MODELING ELECTRON CORRELATION IN QUANTUM CHEMISTRY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Eric Neuscamman

August 2011

MODELING ELECTRON CORRELATION IN QUANTUM CHEMISTRY

Eric Neuscamman, Ph.D.

Cornell University 2011

Accurately predicting the effects of correlations caused by strong interactions between electrons represents one of the key unsolved challenges in quantum chemistry. This thesis details theoretical developments addressing two aspects of this challenge. First, advances in Canonical Transformation (CT) theory address the task of describing weak interactions between the chemically active valence electrons and the chemically inert core electrons when the valence electrons are strongly correlated. Second, new methodologies for Jastrow factor wavefunctions are explored in an attempt to describe strong interactions amongst the valence electrons.

In its original form, CT theory suffered from poorly understood numerical instabilities in strongly interacting systems and poor accuracy relative to coupled cluster theory in weakly interacting systems. Here the numeric issues are identified as consequences of CT theory's central approximation and shown to be similar to the intruder states of many-body perturbation theory. It is demonstrated that these instabilities can be avoided by placing particular restrictions on the theory's degrees of freedom. In weakly interacting systems, we show that using a quadratic, rather than linear, commutator approximation improves accuracy, and that the accuracy in any system can be improved by making use of the system's 3-body reduced density matrix. These improvements are demonstrated in a variety of systems, including N_2, H_2O, NiO, and free base porphin.

Jastrow factor wavefunctions have traditionally been handled using quan-

tum Monte Carlo techniques such as variational Monte Carlo (VMC). Here we explore how VMC can be made to accommodate the unconventionally large numbers of variational parameters present in many-body Jastrow factors and show that the stochastic reconfiguration algorithm can be thus improved by employing the method of conjugate gradients to solve its central linear equation. In addition to extending the reach of VMC, we have also developed non-stochastic methods for working with Jastrow factor wavefunctions that do not rely on random sampling. These methods, similar in design to those used with the coupled cluster wavefunction, produce energies very close to those of VMC in applications to the Heisenberg, spinless Hubbard, and full Hubbard model Hamiltonians.

BIOGRAPHICAL SKETCH

While he was born in Colorado, Eric's childhood was split primarily between Livermore, California, and Beaconsfield, England, due to the itinerant nature of his father's employment. He attended elementary school at the ACS International School in Hillingdon, England before moving back to Livermore, where he attended Mendenhall Middle School and Granada High School. Through the Mathcounts program in middle school, and through subsequent mathematics and science courses, Eric became interested in the use of mathematics to describe the workings of physical systems. Due largely to the spectacular teaching performance of his high school chemistry teacher, Eric decided to pursue chemical engineering at the University of California, Los Angeles, as this field sounded like a good mixture of mathematics and chemistry. During his undergraduate studies, he drifted further and further towards physical chemistry and quantum mechanics, and ended up double majoring in physical chemistry and chemical engineering with a minor in mathematics. Abandoning engineering entirely, Eric applied to graduate schools seeking work in theoretical chemistry. At Cornell, his graduate career has centered on electronic structure theory under the direction of Garnet Chan. Outside of academic life, Eric enjoys spending time with his wife and daughter, reading, and playing strategy games. After the completion of his Ph.D., he will move to the University of California, Berkeley to continue his pursuit of electronic structure theory as a Miller Research Fellow.

For Karen and Peter.

ACKNOWLEDGEMENTS

First and foremost I would like to acknowledge the direction and support I received from my advisor, Dr. Garnet Chan. Garnet provided me with an excellent research environment, offering me the opportunity to explore ideas independently but always standing ready provide direction, explanations, and especially vision to a bewildered young student. Of particular value was Garnet's insuperable optimism, without which I surely would have passed over a number excellent scientific opportunities.

I would also like to thank my wife, Stephanie, for her love, support, and companionship during our time spent in Ithaca. I will never forget her generosity in allowing us to pursue our graduate studies at Cornell, which was not her first choice among the options available to us. For this opportunity, as with so many things in life, I am deeply in her debt.

To all of my group members and collaborators, I would like to extend my earnest gratitude. Specifically, I appreciate the indirect tutelage of Takeshi Yanai, whose beautiful and robust ORZ software taught me most everything I know about writing effective computer programs. Finally, I would like to thank my parents and family, whose unfailing encouragement and support has always been deeply appreciated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The ability to predict the behavior of physical systems from first principles is the ultimate test of scientific understanding. The first step towards meeting this challenge is the cyclical process of making observations and refining hypotheses. Eventually, these hypotheses will be sufficiently accurate to form a theory, at which point the system is often regarded as understood barring any new or unexpected observations. However, distilling the rules governing the system's behavior into a coherent theory is not always sufficient to make accurate predictions about the system, and so the discovery of the correct theory does not necessarily represent a complete understanding. This state of affairs is exemplified in quantum chemistry, where the rules governing electron behavior have been clear for over half a century but have not yet yielded an entirely satisfactory mechanism for predicting the behavior of molecules. In order to effect a complete understanding of quantum chemistry it is therefore necessary to construct such a mechanism, which represents the central task in a field known as electronic structure theory.

Electronic structure theory is the theory of electron behavior in molecules and materials. It can be seen as one corner of the larger realm of quantum physics, in which relativistic effects are ignored due to the relatively slow motion of chemically relevant electrons and atomic nuclei are assumed to be so massive that they can be treated classically. These approximations reduce the mathematics of quantum physics into the non-relativistic Schrödinger equation under the Born-Oppenheimer approximation, which is the foundation upon which electronic structure theory is built. The only variables in this equation

are the coordinates of the electrons, hence the name electronic structure theory. The Hamiltonian operator defining the electronic Schrödinger equation consists of three components: (a) the kinetic energy operator for each electron, (b) the coulombic attraction between electrons and the nuclei, and (c) the coulombic repulsion between each pair of electrons. This final component of the Hamiltonian creates non-trivial many-body correlations in the electronic wavefunction and prevents analytic solutions to the Schrödinger equation for all but the simplest systems. For systems containing significant correlations between electrons, which include high-temperature superconductors, transition metal catalysts, and some light harvesting molecules, even numeric techniques are currently unable to derive accurate predictions from the Schrödinger equation.

Fundamentally, it is the interactions between electrons in the form of the coulomb repulsion that prevents an exact solution to the Schrödinger equation. If these interactions were not present, or if they could be replaced by an average in the form of a repulsive background potential, then the wavefunction would be exactly equal to an antisymmetrized product of single-electron wavefunctions known as a Slater determinant. These one-particle wavefunctions are the ubiquitous orbitals found in chemistry textbooks. In molecular orbital theory, the familiar procedure of creating molecular orbitals through linear combinations of atomic orbitals and then filling them with electrons in the order of increasing energy, the wavefunction is implicitly assumed to be a Slater determinant. Thus molecular orbital theory, and with it a large fraction of chemists' understanding of electronic structure, derives directly from the assumption that electrons can be approximated as non-interacting particles by replacing their coulomb repulsions by some average electrostatic potential.

Is this view valid? From one perspective, the assumption of non-interacting electrons seems absurd. As a thought experiment, consider the probability density for spin-down electrons in a molecule if a spin-up electron has been fixed at some arbitrary point in space. We expect that the probability of finding spin-down electrons in the vicinity of the spin-up electron will be suppressed due to coulombic repulsion, and indeed this is what occurs in nature. This suppression represents a correlation between electrons. However, one can show that in a Slater determinant, there is no correlation between the positions of electrons with opposite spin. Thus this wavefunction will show no suppression of probability density in the vicinity of the fixed electron, which is clearly incorrect and casts doubt on the validity of molecular orbital theory and by extension the chemists' view of electron behavior. However, the success of molecular orbital theory in providing qualitative explanations for many chemical phenomena, such as the all-important formation of chemical bonds, testifies that there must be something to the notion that electrons can be approximated as non-interacting particles.

The resolution to this riddle lies in the relative importance of the three components of the electronic Hamiltonian. In a molecule at equilibrium, the kinetic energy and the strength of the coulombic attraction to the nuclei are so much more favorable in the bonding orbitals than in the anti-bonding orbitals that these two effects overwhelm the repulsive interactions between electrons and yield a wavefunction that for qualitative purposes behaves as if the electrons did not interact with each other at all. Put another way, the energetic separation between the highest occupied and lowest unoccupied orbital (the HOMO/LUMO gap) is often so large that any coulombic repulsion created by doubly occupying the bonding orbitals is negligible compared to the cost of occupying the

LUMO, and so the electrons fill the orbitals according to the Aufbau principle with no regard for electron-electron interactions. The electrons are, however, indistinguishable fermions and thus obey Fermi statistics. The result, for a large HOMO/LUMO gap, is a wavefunction in which the orbitals below the LUMO are placed in a Slater determinant, which creates only the correlations necessary to satisfy Fermi statistics and neglects all other correlations between the electrons. The exact wavefunction could be written as a superposition of all possible Slater determinants (i.e. also those with other choices of which orbitals the electrons occupy), but for systems with a large HOMO/LUMO gap the only determinant in the superposition with any significant weight is the Aufbau determinant. Thus for most molecules at equilibrium, where the bonding and antibonding orbitals are well separated in energy, the Slater determinant approximation in which electrons behave as non-interacting fermions is qualitatively correct.

In general, however, the interactions between electrons and the correlations they produce cannot be neglected. At the very least, making quantitatively accurate predictions about molecular properties requires the inclusion of the correlations that suppress the probability of opposite-spin electrons coming near one another (same-spin electrons are kept apart automatically by the Pauli exclusion principle arising from Fermi statistics). As we have seen, these correlations represent a small correction for molecules with large HOMO/LUMO gaps, and so in these cases it is appropriate and accurate to include them through perturbation theory. One can say that these systems have a simplifying structure (specifically the large HOMO/LUMO gap) that allows their properties to be teased out of the Schrödinger equation through the use of a Slater determinant wavefunction. The pursuit of accurate predictions for the behavior of other sys-

tems can likewise be framed as a search for simplifying structures that can be exploited to extract predictions from the Schrödinger equation. Two types of simplifying structures that are commonly exploited in electronic structure theory are energetic structure and spatial structure.

Most traditional approaches in quantum chemistry take advantage of an underlying energetic order present in the true wavefunction. In the case of molecular orbital theory and large HOMO/LUMO gaps, this order arises from the dominance of the kinetic energy and electron-nuclear attractions over the electron-electron repulsions, which forces the wavefunction to abide by an energetic ordering that dictates the occupation of the molecular orbitals. Improvements to this approach are also based on energetic structure. For example, in practical applications of perturbation theory the perturbed wavefunction is usually written as a linear combination of the Aufbau determinant and all determinants that can be produced by single and double orbital replacements. While the exact wavefunction is a superposition of all possible determinants, the large energetic separation between the bonding and anti-bonding orbitals means that the more electrons that are excited into anti-bonding orbitals, the less energetically favorable and therefore less important the determinant will be. This reasoning is sound for molecules with large HOMO/LUMO gaps, and perturbation theory with single and double replacements typically captures most of the effects of electron correlation in such systems.

Another popular approach based on exploiting energetic structure that is more accurate than molecular orbital theory is the use of an active space. In quantum chemistry, the active space is typically defined as all of a system's valence electrons and orbitals. For example, the active space for the water

molecule could consist of the hydrogen 1s and oxygen 2s and 2p orbitals and the eight electrons that are left over after filling the oxygen 1s orbital, which is called the core. In the active space approach, the wavefunction is not assumed to be dominated by a single determinant, but rather to consist of a superposition of determinants in which the core orbitals are always present and the remaining orbitals are chosen in all possible ways from the active orbitals. This approach exploits the energetic structure deriving from large energy differences between atomic orbitals of different principle quantum numbers. In molecular orbital theory, it was assumed that the behavior of all the electrons was dominated by kinetic energy and electron-nuclear attraction, while in the active space approach only the core electrons are assumed to act in this way. This distinction is critical, as it allows for accurate descriptions of bond dissociation, where there is no longer a large energetic separation between bonding and anti-bonding orbitals to force the electrons into a single determinant. The energetic separation between principle quantum numbers is less dependent on the chemical environment, and so the active space approach is more generally applicable than molecular orbital theory.

The small correlations between the active and core electrons are neglected in the active space wavefunction and usually do not affect the qualitative behavior of a molecule, but they must be accounted for if quantitatively accurate predictions are to be made. Perturbation theory is again a natural approach given that the correlations in question are small. However, the zeroth order wavefunction is now much more complicated than a single Slater determinant, and this complexity limits the applicability of perturbation theory. Unlike a single Slater determinant, the linear combination of determinants that constitutes the active space wavefunction contains non-trivial correlations between the valence

electrons. Thus the probability of simultaneously finding two electrons in two specific positions is no longer a product of two independent probabilities but a non-separable function of the two positions. The same non-separability is true for three and higher particle number probabilities as well. A straightforward application of perturbation theory to an active space wavefunction requires the use of the probability functions for one, two, and three electrons. The fact that these functions do not factorize into one particle functions makes perturbation theory (and most other theories that seek to correct for small correlations between core and active electrons) prohibitively complex for molecules with more than about 20 active valence electrons.

One recent approach to capture the effects of core-active correlations more efficiently is Canonical Transformation theory, which attempts to approximate the troublesome 3-particle probabilities as separable products of 1- and 2-particle probabilities. This approach is a compromise between the exact approach, in which many-particle probabilities do not factorize, and the independent particle approximation, in which they factorize into products of 1-particle probabilities. The theory is written in terms of fermionic quasi-particles that have no correlations between the core and active particles, and thus can be represented exactly using an active space wavefunction. In order for the properties of the quasi-particle system to match those of the electronic system, many-body interactions must be introduced into the quasi-particle Hamiltonian. These interactions are evoked through a unitary (canonical) transformation between electrons and quasi-particles that maps the electronic Hamiltonian onto the quasi-particle Hamiltonian. To make the theory computationally efficient, the quasi-particle Hamiltonian is then simplified by approximating the many-body interactions as products of one and two-body interactions, a process that is closely

related to the factorization of many-body probability functions. While early work on Canonical Transformation theory established that it is capable of providing accurate predictions for systems beyond the reach of more traditional approaches, the original theory was encumbered with numeric difficulties whose origins were not clear. More recently, these subtle difficulties have been shown to arise directly from the theory's central approximation and to correspond broadly with the nefarious intruder states of perturbation theory. Techniques for stabilizing the method's numerics have since been substantially improved, allowing for the theory to be applied to even larger and more complicated systems than its original design allowed.

Although energetic structure is more commonly exploited, a number of approaches to extracting predictions from the Schrödinger equation assume that the wavefunction exhibits a spatial structure instead. One approach has been to assume the opposite extreme from molecular orbital theory: that the coulomb repulsion between electrons dominates over the kinetic energy and electron-nuclear attraction. In a system where this is true, the electrons will localize into different regions of space in order to minimize their repulsion from each other, which creates a wavefunction with significant spatial structure. An example of this localization can be seen when the hydrogen molecule is dissociated: the two nuclei are separated and one electron localizes on each, since the electron repulsion created from placing both electrons between the nuclei outweighs any kinetic or nuclear attraction energy benefits. Wavefunctions known as tensor networks have been developed in the solid state physics and quantum information communities to address systems with this type of spatial structure. The most successful of these wavefunctions, the matrix product state of density matrix renormalization group (DMRG) theory, has also been applied to quantum

chemistry and has proven remarkably effective in systems where the assumptions of molecular orbital theory break down. The DMRG algorithm is able to take special advantage of localization in one spatial dimensional, but is limited when a system extends in two or three dimensions.

Tensor network methods do include generalizations of the matrix product state into more than one dimension, although the topology of these generalizations is fundamentally more complex and precludes the simplifications that make the DRMG algorithm so efficient. Methods such as the pair entangled product state and the multiscale renormalization group ansatz have nonetheless been developed for working with two-dimensional tensor networks, but these methods have so far been too expensive for use in quantum chemistry, with cost-scalings as high as the twelfth power of the system size. For comparison, the cost of the DMRG algorithm scales as the third power of the system size. A relatively simple ansatz, referred to in this thesis as the correlator product state (CPS), has recently been proposed as an alternative to traditional tensor networks for exploiting local structure in the wavefunction. Rather than a tensor contraction algorithm, which forms the basis for most other tensor network methods, correlator product states rely on Monte Carlo methods for their optimization and the evaluation of expectation values. Together, these various tensor network wavefunctions represent attempts to exploit localized spatial structure, which will occur when electron-electron repulsion dominates the other terms in the Hamiltonian.

Naturally, one would like a general method that performs well regardless of what type of underlying structure the Hamiltonian imparts on the system's wavefunction. One way to approach this ideal is by combining traditional

quantum chemistry wavefunctions with tensor networks. The most straightforward such combination has been around for some time and is the Jastrow-Slater wavefunction, in which a Slater determinant is modified by applying Jastrow factors to penalize spatial configurations in which electrons come too close to one another. These Jastrow factors have traditionally been functions of the coordinates of just two electrons, but if they are generalized to functions of multiple electrons' coordinates, they correspond directly to the correlators in the correlator product state. One way forward is therefore to combine the correlator product state, which is designed to efficiently encode spatial structure, with quantum chemistry wavefunctions capable of exploiting energetic structure.

While Jastrow factor wavefunctions have traditionally been optimized using variational Monte Carlo techniques, one must be careful when generalizing to many-body Jastrow factors due to the large increase in the number of wavefunction parameters caused by this generalization. Variational Monte Carlo uses stochastic sampling to construct the energy gradient and possibly the Hessian matrix with respect to the wavefunction's variables. However, many-body Jastrow factors can contain millions of variables, which makes the explicit construction of matrices like the Hessian infeasible and gradient-only techniques such as steepest descent less efficient. Recent research has addressed this difficulty in two ways. First, the approach taken by coupled cluster theory, in which the full Schrödinger equation is projected onto a small but physically important subsection of Hilbert space, was applied to create an optimization algorithm that does not rely on stochastic Monte Carlo techniques. This approach was shown to be highly efficient when the many-body Jastrow factors are local and modest in size. Second, the variational Monte Carlo technique known as stochastic reconfiguration was improved in order to handle much

larger sets of variational parameters. This improvement was achieved by solving the method's central linear equation iteratively using the conjugate gradient algorithm in order to avoid explicitly constructing the respective matrix, which for many-body Jastrow factors is too large to store on a computer. Together, these new optimization strategies have allowed for the successful evaluation of generalized Jastrow factor wavefunctions, delivering a superior ability to simultaneously capture energetic and spatial structure.

As its title suggests, this thesis is focused on capturing the effects of electron correlation on quantum chemistry. This topic is of course much too broad to cover exhaustively, and the discussion will therefore be limited to two particular examples of how correlation is accounted for. In Chapter 2, we will explore Canonical Transformation theory and how to account for the small correlations between core and active electrons that are essential in making accurate quantitative predictions about molecular properties. In Chapter 3, we will consider the many-body generalization of Jastrow factor wavefunctions in order to describe the correlations between valence electrons that alter the qualitative structure of the electronic wavefunction. Together, these chapters will cover both halves of the traditional division of correlation in quantum chemistry.

CHAPTER 2

CANONICAL TRANSFORMATION THEORY

2.1 Introduction

In this chapter we shall explore Canonical Transformation (CT) theory. The presentation is of CT theory as a whole, with the contributions of the current author interspersed throughout the discussion. This approach is taken to provide a holistic perspective of the quadratic commutator approximation, the use of the 3-body RDM, automated derivations, the role of intruder states, and the use of strongly contracted excitations, which together represent the main contributions of the current author. We begin with a relatively broad overview of the difficulty of treating dynamic electron correlation in systems with strongly interacting valence electrons (e.g. multi-reference systems), followed by the motivation behind CT theory and its connections to earlier work. We then proceed to develop the theory in full, exploring its construction as both a theory and a working computational method. Finally, we survey numerical results and compare CT theory to a number of other theories used for multi-reference dynamic correlation.

2.1.1 Dynamic correlation in multi-reference systems

Although accurate descriptions of multi-reference systems' strongly interacting valence electrons are by no means trivial and represent an interesting challenge in their own right (see Chapter 3), such pure-valence descriptions are not sufficient to achieve a quantitatively accurate description of chemical systems. In the

context of quantum chemistry, quantitative accuracy is typically considered to be about 1 kcal mol^{-1} for reaction energies and 0.1 eV for excitation energies. To reaching these exacting accuracies it is critical to capture the electrons' dynamic correlation, which requires a much larger orbital space than the valence-only problem. Indeed, both atomic orbitals with high angular momentum and diffuse Rydberg-like functions are necessary, a requirement that can be understood in several ways. First, the shapes of orbitals should depend on their occupancies: in a configuration where a valence orbital is doubly occupied, we would expect electrons to sit in somewhat different orbits than in the case of single occupancy, reflecting radial and angular correlations. Second, coulomb repulsion can lead to a small probability of electronic configurations with occupancy of non-valence orbitals, so that electrons can better avoid each other. Third, it takes a large number of Gaussian-type functions (quantum chemistry typically uses a Gaussian-type rather than a Slater-type orbital basis) to approximate the sharp wavefunction cusps that occur when two electrons coalesce. These kinds of adjustments to the valence electronic structure constitute "dynamic" correlation, which is not captured by either the Hartree-Fock wavefunction or valence-only wavefunctions such as CASSCF. The quantitative description of electronic structure therefore requires additional theoretical models for dynamic correlation, which can be separated into two groups: those for single-reference systems and those for multi-reference systems. As we shall see, multi-reference dynamic correlation theories have so far been less successful than their single-reference counterparts.

When the valence electronic structure is qualitatively captured by a single electronic configuration, the description of dynamic correlation is well understood. We shall refer to this limit as the single-reference dynamic correlation

problem. Within a wavefunction setting, three common single-reference approaches include

- Møller-Plesset perturbation theory, a Rayleigh-Schrödinger perturbation theory where the Fock operator is the zeroth order Hamiltonian [81]. The perturbation series is usually truncated at second-order (MP2). MP2 calculations have a formal computational scaling of n^5 with the molecular size n.
- Configuration interaction, a variational ansatz formed by a linear combination of configurations including excitations from the reference up to a given excitation rank, typically singles and doubles replacements (CISD) [81]. CISD is not size consistent and is not usually used without a size consistency correction. Approximate size consistent CI theories can be obtained in various ways, such as through the Davidson correction [55, 28], or through the coupled-pair functional [1]. CISD has a formal computational scaling of n^6 with molecular size.
- Coupled cluster theory, which uses an exponential form for the excitation operator [81, 15]. Formally, coupled cluster theory sums many terms in perturbation theory to infinite order, and most importantly all those terms which are necessary for exact size consistency. The most common coupled cluster theory uses singles and doubles excitation operators (CCSD) and has an n^6 computational scaling with molecular size.

Out of these three approaches for single-reference dynamic correlation, coupled cluster theory has established itself as the most satisfactory, both formally and in terms of numerical performance. It is sometimes referred to as the "gold-standard" of dynamic correlation methods [15].

However, this thesis is concerned primarily with systems in which the valence electrons interact strongly, which necessitates a *multi-reference* wavefunction description. We shall refer to the dynamic correlation in this limit as multi-reference dynamic correlation. The process of including corrections for dynamic correlation is now made much more complicated as one cannot exploit the many simplifications that arise from a single-reference starting point. Nonetheless, multi-reference analogues of the single-reference dynamic correlation theories have been considered, including

- Multireference perturbation theory, such as the complete active space perturbation theory (CASPT) [5], multi-reference Møller-Plesset (MRMP) [44], and n-electron valence perturbation theory (NEVPT) [6, 7]. These constitute the most widely applied multi-reference dynamic correlation theories. Unlike in the single-reference limit, some formulations of multi-reference perturbation theory are not exactly size consistent. The computational scaling of multi-reference perturbation theory depends on the specific formulation, but is generally higher than Møller-Plesset theory, with a scaling of at least n^7 for the fully internally contracted variant.
- Multireference configuration interaction (MRCI) [19, 80, 85] with approximate size consistency corrections (MRCI+Q) and averaged coupled pair functionals (MRACPF, MRAQCC) [37, 82]. These are divided into two types: internally contracted (the variational space is formed from excitation operators acting on a single starting multi-reference wavefunction) or externally contracted (the variational space is formed by excitations from all determinants in the active space). The computational scalings are formulation dependent but are typically n^{10} and e^n, respectively.
- Multireference coupled cluster theories (MRCC) [74]. There are many vari-

ants of MRCC which reflects the mathematical difficulties in extending the CC formalism beyond a single-reference starting point. In general, MRCC theories have computational costs that are even higher than MRCI theory. While there is much activity in this field, MRCC methods have yet to be applied to realistic problems.

From this brief analysis, it is clear that the description of dynamic correlation with multi-reference wavefunctions is much less satisfactory than in the single-reference case. Because of this handicap, in certain situations using a multi-reference (e.g. complete active space) description of the valence electrons - in principle a more flexible theoretical framework - can lead to a poorer quantitative accuracy than a single-reference description. Consider, for example, a chemical transformation where the electronic structure changes from single-reference to multi-reference in character. A typical multi-reference description, e.g. through CASPT2, obtains the static correlation of the active valence orbitals exactly (by virtue of the CASSCF treatment) but recovers only a small portion of the dynamic correlation of the external orbitals. This is appropriate when the pure-valence static correlation is larger than the external contributions, but in the single-reference limit all the correlation is dynamic, and the external contributions may in fact be larger, given that there is no degeneracy in the valence space and there are many more external orbitals. A single-reference coupled cluster description provides a (perturbatively) high-order description of dynamic correlation for *all* degrees of freedom. It is likely that in the single-reference limit of the chemical transformation, the coupled cluster treatment would capture a larger portion of the correlation energy than the CASPT2 description, unless a very large active space were to be used. We see that CASPT2 is, in essence, biased against the single-reference limit, due to the inadequate

treatment of dynamic correlation. This problem has recently been demonstrated in studies on the $[Cu_2O_2]^{2+}$ isomerization curve, where CASPT2 predicts an unphysical energy minimum [26, 32, 77, 90]. Even when such behavior is not of concern, multi-reference dynamic correlation theories are much more costly than their single-reference counterparts, limiting their application to very small systems. These various unsatisfactory aspects of existing multi-reference dynamic correlation theories clearly motivate the development of a new theoretical model.

2.1.2 Desirable features for a dynamic correlation theory

Before considering the specifics of canonical transformation theory, we should first consider what features one desires in a multi-reference dynamic correlation theory. Such a theory should

- be size consistent (to allow meaningful calculations on large systems),
- not bias the chemistry towards multi-reference electronic structure by underestimating the dynamic correlation in single-reference settings,
- reduce in single-reference settings to a coupled-cluster like theory,
- have a reasonable scaling with system size, such as CCSD's n^6 scaling.

The canonical transformation theory of dynamical correlation is designed to obtain these features. In particular, it

- uses an exponential excitation operator that preserves size consistency.

- provides a coupled cluster-like description of dynamic correlation, guaranteeing an accurate description in single-reference systems.
- has a favorable computational scaling of n^6 in the system size that is achieved through the use of operator and cumulant decompositions.

2.1.3 Connections to earlier work

Canonical transformation theory derives from several earlier theoretical developments. The main ingredients of canonical transformation theory are (i) a unitary exponential form for the dynamic correlation, (ii) operator and cumulant decompositions to simplify the energy and amplitude equations, and (iii) an emphasis on an effective Hamiltonian picture of the dynamic correlation. The theory's initial motivation came from White's work on numerical canonical transformations [86]. When rewritten in an appropriate form this can be seen to contain both (i) and (iii) above, but without the systematic simplifications presented by operator and cumulant decompositions. However, there are many other developments in quantum chemistry with a direct connection to canonical transformation theory, including

- Coupled cluster theory. While the exponential operator in ordinary CC theory is not unitary, unitary variants of coupled cluster theory were explored in a single-reference setting by Kutzelnigg [51, 52], Bartlett [13, 84, 83], and Pal [73, 72]. Multireference unitary coupled cluster theory was developed by Simons [10, 11, 46].
- Operator decompositions and the generalized Wick's theorem, introduced by Mukherjee and Kutzelnigg [63, 64, 53].

- Density matrix and cumulant decompositions [25, 68, 58, 53, 54]. While their most widespread use in conjunction with the density equation, also known as the contracted Schrödinger equation [25, 68, 58], is quite different from their use in canonical transformation theory, there is a close connection to Kutzelnigg and Mukherjee's irreducible contracted Schrödinger equation theory [65] and the related anti-hermitian contracted Schrödinger equation theory introduced by Mazziotti [59].
- There is much earlier work on effective Hamiltonians, including Freed's effective valence Hamiltonian theory [35], Kirtman and Hoffmann's generalized van Vleck theories [49, 45], and the equation-of-motion coupled cluster theory and symmetry-adapted cluster configuration interaction theory [67, 38].

