{2}x_1\right) \end{bmatrix} \otimes \begin{bmatrix} \cos\left(\frac{\pi}{2}x_2\right) \\ \sin\left(\frac{\pi}{2}x_2\right) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} \cos\left(\frac{\pi}{2}x_N\right) \\ \sin\left(\frac{\pi}{2}x_N\right) \end{bmatrix}. \end{aligned} \tag{8.2}$$

Such a state can be prepared by starting from the computational basis state $|0\rangle^{\otimes N}$, then applying a single qubit unitary to each qubit $n = 1, 2, \ldots, N$.

The model we then propose can be seen as an iterative coarse-graining procedure that parameterizes a CPTP (completely positive trace preserving) map from an N-qubit input space to a small number of output qubits encoding the different possible class labels. The circuit takes the form of a tree, with V qubit lines connecting each subtree to the rest of the circuit. We call such qubit lines "virtual qubits" to connect with the terminology of tensor networks, where tensor indices internal to the network are called virtual indices. A larger V can capture a larger set of functions, just as a tensor network with a sufficiently large bond dimension can parameterize any N-index tensor. More concretely, we can interpret the circuit

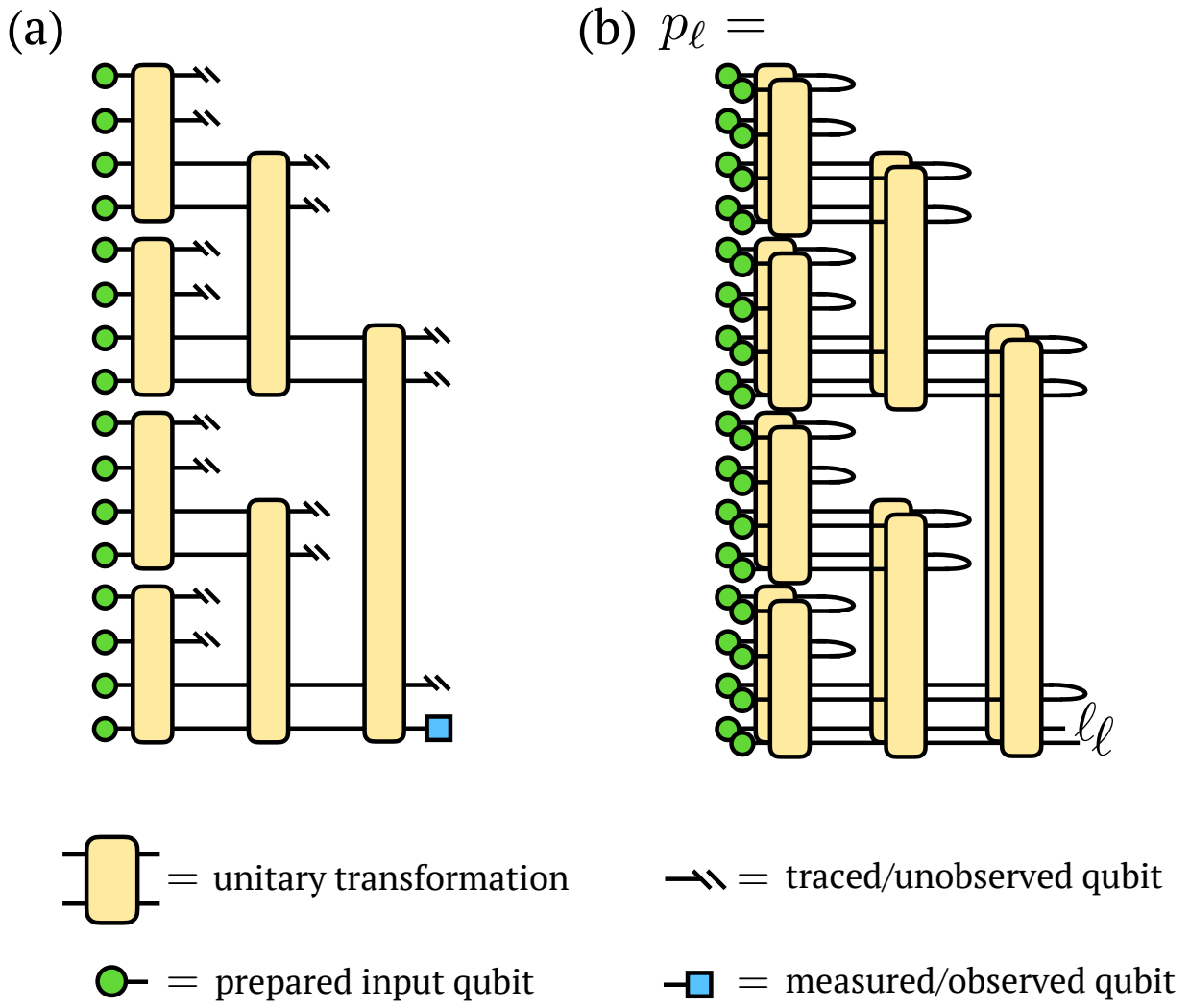

Figure 8.2: Discriminative tree tensor network model architecture, showing an example in which $V = 2$ qubits connect different subtrees. Figure (a) shows the model implementation as a quantum circuit. Circles indicate inputs prepared in a product state as in Eq. 8.2; hash marks indicate qubits that remain unobserved past a certain point in the circuit. A particular pre-determined qubit is sampled (square symbol) and its distribution serves as the output of the model. Figure (b) shows the tensor network diagram for the reduced density matrix of the output qubit.

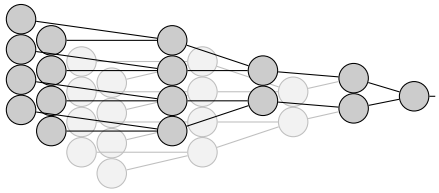

Figure 8.3: The connectivity of nodes of our tree network model, as it would be applied to a 4x4 image. Each step coarse-grains in either the horizontal or the vertical directions, and these steps alternate. The resulting binary tree structure can be easily parameterized by few-qubit unitary operations.

as a tensor network where the bond dimension D of a virtual index with V qubit lines is given by $D = 2^V$.

At each step of implementing the model on a quantum computer, one takes V of the qubits resulting from one of the unitary operations of the previous step, or subtree, and V from another subtree and acts on them with another parameterized unitary transformation (possibly together with some ancilla qubits—not shown). Then V of the qubits are discarded or ignored, while the other V qubits proceed to the next node of the tree, that is, the next step of the circuit. In our classical simulations we trace over all discarded qubits, while on a quantum computer, we would be free to ignore or reset such qubits.

Once all unitary operations defining the circuit have been carried out, one or more qubits serve as the output qubits. (Which qubits are outputs is designated ahead of time.) The most probable state of the output qubits determines the prediction of the model, that is, the label the model assigns to the input. To determine the most probable state of the output qubits, one performs repeated evaluations of the circuit for the same input in order to estimate their probability distribution in the computational basis.

The full discriminative algorithm then consists of the following three steps: (1) the model is evaluated as discussed above for a collection of labeled inputs from a training set. Then (2) a classical algorithm is used to compare the estimated output of the model for each input to the known, correct outputs or labels. Finally, (3) the results of the comparison are used to propose an improved circuit. Starting each time from the improved circuit, one repeats the above steps until convergence. In Section 8.5 we discuss in much more detail a precise implementation of the above steps that can be performed on quantum hardware.

In Fig. 8.2 we show the quantum circuit which implements our proposed discriminative model. In the case of image classification, it is natural to always group input qubits based on pixels coming from nearby regions of the image, with a tree structure illustrated schematically in Fig. 8.3.

A closely related family of models can be devised based on matrix product states. An example is illustrated in Fig. 8.4 showing the case of $V = 2$. Matrix product states (MPS) can be viewed as maximally unbalanced trees, and differ from the binary tree models described

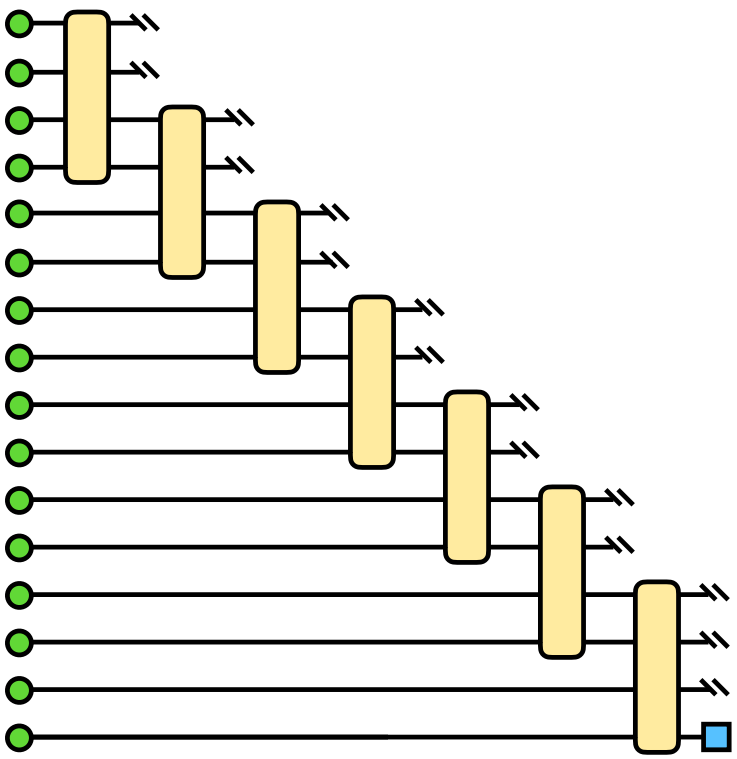

Figure 8.4: Discriminative tensor network model for the case of a matrix product state (MPS) architecture with $V = 2$ qubits connecting each subtree. The symbols have the same meaning as in Fig. 8.2. An MPS can be viewed as a maximally unbalanced tree.

above in that after each unitary operation on $2V$ inputs only one set of V qubits are passed to the next node of the network. Such models are likely a better fit for data that has a one-dimensional pattern of correlations, such as time-series, language, or audio data.

8.4.2 Generative Algorithm

The generative algorithm we propose is nearly the reverse of the discriminative algorithm, in terms of its circuit architecture. The algorithm produces random samples by first preparing a quantum state then measuring it in the computational basis, putting it within the family of algorithms recently dubbed "Born machines" [403, 406, 407, 411, 412]. But rather than preparing a completely general state, we shall consider specific patterns of state preparation corresponding to tree and matrix product state tensor networks. This provides the advantages discussed in the introduction, such as connections to classical tensor network models and the ability to reduce the number of physical qubits required, which will be discussed further in Section 8.6.

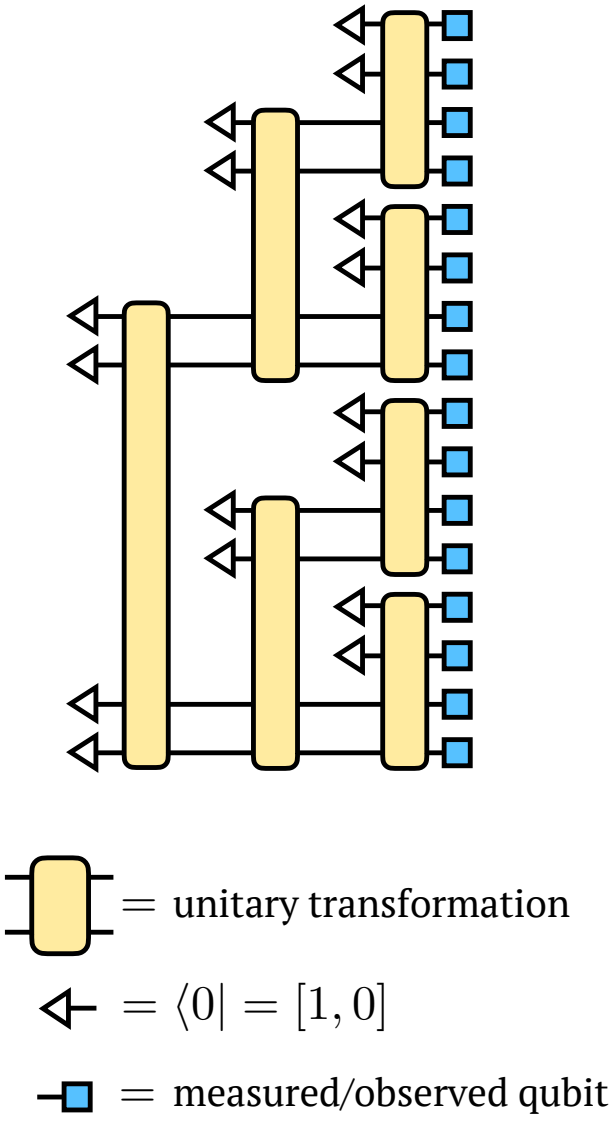

Figure 8.5: Generative tree tensor network model architecture, showing a case with $V = 2$ qubits connecting each subtree. To sample from the model, qubits are prepared in a reference computational basis state $\langle 0|$ (left-hand side of circuit). Then $2V$ qubits are entangled via unitary operations at each layer of the tree as shown. The qubits are measured at the points in the circuit labeled by square symbols (right-hand side of circuit), and the results of these measurements provides the output of the model. While all qubits could be entangled before being measured, we discuss in Section 8.6 the possibility performing opportunistic measurements to reduce the physical qubit overhead.

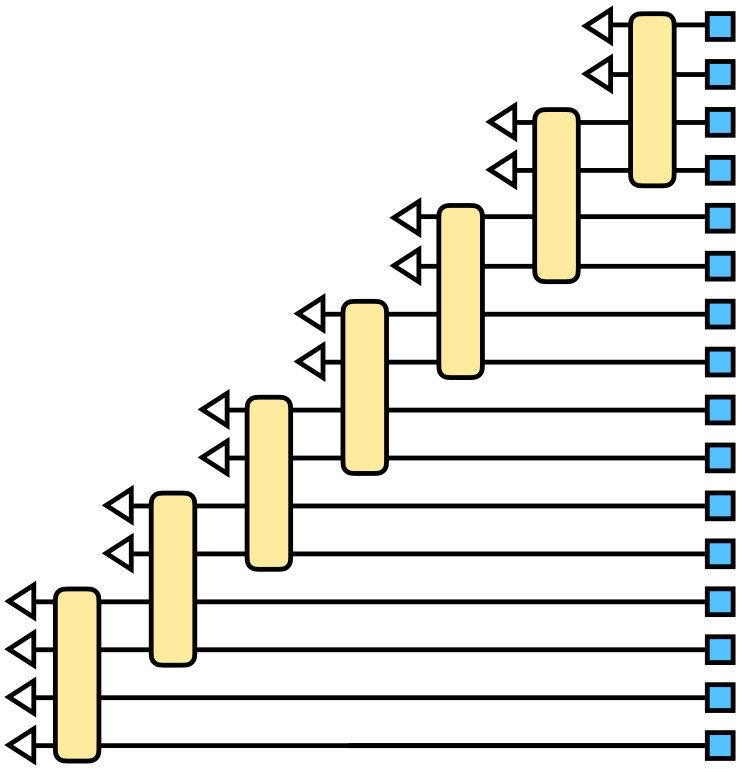

Figure 8.6: Generative tensor network model for the case of a matrix product state (MPS) architecture with $V = 2$ qubits connecting each unitary. The symbols have the same meaning as in Fig. 8.5.

The implementation of the generative circuit based on a tree tensor network (shown in Fig. 8.5) begins by preparing $2V$ qubits in a reference computational basis state $\langle 0|^{\otimes 2V}$, then entangling these qubits by unitary operations. Another set of $2V$ qubits are prepared in the state $\langle 0|^{\otimes 2V}$. Half of these are grouped with the first V entangled qubits, and half with the second V entangled qubits. Two more unitary operations are applied to each new grouping of $2V$ qubits; the outputs are now split into four groups; and the process repeats for each group. The process ends when the total number of qubits processed reaches the size of the output one wants to generate.

Once all unitaries acting on a certain qubit have been applied, this qubit can be measured. The measured output of all of the qubits in the computational basis represents one sample from the generative model. The generative algorithm we then propose consists of three steps: (1) Given a particular circuit specified by a set of unitary operations, one repeatedly prepares the circuit and draws samples from it as described above. Then (2) one uses a classical algorithm to compare the samples drawn from the circuit to samples within a fixed training set representing the desired distribution. Finally, (3) the classical algorithm used to compare

the model and data samples is used to propose an updated generative circuit.

How best to make this comparison between samples drawn from the model and from the data set is an area of continuing research; however, good proposals for this comparison appropriate for a quantum computer have been explored in Refs. 406, 407. For example, one can attempt to minimize the difference between the mean and covariance of the model and data distributions. Depending on the family of unitaries chosen to parameterize the circuit, it may be possible to use a modified circuit to directly sample the gradient of the model parameters [154]. Or in the most general case, one can use gradient-free optimization methods to propose improved circuit parameters [407].

