Title
A Variational Principle for Modeling Electronic Excitations in Gas and Condensed Phase

Author
Zhao, Luning

Publication Date
2019

Peer reviewed|Thesis/dissertation

A Variational Principle for Modeling Electronic Excitations in Gas and Condensed Phase

by

Luning Zhao

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Assistant Professor Eric Neuscamman, Chair
Professor Martin Head-Gordon
Professor Birgitta Whaley
Associate Professor Lin Lin

Summer 2019

A Variational Principle for Modeling Electronic Excitations in Gas and Condensed Phase

Abstract

A Variational Principle for Modeling Electronic Excitations in Gas and Condensed Phase

by

Luning Zhao

Doctor of Philosophy in Chemistry

University of California, Berkeley

Assistant Professor Eric Neuscamman, Chair

Accurate modeling of electronic excited states is one of the most important and challenging problems in electronic structure theory. This thesis focuses on a recently developed excited state variational principle and its applications in gas and condensed phase. In contrast to the widely used excited states method such as linear response (LR) and many-body perturbation theory (MBPT), which find excited states by perturbing around the ground state wave function or a zeroth order particle-hole excitation picture, the new excited state variational principle directly targets excited states with the full flexibility of an approximate wave function ansatz. Due to its non-perturbative nature, this method offers balanced and systematically improvable descriptions to excited states. We will also discuss the efficient implementation of the new excited state variational principle through variational Monte Carlo and the Linear Method optimization algorithm.

The new excited state variational principle is applied to predict both the excitation energies of low lying excited states in small molecules and optical gaps in solids. In molecules, the new method yields order-of-magnitude of improvements over the state-of-art excited state methods based on LR theory in double excitations. In solids, not only is the new method demonstrated to be more accurate than the commonly used MBPT method, but it could also be used to analyze and provide insights into MBPT.

In order to further extend the method's applicability, we introduce a modified optimization method that addresses a fatal memory bottleneck in the original algorithm. With only minor lose in accuracy, the modified algorithm reduces the required memory per parallel process from tens of gigabytes to hundreds of megabytes. With the aid of the new optimization method, we show that the new excited state variational principle could systematically converge the excitation energy in a strongly correlated, Mott-insulating hydrogen ring with respect to increasing flexibility in the wave function ansatzes.

To Siyao Jia, for all her love and support

Contents

List of Figures

List of Tables

Acknowledgements

First of all, I would like to thank my advisor Eric Neuscamman for his excellent mentorship. As Eric's first graduate student, Eric put a lot of efforts into me. He taught me how to identify interesting problems in chemistry, to think creatively as a chemist and a theorist, and to solve these problems using mathematical tools and computational approaches. He has always been supportive to me both in my research work, my personal life, and my applications for postdoc positions. It is truly beyond word to describe how thankful I am to him.

Besides Eric, I would also like to thank professor Martin Head-Gordon. Martin was my Chem 221A instructor and also serves as the chair of my qualifying exam committee and a member of my thesis committee. I truly appreciate the valuable advice he gave me both in terms of my research and my choices for postdoc positions, and his efforts to write me reference letters when I applied for summer internships and postdoc positions.

I am also very grateful to my collaborators in the Center for Predictive Simulation of Functional Materials. I want to thank Paul Kent for organizing the center and for many insightful scientific discussions in my work of optical gaps, and his efforts of advertising my work and writing me reference letters when I applied for postdoc positions last year. I want to thank Ye Luo for helping me understanding and developing QMCPACK. I also thank Miguel Morales-Silva for hosting me when I did my summer internship at Lawrence Livermore National Laboratory.

Berkeley is an amazing and inspirational place to do science, and many of the inspirations come from the talented and knowledgeable teachers. During my PhD, I learned a lot from the classes I took and the instructors of these classes. Therefore I would like to thank Martin Head-Gordon, David Chandler, William H. Miller, Philip Geissler, Lin Lin, James Demmel, Annant Sahai, Jennifer Listgarten, and Sergev Levine for their motivative and insightful teaching. I would also want to thank Leslie Dietterick and Laurie Mason for addressing many administrative issues.

The students and postdocs in Neuscamman group also provide me with a lot of help. I would like to thank Brett Van Der Goetz, Haochuan (Harry) Wei, Jacqueline Shea, Sergio Pineda Flores, Leon Otis, Nick Blunt, Peter Walters, and Lan Tran for countless insightful discussions. I also want to thank Joonho Lee for helping me with many issues that I run into when running QChem.

Finally, I would like to thank my family, especially my parents who raised me and support

me to pursue my PhD in U.S. I thank my partner and soul mate Siyao Jia for her constant love, support, and companionship during our time spent in Berkeley. She is always there for me in joy and frustration, and this journey would not be complete without her.

Chapter 1

Introduction

1.1 Schrödinger's Equation

In quantum mechanics, the behaviors of electrons and atoms are governed by Schrödinger's equation (SE).

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

in which $\hat{H}$ is the Hamiltonian operator, $|\Psi\rangle$ is the "state vector", and E is the state energy.

As one can see, the SE is an eigenvalue problem. The eigenvectors $|\Psi\rangle$ define the stationary state of the system: states that do not evolve with time. For a general state vector that is not an eigenvector of the Hamiltonian, its time evolution is determined by the time-dependent Schrödinger equation (TDSE),

$$i\hbar\frac{\partial\,|\Psi(t)\rangle}{\partial t} = \hat{H}\,|\Psi(t)\rangle \tag{1.2}$$

In this thesis, we only only focus on the time-independent SE and its solutions.

Consider the systems that contain only atoms and electrons, such as molecules and solids, and the Hamiltonian operator is,

$$\hat{H} = -\sum_{i}^{n}\frac{1}{2}\nabla_i^2 - \sum_{A}^{M}\frac{1}{2m_A}\nabla_A^2 - \sum_{i}^{n}\sum_{A}^{M}\frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i>j}^{n}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{A>B}^{M}\frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \tag{1.3}$$

in which i and $\mathbf{r}_i$ correspond to the index and position of electrons, and A, m_A, Z_A, $\mathbf{R}_A$ correspond to the index, mass, charge, and positions of nuclei.

As we can see from Equation 1.3, the Hamiltonian contains five terms. The first two terms are the kinetic energy operators of the electron and nucleus, respectively. The third term is the electron-nucleus attraction energy, and the last two terms are the electron-electron and nucleus-nucleus repulsion energy. In reality, since the nuclei are much heavier than the electrons, their positions are considered as "fixed" so that the nucleus kinetic energy is neglected from the Hamiltonian. The nucleus-nucleus repulsion energy then becomes a constant. This is the so-called "Born-Oppenheimer"(BO) approximation[1].

Within BO, we can write down the electronic part of the Hamiltonian,

$$\hat{H}_{el} = -\sum_{A}^{M} \frac{1}{2m_A}\nabla_A^2 - \sum_{i}^{n}\sum_{A}^{M} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i>j}^{n} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{1.4}$$

and the electronic SE,

$$\hat{H}_{el}\,|\Psi_e\rangle = E_{el}\,|\Psi_{el}\rangle \tag{1.5}$$

Then the total energy is the sum of E_{el} and the nucleus-nucleus repulsion energy, which is a constant with given nucleus positions.

$$E^{BO} = E_{el} + \sum_{A>B}^{M} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \tag{1.6}$$

Although the BO approximation is a sound approximation to make in plenty of situations, there exist cases that it becomes invalid. The basic assumption of BO is that the electrons move much faster than the nuclei so that they can adjust instantaneously to the change of nuclear positions. However, in collision experiments, the nuclei could gain a huge amount of kinetic energy so that the electrons do not have enough time to adjust to nuclear movements and BO breaks down. In other situations in which the movements of nucleus and electron become strongly coupled to each other, such as in conical intersections, separating their movements as done by BO could lead to nonphysical predictions. In this thesis, we will work within the BO approximation.

1.2 Second Quantized Hamiltonian

The electronic Hamiltonian introduced in Equation 1.4 is defined in real space: the space of electron Cartesian positions. The basis for such space are delta-functions in $3N$ space.

$$f_{\mathbf{r}_1^i \ldots \mathbf{r}_n^i}(\mathbf{r}_1 \ldots \mathbf{r}_N) = \delta(\mathbf{r}_i - \mathbf{r}_1^i \ldots \mathbf{r}_N - \mathbf{r}_N^i) \tag{1.7}$$

As one can see, the delta-function basis is of infinite dimension. Since working directly in a infinite dimension space is difficult, it is more convenient to project the Hamiltonian into a finite basis.

First we define a finite, orthonormal one-electron basis functions (orbitals) $\phi_p(\mathbf{x}_i)$, in which $p = 1 \ldots M$ and M is the total number of basis functions. $\mathbf{x}_i$ represents collectively the spatial coordinates $\mathbf{r}_i$ and the spin coordinate σ of the electron i. We can then define a normalized Slater determinant as,

$$|\phi_{p_1}\phi_{p_2}\ldots\phi_{p_N}| = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_{p_1}(\mathbf{x}_1) & \phi_{p_2}(\mathbf{x}_1) & \cdots & \phi_{p_N}(\mathbf{x}_1) \\ \phi_{p_1}(\mathbf{x}_2) & \phi_{p_2}(\mathbf{x}_2) & \cdots & \phi_{p_N}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{p_1}(\mathbf{x}_N) & \phi_{p_2}(\mathbf{x}_N) & \cdots & \phi_{p_N}(\mathbf{x}_N) \end{vmatrix} \tag{1.8}$$

As one can see, a Slater determinant is a anti-symmetric product of the orbitals. The anti-symmetry is a reflection of the particle symmetry of fermions, in which a fermonic wave function would change by a sign if two particles in it are swapped.

From the Slater determinant we can further define the basis of occupation vectors,

$$|\vec{n}\rangle = |n_1, n_2, ..., n_M\rangle \tag{1.9}$$

in which $n_p = 1$ if the orbital ϕ_p is occupied and $n_p = 0$ otherwise. If the orbitals are orthonormal, then one could prove that such a basis is also orthonormal.

$$\left\langle \vec{k}|\vec{m} \right\rangle = \delta_{\vec{k},\vec{m}} \tag{1.10}$$

The Hamiltonian in the occupation number basis is[1],

$$\hat{H} = \sum_{pq}^{M} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs}^{M} g_{pqrs} a_p^\dagger a_r^\dagger a_s a_q + h_{nuc} \tag{1.11}$$

in which h_{nuc} is the nucleus-nucleus repulsion energy. h_{pq} and g_{pqrs} are the one-electron and two-electron integrals respectively[1].

$$\begin{aligned} h_{pq} &= -\frac{1}{2} \int \phi_p^*(\mathbf{x}) \nabla_i^2 \phi_q(\mathbf{x}) d\mathbf{x} - \int \phi_p^*(\mathbf{x}) \sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \phi_q(\mathbf{x}) d\mathbf{x} \\ g_{pqrs} &= \int \int \frac{\phi_p^*(\mathbf{x}_1) \phi_r^*(\mathbf{x}_2) \phi_q(\mathbf{x}_1) \phi_s(\mathbf{x}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{x}_1 d\mathbf{x}_2 \end{aligned} \tag{1.12}$$

The creation (annihilation) operator $a_p^\dagger$ (a_p) creates (annihilates) one electron in orbital p.