2.2 Constructing canonical transformation theory

2.2.1 Basic formalism

CT theory seeks to find the unitary transformation that maps a reference wavefunction $|\Psi_0\rangle$ onto the true wavefunction $|\Psi\rangle$, as shown in Eq. (2.1).

$$|\Psi\rangle = e^A|\Psi_0\rangle \tag{2.1}$$

The reference function is usually assumed to contain the correct static correlation for the problem and thus the transformation operator e^A is thought of as introducing the dynamic correlation. The transformation in Eq. (2.1) is of the coupled cluster form, but unlike in CC theory, CT theory is constructed in such a way that the only information we require from the reference function is its

1- and 2-body reduced density matrices (RDMs). This allows CT theory to efficiently treat dynamic correlation not only for the Hartree-Fock reference but also for multi-configurational references, such as GVB, CASSCF, and DMRG-SCF wavefunctions. In singles and doubles CT, the unitary transformation is built from the exponential of some combination of anti-symmetric single and double excitation operators $\hat{o}$,

$$A = -A^\dagger = \sum_i C_i \hat{o}_i. \tag{2.2}$$

These operators can in principle be any set of one and two body excitation operators $(a_p^\dagger a_q - a_q^\dagger a_p)$ and $(a_p^\dagger a_q^\dagger a_s a_r - a_r^\dagger a_s^\dagger a_q a_p)$. However, if we assume that the reference $|\Psi_0\rangle$ correctly describes the static correlation in some complete active space, it is natural to consider only those excitations which change the occupancies of at least 1 external (non-active) orbital. Denoting core orbitals by c, active by a, and virtual by v, and using the notation $a_q^p = a_p^\dagger a_q$ and $a_{rs}^{pq} = a_p^\dagger a_q^\dagger a_s a_r$, the full set of possible singles and doubles excitations are given by the following, in which all indices are implicitly summed over.

$$\begin{aligned}
A = \quad & A_{c_1c_2}^{a_1a_2}(a_{c_1c_2}^{a_1a_2} - a_{a_1a_2}^{c_1c_2}) + A_{a_3c_1}^{a_1a_2}(a_{a_3c_1}^{a_1a_2} - a_{a_1a_2}^{a_3c_1}) && \text{core-active} \\
& + A_{c_1}^{a_1}(a_{c_1}^{a_1} - a_{a_1}^{c_1}) && \text{core-active} \\
& + A_{a_1a_2}^{v_1v_2}(a_{a_1a_2}^{v_1v_2} - a_{v_1v_2}^{a_1a_2}) + A_{a_1a_2}^{a_3v_1}(a_{a_1a_2}^{a_3v_1} - a_{a_3v_1}^{a_1a_2}) && \text{virtual-active} \\
& + A_{a_1}^{v_1}(a_{a_1}^{v_1} - a_{v_1}^{a_1}) && \text{virtual-active} \\
& + A_{a_1c_1}^{a_2v_1}(a_{a_1c_1}^{a_2v_1} - a_{a_2v_1}^{a_1c_1}) + A_{a_1c_1}^{v_1v_2}(a_{a_1c_1}^{v_1v_2} - a_{v_1v_2}^{a_1c_1}) && \text{core-virtual-active} \\
& + A_{c_1c_2}^{a_1v_1}(a_{c_1c_2}^{a_1v_1} - a_{a_1v_1}^{c_1c_2}) && \text{core-virtual-active} \\
& + A_{c_1c_2}^{v_1v_2}(a_{c_1c_2}^{v_1v_2} - a_{v_1v_2}^{c_1c_2}) + A_{c_1}^{v_1}(a_{c_1}^{v_1} - a_{v_1}^{c_1}) && \text{core-virtual}
\end{aligned} \tag{2.3}$$

Note that each excitation operator (e.g. $a_{a_1}^{v_1}$) must be accompanied by a de-excitation operator ($a_{v_1}^{a_1}$) in order to make CT theory unitary. We will typically

refer to the combined operators $(a_{a_1}^{v_1} - a_{v_1}^{a_1})$ as excitation operators, but the reader should remember that they are really anti-symmetrized excitation operators that contain a de-excitation component. In practice, further restrictions of the operators are used to simplify their optimization (see Sec. 2.4).

Instead of solving the electronic Schrödinger equation in the direct form,

$$He^A|\Psi_0\rangle = Ee^A|\Psi_0\rangle, \tag{2.4}$$

CT theory works with a quasi-particle Schrödinger equation with an effective Hamiltonian,

$$e^{-A}He^A|\Psi_0\rangle = \bar{H}|\Psi_0\rangle = E|\Psi_0\rangle. \tag{2.5}$$

Eq. (2.5) is interpreted as saying that there exists a reference function of quasi-particles, interacting via the effective quasi-particle Hamiltonian $\bar{H}$, which has the same energy as the original system of electrons interacting through the bare Hamiltonian H. This canonical (unitary) transformation from electrons to quasi-particles moves the complexity of the correlation problem from the wavefunction to the Hamiltonian. For electrons, the known Hamiltonian has a number of integrals proportional to the fourth power of the number of electrons, while the exact wavefunction (which must be determined) has a complexity in its determinantal expansion that is exponential in the electron number. For quasi-particles, the exact wavefunction $|\Psi_0\rangle$ is the comparably simple reference wavefunction (which is known) while the unknown effective Hamiltonian has exponential complexity. This complexity of the effective Hamiltonian is revealed by the Baker-Campbell-Hausdorff (BCH) expansion,

$$\bar{H} = H + [H, A] + \frac{1}{2!}[[H, A], A] + \frac{1}{3!}[[[H, A], A], A] + \ldots, \tag{2.6}$$

where each commutator creates a successively higher particle rank operator. We therefore see that the effective Hamiltonian can in general contain interactions

between any number of quasi-particles (up to the maximum number present in the system). In contrast, electrons only interact pairwise in the bare Hamiltonian.

So far in our development, the canonical transformation has only accomplished a transfer of complexity from the wavefunction to the Hamiltonian. Where this conversion becomes useful is in devising approximations based on our intuition about electrons and quasi-particles. For a sufficiently accurate reference wavefunction, A will be small and electrons and quasi-particles will have similar physics. We can therefore make an approximation that the quasi-particles, like electrons, have no direct interactions more complex than pairwise interactions, which ensures that the effective Hamiltonian $\bar{H}$ has the same quartic complexity as the original Hamiltonian. This is the defining approximation of CT theory.

Naturally, we want to retain the effects of the higher than pairwise interactions in some average way. We achieve this in the effective Hamiltonian by approximating 3- and higher-body interactions (operators) with products of 1- and 2-body interactions, a procedure we refer to as operator decomposition. In a similar manner, the 3- and higher-body RDMs that arise in expectations values of the effective Hamiltonian are approximated by neglecting 3- and higher-body cumulants, which can be seen as the deviations from effective 2-body correlations. We will delay discussion of the details of operator and cumulant decompositions until later sections. For now, the reader should understand that these approximations retain the average effects of many-particle interactions and that they are essential for achieving an n^6 cost scaling.

How do we obtain the optimal canonical transformation to $\bar{H}$? To answer

this question, we return to the quasi-particle Schrödinger equation. Recalling that A is parameterized in terms of a set of coefficients C_i, we require that these satisfy Eq. (2.5) in a projective sense with respect to the first-order interacting space $\{\hat{o}_i|\Psi_0\rangle\}$. This yields the coupled cluster style amplitude equations $\langle\Psi_0|\hat{o}_i^\dagger \bar{H}|\Psi_0\rangle = 0$, which we rearrange to form the CT amplitude equations,

$$\langle\Psi_0|[\bar{H}, \hat{o}_i]|\Psi_0\rangle = 0, \tag{2.7}$$

which take the form of generalized Brillouin conditions [50]. In practice, solving these equations is not trivial and requires additional restrictions on the operators $\hat{o}_i$, which we discuss in detail in Sec. 2.4.

Once the amplitude equations have been solved, the amplitude operator A is used to construct the effective Hamiltonian using the BCH expansion, Eq. (2.6), modified by operator decompositions as explained in later sections. The CT energy is then defined by tracing the (approximate) effective integrals with the reference function's RDMs,

$$E = \langle\Psi_0|\bar{H}|\Psi_0\rangle. \tag{2.8}$$

To summarize, the CT energy is found by first solving Eq. (2.7) for the optimal transformation, and then evaluating the trace of the effective Hamiltonian with the reference wavefunction's RDMs. By employing the operator and cumulant decompositions, these steps may be completed in n^6 time using only the reference function's 1- and 2-body RDMs and the 1- and 2-body operators of the electronic Hamiltonian. The details of how this is achieved are the subject of the next several sections.

2.2.2 Operator decomposition

As discussed above, the central approximation in CT theory is to replace the quasi-particles' effective Hamiltonian with an approximation containing only 1- and 2-body operators. Using Mukherjee and Kutzelnigg's formalism of extended normal ordering (ENO) [63, 64, 53], it is possible to construct an approximation for a 3-body operator that has the same expectation value as the original operator. Just as a traditionally normal ordered operator has a zero expectation value with respect to the true vacuum, an ENO operator has a zero expectation value with respect to a general (possibly multi-configurational) reference state. In extended normal ordered theory, a 3-body operator may be written as the sum of an ENO 3-body operator, a combination of ENO 1- and 2-body operators, and the expectation value of the original operator in the form of a 3-body RDM element. This can be seen through the following definitions of the 1-, 2-, and 3-body ENO operators $\tilde{a}^{p_1}_{q_1}$, $\tilde{a}^{p_1p_2}_{q_1q_2}$, $\tilde{a}^{p_1p_2p_3}_{q_1q_2q_3}$, in which γ represents the reference function's RDMs, e.g. $\gamma^p_q = \langle\Psi_0|a^p_q|\Psi_0\rangle$.

$$\tilde{a}^{p_1}_{q_1} = a^{p_1}_{q_1} - \gamma^{p_1}_{q_1} \tag{2.9a}$$

$$\tilde{a}^{p_1p_2}_{q_1q_2} = a^{p_1p_2}_{q_1q_2} - \sum(-1)^x \gamma^{p_1}_{q_1}\tilde{a}^{p_2}_{q_2} - \gamma^{p_1p_2}_{q_1q_2} \tag{2.9b}$$

$$\tilde{a}^{p_1p_2p_3}_{q_1q_2q_3} = a^{p_1p_2p_3}_{q_1q_2q_3} - \sum(-1)^x \gamma^{p_1}_{q_1}\tilde{a}^{p_2p_3}_{q_2q_3} - \sum(-1)^x \gamma^{p_1p_2}_{q_1q_2}\tilde{a}^{p_3}_{q_3} - \gamma^{p_1p_2p_3}_{q_1q_2q_3} \tag{2.9c}$$

These equations make use of the notation $\sum(-1)^x A^{p_1p_2\cdots}_{q_1q_2\cdots} B^{p_kp_{k+1}\cdots}_{q_kq_{k+1}\cdots}\ldots$, which implies that there is one term for each unique partitioning of the indices among the objects $(A, B, \ldots)$ in which p_i are kept on top and q_i on bottom. For each permutation of the indices from their original positions, a factor of (-1) is applied. (See Ref. [69] for some examples of this rule.) We recognize that the 3-body ENO operator represents 3-body fluctuations from the operator's average 1- and 2-particle like behavior, represented by the RDM element and the 1- and

2-body ENO operators. If the final wavefunction is close to the reference function (which is our working assumption in CT theory), we hope that the effect of neglecting these 3-body fluctuations is small as their expectation value with respect to the reference is zero. Thus we may approximate a 3-body operator by neglecting its 3-body ENO component as follows,

$$a^{p_1p_2p_3}_{q_1q_2q_3} \approx \sum (-1)^x \gamma^{p_1}_{q_1} \tilde{a}^{p_2p_3}_{q_2q_3} + \sum (-1)^x \gamma^{p_1p_2}_{q_1q_2} \tilde{a}^{p_3}_{q_3} + \gamma^{p_1p_2p_3}_{q_1q_2q_3}. \tag{2.10}$$

Now only constants and 1- and 2-body operators remain, as originally desired. This decomposition will be used in order to create efficient approximations to the BCH expansion and amplitude equations. Before we do so, however, we must deal with the remaining 3-body RDM elements, which must be removed if we wish to achieve an n^6 cost scaling.

2.2.3 Cumulant decomposition

So far, our commutator approximation requires the use of the reference function's 3-body RDM. In systems with large numbers of strongly interacting valence electrons, the reference wavefunction's 3-body RDM is not always readily available. we would like to limit ourselves to using only the reference's 1- and 2-body RDMs. To avoid using it and to achieve n^6 cost scaling, we turn to the theory of density matrix cumulants, [25, 54, 68] which represent the irreducible correlations of a given particle rank present in the RDMs. If n-body interactions are unimportant, then neglecting the n-body cumulant creates a reasonable approximation for the n-body RDM consisting of antisymmetric products of lower particle rank RDMs. This approximation technique has been most widely used in contracted Schrödinger equation (CSE) theories [58, 23, 66, 25, 59] where the

1- and 2-body RDMs are optimized directly and the 3- and possibly 4-body RDMs are approximated by neglecting the 3- and 4-body cumulants. In CT theory, we neglect the 3-body cumulant in the same manner, reducing the 3-body RDM that arises from the 3-body operator decomposition to a collection of antisymmetric products of 1- and 2-body RDMs. Therefore, the approximation of the 3-body operator shown in Eq. (2.10) becomes [69]

$$\begin{aligned}
\left(a^{p_1p_2p_3}_{q_1q_2q_3}\right)_{1,2} = \quad & \sum (-1)^x \gamma^{p_1}_{q_1} a^{p_2p_3}_{q_2q_3} && (\ 9 \text{ terms}) \\
+ \ & \sum (-1)^x \gamma^{p_1p_2}_{q_1q_2} a^{p_3}_{q_3} && (\ 9 \text{ terms}) \\
-2 & \sum (-1)^x \gamma^{p_1}_{q_1} \gamma^{p_2}_{q_2} a^{p_3}_{q_3} && (18 \text{ terms}) \\
- \ & \sum (-1)^x \gamma^{p_1}_{q_1} \gamma^{p_2p_3}_{q_2q_3} && (\ 9 \text{ terms}) \\
+4 & \sum (-1)^x \gamma^{p_1}_{q_1} \gamma^{p_2}_{q_2} \gamma^{p_3}_{q_3}, && (\ 6 \text{ terms})
\end{aligned} \tag{2.11}$$

where the number of terms in each sum is given at the right. In this thesis we will make extensive use of the notation $X_{1,2}$ to denote the approximation of X in which all 3- and higher-body cumulants and ENO operators have been neglected. This decomposition now allows us to approximate any 3-body operator as a sum of 0-, 1-, and 2-body operators using only the 1- and 2-body RDMs of the reference function. The analogous approximation for spin free operators can be arrived at in the same manner by neglecting the spin free 3-body ENO operator and cumulant.

While the use of cumulants in CT theory is similar to that in CSE theory, there are some important differences that help CT theory avoid common pitfalls associated with CSE methods. One such issue is the dependence of N-representability error in the approximated 3-body RDM on basis set size, as explored by Harris [41] and Herbert [43]. As opposed to CSE methods, which approximate the final wavefunction's 3-body RDM in the entire orbital space,

CT theory approximates only the reference wavefunction's 3-body RDM in the active space. This greatly reduces the problem of basis set dependence, as the reference wavefunction's active space RDMs change little with the basis set. Indeed, Yanai and Chan [89] showed that CT theory's accuracy is not affected when changing from a double- to triple-zeta basis set in the nitrogen dimer. An additional difference from CSE methods is that CT theory optimizes an excitation operator (as in CC theory) that can modify the active space correlation only indirectly via its coupling to the external space, rather than the active space RDMs directly, as is done in CSE. It is therefore reasonable to expect N-representability problems to be less prevalent in CT theory, and indeed this is found to be the case in practice.

2.2.4 An efficient commutator approximation

Our strategy for simplifying the equations of CT theory will be to replace the commutator $[H, A]$ with the approximation $[H, A]_{1,2}$, in which we have decomposed all 3-body operators using Eq. (2.11). The upshot is that $[H, A]_{1,2}$ contains only 0-, 1-, and 2-body operators and can be evaluated in n^6 time. We will now take advantage of this commutator approximation by inserting it into the BCH expansion and amplitude equations. If we apply our approximation to every commutator in Eq. (2.6), we will have

$$\bar{H} \approx H + [H, A]_{1,2} + \frac{1}{2!}[[H, A]_{1,2}, A]_{1,2} + \frac{1}{3!}[[[H, A]_{1,2}, A]_{1,2}, A]_{1,2} + \dots . \qquad (2.12)$$

Notice that this approximate expansion can be evaluated recursively by reusing the result of each term when evaluating the next higher term in the series. Specifically, if we let $\bar{H}^{(n)}$ represent the approximation to the nth term in the

expansion, then we have the recursive formula $\bar{H}^{(n)} = \frac{1}{n}[\bar{H}^{(n-1)}, A]_{1,2}$. The total effective Hamiltonian is then approximated as

$$\bar{H} \approx \sum_{n=0}^{\infty} \bar{H}^{(n)}, \tag{2.13}$$

with $\bar{H}^{(0)}$ equal to the unmodified electronic Hamiltonian H. While this sum is formally infinite, the amplitude operator A is assumed to be small and in practice the terms decay quickly enough that only a finite number need to be evaluated (typically 8 to 10 for a precision of 10^{-9} Hartrees). As each approximate commutator evaluation takes n^6 time, the entire evaluation of the approximate effective Hamiltonian using Eq. (2.13) has a cost scaling of n^6. With a 1- and 2-body approximation to the effective Hamiltonian now in hand, the amplitude equations can be evaluated in n^6 time as well by approximating the commutator in Eq. (2.7) as $[\bar{H}, \hat{o}_i]_{1,2}$. Thus by a recursive application of the commutator approximation we see that the working equations of CT theory can be evaluated in n^6 time using only the 1- and 2-body RDMs of the reference wavefunction and the 1- and 2-body operators of the bare electronic Hamiltonian.

2.2.5 The quadratic commutator approximation

A natural question when studying CT theory is whether a superior operator decomposition exists. One may expect the recursive use of the 3-body operator decomposition in Eq. (2.11) to produce additional error beyond that of the decomposition itself, leading one to wonder if a *quadratic* commutator approximation, in which the decomposition is delayed until after the second commutator, would be more accurate. This approximation could be denoted by $[[H, A], A]_{1,2}$, where we decompose both the 3- and 4-body operators resulting from the dou-

ble commutator into 1- and 2-body operators. The hope is that by delaying the decomposition error to a higher order in A, the effect on the CT energy would be reduced. Testing this hypothesis requires developing a 4-body operator decomposition akin to Eq. (2.11), which we will arrive at in the same way. If one writes a 4-body operator as a sum of 1-, 2-, 3-, and 4-body ENO operators and a 4-body RDM (which involves no approximation) then the natural approach would be to neglect the 3- and 4-body ENO operators and the 3- and 4-body cumulants. Doing so results in the following decomposition [69]

$$\begin{aligned}
\left(a^{p_1p_2p_3p_4}_{q_1q_2q_3q_4}\right)_{1,2} = \quad & \sum (-1)^x \gamma^{p_1p_2}_{q_1q_2} a^{p_3p_4}_{q_3q_4} && \text{(36 terms)} \\
-2 & \sum (-1)^x \gamma^{p_1}_{q_1} \gamma^{p_2}_{q_2} \gamma^{p_3}_{q_3} a^{p_4}_{q_4} && \text{(96 terms)} \\
- & \sum (-1)^x \gamma^{p_1p_2}_{q_1q_2} \gamma^{p_3p_4}_{q_3q_4} && \text{(18 terms)} \\
+6 & \sum (-1)^x \gamma^{p_1}_{q_1} \gamma^{p_2}_{q_2} \gamma^{p_3}_{q_3} \gamma^{p_4}_{q_4} && \text{(24 terms)},
\end{aligned} \tag{2.14}$$

which we see is significantly more complex than the 3-body operator decomposition of Eq. (2.11). Nonetheless, one may proceed by approximating all 3-body operators resulting from $[[H, A], A]$ using Eq. (2.11) and all 4-body operators using Eq. (2.14), which defines the quadratic commutator approximation $[[H, A], A]_{1,2}$.

Applying the quadratic commutator approximation to the BCH expansion is not as straightforward as with its linear counterpart. Our general strategy will be to delay as far as possible the operator decompositions to higher orders of A, in the hopes that increasing orders of A (which we assume is small) will mitigate any decomposition error. Following this strategy results in the following

approximate BCH expansion,

$$\bar{H} \approx H + [H,A]_{1,2} + \frac{1}{2!}[[H,A],A]_{1,2} + \frac{1}{3!}[[[H,A],A]_{1,2},A]_{1,2} + \frac{1}{4!}[[[[H,A],A]_{1,2},A],A]_{1,2} + \ldots \tag{2.15}$$

where we have employed the linear commutator approximation to deal with terms having an odd power of A. We must also modify the approximation to the amplitude equations in order to benefit from using the quadratic commutator. By substituting the BCH expansion into Eq. (2.7) and delaying decompositions to the highest possible order in A, we arrive at the following modified amplitude equations,

$$\begin{aligned} 0 = \langle [H,\hat{o}_i]_{1,2} \rangle + \langle [[H,A],\hat{o}_i]_{1,2} \rangle &+ \frac{1}{2!}\langle [[[H,A],A]_{1,2},\hat{o}_i]_{1,2} \rangle \\ &+ \frac{1}{3!}\langle [[[[H,A],A]_{1,2},A],\hat{o}_i]_{1,2} \rangle \\ &+ \frac{1}{4!}\langle [[[[[H,A],A]_{1,2},A],A]_{1,2},\hat{o}_i]_{1,2} \rangle \\ &+ \ldots, \end{aligned} \tag{2.16}$$

where we have used the notation $\langle X \rangle$ as a shorthand for $\langle \Psi_0 | X | \Psi_0 \rangle$. As with their linear counterparts, the quadratic approximations in Eqs. (2.15) and (2.16) can be evaluated recursively, although the recursion is not as simple as the one from Eq. (2.13). Somewhat surprisingly, the quadratic commutator approximation $[[H,A],A]_{1,2}$ has an evaluation cost that scales as n^6, although the number of tensor contractions involved is far higher than for the linear commutator $[H,A]_{1,2}$ (see Sec. 2.2.7). Thus the cost of quadratic CT theory (denoted QCTSD when using singles and doubles) is only a constant factor higher than that of linear CT theory (denoted LCTSD).

One still must answer the question of whether QCTSD delivers superior accuracy over LCTSD. We shall see in our results that the relative accuracy of the

two methods is system dependent, and that for strongly multi-reference systems QCTSD does not improve on LCTSD and therefore is probably not worth the extra cost. Furthermore, the use of the quadratic commutator does not remove the intruder states that arise when solving the CT amplitude equations, which was is some senses the primary motivation behind the development of QCTSD. For more on the intruder state problem, the reader is referred to Sec. 2.4. In single-reference systems QCTSD is clearly superior to LCTSD, as predicted in the perturbative analysis of Sec. 2.3.2. Specifically, QCTSD is accurate to the same order of perturbation theory as coupled cluster with singles in doubles (CCSD), while LCTSD is accurate to one lower order in the perturbation. Thus with QCTSD, we can realize the goal that CT theory reduces to something comparable to coupled cluster theory in single-reference systems.

2.2.6 Incorporating the 3-body RDM

For systems in which it is feasible to construct and work with the reference wavefunction's 3-body RDM, both the linear and quadratic commutators can be made more accurate. We do so by defining new decompositions for 3- and 4-body operators in which the 3-body cumulant is retained. We will continue to neglect the 4-body cumulant and 3- and 4-body ENO operators. By this pre-

scription, the decompositions become

$$\begin{aligned}
\left(a^{p_1p_2p_3}_{q_1q_2q_3}\right)_{1,2,(3)} = & \sum(-1)^x\gamma^{p_1}_{q_1}a^{p_2p_3}_{q_2q_3} && (\ 9 \text{ terms}) \\
& + \sum(-1)^x\gamma^{p_1p_2}_{q_1q_2}a^{p_3}_{q_3} && (\ 9 \text{ terms}) \\
& -2\sum(-1)^x\gamma^{p_1}_{q_1}\gamma^{p_2}_{q_2}a^{p_3}_{q_3} && (18 \text{ terms}) \\
& -2\sum(-1)^x\gamma^{p_1}_{q_1}\gamma^{p_2p_3}_{q_2q_3} && (\ 9 \text{ terms}) \\
& +6\sum(-1)^x\gamma^{p_1}_{q_1}\gamma^{p_2}_{q_2}\gamma^{p_3}_{q_3} && (\ 6 \text{ terms}) \\
& + \gamma^{p_1p_2p_3}_{q_1q_2q_3},
\end{aligned} \tag{2.17}$$

$$\begin{aligned}
\left(a^{p_1p_2p_3p_4}_{q_1q_2q_3q_4}\right)_{1,2,(3)} = & \sum(-1)^x\gamma^{p_1p_2}_{q_1q_2}a^{p_3p_4}_{q_3q_4} && (\ 36 \text{ terms}) \\
& + \sum(-1)^x\gamma^{p_1p_2p_3}_{q_1q_2q_3}a^{p_4}_{q_4} && (\ 16 \text{ terms}) \\
& - \sum(-1)^x\gamma^{p_1}_{q_1}\gamma^{p_2p_3}_{q_2q_3}a^{p_4}_{q_4} && (144 \text{ terms}) \\
& - \sum(-1)^x\gamma^{p_1p_2}_{q_1q_2}\gamma^{p_3p_4}_{q_3q_4} && (\ 18 \text{ terms}) \\
& +6\sum(-1)^x\gamma^{p_1}_{q_1}\gamma^{p_2}_{q_2}\gamma^{p_3}_{q_3}\gamma^{p_4}_{q_4}, && (\ 24 \text{ terms})
\end{aligned} \tag{2.18}$$

where the notation $X_{1,2,(3)}$ denotes the approximation to X in which all 4- and higher-body cumulants and 3- and higher-body ENO operators have been neglected. We are now in a position to define two further variants of CT theory, L3CTSD and Q3CTSD, that make use of the reference function's 3-body RDM in order to decrease the decomposition error. These variants are constructed by replacing the linear and quadratic commutator approximations with more accurate approximations based on Eqs. (2.17) and (2.18) as follows

$$[H,A]_{1,2} \rightarrow [H,A]_{1,2,(3)} \tag{2.19}$$

$$[[H,A],A]_{1,2} \rightarrow [[H,A],A]_{1,2,(3)}. \tag{2.20}$$

With these replacements, the L3CTSD and Q3CTSD methods derive directly from LCTSD and QCTSD, respectively. In practice, the inclusion of the exact

3-body RDM is found to improve the accuracy of the CT energy, as shown in Sec. 2.5.

2.2.7 Automatic derivation and code generation

The most difficult task when implementing CT theory is deriving and encoding the tensor contractions necessary to compute the commutator approximations. The scale of this task becomes clear when one considers that the approximate quadratic commutator $[[H, A], A]_{1,2}$ (see Sec. 2.2.5) contains over 16,000 unique terms, as shown in Table 2.1. For many of the expressions present in CT theory, we therefore use a computer program to automate the derivation process. This program was developed by the present author and may be found in the erratum of Ref [69]. The central tasks necessary to implement a commutator approximation for use in a CT computer program are:

1. expand all commutators
2. normal order creation and destruction operators
3. replace 3- and/or 4-body operators by their decompositions
4. combine like terms
5. choose an efficient tensor contraction order for each term

All of these tasks were automated to allow for the efficient and accurate implementation of quadratic CT theory.

The most difficult step in deriving the terms involved in a commutator approximation is comparing terms to determine if they may be combined. Term

Table 2.1: Commutators in CT Theory

Commutator	# Unique Terms	Evaluation Cost
$[H,A]_{1,2}$	298	$O(n^6)$
$[H,A]_{1,2,(3)}$	314	$O(n^7)$
$[[H,A],A]_{1,2}$	16,935	$O(n^6)$
$[[H,A],A]_{1,2,(3)}$	23,245	$O(n^8)$

comparison is made difficult by the symmetry of the tensors and the freedom to rename dummy indices. As an example, consider the term

$$\sum_{\substack{a_1 a_2 a_3 \\ i_1 i_2 i_3 i_4 i_5}} \gamma_{i_3}^{a_3} \gamma_{i_1 i_5}^{i_4 a_1} v_{i_3 i_5}^{i_2 i_4} A_{a_2}^{a_1} A_{i_2 i_1}^{a_2 a_3} \tag{2.21}$$

which occurs when deriving the explicit expression for $[[H,A],A]_{1,2}$. The tensors in this term have 2-, 8-, 8-, 1-, and 4-fold symmetries for a total of 512 equivalent index arrangements. Further, because of the summation, all the indices are dummy indices. In order to combine like terms, the program first transforms each term into a unique canonical form (here the word canonical has no relation to CT theory). The rules for writing a term in canonical form are based on a lexicographic ordering of the tensors and their indices, with some special rules in the case of a repeated tensor name. This definition of canonical form is arbitrary. The key is to choose a form such that for each term encountered in CT theory, there is only one unique way to write it in canonical form. The canonical form for our example term is

$$\sum_{\substack{abcd \\ efgh}} A_b^a A_{de}^{bc} \gamma_f^c \gamma_{dh}^{ag} v_{fh}^{eg}. \tag{2.22}$$

The comparison of terms may be seen as a problem of graph isomorphism in which the tensors represent labeled vertices and the summation indices labeled

edges. As the general problem of graph isomorphism has no known solutions of polynomial complexity, we expect that our problem of term combination will become unmanageable for a sufficiently large number of tensors. However, the terms involved in CT theory are simple enough that the cost of converting them to canonical form is not prohibitive. We do note, though, that term comparison is the most expensive step in the automatic derivation process.