We illustrate our proposed generative approach for the case of $V = 2$ and binary outputs in Fig. 8.5. As in the discriminative case, one can also devise an MPS based generative algorithm more suitable for one-dimensional data. The circuit for such an algorithm is shown in Fig. 8.6.

8.5 Numerical Experiments

To show the feasibility of implementing our proposal on a near-term quantum device, we trained a discriminative model based on a tree tensor network for a supervised learning task, namely labeling image data. The specific network architecture we used is shown as a quantum circuit in Fig. 8.7. When viewed as a tensor network, this model has a bond dimension of $D = 2$. This stems from the fact that after each unitary operation entangles two qubits, only one of the qubits is acted on at the next scale (next step of the circuit). All numerical results presented in this work were carried out using code written on top of the TensorFlow Python library [264]. Initial experiments were also conducted using the ITensor C++ library [413].

In the future, it would also be very interesting to conduct numerical experiments with generative tensor network circuit. Though we do not perform such experiments here, we are optimistic that they would produce good results based on both the success of the discriminative experiment below and on the success of related generative experiments (based on more general circuits not equivalent to tensor networks) in the recent literature [407].

8.5.1 Loss Function

Our eventual goal is to select the parameters of our circuit such that we can confidently assign the correct label to a new piece of data by running our circuit a small number of times. To this end, we choose the loss function which we want to minimize starting with the following definitions. Let $\mathbf{\Lambda}$ be the model parameters; $\mathbf{d}$ be an element of the training data set; and let $p_\ell(\mathbf{\Lambda}, \mathbf{x})$ be the probability of the model to output a label ℓ for a given input $\mathbf{x}$. Because we consider the setting of supervised learning, the correct labels are known for the training set inputs, and define $\ell_{\mathbf{x}}$ to be the correct label for the input $\mathbf{x}$. Now define

$$p_{\text{largest false}}(\mathbf{\Lambda}, \mathbf{x}) = \max_{\ell \neq \ell_{\mathbf{x}}} \Big[p_\ell(\mathbf{\Lambda}, \mathbf{x}) \Big] \tag{8.3}$$

as the probability of the incorrect output state which has the highest probability of being observed. Then, define the loss function for a single input $\mathbf{x}$ to be

$$L(\mathbf{\Lambda}, \mathbf{x}) = \max(p_{\text{largest false}}(\mathbf{\Lambda}, \mathbf{x}) - p_{\ell_{\mathbf{x}}}(\mathbf{\Lambda}, \mathbf{x}) + \lambda, 0)^{\eta}, \tag{8.4}$$

and the total loss function to be

$$L(\mathbf{\Lambda}) = \frac{1}{|\text{data}|} \sum_{\mathbf{x} \in \text{data}} L(\mathbf{\Lambda}, \mathbf{x}). \tag{8.5}$$

The "hyper-parameters" λ and η are to be chosen to give good empirical performance on a validation data set. Essentially, we assign a penalty for each element of the training set where the gap between probability of assigning the true label and the probability of assigning the most likely incorrect label is less than λ. The relative weight of the penalty applied to small errors versus large ones in controlled by the hyper-parameter η. This loss function allows us to concentrate our efforts during training on making sure that we are likely to assign the correct label after taking the majority vote of several executions of the model, rather than trying to force the model to always output the correct label in each separate run. Early numerical experiments with simpler loss functions were less satisfactory than the results presented below but it would be interesting to use these circuits as a platform for a more systematic study of possible loss functions and optimization strategies.

8.5.2 Optimization

Of course, we are interested in training our circuit to generalize well to unobserved inputs, so instead of optimizing over the entire distribution of data as in Eq. 8.5, we optimize the loss function over a subset of the training data and compare to a held-out set of test data. Furthermore, because the size of the training set for a typical machine learning problem is so large (60,000 examples in the case of the MNIST data set), it would be impractical to calculate the loss over all of the training data at each optimization step. Instead, we follow a standard approach in machine learning and randomly select a mini-batch of training examples at each iteration. Then, we use the following stochastic estimate of our true training loss (recalling that $\mathbf{\Lambda}$ represents the current model parameters):

$$\tilde{L}(\mathbf{\Lambda}) = \frac{1}{|\text{mini-batch}|} \sum_{\mathbf{x} \in \text{mini-batch}} L(\mathbf{\Lambda}, \mathbf{x}) \tag{8.6}$$

In order to faithfully test how our approach would perform on a near-term quantum computer, we have chosen to minimize our loss function using a variant of the simultaneous perturbation stochastic approximation (SPSA) algorithm which was recently used to find quantum circuits approximating ground states in Ref. 227 and was originally developed in Ref. 414. We note that an alternative approach to optimization could be pursued by constructing quantum circuits which compute the derivatives of the loss function with respect

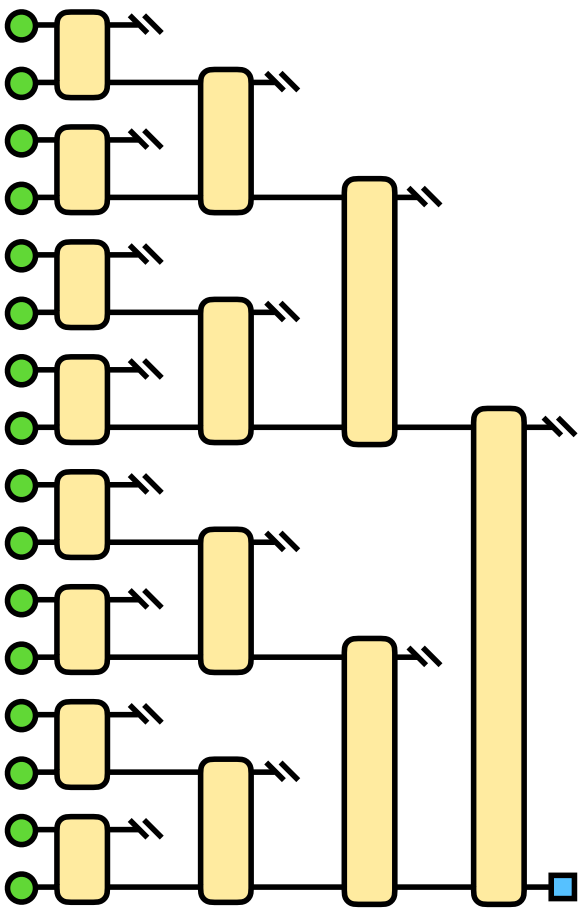

Figure 8.7: Model architecture used in the experiments of Section 8.5, which is a special case of the model of Fig. 8.2 with one virtual qubit connecting each subtree. For illustration purposes we show a model with 16 inputs and 15 two-qubit gates in 4 layers above, whereas the actual model used in the experiments had 64 inputs and 63 two-qubit gates in 6 layers.

to the model parameters using the techniques of Ref. 204 but we defer a more detailed investigation of the optimization strategies for a future work.

Each step of the SPSA algorithm we employed estimates the gradient of the loss function by performing a finite difference calculation along a random direction and updates the parameters accordingly. In our experimentation, we have also found it helpful to include a momentum term $\mathbf{v}$, which mixes a fraction of previous update steps into the current update. We outline the algorithm we used in more detail below.

1. Initialize the model parameters $\mathbf{\Lambda}$ randomly, and set $\mathbf{v}$ to zero.

2. Choose appropriate values for the constants, $a, b, A, s, t, \gamma, n, M$ that define the optimization procedure.

3. For each $k \in \{0, 1, 2, ..., M\}$, set $\alpha_k = \frac{a}{(k+1+A)^s}$ and $\beta_k = \frac{b}{(k+1)^t}$, and randomly partition the training data into mini-batches of n images. Perform the following steps using each mini-batch:

   a) Generate random perturbation $\mathbf{\Delta}$ in parameter space.

   b) Evaluate $g = \frac{\tilde{L}(\mathbf{\Lambda}_{old}+\alpha_k\mathbf{\Delta})-\tilde{L}(\mathbf{\Lambda}_{old}-\alpha_k\mathbf{\Delta})}{2\alpha_k}$, with $\tilde{L}(\mathbf{x})$ defined as in Eq. 8.6.

   c) Set $\mathbf{v}_{new} = \gamma\mathbf{v}_{old} - g\beta_k\mathbf{\Delta}$

   d) Set $\mathbf{\Lambda}_{new} = \mathbf{\Lambda}_{old} + \mathbf{v}_{new}$

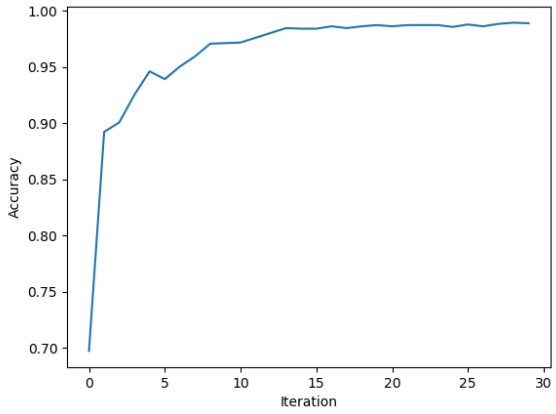

Figure 8.8: Test accuracy as a function of the number of SPSA epochs ($M = 30$, in the language of the previous section) for binary classification of handwritten 0's and 7's from the MNIST data set.

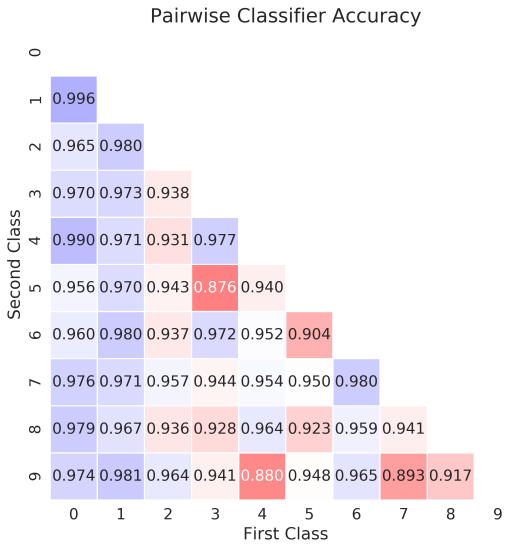

Figure 8.9: The test accuracy for each of the pairwise classifiers trained with the hyper-parameters mentioned in the text. The accuracy for each classifier can be found by choosing the position along the x-axis corresponding to one class and the position on the y-axis corresponding to the other.

8.5.3 Results

We trained circuits with a single output qubit at each node to recognize grayscale images of size 8×8 belonging to one of two classes using the SPSA optimization procedure described above. The images were obtained from the MNIST data set of handwritten digits [415], and we show results below for classifiers trained to distinguish between each of the 45 pairs of handwritten digits 0 through 9.

The unitary operations U applied at each node in the tree were parameterized by writing them as $U = \exp(iH)$ where H is a Hermitian matrix (the matrices H were allowed to be different for each node). The free parameters were chosen to be the elements forming the diagonal and upper triangle of each Hermitian matrix, resulting in exactly 1008 free parameters spread over 63 two-qubit gates for the 8×8 image recognition task. The mini-batch size and the other hyper-parameters for the training procedure and the loss function were hand-tuned by running a small number of experiments, using the SigOpt [416] software package, with the goal of obtaining the most rapid and consistent performance (averaged over the different digit pairs) on a validation data set.

We do not report any of the results here, but we note for the interested reader that preliminary experiments with larger numbers of free parameters led to significantly more

challenging optimization problems. This led us to focus, for this work, on performing binary classification with simpler models rather than attempting ten way classification using circuits with larger numbers of virtual qubits.

Ultimately, we found that networks trained with the choices ($\lambda = .234$, $\eta = 5.59$, $a = 28.0$, $b = 33.0$, $A = 74.1$, $s = 4.13$, $t = .658$, $\gamma = 0.882$, $n = 222$) were able to achieve an average test accuracy above 95%. The accuracies of the individual pairwise classifiers are tabulated in Fig. 8.9, and data from a representative example of the training process for one of the easier pairs to classify is shown in 8.8. We observed significant differences in performance across the different pairs, partly owing, perhaps, to the difficulty of distinguishing similar digits using 64 pixel images. We also note that different choices of hyper-parameters could significantly affect which pairs were classified most accurately.

8.6 Implementation on Near-Term Devices

A key advantage of carrying out machine learning tasks with models equivalent to tree or matrix product tensor networks is that they could be implemented using a very small number of physical qubits. The key requirement is that the hardware must allow the measurement of individual physical qubits without further disturbing the state of the other qubits, a capability also required for certain approaches to quantum error correction [417]. For the case of discriminative learning, we note that explicitly measuring qubits which do not carry information about the label, rather than simply leaving these qubits unobserved, cannot affect the statistics of the final measurement averaged over multiple executions of the circuit. This is a consequence of the "no signaling" principle of quantum mechanics.

Below we will first discuss how the number of qubits needed to implement either a discriminative or generative tree tensor network model can be made to scale only logarithmically in both the data dimension and in the bond dimension of the network. Then we will discuss the special case of matrix product state tensor networks, which can be implemented with a number of physical qubits that is *independent* of the input or output data dimension.

Another key advantage of using tensor network models on near-term devices could be their robustness to noise, which will certainly be present in any near-term hardware. To explore the noise resilience of our models, we present a numerical experiment where we evaluate the model trained in Section 8.5 with random errors, and observe whether it can still produce useful results.

8.6.1 Qubit-Efficient Tree Network Models

To discuss the minimum qubit resources needed to implement general tree tensor network models, recall the notion of the virtual qubit number V from Section 8.4. This is the number of qubit lines connecting each subtree to higher nodes in the tree. Viewed as a tensor network, the bond dimension D, or dimension of the internal tensor indices, is given by $D = 2^V$.

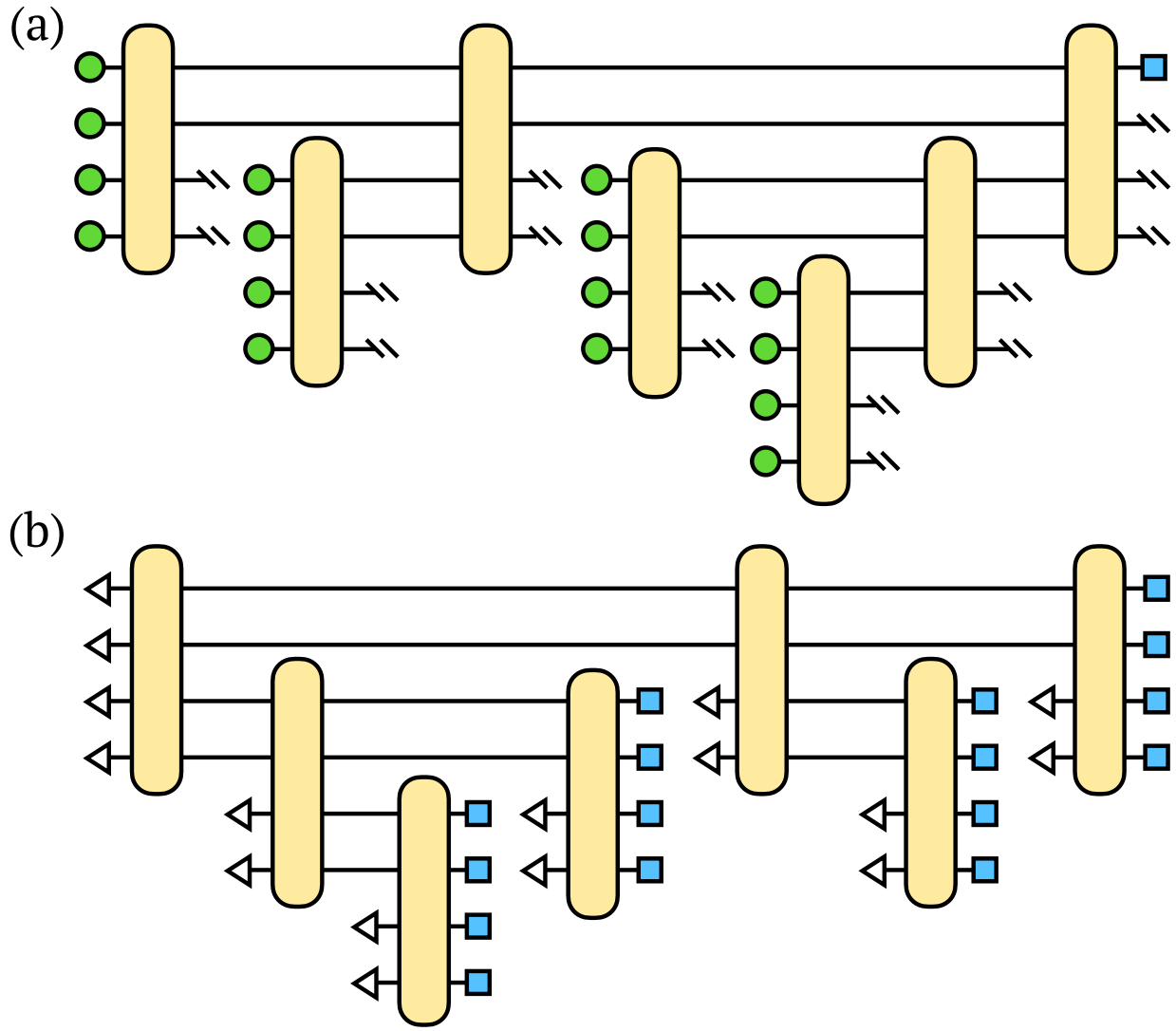

Figure 8.10: Qubit-efficient scheme for evaluating (a) discriminative and (b) generative tree models with $V = 2$ virtual qubits and $N = 16$ inputs or outputs. Note that the two patterns are the reverse of each other. In (a) qubits indicated with hash marks are measured and the measurement results discarded. These qubits are then reset and prepared with additional input states. In (b) measured qubits are recorded and reset to a reference state $\langle 0|$.