1.3 Exact and Approximate Solutions to the Schrödinger Equation

Typically, the exact solutions to SE could only be obtained for one-electron systems, such as the hydrogen atom. It is because of the electron-electron repulsion term in the Hamiltonian that makes SE extremely hard to solve, since the motions of electrons are coupled to each other and separation of variables cannot be performed. Therefore, one needs to find approximate solutions. In electronic structure theory, there are many different methods to find approximate solutions to SE. These methods can be roughly divided into two categories: wave function based methods and density functional based methods. The focus of this thesis will be the wave function based methods.

Hartree-Fock

In wave function based methods, one first guesses a wave function form: an ansatz. The ansatz contains unknown parameters and one needs to optimize this set of parameters such that the optimized ansatz becomes as close to the exact SE solution as possible. The simplest wave function ansatz that preserves the fermonic symmetry is a single Slater determinant, in which the unknown parameters are the one-electron orbitals. By optimizing these orbitals self-consistently, one arrives at Hartree-Fock (HF) theory[2].

The HF wave function can be written as,

$$|\Psi_{\mathrm{HF}}\rangle = \exp(-\hat{\kappa})\,|\Phi\rangle \tag{1.13}$$

in which $|\Phi\rangle$ is a general single determinant. $\exp(-\hat{\kappa})$ is the unitary orbital rotation operator in the following form,

$$\hat{\kappa} = \sum_{pq} \kappa_{pq} a_p^\dagger a_q \tag{1.14}$$

in which κ is an anti-symmetric matrix and the summation is over all orbital pairs.

Although the total energy predicted by HF is very close to the exact energy (the HF energy is usually 99% of the exact energy), the missing 1% of the energy is important to achieve quantitative accuracy. Therefore HF is only referred to as qualitatively correct. The electrons in the HF theory are uncorrelated. In other words, they move in a "mean-field" potential created by other electrons without knowing the explicit positions of other electrons. Therefore, the correlation energy, defined as the energy difference between HF energy and the exact energy, is missing in HF.

The missing correlation energy in HF has severe consequences. Electron correlation can be further divided into "weak" and "strong" correlations. Weak correlation arises due to the instantaneous electron-electron repulsion. Systems whose correlation energy is dominated by weak correlation are referred to as weakly correlated systems.

For weakly correlated systems, the HF determinant is usually the dominant piece of the exact wave function. This is because the electron-electron repulsion is small in these systems and it is valid to use a mean-field treatment. For example, H_2 at equilibrium geometry is a weakly correlated system and a missing of such effects in HF would affect its prediction of bond length. However, as the wave function of a weakly correlated system is dominated by the HF determinant, HF is still qualitatively correct.

If a system has both strong and weak correlations or is dominated by strong correlations, it is called a strongly correlated system. For example, if one stretches H_2 molecule, it gradually transforms from a weakly correlated system to a strongly correlated system. At the stretched geometry, the exact wave function of the molecule should be a linear combination of two determinants and using a single-determinant HF wave function becomes inadequate until it reaches dissociation limit. In summary, for strongly correlated systems, the HF determinant stops being the dominated piece of the wave function. Therefore, HF is not even qualitatively accurate in such cases.

Configuration Interaction

Due to the drawbacks of HF, correlated wave function ansatzes are needed to obtain quantitatively accurate predictions. The most straightforward way to capture the missing correlations in a single determinant wave function is to apply a multi-determinant expansion (configuration interaction) (CI)[2],

$$|\Psi_{CI}\rangle = |\Psi_0\rangle + \sum_{ia} C_i^a |\Psi_i^a\rangle + \sum_{ijab} C_{ij}^{ab} |\Psi_{ij}^{ab}\rangle + ... \tag{1.15}$$

in which $|\Psi_0\rangle$ is the HF determinant, and $|\Psi_i^a\rangle$, $|\Psi_{ij}^{ab}\rangle$ are excited determinants in which the ith and jth occupied orbitals in $|\Psi_0\rangle$ are replaced by ath and bth unoccupied orbitals.

If one includes all possible excitations in the CI expansion, one could in principle obtain the exact solution to SE in the finite orbital basis. This is called full Configuration Interaction (FCI)[3, 4]. However, there are a combinatorial number of terms in the FCI expansion, and so the computational cost scales exponentially as the number of electrons and orbitals gets larger. Therefore FCI can only be applied to small systems.

In order to design a tractable CI method, one usually truncate the FCI expansion at certain excitation levels. Depending on the specific truncation level, one obtains different truncated CI[5] method such as CIS, CISD, CISDT, etc. However, these truncated CI methods are in general not size-consistent: the quality of the wave function deteriorates with system size, making them less useful in large scale systems.

Complete Active Space Self-Consistent Field Theory

Although FCI is not affordable in most systems, sometimes is is possible to identify a small subset of orbitals and electrons that are chemically important, especially in strongly correlated systems. For example, in stretched H_2, it is the bonding and anti-bonding orbitals that are responsible for strong correlation effects. By doing a FCI in this subset of orbitals and electrons, the cost is kept affordable. This is the central idea of the complete active space self-consistent field theory (CASSCF)[1]. The wave function can be written as,

$$|\Psi_{\text{CASSCF}}\rangle = \exp(-\hat{\kappa}) \sum_i C_i |i\rangle \tag{1.16}$$

in which C_i are the configuration interaction coefficients and the operator $\exp(-\hat{\kappa})$ carries out orbital rotations in the same way as for HF wave function. The sum is over all possible Slater determinants $|i\rangle$ that can be formed by arranging the active electrons in the active orbitals.

Unlike HF, CASSCF is capable of describing strongly correlated systems. However, it is not a black-box method. Choosing a correct active space requires considerable amounts of chemical knowledge about the system and is far from trivial. In addition, the cost of CASSCF also scales exponentially with the size of active space so that it is only applicable with a small active space.

Coupled Cluster

One of the most commonly used correlated wave function ansatz is the coupled cluster (CC) wave function. The CC wave function is written as,

$$\begin{aligned} |\Psi_{\mathrm{CC}}\rangle &= \exp(\hat{T})\,|\Psi_0\rangle \\ \hat{T} &= \sum_{ia} t_i^a a_a^\dagger a_i + \sum_{i>j,a>b} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_i a_j + ... \end{aligned} \tag{1.17}$$

in which the $|\Psi_0\rangle$ is the reference determinant (usually taken as the HF determinant) and $\hat{T}$ is the cluster operator.

As with truncated CI, one needs to truncate the cluster operator to certain excitation level to make the method computationally tractable. Truncating the cluster operator at doubles level usually achieves best balance between cost and accuracy. The resulting coupled cluster singles and doubles (CCSD) ansatz, when coupled with perturbative estimate of the effects of triples[1], is considered as one of the most accurate approximate wave function ansatz.

Unlike truncated CI, CC is size consistent. For two non-interacting systems A and B, the system's overall wave function $|\Psi_{A+B}\rangle$ factorizes to $|\Psi_A\rangle\,|\Psi_B\rangle$. Owing to its exponential form, such a property is satisfied by CC. Therefore, the quality of the CC wave function does not deteriorate with increasing system size, making the method suitable to describe large systems, assuming the computational cost is affordable.

1.4 Variational Principle

As discussed before, merely having an approximate wave function ansatz is not enough. One also needs to optimize its parameters so that it closely resembles the exact wave function. One way to optimize the wave function parameters is to minimize the energy functional,

$$E[\Psi] = \frac{\left\langle \Psi|\hat{H}|\Psi \right\rangle}{\langle \Psi|\Psi\rangle} \tag{1.18}$$

as we will prove in the next chapter, the energy functional is lower bounded by the exact ground state energy: the lowest eigenvalue of the Hamiltonian. Therefore the energy functional provides a "driving force" to optimize the wave function parameters towards the ground state. The variational principle is the foundation of both wave function based methods, such as HF, CI, matrix product state (MPS)[6], and variational Monte Carlo (VMC)[7], and also density functional theory (DFT)[8].

We also need to note that the CC wave function is very inefficient to optimize to with the variational principle, as there is no known way to evaluate the energy functional with a CC wave function that does not require exponentially growing cost. Instead, one usually

solves the projected SE for the cluster amplitudes in CCSD wave function.

$$\begin{aligned} E &= \left\langle \Psi_0 | \exp(-\hat{T}) \hat{H} \exp(\hat{T}) | \Psi_0 \right\rangle \\ 0 &= \left\langle \Psi_i^a | \exp(-\hat{T}) \hat{H} \exp(\hat{T}) | \Psi_0 \right\rangle \\ 0 &= \left\langle \Psi_{ij}^{ab} | \exp(-\hat{T}) \hat{H} \exp(\hat{T}) | \Psi_0 \right\rangle \end{aligned} \tag{1.19}$$

One should note even though CC methods do not use the variational principle directly, it also needs to take the variationally optimized HF orbitals as its inputs for both the HF determinant $|\Psi_0\rangle$ and excited determinants $|\Psi_i^a\rangle$, $\left|\Psi_{ij}^{ab}\right\rangle$.

1.5 Variational Monte Carlo

The variational Monte Carlo algorithm is a stochastic algorithm used to evaluate and optimize the energy function $E[\Psi]$. To see how this is performed, we first insert the resolution of identity in the numerator of Equation 1.18,

$$E[\Psi] = \frac{\left\langle \Psi | \hat{H} | \Psi \right\rangle}{\langle \Psi | \Psi \rangle} = \sum_m \frac{\langle \Psi | m \rangle \left\langle m | \hat{H} | \Psi \right\rangle}{\langle \Psi | \Psi \rangle} \tag{1.20}$$

in which we have used the property of the complete basis,

$$\hat{1} = \sum_m |m\rangle \langle m| \tag{1.21}$$

in which $|m\rangle$ loops over every electron configurations.

Multiplying and dividing the above equation by $\langle m|\Psi\rangle$, one obtains,

$$E[\Psi] = \sum_m \frac{|\langle m|\Psi\rangle|^2}{\langle \Psi|\Psi\rangle} \frac{\left\langle m|\hat{H}|\Psi\right\rangle}{\langle m|\Psi\rangle} \tag{1.22}$$

as one can see, now the energy expectation value becomes a weighted sum, and the weights are from the probability distribution $|\langle m|\Psi\rangle|^2 / \langle \Psi|\Psi\rangle$.

The expectation value could be evaluated by a finite Monte Carlo sampling, 1.18,

$$E[\Psi] \approx \frac{1}{N_{MC}} \sum_{m \in \xi} \frac{\left\langle m|\hat{H}|\Psi\right\rangle}{\langle m|\Psi\rangle} = \frac{1}{N_{MC}} \sum_{m \in \xi} E_L(m) \tag{1.23}$$

in which $|m\rangle$ is sampled from $|\langle m|\Psi\rangle|^2 / \langle \Psi|\Psi\rangle$. N_{MC} is the number of samples and ξ is the sample set.

The VMC algorithm evaluates the energy functional stochastically. A wave function ansatz is compatible with VMC as long as we can evaluate the local energy $E_L(m)$ efficiently. The wave function ansatzes that are commonly used in VMC are Jastrow based ansatzes. These wave functions are of the form,

$$|\Psi\rangle = \exp(\hat{J})\,|\Phi\rangle \tag{1.24}$$

in which the Jastrow factor $\exp(\hat{J})$ is able to capture correlations that are missing in the fermonic function $|\Phi\rangle$.