After all of the necessary terms have been derived, it is important to choose an efficient order in which to carry out the contractions for a specific term. In our example term, one must be careful not perform the n^7 cost contraction $\sum_e A^{bc}_{de} v^{eg}_{fh}$. Instead, by first contracting A^a_b and γ^c_f with A^{bc}_{de}, the contractions with γ^{ag}_{dh} and v^{eg}_{fh} can be completed for only an n^6 cost. While one could exhaustively search all possible contraction orders to guarantee maximal efficiency, this process has a potentially unaffordable $N!$ cost for N tensors. Instead, our program looks for the least expensive contraction pair out of all the tensors and chooses it as the first contraction, then repeats this process until all tensors have been assigned a position in the contraction ordering. While this process is not guaranteed to find the most efficient contraction ordering, it can be completed quickly as it has only an N^3 cost. In practice, we tell the program to warn us if it finds a term for which this ordering would produce a contraction cost higher than n^6, which is our target cost for CT theory. Such terms occur infrequently and are inspected manually in order to find a more efficient contraction ordering. After all terms are derived and all contraction orderings chosen, the necessary cache-optimized FORTRAN code is written automatically by the program.

2.3 Formal properties

In this section we will explore some of the desirable formal properties of CT theory. For a survey of CT theory's numerical results and comparisons to other methods, see Sec. 2.5.

2.3.1 Size consistency

CT theory is rigorously size consistent, which we shall now demonstrate. Consider two non-interacting systems X and Y, for which the electronic Hamiltonian can be written as $H = H_X + H_Y$. Treating the systems separately, CT theory will produce two amplitude solutions A_X and A_Y as well as two effective Hamiltonians $\bar{H}_X = e^{-A_X} H_X e^{A_X}$ and $\bar{H}_Y = e^{-A_Y} H_Y e^{A_Y}$. These solutions satisfy the amplitude equations for the systems individually:

$$\langle \Psi_{X_0} | [\bar{H}_X, \hat{o}_i]_{1,2} | \Psi_{X_0} \rangle = 0, \tag{2.23}$$

$$\langle \Psi_{Y_0} | [\bar{H}_Y, \hat{o}_i]_{1,2} | \Psi_{Y_0} \rangle = 0. \tag{2.24}$$

The total energy for the two systems is simply the sum of the expectation values of their effective Hamiltonians,

$$E = \langle \Psi_{X_0} | \bar{H}_X | \Psi_{X_0} \rangle + \langle \Psi_{Y_0} | \bar{H}_Y | \Psi_{Y_0} \rangle = E_X + E_Y. \tag{2.25}$$

If we instead treat the systems together, choose $A = A_X + A_Y$, and make use of the fact that operators from different systems commute, we find that the effective Hamiltonian is the sum of the separate systems' effective Hamiltonians,

$$\bar{H} = e^{-(A_X + A_Y)} H e^{(A_X + A_Y)} = e^{-A_X} H_X e^{A_X} + e^{-A_Y} H_Y e^{A_Y} = \bar{H}_X + \bar{H}_Y. \tag{2.26}$$

This separation allows the amplitude equation,

$$\langle \Psi_{X_0} | \langle \Psi_{Y_0} | [\bar{H}_X + \bar{H}_Y, \hat{o}_i]_{1,2} | \Psi_{X_0} \rangle | \Psi_{Y_0} \rangle = 0, \tag{2.27}$$

to be satisfied for the two possible types of excitation operators $\hat{o}_i$. If $\hat{o}_i$ operates entirely in system X or entirely in system Y, then Eq. (2.27) is satisfied due to Eqs. (2.23-2.24) and the fact that $\hat{o}_i$ commutes with the effective Hamiltonian of the other system. If $\hat{o}_i$ operates on both systems X and Y simultaneously, it either changes the particle number in each system or simultaneously excites into both systems' external spaces. In this case Eq. (2.27) is satisfied because the effective Hamiltonian $\bar{H}_X + \bar{H}_Y$ cannot transfer particles between the systems and cannot simultaneously de-excite from both systems' external spaces. Thus we see that the solution A to the CT equations for the combined system is simply the sum of the solutions A_X and A_Y of the separate systems. Finally, by taking the expectation value of the combined system's effective Hamiltonian, we see that the energy is the same as for the separate systems and that CT theory is rigorously size consistent.

$$\begin{aligned} \langle \Psi_0 | \bar{H} | \Psi_0 \rangle &= \langle \Psi_{Y_0} | \langle \Psi_{X_0} | (\bar{H}_X + \bar{H}_Y) | \Psi_{X_0} \rangle | \Psi_{Y_0} \rangle \\ &= \langle \Psi_{X_0} | \bar{H}_X | \Psi_{X_0} \rangle + \langle \Psi_{Y_0} | \bar{H}_Y | \Psi_{Y_0} \rangle \\ &= E_X + E_Y \end{aligned} \tag{2.28}$$

2.3.2 Perturbative analysis

In this section we analyze the CT energy perturbatively first for the case in which the reference function is a single Slater determinant, and later for the general case of a multi-configurational reference function. This analysis is instructive as it provides insight into the similarities and differences between CT

theory and Coupled Cluster (CC) theory. This section follows the analysis of Bartlett et al in Refs [13, 84, 14]. For the purposes of this section, we separate the CT amplitude operator into its 1- and 2-body components via $A = A_1 + A_2$.

First consider a Hartree-Fock reference function and transform to the Fermi-vacuum (all occupied orbitals are in the vacuum). Then all Fermi-vacuum particle density matrices are zero and the linear commutator approximation corresponds to simply neglecting all 3-body operators. This type of operator truncation is used in the canonical diagonalization theory of White [86].

Now write the Hamiltonian as $H = E_{HF} + F + W$, where F is the 1-body Fock operator and W is the 2-body fluctuation potential. From Brillouin's theorem, we recognize that A_2 is first order in W, while A_1 is second order in W. (To make contact with the analysis of unitary coupled cluster theory in Refs. [13, 84], write A_1 as $(T_1 - T_1^\dagger)$ and $A_2 = (T_2 - T_2^\dagger)$.) Consider now the expectation value of the energy $E = \langle e^{-A} H e^{A} \rangle$ *without* using any commutator approximations. Expanding in powers of the fluctuation operator we have

$$E = E^0 + E^1 + E^2 + E^3 + E^4 + \ldots, \tag{2.29}$$

where the different order energies are defined as

$$E^0 = \langle E_{HF} \rangle + \langle F \rangle \tag{2.30a}$$

$$E^1 = \langle W \rangle \tag{2.30b}$$

$$E^2 = \langle [W, A_2] \rangle + \langle [F, A_1] \rangle + \langle [[F, A_2], A_2] \rangle \tag{2.30c}$$

$$E^3 = \frac{1}{2}\langle [[W, A_2], A_2] \rangle + \langle [W, A_1] \rangle + \frac{1}{2}\langle [[F, A_2], A_1] \rangle + \frac{1}{2}\langle [[F, A_1], A_2] \rangle \tag{2.30d}$$

$$E^4 = \frac{1}{6}\langle [[[W, A_2], A_2], A_2] \rangle + \frac{1}{2}\langle [[W, A_1], A_2] \rangle + \frac{1}{2}\langle [[W, A_2], A_1] \rangle + \frac{1}{2}\langle [[F, A_1], A_1] \rangle + \frac{1}{24}\langle [[[[F, A_2], A_2], A_2], A_2] \rangle \tag{2.30e}$$

Now consider the effect of the commutator approximation on the different orders of energy contribution. Firstly, no approximation is involved in computing E^0 or E^1, since these do not involve any commutators. For E^2, the approximation corresponds to

$$E^2 \approx \langle [W, A_2]_{1,2} \rangle + \langle [F, A_1] \rangle + \langle [[F, A_2], A_2]_{1,2} \rangle \tag{2.31}$$

Note that no approximation is made for commutators like $[F, A_2]$ which generate only 2- and lower-body operators. We see that $\langle [F, A_1] \rangle$ vanishes due to Brillouin's theorem, while both $[W, A_2]$ and $[[F, A_2], A_2]$ generate 3-body operators that are neglected but have no expectation value with the Fermi vacuum and thus would not contribute to the energy. Thus no error is made in Eq. (2.31) for E^2.

In the expression for E^3, we apply the commutator approximation twice for the double commutator $[[W, A_2]_{1,2}, A_2]_{1,2}$. Once again, only fully contracted terms contribute to the energy. The only way fully contracted terms arise is from double contractions in $[W, A_2]$ to produce a two-particle operator, which then fully contracts with A_2 in the outer commutator and contributes to the energy. Since double contractions are involved in each step, the commutator approximation does not affect this term. There is no contribution from the 3-body operators generated by either commutator, and so CT theory evaluates E^3 correctly also.

In the expression for E^4 we find our first error due to the commutator approximation. Here the 3-body operator arising from the inner commutator $[W, A_2]$, which is neglected in the commutator approximation, can contract successively with two other A_2 terms in $[[W, A_2], A_2], A_2]$ to yield a fully contracted term and thus a contribution to the energy. Although CT theory misses this con-

tribution, it *does* contain the contribution that arises from contracting the two-particle operators generated in the inner commutator $[W, A_2]$. By a similar analysis, we find that the commutator approximation also provides an incomplete evaluation of $[[[[F, A_2], A_2], A_2], A_2]$, arising from intermediate 3-body operators.

Can this error in E^4 be avoided? It turns out that we can by using the quadratic commutator approximation $[[W, A_2], A_2] \rightarrow [[W, A_2], A_2]_{1,2}$, in which 3- and 4-body operators are approximated by 1- and 2-body operators after the *second* commutator. This approximation, while more complicated, does not change the computational scaling of the theory. Here the two offending terms are approximated as $[[W, A_2], A_2]_{1,2}, A_2]_{1,2}$ and $[[[[F, A_2], A_2], A_2]_{1,2}, A_2]_{1,2}$, where we neglect only the 3- and 4-body operators that result from the outer two commutators. As these operators do not contribute to any full contractions, they have no energy contribution. Thus by using the quadratic commutator approximation, CT theory's energy can be made correct through fourth order in the perturbation (although errors will appear at fifth order). This version of CT theory, termed QCTSD and discussed in detail in Sec. 2.2.5, is preferable when a single-determinant reference is employed.

In the usual coupled cluster hierarchy $\sum_{i=0}^{2} E_i$ is the MP2 energy functional, while $\sum_{i=0}^{3} E_i$ is the linearized coupled cluster single-doubles (L-CCSD) energy functional. $\sum_{i=0}^{4} E_i$ is the unitary CCSD energy functional. The CTSD (by which we mean LCTSD) energy is correct up to third order in perturbation theory (like linearized CCSD theory), while the QCTSD energy is correct up to fourth order (like CCSD). Unlike linearized CCSD theory, however, fourth order terms are not completely neglected in CTSD but are partly included as discussed above. From this, we might expect the single-reference CTSD theory to perform inter-

mediate between linearized CCSD and the full CCSD theory. But in fact there are an *infinite* number of additional diagrams that are included in CTSD theory as compared to the usual CC and UCC(n) theories, because the energy functional does not terminate at finite order but contains further partial contributions from $E^5, E^6, \ldots, E^\infty$. For example, all terms involving pure orbital rotations (i.e. A_1) are included to *all* orders in the energy functional. Terms involving A_2, where all A_2 operators are at least doubly contracted with one other operator, are also included to all orders. One might speculate that these additional diagrams would yield an improved theory, but in the general case, and certainly when we extend the discussion to cases where a multi-determinantal reference wavefunction is used, the significance of the additional contributions present in CTSD (and QCTSD) can only be assessed numerically.

Finally we briefly analyze the case where the reference function is no longer a single Slater determinant. In this case errors arise not only from neglected operators, but also from neglected cumulants, the most significant being the 3-particle cumulant. The nature of the error arising from neglecting the 3-particle cumulant can be somewhat illuminated by assuming that the reference function itself admits a perturbation expansion in the active space. The corresponding fluctuation operator in the active space we will denote by W_{act}. Then

$$|\Psi_0\rangle = |\Phi^0\rangle + |\Phi^1\rangle + |\Phi^2\rangle + \ldots, \tag{2.32}$$

in which the zeroth order wavefunction $|\Phi^0\rangle$ is the Hartree-Fock solution. As the zeroth order wavefunction contains no correlation between electrons, it makes no contribution to the 3-body cumulant. The structure of $|\Phi^1\rangle$ in Møller-Plesset theory is a sum of doubly excited determinants, which also gives no contribution to the 3-body cumulant. In fact, it is not until we consider $|\Phi^2\rangle$, containing determinants with 1-, 2-, 3-, and some 4-body excitations, that we find a

contribution to the 3-body cumulant. We therefore see that if the Møller-Plesset expansion is a valid representation of our multi-configurational reference wavefunction, the 3-body cumulant is of order W_{act}^2.

Let us now examine the contribution of the 3-body cumulant to the energy. The 3-body cumulant contribution arises in the following component of the first commutator of the BCH expansion, $\langle [W_{ext}, A_2] \rangle$, where W_{ext} reflects the fluctuation potential between the active and external spaces. The operator A_2 is of order W_{ext}, so the product of amplitudes and integrals appearing in $\langle [W_{ext}, A_2] \rangle$ is of order W_{ext}^2, and its trace with the 3-body cumulant yields an energy contribution of order $W_{act}^2 W_{ext}^2$. We see therefore that even in the multi-configurational case (at least in the case where the reference function itself admits a meaningful perturbative expansion), the error in the CTSD energy appears at fourth order in the fluctuation potentials.

2.4 Optimizing the transformation

The excitation operator coefficients C_i that define the CT effective Hamiltonian are optimized by solving the CT amplitude equations, Eq. (2.7). This set of nonlinear equations is analogous to the amplitude equations encountered in single-reference coupled cluster theory. While there are many methods available for solving sets of nonlinear equations, we employ the iterative Newton-Raphson (NR) method in CT theory for two reasons: first, the structure of the CT equations allows for a particularly efficient solution of the linear equation that defines the step in each NR iteration, and second, because the NR Jacobian matrix provides physical insight into why the CT equations are hard to solve and how

they may be simplified. As mentioned in Sec. 2.2.1, the operators $\hat{o}_i$ that define our amplitude operator $A = \sum_i C_i \hat{o}_i$ are not simply taken to be the complete set of singles and doubles antisymmetric excitation operators given in Eq. (2.3). For reasons explained below, the amplitude equations are unsolvable for this set of operators and so we employ more structured definitions of $\hat{o}_i$ instead.

2.4.1 The Newton Raphson algorithm

When solving for the root of a single nonlinear equation $f(x)$, the NR approach is to make an initial guess $x^{(0)}$ and then extrapolate to the solution by following the function's slope $f'(x)$. This procedure leads to an iterative method in which the current solution is updated by the equation

$$x^{(i+1)} = x^{(i)} - \frac{f(x^{(i)})}{f'(x^{(i)})}. \tag{2.33}$$

This equation is evaluated repeatedly until the value of the function $f(x)$ is sufficiently small. The procedure for solving a system of nonlinear equations is analogous, except that now extrapolating to the solution using the functions' first derivatives requires solving a linear equation. The matrix in this equation, known as the Jacobian, consists of the derivative of each function with respect to each variable. In CT theory, our nonlinear equations are

$$R_i = \langle \Psi_0 | [\bar{H}, \hat{o}_i]_{1,2} | \Psi_0 \rangle = 0, \tag{2.34}$$

and so the Jacobian matrix is

$$J_{ij} = \frac{\partial R_i}{\partial C_j} = \langle \Psi_0 | [[\bar{H}, \hat{o}_j]_{1,2}, \hat{o}_i]_{1,2} | \Psi_0 \rangle + \mathrm{O}(A). \tag{2.35}$$

Given an initial guess $C_i^{(0)}$ for the excitation operator coefficients, one first evaluates the effective Hamiltonian $\bar{H}$ (using a BCH expansion approximated by operator decomposition) and determines the values R_i of the amplitude equations,

which we often refer to as the residuals. An improved guess for the coefficients, $C_i^{(1)}$, is then found by solving the linear equation

$$\sum_j J_{ij} C_j^{(1)} = -R_i, \tag{2.36}$$

which extrapolates along the functions' derivatives in an analogous manner to Eq. (2.33). Solving this linear equation is not trivial, however. Building the Jacobian is infeasible for all but the smallest systems, because it has a number of elements that scales as n^8 when the number of core, active, and virtual orbitals are assumed to be proportional to n. Instead we solve Eq. (2.36) iteratively using a Krylov subspace method, which requires only the ability to compute the Jacobian matrix's action on a trial vector. This action can be evaluated efficiently if we neglect the last term of Eq. (2.35), whose magnitude is linear in A and therefore should be small. Note that approximating the Jacobian in this way introduces no errors into the CT energy so long as the NR iteration leads to the solution of the amplitude equations. The NR equation can now be rewritten as

$$\langle \Psi_0 | [[\bar{H}, \sum_j C_j^{(1)} \hat{o}_j]_{1,2}, \hat{o}_i]_{1,2} | \Psi_0 \rangle = -R_i. \tag{2.37}$$

This form reveals how the Jacobian's action on the vector $C_j^{(1)}$ can be computed efficiently. The inner commutator of Eq. (2.37) has the same form as the approximate commutator $[H, A]_{1,2}$ used to evaluate the BCH expansion and can therefore be evaluated in n^6 time. The result of this inner commutator is a 2-body operator with the same symmetries as the electronic Hamiltonian, and so the outer commutator has precisely the same form as the amplitude equations, Eq. (2.7), which can also be evaluated in n^6 time. Thus by evaluating and storing the intermediate operator resulting from the inner commutator of Eq. (2.37), the action of the Jacobian on a trial vector can be evaluated efficiently and a Krylov subspace method can be used to solve Eq. (2.36) in n^6 time.

A difficulty arises if any of the Jacobian's eigenvalues vanish or are very small, because linear equations involving singular or near singular matrices cannot in practice be solved using a Krylov subspace. Even if we could invert such a Jacobian, the result would not be physically meaningful, as one or more excitation operators would acquire a near-infinite coefficient. Recall that the working assumption for CT theory is that the reference function is close to the true wavefunction, which means that the excitation operators' coefficients should be small. As long as this assumption is satisfied (i.e. for a sufficiently large active space), any small eigenvalues in the Jacobian matrix are unphysical in nature and must be artifacts of CT theory's approximations. We call such eigenvalues and their corresponding excitation operators intruder states, because they are closely related to the intruder states encountered in second order perturbation theory (PT2) [33]. To see how, replace the effective Hamiltonian on the left hand side of Eq. (2.37) with a zeroth order Hamiltonian satisfying $H_0|\Psi_0\rangle = E_0|\Psi_0\rangle$. By ignoring the commutator approximations, setting $C_i^{(0)} = 0$, and recalling that $\hat{o}_i^\dagger = -\hat{o}_i$, we may simplify the resulting equation to obtain

$$\sum_j \langle\Psi_0|\hat{o}_i^\dagger(H_0 - E_0)\hat{o}_j|\Psi_0\rangle C_j^{(1)} = \langle\Psi_0|\hat{o}_i^\dagger H|\Psi_0\rangle. \tag{2.38}$$

This is the defining equation for the coefficients $C_j^{(1)}$ of the perturber states $\hat{o}_j|\Psi_0\rangle$ that appear in the first order wavefunction of PT2 theory. Traditional intruder states in perturbation theory are those with unphysically large coefficients, and so we see that there is a strong analogy between intruder states in CT and PT2.

The key difference between intruder states in CT and PT2 theory is what causes them. For perturbation theory, Eq. (2.38) tells us that intruder states will occur for inaccurate zeroth order Hamiltonians for which a state in the basis

of perturber functions has an energy close to the reference state. In CT theory, Eq. (2.35) contains no zeroth order Hamiltonian and so the origin of intruder states must be different. The three possibilities are errors in the effective Hamiltonian $\bar{H}$, neglecting the $\mathrm{O}(A)$ terms present in Eq. (2.35), and the commutator approximations present in Eq. (2.35). The first two possibilities can be ruled out if we consider the initial guess $C_i^{(0)} = 0$ (this is a reasonable guess for the coefficients as they should be small). For this guess the $\mathrm{O}(A)$ terms vanish and the initial effective Hamiltonian $\bar{H}$ is exactly equal to the bare electronic Hamiltonian H, which should give correct energies for states in the first order interacting basis. As CT theory typically uses $C_i = 0$ as its initial guess, any intruder states encountered are therefore caused by the commutator approximations present in Eq. (2.35). Unfortunately, if one formulates CT theory using a multi-configurational reference function and the full set of excitation operators given in Eq. (2.3), these commutator approximations create intruder states for all but the simplest systems. In the next two sections we discuss alternative choices for the excitation operators that are more successful at avoiding intruder states.

2.4.2 Overlap matrix truncation

To address the problem of intruder states in CT theory, one must first recognize that for the standard choice of single and double excitation operators $\hat{p}_i \in \{a_q^p, a_{rs}^{pq} \ldots\}$, the first order interacting basis $\{\hat{p}_i|\Psi_0\rangle\}$ is not orthogonal if the reference is multi-configurational. Instead there is a dense, non-diagonal overlap matrix $S_{ij} = \langle\Psi_0|\hat{p}_i^\dagger \hat{p}_j|\Psi_0\rangle$. The eigenvalues of the Jacobian defined in Eq. (2.35), which govern the presence of intruder states, are therefore represented

by a generalized eigenvalue equation in which the overlap matrix appears:

$$\sum_j J_{ij} B_{jk} = \sum_l S_{il} B_{lk} \epsilon_k. \tag{2.39}$$

Here ϵ_k are the Jacobian eigenvalues and B is the matrix whose columns are the Jacobian eigenvectors. In order to improve the conditioning of the Newton-Raphson equation and remove linear degeneracies from the first order interacting basis, we can transform to the set of "orthonormal" excitation operators $\hat{o}_i = \sum_j (S^{-1/2})_{ij} \hat{p}_j$ that generate an orthonormal first order interacting basis $\{\hat{o}_i|\Psi_0\rangle\}$. For these operators the Jacobian eigenvalues ϵ_k are defined by the simple eigenvalue equation

$$\sum_j \tilde{J}_{ij} \tilde{B}_{jk} = \tilde{B}_{ik} \epsilon_k, \tag{2.40}$$

in which $\tilde{J} = S^{-1/2} J S^{-1/2}$ and $\tilde{B} = S^{1/2} B$ are the Jacobian and its eigenvectors in the orthonormal basis $\{\hat{o}_i|\Psi_0\rangle\}$. To prevent intruder states, the small eigenvalues present in S and J must cancel in the product $S^{-1/2} J S^{-1/2}$. Errors in the eigenvalue spectrums of either S or J can disrupt this cancellation and produce unphysically small values for ϵ_k, which as described in the previous section will correspond to intruder states. There are two possible sources for such errors. First, for semi-internal excitation operators (double excitations that involve one excitation into or out of the active space and one excitation within the active space) the overlap matrix S depends on the reference function's 3-body RDM. If we approximate S by neglecting the 3-body cumulant in order to avoid using the 3-body RDM, then we introduce errors in S that can lead to intruder states. Second, the commutator approximations present in the definition of J create errors in its eigenvalue spectrum. In practice we find that even if the exact 3-body RDM is used to avoid errors in S, the errors in J are sufficient to produce intruder states for most multi-configurational references. Therefore, in order to

circumvent the need for a delicate cancellation of small eigenvalues in the product $S^{-1/2}JS^{-1/2}$, we restrict our operators $\hat{o}_i$ by discarding the eigenvalues of S below some threshold τ. In practice we use two thresholds, τ_1 for the single and semi-internal excitations' overlap matrices and τ_2 for the double excitations' overlap matrices. While all multi-configurational dynamic correlation methods employing an internally contracted form must truncate the overlap matrices to some extent (e.g. $\tau = 10^{-6}$) in order to remove true linear degeneracies from the first order interacting basis, CT theory requires a much larger truncation threshold (e.g. $\tau = 10^{-2}$) in order to prevent intruder states. Because of the importance of accurately truncating the overlap matrices, the standard CTSD version of CT theory, reported previously as L-CTSD(MK) [89] and LCTSD [69], uses the exact 3-body RDM in order to compute the semi-internal overlap matrices. We note that CTSD does not, however, use the 3-body RDM anywhere else.

The use of overlap truncation in CT theory has been very successful, producing accuracies competitive with expensive methods such as MRCI+Q. However, the need to accurately truncate the overlap matrices creates three significant disadvantages. First, the choice of the truncation thresholds is arbitrary and in difficult systems can affect the CT energy, as seen in the results for NiO in Sec. 2.5.5. Second, constructing the overlap matrices exactly with the reference wavefunction's 3-body RDM is necessary to achieve these accuracies. While overlap truncation can use the cumulant-approximated overlap matrices, the accuracy suffers because higher truncation thresholds are necessary to prevent the errors in S from creating intruder states. Finally, diagonalizing the semi-internal overlap matrices to produce $S^{-1/2}$ has a cost that scales as n_{act}^9, where n_{act} is the number of active orbitals. This cost is trivial for small active spaces, but will become infeasible for the very large active spaces accessible with DMRG theory. There

is another way to construct the operator set $\{\hat{o}_i\}$ that avoids these issues at the cost of further limiting the ansatz's freedom, which we shall now discuss.

2.4.3 Strong contraction

Strongly contracted (SC) excitation operators were first introduced by Malrieu et al in the context of n-electron valence perturbation theory (NEVPT2) [6, 7]. They consist of a drastic simplification of the first order interaction basis in which each external orbital (for singles) or orbital pair (for doubles) has only one excitation operator that connects it to the active space. Such an operator is of course a linear combination of many of the basic excitation operators, but all terms in the combination involve the same external indices. An immediate consequence of this formulation is that the SC operators are mutually orthogonal. These operators thus avoid completely the difficulties of building, diagonalizing, and truncating overlap matrices and therefore require neither the n_{act}^9 cost diagonalization step nor the reference function's 3-body RDM. While SC operators are certainly simple to work with, one must construct them carefully in order to retain accuracy in such a restricted set of excitations. Each SC operator is therefore formed as the sum of all contracted operators of its type (e.g. double excitations from the active space into the virtual orbitals v_1 and v_2) weighted by their coefficients in the electronic Hamiltonian H. For the example of double excitations between the active and virtual orbitals, the operator corresponding to the pair of virtual orbitals (v_1, v_2) is

$$\hat{o}^{v_1 v_2} = \sum_{a_1 a_2} g_{a_1 a_2}^{v_1 v_2} (a_{a_1 a_2}^{v_1 v_2} - a_{v_1 v_2}^{a_1 a_2}), \tag{2.41}$$

where g is the usual 2-body integral tensor. In total there are eight types of SC operators, the precise definitions for which are given in Appendix A. The justification for weighting the operators based on their Hamiltonian coefficients is that it makes the states $\hat{o}_i|\Psi_0\rangle$ components of the system's first Krylov vector, $H|\Psi_0\rangle$. The operation of H on the reference function emphasizes the eigenstates of H in $|\Psi_0\rangle$ with the largest magnitude eigenvalues. As states with large positive eigenvalues will not be present in $|\Psi_0\rangle$, the emphasis will be put on those eigenstates with large negative eigenvalues, namely the ground state and low-lying excited states. Thus the choice of Hamiltonian coefficients for constructing a SC excitation operator is good in the sense that it is tailored towards describing the system's lowest eigenstates. Even with this intelligent choice for the form of the SC operators, they offer far less freedom than is available through the overlap truncation method, and the accuracy of CT suffers as a result. The motivation for using SC operators is of course not to improve accuracy (which for CTSD was already excellent) but to avoid intruder states.

In practice, the orthogonality of the SC operators is sufficient to prevent intruder states in many systems. To detect and remove any remaining intruder states, we construct an approximation of the CT Jacobian and inspect it for any unphysically small eigenvalues. This approximation consists of neglecting the last term in Eq. (2.35) and replacing the effective Hamiltonian with Dyall's zeroth order Hamiltonian [31],

$$H_0 = C + \sum_{c_1} \bar{t}^{c_1}_{c_1} a^{c_1}_{c_1} + \sum_{v_1} \bar{t}^{v_1}_{v_1} a^{v_1}_{v_1} + \sum_{a_1 a_2} \bar{t}^{a_1}_{a_2} a^{a_1}_{a_2} + \frac{1}{2} \sum_{\substack{a_1 a_2 \\ a_3 a_4}} g^{a_1 a_2}_{a_3 a_4} a^{a_1 a_2}_{a_3 a_4}, \tag{2.42}$$

in which C is a constant, $\bar{t}$ is a set of effective 1-body integrals, and g is the usual 2-body integral tensor. Note that C and $\bar{t}$ are defined such that $H_0|\Psi_0\rangle = E_0|\Psi_0\rangle$.

Our approximation for the CT Jacobian is now

$$J_{ij} = \langle \Psi_0 | [[H_0, \hat{o}_j]_{1,2}, \hat{o}_i]_{1,2} | \Psi_0 \rangle, \tag{2.43}$$

which is a diagonal matrix because the Dyall Hamiltonian cannot connect operators that excite into different external orbitals (recall that each SC operator corresponds to a different set of external orbitals). As in the previous section, we desire the Jacobian eigenvalues in an orthonormal basis of excitation operators, and although the SC operators are orthogonal by construction they are not normalized. We therefore evaluate their norms approximately via

$$||\hat{o}_i||^2 \simeq \langle \Psi_0 | \hat{o}_i^\dagger \hat{o}_i | \Psi_0 \rangle_{1,2}, \tag{2.44}$$

where we neglect the 3-body cumulant to avoid using the 3-body RDM. With these approximate norms, we may evaluate the approximate eigenvalues of the orthonormal Jacobian as

$$\epsilon_i = \frac{1}{||\hat{o}_i||^2} \langle \Psi_0 | [[H_0, \hat{o}_i]_{1,2}, \hat{o}_i]_{1,2} | \Psi_0 \rangle. \tag{2.45}$$

After obtaining these approximate Jacobian eigenvalues (which can be done in n^6 time), we inspect them and remove from our excitation operator basis any operators with unphysically small values for ϵ_i. These eigenvalues should be bounded from below by the Dyall Hamiltonian's smallest excitation energy between the reference state and states in the first order interacting basis. For a sufficiently large active space, these excitation energies will be quite sizable as the cost to occupy a virtual orbital or create a hole among the core orbitals will be high. In practice, we typically discard operators with ϵ_i below a threshold of $\tau_\epsilon = 0.1$ Hartrees. This truncation threshold is more physically intuitive than the overlap truncation thresholds presented in the previous section, and in large active spaces the CT energy is insensitive to its value (e.g. all excitation operators have ϵ_i much larger than 0.1 Hartrees). In summary, SC operators prevent

intruder states in CT theory without building or diagonalizing overlap matrices or using the 3-body RDM.