For example, the tree shown in Fig. 8.7 has $V = 1$ and a bond dimension of $D = 2$. The tree shown in Fig. 8.10 has $V = 2$ and $D = 4$. When discussing these models in general terms, it suffices to consider only unitary operations acting on $2V$ qubits, since at each node of the tree, two subtrees (two sets of V qubits) are entangled together.

Given only the ability to perform state preparation and unitary operations, it would take N physical qubits to evaluate a discriminative tree network model on N inputs. However, if we also allow the step of measurement and resetting of certain qubits, then the number of physical qubits Q required to process N inputs given V virtual states passing between each node can be significantly reduced to just $Q(N, V) = V \lg(2N/V)$.

To see why, consider the circuit showing the most qubit-efficient scheme for implementing the discriminative case Fig. 8.10(a). For a given V, the number of inputs that can be processed by a single unitary is $2V$. Then V of the qubits can be measured and reused, but the other V qubits must remain entangled. So only V new qubits must be introduced to process $2V$ more inputs. From this line of reasoning and the observation that $Q(2V, V) = 2V$, one can deduce the result $Q(N, V) = V \lg(2N/V)$.

For generative tree network models, generating N outputs with V virtual qubits requires the same number of physical qubits as for the discriminative case; this can be seen by observing that the pattern of unitaries is just the reverse of the discriminative case for the same N and V. Fig. 8.10 shows the most qubit-efficient way to sample a generative tree models for the case of $V = 2$ virtual and $N = 16$ output qubits, requiring only $Q = 8$ physical qubits.

Though a linear growth of the number of physical qubits as a function of virtual qubit number V may seem more prohibitive compared to the logarithmic scaling with N, even a small increase in V would lead to a significantly more expressive model. From the point of view of tensor networks the expressivity of the model is usually measured by the bond dimension $D = 2^V$. In terms of the bond dimension, the number of qubits needed thus scales only as $Q(N, D) \sim \lg(D) \lg(N)$. The largest bond dimensions used in state-of-the-art classical tensor network calculations are around $D = 2^{15}$ or about $30,000$. So for $V = 16$ or more virtual qubits one could quickly exceed the power of any classical tensor network calculation we are aware of, provided that parameterized unitaries involved could be efficiently implemented and optimized.

8.6.2 Qubit-Efficient Matrix Product Models

A matrix product state (MPS) tensor network is a special case of a tree tensor network that is maximally unbalanced. This gives an MPS certain advantages without sacrificing expressivity for one-dimensional distributions, as measured by the maximum entanglement entropy it can carry across bipartitions of the input or output space, meaning a division of $(x_1, \ldots, x_j)$ from $(x_{j+1}, \ldots, x_N)$.

Given the ability to measure and reset a subset of physical qubits, a key advantage of implementing a discriminative or generative tensor network model based on an MPS is that for a model with V virtual qubits, an *arbitrary* number of inputs or outputs can be processed

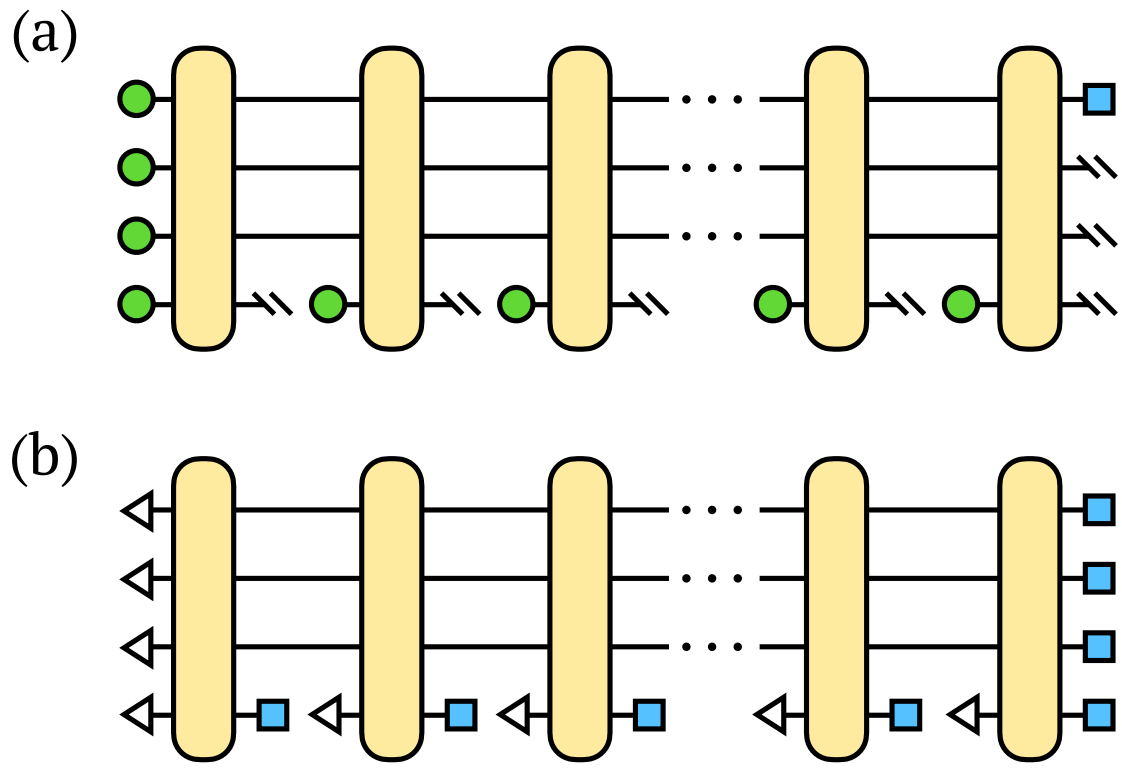

Figure 8.11: Qubit-efficient scheme for evaluating (a) discriminative and (b) generative matrix product state models for an arbitrary number of inputs or outputs. The figure shows the case of $V = 3$ qubits connecting each node of the network. When evaluating the discriminative model, one of the qubits is measured after each unitary is applied and the result discarded; the qubit is then prepared with the next input component. To implement the generative model, one of the qubits is measured after each unitary operation and the result recorded. The qubit is then reset to the state $\langle 0|$.

by using only $V+1$ physical qubits. The circuits illustrating how this can be done are shown in Fig. 8.11.

The implementation of the discriminative algorithm shown in Fig. 8.11(a) begins by preparing and entangling V input qubit states. One of the qubits is measured and reset to the next input state. Then all $V+1$ qubits are entangled and a single qubit measured and re-prepared. Continuing in this way, one can process all of the inputs. Once all inputs are processed, the model output is obtained by sampling one or more of the physical qubits.

To implement the generative MPS algorithm shown in Fig. 8.11(b), one prepares all qubits to a reference state $|0\rangle^{\otimes V+1}$ and after entangling the qubits, one measures and records a single qubit to generate the first output value. This qubit is reset to the state $|0\rangle$ and all the qubits are then acted on by another $(V+1)$ qubit unitary. A single qubit is again measured to generate the second output value, and the algorithm continues until N outputs have been generated.

To understand the equivalence of the generative circuit of Fig. 8.11(b) to conventional tensor diagram notation for an MPS, interpret the circuit diagram Fig. 8.12(a) as a tensor network diagram, treating elements such as reference states $\langle 0|$ as tensors or vectors $[1, 0]$. One can contract or sum over the reference state indices and merge any V qubit indices into a single index of dimension $D = 2^V$. The result is a standard MPS tensor network diagram Fig. 8.12(d) for the amplitude of observing a particular set of values of the measured qubits.

8.6.3 Noise Resilience

Any implementation of our proposed approach on near-term quantum hardware will have to contend with a significant level of noise due to qubit and gate imperfections. But one intuition about noise effects in our tree models is that an error which corrupts a qubit only scrambles the information coming from the patch of inputs belonging to the past "causal cone" of that qubit. And because the vast majority of the operations occur near the leaves of the tree, the most likely errors therefore correspond to scrambling only small patches of the input data. We note that a good classifier should naturally be robust to small deformations and corruptions of the input, and, in fact, adding various kinds of noise during training is a commonly used strategy in classical machine learning. Based on these intuitions, we expect our circuits could demonstrate a high level of tolerance to noise.

In order to quantitatively understand the robustness of our proposed approach to noise on quantum hardware, we study how performance is affected by independent amplitude-damping and dephasing channels applied to each qubit. In particular, we investigate how this error model would affect the pairwise tree network discriminative models of the type described in Section 8.4.1 and shown in Fig. 8.7.

The specific error model we implemented is the following: during the contraction step of node i in the model evaluation, we compose amplitude damping and dephasing noise channels acting on its left and right children ρ_{iL} and ρ_{iR}, mapping $\rho_{iL} \to \mathcal{E}_a(\mathcal{E}_d(\rho_{iL}))$ and $\rho_{iR} \to \mathcal{E}_a(\mathcal{E}_d(\rho_{iR}))$. Any completely positive trace-preserving noise channel $\mathcal{E}(\rho)$ can be expressed in the operator-sum representation as $\mathcal{E}(\rho) = \sum_a M_a \rho M_a^\dagger$, where $\sum_a M_a M_a^\dagger = I$.

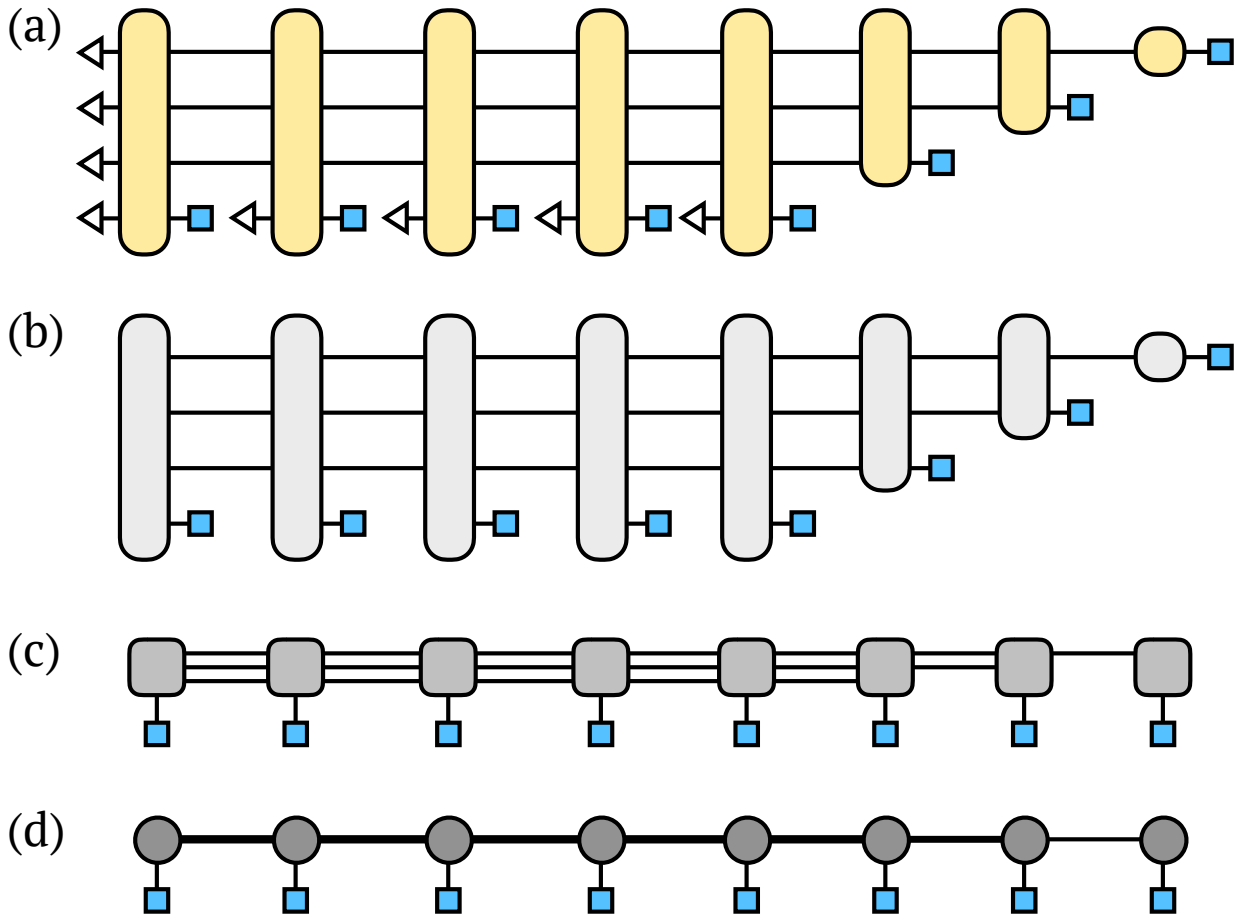

Figure 8.12: Mapping of the generative matrix product state (MPS) quantum circuit with $V = 3$ to a bond dimension $D = 2^3$ MPS tensor network diagram. First (a) interpret the circuit diagram as a tensor diagram by interpreting reference states $\langle 0|$ as vectors $[1, 0]$; qubit lines as dimension 2 tensor indices; and measurements as setting indices to fixed values. Then (b) contract the reference states into the unitary tensors and (c) redraw the tensors in a linear chain. Finally, (d) merge three $D = 2$ indices into a single $D = 8$ dimensional index on each bond.

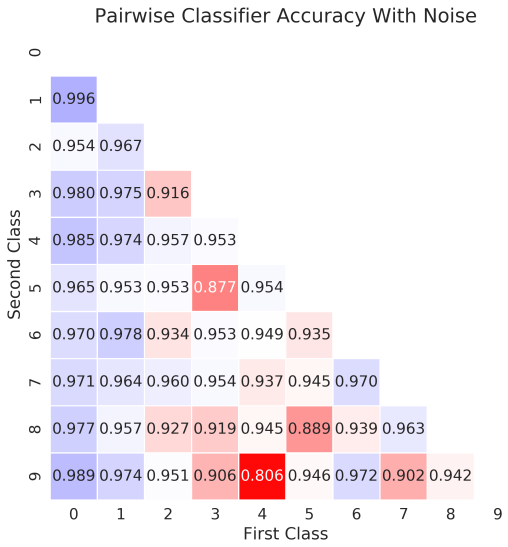

Figure 8.13: The test accuracy for each of the pairwise classifiers under noise corresponding to a T_1 of 5 μs, a T_2 of 7 μs, and a gate time of 200 ns. In most cases, the accuracy is comparable to the results from training without noise. Note that it was necessary to choose a different set of hyper-parameters to enable successful training under noise.

Here the Kraus operators M_a for the amplitude damping channel $\mathcal{E}_a$ are (in the z-basis)

$$M_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p_a} \end{pmatrix}, \quad M_1 = \begin{pmatrix} 0 & \sqrt{p_a} \\ 0 & 0 \end{pmatrix},$$

while for the dephasing channel $\mathcal{E}_d$ the Kraus operators are

$$M_0 = \sqrt{1-p_d}\, I, \quad M_1 = \begin{pmatrix} \sqrt{p_d} & 0 \\ 0 & 0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{p_d} \end{pmatrix}.$$

To evaluate model performance under realistic values of p_a and p_d on current hardware, we determine p_a and p_d based on the continuous-time Kraus operators of these channels, which depend on the duration of the two-qubit gate T_g, the coherence time T_1 of the qubits, and the dephasing time T_2 of the qubits. Specifically, $p_a = 1 - e^{-T_g/T_1}$ and $p_d = 1 - e^{-Tg/T_2}$. Realistic values for the time scales are $T_g = 200$ ns and $T_1 = 50\,\mu$s, $T_2 = 70\,\mu$s, corresponding to $p_a = 0.004$ and $p_d = 0.003$. But numerical experiments with these values showed almost no observable noise effects, so we consider an even more conservative parameter set with T_1 and T_2 reduced by an order of magnitude, such that $p_a = 0.039$ and $p_d = 0.028$.