At first glance, the Jastrow factor is very similar to the cluster operator in CC wave functions. However, there is a major difference between these two. The Jastrow factor is defined in a local basis, such as electron positions in real space and atomic orbital basis. In contrast, the cluster operator is often defined in delocalized, molecular orbital basis.

Besides capturing correlations, another major usage of the Jastrow factor is to enforce correct cusp conditions in the vicinity of an electron and a nucleus. As an electron is close to another electron or a nucleus, its potential energy diverges. Therefore, the kinetic energy must compensate for the diverging potential energy so that the total energy is still finite. It is shown that the exact wave function must satisfy the following cusp conditions[1].

$$\begin{aligned} \lim_{r_i \to 0} \left(\frac{\partial \Psi}{\partial r_i} \right)_{\text{ave}} &= -Z\Psi(r_i = 0) \\ \lim_{r_{ij} \to 0} \left(\frac{\partial \Psi}{\partial r_{ij}} \right)_{\text{ave}} &= \frac{1}{2}\Psi(r_{ij} = 0) \end{aligned} \tag{1.25}$$

in which the "ave" subscript denotes that the radial derivative is taken while performing a spherical averaging over all directions.

The electron-electron cusp condition could be easily built into Jastrow factors[7], and the electron-nucleus cusp condition is usually satisfied by using Slater orbitals[2] or augmented Gaussian orbitals[9]. The Jastrow based wave functions are generally hard to evaluate analytically by Equation 1.18, however, their local energies are usually trivial to compute so that they admit efficient implementations via VMC.

Another useful quantity that could be obtained via VMC is the variance of the local energy,

$$\sigma^2 = \frac{1}{N_{MC}} \sum_{m \in \xi} (E_L(m) - \bar{E})^2 \tag{1.26}$$

in which $\bar{E}$ is the average of the local energy. If the wave function $|\Psi\rangle$ is exact, then one could show that each local energy, no matter the configuration $|m\rangle$, will be the eigen-energy and the energy variance will be zero. Therefore, the energy variance offers us a way to tell how close the wave function is from an exact eigenstate. As we shall see later in Chapter 4, by matching the energy variance of two different states, one could obtain balanced descriptions for properties that depends on both states.

1.6 Electronic Excitations

Solving the electronic SE in full is equivalent to solving an eigenvalue problem. Suppose dimension of the electronic Hamiltonian is M, one then could obtain M eigenstates. The eigenstate with the lowest energy is referred to as the electronically ground state, and the eigenstates with higher energy correspond to electronically excited states.

Electronic excitation is one of the most important phenomena in nature. Governed by the rules of quantum mechanics, electrons in molecules and solids form either discrete energy levels or continuous energy bands separated by "energy gap". Without external perturbations, electrons live in the state with the lowest energy: the electronically ground state at equilibrium. However, once the system absorbs external energy, such as light irradiation, collision with other molecules and atoms, and thermal fluctuations, electrons could be promoted to excited states with higher energies.

The formation of electronically excited states is the foundation of ultraviolet and visible spectroscopy techniques. The peaks in the spectra of molecules and solids correspond to different types of electronic transitions from ground to excited states. Therefore with detailed knowledge of electronic excited states, such as vertical excitation energies, spin multiplicity, and state symmetry, chemists can identify transition states and study the rate of light-initiated chemical reactions with high spectral and time resolution.

Besides spectroscopy, the entire field of photochemistry is built upon electronic excitations. For instance, solar cells[10] convert solar energy to electricity by first forming electronically excited states. By taking advantage of the excited electron and the hole being left behind, a photocatalytic process could be performed to split water into oxygen and hydrogen and to reduce carbon dioxide to methane and carbon monoxide[11]. Given the importance of water splitting and CO_2 reduction in terms of renewable energy and climate change, deep understanding the first step, the excitation of electrons, becomes extremely crucial for people to design new light-harvesting materials.

Even without external perturbation such as light irradiation, electronically excited states can still play vital roles in circumstances where the motion of electrons and nuclei become strongly coupled.[12] In such cases, the BO approximation breaks down. The BO approximation assumes that the electrons respond instantaneously to the change of nuclear configurations since they are much lighter than the nuclei. However, when the atoms have high kinetic energies, such as in collision experiments,[13] the electrons are unable to adjust to the nuclear motion instantaneously, causing "hoppings" of the electrons between multiple electronic states. Another well-known failure of the BO approximation is at conical intersection, where two potential energy surfaces become exactly degenerate. It is widely accepted that conical intersections play vital roles in photochemistry and photobiology[14]. Consequently, gaining knowledge about the electronic excited states would help us understand the mechanisms of these phenomena.

1.7 Current Excited State Methods

Given the importance of electronic excitations, having an accurate theoretical approach to model them is highly desired. In principle, electronically excited states are just interior eigenstates of the electronic Hamiltonian. Hence, the simplest way is to construct the electronic Hamiltonian in the complete basis and solve for its interior eigenstates. Such a method is the aforementioned FCI approach or exact diagonalization (ED)[15]. However, the dimension of the FCI Hamiltonian scales exponentially with respect to the system size, which limits the method to only small systems with 20 electrons or less. Albeit with improvements made by selective CI methods[16], the applicabilities of FCI and its variants in large scale systems are still highly limited, and one needs to apply approximations to describe excited states.

One type of approximate excited state approach is based on CI, including truncated CI as well as the complete active space self-consistent field (CASSCF) method and its CI-based (MRCI)[17] and perturbative (CASPT2)[1] extensions. Excited states can be obtained by either solving the interior eigenstates of the truncated CI Hamiltonian or by performing state-averaged calculations in CASSCF. CASSCF-based methods are among the most robust available and have the advantage of systematic convergence via expansions of the active space, but they suffer from the need for state-averaging and combinatorially growing costs. More recent methods offering systematic convergence include full configuration interaction quantum Monte Carlo (FCI-QMC) [18, 19] and the density matrix renormalization group (DMRG) [20, 21], although these also have combinatorially growing costs in general.

Linear Response / Equation of Motion

Another approach to excited states is to apply linear response (LR) or equation-of-motion (EOM) theory. In LR or EOM, excited states are built using linear combinations of an approximate ground state wave function ansatz's first derivatives. Depending on the specific form of the ground state wave function, LR/EOM methods admit different formulations as well. Taking the HF wave function for example, its first order derivatives with respect to the ith occupied to ath unoccupied orbital rotation parameter ,

$$\frac{\partial \left|\Psi_{\mathrm{HF}}\right\rangle}{\partial \kappa_{ia}} = \left|\Psi_i^a\right\rangle \tag{1.27}$$

one immediately sees that these first derivatives are singly excited determinants. Excited states can be obtained by projecting and diagonalizing the Hamiltonian in the first derivative basis. The LR/EOM approach based on HF wave function is called the configuration interaction singles (CIS) or time-dependent Hartree-Fock (TDHF)[22].

Besides CIS, LR/EOM based methods also include the time-dependent density functional theory (TDDFT),[22] equation-of-motion coupled cluster with singles and doubles (EOM-CCSD),[23] and linear-response DMRG[24], depending on the choice of the ground state wave function. These methods typically have more favorable cost-scalings than those based

on CASSCF, allowing them to reach larger molecules, therefore they are routinely used to study medium to large size molecular systems.

In most cases, we are only interested in the relative properties of the excited states with respect to the ground state, such as the excitation energy and the transition dipole moment of excited states. These quantities depend on the accuracy of both excited state and ground state. Therefore, obtaining a balanced description for these two states is of great importance. However, LR and EOM based methods often suffer from a bias in favor of the ground state.[25] For example, in CIS, the LR space contains the freedom to shape the excitation's orbital, but cannot achieve the second-order effect of relaxing the shapes of other orbitals in the presence of the excitation. This lack of orbital relaxation is perhaps in practice the most common and important source of ground state bias, showing up also in EOM-CCSD in the case of doubly excited states (for which EOM-CCSD has no triples to use to couple in relaxation in the way it can via its doubles for singly excited states). Consequently, the predicted excitation energy are usually overestimated. One notoriously hard problem for LR is the controversial 2^1A_g excited state in *trans*-Butadiene[26, 27]. Such a low-lying excited state contributes significantly to photodynamics, but its double excitation character prevents the applications using low order reference states for LR methods.

The limitations of the LR/EOM based methods are mainly due to their perturbative nature. As we shall show rigorously in Chapter 3, the excited states of EOM/LR are found through the response of the ground state due to a time-dependent perturbation. Hence, it becomes intuitive to picture that when certain excited states are far away from the ground state so that they do not or only partially live in the ground state perturbative space, EOM/LR methods become unreliable. Doubly excited states are notoriously hard to describe by CIS and EOM-CCSD since they do not live in the perturbative space of a single determinant and only partially live in such space of the CCSD wave function.

Many-Body Perturbation Theory and the GW Approximation

Instead of CASSCF based or LR/EOM based methods, many-body perturbation theory (MBPT) in the form of GW[28] and Bethe-Salpeter equation (BSE)[29] are the most successful[30] methods used to study charged and neutral electronic excitations in solids. Properties including quasi-particle band structures, optical spectra, and even dynamical properties such as excited state life time have been predicted accurately with MBPT, especially for *sp* bonded semi-conductors, conventional band insulators, and semi-metals. However, there still remain many materials of great technological interest, especially within the transition metal oxides such as Mott insulators[31], whose low-energy excitations are poorly described by density functional theory (DFT)[8] and MBPT.

Although MBPT does not need rely on input from DFT, some of its most widely used practical incarnations (e.g. G_0W_0) assume a zeroth order picture in which electronic excitations are simple particle-hole transitions between the one-particle eigenstates of Kohn-Sham DFT with transition energies given by differences between these Kohn-Sham orbitals' energies. In this picture, the lowest excited state corresponds to a single open-shell Slater

determinant in which one electron has been promoted from the valence band maximum (VBM) orbital to the conduction band minimum (CBM) orbital. Although the DFT orbital energy difference is known to underestimate the corresponding band gap, [30, 32] this zeroth order picture is nonetheless quite close to reality when solids like diamond and silicon are treated with standard LDA [33] or GGA [34] density functionals. In these situations, the DFT orbitals closely resemble the excited electron and hole states and the orbital energy differences, although not perfect, are close enough to reality that MBPT variants that perturb around them can be quite accurate. [35]

The story can be strikingly different when a solid/functional pairing produces one-particle states that differ significantly from the true electron and hole states and/or the orbital energy differences stray too far from reality. For example, zinc oxide represents a traditionally challenging case for MBPT. It has been shown that the G_0W_0 quasi-particle gap is extremely difficult to converge[36]. Even though the gap predictions could be improved by using hybrid functionals [37, 38], the sensitivity of the results to the fraction of exact exchange and the degree of self-consistency that is sought in the GW equations puts MBPT's reliability in question.

As we see, although the current approaches for excited states have been quite successful in both molecules and solids, there is still room for the development of new methods that are free from the major disadvantages of the current methods. One of the main drawbacks of both LR/EOM and MBPT method is their perturbative nature. In molecules, LR/EOM methods build excited states in the perturbative space of the ground state. In solids, MBPT makes perturbations around a zeroth order picture in which excitations are simple particle-hole transitions. It is the breakdown of the perturbation assumptions that results in the aforementioned challenging examples of these methods.