2.5 Applications and comparisons to other methods

In this section we will review the results of applying CT theory to a number of chemical systems. In particular, we will investigate the effects of using the quadratic commutator, the 3-body RDM, and strongly contracted excitation operators. As CT theory is meant to capture dynamic correlation in order to achieve chemical accuracy, it is helpful at this point to explain precisely what is meant by chemical accuracy. When breaking a chemical bond, moving along a chemical reaction coordinate, or studying the excitation spectrum of a molecule, one is always considering energy differences between one or more states. Chemical accuracy is commonly regarded to imply an error in these energy differences of less than 1 kcal/mol for bond breaking and reactions and 0.1 eV for excitations. As quantum chemistry is chiefly concerned with energy differences, it is more important for a method to have the correct relative energies between states than to have the correct absolute energy for any given state taken alone. As such, a useful measure for a method's accuracy is its relative error, which in bond stretching is defined as (the absolute value of) the worst possible error for the difference between the energies of any two geometries. The results below will make heavy use of relative error when comparing the accuracies of different methods.

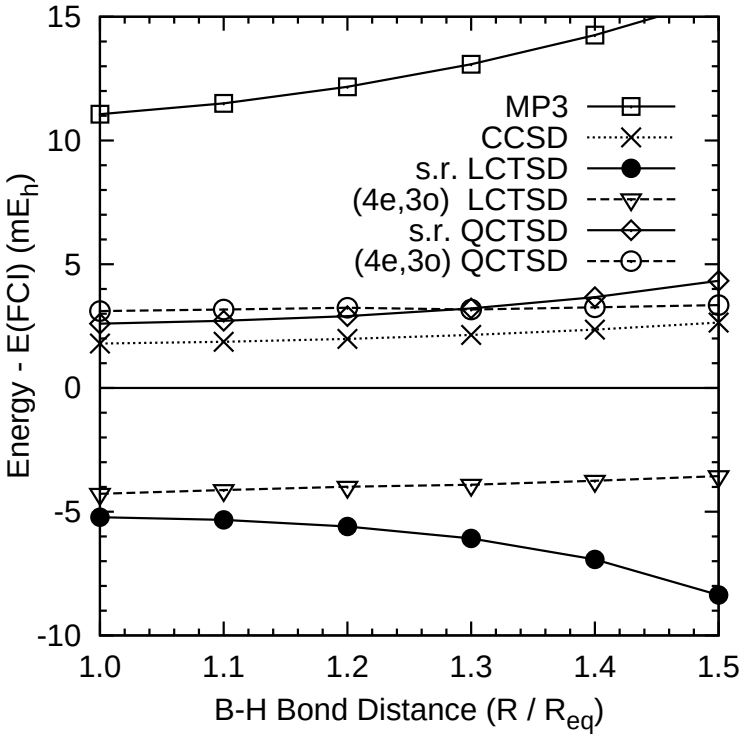

Figure 2.1: BH ground state energy errors relative to FCI.

2.5.1 Boron hydrogen

The first molecule we study is BH, which is treated in the Dunning double zeta with polarization (DZP) basis [29] using Cartesian d orbitals, with the boron d orbital exponent changed to 0.5 as in Ref. [88]. We have carried out two types of CT calculations, one using the Hartree-Fock determinant as the reference function and the other using a 4-electron 3-orbital (4e,3o) CASSCF reference. In both cases, all electrons were correlated. Notice that when a single determinant is used as the CT reference function, the large truncation thresholds discussed in Sec. 2.4.2 are unnecessary as the overlap matrices' eigenvalues are either 0 or ≥ 1. Also recall that the cumulant approximation is exact for a single determi-

Table 2.2: Results for the BH potential energy curve.

The FCI energy is reported in $\mathrm{E_h}$, with other methods reported as the difference from FCI in $\mathrm{mE_h}$.

$\mathrm{R/R_{eq}}^*$	FCI	HF	MP2	MP3	CCSD	LCTSD†	QCTSD†	LCTSD$^\diamond$	QCTSD$^\diamond$
1.0	-25.227 627	102.366	28.639	11.060	1.792	-5.225	2.595	-4.274	3.105
1.1	-25.223 980	103.563	29.430	11.502	1.868	-5.333	2.712	-4.124	3.170
1.2	-25.214 410	105.299	30.553	12.166	1.980	-5.595	2.906	-3.987	3.239
1.3	-25.202 124	107.575	32.029	13.077	2.140	-6.084	3.213	-3.912	3.173
1.4	-25.188 960	110.393	33.875	14.258	2.360	-6.930	3.672	-3.752	3.258
1.5	-25.175 976	113.763	36.107	15.727	2.644	-8.366	4.331	-3.562	3.352
Relative Error $(\mathrm{mE_h})$		11.397	7.468	4.667	0.852	3.141	1.736	0.712	0.247

* $\mathrm{R_{eq}} = 2.329$ Bohr † HF Reference $^\diamond$ CASSCF (4e,3o) Reference, $\tau_1 = \tau_2 = 0.05$

nant, so the L3CTSD and Q3CTSD methods become redundant.

One reason for developing the QCTSD method was to improve accuracy when the reference function is a single electronic determinant. Fig. 2.1 and Table 2.2 show that QCTSD is indeed an improvement over LCTSD when the Hartree-Fock reference is used. QCTSD is also an improvement when starting from the CASSCF solution, although this behavior appears to be unique to the BH molecule. In our other systems, QCTSD appears to be less accurate than LCTSD when a CASSCF solution is used as the reference function. We also note that QCTSD behaves similarly to CCSD when the Hartree-Fock reference is employed, as may be expected from the fact that their energies are formally accurate to the same order in the fluctuation potential. This similarity is most pronounced near the equilibrium geometry. As the bond is stretched, QCTSD shows a larger relative error. Overall, the BH molecule supports the analysis that QCTSD should be more accurate than LCTSD in single-reference systems.

2.5.2 Hydrofluoric acid

HF is treated in the Dunning DZP basis [29] using spherical d orbitals, with the hydrogen p and fluorine d orbital exponents changed to 0.75 and 1.6, respectively, as in Ref. [88]. We carried out CT calculations using both Hartree-Fock and (2e,2o) CASSCF solutions as reference functions. All orbitals were correlated.

QCTSD again makes an improvement on LCTSD when using the Hartree-Fock reference, as shown in Fig. 2.2 and Table 2.3. Its energy is also again very similar to CCSD, although it fails to converge for bond lengths greater than

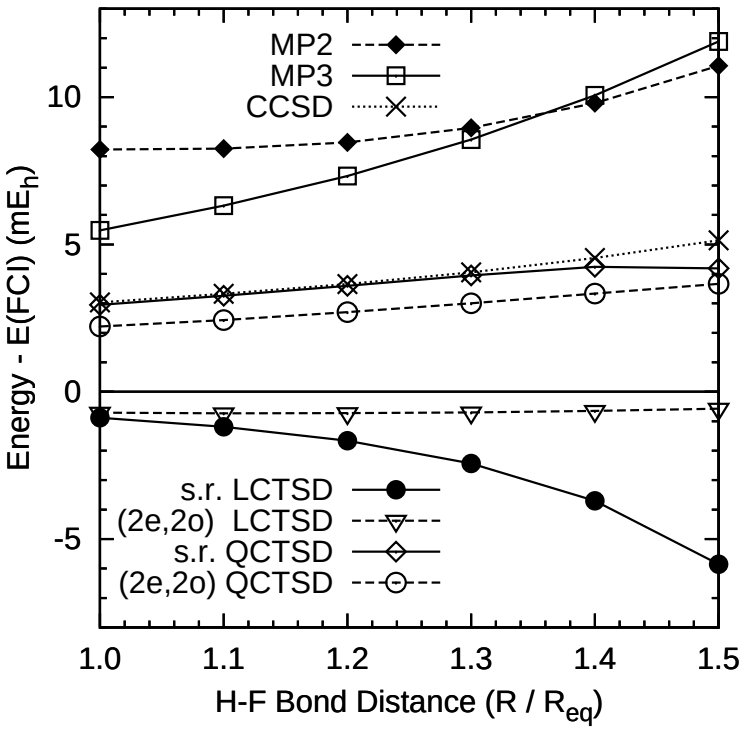

Figure 2.2: HF ground state energy errors relative to FCI.

150% of the equilibrium distance (both CT methods converge across the entire dissociation curve when using the CASSCF reference).

The main difference from BH can be seen when using a CASSCF reference function where, unlike in the Hartree-Fock case, QCTSD is less accurate than LCTSD. This unexpected behavior is also displayed in the H_2O and N_2 molecules. While LCTSD and QCTSD both enjoy improvements in absolute error when moving from a Hartree-Fock to CASSCF reference, the improvements in LCTSD produce better parallelity with the FCI bonding curve.

Table 2.3: Results for the HF potential energy curve.

The FCI energy is reported in $\mathrm{E_h}$, with other methods reported as the difference from FCI in $\mathrm{mE_h}$.

$\mathrm{R/R_{eq}}^*$	FCI	HF	MP2	MP3	CCSD	LCTSD†	QCTSD†	LCTSD$^\diamond$	QCTSD$^\diamond$
1.0	-100.264 981	217.894	8.226	5.477	3.036	-0.881	2.951	-0.712	2.214
1.1	-100.257 467	221.981	8.249	6.314	3.322	-1.187	3.255	-0.729	2.436
1.2	-100.240 154	226.213	8.465	7.326	3.656	-1.666	3.596	-0.727	2.701
1.3	-100.218 661	230.720	8.956	8.558	4.056	-2.434	3.950	-0.701	3.005
1.4	-100.196 105	235.613	9.800	10.062	4.542	-3.700	4.238	-0.648	3.335
1.5	-100.174 219	240.989	11.062	11.889	5.137	-5.857	4.185	-0.575	3.663
Relative Error $(\mathrm{mE_h})$		23.095	2.836	6.412	2.101	4.976	1.287	0.154	1.449

* $\mathrm{R_{eq}} = 1.733$ Bohr † HF Reference $^\diamond$ CASSCF (2e,2o) Reference, $\tau_1 = \tau_2 = 0.001$

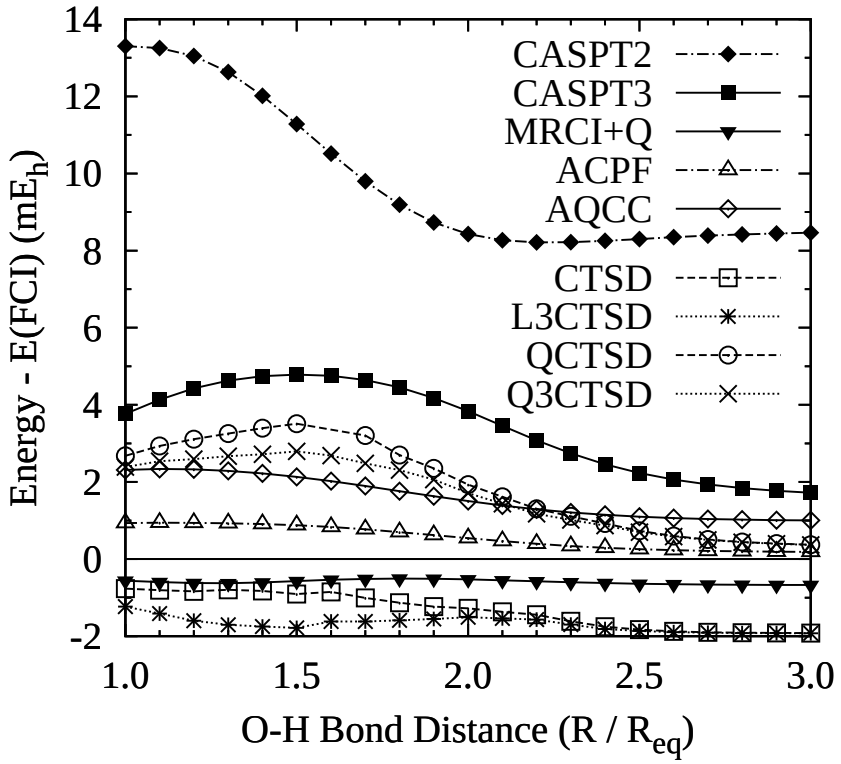

Figure 2.3: Quadratic commutator and 3-body RDM results in H_2O.

2.5.3 Water

The H_2O molecule was used as a benchmark system for evaluating the performance of both alternative commutator approximations and strongly contracted excitation operators. In both cases, the symmetric stretch of the two O-H bonds was studied, with the bond angle fixed at $109.57°$ and the equilibrium bond distance R_{eq} taken to be 1.876 Bohr (0.9929 Angstroms). The basis set used for both studies was the Dunning cc-pVDZ basis set [30] with spherical d orbitals. Note, however, that the oxygen 1s orbital was frozen for the quadratic commutator study but not for the strong contraction study. We will review the results of both studies in this section, beginning with the results for alternative commutators.

Table 2.4: Quadratic commutator and 3-body RDM results in H_2O.

The FCI energy is reported in E_h, with other methods reported as the difference from FCI in mE_h.

R/R_{eq}*	FCI	CASSCF	CASPT2	CASPT3	MRCI+Q	ACPF	AQCC	LCTSD†	L3CTSD†	QCTSD†	Q3CTSD†
1.0	-76.238 851	162.986	13.302	3.767	-0.561	0.933	2.313	-0.768	-1.232	2.680	2.387
1.5	-76.061 811	149.701	11.282	4.783	-0.572	0.876	2.133	-0.907	-1.783	3.509	2.795
2.0	-75.945 588	131.954	8.431	3.835	-0.525	0.542	1.504	-1.279	-1.510	1.933	1.712
2.5	-75.915 266	124.574	8.301	2.233	-0.639	0.257	1.100	-1.829	-1.865	0.733	0.700
3.0	-75.910 028	123.011	8.464	1.721	-0.673	0.184	1.001	-1.919	-1.923	0.371	0.368
Relative Error (mE_h)		39.975	5.091	3.061	0.169	0.757	1.335	1.152	0.691	3.138	2.427

* $R_{eq} = 1.876$ Bohr, $\angle HOH = 109.57°$

† $\tau_1 = \tau_2 = 0.01$

In studying the quadratic commutator and the inclusion of the 3-body RDM, all dynamic correlation methods used as a reference the (6e,5o) CASSCF wavefunction, in which the oxygen 1s orbital was frozen after the CASSCF orbital optimization. The results of the different methods' performance can be found in Figs. 2.3 and Table 2.4.

The most notable result is that Q3CTSD and QCTSD, which use the quadratic commutator approximation with and without the exact 3-body RDM, have larger relative errors and are therefore less accurate than L3CTSD and CTSD, which use the linear commutator approximation. This result shows that the (single-reference) perturbative arguments advocating the use of the quadratic commutator break down for the multi-reference case of the symmetric stretching of H_2O. As seen in Fig. 2.3, the deviation from FCI in QCTSD, and to a lesser extent Q3CTSD, increases as the bonds are stretched up to 150% of their equilibrium distances, beyond which the error diminishes. This behavior, which is similar to that of CASPT3, is not present in CTSD or L3CTSD, which produced potential energy curves significantly more parallel to that of FCI.

A less surprising result is that the use of the exact 3-body RDM reduced CT theory's relative error for both the linear and quadratic commutator formulations of the theory. L3CTSD was more accurate than CTSD, and Q3CTSD was more accurate than QCTSD. This improved accuracy is to be expected, as the use of the exact 3-body RDM removes the portion of the decomposition error due to the neglect of the 3-body cumulant.

Compared to CASPT2, all of the CT variants showed superior accuracy. This comparison is especially significant for CTSD, as it has a lower cost scaling than multi-reference perturbation theory. Furthermore, CTSD and L3CTSD were

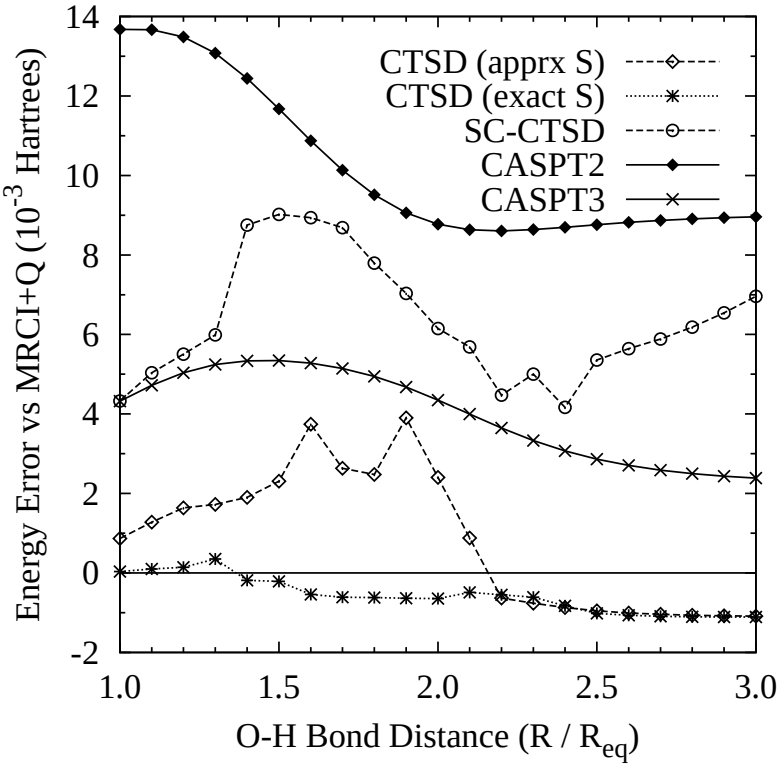

Figure 2.4: Relative energy results for strongly contracted CTSD in $\mathrm{H_2O}$.

comparable in accuracy to the expensive corrected MRCI methods. In fact, both CTSD and L3CTSD were superior in accuracy to AQCC.

The study of strongly contracted excitation operators used the same geometry and basis set as noted above, but no orbitals were held frozen during the dynamic correlation calculations. Comparisons between strongly contracted CT theory (SC-CTSD) and other methods can be found in Fig. 2.4 and 2.5 and Table 2.5.

The CTSD method shows a relative error of 1.5 $\mathrm{mE_h}$ when the semi-internal overlap matrix S_{int} is evaluated exactly using the 3-body RDM. (Note that other than for the truncation of S_{int}, all other terms in CTSD use only the 1- and 2-body

Table 2.5: Strongly contracted CTSD results in H_2O.

The MRCI+Q energy is reported in E_h, with other methods reported as the difference from MRCI+Q in mE_h.

R/R_{eq} [a]	MRCI+Q	CASSCF	CASPT2	CASPT3	CTSD [b]	CTSD [c]	SC-CTSD [d]
1.0	-76.241 466	165.601	13.674	4.321	0.866	0.034	4.326
1.2	-76.182 723	161.881	13.487	5.038	1.638	0.145	5.505
1.4	-76.101 419	155.832	12.442	5.336	1.900	-0.187	8.753
1.6	-76.031 316	148.035	10.874	5.278	3.741	-0.541	8.937
1.8	-75.980 391	140.359	9.514	4.945	2.477	-0.620	7.796
2.0	-75.947 828	134.194	8.774	4.350	2.402	-0.643	6.150
2.2	-75.929 506	130.039	8.606	3.646	-0.633	-0.549	4.475
2.4	-75.920 239	127.638	8.697	3.071	-0.869	-0.824	4.165
2.6	-75.915 748	126.371	8.820	2.705	-1.007	-1.064	5.641
2.8	-75.913 526	125.713	8.908	2.499	-1.065	-1.102	6.185
3.0	-75.912 381	125.364	8.959	2.387	-1.088	-1.109	6.960
Relative Error (mE_h)		40.237	5.068	2.956	4.988	1.458	4.855

[a] $R_{eq} = 0.9929$ Angstroms

[b] $\tau_1 = 0.2$, $\tau_2 = 0.1$, approximate S_{int}

[c] $\tau_1 = 0.1$, $\tau_2 = 0.01$, exact S_{int}

[d] $\tau_\epsilon = 0.1\ E_h$

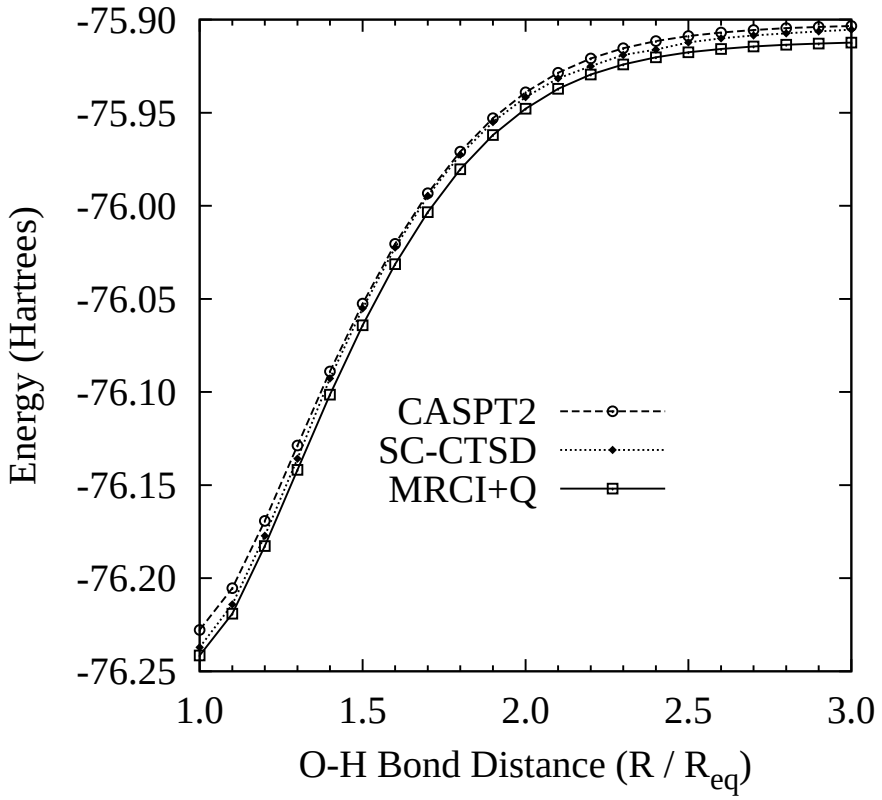

Figure 2.5: Absolute energy results for strongly contracted CTSD in H_2O.

RDMs.) However, when S_{int} is approximated with the cumulant decomposition, CTSD has a significantly larger relative error of 5.0 mE_h. The accuracy suffers because the truncation thresholds τ_1 and τ_2 must be increased to prevent the cumulant decomposition approximations in S_{int} from creating intruder states. This more aggressive truncation results in a more limited first order interacting basis which reduces accuracy. The SC-CTSD method, which does not require the 3-body RDM, shows relative error of 4.9 mE_h. All of the CT methods are more accurate than CASPT2, whose relative error is 5.1 mE_h. The discontinuity problem discussed above is more severe for SC-CTSD than for CTSD, although this is a somewhat unfair comparison as the latter uses the exact 3-body RDM to evaluate its overlap matrix. Indeed, the SC-CTSD discontinuities are less severe

than those that occur in CTSD when the cumulant-approximated overlap matrix is used. Fortunately, as shown in Fig. 2.5, the discontinuities present in SC-CTSD are small enough that they do not affect the shape of the potential energy curve.

2.5.4 Nitrogen

Correctly breaking the bond of the nitrogen dimer is a particularly challenging problem for quantum chemical theories due to the strong role of static correlation. Near equilibrium, the N_2 wavefunction is dominated by a single configuration in which the bonding π orbitals are doubly occupied. However, as the bond is broken, excited configurations play an increasingly important role, leading to the qualitative failure of single-reference methods. The use of a multireference method such as CASSCF, which includes all possible configurations of the electrons in the 2p orbitals, restores qualitative agreement with the exact theoretical (FCI) result, but still fails to achieve quantitative agreement due to the omission of dynamic correlation.

In this section we discuss the results from applying a number of different dynamic correlation methods to the stretching of the N_2 bond. In all calculations we employ Dunning's [30] cc-pVDZ basis set with spherical d orbitals. The reference function is a (6e,6o) CASSCF wavefunction, with the active space consisting of the 2p orbitals. After the CASSCF orbital optimization, the 1s orbitals were held frozen.

Significantly, not all multireference dynamic correlation theories are able to achieve chemical accuracy in the N_2 binding curve, despite its small size and the affordability of constructing and diagonalizing its 3-body RDM. Most notable is

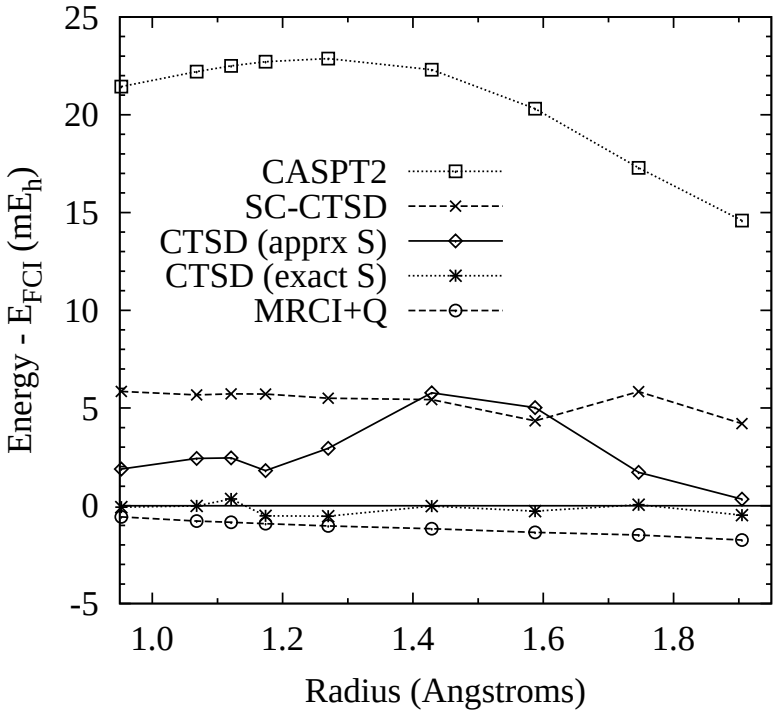

Figure 2.6: N_2 ground state energy errors relative to FCI.

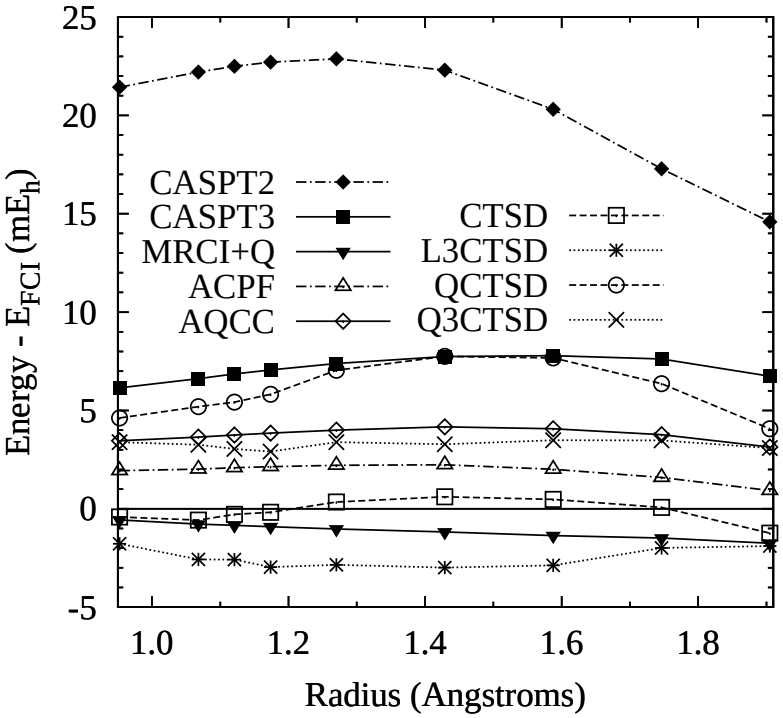

Figure 2.7: Quadratic commutator and 3-body RDM results in N_2.

Table 2.6: Results for the N_2 potential energy curve.

The FCI energy is reported in E_h, with other methods reported as the difference from FCI in mE_h.