We plot the resulting test accuracies in Fig. 8.13, noting that the Kraus operator formalism allows us to directly calculate the reduced density matrix of the labeling qubit under the

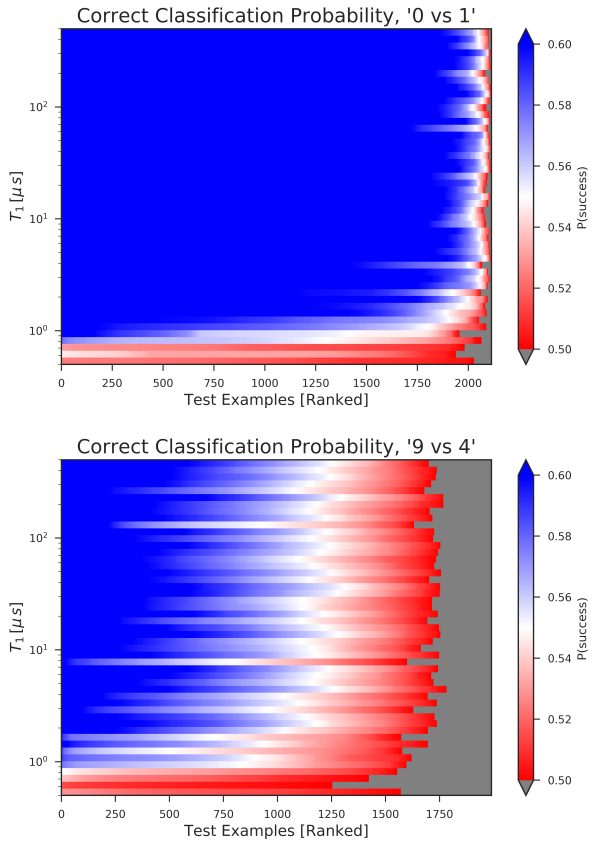

Figure 8.14: Success probability of two different pairwise classification circuits prediction on their test sets (sorted by decreasing probability of success along the x-axis) over a wide range of T_1 values (y-axis). For each T_1 shown, the probability of successfully classifying each member of the test set is indicated. Note that success probabilities which are larger than .5 even by a relatively small margin imply that the corresponding test example could be correctly classified with a majority voting scheme. Gate time $T_g = 200\,ns$ was held fixed while T_2 was set to be $\frac{7}{5}T_1$. Noise levels corresponding to current hardware are approximately two thirds of the way up the chart. Grey areas indicate regions where the model would misclassify the test example.

effects of our noise model, therefore no explicit sampling of noise realizations is needed. Given that the coherence times used for the plot are easily achievable even on today's very early hardware platforms, the results shown in Fig. 8.13 are encouraging: many of the models give a test accuracy only slightly reduced from the noiseless case Fig. 8.9. The largest reduction was for the digit '4' versus digit '9' model, which dropped from a test accuracy of 0.88 to 0.806. Interestingly this was also the model with the worst performance in the noiseless case. The typical change in test accuracy across all of the models due to the noise was about 0.004.

To mitigate the effect of noise when classifying a particular image, one can evaluate the quantum circuit some small number of times and choose the label which is most frequently observed. For example, one could take a majority vote from 500 executions and classify and correctly classify an image whose individual probability of success is .55 with almost 99% accuracy. In order to shed a more detailed light on our approach's robustness to noise, we plot in Fig. 8.14 the individual success probabilities for classifying each test example (x-axis), sorted by their probabilities for ease of visualization, over a range of decoherence times (y-axis). The two panels show two different models, one trained to distinguish images of digits '0' versus '1' ; the other digits '9' versus '4'. These models were trained and evaluated at various levels of noise using the same training hyper-parameters that were found to give a good performance at $\{p_a, p_d\} = \{0.039, 0.028\}$. The ratio between the dephasing time and the coherence time was held at a fixed ratio $T_2/T_1 = 7/5$.

We see that, for the examples which our models correctly classify in the low noise limit, the success probability remains appreciably greater than .5 for a wide range of noise levels. In both diagrams the y-axis is scaled so that coherence times achievable by today's hardware occur two thirds of the way up from the bottom. Interestingly, we note that the success probabilities saturate at coherence times much shorter than this, and only drop off dramatically at T_1 values near $T_1 \sim 1\,\mu\,s$. The high performance of our model over a broad swath of tested coherence and dephasing times suggests that the effects of noise on our approach can be dramatically mitigated by the combination of the hybrid quantum/classical training procedure and a small number of repetitions with a majority voting scheme. We find these results encouraging as empirical evidence that the limited-width "causal cone" structure possessed by models of this type may have inherent noise robustness properties.

8.7 Discussion

Many of the features that make tensor networks appealing for classical algorithms also make them a promising framework for quantum computing. Tensor networks provide a natural hierarchy of increasingly complex quantum states, allowing one to choose the appropriate amount of resources for a given task. They also enable specialized algorithms which can make efficient use of valuable resources, such as reducing the number of qubits needed to process high dimensional data. An optimized, classically tractable tensor network can be used to initialize the parameters of a more powerful model implemented on quantum hardware. Doing so would alleviate issues associated with random initial parameters, which can place

circuits in regions of parameter space with vanishing gradients [217].

While the approach to optimization we considered in our numerical experiments worked well, algorithms which are more specialized to the tensor network architecture could be devised. For example, by defining an objective for each subtree of a tree network it could be possible to train subtrees separately [402]. Likewise, the MPS architecture has certain orthogonality or light-cone properties which mean that only the tensors to the left of a certain physical index determine its distribution; this property could also be exploited for better optimization.

Another very interesting future direction would be to gain a better understanding of the noise resilience of tensor network machine learning algorithms. We performed some simple numerical experiments to show that these algorithms can tolerate a high level of noise, but additional empirical demonstrations as well as a theoretical explanation of how generic this property is would be very useful. In an interesting recent work, Kim and Swingle investigated tensor networks within a quantum computing framework for finding ground states of local Hamiltonians [227]. One of their results was a rigorous bound on the sensitivity of the algorithm output to noise, which relied on specific properties of tensor networks. It would be very interesting adapt their bound to the machine learning context.

Other tensor network architectures besides trees and MPS also deserve further investigation in the context of quantum algorithms. The PEPS family of tensor networks are specially designed to capture two-dimensional patterns of correlations [418, 419]. The MERA family of tensor networks, retain certain benefits of tree tensor networks but have more expressive power, and admit a natural description as a quantum circuit [227, 358].

Tensor networks strike a careful balance between expressive power and computational efficiency, and can be viewed as a particularly useful and natural class of quantum circuits. Based on the rich theoretical understanding of their properties and powerful algorithms for optimizing them, we are optimistic they will provide many interesting avenues for quantum machine learning research.

Chapter 9

Conclusion

This thesis has been concerned with the challenges facing variational quantum algorithms in the NISQ era. In Section 3.6, we introduced the notions of coherent time complexity and total time complexity. We explained how the number of two-qubit gates and the number of circuit repetitions serve as good proxies for these two kind of resources. We used these two notions of time complexity to frame our discussion of the difficulties that practitioners of near-term quantum algorithms encounter and some of the solutions that have been put forward. In the main body of the thesis, we laid out several of our previously published contributions in this direction. We now briefly review these contributions and explain explicitly how they fit into this framework of the two different kinds of resources.

We began by introducing the thesis itself in Chapter 1. Applications of near-term quantum computers to quantum chemistry are a central part of several of the later chapters. Because of this, in Chapter 2, we presented a pedagogical review of aspects of the electronic structure problem necessary to provide context for the rest of the thesis. In Chapter 3, we reviewed the formalism of quantum computing and the prospects of different hardware platforms for near-term quantum computing. We introduced the notion of variational quantum algorithms and reviewed the variational quantum eigensolver in detail, both as an example and to provide some background material to support the later chapters.

The main body of the thesis began in earnest with Chapter 4, where we explored several classes of unitary coupled cluster (UCC) wavefunction ansatze for performing quantum chemical calculations on a quantum computer. We made a number of contributions in this chapter, including providing numerical evidence for the utility of generalized excitation operators and benchmarking a straightforward method to calculate excited states. From the perspective of the two kinds of resources constraining NISQ algorithms, our most important contribution was the development of the $k-$fold unitary pair coupled cluster with generalized singles and doubles ($k-$UpCCGSD) ansatz. We showed that by using repeated applications of a sparse unitary coupled cluster operator we could achieve chemical accuracy on a number of model systems using fewer two-qubit gates and a shorter circuit depth than would be required with a more standard formulation of UCC. Stated more simply, the use of this ansatz reduces the coherent time complexity required for quantum chemical calculations.

In Chapter 5, we proposed a strategy for trading off between the coherent complexity of a variational ansatz and the overall number of measurements required. By measuring the matrix elements between a collection of parameterized wavefunctions and solving a small generalized eigenvalue problem on a classical computer, we were able to estimate the ground state energy of two model systems to within chemical accuracy. The technical tools we developed in this chapter minimized the cost of making these matrix element measurements and the overall framework we developed allows for the quality of the approximate ground state to be systematically increased by adding additional wavefunctions. Importantly, the number of two-qubit gates stays constant as additional wavefunctions are added. Therefore, at the cost of an increase in the total time complexity (number of overall circuit repetitions), the quality of the solution can be improved without using additional coherent quantum resources (two-qubit gates).

In Chapter 6, we showed how a tensor factorization of the quantum chemical Hamiltonian can dramatically reduce the number of circuit repetitions required to measure the energy within a fixed precision. This directly addresses the challenge posed by the high overall time complexity of variational quantum algorithms for quantum chemistry. In Table 6.3, we provide evidence that this reduction leads to a lower asymptotic scaling in the number of measurement repetitions required as well. In order to perform these more sophisticated energy measurements, our scheme uses some additional coherent quantum resources (quantified by the number of two-qubit gates). We explained how this tradeoff is more favorable than it might appear at first because performing measurements using our approach automatically reduces susceptibility to measurement errors and also enables another form of error mitigation based on efficient post-selection.

In Chapter 7, we presented some background on the tensor network formalism that has informed our understanding of quantum computing, along with an example of a classical calculation using this formalism with a similar tradeoff between two types of classical computational resources. We applied previously developed techniques to turn a deterministic tensor network algorithm for evaluating the partition function of the classical 2d Ising model into an unbiased Monte Carlo algorithm. This enabled us to have an additional refinement parameter for the calculation. The accuracy of the deterministic algorithm can be systematically improved by increasing the bond dimension, resulting in the additional use of serial computational resources. The Monte Carlo approach possess this same parameter, but its accuracy can also be improved by using additional parallel computational resources to generate additional samples for averaging. This echoes the tradeoff between coherent time complexity and overall time complexity that we address in the rest of the thesis.

In Chapter 8, we applied ideas from the tensor network literature to design a strategy for machine learning with a NISQ computer. In part, this chapter addresses the challenges of the limited resources available in the NISQ era by choosing an application that does not demand high accuracy. By using our parameterized circuit to learn a qualitative feature of the data, we were able to tolerate significant errors, both arising from the operation of the gates and from the statistical noise associated with a finite number of measurements. We also provided some arguments and numerical evidence that the particular structure of our circuits should

be especially resilient to noise for the machine learning application we considered. We also showed how our proposal could reduce the requirements for a third resource that we have mostly set aside throughout this thesis, the number of qubits.

Ultimately, the work of this thesis constitutes a few small pieces of the ongoing effort to make variational quantum algorithms practically useful. We've made an attempt to cite the foundational works in this subfield along with those that have shaped our thinking, but we have surely neglected many. A tremendous amount of progress has happened during the last few years. The list of new techniques and algorithms for use with near-term quantum computers is manifold, and any exclusions on our part are unintentional. It remains an open question whether or not NISQ quantum computing will provide a benefit over classical techniques for problems of interest outside of the quantum computing community but surely there is much to learn in the attempt.

It is our belief that the most promising path towards answering this question is to keep experimenting. Future developments in algorithms, error mitigation, and optimization will all be necessary. Parallel to these efforts, as the hardware continues to improve, existing approaches may become more viable. We look forward to seeing what the future holds, and hope that this thesis is an interesting and helpful window into the present state of the field.

Bibliography

[1]R. P. Feynman, "Simulating physics with computers", Int. J. Theor. Phys. **21**, 467–488 (1982).

[2]S. Lloyd, "Universal quantum simulators", Science **273**, 1073–1078 (1996).

[3]P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring", in Proceedings 35th annual symposium on foundations of computer science (Nov. 1994), pp. 124–134.

[4]L. K. Grover, "A fast quantum mechanical algorithm for database search", in Proceedings of the twenty-eighth annual ACM symposium on theory of computing, STOC '96 (July 1996), pp. 212–219.

[5]L. K. Grover, "From schrödinger's equation to the quantum search algorithm", Pramana **56**, 333–348 (2001).

[6]E. Bernstein and U. Vazirani, "Quantum complexity theory", SIAM J. Comput. **26**, 1411–1473 (1997).

[7]F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum supremacy using a programmable superconducting processor", Nature **574**, 505–510 (2019).

[8]D. Aharonov and M. Ben-Or, "Fault-tolerant quantum computation with constant error", in Proceedings of the twenty-ninth annual ACM symposium on theory of computing, STOC '97 (May 1997), pp. 176–188.

[9]D. Gottesman, "An introduction to quantum error correction and Fault-Tolerant quantum computation", arXiv:0904.2557 (2009).

[10]J. Preskill, "Quantum computing in the NISQ era and beyond", Quantum **2**, 79 (2018).

[11]D. Litinski, "A game of surface codes: large-scale quantum computing with lattice surgery", Quantum **3**, 128 (2019).

[12]D. Aharonov, "Quantum to classical phase transition in noisy quantum computers", Phys. Rev. A **62**, 062311 (2000).

[13]E. Pednault, J. A. Gunnels, G. Nannicini, L. Horesh, and R. Wisnieff, "Leveraging secondary storage to simulate deep 54-qubit sycamore circuits", arXiv:1910.09534 (2019).

[14]H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, "Probing many-body dynamics on a 51-atom quantum simulator", Nature **551**, 579–584 (2017).

[15]S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, "Characterizing quantum supremacy in near-term devices", Nat. Phys. **14**, 595–600 (2018).

[16]S. Aaronson and L. Chen, "Complexity-Theoretic foundations of quantum supremacy experiments", arXiv:1612.05903 (2016).

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

[18]D. Wecker, M. B. Hastings, and M. Troyer, "Progress towards practical quantum variational algorithms", Phys. Rev. A **92**, 042303 (2015).

[19]E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm", arXiv:1411.4028 (2014).

[20]E. Farhi and H. Neven, "Classification with quantum neural networks on near term processors", arXiv:1802.06002 (2018).

[21]P. A. M. Dirac and R. H. Fowler, "Quantum mechanics of many-electron systems", Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **123**, 714–733 (1929).

[22]W. D. Loveland, D. J. Morrissey, and G. T. Seaborg, *Modern nuclear chemistry* (John Wiley & Sons, Apr. 2017).

[23]R. Shankar, *Principles of quantum mechanics* (Springer Science & Business Media, Dec. 2012).

[24]A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory* (Courier Corporation, June 2012).

[25]G. A. Worth and L. S. Cederbaum, "Beyond Born-Oppenheimer: molecular dynamics through a conical intersection", Annu. Rev. Phys. Chem. **55**, 127–158 (2004).

[26]R. Mota, R. Parafita, A. Giuliani, M.-J. Hubin-Franskin, J. M. C. Lourenço, G. Garcia, S. V. Hoffmann, N. J. Mason, P. A. Ribeiro, M. Raposo, and P. Limão-Vieira, "Water VUV electronic state spectroscopy by synchrotron radiation", Chem. Phys. Lett. **416**, 152–159 (2005).

[27]P. W. Atkins and R. S. Friedman, *Molecular quantum mechanics* (OUP Oxford, 2011).

[28]A. D. McNaught, A. Wilkinson, et al., *Compendium of chemical terminology*, Vol. 1669 (Blackwell Science Oxford, 1997).

[29]J. I. Steinfeld, J. S. Francisco, and W. L. Hase, *Chemical kinetics and dynamics*, Vol. 3 (Prentice Hall Englewood Cliffs (New Jersey), 1989).

[30]W. Pauli, "The connection between spin and statistics", Phys. Rev. **58**, 716–722 (1940).

[31]D. J. Griffiths and D. F. Schroeter, *Introduction to quantum mechanics* (Cambridge University Press, Aug. 2018).

[32]J. C. Slater, "The theory of complex spectra", Phys. Rev. **34**, 1293–1322 (1929).

[33]T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular Electronic-Structure theory* (John Wiley & Sons, Aug. 2014).

[34]R. M. Martin, *Electronic structure: basic theory and practical methods* (Cambridge University Press, Aug. 2020).