Excited State Variational Principle

One way of modeling excited states without relying on perturbation theory is through excited state variational principles. This includes the early work done by Messmer[39] for harmonic oscillators and the Harmonic Davidson algorithm applied by Chan and co-workers[40] with DMRG. The underlying idea of excited state variational principle is to optimize a functional of approximate wave function ansatzes whose global minimum corresponds to the excited state of interest. The main of advantage of the excited state variational principle is its non-perturbative nature. In other words, instead of finding excited states by making small changes to the ground state, it tries to target excited states with the full flexibility of the ansatz. Therefore, it should yield more accurate predictions in cases where LR/EOM based methods perform poorly.

The focus of this thesis is a newly developed excited state variational principle[41] and its applications in modeling excited states. The method introduces a functional whose global minimum corresponds to an excited eigenstate.[41] With efficient quantum Monte Carlo (QMC)[7] algorithm to evaluate and optimize it and its independence of boundary conditions, this method is capable of describing excited states in both molecules and solids in a non-

perturbative way. Consequently, it does not suffer from the breakdown of perturbation theory as in LR/EOM and MBPT approaches. Furthermore, although the focus of this thesis is electronically excited states, the applicability of this new excited state variational principle is not limited to electronic excitations. With the ability to target interior eigenstates of arbitrary Hamiltonians, this method could also be applied to study vibrational and rotational excited states.

Targeting excited states with this new excited state variational principle requires user to optimize both linear and nonlinear parameters in approximate wave function ansatzes. Therefore the optimization algorithm is as important as the variational principle itself. As we will show, the commonly used energy optimization techniques in QMC: the linear method (LM)[42, 43, 44, 45] could be trivially adjusted to optimize excited states, with no additional cost. The LM method can be regarded as quasi Newton-Raphson (NR) minimization algorithm with an approximate Hessian matrix. In a variety of different applications, the performance of the LM method is as robust as the exact NR algorithm[44], although the former needs not to compute second-order derivatives explicitly.

Although the LM method is effective to optimize a small number of parameters, it suffers from fatal problems. Since it needs to construct the Hamiltonian and overlap matrices in the full parameter basis stochastically, every matrix element needs to be updated at each Monte Carlo sampling, and on each CPU process. This inefficient updating operation needs to be performed for millions of times on each process and slows down the program dramatically. Furthermore, once the matrix dimension becomes large (exceed a few thousand), the amount of memories required to store these matrices becomes too large to fit in memory. However, in strongly correlated systems, such as transition metal oxides with metal-insulator transition, tens of thousands of parameters are needed to achieve quantitative accuracy. The accelerated gradient descent approach developed by Booth[46] is an effective solution to the memory bottleneck of LM in determinant space. Similar techniques have been introduced to real-space QMC recently[47]. Although these methods consume significantly less amount of memory comparing to LM, they are also less effective in optimizing highly non-linear wave function parameters, such as orbital rotation parameters[47] and Jastrow factors[48].

In order to address the memory bottleneck without losing the effectiveness of LM, I will discuss the blocked LM (BLM)[49]. BLM separates the variable space into blocks, within each of which we estimate a small number of important update directions that can be used to construct a relatively small LM eigenvalue problem. As demonstrated in a various of different cases, this method drastically reduces the memory requirements without significantly affecting the accuracy of the optimization.

1.8 Outline

Chapter 2

In Chapter 2, we will introduce a new excited state variational principle and its evaluation and optimization with VMC algorithm. We will implement this method with two wave function ansatzes in Fock space and in real space. In order to show its predictive power, we compute excitation energies of various small molecules with low lying doubly excited states. Both in Fock space and real space, this new excited state variational principle yields order-of-magnitude of improvements over the state-of-art LR/EOM methods.

Chapter 3

In Chapter 3, we slightly change our focus to the LR/EOM based excited state methods. We will derive the LR/EOM theory based on the time-dependent Schrödinger's equation and discuss its implementation using VMC and the same wave function used in Chapter 2. We illustrate the accuracy of EOM-VMC method by comparing it with CIS and EOM-CCSD, and present numerical results of single and double excitations for a variety of different small molecules. The comparison of Chapter 2 and 3 will paint a clear picture of the advantages and disadvantages of both the new excited state variational principle and the LR/EOM methods.

Chapter 4

In Chapter 4, we will present further applications of this new excited state variational principle in terms of optical band gap predictions in real solids. We will start with a brief overview of MPBT method in the form of G_0W_0 approximation. Then we discuss how we construct the wave function for the exciton, along with techniques we used to address finite size effects. After showing the predicted optical gaps for a wide range of solids, we take zinc oxide as an example to provide insights into the MBPT methods.

Chapter 5

In Chapter 5, we describe how we address a fatal memory bottleneck in the standard VMC LM optimizer by introducing the BLM method. We will also present numerical examples showing that the excitation energy of a Mott-insulating hydrogen ring could be systematically converged with tens of thousands of wave function parameters.

Chapter 6

In this last chapter, we conclude our discussion with a summary of the current progresses and future research directions of the approach.

Chapter 2

A New Excited State Variational Principle

2.1 Introduction

In this chapter I will explore a new excited state variational principle. In fact, there is no unique form of the excited state variational principle and different formulations have been studied before. Messmer and co-workers[39] tried to target the excited states of harmonic and anharmonic oscillators by minimizing the distance between the trial wave function and targeted eigenstate with a specific energy. Similar formulation has been used in recent years by Troy Van Voorhis[50] with *ab initio* Hamiltonian and single-determinant wave function. Chan and co-workers[40] applied the Harmonic Davidson formulation, which is a powerful tool in mathematics used to find interior eigenstate of a matrix, and targeted excited states with DMRG.

Although the results of the previous studies are promising, their applicabilities are also limited by different types of problems. Firstly, owing to the prohibitively high computational cost, the deterministic evaluation of the formulation applied by Messmer and Van Voorhis is only applicable to simple Hamiltonians such as the harmonic oscillator, or uncorrelated single determinant wave functions for *ab initio* Hamiltonian. Secondly, the Harmonic Davidson algorithm utilized by Chan is limited only to linear wave function parameters. Considering the fact that almost every single widely used wave function ansatz contains nonlinear parameters, the flexibility of a wave function ansatz can never be fully explored without a formulation that could optimize both linear and nonlinear parameters.

Both of the aforementioned limitations could be fixed by the new excited state variational principle. We will use a formulation that is similar to the one used by Messmer. We shall demonstrate that although it is unclear whether there exists an efficient deterministic algorithm to evaluate this formula, its evaluation does admit an efficient stochastic algorithm through VMC. Combined with VMC's state-of-art optimizer, this approach could optimize both linear and nonlinear parameters. These two properties, along with the fact that the

method is independent of boundary conditions and the underlying basis, make the method a very promising candidate for the description of excited states in gas and condensed phase.

We begin our discussion with a overview of the ground state variational principle and point out its unique importance in electronic structure theory. We then present a new variational method consisting of two parts: first, a function $\Omega[\Psi]$ whose global minimum is an excited eigenstate, and second, a method for evaluating and minimizing Ω whose cost scales polynomially for a wide class of approximate wave functions. We will begin by proving that Ω has the necessary properties to be the basis of an excited state variational principle, after which we detail our method for minimizing it. During this discussion, we will explain which wave functions are compatible with the approach and introduce two specific wave functions that will be used in this thesis. We will also discuss the method's general applicability in molecules and solids. Finally, we will present numerical examples that demonstrate the method's potential both in Hilbert space and in real space.

2.2 Ground Variational Principle

The ground state variational principle applies the energy functional[2]:

$$E[\Psi] = \frac{\left\langle \Psi | \hat{H} | \Psi \right\rangle}{\langle \Psi | \Psi \rangle} \tag{2.1}$$

If we write the exact ansatz as a linear combination of all eigenstates of H, $|\Psi_e\rangle = \sum_i c_i |i\rangle$, this functional becomes:

$$E[\Psi] = \frac{\sum_i c_i^2 E_i}{\sum_i c_i^2} \geq E_0 \tag{2.2}$$

in which E_0 is the exact ground state energy.

The above proof shows that the global minimum of the energy functional corresponds to the exact ground state, given the exact wave function ansatz. For an approximate ansatz, searching for the global minimum of the energy functional in the ansatzes' parameter space would yield the most optimal approximate ground state wave function.

The ground state variational principle is of great importance in the field of electronic structure theory. By taking $|\Psi\rangle$ as a single determinant in Equation 2.1 and variationally optimizing the orbitals of the determinant, one gets the Roothann's equation of Hartree-Fock. Even though HF is not a quantitatively accurate method on its own, it is the first step of many highly accurate state-of-art quantum chemistry methods such as Moller-Plesset perturbation method (MP), CC methods, CASSCF, MRCI, QMC, and DMRG.

The success of DFT also relies on the ground state variational principle. The foundation of DFT, the Hohenberg-Kohn theory, which establishes the one-to-one mapping between the electron density and the external potential, is proved by the ground state variational principle. Furthermore, the KS equation of DFT is also derived by recasting Equation 2.1

into a search over densities of single determinant that give the lowest energy. Given the fact that DFT and many post-DFT methods such as GW and dynamical mean-field theory (DMFT)[51] are routinely used in chemistry, physics, and materials science, their successes would not be possible without the ground state variational principle.

2.3 Excited State Variational Principle

For excited states, we employ the functional,[41]

$$\Omega[\Psi] = \frac{\left\langle \Psi|(\omega - \hat{H})|\Psi \right\rangle}{\left\langle \Psi|(\omega - \hat{H})^2|\Psi \right\rangle} = \frac{\omega - E}{(\omega - E)^2 + \sigma^2} \tag{2.3}$$

where $\sigma^2 = \left\langle \Psi|(\hat{H} - E)^2|\Psi \right\rangle / \langle \Psi|\Psi \rangle$ is the variance and the energy shift ω is assumed to be placed in between distinct eigenvalues of $\hat{H}$ in order to target the eigenstate whose energy is immediately above it. Assuming real numbers for brevity, we proceed to prove that this eigenstate is the global minimum of Ω as follows. First, as in the ground state variational principle, we write an exact ansatz as a linear combination of all eigenstates of $\hat{H}$, $|\Psi_e\rangle = \sum_i c_i |i\rangle$, and rewrite Ω in terms of $\hat{H}$'s eigenvalues.

$$\Omega(\vec{c}) = \frac{\sum_i c_i^2(\omega - E_i)}{\sum_i c_i^2(\omega - E_i)^2} \tag{2.4}$$

Differentiating with respect to the elements of $\vec{c}$, we see that $\vec{c}$ is a stationary point (SP) if and only if

$$0 = c_i(\omega - E_i)(1 - (\omega - E_i)\Omega) \quad \forall i \tag{2.5}$$

Recalling that ω is assumed to be distinct from any of $\hat{H}$'s eigenvalues, we see that $\vec{c}$ cannot be a SP if any two of its elements that correspond to distinct Hamiltonian eigenvalues are nonzero, as this would prevent $(1 - (\omega - E_i)\Omega)$ from vanishing for both of them. In other words, $|\Psi_e\rangle$ cannot be a SP of Ω unless the nonzero values in $\vec{c}$ all correspond to one (possibly degenerate) eigenvalue of $\hat{H}$. As the eigenstates of $\hat{H}$ are clearly SPs of $\Omega(\vec{c})$, we see that $|\Psi_e\rangle$ is a SP if and only if it is an eigenstate. At one of these SPs, $\Omega(\vec{c})$ thus simplifies to $1/(\omega - E_i)$ and takes on negative values for SPs with $E_i > \omega$. The global minimum (i.e., most negative value) of $\Omega(\vec{c})$ therefore corresponds to the SP with E_i immediately above ω, as this maximizes the magnitude of $1/(\omega - E_i)$ while keeping its sign negative. (Note that this directionality, i.e. that we target the state above ω, is in our view the key difference between $\Omega(\Psi)$ and the related, nondirectional $\tilde{\Delta}_x^{(2)} = \left\langle \Psi|(\omega - \hat{H})^2|\Psi \right\rangle$ form considered by Messmer[39]). As $|\Psi_e\rangle$ can describe any state in Hilbert space, this value will be less than or equal to that of any approximate ansatz, thus achieving the variational property we desire.