R[a]	FCI[b]	CASSCF	MRCI+Q	CASPT2	CTSD[c]	CTSD[d]	SC-CTSD[e]
0.9525	-109.167 573	182.072	-0.564	21.437	1.883	-0.073	5.850
1.0679	-109.270 384	186.030	-0.782	22.202	2.423	-0.008	5.668
1.1208	-109.278 339	187.644	-0.845	22.497	2.448	0.352	5.723
1.1737	-109.271 915	189.155	-0.912	22.713	1.801	-0.519	5.713
1.2700	-109.238 397	191.715	-1.029	22.873	2.940	-0.535	5.497
1.4288	-109.160 305	195.376	-1.174	22.300	5.776	-0.015	5.428
1.5875	-109.086 211	197.685	-1.363	20.309	5.019	-0.272	4.344
1.7463	-109.030 31	197.660	-1.491	17.287	1.709	0.053	5.837
1.9050	-108.994 81	194.696	-1.750	14.587	0.339	-0.482	4.202
Relative Error (mE_h)		15.612	1.186	8.286	5.437	0.887	1.648

[a] Radius in Angstroms

[b] Taken from Ref. [56]

[c] $\tau_1 = 0.3$, $\tau_2 = 0.1$, approximate overlap

[d] $\tau_1 = 0.1$, $\tau_2 = 0.01$, exact overlap

[e] $\tau_\epsilon = 0.1$ E_h

Table 2.7: Quadratic commutator and 3-body RDM results in N_2.

The FCI energy is reported in E_h, with other methods reported as the difference from FCI in mE_h.

R[a]	FCI[b]	CASSCF	CASPT2	CASPT3	MRCI+Q	ACPF	AQCC	CTSD[c]	L3CTSD[c]	QCTSD[c]	Q3CTSD[c]
0.9525	-109.167 573	182.072	21.437	6.146	-0.564	1.932	3.458	-0.421	-1.781	4.620	3.387
1.0679	-109.270 384	186.030	22.202	6.618	-0.782	2.016	3.642	-0.576	-2.575	5.191	3.257
1.1208	-109.278 339	187.644	22.497	6.852	-0.845	2.083	3.752	-0.281	-2.582	5.426	3.043
1.1737	-109.271 915	189.155	22.713	7.062	-0.912	2.136	3.847	-0.178	-2.969	5.818	2.913
1.2700	-109.238 397	191.715	22.873	7.387	-1.029	2.207	3.996	0.342	-2.852	7.044	3.387
1.4288	-109.160 305	195.376	22.300	7.743	-1.174	2.240	4.165	0.613	-2.993	7.745	3.280
1.5875	-109.086 211	197.685	20.309	7.782	-1.363	2.004	4.069	0.474	-2.873	7.672	3.485
1.7463	-109.030 31	197.660	17.287	7.616	-1.491	1.591	3.768	0.070	-1.995	6.359	3.473
1.9050	-108.994 81	194.696	14.587	6.740	-1.750	0.928	3.142	-1.241	-1.891	4.066	3.086
Relative Error (mE_h)		15.612	8.286	1.635	1.186	1.312	1.023	1.854	1.213	3.679	0.571

[a] Radii in Angstroms. [b] FCI results of Larsen et al [56] [c] $\tau_1 = 0.02$ and $\tau_2 = 0.01$

the failure of perturbation theory, which for example produces a 5.2 kcal/mol relative error when implemented as CASPT2, as shown in Table 2.6 and Fig. 2.6. In contrast, the two standard CT methods, CTSD with exact overlap and SC-CTSD, have relative errors of 0.6 and 1.0 kcal/mol, respectively. Both of these methods have lower cost scalings than CASPT2 and can therefore be applied to much larger systems. Of particular note is that the SC-CTSD method achieves chemical accuracy in N_2 using only the 1- and 2-body RDMs of the reference wavefunction, an unprecedented feat. Other methods with comparable accuracy, such as MRCI+Q and CASPT3, require access to either the CASSCF wavefunction's CI coefficients or the 3-body RDM.

The breaking of the N_2 bond has been extensively studied with CT theory and provides an excellent pedagogical example of the theory's strengths and weaknesses. The relatively poor performance of CTSD when using the cumulant-approximated overlap matrices is a reminder of CT theory's intruder states (Fig. 2.6 and Table 2.6). The accuracy is damaged in this case by the necessity of using larger truncation thresholds to prevent errors in the overlap matrices from creating intruder states. The result is a relative error of 3.4 kcal/mol, more than five times larger than the 0.6 kcal/mol error produced when the exact 3-body RDM is used to construct the overlap matrices. Strongly contracted operators restrict the ansatz's freedom in a more intelligent way than the approximate overlap truncation and are effective at removing intruder states without using the 3-body RDM. SC-CTSD's relative error of 1.0 kcal/mol is a significant improvement versus 3.4 kcal/mol and is almost as good as the 0.6 kcal/mol achieved by using the exact 3-body RDM. This hierarchy of accuracy (CTSD with 3-body RDM $>$ SC-CTSD $>$ CTSD without 3-body RDM) is typical and has also been observed in H_2O and NiO.

The results of alternative commutator approximations in N_2 are shown in Table 2.7 and Fig. 2.7, and are broadly similar to the case of H_2O. As in H_2O, the QCTSD method is less accurate than standard CTSD, and using the exact 3-body RDM to improve the commutator approximation results in an improvement in accuracy. Unlike in H_2O, the Q3CTSD method was the most accurate of the CT methods, and in fact the most accurate of all the dynamic correlation methods tested. While Q3CTSD's accuracy is not as reliable as that of the less expensive CTSD (it performed poorly for H_2O), it is nonetheless interesting that Q3CTSD outperforms MRCI+Q. This is significant because although Q3CTSD has a higher cost scaling than CTSD and CASPT2, its cost scaling is lower than MRCI+Q.

2.5.5 Transition metal oxides

One area in which CT theory can make important contributions is the quantitative study of transition metal chemistry. Transition metal electronic structure is in general difficult to treat with single-reference methods due to the strong correlations associated with the locality of the d orbitals. It also poses a significant problem for traditional multireference techniques due to the large number of valence orbitals involved. To explicitly correlate the 4s, 3d, and 4p orbitals of just two transition metals requires 18 active orbitals, which is already beyond the reach of CASSCF, CASPT2, and MRCI. Some interesting enzyme active sites, such as that found in aconitase, contain four or more transition metals, not to mention neighboring O and S atoms. Although advanced static correlation methods such as DMRG cannot yet treat such large active sites, they have been applied to $[Cu_2O_2]^{2+}$ clusters containing 32 active orbitals [90]. CT the-

ory is the only feasible method for treating dynamic correlation on top of such large active spaces. To assess the capabilities of CT theory for transition metal chemistry, it has been applied to both FeO by Yanai and Chan [89] and to NiO by the present author [70]. These single-metal oxides were chosen because they are small enough to allow a comparison with traditional dynamic correlation methods. The active spaces for both systems consisted of the O 2p and metal 4s, 4p, and 3d orbitals.

In both FeO and NiO, the CTSD method with exact overlap had a relative error of less than 2 kcal/mol when compared to MRCI+Q, as shown in Table 2.8. For comparison, CASPT2 had a 5.5 kcal/mol error in FeO and a 1.8 kcal/mol error in NiO. SC-CTSD, which unlike CTSD and CASPT2 does not use the 3-body RDM, was also applied to NiO and produced a relative error of 3.7 kcal/mol. These results suggest that CTSD is more reliable in metal oxides than CASPT2 and give hope that it will be effective at describing larger transition metal clusters. While SC-CTSD was not as accurate as CTSD, it has an even greater potential to treat large clusters by avoiding the use of the 3-body RDM. The study of NiO also revealed the importance of using the full set of valence orbitals, which becomes difficult when more than one transition metal is present. When the Ni 4p orbitals were excluded from the active space, the relative error of CASPT2 increased to 10.1 kcal/mol. In multi-metal clusters, methods such as CASPT2 and CTSD which require the 3-body RDM will typically be forced to use an active space smaller than the full valence space and will likely suffer similar accuracy penalties. Another advantage of CT theory is its relatively low cost scaling, which in metal oxides begins to have a significant effect. The CASPT2 and CTSD calculations on FeO required a comparable amount of computation time, while MRCI+Q was more than an order of magnitude more expensive. In

Table 2.8: CT results for the FeO and NiO potential energy curves.

The MRCI+Q energy is reported in $\mathrm{E_h}$, with other methods reported as the difference from MRCI+Q in $\mathrm{mE_h}$.

FeO[a]					NiO[b]				
R[c]	MRCI+Q	CASSCF	CASPT2	CTSD[d]	R[c]	MRCI+Q	CASSCF	CASPT2	CTSD[e]
1.50	-1337.65843	299.22	-8.20	3.09	1.50	-1582.54099	751.29	-74.86	-3.11
1.65	-1337.67302	290.00	-5.73	3.00	1.60	-1582.55611	746.61	-73.25	-2.91
1.72	-1337.66923	284.82	-4.12	3.85	1.70	-1582.55476	742.20	-72.33	-2.79
2.00	-1337.62980	265.74	0.57	3.52	1.80	-1582.54261	736.47	-71.98	-5.20
Relative Error $(\mathrm{mE_h})$		33.48	8.77	1.12	Relative Error $(\mathrm{mE_h})$		14.82	2.88	2.40

[a] Data for FeO reprinted with permission from Ref. [89], copyright 2007, American Institute of Physics.

[b] Data for NiO reprinted with permission from Ref. [70], copyright 2010, American Institute of Physics.

[c] Bond length in Angstroms

[d] $\tau_1 = 0.3$, $\tau_2 = 0.05$, exact overlap

[e] $\tau_1 = 0.3$, $\tau_2 = 0.1$, exact overlap

NiO, the MRCI+Q calculations were more than 50 times more expensive than those for SC-CTSD. In summary, CT theory has proven both affordable and accurate in initial studies on transition metal systems.

2.5.6 Effect of truncation thresholds

As described in Sec. 2.4.2, the standard CTSD method requires the choice of overlap matrix truncation thresholds τ_1 and τ_2 in order to prevent intruder states. Unfortunately, there is no way of knowing good values of these thresholds a priori, and in some systems good values may not exist at all. It is therefore important to analyze the effect of these thresholds on the CT energy to avoid creating artificially accurate or inaccurate results. In what we term a well behaved system, there is a range of values for both thresholds over which the CT energy changes little or not at all. We say that the energy of such a system is independent of the truncation thresholds. In some difficult systems, there may be no obvious choice for the thresholds (especially for τ_1), and one then has to investigate the effect of thresholds on the potential energy surface and report the results. This problem arose when applying CTSD to NiO, where we found that there was no region of values for τ_1 over which the CT energy was stable. We therefore produced three potential energy curves for three different choices of τ_1 so as to assess the effect of the threshold. These curves can be seen in Fig. 2.8, where the curves have been shifted so that the energies are equal at 1.5 Angstroms. In this case it happened that all three choices gave results with relative errors of less than 2 kcal/mol when compared to MRCI+Q, so although absolute energies were sensitive to the threshold, relative energies were not.

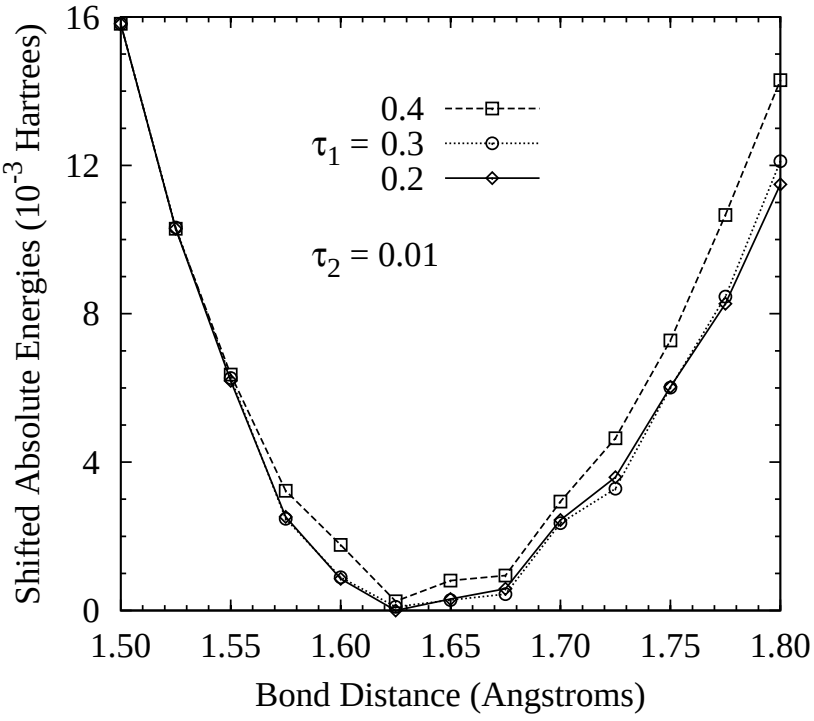

Figure 2.8: NiO energy dependence on CT overlap truncation thresholds.

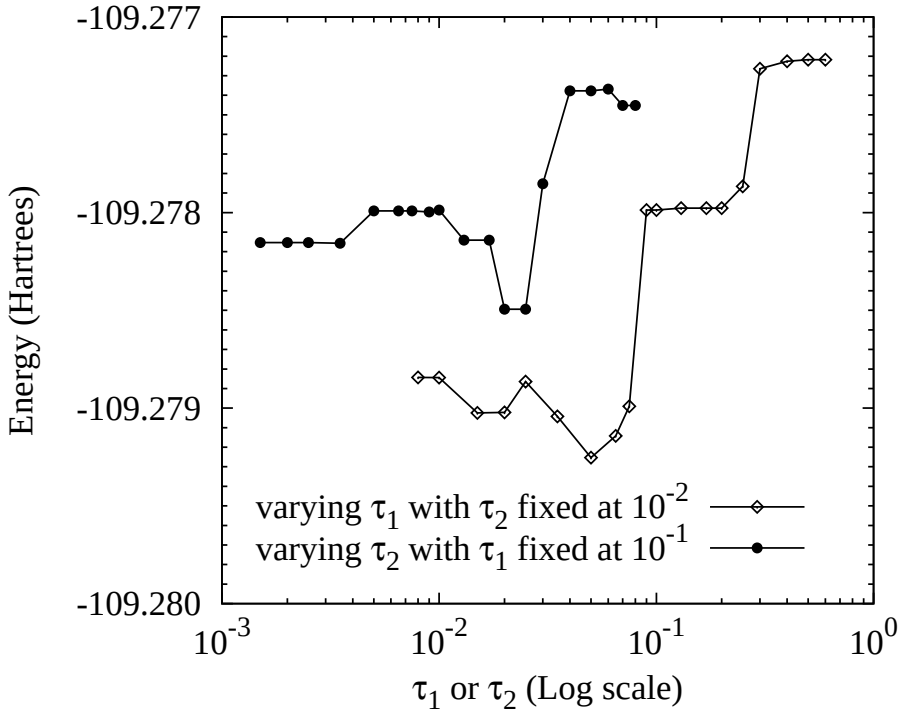

Figure 2.9: N_2 energy dependence on CT overlap truncation thresholds.

A typical approach for analyzing the threshold-sensitivity of the CTSD energy is shown in Fig. 2.9, where we have plotted the energy of the N_2 dimer against both thresholds. Our technique is to hold one threshold fixed while varying the other, and to repeat this process until a region of stability is found for both thresholds. The bond length used is 1.1208 Angstroms, and the basis set and CASSCF reference function are the same as those used in Sec. 2.5.4. We see that τ_2 has a region of stable energy between 0.01 and 0.002, below which the CT equations are too poorly conditioned to solve numerically. Thus we say that the double excitations necessary to describe dynamic correlation have all been included at a threshold of 0.1. For τ_1, there is a region of stability between 0.2 and 0.09, below which the energy fluctuates until the amplitude equations become unsolvable at 0.008. We are therefore justified in the choices $\tau_1 = 0.1$ and $\tau_2 = 0.01$ used in the strong contraction study of Sec. 2.5.4. Were we instead to choose τ_1 in the fluctuating region, say 0.03, we would introduce unnecessary error in our bonding curve due to the fact that the intruder states causing the fluctuations change with geometry. The necessity of validating the choice of τ_1 and τ_2 is bothersome in small systems and potentially unaffordable in large ones. In these cases it becomes desirable to first attempt the SC-CTSD method, whose energy is much less sensitive to changes in its single τ_ϵ threshold [70].

2.5.7 Free base porphin

Strongly contracted CT theory was used to evaluate the singlet-triplet gap of free base porphin ($C_{20}H_{14}N_4$) in two basis sets: 6-31G [42] and an ANO basis [87, 76] with spherical d orbitals and contractions of (3s,2p,1d) for C and N and (2s) for H. The active space was taken as the 24 out-of-plane 2p orbitals of C and

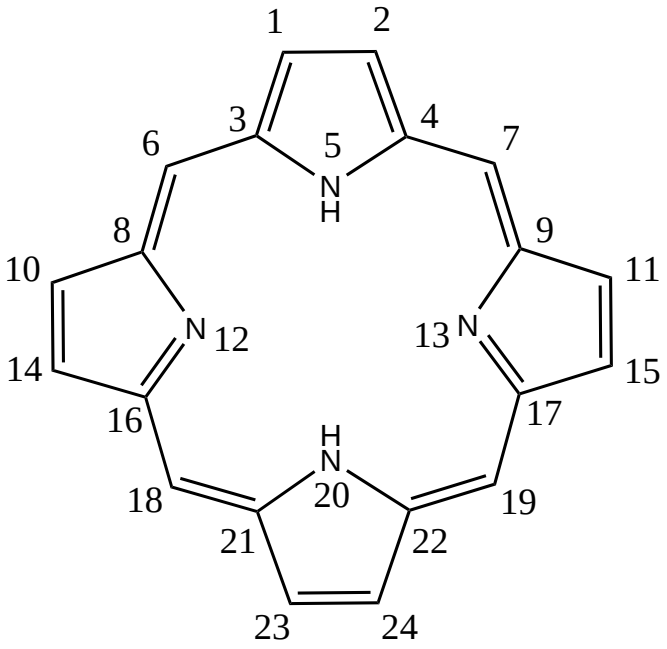

Figure 2.10: Ordering of porphin 2p orbitals on the DMRG orbital lattice.

N, necessitating a DMRG-SCF reference function in place of the usual CASSCF reference. For the DMRG-SCF calculations, a Pipek-Mezey [75] localization was applied to the out-of-plane 2p orbitals obtained from a Hartree-Fock calculation in PSI3 [27], which were then arranged on the orbital lattice as shown in Fig. 2.10. The orbitals were then optimized using 1200 DMRG states, after which the final energies and 1- and 2-body RDMs were evaluated using 2400 states. In the SC-CTSD calculations, the C and N 1s orbitals were not correlated, and the strongly contracted excitation operators were defined using the Hamiltonian in the DMRG-SCF natural orbital basis. A threshold of $\tau_\epsilon = 0.1\ \mathrm{E_h}$ was employed, although it proved unnecessary as none of the strongly contracted excitation operators displayed intruder state character (their approximate Jacobian eigenvalues were all larger than the threshold). In both the SC-CTSD and CASPT2 results discussed, the lowest lying triplet state was of $\mathrm{B_{2u}}$ symmetry.

Table 2.9: Singlet-triplet gaps for free base porphin.

Method	DMRG States	Basis Set	Gap (eV)
	Vertical Gaps		
DMRG-SCF	1200	6-31G	1.56
SC-CTSD	1200	6-31G	1.85
DMRG-SCF	2400	6-31G	1.57
SC-CTSD	2400	6-31G	1.87
DMRG-SCF	1200	ANO	1.63
DMRG-SCF	2400	ANO	1.65
SC-CTSD	2400	ANO	1.95
CASPT2[a]	-	ANO	1.52
B3LYP	-	6-31G*	1.74
	Non-Vertical Gaps		
SC-CTSD[b]	2400	6-31G	1.65
SC-CTSD[b]	2400	ANO	1.73
CASPT2[ab]	-	ANO	1.30
B3LYP	-	6-31G*	1.53
Experiment[c]	-	-	1.58

[a] Ref. [79].

[b] Approximated using the B3LYP geometry relaxation.

[c] Ref. [39].

The DMRG and SC-CTSD calculations were performed using the geometry optimized by Haeser et. al. through density functional theory [3]. These calculations, like the CASPT2 results of Roos et. al. [79], correspond to a vertical excitation in which the triplet state's geometry is not allowed to relax. However, the measurement of the experimental gap was performed by observing phosphorescence emission [39], which, due to the millisecond time scale separating excitation and emission, measures the non-vertical gap (the gap after the triplet geometry has relaxed). Therefore, to compare to experiment, it would have been more appropriate to calculate the non-vertical singlet-triplet gap. To approximately correct for this disparity, we optimized the geometry of both the singlet and triplet states with spin-unrestricted B3LYP density functional theory [16] in the 6-31G* basis set [40] using the GAUSSIAN-03 program package [36]. The change in the B3LYP singlet-triplet gap due to geometry relaxations was then combined with the vertical gaps of the other methods to produce approximate non-vertical gaps, which can be more appropriately compared with experiment. The results of these calculations are shown in Table 2.9, while the geometries involved can be found in the supplemental information of Ref. [70].

The approximate non-vertical SC-CTSD singlet-triplet gaps were 1.65 and 1.73 eV for the 6-31G and ANO basis sets, respectively. These gaps are both within 0.15 eV of the 1.58 eV experimental value. After accounting for geometry relaxation, the CASPT2 gap is 1.30 eV, which has an error of 0.28 eV when compared to experiment. Finally, B3LYP density functional theory produced a 1.53 eV non-vertical gap, in error by only 0.05 eV. Note that none of the theoretical methods account for solvent effects, which should be kept in mind when comparing to the experiment.

Table 2.10: Orbital occupations for porphin's active space natural orbitals.

Orbital	Singlet Occupation	Triplet Occupation
1	1.9859	1.9877
2	1.9854	1.9876
3	1.9636	1.9567
4	1.9612	1.9548
5	1.9488	1.9427
6	1.9456	1.9393
7	1.9405	1.9369
8	1.9301	1.9238
9	1.9167	1.9069
10	1.9149	1.8933
11	1.9147	1.8754
12	1.8197	1.7229
13	1.7955	1.1608
14	0.2286	0.8502
15	0.2069	0.3104
16	0.1075	0.1436
17	0.0779	0.1013
18	0.0649	0.0724
19	0.0620	0.0719
20	0.0591	0.0625
21	0.0444	0.0504
22	0.0429	0.0498
23	0.0424	0.0497
24	0.0407	0.0490

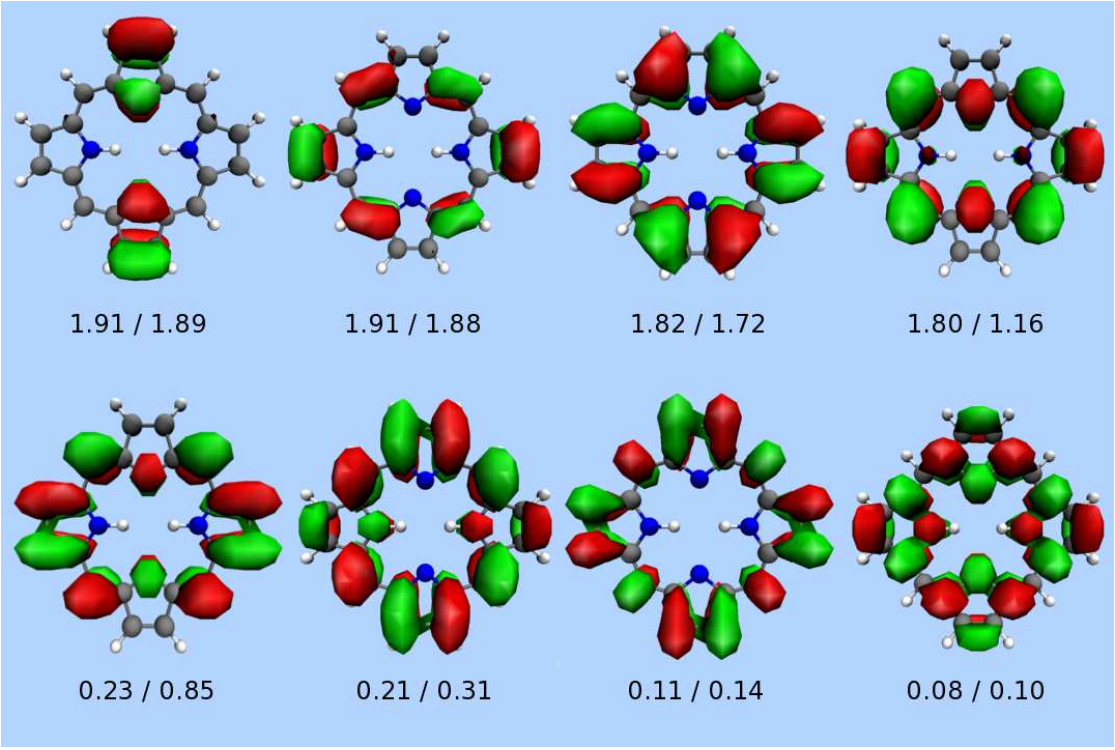

Figure 2.11: Isosurface plots of eight porphin orbitals.

An important difference between the SC-CTSD and CASPT2 results is found in the role of intruder states. Limited to a 14-orbital active space, the CASPT2 calculation required a 0.4 Hartree level shift in order to avoid intruder states, which the authors cautioned could produce up to 0.2 eV of error in the excitation energy [79]. In contrast, the strongly contracted operators used in CT theory showed no intruder state characteristics, allowing an unambiguous singlet-triplet gap to be obtained.

We have included in Table 2.10 the singlet and triplet natural orbital occupations for the 2400 state DMRG solutions in the ANO basis set. All orbitals of the singlet and triplet states had occupations differing from single reference behavior by more than 0.01, while 16 of the singlet and 17 of the triplet orbitals had

occupations differing by more than 0.05. Isosurface plots of eight of the singlet state's orbitals are shown in Fig. 2.11, along with the orbital occupations in the single/triplet states (the triplet orbitals were of the same qualitative shapes as the singlet orbitals). As a final note, we observe that the contribution of correlations between the active (out-of-plane 2p) and external (everything else) orbitals to the SC-CTSD singlet-triplet gap was not sensitive to either the number of DMRG states retained or the presence of polarization functions in the basis set. This can be seen by recognizing that the difference between the DMRG-SCF and SC-CTSD vertical energy gaps (0.30 eV) changes by less than 0.01 eV across the three SC-CTSD calculations that were performed.

2.6 Conclusion

Canonical transformation (CT) theory is a dynamic correlation theory formulated for efficient application to multireference systems. As we have seen, the theory achieves many of the desirable properties that one looks for in a multireference dynamic correlation theory, including size consistency, the ability to treat large active spaces, and an n^6 cost scaling proportional to that of CCSD. In addition to these good formal properties, CT theory has demonstrated good performance in both benchmark systems, such as $\mathrm{N_2}$, FeO, and NiO, and systems at the limits of multireference quantum chemical description, such as $[\mathrm{Cu_2O_2}]^{2+}$ and free base porphin, with a typical accuracy significantly exceeding that of multireference perturbation theory. CT theory is a relatively young theory, however, and there remain a number of unanswered questions and possible improvements.

1. Two improvements that can be made to a dynamic correlation theory are the use of explicit correlation and density fitting techniques. While we foresee no fundamental difficulties in augmenting the theory with an explicit correlation method, the use of density fitting is not so straightforward. Initial investigations on the N_2 molecule have revealed that the CT effective Hamiltonian may lack the considerable 2-body integral sparsity that is typically exploited by density fitting techniques. It is therefore not obvious how to employ density fitting in CT theory, but we believe the topic to be of significant research interest because it may simplify the challenge of constructing and storing the effective Hamiltonian.

2. Another common feature of dynamic correlation theories is the ability to evaluate analytic derivatives. This capability is under development for CT theory, which will allow the evaluation of nuclear gradients as well as response properties like polarizabilities.

3. A less well understood but no less interesting topic is whether the canonical transformation can be optimized *simultaneously* with the underlying static correlation wavefunction in order to better capture the effect of active space relaxations that result from the inclusion of dynamic correlation. This simultaneous procedure would also eliminate the dichotomy between the "diagonalize and perturb" and "perturb and diagonalize" approaches to multireference models. However, little is understood as to how the cumulant and operator decompositions would affect the stability of such an optimization. As we have seen, great care is required to prevent these approximations from undermining the stability of CT theory itself, so it is natural to be cautious when entangling them with a static correlation method.

4. Although CT theory is primarily intended for application to large multireference systems that are out of reach of traditional methods, it has proven remarkably competitive with existing methods in small molecules such as N_2 and H_2O. It would therefore be prudent to formulate a version of the theory that can make use of not only the exact 3-body RDM but also a pre-contracted 4-body RDM of the sort that is commonly employed in multireference perturbation theory. This would create a more accurate and robust method for small systems, and perhaps even large systems with small active spaces. Along similar lines, it would be interesting to develop a single-reference version of CT theory, in which many simplifications and efficiency gains should be possible. Research in these directions would allow theorists to select the type of canonical transformation theory most suitable to their specific application. It would also allow for more direct and informative comparisons between the behavior of CT theory, single-reference coupled cluster theories, and multireference theories that make use of higher order wavefunction information.

5. Another area in which CT theory should be more extensively studied is the modeling of excited states. Preliminary investigations into this topic have produced promising results but have also raised a number of questions about how best to approach the problem. For example, should one canonical transformation be defined for all excited states, akin to the similarity transform in equation-of-motion coupled cluster theory, or should each excited state be modeled with its own transformation that is specifically tailored to its reference wavefunction? In addition to answering such questions, the more basic question of how well CT theory performs for excited states has yet to be systematically investigated. In short, more ex-

tensive studies of excited states using CT theory are called for.