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

[36]J. J. Sakurai and A. W. Longman, *Quantum mechanics* (Addison-Wesley, 1976).

[37]B. O. Roos, P. R. Taylor, and P. E. M. Sigbahn, "A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach", Chem. Phys. **48**, 157–173 (1980).

[38]B. O. Roos, P. Linse, P. E. M. Siegbahn, and M. R. A. Blomberg, "A simple method for the evaluation of the second-order-perturbation energy from external double-excitations with a CASSCF reference wavefunction", Chem. Phys. **66**, 197–207 (1982).

[39]J. A. Pople, J. S. Binkley, and R. Seeger, "Theoretical models incorporating electron correlation", Int. J. Quantum Chem., NBS Circular 467 **10**, 1–19 (1976).

[40]J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, "Electron correlation theories and their application to the study of simple reaction potential surfaces", Int. J. Quantum Chem. **14**, 545–560 (1978).

[41]T. Van Voorhis and M. Head-Gordon, "Benchmark variational coupled cluster doubles results", J. Chem. Phys. **113**, 8873–8879 (2000).

[42]Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package", Mol. Phys. **113**, 184–215 (2015).

[43]G. E. Scuseria and P. Y. Ayala, "Linear scaling coupled cluster and perturbation theories in the atomic orbital basis", J. Chem. Phys. **111**, 8330–8343 (1999).

[44]C. Riplinger and F. Neese, "An efficient and near linear scaling pair natural orbital based local coupled cluster method", J. Chem. Phys. **138**, 034106 (2013).

[45]M. Schütz and H.-J. Werner, "Low-order scaling local electron correlation methods. IV. linear scaling local coupled-cluster (LCCSD)", J. Chem. Phys. **114**, 661–681 (2001).

[46]C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, and F. Neese, "Sparse maps–a systematic infrastructure for reduced-scaling electronic structure methods. II. linear scaling domain based pair natural orbital coupled cluster theory", J. Chem. Phys. **144**, 024109 (2016).

[47]Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki, and M. Kállay, "An efficient linear-scaling CCSD(T) method based on local natural orbitals", J. Chem. Phys. **139**, 094105 (2013).

[48]J. Lee, D. W. Small, and M. Head-Gordon, "Open-shell coupled-cluster valence-bond theory augmented with an independent amplitude approximation for three-pair correlations:

application to a model oxygen-evolving complex and single molecular magnet", J. Chem. Phys. **149**, 244121 (2018).

[49]J. Lee, D. W. Small, E. Epifanovsky, and M. Head-Gordon, "Coupled-Cluster Valence-Bond Singles and Doubles for Strongly Correlated Systems: Block-Tensor Based Implementation and Application to Oligoacenes", J. Chem. Theory Comput. **13**, 602–615 (2017).

[50]T. Stein, T. M. Henderson, and G. E. Scuseria, "Seniority zero pair coupled cluster doubles theory", J. Chem. Phys. **140**, 214113 (2014).

[51]J. A. Gaunt, "A theory of hartree's atomic fields", Math. Proc. Cambridge Philos. Soc. **24**, 328–342 (1928).

[52]J. C. Slater, "The self consistent field and the structure of atoms", Phys. Rev. **32**, 339–348 (1928).

[53]D. R. Hartree, "The wave mechanics of an atom with a Non-Coulomb central field. part II. some results and discussion", Math. Proc. Cambridge Philos. Soc. **24**, 111–132 (1928).

[54]C. Møller and M. S. Plesset, "Note on an approximation treatment for Many-Electron systems", Phys. Rev. **46**, 618–622 (1934).

[55]F. Coester and H. Kümmel, "Short-range correlations in nuclear wave functions", Phys. Rev. C Nucl. Phys. **17**, 477–485 (1960).

[56]J. Čıžek, "On the correlation problem in atomic and molecular systems. calculation of wavefunction components in Ursell-Type expansion using Quantum-Field theoretical methods", J. Chem. Phys. **45**, 4256–4266 (1966).

[57]J. Čıžek and J. Paldus, "Correlation problems in atomic and molecular systems III. rederivation of the coupled-pair many-electron theory using the traditional quantum chemical methodst", Int. J. Quantum Chem. **5**, 359–379 (1971).

[58]P. Hohenberg and W. Kohn, "Inhomogeneous electron gas", Phys. Rev. **136**, B864–B871 (1964).

[59]W. Kohn and L. J. Sham, "Self-Consistent equations including exchange and correlation effects", Phys. Rev. **140**, A1133–A1138 (1965).

[60]N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals", Mol. Phys. **115**, 2315–2372 (2017).

[61]D. M. Ceperley and B. J. Alder, "Ground state of the electron gas by a stochastic method", Phys. Rev. Lett. **45**, 566–569 (1980).

[62]P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, "Fixed-node quantum monte carlo for moleculesa) b)", J. Chem. Phys. **77**, 5593–5603 (1982).

[63]J. Toulouse, R. Assaraf, and C. J. Umrigar, "Introduction to the variational and diffusion monte carlo methods", in *Advances in quantum chemistry*, Vol. 73 (Elsevier, 2016), pp. 285–314.

[64]J. E. Hirsch, "Discrete Hubbard-Stratonovich transformation for fermion lattice models", Phys. Rev. B Condens. Matter **28**, 4059–4061 (1983).

[65]S. Zhang and H. Krakauer, "Quantum monte carlo method using phase-free random walks with slater determinants", Phys. Rev. Lett. **90**, 136401 (2003).

[66]M. Motta and S. Zhang, "Ab initio computations of molecular systems by the auxiliary-field quantum monte carlo method", WIREs Comput Mol Sci, Dover Books on Physics **8**, e1364 (2018).

[67]S. R. White, "Density matrix formulation for quantum renormalization groups", Phys. Rev. Lett. **69**, 2863–2866 (1992).

[68]S. R. White and R. L. Martin, "Ab initio quantum chemistry using the density matrix renormalization group", J. Chem. Phys. **110**, 4127–4130 (1999).

[69]R. Olivares-Amaya, W. Hu, N. Nakatani, S. Sharma, J. Yang, and G. K.-L. Chan, "The ab-initio density matrix renormalization group in practice", J. Chem. Phys. **142**, 034102 (2015).

[70]Simons Collaboration on the Many-Electron Problem, M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, A. J. Millis, N. V. Prokof'ev, U. Ray, G. E. Scuseria, S. Sorella, E. M. Stoudenmire, Q. Sun, I. S. Tupitsyn, S. R. White, D. Zgid, and S. Zhang, "Towards the solution of the Many-Electron problem in real materials: equation of state of the hydrogen chain with State-of-the-Art Many-Body methods", Phys. Rev. X **7**, 031059 (2017).

[71]Simons Collaboration on the Many-Electron Problem, J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, E. Kozik, X.-W. Liu, A. J. Millis, N. V. Prokof'ev, M. Qin, G. E. Scuseria, H. Shi, B. V. Svistunov, L. F. Tocchio, I. S. Tupitsyn, S. R. White, S. Zhang, B.-X. Zheng, Z. Zhu, and E. Gull, "Solutions of the Two-Dimensional hubbard model: benchmarks and results from a wide range of numerical algorithms", Phys. Rev. X **5**, 041041 (2015).

[72]J. Greeley, J. K. Nørskov, and M. Mavrikakis, "Electronic structure and catalysis on metal surfaces", Annu. Rev. Phys. Chem. **53**, 319–348 (2002).

[73]A. Veillard, *Quantum chemistry: the challenge of transition metals and coordination chemistry* (Springer Science & Business Media, Dec. 2012).

[74]H. Rabitz, de Vivie-Riedle R, M. Motzkus, and K. Kompa, "Whither the future of controlling quantum phenomena?", Science **288**, 824–828 (2000).

[75]J. Flick, N. Rivera, and P. Narang, "Strong light-matter coupling in quantum chemistry and quantum photonics", Nanophotonics **7**, 1479–1501 (2018).

[76]A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, "Simulated quantum computation of molecular energies", Science **309**, 1704–1707 (2005).

[77]D. W. Berry, C. Gidney, M. Motta, J. R. McClean, and R. Babbush, "Qubitization of arbitrary basis quantum chemistry leveraging sparsity and low rank factorization", Quantum **3**, 208 (2019).

[78]A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, "Toward the first quantum simulation with quantum speedup", Proc. Natl. Acad. Sci. U. S. A. **115**, 9456–9461 (2018).

[79]R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, "Encoding electronic spectra in quantum circuits with linear T complexity", Phys. Rev. X **8**, 041015 (2018).

[80]A. Baiardi and M. Reiher, "Large-Scale quantum dynamics with matrix product states", J. Chem. Theory Comput. **15**, 3481–3498 (2019).

[81]M. R. Provorse and C. M. Isborn, "Electron dynamics with real-time time-dependent density functional theory", Int. J. Quantum Chem. **116**, 739–749 (2016).

[82]G. Cohen, E. Gull, D. R. Reichman, and A. J. Millis, "Taming the dynamical sign problem in Real-Time evolution of quantum Many-Body problems", Phys. Rev. Lett. **115**, 266802 (2015).

[83]R. Lin, P. Molignini, L. Papariello, M. C. Tsatsos, C. Lévêque, S. E. Weiner, E. Fasshauer, R. Chitra, and A. U. J. Lode, "MCTDH-X: the multiconfigurational time-dependent hartree method for indistinguishable particles software", arXiv:1911.00525 (2019).

[84]A. Yu. Kitaev, "Quantum measurements and the abelian stabilizer problem", arXiv:quant-ph/9511026 (1995).

[85]D. Deutsch and R. Penrose, "Quantum theory, the Church–Turing principle and the universal quantum computer", Proc. R. Soc. Lond. A Math. Phys. Sci. **400**, 97–117 (1985).

[86]M. A. Nielsen and I. L. Chuang, "Quantum information and quantum computation", Cambridge: Cambridge University Press **2**, 23 (2000).

[87]N. M. Tubman, C. Mejuto-Zaera, J. M. Epstein, et al., "Postponing the orthogonality catastrophe: efficient state preparation for electronic structure simulations on quantum devices", arXiv:1809.05523 (2018).

[88]M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, "Elucidating reaction mechanisms on quantum computers", Proc. Natl. Acad. Sci. U. S. A. **114**, 7555–7560 (2017).

[89]M. Mitrano, A. Cantaluppi, D. Nicoletti, S. Kaiser, A. Perucchi, S. Lupi, P. Di Pietro, D. Pontiroli, M. Riccò, S. R. Clark, D. Jaksch, and A. Cavalleri, "Possible light-induced superconductivity in K3C60 at high temperature", Nature **530**, 461–464 (2016).

[90]J. I. Cirac and P. Zoller, "Goals and opportunities in quantum simulation", Nat. Phys. **8**, 264–266 (2012).

[91]J. Du, N. Xu, X. Peng, P. Wang, S. Wu, and D. Lu, "NMR implementation of a molecular hydrogen quantum simulation with adiabatic state preparation", Phys. Rev. Lett. **104**, 030502 (2010).

[92]A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, "Surface codes: towards practical large-scale quantum computation", Phys. Rev. A **86**, 032324 (2012).

[93]K. X. Wei, I. Lauer, S. Srinivasan, N. Sundaresan, D. T. McClure, D. Toyli, D. C. McKay, J. M. Gambetta, and S. Sheldon, "Verifying multipartite entangled Greenberger-Horne-Zeilinger states via multiple quantum coherences", Phys. Rev. A **101**, 032343 (2020).

[94]Y. Nam, J.-S. Chen, N. C. Pisenti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, J. Apisdorf, K. M. Beck, A. Blinov, V. Chaplin, M. Chmielewski, C. Collins, S. Debnath, K. M. Hudek, A. M. Ducore, M. Keesan, S. M. Kreikemeier, J. Mizrahi, P. Solomon, M. Williams, J. D. Wong-Campos, D. Moehring, C. Monroe, and J. Kim, "Ground-state energy estimation of the water molecule on a trapped-ion quantum computer", npj Quantum Inf **6**, 303 (2020).

[95]D. E. Deutsch and R. Penrose, "Quantum computational networks", Proc. R. Soc. Lond. A Math. Phys. Sci. **425**, 73–90 (1989).

[96]D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, "Adiabatic quantum computation is equivalent to standard quantum computation", arXiv:quant-ph/0405098 (2004).

[97]R. Raussendorf and H. J. Briegel, "A one-way quantum computer", Phys. Rev. Lett. **86**, 5188–5191 (2001).

[98]A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, "Elementary gates for quantum computation", Phys. Rev. A **52**, 3457–3467 (1995).

[99]A. Yu Kitaev, "Quantum computations: algorithms and error correction", Russian Math. Surveys **52**, 1191 (1997).

[100]A. Y. Kitaev, "Quantum computations: algorithms and error correction", Russian Math. Surveys **52**, 1191 (1997).

[101]A. W. Harrow, B. Recht, and I. L. Chuang, "Efficient discrete approximations of quantum gates", arXiv:quant-ph/0111031 (2001).

[102]C. M. Dawson and M. A. Nielsen, "The Solovay-Kitaev algorithm", arXiv:quant-ph/0505030 (2005).

[103]M. Suzuki, "General theory of fractal path integrals with applications to many-body theories and statistical physics", J. Math. Phys. **32**, 400–407 (1991).

[104]N. Hatano and M. Suzuki, "Finding exponential product formulas of higher orders", in *Quantum annealing and other optimization methods*, edited by A. Das and B. K. Chakrabarti (Springer Berlin Heidelberg, Berlin, Heidelberg, 2005), pp. 37–68.

[105]A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, "Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics", in Proceedings of the 51st annual ACM SIGACT symposium on theory of computing, STOC 2019 (June 2019), pp. 193–204.

[106]G. H. Low and I. L. Chuang, "Hamiltonian simulation by qubitization", Quantum **3**, 163 (2019).

[107]A. M. Childs and N. Wiebe, "Hamiltonian simulation using linear combinations of unitary operations", arXiv:1202.5822 (2012).

[108]G. H. Low and I. L. Chuang, "Optimal hamiltonian simulation by quantum signal processing", Phys. Rev. Lett. **118**, 010501 (2017).

[109]A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman, "Exponential algorithmic speedup by a quantum walk", in Proceedings of the thirty-fifth annual ACM symposium on theory of computing, STOC '03 (June 2003), pp. 59–68.

[110]E. Knill, G. Ortiz, and R. D. Somma, "Optimal quantum measurements of expectation values of observables", Phys. Rev. A **75**, 012328 (2007).

[111]A. Asfaw, L. Bello, Y. Ben-Haim, S. Bravyi, L. Capelluto, A. C. Vazquez, J. Ceroni, J. Gambetta, S. Garion, L. Gil, et al., "Learn quantum computation using qiskit", (2019).

[112]D. Coppersmith, "An approximate fourier transform useful in quantum factoring", arXiv:quant-ph/0201067 (1994).

[113]L. Hales and S. Hallgren, "An improved quantum fourier transform algorithm and applications", in Proceedings 41st annual symposium on foundations of computer science (Nov. 2000), pp. 515–525.

[114]K. M. Svore, M. B. Hastings, and M. Freedman, "Faster phase estimation", Quantum Inf. Comput. **14**, 306–328 (2014).

[115]M. Zwierz, C. A. Pérez-Delgado, and P. Kok, "General optimality of the heisenberg limit for quantum metrology", Phys. Rev. Lett. **105**, 180402 (2010).

[116]M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, "Superconducting qubits: current state of play", Annu. Rev. Condens. Matter Phys. **11**, 369–395 (2020).

[117]D. Kielpinski, C. Monroe, and D. J. Wineland, "Architecture for a large-scale ion-trap quantum computer", Nature **417**, 709–711 (2002).

[118]C. Müller, J. Lisenfeld, A. Shnirman, and S. Poletto, "Interacting two-level defects as sources of fluctuating high-frequency noise in superconducting circuits", Phys. Rev. B Condens. Matter **92**, 035442 (2015).

[119]Z. Chen, J. Kelly, C. Quintana, R. Barends, B. Campbell, Y. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Lucero, E. Jeffrey, A. Megrant, J. Mutus, M. Neeley, C. Neill, P. J. J. O'Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Korotkov, and J. M. Martinis, "Measuring and suppressing quantum state leakage in a superconducting qubit", Phys. Rev. Lett. **116**, 020501 (2016).

[120]J. Ghosh, A. G. Fowler, J. M. Martinis, and M. R. Geller, "Understanding the effects of leakage in superconducting quantum-error-detection circuits", Phys. Rev. A **88**, 062329 (2013).

[121]C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, "Trapped-ion quantum computing: progress and challenges", Applied Physics Reviews **6**, 021314 (2019).

[122]C. Monroe, R. Raussendorf, A. Ruthven, K. R. Brown, P. Maunz, L.-M. Duan, and J. Kim, "Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects", Phys. Rev. A **89**, 022317 (2014).

[123]E. Knill, R. Laflamme, and G. J. Milburn, "A scheme for efficient quantum computation with linear optics", Nature **409**, 46–52 (2001).