Note that this proof requires no assumption about the system's boundary conditions, and so this variational principle has the potential to be applied to periodic systems as well as to molecules, for example to calculate band gaps by optimizing the wave function of the eigenstate at the bottom of the conduction band in a semiconductor.

2.4 Evaluation of Ω

While formally interesting, the mere existence of a variational function for excited states is not useful without an efficient way to evaluate and minimize it. Indeed, the presence of H^2 makes the straightforward evaluation of Ω drastically more expensive than the ground state function E, which is why studies that have worked implicitly with this function in the past[40, 52] have, to the best of our knowledge, always approximated this term (see discussion of harmonic Ritz methods below). As is done in variance evaluation[53] (and also by Booth and Chan,[54] but in a projector context) we avoid explicitly squaring $\hat{H}$ by resolving identities via complete sums over states,

$$\Omega[\Psi] = \frac{\sum_m \langle \Psi | m \rangle \left\langle m | (\omega - \hat{H}) | \Psi \right\rangle}{\sum_m \left\langle \Psi | (\omega - \hat{H}) | m \right\rangle \left\langle m | (\omega - \hat{H}) | \Psi \right\rangle} \tag{2.6}$$

We may evaluate this sum (up to a controllable statistical uncertainty that obeys the zero variance principle) through Monte Carlo integration as,

$$\Omega_{MC}[\Psi] = \frac{\sum_{m \in \xi} W_m}{\sum_{m \in \xi} |W_m|^2}, \quad W_m \equiv \frac{\left\langle m | (\omega - \hat{H}) | \Psi \right\rangle}{\langle m | \Psi \rangle} \tag{2.7}$$

where the elements of ξ are sampled from $|\langle m|\Psi\rangle|^2$ via a Metropolis walk (note that the normalization constants for numerator and denominator cancel and that we have corrected for the ratio-of-mean bias using Tin's modified ratio estimator[55]). Thus, any ansatz admitting efficient evaluations for W_m will be compatible with our approach. This includes the wide class of wave functions already used in ground state VMC for molecules and solids, such as Slater-Jastrow (SJ)[7], multi-Slater-Jastrow (MSJ)[56], the Jastrow antisymmetric geminal power (JAGP)[57, 58, 59, 60, 61, 62], and in principle even matrix product states[63, 64]. Moreover, the method is applicable to both real space, in which case m is a position vector, and Fock space, in which case m is an occupation number vector.

We need to point out that the denominator of Ω is similar to the variance σ^2 of the local energy. A naive evaluation of it can have infinite uncertainty[65], resulting in poor convergence behaviors in optimizations. This is because of the vanishing denominator of W_m in Equation 2.7 at the nodes of the wave function. To avoid this problem, instead of drawing samples directly from $|\Psi(R)|^2$, one could draw samples from a modified nodeless guiding function $|\Psi_G(R)|^2$[66, 67]. Such nodeless guiding functions have been demonstrated to reduce the statistical uncertainties of the variance[67].

2.5 Optimization of Ω

We now demonstrate that $\Omega[\Psi]$ could be optimized by a generalization of the ground state linear method (LM)[43]. Performing a linear expansion of $|\Psi\rangle$ with respect to its variational parameters $\vec{u}$,

$$\begin{aligned} |\Psi\rangle &\to \sum_i a_i \left|\Psi^i\right\rangle \\ \left|\Psi^0\right\rangle &\equiv |\Psi\rangle \,, \left|\Psi^i\right\rangle \equiv \partial |\Psi\rangle / \partial u_i \end{aligned} \tag{2.8}$$

we plug $|\Psi\rangle$ in Equation 2.8 into Equation 2.3 and find,

$$\Omega[\Psi] = \frac{\sum_{ij=0} a_i a_j \langle \Psi^i | \omega - H | \Psi^j \rangle}{\sum_{ij=0} a_i a_j \langle \Psi^i | (\omega - H)^2 | \Psi^j \rangle} \tag{2.9}$$

We then take derivative with respect to a_i and set it to be zero, and we arrive at the generalized eigenproblem

$$\sum_j \frac{\langle \Psi^i | [(\omega - H) - \lambda(\omega - H)^2] | \Psi^j \rangle}{\langle \Psi | \Psi \rangle} a_j = \sum_j \tilde{H}_{ij} a_j - \lambda \tilde{S}_{ij} a_j = 0 \tag{2.10}$$

in which we have defined the generalized Hamiltonian and overlap matrices as,

$$\begin{aligned} \tilde{H}_{ij} &= \frac{\langle \Psi^i | (\omega - H) | \Psi^j \rangle}{\langle \Psi | \Psi \rangle} \\ \tilde{S}_{ij} &= \frac{\langle \Psi^i | (\omega - H)^2 | \Psi^j \rangle}{\langle \Psi | \Psi \rangle} \end{aligned} \tag{2.11}$$

Assuming we were already near the minimum, in which case all $\vec{a}$ elements except a_0 will be small, then we may use $\vec{a}$ to update $\vec{u}$ through a reverse Taylor expansion exactly as in the ground state LM,

$$u_i \leftarrow u_i + a_i / a_0 \tag{2.12}$$

In practice, we may shift the eigenproblem[43, 45] to ensure this assumption is valid even when far from the minimum. The Hamiltonian matrix is shifted as[68],

$$\tilde{H} \to \tilde{H} + \alpha A + \beta B \tag{2.13}$$

in which the A matrix is defined as,

$$A_{ij} = \delta_{ij}(1 - \delta_{i0}) \tag{2.14}$$

This shift is called the "identity" shift. The identity shift raises the Ω value along each wave function derivative direction while keeping the current wave function $|\Psi\rangle = |\Psi^0\rangle$ unchanged.

While the identity shift itself could be very effective in many cases, it could struggle if the wave function first derivative vector have significantly different norms. In other words, it is entirely possible that changing one "steep" parameter by a small amount would lead to drastic change in the Ω value while doing the same for a "gentle" parameter only changes Ω modestly. Therefore, it is unfair to shift each parameter by the same amount, as done by the identity shift.

In order to address this problem, we introduce the "overlap" B shift. The B matrix is defined as,

$$\begin{aligned} Q_{ij} &= \delta_{ij} - \delta_{i0}(1-\delta_{j0})\tilde{S}_{0j} \\ T_{ij} &= (1-\delta_{i0}\delta_{j0})[Q^T\tilde{S}Q] \\ B &= (Q^T)^{-1}TQ^{-1} \end{aligned} \tag{2.15}$$

in which Q transforms into a basis in which all update directions are orthogonal to the current wave function. T is the overlap matrix in this basis with its first elements zeroed out so that the current wave function is not shifted. Finally, $(Q^T)^{-1}$ and Q^{-1} transform back to the original basis.

The shifting technique is essentially a trust region method: it would force the optimizer make small changes to the wave function parameters so that the first order Taylor expansion is valid. This linearize, diagonalize, and update procedure may be iterated to convergence in the same manner as Newton's method, allowing us to optimize the ansatz's linear and nonlinear parameters variationally via the minimization of Ω.

As the matrix elements for the eigenproblem can be evaluated through the same stochastic identity resolution as described above,

$$\begin{aligned} \frac{\left\langle \Psi^i|\omega-\hat{H}|\Psi^j\right\rangle}{\langle\Psi|\Psi\rangle} &= \sum_m \frac{|\langle\Psi|m\rangle|^2}{\langle\Psi|\Psi\rangle}\frac{\langle\Psi^i|m\rangle}{\langle\Psi|m\rangle}\frac{\left\langle m|\omega-\hat{H}|\Psi^j\right\rangle}{\langle m|\Psi\rangle} \\ &\approx \sum_{m\in\xi}\frac{\langle\Psi^i|m\rangle}{\langle\Psi|m\rangle}\frac{\left\langle m|\omega-\hat{H}|\Psi^j\right\rangle}{\langle m|\Psi\rangle} \\ \frac{\left\langle \Psi^i|(\omega-\hat{H})^2|\Psi^j\right\rangle}{\langle\Psi|\Psi\rangle} &= \sum_m \frac{|\langle\Psi|m\rangle|^2}{\langle\Psi|\Psi\rangle}\frac{\left\langle \Psi^i|\omega-\hat{H}|m\right\rangle}{\langle\Psi|m\rangle}\frac{\left\langle m|\omega-\hat{H}|\Psi^j\right\rangle}{\langle m|\Psi\rangle} \\ &\approx \sum_{m\in\xi}\frac{\left\langle \Psi^i|\omega-\hat{H}|m\right\rangle}{\langle\Psi|m\rangle}\frac{\left\langle m|\omega-\hat{H}|\Psi^j\right\rangle}{\langle m|\Psi\rangle} \end{aligned} \tag{2.16}$$

we arrive at a full-fledged and efficient method for the evaluation and minimization of Ω for any ansatz that can be efficiently used with the ground state LM in either molecules or solids. The precise cost scaling will of course depend on the choice of Ψ, with examples including $N_s N_e^3$ for real space SJ and JAGP and $N_s N_e^4$ for Hilbert space JAGP, where N_s and N_e

are the number of samples and electrons, respectively, both of which will grow linearly with system size.

Note the similarity of this eigenvalue equation to the harmonic Davidson equation that arises in applications[40, 52, 69, 70] of the harmonic Ritz principle[71, 72] for targeting interior eigenvalues of a matrix. In fact, some of these approaches[40, 52] appear to have been minimizing an approximation to Ω with respect to linear parameters, in which PH^2P was approximated by $PHPHP$, where P is the projector into the subspace corresponding to the linear parameters in question. Except for its controllable statistical uncertainty, the present approach makes no approximation when evaluating Ω and can optimize both linear and nonlinear parameters.

2.6 Wave Function Ansatzes

As an ansatz based method, choosing the most appropriate wave function ansatz is one of the most important steps in obtaining accurate predictions to excited states. As discussed in the previous chapters, a wide range of different ansatzes could be used in VMC. In general, an ansatz is compatible with VMC as long as the the local quantities such as W_m in Equation 2.7 could be evaluated efficiently. One of the main advantages of the Jastrow based ansatzes in real space VMC is that the electron-electron and electron-nucleus cusp conditions can be trivially satisfied, a property that conventional quantum chemistry ansatz such as CC does not process. In addition, the development of MSJ and JAGP ansatzes facilitate our understanding of strongly correlated systems, such as Hubbard model, bond breaking, Van-der Waals interactions, and metal-insulator transition.

Furthermore, VMC is not the only place that would benefit from an accurate wave function ansatz. As one of the most accurate electronic structure methods, diffusion Monte Carlo (DMC) takes the excited state nodal surface of optimized VMC wave function as input. Since the only error in DMC comes from the error of the trial wave function's nodal surface, obtaining more accurate wave functions on the VMC side would greatly improve the accuracy of DMC. We will use two commonly used wave function ansatzes in this chapter: the Jastrow antisymmetric geminal power (JAGP) and multi-Slater Jastrow (MSJ).