6. Finally, we would like to raise the question of whether there is a better way to formulate efficient models for multireference dynamic correlation than through the use of internally contracted excitations. In this review we have advocated the use of even more restricted formulations, such as strong contraction, that increase the stability of CT theory. However, these excitations are still ultimately based on the internally contracted excitations employed in perturbation theory and configuration interaction. The usual justification for using internal contraction is that the obvious alternative of considering excitations from individual determinants in the active space becomes impractical for large active spaces, as the number of such determinants grows exponentially. However, the advent of advanced wavefunctions such as DMRG's matrix product state allow the active space wavefunction to be expressed in forms that require only a polynomial amount of information. It would be interesting to search for external excitations that are more suitable for use with the matrix product state and other advanced wavefunctions, for which the "natural" starting point is something other than the uncontracted excitations that arise from a CASSCF wavefunction. Such excitations could have important implications for CT theory, but also for the general problem of modeling dynamic correlation in multireference systems.

CHAPTER 3

JASTROW FACTOR AND CORRELATOR PRODUCT STATE WAVEFUNCTIONS

3.1 Introduction

Describing the qualitative electronic structure of molecules and materials containing large numbers of strongly interacting valence electrons represents one of the key unsolved challenges in quantum chemistry and electronic structure theory. In the previous chapter, we assumed that the qualitative behavior of the valence electrons was known, and focused instead on the effects of strong interactions on the fine details of the wavefunction known as dynamic correlation. In this chapter, we will address the equally essential task of describing the static correlation between valence electrons responsible for their qualitative behavior.

Before describing how we will address the challenge of modeling static correlation, it is useful to create a clear view of exactly what is meant by the phrase "strong interactions". In the absence of any interactions between electrons, a quantum system will populate its low-lying single-particle eigenstates (orbitals) according to the Aufbau principle. The result, for fermions, is a Slater determinant consisting of the N orbitals with the lowest single-particle energies, N being the number of electrons present. We see that the lack of any interactions gives rise to a type of energetic order. When electron interactions are very weak compared to the single particle energetics, the wavefunction will be dominated by a single, energetically structured Slater determinant (hence the name single-reference) and will behave qualitatively like a system without any interactions at all.

A strongly interacting system can be defined as one whose interactions between electrons are large enough to qualitatively disrupt this energetic order. In such systems, the single-particle orbitals do a very poor job of keeping the electrons away from each other, leading to large coulomb repulsions (interactions) between electrons in the low-lying orbitals. From the standpoint of minimizing the system's energy, these interactions make it desirable to separate electrons by placing them in configurations other than the one prescribed by the Aufbau principle. Indeed, keeping electrons away from each other often requires the wavefunction to be a linear combination of many Slater determinants without any one being dominant, and so strongly interacting systems are commonly called multi-reference systems. In the extreme case, known as the strongly interacting limit, the coulomb repulsion is so much larger than any one-particle energy that the energetic order gives way completely to a spatial order in which the electrons are separated by being localized in different regions of space. Most multi-reference systems are not so strongly interacting as this, however, and so neither spatial nor energetic order dominate their electronic structure.

Traditional approaches in quantum chemistry have focused on exploiting energetic order and are most reliable in weakly interacting, single-reference systems. More recently, a class of wavefunctions known as tensor networks has been developed to exploit the spatial order present in very strongly interacting systems. The approximations made by these two approaches bias them towards either spatial or energetic structure, however, and so they are often unable to accurately model systems that transition from one structure to the other. A very common example of this transition in chemistry is the breaking of a chemical bond, in which the system moves from a bonding geometry whose wavefunction is dominated by energetic structure to the dissociation limit in which

electrons are localized on their respective atoms. Another common example is found in excitation processes in which the ground state is single-reference but the excited state is multi-reference. To provide an unbiased prediction of energy differences during such processes, a method should be capable of describing both energetically and spatially structured wavefunctions. One common approach to this challenge is to augment an energetically structured wavefunction such as a Slater determinant with a set of Jastrow factors, which work to add spatial structure.

In this chapter we will explore a generalization of the traditional Jastrow factor wavefunction. The vast majority of Jastrow factors used in electronic structure theory have been of a pairwise form, meaning that they only take into account the relative positions of two electrons at a time. While this type of Jastrow factor is undoubtedly the most important, the spatial structure of strongly interacting systems is not limited to creating only pairwise correlations. One may therefore attempt to augment the wavefunction's spatial structure based on the relative positions of more than two electrons at a time, at which point the Jastrow factors begin to resemble the tensor network wavefunction known as the correlator product state. As we will show in Sec. 3.2.1, Jastrow factors and correlators are mathematically identical, and the generalization to many-body Jastrow factors can be seen as a fusion of a tensor network wavefunction that exploits spatial structure with a more traditional wavefunction designed to exploit energetic structure.

While the combination of Jastrow factors and a more traditional wavefunction such as a Slater determinant allows one to exploit both spatial and energetic structure, working with these constructions is not trivial. Traditionally,

Monte Carlo integration has been required to extract information from Jastrow factor wavefunctions, because the sums and integrals involved do not simplify to a tractable closed form. For example, while a Slater determinant can be manipulated and optimized efficiently, by which we mean in a time that scales polynomially with the size of the system, the addition of Jastrow factors creates non-trivial correlations between electrons, preventing the usual simplifications associated with Slater determinants.

While Monte Carlo techniques remain applicable when the Jastrow factors are generalized into a many-body form, one must be careful to formulate them in order to accommodate the much larger number of variational parameters that result from this generalization. We will explain how two particular Monte Carlo based approaches, steepest descent and stochastic reconfiguration, can be efficiently applied to generalized Jastrow factor wavefunctions. Furthermore, we will show that it is possible in some cases to avoid the use of stochastic techniques entirely. This approach, detailed in Sec. 3.4, creates a new set of criteria for identifying efficient wavefunctions, and in some cases proves to be more efficient than the use of stochastic sampling.

3.2 Ansatz

3.2.1 Jastrow factors

Consider an orthonormal one-particle basis of size k. As the applications in this chapter are to lattice models, we will call this basis the lattice basis, although the formalism generalizes directly to a basis of orthonormal orbitals. In this

case the orbitals would typically be chosen to be localized, so that the Jastrow factor takes on a spatial meaning, although localized orbitals are not strictly required. In our lattice (one-particle) basis, an arbitrary quantum wavefunction can be written as

$$|\Psi\rangle = \sum_{n_1 n_2 \ldots n_k} \Psi_{n_1 n_2 \ldots n_k} |n_1 n_2 \ldots n_k\rangle = \sum_{\mathbf{n}} \Psi_{\mathbf{n}} |\mathbf{n}\rangle, \tag{3.1}$$

where $\mathbf{n}$ denotes the vector of occupancies $n_1 n_2 \ldots n_k$. In a spin 1/2 system,

$$|n_i\rangle \in \{|\uparrow\rangle, |\downarrow\rangle\}, \tag{3.2}$$

while in a fermion system,

$$|n_i\rangle \in \{|-\rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle\}. \tag{3.3}$$

In order to describe how we will approximate $|\Psi\rangle$, we must first define a Jastrow factor. A Jastrow factor is an operator that is diagonal in the lattice basis $|\mathbf{n}\rangle$. It is usually written as an exponential of an expansion in number operators,

$$\hat{J} = \exp(j + \sum_i j_i \hat{n}_i + \sum_{ij} j_{ij} \hat{n}_i \hat{n}_j + \ldots). \tag{3.4}$$

In most applications, the Jastrow amplitudes $j, j_i, j_{ij} \ldots$, are restricted to be real numbers, which makes the Jastrow factor $\hat{J}$ positive definite. For full generality, however, we should regard the amplitudes as possibly being complex.

A correlator, the building block of the correlator product state, is also an operator that is diagonal in the lattice basis and can be viewed as a Jastrow factor written in non-exponential form. A single correlator acts on a domain of sites. For example, a correlator on sites i, j takes the form

$$\hat{c}_{ij} = \sum_{n_i n_j} |n_i n_j\rangle \, c_{n_i n_j} \, \langle n_i n_j| = \sum_{n_i n_j} c_{n_i n_j} \hat{P}_{n_i n_j}, \tag{3.5}$$

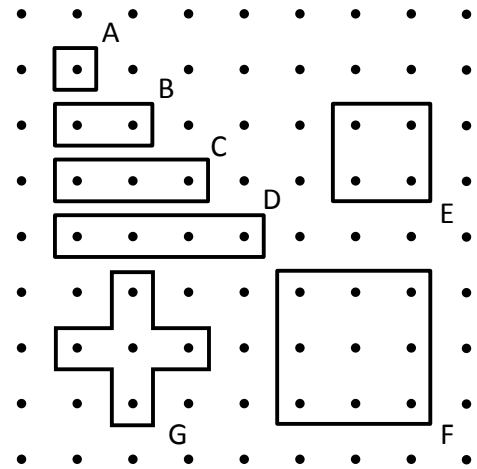

Figure 3.1: Examples of correlators on the square lattice. A: a 1-site correlator, similar to a Gutzwiller factor. B: a nearest-neighbor 2-site correlator. C: a 3-site line correlator. D: a 4-site line correlator. E: a 4-site square correlator. F: a 9-site square correlator. G: a 5-site cross correlator.

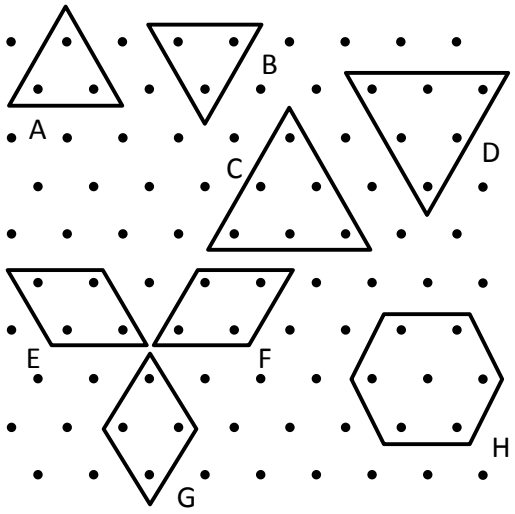

Figure 3.2: Examples of correlators on the triangular lattice. A and B: the two orientations of 3-site triangle correlators. C and D: the two orientations of 6-site triangle correlators. E, F, and G: the three orientations of 4-site rhombus correlators. H: a 7-site hexagon correlator.

where $c_{n_i n_j}$ are the correlator amplitudes, and we have introduced the projection operator $\hat{P}_{n_i n_j} = |n_i n_j\rangle\langle n_i n_j|$. The equivalence of the correlator and Jastrow factor is seen by recognizing that a two-site correlator $\hat{c}_{ij}$ is exactly representable by a two-site Jastrow factor $\exp(j + j_i \hat{n}_i + j_{ij} \hat{n}_i \hat{n}_j)$, so long as the Jastrow amplitudes are allowed to be complex. The choice of using the exponentiated Jastrow representation or the correlator representation is a matter of numerical expediency. In the current work, we henceforth use the term "correlator" to refer to both representations.

A correlator can be chosen to act on an arbitrary number of sites (see Figs. 3.1 and 3.2). Such a general correlator is written as

$$\hat{c}_\lambda = \sum_{\mathbf{n}_\lambda} c_{\mathbf{n}_\lambda} \hat{P}_{\mathbf{n}_\lambda}, \tag{3.6}$$

where $\mathbf{n}_\lambda$ is the occupancy vector of the sites in the domain of the correlator, and $\hat{P}_{\mathbf{n}_\lambda}$ is the corresponding projector.

3.2.2 Reference functions

Both Jastrow factors and correlators can be applied to reference wavefunctions $|\Phi\rangle$ to generate approximations to $|\Psi\rangle$. For the overall wavefunction to be useful, the reference function must be chosen such that expectation values and any information necessary for the wavefunction's optimization can be extracted efficiently. For the purposes of variational Monte Carlo, it is sufficient that the overall wavefunction's coefficient $\Psi_{\mathbf{n}}$ can be evaluated efficiently for any lattice configuration $\mathbf{n}$. All of the reference functions described below meet this criterion. For use with the non-stochastic algorithms described in Sec. 3.4, there is a more stringent requirement that prevents some of these references below from

being employed. The references that are compatible with the non-stochastic methods are the uniform reference and the Slater determinant, as well as the BCS (but not AGP) wavefunction.

The uniform reference

The correlator product state (CPS) [22], also called the entangled plaquette state [61, 62] and the complete graph tensor network [57], is defined by applying a product of correlators to the uniform reference $|\Phi_U\rangle$, which is an equally weighted sum over the lattice quantum basis,

$$|\Phi_U\rangle = \sum_{\mathbf{n}} |\mathbf{n}\rangle. \tag{3.7}$$

The summation in Eq. (3.7) may be chosen with symmetry constraints. For example, in this thesis, we always use a uniform reference such that for spin systems, the summation in Eq. (3.7) refers to states only with given S_z, while for fermionic system, to states with only given N and given S_z.

The CPS is then obtained by applying a product of correlators to $|\Phi_U\rangle$,

$$|\Psi\rangle = \hat{C}|\Phi_U\rangle = \prod_{\lambda} \hat{c}_\lambda |\Phi_U\rangle. \tag{3.8}$$

The correlator amplitudes $c_{\mathbf{n}_\lambda}$ provide a product approximation to the wavefunction amplitudes $\Psi_{\mathbf{n}}$ in Eq. (3.1)

$$\Psi_{\mathbf{n}} = \prod_{\lambda} c_{\mathbf{n}_\lambda}, \tag{3.9}$$

and this expression can be evaluated efficiently in a time proportional to the number of correlators. Note that the domains λ of the different correlators will

usually contain overlapping sites. For example, for a CPS wavefunction in one-dimension with nearest neighbor correlators, we have

$$|\Psi\rangle = \hat{c}_{12}\hat{c}_{23}\hat{c}_{34}\ldots\hat{c}_{k-1k}|\Phi_U\rangle. \tag{3.10}$$

By using correlators that cover increasingly larger numbers of sites, CPS become an exact family of states.

Slater determinants

The most common reference function used in conjunction with Jastrow factors is a Slater determinant [34], a formulation originally introduced by Robert Jastrow [48]. The determinant is defined by a set of orbitals $\{\phi_1, \phi_2, \ldots, \phi_N\}$ as

$$|\Phi_D\rangle = \det|\phi_1\phi_2\ldots\phi_N|, \tag{3.11}$$

which may be expressed in our lattice basis as

$$|\Phi_D\rangle = \prod_{i=1}^{N}\left(\sum_{j=1}^{2k} X_{ij}a_j^\dagger\right)|0\rangle, \tag{3.12}$$

where $a_j^\dagger$ is the creation operator for the jth one-particle state, $|0\rangle$ is the true vacuum, and $\mathbf{X}$ is the matrix of orbital coefficients whose ith row defines the orbital ϕ_i. Note that here j indexes spin-labeled one-particle states, with $j \leq k$ referring to spin $\uparrow$ states and $j > k$ referring to spin $\downarrow$ states. This form gives us a general determinant, from which spin-restricted determinants can be constructed by placing the relevant restrictions on the matrix $\mathbf{X}$.

The overall Jastrow-Slater wavefunction is $|\Psi\rangle = \hat{C}|\Phi_D\rangle$, and the coefficients $\Psi_{\mathbf{n}} = \langle\mathbf{n}|\Psi\rangle$ can be evaluated efficiently as a product of the correlators' contributions, Eq. (3.9), and the Slater determinant contribution, $\langle\mathbf{n}|\Phi_D\rangle$. The latter

contribution is equal to a determinant of orbital coefficients,

$$\langle \mathbf{n}|\Phi_D\rangle = \langle 0|\, a_{p_N} \dots a_{p_2} a_{p_1} \prod_{i=1}^{N} \left(\sum_j X_{ij} a_j^\dagger \right) |0\rangle$$

$$= \begin{vmatrix} X_{1p_1} & X_{1p_2} & \cdots & X_{1p_N} \\ X_{2p_1} & X_{2p_2} & & \vdots \\ \vdots & & \ddots & \vdots \\ X_{Np_1} & \cdots & \cdots & X_{Np_N} \end{vmatrix}, \tag{3.13}$$

where $\{p_1, p_2, \dots, p_N\}$ are the occupied one-particle states in $|\mathbf{n}\rangle$. This determinant can be evaluated efficiently in $O(N^3)$ time via the LU decomposition or other matrix factorizations.

Singlet pairing functions

Jastrow factors are also commonly used in conjunction with pairing wavefunctions. A singlet pairing wavefunction can be written in the form of the Bardeen Cooper Schrieffer (BCS) wavefunction [12],

$$|\Phi_{BCS}\rangle = \exp(\hat{F})|0\rangle, \tag{3.14}$$

where the pairing operator $\hat{F}$ creates a linear combination of all possible singlet pairings of the one-particle states,

$$\hat{F} = \sum_{i,j=1}^{k} f_{ij} a_{i\uparrow}^\dagger a_{j\downarrow}^\dagger. \tag{3.15}$$

Note that here our labels i and j refer to spin-independent one-particle states, to which we add the spin label explicitly. For example, $a_{i\uparrow}^\dagger$ creates a spin $\uparrow$ electron in state i. In order for the pairing to be made of singlets, the pairing matrix must be symmetric, $f_{ij} = f_{ji}$.

One aspect to note about the BCS wavefunction is that it contains components with all possible even particle numbers up to the maximum number allowed by our one-particle basis. This superposition of states with different particle numbers is reasonable for a bulk system connected to an electrode, but is not appropriate in isolated systems like molecules. In the latter case, we may force the system to have a specific number of electrons by applying the particle number projectors $\hat{P}^{\uparrow}_{N/2}$ and $\hat{P}^{\downarrow}_{N/2}$, which project to the subspace of the system's Hilbert space with the appropriate number of $\uparrow$ and $\downarrow$ electrons. The result of this projection is the Antisymmetrized Geminal Power (AGP) wavefunction [47, 24, 4],

$$|\Phi_{AGP}\rangle = \hat{P}^{\uparrow}_{N/2}\hat{P}^{\downarrow}_{N/2}|\Phi_{BCS}\rangle = \frac{1}{(N/2)!}\hat{F}^{N/2}|0\rangle. \tag{3.16}$$

Both the BCS [18] and AGP [21] wavefunctions have seen use in conjunction with Jastrow factors in the context of quantum Monte Carlo. Similar to the Slater determinant, the coefficient $\Psi_{\mathbf{n}}$ of these wavefunctions can be evaluated efficiently as the product between the correlator contribution, Eq. (3.9), and the BCS/AGP contribution, which takes the form of a determinant of pairing matrix elements,

$$\begin{aligned}\langle \mathbf{n}|\Phi_{AGP}\rangle = \langle \mathbf{n}|\Phi_{BCS}\rangle &= \frac{1}{(N/2)!}\langle 0|\prod_{l=1}^{N/2} a_{p_l\uparrow}a_{q_l\downarrow}\left(\sum_{i,j=1}^{k} f_{ij}a^{\dagger}_{i\uparrow}a^{\dagger}_{j\downarrow}\right)^{N/2}|0\rangle \\ &= \begin{vmatrix} f_{p_1q_1} & f_{p_1q_2} & \cdots & f_{p_1q_{N/2}} \\ f_{p_2q_1} & f_{p_2q_2} & & \vdots \\ \vdots & & \ddots & \vdots \\ f_{p_{N/2}q_1} & \cdots & \cdots & f_{p_{N/2}q_{N/2}} \end{vmatrix},\end{aligned} \tag{3.17}$$

where $\{p_1, p_2, \ldots, p_{N/2}\}$ and $\{q_1, q_2, \ldots, q_{N/2}\}$ are the positions in $|\mathbf{n}\rangle$ of the $\uparrow$ and $\downarrow$ electrons, respectively.

General pairing functions and pfaffians

For systems whose ground states are not singlets or which display significant spin-polarization, the singlet pairing operator that defines the BCS and AGP wavefunctions is no longer adequate. Instead, we can employ a generalized pairing operator

$$\hat{G} = \sum_{i=1}^{2k} \sum_{j=i+1}^{2k} g_{ij} a_i^\dagger a_j^\dagger, \tag{3.18}$$

where $i, j \leq k$ refer to spin $\uparrow$ and $i, j > k$ refer to spin $\downarrow$ one-particle states. While $\hat{F}$ creates a linear combination of all singlet pairings, $\hat{G}$ creates a linear combination of all singlet ($\uparrow\downarrow - \downarrow\uparrow$) and triplet ($\uparrow\downarrow + \downarrow\uparrow,\ \uparrow\uparrow,\ \downarrow\downarrow$) pairings.

For the $\uparrow\downarrow - \downarrow\uparrow$ and $\uparrow\downarrow + \downarrow\uparrow$ pairings, whether the singlet or triplet version is created depends on the values of the pairing matrix elements. For example, the pairing of the one-particle states 1 and 2 through the operator $a_1^\dagger a_{k+2}^\dagger \pm a_2^\dagger a_{k+1}^\dagger$ (i.e. $a_{1\uparrow}^\dagger a_{2\downarrow}^\dagger \pm a_{2\uparrow}^\dagger a_{1\downarrow}^\dagger$) is a singlet for $+$ and a triplet for $-$. If the pairing matrix elements $g_{1,k+2}$ and $g_{2,k+1}$ are real and equal in magnitude, then the singlet pair will be created if their signs are equal and the triple pair if they are not. If the magnitudes differ, or if $\boldsymbol{g}$ is allowed to be complex, then in general a linear combination of the singlet and triplet pairings will be created.

Unlike the Slater determinant, BCS, and AGP wavefunctions, the wavefunction coefficients $\Psi_{\mathbf{n}}$ of the generalized pairing wavefunction do not work out to be a determinant. Instead, the reference function's contribution to the wavefunction coefficient will be a pfaffian [17, 8, 9], which is a generalization of a determinant. We therefore denote the generalized pairing reference as $|\Phi_{PF}\rangle$ and define it as

$$|\Phi_{PF}\rangle = \exp(\hat{G})|0\rangle. \tag{3.19}$$

The coefficient of this wavefunction for a given lattice configuration $|\mathbf{n}\rangle$ is

$$\langle \mathbf{n}|\Phi_{PF}\rangle = \frac{1}{(N/2)!}\,\langle 0|\, a_{p_N}\ldots a_{p_2}a_{p_1}\,\hat{G}^{N/2}\,|0\rangle, \tag{3.20}$$

where $\{p_1, p_2, \ldots, p_N\}$ are the occupied one-particle states in $|\mathbf{n}\rangle$. Note that our definition of $|\Phi_{PF}\rangle$ only makes use of the upper triangle of the pairing matrix (i.e. g_{ij} for $i < j$), and so we are free to assign whatever values we want to the lower triangle without affecting $|\Phi_{PF}\rangle$. The choice $g_{ij} = -g_{ji}$ is expedient, for it allows us to write Eq. (3.20) as

$$\langle \mathbf{n}|\Phi_{PF}\rangle = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} \mathrm{sign}(\boldsymbol{\alpha},\boldsymbol{\beta})\; g_{p_{\alpha_1}p_{\beta_1}}\; g_{p_{\alpha_2}p_{\beta_2}}\; \cdots\; g_{p_{\alpha_{N/2}}p_{\beta_{N/2}}}. \tag{3.21}$$

Here the sum is over all $(N-1)!!$ partitionings of the numbers $\{1, 2, \ldots, N\}$ into two series $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_{N/2}\}$ and $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_{N/2}\}$ such that $\alpha_i < \beta_i$. The function $\mathrm{sign}(\boldsymbol{\alpha},\boldsymbol{\beta})$ is equal to the sign produced by permuting $\{1, 2, \ldots, N\}$ into the order $\{\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_{N/2}, \beta_{N/2}\}$. Up to a factor of $(N/2)!$, Eq. (3.21) is the defining equation for the pfaffian of what we will call the occupied pairing matrix,

$$\langle \mathbf{n}|\Phi_{PF}\rangle = \mathrm{pf}\begin{bmatrix} 0 & g_{p_1p_2} & g_{p_1p_3} & \cdots & g_{p_1p_N} \\ g_{p_2p_1} & 0 & g_{p_2p_3} & & \\ g_{p_3p_1} & g_{p_3p_2} & 0 & & \\ \vdots & & & \ddots & \\ g_{p_Np_1} & & & & 0 \end{bmatrix}. \tag{3.22}$$

As explained in Ref. [9], this pfaffian can be computed in $O(N^3)$ time by transforming the matrix into a block diagonal form. We therefore see that the general pairing reference, like the Slater determinant and the singlet pairing reference, can be efficiently manipulated through variational Monte Carlo techniques.

One can easily see that $|\Phi_{PF}\rangle$ contains $|\Phi_{BCS}\rangle$ by constraining the pairing matrix to satisfy $g_{i,j+k} = g_{j,i+k}$ and $g_{i,j} = g_{i+k,j+k} = 0$ for all $i, j \in \{1, 2, \ldots, k\}$.

Less obviously, $|\Phi_{PF}\rangle$ also contains all Slater determinants $|\Phi_D\rangle$ of the form given in Eq. (3.12). This includes both the restricted and unrestricted Hartree-Fock determinants, as well as more general determinants whose orbitals mix the spin $\uparrow$ and $\downarrow$ one-particle states. To see how, consider the following pfaffian identity [9],

$$\mathrm{pf}[\boldsymbol{B}^T\boldsymbol{A}\boldsymbol{B}] = \det[\boldsymbol{B}]\mathrm{pf}[\boldsymbol{A}], \tag{3.23}$$

in which $\boldsymbol{B}$ is any square matrix and $\boldsymbol{A}$ is a skew-symmetric matrix of the same dimension. By choosing $\boldsymbol{A}$ as the block diagonal matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ -1 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 1 & \\ 0 & 0 & -1 & 0 & \\ \vdots & & & & \ddots \end{bmatrix}, \tag{3.24}$$

we have $\mathrm{pf}[\boldsymbol{A}] = 1$ and can thus write the determinant of any square matrix as $\det[\boldsymbol{B}] = \mathrm{pf}[\boldsymbol{B}^T\boldsymbol{A}\boldsymbol{B}]$. If we can arrange for our occupied pairing matrix, Eq. (3.22), to be equal to $\boldsymbol{B}^T\boldsymbol{A}\boldsymbol{B}$, with $\boldsymbol{B}$ now the matrix from Eq. (3.2.2), then the coefficients $\langle \mathbf{n}|\Phi_{PF}\rangle$ and $\langle \mathbf{n}|\Phi_D\rangle$ will be equal. We can accomplish this by choosing our pairing matrix as $\boldsymbol{g} = \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$, and thus we see that any Slater determinant of the form given in Eq. (3.12) can be written as a generalized pairing wavefunction.

3.3 Variational Monte Carlo

Jastrow factor wavefunctions, both of the CPS and more traditional Jastrow-Slater and Jastrow-AGP varieties, have so far been used almost exclusively in

conjunction with Monte Carlo algorithms. The reason for this state of affairs is relatively simple: many of the integrals (when working in real space) and summations (when working in Fock space) that arise in computing expectation values for these wavefunctions can only be practically evaluated by using stochastic sampling. In this section we will overview the use of variational Monte Carlo [60, 71], which uses stochastic sampling to compute and minimize the energy of the chosen wavefunction.

3.3.1 Variational energies

The energy of any wavefunction can be written as

$$E = \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle} = \sum_{\mathbf{n}} \frac{|\Psi_{\mathbf{n}}|^2}{\langle\Psi|\Psi\rangle} E_{\mathrm{L}}(\mathbf{n}), \tag{3.25}$$

where the local energy $E_{\mathrm{L}}(\mathbf{n})$ is defined by

$$E_{\mathrm{L}}(\mathbf{n}) = \sum_{\mathbf{n}'} \frac{\Psi_{\mathbf{n}'}}{\Psi_{\mathbf{n}}} \langle\mathbf{n}|H|\mathbf{n}'\rangle. \tag{3.26}$$

As long as $\Psi_{\mathbf{n}}$ can be evaluated efficiently, which is the case when combining Jastrow factors with any of the reference functions in Sec. 3.2.2, a Markov chain can be used to sample the probability distribution $|\Psi_{\mathbf{n}}|^2/\langle\Psi|\Psi\rangle$ and to efficiently compute the overall energy as an average of the sampled local energies.

In addition to computing the energy, stochastic sampling can be used to minimize the energy by optimizing the wavefunction's variational parameters. One approach is to use a stochastic estimate of the energy gradient to perform a steepest descent optimization. Alternatively, one may minimize the energy by using stochastic sampling to approximately project onto the Hamiltonian's ground state, a procedure known as stochastic reconfiguration. In either case,

one must take care that the optimization procedure can handle the large numbers of variational parameters that are introduced by many-body Jastrow factors and correlators. Traditionally, variational Monte Carlo techniques have been applied to wavefunctions with at most a few thousand variational parameters. In contrast, many-body correlators can contain hundreds of thousands or even millions of parameters. While applying the steepest descent method for such large parameter sets is relatively straightforward, the use of stochastic reconfiguration is less so. We will therefore devote the next section to explaining the stochastic reconfiguration method and how it can be formulated to tackle large parameter sets.