[124]P. Walther, K. J. Resch, T. Rudolph, E. Schenck, H. Weinfurter, V. Vedral, M. Aspelmeyer, and A. Zeilinger, "Experimental one-way quantum computing", Nature **434**, 169–176 (2005).

[125]S. Slussarenko and G. J. Pryde, "Photonic quantum information processing: a concise review", Applied Physics Reviews **QUC2019**, 041303 (2019).

[126]G.-Q. Liu and X.-Y. Pan, "Quantum information processing with nitrogen–vacancy centers in diamond*", Chin. Physics B **27**, 020304 (2018).

[127]M. Saffman, "Quantum computing with atomic qubits and rydberg interactions: progress and challenges", J. Phys. B At. Mol. Opt. Phys. **49**, 202001 (2016).

[128]D. Aasen, M. Hell, R. V. Mishmash, A. Higginbotham, J. Danon, M. Leijnse, T. S. Jespersen, J. A. Folk, C. M. Marcus, K. Flensberg, and J. Alicea, "Milestones toward Majorana-Based quantum computing", Phys. Rev. X **6**, 031016 (2016).

[129]Warren S. Warren, N. Gershenfeld, and I. Chuang, "The usefulness of NMR quantum computing", Science **277**, 1688–1690 (1997).

[130]R. Maurand, X. Jehl, D. Kotekar-Patil, A. Corna, H. Bohuslavskyi, R. Laviéville, L. Hutin, S. Barraud, M. Vinet, M. Sanquer, and S. De Franceschi, "A CMOS silicon spin qubit", Nat. Commun. **7**, 13575 (2016).

[131]G. Adesso, S. Ragy, and A. R. Lee, "Continuous variable quantum information: gaussian states and beyond", arXiv:1401.4679 (2014).

[132]Sah Chih-Tang, "Evolution of the MOS transistor-from conception to VLSI", Proc. IEEE **76**, 1280–1326 (1988).

[133]S. J. Devitt, W. J. Munro, and K. Nemoto, "Quantum error correction for beginners", Rep. Prog. Phys. **76**, 076001 (2013).

[134]L. Viola, E. Knill, and S. Lloyd, "Dynamical decoupling of open quantum systems", Phys. Rev. Lett. **82**, 2417–2421 (1999).

[135]M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, "Optimized dynamical decoupling in a model quantum memory", Nature **458**, 996–1000 (2009).

[136]Y. R. Sanders, J. J. Wallman, and B. C. Sanders, "Bounding quantum gate error rate based on reported average fidelity", New J. Phys. **18**, 012002 (2015).

[137]J. Preskill, "Lecture notes for physics 229: quantum information and computation", (1998).

[138]G. G. Carlo, G. Benenti, G. Casati, and C. Mejıa-Monasterio, "Simulating noisy quantum protocols with quantum trajectories", Phys. Rev. A **69**, 062317 (2004).

[139]M. Gutiérrez and K. R. Brown, "Comparison of a quantum error-correction threshold for exact and approximate errors", Phys. Rev. A **91**, 022335 (2015).

[140]J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms", New J. Phys. **18**, 023023 (2016).

[141]J. J. Wallman and J. Emerson, "Noise tailoring for scalable quantum computation via randomized compiling", Phys. Rev. A **94**, 052325 (2016).

[142]F. B. Maciejewski, Z. Zimborás, and M. Oszmaniec, "Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography", arXiv:1907.08518 (2019).

[143]M. A. Fogarty, M. Veldhorst, R. Harper, C. H. Yang, S. D. Bartlett, S. T. Flammia, and A. S. Dzurak, "Nonexponential fidelity decay in randomized benchmarking with low-frequency noise", Phys. Rev. A **92**, 022326 (2015).

[144]P. V. Klimov, J. Kelly, Z. Chen, M. Neeley, A. Megrant, B. Burkett, R. Barends, K. Arya, B. Chiaro, Y. Chen, A. Dunsworth, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, T. Huang, E. Jeffrey, E. Lucero, J. Y. Mutus, O. Naaman, C. Neill, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, S. Boixo, R. Babbush, V. N. Smelyanskiy, H. Neven, and J. M. Martinis, "Fluctuations of Energy-Relaxation times in superconducting qubits", Phys. Rev. Lett. **121**, 090502 (2018).

[145]V. V. Ramasesh, "Quantum simulation with superconducting circuits", PhD thesis (UC Berkeley, 2019).

[146]J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, "Computation of molecular spectra on a quantum processor with an Error-Resilient algorithm", Phys. Rev. X **8**, 011021 (2018).

[147]E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, "Randomized benchmarking of quantum gates", Phys. Rev. A **77**, 012307 (2008).

[148]P. W. Shor, "Scheme for reducing decoherence in quantum computer memory", Phys. Rev. A **52**, R2493–R2496 (1995).

[149]P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, "Scalable Quantum Simulation of Molecular Energies", Phys. Rev. X **6**, 031007 (2016).

[150]J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven, "Decoding quantum errors with subspace expansions", Nat. Commun. **11**, 636 (2020).

[151]K. Sharma, S. Khatri, M. Cerezo, and P. J. Coles, "Noise resilience of variational quantum compiling", New J. Phys. **22**, 043006 (2020).

[152]C. Durr and P. Hoyer, "A quantum algorithm for finding the minimum", arXiv:quant-ph/9607014 (1996).

[153]E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, "Quantum computation by adiabatic evolution", arXiv:quant-ph/0001106 (2000).

[154]K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning", Phys. Rev. A **98**, 032309 (2018).

[155]M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, "Parameterized quantum circuits as machine learning models", Quantum Sci. Technol. **4**, 043001 (2019).

[156]M. Schuld and N. Killoran, "Quantum machine learning in feature hilbert spaces", Phys. Rev. Lett. **122**, 040504 (2019).

[157]I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning* (MIT Press, Nov. 2016).

[158]Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", Nature **521**, 436–444 (2015).

[159]W. Huggins, P. Patil, B. Mitchell, K. B. Whaley, et al., "Towards quantum machine learning with tensor networks", Quantum Science (2019).

[160]Y. Li and S. C. Benjamin, "Efficient variational quantum simulator incorporating active error minimization", Phys. Rev. X **7**, 021050 (2017).

[161]T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo, and L. Visscher, "Calculating energy derivatives for quantum chemistry on a quantum computer", npj Quantum Information **5**, 113 (2019).

[162]S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, "Quantum-assisted quantum compiling", Quantum **3**, 140 (2019).

[163]S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, "Quantum computational chemistry", Rev. Mod. Phys. **92**, 015003 (2020).

[164]Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, "Quantum chemistry in the age of quantum computing", Chem. Rev. **119**, 10856–10915 (2019).

[165]A. Y. Kitaev, A. Shen, M. N. Vyalyi, and M. N. Vyalyi, *Classical and quantum computation* (American Mathematical Soc., 2002).

[166]B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, "Quantum algorithms for quantum chemistry and quantum materials science", arXiv:2001.03685 (2020).

[167]P. Jordan and E. Wigner, "Über das paulische äquivalenzverbot", Z. Angew. Phys. **47**, 631 (1928).

[168]G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme, "Quantum algorithms for fermionic simulations", Phys. Rev. A **64**, 022319 (2001).

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

[170]I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, "Quantum simulation of electronic structure with linear depth and connectivity", Phys. Rev. Lett. **120**, 110501 (2018).

[171]F. Verstraete, J. I. Cirac, and J. I. Latorre, "Quantum circuits for strongly correlated quantum systems", Phys. Rev. A **79**, 032316 (2009).

[172]B. O'Gorman, W. J. Huggins, E. G. Rieffel, and K. Birgitta Whaley, "Generalized swap networks for near-term quantum computing", arXiv:1905.05118 (2019).

[173]D. J. Thouless, "Stability conditions and nuclear rotations in the Hartree-Fock theory", Phys. Rev. C Nucl. Phys. **21**, 225–232 (1960).

[174]Z. Jiang, K. J. Sung, K. Kechedzhi, V. N. Smelyanskiy, and S. Boixo, "Quantum algorithms to simulate Many-Body physics of correlated fermions", Phys. Rev. Applied **9**, 044036 (2018).

[175]M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush, and G. K.-L. Chan, "Low rank representations for quantum simulation of electronic structure", arXiv:1808.02625 (2018).

[176]S. B. Bravyi and A. Y. Kitaev, "Fermionic quantum computation", Ann. Phys. **298**, 210–226 (2002).

[177]J. T. Seeley, M. J. Richard, and P. J. Love, "The Bravyi-Kitaev transformation for quantum computation of electronic structure", J. Chem. Phys. **137**, 224109 (2012).

[178]M. Steudtner and S. Wehner, "Fermion-to-qubit mappings with varying resource requirements for quantum simulation", New J. Phys. **20**, 063010 (2018).

[179]K. Setia and J. D. Whitfield, "Bravyi-Kitaev superfast simulation of electronic structure on a quantum computer", J. Chem. Phys. **148**, 164104 (2018).

[180]K. Setia, S. Bravyi, A. Mezzacapo, and J. D. Whitfield, "Superfast encodings for fermionic quantum simulation", Phys. Rev. Research, 033033 (2019).

[181]Z. Jiang, J. McClean, R. Babbush, and H. Neven, "Majorana loop stabilizer codes for error mitigation in fermionic quantum simulations", Phys. Rev. Applied **12**, 064041 (2019).

[182]Z. Jiang, A. Kalev, W. Mruczkiewicz, and H. Neven, "Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning", arXiv:1910.10746 (2019).

[183]M. Nooijen, "Can the eigenstates of a many-body hamiltonian be represented exactly using a general two-body cluster expansion?", Phys. Rev. Lett. **84**, 2108–2111 (2000).

[184]W. Kutzelnigg, "Quantum chemistry in Fock space. I. The universal wave and energy operators", J. Chem. Phys. **77**, 3081–3097 (1982).

[185]W. Kutzelnigg and S. Koch, "Quantum chemistry in Fock space. II. Effective Hamiltonians in Fock space", J. Chem. Phys. **79**, 4315–4335 (1983).

[186]W. Kutzelnigg, "Quantum chemistry in Fock space. III. Particle-hole formalism", J. Chem. Phys. **80**, 822–830 (1984).

[187]W. Kutzelnigg, "Quantum chemistry in Fock space. IV. The treatment of permutational symmetry. Spin-free diagrams with symmetrized vertices", J. Chem. Phys. **82**, 4166–4186 (1985).

[188]A. G. Taube and R. J. Bartlett, "New perspectives on unitary coupled-cluster theory", Int. J. Quantum Chem. **106**, 3393–3401 (2006).

[189]Y. Matsuzawa and Y. Kurashige, "Jastrow-type decomposition in quantum chemistry for Low-Depth quantum circuits", J. Chem. Theory Comput. **16**, 944–952 (2020).

[190]P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, "Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer", Quantum Sci. Technol. **4**, 045005 (2019).

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

[192]H. L. Tang, V. O. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, "qubit-ADAPT-VQE: an adaptive algorithm for constructing hardware-efficient ansatze on a quantum processor", arXiv:1911.10205 (2019).

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

[194]C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, and P. Zoller, "Self-verifying variational quantum simulation of lattice models", Nature **569**, 355–360 (2019).

[195]R. Assaraf and M. Caffarel, "Zero-Variance principle for monte carlo algorithms", Phys. Rev. Lett. **83**, 4682–4685 (1999).

[196]V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, "Measurement optimization in the variational quantum eigensolver using a minimum clique cover", J. Chem. Phys. **152**, 124114 (2020).

[197]N. C. Rubin, R. Babbush, and J. McClean, "Application of fermionic marginal constraints to hybrid quantum algorithms", New J. Phys. **20**, 053020 (2018).

[198]K. J. Sung, M. P. Harrigan, N. C. Rubin, Z. Jiang, R. Babbush, and J. R. McClean, "An exploration of practical optimizers for variational quantum algorithms on superconducting qubit processors", arXiv:2005.11011 (2020).

[199]M. J. D. Powell, "A direct search optimization method that models the objective and constraint functions by linear interpolation", in *Advances in optimization and numerical analysis*, Vol. 7, edited by S. Gomez and J.-P. Hennart (Springer Netherlands, Dordrecht, 1994), pp. 51–67.

[200]M. J. D. Powell, "The BOBYQA algorithm for bound constrained optimization without derivatives", Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, 26–46 (2009).

[201]J. A. Nelder and R. Mead, "A simplex method for function minimization", Comput. J. **7**, 308–313 (1965).

[202]J. Kennedy and R. Eberhart, "Particle swarm optimization", in Proceedings of ICNN'95 - international conference on neural networks, Vol. 4 (Nov. 1995), 1942–1948 vol.4.

[203]J. Li, X. Yang, X. Peng, and C.-P. Sun, "Hybrid Quantum-Classical approach to quantum optimal control", Phys. Rev. Lett. **118**, 150503 (2017).

[204]M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, "Evaluating analytic gradients on quantum hardware", Phys. Rev. A **99**, 032331 (2019).

[205]G. E. Crooks, "Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition", arXiv:1905.13311 (2019).

[206]A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: a survey", J. Mach. Learn. Res. **18**, 5595–5637 (2017).

[207]A. Harrow and J. Napp, "Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms", arXiv:1901.05374 (2019).

[208]K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for Short-Depth quantum circuits", Phys. Rev. Lett. **119**, 180509 (2017).

[209]A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor", Nature **567**, 491–495 (2019).

[210]M. Otten and S. K. Gray, "Recovering noise-free quantum observables", Phys. Rev. A **99**, 012338 (2019).

[211]A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, "Learning-based quantum error mitigation", arXiv:2005.07601 (2020).

[212]J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, "Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states", Phys. Rev. A **95**, 042308 (2017).

[213]X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien, "Low-cost error mitigation by symmetry verification", Phys. Rev. A **98**, 062339 (2018).

[214]S. McArdle, X. Yuan, and S. Benjamin, "Error-Mitigated digital quantum simulation", Phys. Rev. Lett. **122**, 180501 (2019).

[215]R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O'Brien, and L. DiCarlo, "Experimental error mitigation via symmetry verification in a variational quantum eigensolver", Phys. Rev. A **100**, 010302 (2019).

[216]F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. Dunsworth, E. Farhi, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, S. Habegger, M. P. Harrigan, A. Ho, S. Hong, T. Huang, W. J. Huggins, L. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. V. Klimov, A. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, M. Lindmark, E. Lucero, O. Martin, J. M. Martinis, J. R. McClean, M. McEwen, A. Megrant, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, H. Neven, M. Y. Niu, T. E. O'Brien, E. Ostby, A. Petukhov, H. Putterman, C. Quintana, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, D. Strain, K. J. Sung, M. Szalay, T. Y. Takeshita, A. Vainsencher, T. White, N. Wiebe, Z. Jamie Yao, P. Yeh, and A. Zalcman, "Hartree-Fock on a superconducting qubit quantum computer", arXiv:2004.04174 (2020).

[217]J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes", Nat. Commun. **9**, 4812 (2018).

[218]E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, "An initialization strategy for addressing barren plateaus in parametrized quantum circuits", arXiv:1903.05076 (2019) 10.22331/q-2019-12-09-214.

[219]T. Volkoff and P. J. Coles, "Large gradients via correlation in random parameterized quantum circuits", arXiv:2005.12200 (2020).

[220]Fernando G S, M. Broughton, E. Farhi, S. Gutmann, and H. Neven, "For fixed control parameters the quantum approximate optimization algorithm's objective function value concentrates for typical instances", arXiv:1812.04170 (2018).

[221]R. M. Parrish and P. L. McMahon, "Quantum filter diagonalization: quantum eigendecomposition without full quantum phase estimation", arXiv:1909.08925 (2019).

[222]N. H. Stair, R. Huang, and F. A. Evangelista, "A multireference quantum krylov algorithm for strongly correlated electrons", J. Chem. Theory Comput. **16**, 2236–2245 (2020).

[223]D. Wang, O. Higgott, and S. Brierley, "Accelerated variational quantum eigensolver", Phys. Rev. Lett. **122**, 140504 (2019).

[224]J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, "Generalized unitary coupled cluster wave functions for quantum computation", J. Chem. Theory Comput. **15**, 311–324 (2019).

[225]A. W. Harrow and A. Montanaro, "Quantum computational supremacy", Nature **549**, 203–209 (2017).

[226]F. A. Evangelista, "Alternative single-reference coupled cluster approaches for multireference problems: the simpler, the better", J. Chem. Phys. **134**, 224102 (2011).

[227]I. H. Kim and B. Swingle, "Robust entanglement renormalization on a noisy quantum computer", arXiv:1711.07500 (2017).

[228]I. G. Ryabinkin, T.-C. Yen, S. N. Genin, and A. F. Izmaylov, "Qubit coupled cluster method: a systematic approach to quantum chemistry on a quantum computer", J. Chem. Theory Comput. **14**, 6317–6326 (2018).