The JAGP Ansatz

The JAGP ansatz in orbital space is defined as:

$$|\Psi_{\mathrm{JAGP}}\rangle = \exp\left(\hat{J}\right)|\Psi_{\mathrm{AGP}}\rangle \tag{2.17}$$

$$|\Psi_{\mathrm{AGP}}\rangle = \left(\sum_{r\bar{s}} F_{r\bar{s}} a_r^\dagger a_{\bar{s}}^\dagger\right)^{N/2}|0\rangle \tag{2.18}$$

$$\hat{J} = \sum_{p\leq q} J_{pq}^{\alpha\alpha}\hat{n}_p\hat{n}_q + \sum_{\bar{p}\leq\bar{q}} J_{\bar{p}\bar{q}}^{\beta\beta}\hat{n}_{\bar{p}}\hat{n}_{\bar{q}} + \sum_{p\bar{q}} J_{p\bar{q}}^{\alpha\beta}\hat{n}_p\hat{n}_{\bar{q}} \tag{2.19}$$

where $N/2$ is the number of α (and β) electrons, unbarred and barred indices represent alpha and beta orbitals, respectively, $\hat{n}_p$ and $a_p^\dagger$ are the number and creation operator for the pth α orbital in the orthonormal orbital basis, and $|0\rangle$ is the vacuum.

Although the AGP part on its own is already a highly multi-referenced wave function, it suffers from a severe size-consistency problem resulting from the fact that only one geminal ($\sum_{r\bar{s}} F_{r\bar{s}} a_r^\dagger a_{\bar{s}}^\dagger$) is used and repeated, creating nonphysical charge fluctuations by placing 4 or more electrons in a single bond.

These nonphysical charge transfer can be eliminated by the number counting Jastrow factor defined in Equation 2.19. As showed before, such a Jastrow factor is able to control the number of electrons in certain region in space and deletes the charge transfer terms generated by AGP[62]. In this way the size inconsistency problem would be solved. In a variety of different applications involving multiple bond breaking, strongly correlated transition metal oxides, and barrier heights, JAGP and its extensions have been demonstrated to yield highly accurate results.

The MSJ Ansatz

The multi-Slater Jastrow [73, 56] ansatz is:

$$\Psi(\mathbf{r}_1, ..., \mathbf{r}_N) = e^{U(\mathbf{r}_1,...,\mathbf{r}_N)} \sum_I C_I \Phi_I(\mathbf{r}_1, ..., \mathbf{r}_N) \tag{2.20}$$

where $U(\mathbf{r}_1, ..., \mathbf{r}_N)$ is a correlation factor [68]

$$U(\mathbf{r}_1, ..., \mathbf{r}_N) = \sum_{ip} V_p(r_{ip}) + \sum_{i<j} W(r_{ij}) \tag{2.21}$$

The one-dimensional functions V and W are represented by 10-point cubic B-splines of the electron-nucleus (r_{ip}) and electron-electron (r_{ij}) distances. The electron-electron cusp condition is included in $W(r_{ij})$. Although it is also possible to satisfy the electron-nucleus cusp condition with $V_p(r_{ip})$, we instead include the cusp condition by using augmented orbitals [9].

The multi-Slater part of the wave function ($\sum_I C_I \Phi_I(\mathbf{r}_1, .., \mathbf{r}_N)$) is mainly used to account for strong correlation effects. It includes a reference determinant, which is usually taken as the Hartree-Fock determinant, and "excited" determinants in which one or more occupied orbitals in the reference determinant have been replaced by unoccupied ones. In strongly correlated systems, high order excitations are required to obtain accurate predictions, and the resulting multi-Slater expansion would be rather long. However, thanks to the development of the table method[74], the multi-Slater expansion could be evaluated efficiently and optimizations of tens of thousands of determinant expansions have been reported[74, 49, 66].

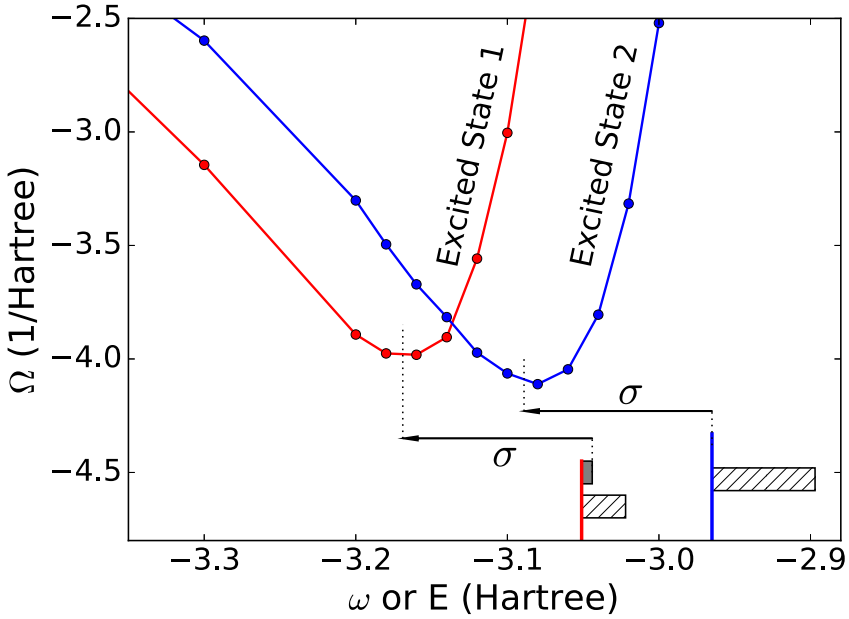

Figure 2.1: Ω vs ω for the first two excited states of an H_6 ring in the 6-31G basis, where Ψ_{JAGP} is optimized to minimize Ω at each ω value. At bottom right, solid vertical lines show FCI energies for these two states, while filled and slashed bars show deviations from FCI for JAGP and EOM-CCSD, respectively (note JAGP's deviation for state 2 is too small to be visible). The length-σ arrows give graphical confirmation that the ω value that minimizes Ω is roughly $E - \sigma$, as expected from our analysis of Eq. (2.3).

2.7 Effects of Finite Variance

As the excited state variational principle is expected to be used with an approximate ansatz in practice, it is important to consider the consequences of a nonzero value for the variance σ^2 in Equation 2.3. To guide this discussion, we plot in Figure 2.1 the value of Ω against the shift ω for the JAGP applied to the first two excited states of a regular, 1.5 $\mathring{A}$-edge-length H_6 hexagon. We first notice that instead of diverging at $\omega = E$ as would occur for an exact ansatz, Ω instead has a finite minimum near $\omega = E - \sigma$ which is in fact the analytic solution for the minimum of Equation 2.3 when E and σ are held fixed. A direct consequence of this downward shift of Ω's minimums is that the value of ω at which the character of the global minimum switches states is shifted downward as well, and no longer occurs at the energy of the lower state. The practical consequence is that the range of ω values that target a given state (i.e., those for which that state is the global minimum of Ω) gets shifted downward due to the nonzero variance of the wave function. Happily, the variance is readily evaluated in VMC, and so the magnitude of these shifts may be readily estimated and accounted for when selecting an ω with which to target a state.

A second consequence of using an approximate rather than exact ansatz is that the optimized wave function, and thus also its energy, may now depend on the precise choice of ω. In our tests so far, we have observed this energy dependence to be quite small (for an example, see Figure A.1 in the Appendix), and so the precise choice of ω does not meaningfully affect our results. This observation can be explained by considering that as a wave function becomes more accurate, the point in wave function variable space that minimizes Ω must align with the corresponding stationary point for the energy, and as the energy's first derivatives are zero at such a point, small changes to the wave function induced by adjusting ω will have only a very small effect on the energy value. However, in cases where the wave function approximation is poor, this mechanism will likely break down, and a strong dependence of the energy on ω may arise. Such a issue could be eliminated by the "updating ω" techniques[75] developed in recent years. This method automatically optimizes ω alongside the wave function so as to minimize Ω, thus it removes ω as a free parameter and prevents the user for adjusting it to select a desired result.

2.8 Numerical Results

Symmetry Breaking

Other consequences of using an approximate ansatz include the possibility of symmetry breaking, as occurs in ground state variational methods, and also the loss of orthogonality between the approximate eigenstates. If desired, these issues can be addressed by performing a configuration interaction between the optimized wave functions, which is trivial to perform within VMC for any ansatzes that may be efficiently optimized via Ω. Note that we only observed symmetry breaking in the case of CH_2's fifth and sixth excited states, for which we report the energies after a 2 state rediagonalization. Finally, note that all the effects of using an approximate ansatz may be systematically eliminated by increasing the ansatz's flexibility, ensuring the same systematically improvable accuracy that makes the ground state variational principle so powerful.

Difference Between Optimizing E and Ω

Before presenting results, we should also point out an important difference between optimizing E and Ω. Using Ω, the quality of the wave function depends strongly on both the value of its energy and its variance, as they are both important for shrinking the magnitude of the denominator in Equation 2.3. Ω minimization thus has as much in common with variance minimization[53] as it does with pure energy minimization. Therefore, just as it is biased (in some cases[76] by 0.5 eV) to take an energy difference between states when one is optimized for σ^2 and the other for E, it would be biased to take such a difference between states when one was optimized for E and the other for Ω. For this reason, we report results solely for

wave functions optimized via Ω, even when energy minimization is possible due to a state being the ground state or the lowest in its symmetry.

Results in Fock space

As a demonstration in Fock space, we applied the method to optimize the Hilbert space JAGP[61] for singlet excited states in an H_6 ring (Figure 2.1), CH_2 (Figure 2.2), and C_2 (Figure 2.3). In CH_2, the two doubly excited states are absent in CIS due to HF's limited LR space and are treated poorly by EOM-CCSD. While CCSD's LR space contains doubles, it lacks the triples necessary to describe the orbital relaxations that should accompany the excitation. Although JAGP's LR space also lacks triples, which becomes clear when one considers that its Jastrow factor can be written as a constrained CC doubles operator,[77, 78] it agrees much better with full CI (FCI)[79], because the variational minimization of Ω explores regions of parameter space beyond the LR regime.

The excited states of H_6 are even more challenging, each having 12 or more normalized FCI coefficients above 0.1 as compared to 8 or fewer for CH_2. Nonetheless, the same pattern emerges: large errors in EOM-CCSD are reduced by an order of magnitude in variationally optimized JAGP. We should stress that multi-reference perturbation theory is preferred over EOM-CCSD for double excitations, and that we compare to EOM-CCSD not because it is a benchmark, but because it is among the most reliable polynomial cost methods in wide use.

C_2 provides further evidence of JAGP's superiority to EOM-CCSD for double excitations while also revealing the limits of the ansatz's flexibility. While JAGP delivers 0.1 eV accuracy versus FCI for excited states 1, 2, 4, and 5, it shows an error almost as large as EOM-CCSD for state 3, a complicated excitation involving four different electrons in a mixture of double excitations. Moreover, JAGP's accuracy (and that of EOM-CCSD for the single excitations) is more dependent on error cancellation in this case, as seen in the total energy data provided in the Appendix. This raises the important point that, just like selecting a ground state ansatz to be balanced at, say, both equilibrium and stretched geometries, it is important in the present approach to select an ansatz that is not obviously unbalanced for the different excited states involved.