3.3.2 Stochastic reconfiguration

Stochastic Reconfiguration (SR) [20, 21] is an optimization method that attempts to project an initial guess for the wavefunction onto the ground state. The projection operator used for the optimization is the difference between a diagonal shift Λ and the Hamiltonian,

$$\hat{P} = \Lambda - \hat{H}. \tag{3.27}$$

For a sufficiently large shift, the Hamiltonian's ground state will be the eigenstate of the projector with the largest magnitude eigenvalue, so that repeated applications of the projector will yield the ground state. If the shift is large, the change to the wavefunction from a single application of the projector will be small. In this case, it is reasonable to approximate the result of the projection as

a linear combination of the current wavefunction and its first derivatives,

$$(\Lambda - \hat{H})|\Psi\rangle \approx \sum_{i=0}^{n_v} z_i |\Psi^i\rangle, \tag{3.28}$$

$$|\Psi^0\rangle \equiv |\Psi\rangle, \tag{3.29}$$

$$|\Psi^i\rangle \equiv \frac{\partial}{\partial x_i}|\Psi\rangle \qquad i = 1, 2, \ldots, n_v, \tag{3.30}$$

where n_v is the length of the vector of variable parameters $\boldsymbol{x}$. If we now left-multiply by each state $\langle\Psi^j|$, we arrive at a system of linear equations that may be solved for the values of $\boldsymbol{z}$,

$$\langle\Psi^j|(\Lambda - \hat{H})|\Psi\rangle = \sum_{i=0}^{n_v} \langle\Psi^j|\Psi^i\rangle z_i. \tag{3.31}$$

Once we have obtained the expansion coefficients z_i, we use them to update the wavefunction's variables as follows,

$$x_i \to x_i + \frac{z_i}{z_0} \qquad i = 1, 2, \ldots, n_v. \tag{3.32}$$

The wavefunction based on these updated variables is equivalent to the right hand side of Eq. (3.28) up to order Λ^{-2} by virtue of the fact that the states $|\Psi^i\rangle$ are the wavefunction's first derivatives. In summary, the SR method is defined by projecting by $\hat{P}$, solving for $\boldsymbol{z}$, updating the wavefunction's variables via Eq. (3.32), and then repeating these steps until the energy has converged.

The method is called *stochastic* reconfiguration because for Jastrow factor wavefunctions, Monte Carlo sampling is the only way to evaluate the expectation values in Eq. (3.31). Traditionally, the overlap matrix $\langle\Psi^j|\Psi^i\rangle$ and the vector $\langle\Psi^j|(\Lambda - \hat{H})|\Psi\rangle$ are constructed explicitly via stochastic sampling, after which Eq. (3.31) is solved by direct inversion. Let us inspect the stochastic construction of the overlap matrix, which may be written as

$$\langle\Psi^j|\Psi^i\rangle = \sum_{\mathbf{n}} \frac{|\Psi_{\mathbf{n}}|^2}{\sum_{\mathbf{n}'} |\Psi_{\mathbf{n}'}|^2} \left(\frac{\Psi^j_{\mathbf{n}}}{\Psi_{\mathbf{n}}}\right)^* \left(\frac{\Psi^i_{\mathbf{n}}}{\Psi_{\mathbf{n}}}\right), \tag{3.33}$$

where $\Psi_{\mathbf{n}}$ and $\Psi^i_{\mathbf{n}}$ are the wavefunction's coefficient for configuration $\mathbf{n}$ and the derivative of this coefficient with respect to x_i. The cost to construct the overlap matrix in this way is $n_v^2 n_s$, where n_s is the number of configurations visited during the stochastic sampling process.

We seek to increase the number of variational parameters that can be optimized by stochastic reconfiguration by solving the linear equation more efficiently. To do so, we will rely on an iterative solver (in our case the conjugate gradient method), which requires only that we compute the overlap matrix's action on a trial vector. While this can be accomplished by building the matrix and multiplying by it explicitly, there is a more efficient way. By inserting the stochastic formula for the overlap matrix, Eq. (3.33), into the right hand side of Eq. (3.31), we see that the overlap matrix's action is

$$\begin{aligned}\sum_{i=0}^{n_v} \langle \Psi^j | \Psi^i \rangle z_i &= \sum_{i=0}^{n_v} \sum_{\mathbf{n}} \frac{|\Psi_{\mathbf{n}}|^2}{\sum_{\mathbf{n}'} |\Psi_{\mathbf{n}'}|^2} \left(\frac{\Psi^j_{\mathbf{n}}}{\Psi_{\mathbf{n}}}\right)^* \left(\frac{\Psi^i_{\mathbf{n}}}{\Psi_{\mathbf{n}}}\right) z_i \\ &= \sum_{\mathbf{n}} \frac{|\Psi_{\mathbf{n}}|^2}{\sum_{\mathbf{n}'} |\Psi_{\mathbf{n}'}|^2} \left(\frac{\Psi^j_{\mathbf{n}}}{\Psi_{\mathbf{n}}}\right)^* \left[\sum_{i=0}^{n_v} \left(\frac{\Psi^i_{\mathbf{n}}}{\Psi_{\mathbf{n}}}\right) z_i\right].\end{aligned} \tag{3.34}$$

By exchanging the orders of summation, we see that the overlap matrix is just a sum of outer products and that its action can be computed at a cost of $n_v n_s$, which is a factor of n_v less than the traditional approach. Note that this assumes that the ratios $\Psi^i_{\mathbf{n}}/\Psi_{\mathbf{n}}$ have been computed in advance and stored either in memory or on disk. As the sampling procedure is perfectly parallel, very large sample lengths can be accomplished by dividing the sampling over many processors. After each processor computes its contribution to the matrix's action, the results can be combined at a cost of only $n_v \ln n_s$. Using this approach, we are able to optimize by SR the large numbers of variational parameters inherent to many-body correlators and Jastrow factors.

3.3.3 Efficient wavefunction ratios

While the descriptions of the various reference functions in Sec. 3.2.2 and the simple form of Jastrow factors make clear that the coefficients of the wavefunctions we are considering can be computed in polynomial time for arbitrary configurations of the system, there is a clever technique for speeding up these calculations in the context of quantum Monte Carlo. Typically during a Monte Carlo calculation, one deals with a walker that moves from one system configuration to another as part of a stochastic process such as a Markov chain. In quantum chemistry, the walker's moves are usually restricted to changing the position of a single electron, because this simplifies the procedure of updating the wavefunction's coefficient for the new configuration. Specifically, when one electron's position changes, only one column or row of the matrix whose determinant defines the reference function's coefficient will change. Evaluating the ratio of two determinants differing by only one row or column can be accomplished quickly through the use of the matrix determinant lemma,

$$\frac{\det(\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}})}{\det(\mathbf{A})} = 1 + \mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u}, \tag{3.35}$$

where $\mathbf{u}\mathbf{v}^{\mathrm{T}}$ is the outer product of the vectors $\mathbf{u}$ and $\mathbf{v}$. By choosing $\mathbf{u}$ and $\mathbf{v}$ correctly, this formula can be made equal to the ratio of two determinants differing by one row or one column. In that case, the right hand side of Eq. (3.35) becomes the dot product of the row or column difference and the appropriate column or row of $\mathbf{A}^{-1}$. This operation's cost is linear in the dimension of the matrix, as opposed to the cubic cost of directly re-evaluating the determinant through matrix factorization. This speedup is of course only possible if the matrix's inverse is known, but the inverse need only be directly computed once, as it too can be efficiently updated when one row or one column of the matrix

changes by using the Sherman-Morrison formula,

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}})^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}}{1 + \mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u}}. \tag{3.36}$$

When changing a single row or column of the matrix, the right hand side of Eq. (3.36) has an evaluation cost that scales as the square of the matrix dimension, which is again significantly faster than the cubic cost of direct matrix inversion through matrix factorization. Thus by having each Monte Carlo walker carry with it the inverse of its reference function's matrix, the cost of computing the wavefunction coefficients during a random walk can by significantly reduced. For the case of the generalized pairing reference, Ref. [9] shows that the same speedup is possible using the pfaffian relations analogous to Eqs. (3.35) and (3.36).

3.4 Non-stochastic methods

While the form of the Jastrow-Slater and CPS wavefunctions make them a natural fit for Monte Carlo algorithms, it is interesting to ask whether or not statistical algorithms, with their associated drawbacks of statistical error, are the only way to manipulate these states. We will now show that efficient non-stochastic algorithms exist both to evaluate observables and to optimize the CPS and Jastrow-Slater wavefunction, so long as we are willing to sacrifice the strict variational principle. The non-stochastic algorithms rely on the common structure of the Jastrow-Slater and CPS wavefunctions, namely that they are a product of commuting invertible operators acting on a simple reference state. Other wavefunctions, such as the coupled cluster wavefunction, also take this form, and the algorithms we describe are similar to the techniques used for observ-

able evaluation and wavefunction optimization in the coupled cluster literature [15].

3.4.1 Non-stochastic energy evaluation

The idea behind non-stochastic energy evaluation is the following. First, we assume that the CPS or Jastrow wavefunction $|\Psi\rangle = \hat{C}|\Phi\rangle$ (here we use $|\Phi\rangle$ to refer to either the uniform or Slater determinant reference) is an eigenstate of the Hamiltonian H,

$$H\hat{C}|\Phi\rangle = E\hat{C}|\Phi\rangle. \tag{3.37}$$

The energy is then obtained as the asymmetric expectation value

$$E = \langle\Phi|\hat{C}^{-1}H\hat{C}|\Phi\rangle = \langle\Phi|\bar{H}|\Phi\rangle, \tag{3.38}$$

where we define the similarity transformed effective Hamiltonian, $\bar{H} = \hat{C}^{-1}H\hat{C}$.

Efficient non-stochastic energy evaluation now reduces to whether or not $\langle\bar{H}\rangle$ can be efficiently obtained with the reference state $|\Phi\rangle$, where $|\Phi\rangle$ is a uniform reference in the case of CPS, or a determinant in the case of the Jastrow-Slater wavefunction. By efficient, we mean that the cost is polynomial in the lattice size. The standard expectation value $\langle\Phi|H|\Phi\rangle$ can be efficiently evaluated in either case because the individual terms in the Hamiltonian act on a small number of sites, independent of lattice size. For example, the Heisenberg and Hubbard model Hamiltonians contain terms which only act at most on a pair of sites. The corresponding terms in the effective Hamiltonian $\bar{H}$ act on a larger number of sites due to the similarity transformation by the correlators. However, if the size and range of the correlators are independent of lattice size, then

even after similarity transformation, the terms in $\bar{H}$ still act on a number of sites that will be independent of lattice size. Consequently, the expectation value $\langle \bar{H} \rangle$ can still be efficiently evaluated.

More concretely, consider the energy contribution E_{xy} from a hopping operator $a_x^\dagger a_y$ associated with sites x and y,

$$E_{xy} = \langle \Phi | \hat{C}^{-1} a_x^\dagger a_y \hat{C} | \Phi \rangle. \tag{3.39}$$

We can divide the correlators in $\hat{C}$ (and their inverses in $\hat{C}^{-1}$) into two classes: those which involve (touch) sites x or y and those which do not. Denote the product of all the correlators which involve x or y as $\hat{C}_{xy}$, and the product of the remaining correlators as $\hat{C}_{\overline{xy}}$. Then we have

$$\hat{C} = \hat{C}_{xy}\, \hat{C}_{\overline{xy}}. \tag{3.40}$$

As an example, consider the hopping operator $a_3^\dagger a_4$, and the one dimensional CPS in Eq. (3.10). Then, $\hat{C}_{34}$ is given by

$$\hat{C}_{34} = \hat{c}_{23} \hat{c}_{34} \hat{c}_{45}. \tag{3.41}$$

Because $\hat{C}_{\overline{xy}}$ only contains correlators which do *not* involve sites x or y, it can be commuted past $a_x^\dagger a_y$ in Eq. (3.39) to cancel with its corresponding inverse $\hat{C}_{\overline{xy}}^{-1}$,

$$E_{xy} = \langle \Phi | \hat{C}_{xy}^{-1}\, \hat{C}_{\overline{xy}}^{-1}\, a_x^\dagger a_y\, \hat{C}_{xy}\, \hat{C}_{\overline{xy}} | \Phi \rangle = \langle \Phi | \hat{C}_{xy}^{-1}\, a_x^\dagger a_y\, \hat{C}_{xy} | \Phi \rangle. \tag{3.42}$$

Thus, the similarity transform of $a_x^\dagger a_y$ involves only $\hat{C}_{xy}$, not the whole $\hat{C}$ operator. The correlators in $\hat{C}_{xy}$ define a cluster of k_{xy} sites, where the size depends on the sizes and ranges of the correlators and the geometry of the cluster, but is independent of lattice size as long as the correlators are local. The transformed

hopping operator $\hat{C}_{xy}^{-1} a_x^\dagger a_y \hat{C}_{xy}$ now acts on k_{xy} sites. The energy contribution E_{xy} thus requires evaluating the expectation value of a k_{xy} site operator with the reference function $|\Phi\rangle$.

To see the explicit dependence of the evaluation of $\langle a_x^\dagger a_y \rangle$ on the reference function, we separate $\hat{C}_{xy}$ into its amplitude and projection operator components

$$\hat{C}_{xy} = \sum_{\mathbf{n}_{xy}} C_{\mathbf{n}_{xy}} \hat{P}_{\mathbf{n}_{xy}}, \tag{3.43}$$

where $\mathbf{n}_{xy}$ is the occupancy vector for the cluster of sites defined by $\hat{C}_{xy}$. The expectation values of the projection operators define a many-body reduced density-matrix (RDM) γ on k_{xy} sites,

$$\gamma_{\mathbf{n}_{xy},\mathbf{n}'_{xy}} = \langle \Phi | \hat{P}_{\mathbf{n}_{xy}} a_x^\dagger a_y \hat{P}_{\mathbf{n}'_{xy}} | \Phi \rangle. \tag{3.44}$$

The evaluation of γ depends on the form of the reference function $|\Phi\rangle$. In the case of the CPS, each element of the RDM is obtained in $O(1)$ time (even with particle number and S_z restrictions). For a Jastrow-Slater wavefunction, the corresponding RDM element can be evaluated as a determinant of one-body RDM elements in $O(k^3)$ time. Once the RDM is obtained, the combination with the amplitudes is independent of the reference. The expectation value E_{xy} becomes

$$E_{xy} = \sum_{\mathbf{n}_{xy}\,\mathbf{n}'_{xy}} C^{-1}_{\mathbf{n}_{xy}}\, \gamma_{\mathbf{n}_{xy},\mathbf{n}'_{xy}}\, C_{\mathbf{n}'_{xy}}. \tag{3.45}$$

Due to the summation over the occupancy vector, the cost of Eq. (3.45) is exponential in the cluster (but not lattice) size, k_{xy}. Note that the cost is effectively the cost of a single summation over $\mathbf{n}_{xy}$, rather than the formal double summation shown above, because the sparsity of γ means that it has $O(d^{k_{xy}})$ non-zero elements, where d is dimension of a single site. For sufficiently small correlators, which lead to small clusters, the summation can be carried out affordably.

As each operator in the Hamiltonian involves a similar contribution, the entire energy may be efficiently evaluated.

The non-stochastic energy evaluation algorithm relies on little of the detailed structure of the Jastrow-Slater and CPS wavefunctions. The key steps require only that (i) $\hat{C}$ is made of a product of commuting invertible operators, (ii) only a small number of these operators do not commute with a given term in the Hamiltonian, and (iii) the reference function $|\Phi\rangle$ is sufficiently simple that the expectation values in Eq. (3.44) can be efficiently obtained. This structure is obeyed by many other wavefunctions, such as the coupled cluster wavefunction [15], and indeed for any wavefunction with this structure, an efficient non-stochastic energy evaluation algorithm may be formulated.

3.4.2 Non-stochastic optimization

Above we showed that we can evaluate the approximate energy of a CPS or Jastrow-Slater wavefunction in a non-stochastic way, as long as the individual correlators or Jastrow factors do not cover too many sites. For a complete calculation, we need to also determine the correlator or Jastrow amplitudes. Since the non-stochastic energy is not variational, we cannot obtain the optimal parameters by minimizing the energy. Instead, we require that the CPS or Jastrow wavefunction satisfies a set of non-linear projected Schrödinger equations. Solving these equations yields exactly the same solution as the minimum of the variational energy if the wavefunction provides an exact parameterization, although this is not the case for approximate CPS and Jastrow wavefunctions.

As in the previous section, we assume that the CPS or Jastrow wavefunc-

tion is a true eigenstate of the Hamiltonian, Eq. (3.37). From the Schrödinger equation, we need to obtain a set of non-linear equations equal in number to the number of amplitudes in $\hat{C}$. We obtain sufficient equations by projecting with the bras $\langle\Phi|\hat{P}_{\mathbf{n}_\lambda}$, where $\hat{P}_{\mathbf{n}_\lambda}$ are the projectors used to define the correlator operators in Eq. (3.6). Applying these bra states gives us

$$\langle\Phi|\hat{P}_{\mathbf{n}_\lambda}(\bar{H}-E)|\Phi\rangle = R_{\mathbf{n}_\lambda} = 0. \tag{3.46}$$

By requiring the residuals $R_{\mathbf{n}_\lambda}$ to vanish we determine all the correlator amplitudes $c_{\mathbf{n}_\lambda}$. As with energy evaluation, the expectation value in Eq. (3.46) can be obtained in a non-stochastic manner by tracing over clusters of sites associated with each similarity transformed term in the effective Hamiltonian $\bar{H}$.

In order to solve the simultaneous set of equations (3.46), we have taken two approaches. In the first case, we use a standard Newton-Raphson procedure to find the simultaneous zeroes of the residuals $R_{\mathbf{n}_\lambda}$. This requires evaluating the Jacobian matrix $\partial R_{\mathbf{n}_\lambda}/\partial c_{\mathbf{n}_\mu}$. Alternatively, we may take the approach of constructing and diagonalizing a local Hamiltonian for each correlator's amplitudes. The local Hamiltonian and overlap matrices that determine $c_{\mathbf{n}_\lambda}$ are defined as

$$\bar{H}_{\mathbf{n}_\lambda,\mathbf{n}'_\lambda} = \langle\Phi|\hat{P}_{\mathbf{n}_\lambda}\, c_{\mathbf{n}_\lambda}\, \bar{H}\, c^{-1}_{\mathbf{n}'_\lambda}\, \hat{P}_{\mathbf{n}'_\lambda}|\Phi\rangle, \tag{3.47}$$

$$S_{\mathbf{n}_\lambda,\mathbf{n}'_\lambda} = \langle\Phi|\hat{P}_{\mathbf{n}_\lambda}\, \hat{P}_{\mathbf{n}'_\lambda}|\Phi\rangle. \tag{3.48}$$

The correlator amplitudes are obtained from solving an eigenvalue problem for each correlator $\hat{c}_\lambda$,

$$\sum_{\mathbf{n}'_\lambda} \bar{H}_{\mathbf{n}_\lambda,\mathbf{n}'_\lambda} c_{\mathbf{n}'_\lambda} = E \sum_{\mathbf{n}'_\lambda} S_{\mathbf{n}_\lambda,\mathbf{n}'_\lambda} c_{\mathbf{n}'_\lambda}. \tag{3.49}$$

The local Hamiltonian matrix $\bar{H}_{\mathbf{n}_\lambda,\mathbf{n}'_\lambda}$ depends on the amplitudes of all the correlators, $\hat{c}_{\mu\neq\lambda}$. Thus after each correlator amplitude is obtained from the respec-

tive eigenvalue problem, Eq. (3.49), the local Hamiltonians are updated, and the procedure is iterated until convergence is achieved.

Much as in the case of energy evaluation, the formulation of the amplitude equations relies only on generic elements of the product structure of the Jastrow-Slater and CPS wavefunctions. By analogy with methods for the coupled cluster wavefunction, we can also write down a non-stochastic algorithm to obtain expectation values of arbitrary operators. Starting from the amplitude equations, we first define a Lagrangian as

$$L = \langle\Phi|\bar{H} + \sum_{\mathbf{n}_\mu} \Lambda_{\mathbf{n}_\mu} \hat{P}_{\mathbf{n}_\mu} \left(\bar{H} - \langle\Phi|\bar{H}|\Phi\rangle\right) |\Phi\rangle, \tag{3.50}$$

from which the amplitude equations arise from the stationary conditions

$$\frac{\partial L}{\partial \Lambda_{\mathbf{n}_\mu}} = 0. \tag{3.51}$$

The values of the Lagrange multipliers $\Lambda_{\mathbf{n}_\mu}$ are found by requiring the Lagrangian to be stationary with respect to the correlator variables $c_{\mathbf{n}_\mu}$,

$$\frac{\partial L}{\partial c_{\mathbf{n}_\mu}} = 0. \tag{3.52}$$

Then, derivatives of the Lagrangian with respect to the Hamiltonian's parameters define reduced density matrices as

$$\Gamma_{ij} = \langle\Phi|\overline{a_i^\dagger a_j} + \sum_{\mathbf{n}_\mu} \Lambda_{\mathbf{n}_\mu} \hat{P}_{\mathbf{n}_\mu} \left(\overline{a_i^\dagger a_j} - \langle\Phi|\overline{a_i^\dagger a_j}|\Phi\rangle\right) |\Phi\rangle, \tag{3.53}$$

where $\overline{a_i^\dagger a_j} = \hat{C}^{-1} a_i^\dagger a_j \hat{C}$. The two body reduced density matrix Γ_{ijkl} is defined likewise using $\overline{a_i^\dagger a_j^\dagger a_l a_k}$. These density matrices allow us to obtain expectation values of arbitrary one and two body operators.

3.4.3 Heisenberg model with non-stochastic methods

We have applied the correlator product state to the antiferromagnetic spin-$\frac{1}{2}$ Heisenberg model on a periodic 8x8 square lattice and a periodic 6x6 triangular lattice using both the non-stochastic and variational MC frameworks. The Hamiltonian is written as

$$H = J \sum_{\langle ij \rangle} \vec{S}_i \cdot \vec{S}_j, \tag{3.54}$$

where $\langle ij \rangle$ indicates nearest neighbor pairs and $J > 0$. The results are summarized in Tables 3.1 and 3.2. In the case of the square lattice, we have essentially exact stochastic series expansion (SSE) results with which to compare [78]. While the accuracy of the CPS ansatz is not the main question we are studying here (such studies can be found in Refs. [22, 61, 62]), we see that both the non-stochastic and variational Monte Carlo energies are within 2% of the SSE result for 4-site square correlators and about 1% for 9-site square correlators. The more central question is the relative difference between the non-stochastic and variational Monte Carlo energies. We see that in all cases, the relative difference is comparable to the intrinsic energy error associated with the wavefunction, and in one case it is significantly smaller.

As expected, the non-stochastic energy is not variational, and for small correlators it tends to be slightly below the variational energy. More surprisingly, the convergence in accuracy for the non-stochastic energy is not monotonic with correlator size. We see both larger deviations from the variational Monte Carlo results, as well as lower accuracy in the total energy, for 5-site crosses (in the square lattice) and 6-site triangles (in the triangular lattice), than for some smaller correlators. We find that square and rhombus correlators do particularly well on the square and triangular lattices, respectively, with relative differences

Table 3.1: 8x8 square Heisenberg lattice

Correlators	Non-Stochastic	Variational MC	% Difference	VMC % Error
nearest neighbor	-0.6883	-0.6534(6)	-5.34	2.98
4-site squares	-0.6659	-0.6617(3)	-0.64	1.76
5-site crosses	-0.6449	-0.6637(3)	2.82	1.46
9-site squares	-0.6651	-0.6699(2)	0.72	0.53

Comparison of non-stochastic and variational MC results for the antiferromagnetic spin-$\frac{1}{2}$ Heisenberg model on the 8x8 square lattice with periodic boundary conditions and total $S_z = 0$. The energies per site are reported in units of J, with the number in parentheses representing the uncertainty in the final digit. The variational Monte Carlo error is reported relative to the essentially exact stochastic series expansion (SSE) [78]. Translationally invariant correlators are employed, with each of the two sublattices having independent correlators. When optimizing the correlators, Marshall's sign rule is used as an initial guess.

Table 3.2: 6x6 triangular Heisenberg lattice

Correlators	Non-Stochastic	Variational MC	% Difference
nearest neighbor	-0.5253	-0.5184(2)	-1.33
3-site triangles	-0.5253	-0.5184(2)	-1.33
4-site rhombuses	-0.5353	-0.5383(3)	0.55
6-site triangles	-0.5289	-0.5419(2)	2.40
7-site hexagons	-0.5390	-0.5435(2)	0.83

Comparison of non-stochastic and variational MC results for the antiferromagnetic spin-$\frac{1}{2}$ Heisenberg model on the 6x6 triangular lattice with periodic boundary conditions and total $S_z = 0$. The energies per site are reported in units of J, with the number in parentheses representing the uncertainty in the final digit. Translationally invariant correlators are employed, with each of the three sublattices having independent correlators. When optimizing the correlators, the solution to the classical Heisenberg model is used as an initial guess.

from the variational energy all within 1%. We speculate that these correlators success is due in part to the fact that the amplitude equations used in their optimization (Eq. (3.46)) are constructed using projectors that share the translational symmetry of the underlying lattice.

One appealing aspect of the non-stochastic method for CPS on periodic spin lattices is that the cost of the method can be made independent of the lattice size by using translationally invariant correlators and by taking advantage of the uniform reference's particularly simple RDM elements (Eq. (3.44)). In both the 8x8 square and 6x6 triangular lattice, we found that the non-stochastic method was much faster than variational MC for small correlators (4-site squares/rhombuses and smaller), while being slower for larger correlators due to the exponential increase of the cost with correlator size.

3.4.4 Spinless Hubbard model with non-stochastic methods

We have studied a 20-site (4x5) spinless Hubbard lattice with open boundary conditions using the Jastrow-Slater wavefunction and the non-stochastic algorithms. We have also performed variational Monte Carlo calculations for comparison. While we could treat much larger lattices, we chose this lattice size in order to compare to exact results. The Hamiltonian for the spinless Hubbard model is

$$H = \sum_{\langle ij \rangle} -t(a_i^\dagger a_j + a_j^\dagger a_i) + U a_i^\dagger a_i a_j^\dagger a_j, \tag{3.55}$$

in which $a_i^\dagger$ and a_i are the fermionic particle creation and destruction operators on site i, and $\langle ij \rangle$ represents nearest neighbors.

Table 3.3: 4x5 spinless Hubbard lattice at half filling

U / t	Non-Stochastic	Variational MC	% Difference	Hartree Fock	Exact
0.1	-13.8157	-13.8158(1)	0.00	-13.8117	-13.8166
0.2	-13.2437	-13.2438(2)	0.00	-13.2276	-13.2475
0.4	-12.1367	-12.1368(4)	0.00	-12.0709	-12.1544
0.6	-11.0591	-11.0611(2)	0.02	-10.9539	-11.1267
0.8	-10.0850	-10.0884(2)	0.03	-9.9598	-10.1737
1.0	-9.2230	-9.2273(2)	0.05	-9.0917	-9.3066
1.2	-8.4660	-8.4712(2)	0.06	-8.3382	-8.5326
1.4	-7.8022	-7.8072(2)	0.06	-7.6827	-7.8507
2.0	-6.2491	-6.2529(2)	0.06	-6.1622	-6.2658
4.0	-3.6139	-3.6141(1)	0.00	-3.5885	-3.6151
6.0	-2.4989	-2.4988(1)	-0.01	-2.4895	-2.4991
8.0	-1.9007	-1.9000(1)	-0.03	-1.8963	-1.9007
10.0	-1.5308	-1.5304(1)	-0.03	-1.5285	-1.5308

Total ground state energies, in units of t, for the 4x5 spinless Hubbard lattice at half filling with open boundary conditions. Both the non-stochastic and variational MC methods use 4-site square Jastrow factors (correlators). Exact results were computed using the ALPS program [2]. Numbers in parentheses represent the uncertainty in the final digit.

Table 3.4: 4x5 spinless Hubbard lattice with single hole doping

U / t	Non-Stochastic	Variational MC	% Difference	Hartree Fock	Exact
0.1	-13.8554	-13.8554(1)	0.00	-13.8522	-13.8558
0.2	-13.4341	-13.4339(1)	0.00	-13.4214	-13.4359
0.4	-12.6195	-12.6194(4)	0.00	-12.5685	-12.6268
0.6	-11.8410	-11.8411(2)	0.00	-11.7262	-11.8583
0.8	-11.0976	-11.0985(3)	0.01	-10.8940	-11.1303
1.0	-10.3886	-10.3921(3)	0.03	-10.0718	-10.4434
1.2	-9.6456	-9.6579(4)	0.13	-9.2971	-9.7983
1.4	-8.9322	-8.9483(4)	0.18	-8.6411	-9.1958
2.0	-7.4356	-7.4486(3)	0.17	-7.1974	-7.6473
4.0	-4.9019	-4.9105(3)	0.18	-4.7707	-4.9959
6.0	-3.8209	-3.8238(2)	0.08	-3.7374	-3.8793
8.0	-3.2430	-3.244(1)	0.03	-3.1820	-3.2846
10.0	-2.8865	-2.885(1)	-0.05	-2.8384	-2.9185

Total ground state energies, in units of t, for the 4x5 spinless Hubbard lattice with single hole doping and open boundary conditions. Both the non-stochastic and variational MC methods use 4-site square Jastrow factors (correlators). Exact results were computed using the ALPS program [2]. Numbers in parentheses represent the uncertainty in the final digit.