[229]E. M. Stoudenmire and S. R. White, "Studying Two-Dimensional systems with the density matrix renormalization group", Annu. Rev. Condens. Matter Phys. **3**, 111–128 (2012).

[230]M. P. Nightingale and C. J. Umrigar, *Quantum monte carlo methods in physics and chemistry*, edited by M. P. Nitingale and C. J. Umrigar (Springer Science & Business Media, Dec. 1998), p. 129.

[231]O. Higgott, D. Wang, and S. Brierley, "Variational quantum computation of excited states", Quantum **3**, 156 (2019).

[232]J. Čızek and J. Paldus, "Coupled cluster approach", Phys. Scr. **21**, 251 (1980).

[233]R. J. Bartlett and M. Musiał, "Coupled-cluster theory in quantum chemistry", Rev. Mod. Phys. **79**, 291–352 (2007).

[234]J. Paldus, J. Čızek, and M. Takahashi, "Approximate account of the connected quadruply excited clusters in the coupled-pair many-electron theory", Phys. Rev. A **30**, 2193–2209 (1984).

[235]P. Piecuch, R. Tobo?a, and J. Paldus, "Coupled-Cluster approaches with an approximate account of triply and quadruply excited clusters: Implementation of the orthogonally spin-adaptedCCD +ST(CCD),CCSD +T(CCSD), andACPQ +ST(ACPQ) formalisms", Int. J. Quantum Chem., Lecture Notes **55**, 133–146 (1995).

[236]D. W. Small and M. Head-Gordon, "A fusion of the closed-shell coupled cluster singles and doubles method and valence-bond theory for bond breaking", J. Chem. Phys. **137**, 114103 (2012).

[237]D. W. Small, K. V. Lawler, and M. Head-Gordon, "Coupled cluster valence bond method: efficient computer implementation and application to multiple bond dissociations and strong correlations in the acenes", J. Chem. Theory Comput. **10**, 2027–2040 (2014).

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

[239]H. Nakatsuji, "Equation for the direct determination of the density matrix", Phys. Rev. A **14**, 41–50 (1976).

[240]H. Nakatsuji, "Structure of the exact wave function", J. Chem. Phys. **113**, 2949–2956 (2000).

[241]T. Van Voorhis and M. Head-Gordon, "Two-body coupled cluster expansions", J. Chem. Phys. **115**, 5033–5040 (2001).

[242]P. Piecuch, K. Kowalski, P.-D. Fan, and K. Jedziniak, "Exactness of two-body cluster expansions in many-body quantum theory", Phys. Rev. Lett. **90**, 113001 (2003).

[243]E. R. Davidson, "Exactness of the general two-body cluster expansion in many-body quantum theory", Phys. Rev. Lett. **91**, 123001 (2003).

[244]S. Ronen, "Can the eigenstates of a many-body hamiltonian be represented exactly using a general two-body cluster expansion?", Phys. Rev. Lett. **91**, 123002 (2003).

[245]D. A. Mazziotti, "Exactness of wave functions from two-body exponential transformations in many-body quantum theory", Phys. Rev. A **69**, 012507 (2004).

[246]H. Nakatsuji, "Scaled schrödinger equation and the exact wave function", Phys. Rev. Lett. **93**, 030403 (2004).

[247]D. Mukherjee and W. Kutzelnigg, "Some comments on the coupled cluster with generalized singles and doubles (CCGSD) ansatz", Chem. Phys. Lett. **397**, 174–179 (2004).

[248]P. A. Limacher, P. W. Ayers, P. A. Johnson, S. De Baerdemacker, D. Van Neck, and P. Bultinck, "A New Mean-Field Method Suitable for Strongly Correlated Electrons: Computationally Facile Antisymmetric Products of Nonorthogonal Geminals", J. Chem. Theory Comput. **9**, 1394–1401 (2013).

[249]W. A. Goddard, T. H. Dunning, W. J. Hunt, and P. J. Hay, "Generalized valence bond description of bonding in low-lying states of molecules", Acc. Chem. Res. **6**, 368–376 (1973).

[250]R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, D. P. Tew, X. Zhou, J. L. O'Brien, and M. G. Thompson, "Witnessing eigenstates for quantum simulation of hamiltonian spectra", Sci Adv **4**, eaap9646 (2018).

[251]Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, "Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure", Phys. Rev. A **95**, 020501 (2017).

[252]J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, "Robust determination of molecular spectra on a quantum processor", arXiv:1707.06408 (2017).

[253]T. Jones, S. Endo, S. McArdle, X. Yuan, and S. C. Benjamin, "Variational quantum algorithms for discovering hamiltonian spectra", Phys. Rev. A **99**, 062304 (2019).

[254]M. A. Watson and G. K.-L. Chan, "Excited states of butadiene to chemical accuracy: reconciling theory and experiment", J. Chem. Theory Comput. **8**, 4013–4018 (2012).

[255]J. H. Choi, C. F. Lebeda, and R. P. Messmer, "Variational principle for excited states: Exact formulation and other extensions", Chem. Phys. Lett. **5**, 503–506 (1970).

[256]J. S. Andrews, D. Jayatilaka, R. G. A. Bone, N. C. Handy, and R. D. Amos, "Spin contamination in single-determinant wavefunctions", Chem. Phys. Lett. **183**, 423–431 (1991).

[257]F. A. Evangelista, P. Shushkov, and J. C. Tully, "Orthogonality constrained density functional theory for electronic excited states", J. Phys. Chem. A **117**, 7378–7392 (2013).

[258]V. N. Glushkov and X. Assfeld, "Orthogonality-constrained Hartree–Fock and perturbation theory for high-spin open-shell excited states", Theor. Chem. Acc. **135**, 3 (2015).

[259]F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller 3rd, "A Simple, Exact Density-Functional-Theory Embedding Scheme", J. Chem. Theory Comput. **8**, 2564–2568 (2012).

[260]S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, "Tapering off qubits to simulate fermionic hamiltonians", arXiv:1701.08213 (2017).

[261]P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, S. Filipp, and I. Tavernelli, "Quantum algorithms for electronic structure calculations: particle/hole Hamiltonian and optimized wavefunction expansions", arXiv:1805.04340 (2018).

[262]R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince 3rd, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer 3rd, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, "Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability", J. Chem. Theory Comput. **13**, 3185–3197 (2017).

[263]J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. Schuyler Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. O'Brien, B. O'Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. P. D. Sawaya, B. Senjean, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush, "OpenFermion: the electronic structure package for quantum computers", Quantum Sci. Technol. **5**, 034014 (2020).

[264]M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., "Tensorflow: a system for large-scale machine learning", in 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16) (2016), pp. 265–283.

[265]P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa,

P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: fundamental algorithms for scientific computing in python", Nat. Methods **17**, 261–272 (2020).

[266]J. Nocedal and S. Wright, *Numerical optimization* (Springer Science & Business Media, Dec. 2006), pp. 245–269.

[267]J. Paldus, P. Piecuch, L. Pylypow, and B. Jeziorski, "Application of hilbert-space coupled-cluster theory to simple (h2)2 model systems: planar models", Phys. Rev. A **47**, 2738–2782 (1993).

[268]U. S. Mahapatra, B. Datta, and D. Mukherjee, "A size-consistent state-specific multireference coupled cluster theory: Formal developments and molecular applications", J. Chem. Phys. **110**, 6171–6188 (1999).

[269]K. Kowalski and P. Piecuch, "Complete set of solutions of multireference coupled-cluster equations: The state-universal formalism", Phys. Rev. A **61**, 052506 (2000).

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

[271]F. A. Evangelista, W. D. Allen, and H. F. Schaefer 3rd, "High-order excitations in state-universal and state-specific multireference coupled cluster theories: model systems", J. Chem. Phys. **125**, 154113 (2006).

[272]W. J. Hehre, R. F. Stewart, and J. A. Pople, "Self-Consistent Molecular-Orbital methods. i. use of gaussian expansions of Slater-Type atomic orbitals", J. Chem. Phys. **51**, 2657–2664 (1969).

[273]J. B. Collins, P. von R. Schleyer, J. S. Binkley, and J. A. Pople, "Self-consistent molecular orbital methods. XVII. geometries and binding energies of second-row molecules. a comparison of three basis sets", J. Chem. Phys. **64**, 5142–5151 (1976).

[274]T. Van Voorhis and M. Head-Gordon, "Implementation of generalized valence bond-inspired coupled cluster theories", J. Chem. Phys. **117**, 9190–9201 (2002).

[275]R. Ditchfield, W. J. Hehre, and J. A. Pople, "Self-Consistent Molecular-Orbital methods. IX. an extended Gaussian-Type basis for Molecular-Orbital studies of organic molecules", J. Chem. Phys. **54**, 724–728 (1971).

[276]F. B. Brown, I. Shavitt, and R. Shepard, "Multireference configuration interaction treatment of potential energy surfaces: symmetric dissociation of H2O in a double-zeta basis", Chem. Phys. Lett. **105**, 363–369 (1984).

[277]J. Olsen, P. Jo/rgensen, H. Koch, A. Balkova, and R. J. Bartlett, "Full configuration–interaction and state of the art correlation calculations on water in a valence double-zeta basis with polarization functions", J. Chem. Phys. **104**, 8007–8015 (1996).

[278]X. Li and J. Paldus, "Reduced multireference couple cluster method. II. application to potential energy surfaces of HF, f2, and H2O", J. Chem. Phys. **108**, 637–648 (1998).

[279]J. Ma, S. Li, and W. Li, "A multireference configuration interaction method based on the separated electron pair wave functions", J. Comput. Chem., Modern Theore **27**, 39–47 (2006).

[280]P. E. M. Siegbahn, "The externally contracted CI method applied to N2", Int. J. Quantum Chem., Proceedings o **23**, 1869–1889 (1983).

[281]A. Dreuw and M. Head-Gordon, "Single-reference ab initio methods for the calculation of excited states of large molecules", Chem. Rev. **105**, 4009–4037 (2005).

[282]J. Cullen, M. Krykunov, and T. Ziegler, "The formulation of a self-consistent constricted variational density functional theory for the description of excited states", Chem. Phys. **391**, 11–18 (2011).

[283]J. D. Watts, G. W. Trucks, and R. J. Bartlett, "The unitary coupled-cluster approach and molecular properties. Applications of the UCC(4) method", Chem. Phys. Lett. **157**, 359–366 (1989).

[284]R. J. Bartlett, S. A. Kucharski, and J. Noga, "Alternative coupled-cluster ansätze II. The unitary coupled-cluster method", Chem. Phys. Lett. **155**, 133–140 (1989).

[285]W. Kutzelnigg, "Error analysis and improvements of coupled-cluster theory", Theor. Chim. Acta **80**, 349–386 (1991).

[286]E. Xu, M. Uejima, and S. L. Ten-No, "Full Coupled-Cluster reduction for accurate description of strong electron correlation", Phys. Rev. Lett. **121**, 113001 (2018).

[287]C. E. Dykstra, "An examination of the Brueckner condition for the selection of molecular orbitals in correlated wavefunctions", Chem. Phys. Lett. **45**, 466–469 (1977).

[288]N. C. Handy, J. A. Pople, M. Head-Gordon, K. Raghavachari, and G. W. Trucks, "Size-consistent Brueckner theory limited to double substitutions", Chem. Phys. Lett. **164**, 185–192 (1989).

[289]A. I. Krylov, C. D. Sherrill, E. F. C. Byrd, and M. Head-Gordon, "Size-consistent wave functions for nondynamical correlation energy: the valence active space optimized orbital coupled-cluster doubles model", J. Chem. Phys. **109**, 10669–10678 (1998).

[290]R. C. Lochan and M. Head-Gordon, "Orbital-optimized opposite-spin scaled second-order correlation: an economical method to improve the description of open-shell molecules", J. Chem. Phys. **126**, 164101 (2007).

[291]J. Lee and M. Head-Gordon, "Regularized Orbital-Optimized Second-Order Møller-Plesset perturbation theory: a reliable Fifth-Order-Scaling electron correlation model with orbital energy dependent regularizers", J. Chem. Theory Comput. **14**, 5203–5219 (2018).

[292]J. Lee, "Postmodern electronic structure theory", PhD thesis (UC Berkeley, 2019).

[293]R. Fletcher, *Practical methods of optimization* (John Wiley & Sons, June 2013).

[294]D. F. Shanno, "Conditioning of quasi-newton methods for function minimization", Math. Comput. **24**, 647–647 (1970).

[295]D. Goldfarb, "A family of variable-metric methods derived by variational means", Math. Comput. **24**, 23–23 (1970).

[296]R. Fletcher, "A new approach to variable metric algorithms", Comput. J. **13**, 317–322 (1970).

[297]C. G. Broyden, "The convergence of a class of double-rank minimization algorithms 1. general considerations", IMA J Appl Math **6**, 76–90 (1970).

[298]W. J. Huggins, J. Lee, U. Baek, B. O'Gorman, and K. Birgitta Whaley, "A non-orthogonal variational quantum eigensolver", New J. Phys. **22**, 073009 (2020).

[299]W. J. Huggins, J. McClean, N. Rubin, Z. Jiang, N. Wiebe, K. Birgitta Whaley, and R. Babbush, "Efficient and noise resilient measurements for quantum chemistry on Near-Term quantum computers", arXiv:1907.13117 (2019).

[300]M. Otten and S. K. Gray, "Accounting for errors in quantum algorithms via individual error reduction", npj Quantum Information **5**, 11 (2019).

[301]M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, "Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution", Nat. Phys. **16**, 205–210 (2020).

[302]O. Kyriienko, "Quantum inverse iteration algorithm for programmable quantum simulators", npj Quantum Information **6**, 7 (2020).

[303]R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martınez, "Quantum computation of electronic transitions using a variational quantum eigensolver", Phys. Rev. Lett. **122**, 230401 (2019).

[304]E. U. Condon, E. U. Condon, and G. H. Shortley, *The theory of atomic spectra* (Cambridge University Press, 1951).

[305]B. Jeziorski and H. J. Monkhorst, "Coupled-cluster method for multideterminantal reference states", Phys. Rev. A **24**, 1668–1681 (1981).

[306]H. Werner and E. Reinsch, "The self-consistent electron pairs method for multiconfiguration reference state functions", J. Chem. Phys. **76**, 3144–3156 (1982).

[307]H. Koch and E. Dalgaard, "Linear superposition of optimized non-orthogonal slater determinants for singlet states", Chem. Phys. Lett. **212**, 193–200 (1993).

[308]P. Malmqvist, "Calculation of transition density matrices by nonunitary orbital transformations", Int. J. Quantum Chem., Lecture Notes **30**, 479–494 (1986).

[309]A. J. W. Thom and M. Head-Gordon, "Hartree-Fock solutions as a quasidiabatic basis for nonorthogonal configuration interaction", J. Chem. Phys. **131**, 124113 (2009).

[310]C. A. Jiménez-Hoyos, R. Rodrıguez-Guzmán, and G. E. Scuseria, "Multi-component symmetry-projected approach for molecular ground state correlations", J. Chem. Phys. **139**, 204102 (2013).

[311]J. R. McClean and A. Aspuru-Guzik, "Compact wavefunctions from compressed imaginary time evolution", RSC Adv. **5**, 102277–102283 (2015).

[312]E. J. Sundstrom and M. Head-Gordon, "Non-orthogonal configuration interaction for the calculation of multielectron excited states", J. Chem. Phys. **140**, 114103 (2014).

[313]E. J. Landinez Borda, J. Gomez, and M. A. Morales, "Non-orthogonal multi-slater determinant expansions in auxiliary field quantum monte carlo", J. Chem. Phys. **150**, 074105 (2019).

[314]M. Araújo, A. Feix, F. Costa, and Č. Brukner, "Quantum circuits cannot control unknown operations", New J. Phys. **16**, 093026 (2014).

[315]A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi, "Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method", J. Chem. Theory Comput. **16**, 190–195 (2020).

[316]V. V. Shende and I. L. Markov, "On the CNOT-cost of TOFFOLI gates", arXiv:0803.2316 (2008).

[317]M. Shinozuka and C. J. Astill, "Random eigenvalue problems in structural analysis", AIAA Journal **10**, 456–462 (1972).

[318]H. Benaroya, "Random eigenvalues, algebraic methods and structural dynamic models", Appl. Math. Comput. **52**, 37–66 (1992).

[319]C. Soize, "Random matrix theory for modeling uncertainties in computational mechanics", Comput. Methods Appl. Mech. Eng. **194**, 1333–1366 (2005).

[320]K. Kowalski, "Properties of coupled-cluster equations originating in excitation subalgebras", J. Chem. Phys. **148**, 094104 (2018).

[321]The Cirq Developers, *Cirq*, July 2019.

[322]I. G. Ryabinkin, R. A. Lang, S. N. Genin, and A. F. Izmaylov, "Iterative qubit coupled cluster approach with efficient screening of generators", J. Chem. Theory Comput. **16**, 1055–1063 (2020).