Results in Real Space

To show the method's systematic improvability and compatibility with a real space Monte Carlo walk, we have also treated C_2 with a MSJ ansatz consisting of short configuration state function (CSF) expansions and spline-based 1- and 2-body Jastrow factors (Figure 2.4). For each state, we selected CSFs with coefficients above a given threshold from a complete active space (CAS) wave function, leading to fewer than 10 (65) CSFs per state for a threshold of0.1 (0.01). Under variational optimization (with the random walk now in real space), the worst-case MSJ excitation energy error is found to drop from 0.3 to 0.1 eV upon lowering the threshold, as expected for a systematically improvable method. As a benchmark we use Davidson-corrected multireference CI (MRCI+Q) in a triple-ζ basis, which for excited state

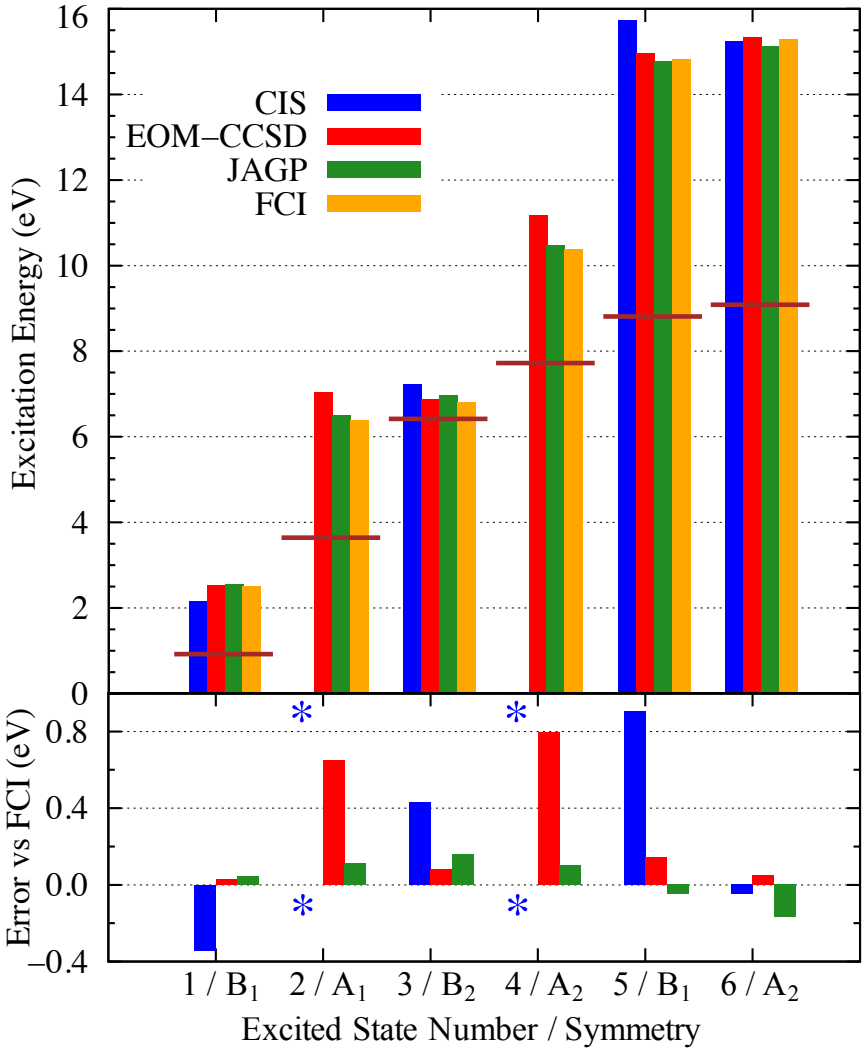

Figure 2.2: Singlet excitations for CH_2 in a STO-3G basis. Lines mark ω values. Asterisks mark doubly excited states. Note that the 5th and 6th states' energies are reported after the symmetry-restoring 2x2 re-diagonalization.

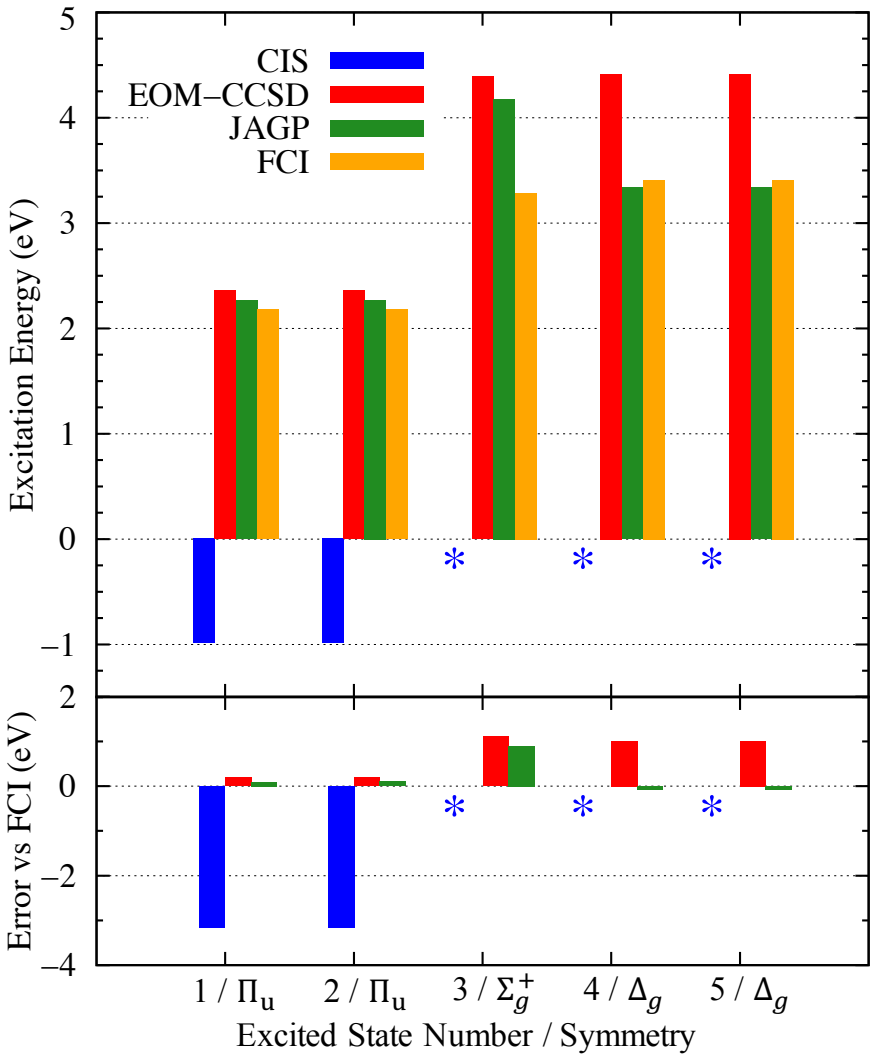

Figure 2.3: Singlet excitations for C_2 in a 6-31G basis. Asterisks mark doubly excited states.

5 (the $^1\Sigma_g^+$ state) is within 0.03 eV of the recent quadruple-ζ DMRG[80] and FCIQMC[81] benchmarks. Significantly, our MSJ result for this state (2.57 eV) is within 0.1 eV of these benchmarks (2.47 eV) and cc-pVTZ auxiliary field results (2.65 eV),[82] despite containing fewer than 100 variational parameters, compared to more than 4,000 in EOM-CCSD, millions in DMRG, and 2,000 in the FCIQMC trial function. This success, along with MSJ's high accuracy for C_2's other excited states, demonstrates the advantage of optimizing an ansatz directly and variationally for an individual excited state.

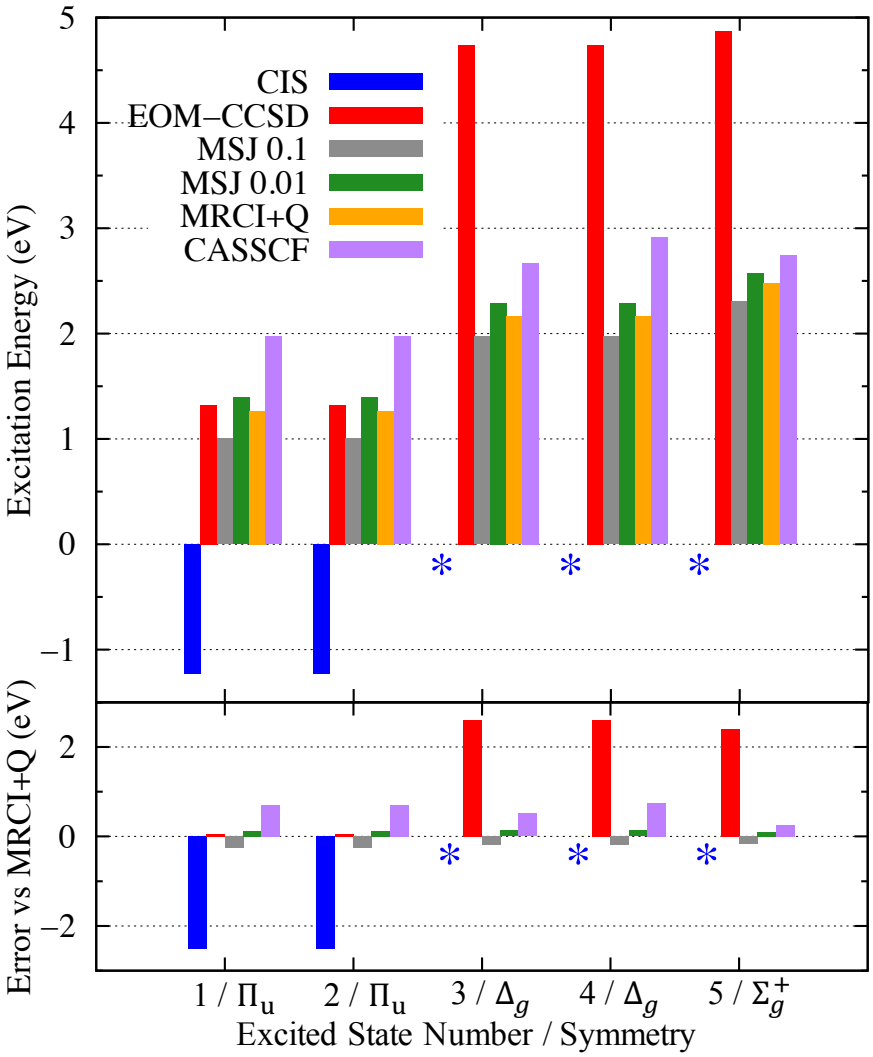

Figure 2.4: Singlet excitations for C_2 in a cc-pVTZ basis with MSJ in real space. Asterisks mark doubly excited states.

2.9 Conclusions

In this chapter, we have introduced a new excited state variational principle. It consists of two parts 1)a function Ω whose global minimum can be tuned to target individual excited states 2)a Monte Carlo method for the efficient, variational optimization of this function at polynomial cost with a wide range of approximate wave functions. We have also discussed necessary numerical techniques such as trust radius method used to stabilize the optimization. We also have addressed problems of using approximate wave function ansatzes, symmetry breaking, and dependence of the optimized energy on pre-choosen parameters.