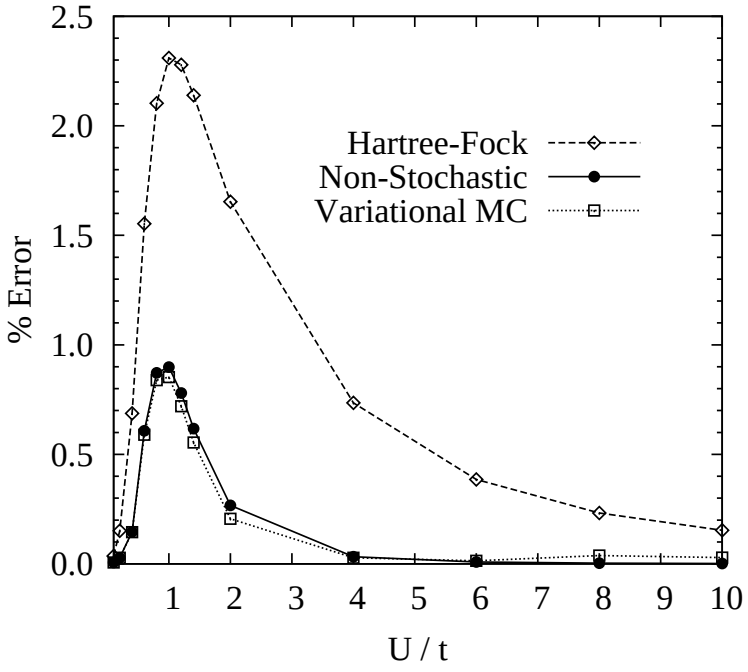

Figure 3.3: 4x5 spinless Hubbard lattice at half filling

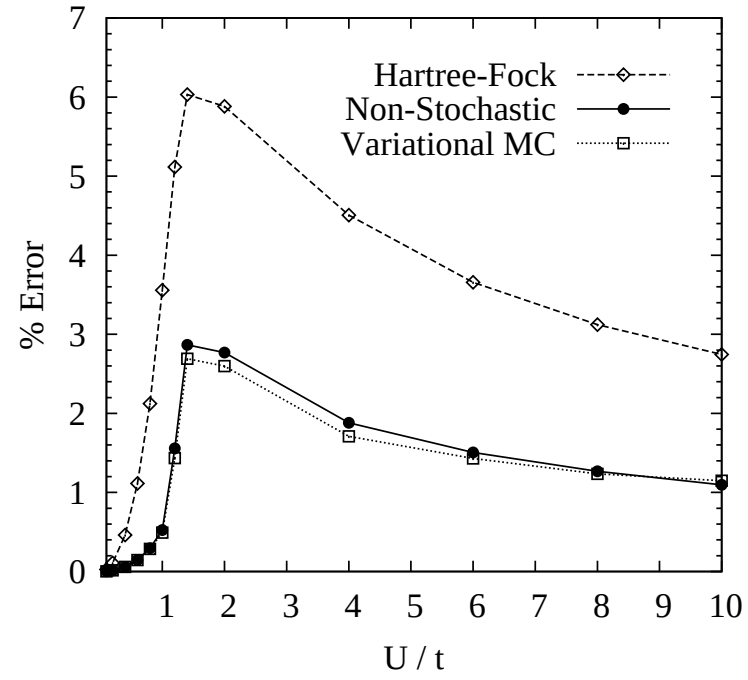

Figure 3.4: 4x5 spinless Hubbard lattice with single hole doping

Results for half-filling and single hole doping are presented in Figs. 3.3 and 3.4 and Tables 3.3 and 3.4, respectively. We find that the difference between the non-stochastic and variational Monte Carlo energies is small for all ratios of U/t at both half-filling and single hole doping. At half-filling, the largest difference is 0.06%, while for single hole doping, it is 0.18%. The energy errors of the Jastrow-Slater form (compared to the exact energy) are less than 1% for all values of U/t at half-filling, and below 3% for single hole doping. We see that the difference between the non-stochastic and variational Monte Carlo energies is here much smaller than the intrinsic error associated with the quality of the wavefunction.

3.4.5 Full Hubbard model with non-stochastic methods

We have also studied the Hubbard model at half filling with open boundary conditions, in one and two dimensions, using the non-stochastic and variational Monte Carlo algorithms for the Jastrow-Slater wavefunction. The Hubbard Hamiltonian is

$$H = -t \sum_{\langle ij \rangle} \sum_{\sigma=\uparrow,\downarrow} (a^\dagger_{i\sigma} a_{j\sigma} + a^\dagger_{j\sigma} a_{i\sigma}) + U \sum_i a^\dagger_{i\uparrow} a_{i\uparrow} a^\dagger_{i\downarrow} a_{i\downarrow}, \tag{3.56}$$

in which $a^\dagger_{i\uparrow(\downarrow)}$ and $a_{i\uparrow(\downarrow)}$ are the fermionic creation and destruction operators for particles with spin $\uparrow$ ($\downarrow$), and $\langle ij \rangle$ refers to nearest neighbors.

Since the fermions have spin, there are several choices of Slater determinant possible. We use as our Slater determinant the restricted Hartree-Fock (RHF) Slater determinant. While better energies could be obtained with an unrestricted or generalized Slater determinant, the restricted Hartree-Fock determinant is sufficient for the comparison between the non-stochastic and variational

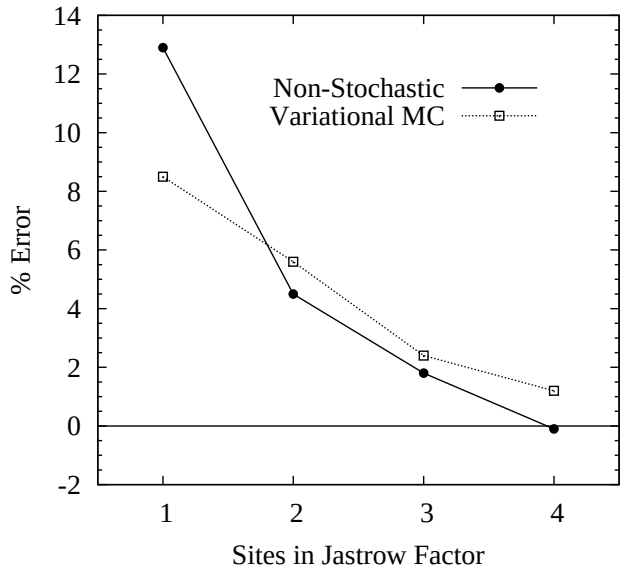

Figure 3.5: Energy errors in a 1x22 Hubbard lattice at U/t = 4

Monte Carlo algorithms that is our primary concern.

Results for the ratios $U/t = 2$ and $U/t = 4$ are presented in Table 3.5. For $U/t = 2$, the RHF Slater determinant alone produces energies in error by 6-10%, which are reduced to 1% or less after the inclusion of Jastrow factors, optimized either through the non-stochastic or variational Monte Carlo algorithms. Importantly, the non-stochastic and variational CPS energies differ from each other by less than 0.02% for both the one and two dimensional lattices. For the case of $U/t = 4$, the RHF reference is qualitatively incorrect with relative errors as high as 60%. The inclusion of Jastrow factors reduces the error to 2% and 7% in one and two dimensions, respectively. Despite the poor quality of the wavefunction in this problem, the non-stochastic energy reproduces the variational CPS energy quite well, with the relative differences in Table 3.5 never exceeding 1%. This is much less than the intrinsic error due to the quality of the wavefunction.

Finally, for the 22-site chain with $U/t = 4$ and open boundary conditions,

Table 3.5: Non-stochastic results in the Hubbard model

$U/t = 2$

Lattice	Non-Stochastic	Variational MC	% Difference	Restricted HF	DMRG
1x14	-11.241 776	-11.240(1)	-0.02	-10.133 544	-11.279 897
1x18	-14.592 961	-14.591(1)	-0.01	-13.219 131	-14.653 987
1x22	-17.946 260	-17.947(2)	0.00	-16.307 287	-18.029 379
4x5	-19.920 320	-19.917(1)	-0.02	-18.800 678	-20.127 521

$U/t = 4$

Lattice	Non-Stochastic	Variational MC	% Difference	Restricted HF	DMRG
1x14	-7.631 100	-7.556(1)	-0.99	-3.133 544	-7.672 349
1x18	-9.842 409	-9.770(3)	-0.74	-4.219 131	-9.965 398
1x22	-12.042 344	-11.968(4)	-0.62	-5.307 287	-12.259 082
4x5	-13.384 297	-13.350(1)	-0.26	-8.800 678	-14.404 488

Total ground state energies for the Hubbard model at half filling with open boundary conditions and total $S_z = 0$. The DMRG results used m=1600 renormalized states. The Jastrow factors (correlators) employed were 3-site lines for the 1D lattices and 4-site squares for the 4x5 lattice. Energies are in units of t, with the number in parentheses representing the uncertainty in the final digit.

we investigated the effect of varying the size of the Jastrow factor. As shown in Fig. 3.5, both the non-stochastic and variational energies improve when extending the Jastrow factor from one to four sites. We observe in all cases that the non-stochastic and variational energies differ by an amount significantly less than the intrinsic energy error of the wavefunction, except for the 4-site Jastrows, where the non-stochastic energy lies slightly below (-0.1%) below the true energy, while the variational energy is above (1.2%) and thus the difference between the non-stochastic and variational energies is almost exactly the same as the intrinsic variational energy error.

3.5 Conclusions

Jastrow factor wavefunctions offer a powerful framework for describing both weakly and strongly interacting systems. A reference function such as a Slater determinant or one of a number of pairing functions captures the mean-field behavior of weakly interacting electrons, while many-body Jastrow factors handle strong interactions in a robust, non-perturbative manner. While the use of many-body Jastrow factors improves the wavefunction's ability to handle strong interactions, it also greatly increases the number of variational parameters in the wavefunction. We have discussed how to accommodate this increase in variational Monte Carlo simulations and have shown that the stochastic reconfiguration method in particular can be improved in this context by introducing an iterative solver for its linear equation.

We have also shown that efficient non-stochastic algorithms exist both to evaluate the energy and expectation values of Jastrow-Slater and correlator

product state wavefunctions, as well as to optimize the wavefunction parameters. We have applied the non-stochastic methods to three model systems: the spin-$\frac{1}{2}$ antiferromagnetic Heisenberg model, the spinless Hubbard model, and the full Hubbard model. While unlike the variational Monte Carlo energy, the non-stochastic energy is not a strict upper bound, the difference between the two energies is comparable to and often significantly less than the intrinsic error associated with the quality of the wavefunction. In practice we find that the non-stochastic algorithms are faster than the variational Monte Carlo algorithms for small correlator (or Jastrow) sizes, but become more expensive for larger correlators.

The non-stochastic algorithms we have described rely on the mathematical form of the Jastrow-Slater and correlator product state wavefunctions as a product of commuting invertible operators acting on a simple reference wavefunction. Any wavefunction with this mathematical form may be studied with analogous efficient non-stochastic techniques. This possibility can guide the construction of efficient new classes of wavefunctions in the future.

CHAPTER 4

CONCLUSION

Explicit approximations of a molecule's wavefunction offer a powerful technique for extracting information from the Schrödinger equation. By exploiting the underlying energetic and spatial structures present in molecular physics, these approximations simplify the wavefunction in order to permit the efficient calculation of molecular properties. Here we have considered two specific wavefunction approximations intended to tackle separate aspects of the challenge of describing correlation between electrons. First, Canonical Transformation theory was explored as a way to approximate the effects of small correlations between the strongly interacting valence electrons and the weakly interacting core electrons. Through an approximate canonical transformation to a system of quasi-particles, in which many-body interactions are reduced to combinations of 1- and 2-body interactions, this technique achieves a remarkably efficient treatment of multi-reference dynamic correlation. Second, the theory of Jastrow factor wavefunctions was generalized to include many-body Jastrow factors, improving the ability of the ansatz to capture strong static correlations between spatially localized electrons. To overcome the substantial increase in wavefunction complexity associated with many-body Jastrow factors, two new optimization algorithms were introduced. Together, these advances in Canonical Transformation theory and Jastrow factor wavefunctions bring electronic structure theory one small step closer to a complete, predictive understanding of molecular behavior.

APPENDIX A

STRONGLY CONTRACTED EXCITATION OPERATORS

Here we present definitions for the 8 types of strongly contracted excitation operators in terms of the spin free excitation operators $E^p_q = \sum_{\sigma=\alpha}^{\beta} a^\dagger_{p_\sigma} a_{q_\sigma}$ and $E^{pq}_{rs} = \sum_{\sigma,\tau=\alpha}^{\beta} a^\dagger_{p_\sigma} a^\dagger_{q_\tau} a_{s_\tau} a_{r_\sigma}$. See Sec. 2.4.3 for a description of how we arrive at these definitions.

$$\hat{o}^{v_1 v_2} = \sum_{a_1 a_2} g^{v_1 v_2}_{a_1 a_2} (E^{v_1 v_2}_{a_1 a_2} - E^{a_1 a_2}_{v_1 v_2}) \tag{A.1}$$

$$\hat{o}_{c_1 c_2} = \sum_{a_1 a_2} g^{a_1 a_2}_{c_1 c_2} (E^{a_1 a_2}_{c_1 c_2} - E^{c_1 c_2}_{a_1 a_2}) \tag{A.2}$$

$$\hat{o}^{v_1}_{c_1 c_2} = \sum_{a_1} g^{v_1 a_1}_{c_1 c_2} (E^{v_1 a_1}_{c_1 c_2} - E^{c_1 c_2}_{v_1 a_1}) \tag{A.3}$$

$$\hat{o}^{v_1 v_2}_{c_1} = \sum_{a_1} g^{v_1 v_2}_{a_1 c_1} (E^{v_1 v_2}_{a_1 c_1} - E^{a_1 c_1}_{v_1 v_2}) \tag{A.4}$$

$$\hat{o}^{v_1 v_2}_{c_1 c_2} = g^{v_1 v_2}_{c_1 c_2} (E^{v_1 v_2}_{c_1 c_2} - E^{c_1 c_2}_{v_1 v_2}) \tag{A.5}$$

$$\begin{aligned} \hat{o}^{v_1}_{c_1} = & \left(t^{v_1}_{c_1} + \sum_{c_2} \left(2 g^{v_1 c_2}_{c_1 c_2} - g^{c_2 v_1}_{c_1 c_2} \right) \right) \left(E^{v_1}_{c_1} - E^{c_1}_{v_1} \right) \\ & + \sum_{a_1 a_2} g^{v_1 a_2}_{c_1 a_1} \left(E^{v_1 a_2}_{c_1 a_1} - E^{c_1 a_1}_{v_1 a_2} \right) \\ & + \sum_{a_1 a_2} g^{a_2 v_1}_{c_1 a_1} \left(E^{a_2 v_1}_{c_1 a_1} - E^{c_1 a_1}_{a_2 v_1} \right) \end{aligned} \tag{A.6}$$

$$\begin{aligned} \hat{o}^{v_1} = & \sum_{a_1} \left(t^{v_1}_{a_1} + \sum_{c_2} \left(2 g^{v_1 c_2}_{a_1 c_2} - g^{c_2 v_1}_{a_1 c_2} \right) \right) \left(E^{v_1}_{a_1} - E^{a_1}_{v_1} \right) \\ & + \sum_{a_1 a_2 a_3} g^{v_1 a_3}_{a_1 a_2} \left(E^{v_1 a_3}_{a_1 a_2} - E^{a_1 a_2}_{v_1 a_3} \right) \end{aligned} \tag{A.7}$$

$$\begin{aligned} \hat{o}_{c_1} = & \sum_{a_1} \left(t^{a_1}_{c_1} + \sum_{c_2} \left(2 g^{a_1 c_2}_{c_1 c_2} - g^{c_2 a_1}_{c_1 c_2} \right) \right) \left(E^{a_1}_{c_1} - E^{c_1}_{a_1} \right) \\ & + \sum_{a_1 a_2 a_3} g^{a_1 a_2}_{c_1 a_3} \left(E^{a_1 a_2}_{c_1 a_3} - E^{c_1 a_3}_{a_1 a_2} \right) \end{aligned} \tag{A.8}$$

BIBLIOGRAPHY

[1] R. Ahlrichs, P. Scharf, and C. Ehrhard. The coupled pair functional (CPF). a size consistent modification of the CI(SD) based on an energy functional. *J. Chem. Phys.*, 82:890, 1985.

[2] A. Albuquerque, F. Alet, P. Corboz, P. Dayal, A. Feiguin, S. Fuchs, L. Gamper, E. Gull, S. Gürtler, A. Honecker, R. Igarashi, M. Körner, A. Kozhevnikov, A. Läuchli, S. R. Manmana, M. Matsumoto, I. P. McCulloch, F. Michel, R. M. Noack, G. Pawłowski, L. Pollet, T. Pruschke, U. Schollwöck, S. Todo, S. Trebst, M. Troyer, P. Werner, S. Wessel, and the ALPS collaboration. The ALPS project release 1.3: open-source software for strongly correlated systems. *J. Magn. Magn. Mater.*, 310:1187, 2007.

[3] Jan Almlof, Thomas H. Fischer, Paul G. Gassman, Abhik Ghosh, and Marco Haeser. Electron correlation in tetrapyrroles: ab initio calculations on porphyrin and the tautomers of chlorin. *J. Phys. Chem.*, 97:10964, 1993.

[4] P. W. Anderson. The resonating valence bond state in La_2CuO_4 and superconductivity. *Science*, 235:1196, 1987.

[5] K. Andersson, P.-Å. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski. Second-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.*, 94:5483, 1990.

[6] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu. Introduction of *n*-electron valence states for multireference perturbation theory. *J. Chem. Phys.*, 114:10252, 2001.

[7] C. Angeli, R. Cimiraglia, and J.-P. Malrieu. *J. Chem. Phys.*, 117:9138, 2002.

[8] M. Bajdich, L. Mitas, G. Drobný, L. K. Wagner, and K. E. Schmidt. Pfaffian pairing wave functions in electronic-structure quantum Monte Carlo simulations. *Phys. Rev. Lett.*, 96:130201, 2006.

[9] M. Bajdich, L. Mitas, L. K. Wagner, and K. E. Schmidt. Pfaffian pairing and backflow wavefunctions for electronic structure quantum Monte Carlo methods. *Phys. Rev. B.*, 77:115112, 2008.

[10] Ajit Banerjee and Jack Simons. *Int. J. Quantum. Chem.*, 14:207, 1981.

[11] Ajit Banerjee and Jack Simons. *J. Chem. Phys.*, 76:4548, 1982.

[12] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Theory of superconductivity. *Phys. Rev.*, 108:1175, 1957.

[13] R. J. Bartlett, S. A. Kucharski, and J. Noga. *Chem. Phys. Lett.*, 155:133, 1989.

[14] R. J. Bartlett, S. A. Kucharski, J. Noga, J. D. Watts, and G. W. Trucks. In U. Kaldor, editor, *Many-Body Methods in Quantum Chemistry*, page 125. Springer, Berlin, 1989.

[15] R. J. Bartlett and M. Musial. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.*, 79(1):291–352, 2007.

[16] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.

[17] Somendra M. Bhattacharjee. Exact matrix representation of the RVB wavefunction. *Z. Phys. B*, 82:323, 1991.

[18] J. P. Bouchand and C. Lhuillier. New microscopic description of liquid ^{3}He. *Z. Phys. B: Condens. Matter*, 75:283, 1989.

[19] R. J. Buenker and S. D. Peyerimhoff. Individualized configuration selection in ci calculations with subsequent energy extrapolation. *Theor. Chim. Acta*, 35:33, 1974.

[20] Michele Casula, Claudio Attaccalite, and Sandro Sorella. Correlated geminal wave function for molecules: an efficient resonating valence bond approach. *J. Chem. Phys.*, 121:7110, 2004.

[21] Michele Casula and Sandro Sorella. Geminal wave functions with Jastrow correlation: a first application to atoms. *J. Chem. Phys.*, 119:6500, 2003.

[22] Hitesh J. Changlani, Jesse M. Kinder, Cyrus J. Umrigar, and Garnet Kin-Lic Chan. Approximating strongly correlated wave functions with correlator product states. *Phys. Rev. B.*, 80:245116, 2009.

[23] L. Cohen and C. Frishberg. *Phys. Rev. A.*, 13:927, 1976.

[24] A. J. Coleman. Structure of fermion density matrices. II. Antisymmetrized geminal powers. *J. Math. Phys.*, 6:1425, 1965.

[25] F. Colmenero and C. Valdemoro. *Phys. Rev. A.*, 47:979, 1993.

[26] Christopher J. Cramer, Armagan Kinal, Marta Włoch, Piotr Piecuch, and Laura Gagliardi. Theoretical models on the Cu_2O_2 torture track: Mechanistic implications for oxytyrosinase and small-molecule analogues. *J. Phys. Chem. A*, 110:1991, 2006.

[27] T. Daniel Crawford, C. David Sherrill, Edward F. Valeev, Justin T. Fermann, Rollin A. King, Matthew L. Leininger, Shawn T. Brown, Curtis L. Janssen, Edward T. Seidl, Joseph P. Kenny, and Wesley D. Allen. Psi3: An open-source ab initio electronic structure package. *J. Comput. Chem.*, 28:1610, 2007.

[28] E. R. Davidson and D. W. Silver. *Chem. Phys. Lett.*, 52:403, 1977.

[29] Thom H. Dunning. Gaussian basis functions for use in molecular calculations. i. contraction of (9s5p) atomic basis sets for the first-row atoms. *J. Chem. Phys.*, 53:2823, 1970.

[30] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90:1007, 1989.

[31] Kenneth G. Dyall. *J. Chem. Phys.*, 102:4909, 1995.

[32] Michaela Flock and Kristine Pierloot. Theoretical study of the interconversion of O_2-binding dicopper complexes. *J. Phys. Chem. A*, 103:95, 1999.

[33] Niclas Forsberg and P.-Å. Malmqvist. *Chem. Phys. Lett.*, 274:196, 1997.

[34] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.*, 73:33, 2001.

[35] K. F. Freed. In U. Kaldor, editor, *Many-Body Methods in Quantum Chemistry*, page 1. Springer, Berlin, 1989.

[36] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala,

K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

[37] R. J. Gdanitz and R. Ahlrichs. *Chem. Phys. Lett.*, 143:413, 1988.

[38] J. Geertsen, M. Rittby, and R. J. Bartlett. *Chem. Phys. Lett.*, 164:57, 1989.

[39] Martin Gouterman and Gamal-Eddin Khalil. Porphyrin free base phosphorescence. *J. Mol. Spectrosc.*, 53:88, 1974.

[40] P. C. Hariharan and J. A. Pople. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoret. chim. Acta*, 28:213, 1973.

[41] Frank E. Harris. Cumulant-based approximations to reduced density matrices. *Int. J. Quantum Chem.*, 90:105, 2002.

[42] W. J. Hehre, R. Ditchfield, and J. A. Pople. Selfconsistent molecular orbital methods. xii. further extensions of gaussiantype basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, 56:2257, 1972.

[43] John M. Herbert. Magnitude and significance of the higher-order reduced density matrix cumulants. *Int. J. Quantum Chem.*, 107:703, 2006.

[44] K. Hirao. Multireference Møller-Plesset method. *Chem. Phys. Lett.*, 190:374, 1992.

[45] M. R. Hoffmann. Third-order complete active space self-consistent field based generalized Van Vleck perturbation theory. *Chem. Phys. Lett.*, 210(1-3):193, 1993.

[46] M. R. Hoffmann and J. Simons. *J. Chem. Phys.*, 88:993, 1988.

[47] A. C. Hurley, John Lennard-Jones, and J. A. Pople. The molecular orbital theory of chemical valency. XVI. A theory of paired-electrons in polyatomic molecules. *Proc. R. Soc. London, Ser. A*, 220:446, 1953.

[48] Robert Jastrow. Many-body problem with strong forces. *Phys. Rev.*, 98:1479, 1955.

[49] B. Kirtman. *J. Chem. Phys.*, 75:798, 1981.

[50] W. Kutzelnigg. *Chem. Phys. Lett.*, 64:383, 1979.

[51] W. Kutzelnigg. *J. Chem. Phys.*, 77:3081, 1982.

[52] W. Kutzelnigg. *J. Chem. Phys.*, 80:822, 1984.

[53] Werner Kutzelnigg and Dabashis Mukherjee. Normal order and extended Wick theorem for a multiconfiguration reference wave function. *J. Chem. Phys.*, 107:432, 1997.

[54] Werner Kutzelnigg and Dabashis Mukherjee. Cumulant expansion of the reduced density matrices. *J. Chem. Phys.*, 110:2800, 1999.

[55] S. R. Langhoff and E. R. Davidson. *Int. J. Quantum Chem.*, 8:61, 1974.

[56] Helena Larsen, Jeppe Olsen, Poul Jørgensen, and Ove Christiansen. Full configuration interaction benchmarking of coupled-cluster models for the lowest singlet energy surfaces of N_2. *J. Chem. Phys.*, 113:6677, 2000.

[57] Konrad H Marti, Bela Bauer, Markus Reiher, Matthias Troyer, and Frank Verstraete. Complete-graph tensor network states: A new fermionic wave function ansatz for molecules. *New J. Phys.*, 12:103008, 2010.

[58] D. A. Mazziotti. *Phys. Rev. A.*, 57:4219, 1998.

[59] D. A. Mazziotti. *Phys. Rev. Lett.*, 97:143002, 2006.

[60] W. L. McMillan. Ground state of liquid He^4. *Phys. Rev.*, 138:A442, 1965.

[61] F Mezzacapo, N Schuch, M Boninsegni, and J I Cirac. Ground-state properties of quantum many-body systems: entangled-plaquette states and variational Monte Carlo. *New J. Phys.*, 11:083026, 2009.

[62] Fabio Mezzacapo and J. Ignacio Cirac. Ground-state properties of the spin-1/2 antiferromagnetic Heisenberg model on the triangular lattice: A variational study based on entangled-plaquette states. *New J. Phys.*, 12:103039, 2010.

[63] D. Mukherjee. In E. Schachinger, editor, *Recent Progress in Many-Body Theories*, volume 4, page 127. Plenum, New York, 1995.

[64] D. Mukherjee. *Chem. Phys. Lett.*, 274:561, 1997.

[65] D. Mukherjee and W. Kutzelnigg. Irreducible Brillouin conditions and contracted Schrödinger equations for n-electron systems. i. the equations satisfied by the density cumulants. *J. Chem. Phys.*, 114:2047, 2001.

[66] H. Nakatsuji. *Phys. Rev. A.*, 14:41, 1976.

[67] H. Nakatsuji. *Chem. Phys. Lett.*, 59:362, 1978.

[68] Hiroshi Nakatsuji and Koji Yasuda. *Phys. Rev. Lett.*, 76:1039, 1996.

[69] Eric Neuscamman, Takeshi Yanai, and Garnet K.-L. Chan. Quadratic canonical transformation theory and higher order density matrices. *J. Chem. Phys.*, 130:124102, 2009.

[70] Eric Neuscamman, Takeshi Yanai, and Garnet K.-L. Chan. Strongly contracted canonical transformation theory. *J. Chem. Phys.*, 132:024106, 2010.

[71] M. P. Nightingale and C. J. Umrigar. *Quantum Monte Carlo Methods in Physics and Chemistry*. NATO ASI Ser. C 525. Kluwer Academic, Boston, 1999.

[72] S. Pal. *Theor. Chim. Acta*, 66:207, 1984.

[73] S. Pal, M. D. Prasad, and D. Mukherjee. *Theor. Chim. Acta*, 62:523, 1983.

[74] J. Paldus and X. Li. *Adv. Chem. Phys.*, 110:1, 1999.

[75] János Pipek and Paul G. Mezey. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.*, 90:4916, 1989.

[76] R. Pou-Amerigo, M. Merchan, I. Nebot-Gil, P.O. Widmark, and B. Roos. *Theor. Chim. Acta*, 92:149, 1995.

[77] Michal F. Rode and H.-J. Werner. Ab initio study of the O_2 binding in dicopper complexes. *Theor. Chem. Acc.*, 114:309, 2005.

[78] Anders W. Sandvik. Finite-size scaling of the ground-state parameters of the two-dimensional Heisenberg model. *Phys. Rev. B*, 56:11678, 1997.

[79] Luis Serrano-Andrés, Manuela Merchán, Mercedes Rubio, and Björn O. Roos. Interpretation of the electronic absorption spectrum of free base porphin by using multiconfigurational second-order perturbation theory. *Chem. Phys. Lett.*, 295:195, 1998.

[80] Per E. M. Siegbahn. Generalizations of the direct ci method based on the graphical unitary group approach. ii. single and double replacements from any set of reference configurations. *J. Chem. Phys.*, 72:1647, 1980.

[81] A. Szabo and N.S. Ostlund. *Modern quantum chemistry*. McGraw-Hill New York, 1989.

[82] P. G. Szalay and R. J. Bartlett. *Chem. Phys. Lett.*, 214:481, 1993.

[83] A. G. Taube and R. J. Bartlett. *Int. J. Quantum Chem.*, 106:3393, 2006.

[84] J. D. Watts, G. W. Trucks, and R. J. Bartlett. *Chem. Phys. Lett.*, 157:359, 1989.

[85] H.-J. Werner and P. J. Knowles. An efficient internally contracted multiconfiguration-reference configuration interaction method. *J. Chem. Phys.*, 89:5803, 1988.

[86] S. R. White. Numerical canonical transformation approach to quantum many-body problems. *J. Chem. Phys*, 117:7472, 2002.

[87] P.O. Widmark, P.A. Malmqvist, and B. Roos. *Theor. Chim. Acta*, 77:291, 1990.

[88] Takeshi Yanai and Garnet K.-L. Chan. Canonical transformation theory for multireference problems. *J. Chem. Phys.*, 124:194106, 2006.

[89] Takeshi Yanai and Garnet K.-L. Chan. Canonical transformation theory from extended normal ordering. *J. Chem. Phys.*, 127:104107, 2007.

[90] Takeshi Yanai, Yuki Kurashige, Eric Neuscamman, and Garnet K.-L. Chan. Multireference quantum chemistry through a joint density matrix renormalization group and canonical transformation theory. *J. Chem. Phys.*, 132:024105, 2010.