[323]C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization", ACM Trans. Math. Softw. **23**, 550–560 (1997).

[324]J. Lee, F. D. Malone, and M. A. Morales, "An auxiliary-field quantum monte carlo perspective on the ground state of the dense uniform electron gas: an investigation with Hartree-Fock trial wavefunctions", arXiv:1905.04361, 064122 (2019).

[325]H. R. Grimsley, D. Claudino, S. E. Economou, E. Barnes, and N. J. Mayhall, "Is the trotterized UCCSD ansatz chemically Well-Defined?", J. Chem. Theory Comput., 1–6 (2020).

[326]H.-Y. Huang, K. Bharti, and P. Rebentrost, "Near-term quantum algorithms for linear systems of equations", arXiv:1909.07344 (2019).

[327]J.-D. Chai and M. Head-Gordon, "Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections", Phys. Chem. Chem. Phys. **10**, 6615–6620 (2008).

[328]T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo, and L. Visscher, "Calculating energy derivatives for quantum chemistry on a quantum computer", arXiv:1905.03742 (2019).

[329]A. Jena, S. Genin, and M. Mosca, "Pauli partitioning with respect to gate sets", arXiv:1907.07859 (2019).

[330]T.-C. Yen, V. Verteletskyi, and A. F. Izmaylov, "Measuring all compatible operators in one series of a single-qubit measurements using unitary transformations", arXiv:1907.09386 (2019).

[331]A. F. Izmaylov, T.-C. Yen, and I. G. Ryabinkin, "Revising the measurement process in the variational quantum eigensolver: is it possible to reduce the number of separately measured operators?", Chem. Sci. **10**, 3746–3755 (2019).

[332]P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, "Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families", arXiv:1907.13623 (2019).

[333]R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, "Low-Depth Quantum Simulation of Materials", Phys. Rev. X **8**, 011044 (2018).

[334]D. Poulin, M. B. Hastings, D. Wecker, N. Wiebe, A. C. Doherty, and M. Troyer, "The Trotter Step Size Required for Accurate Quantum Simulation of Quantum Chemistry", arXiv:1406.4920, 361–384 (2014).

[335]J. L. Whitten, "Coulombic potential energy integrals and approximations", J. Chem. Phys. **58**, 4496–4501 (1973).

[336]F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-A. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitonák, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov, and R. Lindh, "MOLCAS 7: the next generation", J. Comput. Chem. **31**, 224–247 (2010).

[337]T. B. Pedersen, F. Aquilante, and R. Lindh, "Density fitting with auxiliary basis sets from cholesky decompositions", Theor. Chem. Acc. **124**, 1–10 (2009).

[338]N. H. F. Beebe and J. Linderberg, "Simplifications in the generation and transformation of two-electron integrals in molecular calculations", Int. J. Quantum Chem. **12**, 683–705 (1977).

[339]H. Koch, A. Sánchez de Merás, and T. B. Pedersen, "Reduced scaling in electronic structure calculations using cholesky decompositions", J. Chem. Phys. **118**, 9481–9484 (2003).

[340]W. Purwanto, H. Krakauer, Y. Virgus, and S. Zhang, "Assessing weak hydrogen binding on ca+ centers: an accurate many-body study with large basis sets", J. Chem. Phys. **135**, 164105 (2011).

[341]N. Mardirossian, J. D. McClain, and G. K.-L. Chan, "Lowering of the complexity of quantum chemistry methods by choice of representation", J. Chem. Phys. **148**, 044106 (2018).

[342]B. Peng and K. Kowalski, "Highly efficient and scalable compound decomposition of Two-Electron integral tensor and its application in coupled cluster calculations", J. Chem. Theory Comput. **13**, 4179–4192 (2017).

[343]I. Røeggen and E. Wisløff-Nilssen, "On the Beebe-Linderberg two-electron integral approximation", Chem. Phys. Lett. **132**, 154–160 (1986).

[344]I. Røeggen and T. Johansen, "Cholesky decomposition of the two-electron integral matrix in electronic structure calculations", J. Chem. Phys. **128**, 194107 (2008).

[345]L. Boman, H. Koch, and A. Sánchez de Merás, "Method specific cholesky decomposition: coulomb and exchange energies", J. Chem. Phys. **129**, 134107 (2008).

[346]W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. Steven Kolthammer, and I. A. Walmsley, "Optimal design for universal multiport interferometers", Optica, OPTICA **3**, 1460–1465 (2016).

[347]C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, "Robust online hamiltonian learning", New J. Phys. **14**, 103013 (2012).

[348]J. Heinsoo, C. K. Andersen, A. Remm, S. Krinner, T. Walter, Y. Salathé, S. Gasparinetti, J.-C. Besse, A. Potočnik, A. Wallraff, and C. Eichler, "Rapid high-fidelity multiplexed readout of superconducting qubits", Phys. Rev. Applied **10**, 034040 (2018).

[349]A. Kitaev, *A Simple Model of Quantum Holography*, Santa Barbara, 2015.

[350]R. Babbush, D. W. Berry, and H. Neven, "Quantum simulation of the Sachdev-Ye-Kitaev model by asymmetric qubitization", Phys. Rev. A **99**, 040301 (2019).

[351]W. Huggins, C. D. Freeman, M. Stoudenmire, et al., "Monte carlo tensor network renormalization", arXiv:1710.03757 (2017).

[352]N. Jeevanjee, *An introduction to tensors and group theory for physicists* (Birkhäuser, Cham, 2015).

[353]J. C. Bridgeman and C. T. Chubb, "Hand-waving and interpretive dance: an introductory course on tensor networks", J. Phys. A: Math. Theor. **50**, 223001 (2017).

[354]R. Penrose, "Applications of negative dimensional tensors", Combinatorial mathematics and its applications **1**, 221–244 (1971).

[355]M. B. Hastings, "An area law for one-dimensional quantum systems", J. Stat. Mech. **2007**, P08024–P08024 (2007).

[356]J. Eisert, M. Cramer, and M. B. Plenio, "Colloquium: area laws for the entanglement entropy", Rev. Mod. Phys. **82**, 277–306 (2010).

[357]R. Orus, "A practical introduction to tensor networks: matrix product states and projected entangled pair states", Ann. Phys. **349**, 117–158 (2014).

[358]G. Vidal, "Class of quantum many-body states that can be efficiently simulated", Phys. Rev. Lett. **101**, 110501 (2008).

[359] *Tensor network*, `http://tensornetwork.org/`, Accessed: 2020-7-16.

[360]M. Levin and C. P. Nave, "Tensor renormalization group approach to two-dimensional classical lattice models", Phys. Rev. Lett. **99**, 120601 (2007).

[361]U. Schollwöck, "The density-matrix renormalization group in the age of matrix product states", Ann. Phys. **326**, 96–192 (2011).

[362]S. R. White, "Density-matrix algorithms for quantum renormalization groups", Phys. Rev. B Condens. Matter **48**, 10345–10356 (1993).

[363]Y. Kurashige and T. Yanai, "High-performance ab initio density matrix renormalization group method: applicability to large-scale multireference problems for metal compounds", J. Chem. Phys. **130**, 234114 (2009).

[364]H. J. Changlani, N. M. Tubman, and T. L. Hughes, "Charge density waves in disordered media circumventing the Imry-Ma argument", Sci. Rep. **6**, 31897 (2016).

[365]J. Dukelsky, M. A. Martın-Delgado, T. Nishino, and G. Sierra, "Equivalence of the variational matrix product method and the density matrix renormalization group applied to spin chains", EPL **43**, 457 (1998).

[366]F. Verstraete, V. Murg, and J. I. Cirac, "Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems", Adv. Phys. **57**, 143–224 (2008).

[367]P. Corboz, R. Orús, B. Bauer, and G. Vidal, "Simulation of strongly correlated fermions in two spatial dimensions with fermionic projected entangled-pair states", Phys. Rev. B Condens. Matter **81**, 165104 (2010).

[368]P. Corboz, "Improved energy extrapolation with infinite projected entangled-pair states applied to the two-dimensional Hubbard model", Phys. Rev. B Condens. Matter **93**, 045116 (2016).

[369]B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehlers, M.-P. Qin, R. M. Noack, H. Shi, S. R. White, S. Zhang, and G. K.-L. Chan, "Stripe order in the underdoped region of the two-dimensional Hubbard model", Science **358**, 1155–1160 (2017).

[370]Z. Y. Xie, J. Chen, M. P. Qin, J. W. Zhu, L. P. Yang, and T. Xiang, "Coarse-graining renormalization by higher-order singular value decomposition", Phys. Rev. B Condens. Matter **86**, 045139 (2012).

[371]G. Evenbly and G. Vidal, "Tensor network renormalization", Phys. Rev. Lett. **115**, 180405 (2015).

[372]T. Nishino and K. Okunishi, "Corner Transfer Matrix Renormalization Group Method", J. Phys. Soc. Jpn. **65**, 891–894 (1996).

[373]B. Bruognolo, Z. Zhu, S. R. White, and E. Miles Stoudenmire, "Matrix product state techniques for two-dimensional systems at finite temperature", arXiv:1705.05578 (2017).

[374]S. Yan, D. A. Huse, and S. R. White, "Spin-liquid ground state of the S = 1/2 kagome heisenberg antiferromagnet", Science **332**, 1173–1176 (2011).

[375]G. K.-L. Chan, "An algorithm for large scale density matrix renormalization group calculations", J. Chem. Phys. **120**, 3172–3178 (2004).

[376]E. M. Stoudenmire and S. R. White, "Real-space parallel density matrix renormalization group", Phys. Rev. B Condens. Matter **87**, 155137 (2013).

[377]R. J. Needs, M. D. Towler, N. D. Drummond, and P. López Rıos, "Continuum variational and diffusion quantum monte carlo calculations", J. Phys. Condens. Matter **22**, 023201 (2010).

[378]N. M. Tubman, J. L. DuBois, R. Q. Hood, and B. J. Alder, "Prospects for release-node quantum monte carlo", J. Chem. Phys. **135**, 184109 (2011).

[379]N. M. Tubman, I. Kylänpää, S. Hammes-Schiffer, and D. M. Ceperley, "Beyond the Born-Oppenheimer approximation with quantum monte carlo methods", Phys. Rev. A **90**, 042507 (2014).

[380]E. W. Brown, B. K. Clark, J. L. DuBois, and D. M. Ceperley, "Path-integral monte carlo simulation of the warm dense homogeneous electron gas", Phys. Rev. Lett. **110**, 146405 (2013).

[381]A. J. Ferris, "Unbiased Monte Carlo for the age of tensor networks", arXiv:1507.00767 (2015).

[382]M. H. Kalos, *Monte carlo methods*, 2nd rev. and enl. ed. (Wiley-Blackwell, Weinheim, 2008).

[383]J. R. Trail, "Heavy-tailed random error in quantum monte carlo", Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **77**, 016703 (2008).

[384]D. Tamascelli, R. Rosenbach, and M. B. Plenio, "Improved scaling of time-evolving block-decimation algorithm through reduced-rank randomized singular value decomposition", Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **91**, 063306 (2015).

[385]A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks", arXiv:1511.06434 (2015).

[386]H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks", IEEE Transactions on Computational Imaging **3**, 47–57 (2017).

[387]D. Duvenaud, D. Maclaurin, and R. Adams, "Early stopping as nonparametric variational inference", in Artificial intelligence and statistics (May 2016), pp. 1070–1077.

[388]J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization", J. Mach. Learn. Res. **13**, 281–305 (2012).

[389]M. Feurer and F. Hutter, "Hyperparameter optimization", in *Automated machine learning* (Springer, Cham, 2019), pp. 3–33.

[390]S. Östlund and S. Rommer, "Thermodynamic limit of density matrix renormalization", Phys. Rev. Lett. **75**, 3537–3540 (1995).

[391]M. Fannes, B. Nachtergaele, and R. F. Werner, "Finitely correlated states on quantum spin chains", Commun. Math. Phys. **144**, 443–490 (1992).

[392]M.-B. Lepetit, M. Cousy, and G. M. Pastor, "Density-matrix renormalization study of the hubbard model [4]on a bethe lattice", The European Physical Journal B - Condensed Matter and Complex Systems **13**, 421–427 (2000).

[393]L. Tagliacozzo, G. Evenbly, and G. Vidal, "Simulation of two-dimensional quantum systems using a tree tensor network that exploits the entropic area law", Phys. Rev. B Condens. Matter **80**, 235127 (2009).

[394]W. Hackbusch and S. Kühn, "A new scheme for the tensor representation", J. Fourier Anal. Appl. **15**, 706–722 (2009).

[395]G. Vidal, "Efficient classical simulation of slightly entangled quantum computations", Phys. Rev. Lett. **91**, 147902 (2003).

[396]N. Cohen, O. Sharir, and A. Shashua, "On the expressive power of deep learning: a tensor analysis", in Conference on learning theory (June 2016), pp. 698–728.

[397]A. Novikov, M. Trofimov, and I. Oseledets, "Exponential machines", arXiv:1605.03795 (2016).

[398]E. Stoudenmire and D. J. Schwab, "Supervised learning with tensor networks", in *Advances in neural information processing systems 29*, edited by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc., 2016), pp. 4799–4807.

[399]Y. Levine, D. Yakira, N. Cohen, and A. Shashua, "Deep learning and quantum entanglement: fundamental connections with implications to network design", arXiv:1704.01552 (2017).

[400]D. Liu, S.-J. Ran, P. Wittek, C. Peng, R. B. Garcıa, G. Su, and M. Lewenstein, "Machine learning by Two-Dimensional hierarchical tensor networks: a quantum information theoretic perspective on deep architectures", arXiv preprint arXiv:1710.04833 (2017).

[401]V. Khrulkov, A. Novikov, and I. Oseledets, "Expressive power of recurrent neural networks", arXiv:1711.00811 (2017).

[402]E. M. Stoudenmire, "Learning relevant features of data with multi-scale tensor networks", Quantum Science and Technology **3**, 034003 (2018).

[403]Z.-Y. Han, J. Wang, H. Fan, L. Wang, and P. Zhang, "Unsupervised generative modeling using matrix product states", Phys. Rev. X **8**, 031012 (2018).

[404]P. W. Shor, "Polynomial-Time algorithms for prime factorization and discrete logarithms on a quantum computer", SIAM Rev. **41**, 303–332 (1999).

[405]M. Schuld, A. Bocharov, K. Svore, and N. Wiebe, "Circuit-centric quantum classifiers", arXiv:1804.00633 (2018).

[406]X. Gao, Z. Zhang, and L. Duan, "An efficient quantum algorithm for generative machine learning", arXiv:1711.02038 (2017).

[407]M. Benedetti, D. Garcia-Pintos, O. Perdomo, V. Leyton-Ortega, Y. Nam, and A. Perdomo-Ortiz, "A generative modeling approach for benchmarking and training shallow quantum circuits", npj Quantum Information **5**, 45 (2019).

[408]J. Biamonte, A. Kardashin, and A. Uvarov, "Quantum machine learning tensor network states", arXiv:1804.02398 (2018).

[409]G. Evenbly and G. Vidal, "Tensor network states and geometry", J. Stat. Phys. **145**, 891–918 (2011).

[410]P. Mehta, C.-H. Wang, A. G. R. Day, C. Richardson, M. Bukov, C. K. Fisher, and D. J. Schwab, "A high-bias, low-variance introduction to machine learning for physicists", Phys. Rep. **810**, 1–124 (2019).

[411]S. Cheng, J. Chen, and L. Wang, "Information perspective to probabilistic modeling: boltzmann machines versus born machines", Entropy **20**, 583 (2018).

[412]J.-G. Liu and L. Wang, "Differentiable learning of quantum circuit born machine", arXiv:1804.04168 (2018).

[413]ITensor Library (version 2. 1. 1) https://itensor. org/.

[414]J. C. Spall, "An overview of the simultaneous perturbation method for efficient optimization", Johns Hopkins APL Tech. Dig. **19**, 482–492 (1998).

[415]C. J. C. B. Yann LeCun Corinna Cortes, "MNIST handwritten digit database", http://yann. lecun. com/exdb/mnist/.

[416]S. Inc., *SigOpt reference manual*, 2014.

[417]A. D. Córcoles, E. Magesan, S. J. Srinivasan, A. W. Cross, M. Steffen, J. M. Gambetta, and J. M. Chow, "Demonstration of a quantum error detection code using a square lattice of four superconducting qubits", Nat. Commun. **6**, 6979 (2015).

[418]M. Schwarz, K. Temme, and F. Verstraete, "Preparing projected entangled pair states on a quantum computer", Phys. Rev. Lett. **108**, 110502 (2012).

[419]M. Schwarz, K. Temme, F. Verstraete, D. Perez-Garcia, and T. S. Cubitt, "Preparing topological projected entangled pair states on a quantum computer", Phys. Rev. A **88**, 032321 (2013).