Having laid out the theoretical basics of the new excited state variational principe, we implemented it with two different wave functions: JAGP in Fock space and MSJ in real space. In demonstrations on three molecules with low-lying doubly excited states, the method's ability to explore an ansatz's full variational freedom allows for drastic improvements in accuracy compared to linear response theories such as EOM-CCSD, which is among the most reliable polynomial-cost methods for excited states in chemistry (of course, when affordable, exponentially scaling multireference methods are preferred). Further, we have shown that for the notoriously difficult double excitations of the carbon dimer, variational optimization allows a very modest multi-Slater Jastrow expansion to achieve accuracies on par with the much more cumbersome DMRG and FCIQMC benchmarks. Given the importance of double excitations in light harvesting and excited state absorption experiments, the method's compatibility with periodic boundary conditions and thus the solid state, its systematically improvable nature, and its strong similarities to the ground state variational principle, we look forward to its further development and application.

Chapter 3

Equation of Motion Variational Monte Carlo Excited State Theory

3.1 Introduction

Having introduced the new excited state variational principle in last chapter, we turn our attention to a different VMC excited state approach in this chapter: the equation-of-motion variational Monte Carlo (EOM-VMC) excited state theory. Like other EOM/LR based excited state methods, EOM-VMC builds excited state approximations using linear combinations of an approximate ground state ansatz's first derivatives. Therefore, it inherits both the advantages and limitations of EOM/LR. As we shall see, the excitation energy of EOM-VMC is generally less accurate than that of the excited variational principle using the same wave function ansatz. The inaccuracy of LR/EOM could be attributed to its perturbative nature. Namely, LR/EOM tries to find excited states in the perturbative space of the ground state. For a highly compact wave function ansatz, such as single determinant, CCSD, and JAGP, its perturbative space is usually limited. Consequently, if the state of interest does not live in such space, or only partially live in such space, LR/EOM based method could become highly inaccurate, if not impossible.

One of the well known failure of EOM/LR are doubly excited states. As mentioned in the introduction, EOM-RHF (CIS) is incapable of describing such states, since the first derivative space of a single determinant only contains single excitations. Even though the first derivative space of CCSD wave function contains doubles, the performance of EOM-CCSD in terms of doubly excited states is often poor. The main reason is that the triple excitation contributions, which are needed for a double excitation to relax its orbitals, are missing in the first derivative space of CCSD.

With the limitations of EOM/LR based excited state methods, one may raise the question that what is point of studying these methods other than showing the superiority of the state specific method? One should note that although the EOM/LR based methods are in general less accurate, they are significantly more efficient than state specific methods. While the

latter needs to target excited state once at a time, the former yield the entire spectrum (or a large part of the spectrum) with a single calculation. Furthermore, albeit its failure on double excitations, EOM-CCSD is considered as one of the most accurate method for singly excited states. Even though CIS lacks quantitative accuracy, it does offer a cheap way to obtain qualitatively correct predictions to single excitations.

EOM/LR based excited state method with wave function ansatzes used in VMC is largely an unexplored regime. In this chapter we would like to derive the EOM-VMC method and show its performance with the JAGP wave function. While we limit our investigation in the current chapter to the JAGP ansatz, we stress that the EOM-VMC formalism itself is amenable to use with any wave function ansatz to which the VMC linear method [42, 76, 43, 44, 45] optimization is applicable. We thus seek to use the EOM approach as a direct comparison to the excited state variational approach to illustrate both the limitations and advantages of EOM excited state theory.

We begin with a review of the EOM formalism as derived from time dependent perturbation theory (Section 3.2). We then explain the use of LM technology to evaluate the EOM Hamiltonian and overlap matrices and discuss why a naive approach to Monte Carlo sampling can be problematic in this evaluation (Section 3.2). Having laid out the general formalism, we discuss the nature of the linear response space of the JAGP ansatz (Sections 2.6 and 3.2). We then compare this space to those of restricted Hartree-Fock (RHF) (Section 3.2) and CCSD (Section 3.2) to help illuminate the differences between EOM-JAGP, CIS, and EOM-CCSD, after which we conclude our theoretical analysis by discussing the origin and prevalence of ground-state bias in these methods (Section 3.2). Results are presented for singlet and triplet excitations in LiH (Section 3.3), H_2O (Section 3.3), C_2 (Section 3.3), and a collection of other small molecules (Section 3.3). We conclude (Section 3.4) with a summary of our findings and comments on the future development of EOM-VMC. In this chapter, we will be using acronym LR and EOM interchangeably.

3.2 Theory

EOM Linear Response

The general EOM linear response formalism for an approximate ansatz may be derived by assuming that the effects of a time-dependent perturbation to the Hamiltonian may be accounted for by adding time-dependent perturbations to the wave function's variational parameters. We therefore assume a Hamiltonian and wave function of the form

$$\hat{H} = \hat{H}_0 + e^{-i\omega t}\hat{H}_1 \tag{3.1}$$

$$|\Psi\rangle = e^{-iE_0 t/\hbar}\,|\Psi\left(\vec{u}(t)\right)\rangle \tag{3.2}$$

$$\vec{u}(t) = \vec{x}_0 + e^{-i\omega t}\vec{a} \tag{3.3}$$

$$E_0 = \langle\Psi\left(\vec{u}_0\right)|\,\hat{H}_0\,|\Psi\left(\vec{u}_0\right)\rangle \tag{3.4}$$

where $\hat{H}_1$ and $\vec{a}$ are assumed to be small, $\vec{u}_0$ are the variational parameter values for this ansatz's estimate of the unperturbed ground state, and E_0 is the corresponding estimate of the ground state energy. Using the shorthand notation

$$\left|\Psi^0\right\rangle \equiv \left|\Psi\left(\vec{u}_0\right)\right\rangle \tag{3.5}$$

$$\left|\Psi^k\right\rangle \equiv \left[\partial \left|\Psi\left(\vec{u}\right)\right\rangle / \partial u_k\right]_{\vec{u}\to\vec{u}_0} \qquad (\text{for } k>0) \tag{3.6}$$

this wave function may be Taylor-expanded as

$$\left|\Psi\right\rangle = e^{-iE_0t/\hbar}\left(\left|\Psi^0\right\rangle + e^{-i\omega t}\sum_k \mu_k \left|\Psi^k\right\rangle + \mathcal{O}\left(|\vec{a}|^2\right)\right). \tag{3.7}$$

This expansion may be inserted into the Schrödinger equation $i\hbar\partial|\Psi\rangle/\partial t = \hat{H}|\Psi\rangle$ to give

$$\begin{aligned}\hat{H}_1\left|\Psi^0\right\rangle + (\hat{H}_0 - E)\sum_k a_k\left|\Psi^k\right\rangle \\ = e^{i\omega t}\left(E_0 - \hat{H}_0\right)\left|\Psi^0\right\rangle\end{aligned} \tag{3.8}$$

where $E \equiv \hbar\omega + E_0$ and we have dropped terms quadratic in the perturbation or smaller. Assuming the ground state variational principle is satisfied, and thus

$$\left\langle\Psi^k\right|\left(\hat{H}_0 - E_0\right)\left|\Psi^0\right\rangle = 0, \tag{3.9}$$

we may project Eq. (3.8) into the span of the ansatz's first derivatives (i.e. left-project by $\langle\Psi^j|$) to obtain

$$\sum_k \left\langle\Psi^j\right|(\hat{H}_0 - E)\left|\Psi^k\right\rangle a_k = -\left\langle\Psi^j\right|\hat{H}_1\left|\Psi^0\right\rangle. \tag{3.10}$$

Eq. (3.10) is the EOM approximation for the response $\vec{a}$ to a small perturbation $\exp(-i\omega t)\hat{H}_1$ to the Hamiltonian. Note that in particular, this approximation gives the resonances, i.e. the frequencies ω at which the response may be large even for a small perturbation, as the eigenvalues of the unperturbed Hamiltonian $\hat{H}_0$ in the subspace of Hilbert space spanned by the ansatz's first derivatives with respect to its variational parameters. In conclusion, obtaining the EOM estimates of the excitation energies $\hbar\omega = E - E_0$ therefore requires only that the Hamiltonian be diagonalized in this subspace by solving

$$\sum_k \langle\Psi^j|\hat{H}_0|\Psi^k\rangle a_k = E\sum_k \langle\Psi^j|\Psi^k\rangle a_k. \tag{3.11}$$

EOM-VMC

While deterministic methods exist to solve Eq. (3.11) for RHF and CCSD (giving the CIS and EOM-CCSD methods, respectively), a stochastic approach is more efficient in the case of wave functions with Jastrow factors like the JAGP. Happily, the matrices involved are already available in many QMC software packages as they are the same matrices required for the ground state LM. Here we briefly review how these matrices are estimated stochastically, and also point out some potential pitfalls when generalizing the methodology for use in EOM.

To obtain our stochastic estimate to Eq. (3.11), we first insert resolutions of the identity, either in Fock space or real space,

$$I = \sum_{\vec{n}} |\vec{n}\rangle \frac{|\langle\vec{n}|\Phi\rangle|^2}{|\langle\vec{n}|\Phi\rangle|^2} \langle\vec{n}| \tag{3.12}$$

$$I = \int d\vec{r}\, |\vec{r}\rangle \frac{|\langle\vec{r}|\Phi\rangle|^2}{|\langle\vec{r}|\Phi\rangle|^2} \langle\vec{r}| \tag{3.13}$$

on both sides (note we will work in Fock space but the approach is equally well defined in real space) to obtain

$$\begin{aligned} &\sum_{\vec{n}} \sum_{j} \frac{|\langle\vec{n}|\Phi\rangle|^2}{\langle\Phi|\Phi\rangle} \frac{|\langle\vec{n}|\Psi^0\rangle|^2}{|\langle\vec{n}|\Phi\rangle|^2}\, \mathcal{D}^*_{\vec{n},i}\, \mathcal{G}_{\vec{n},j}\, \mu_j \\ &\quad = E \sum_{\vec{n}} \sum_{j} \frac{|\langle\vec{n}|\Phi\rangle|^2}{\langle\Phi|\Phi\rangle} \frac{|\langle\vec{n}|\Psi^0\rangle|^2}{|\langle\vec{n}|\Phi\rangle|^2}\, \mathcal{D}^*_{\vec{n},i}\, \mathcal{D}_{\vec{n},j}\, \mu_j \end{aligned} \tag{3.14}$$

where

$$\mathcal{D}_{\vec{n},j} \equiv \frac{\langle\vec{n}|\Psi^j\rangle}{\langle\vec{n}|\Psi^0\rangle} \tag{3.15}$$

$$\mathcal{G}_{\vec{n},j} \equiv \frac{\left\langle\vec{n}|\hat{H}_0|\Psi^j\right\rangle}{\langle\vec{n}|\Psi^0\rangle} \tag{3.16}$$

and we have introduced the importance sampling function $|\Phi\rangle$. Eq. (3.14) may be evaluated stochastically by an average on the Monte Carlo sample Ω drawn from Φ's probability distribution by a Metropolis-Hastings walk, yielding

$$\begin{aligned} &\sum_{\vec{n}\in\Omega} \sum_{j} \frac{|\langle\vec{n}|\Psi^0\rangle|^2}{|\langle\vec{n}|\Phi\rangle|^2}\, \mathcal{D}^*_{\vec{n},i}\, \mathcal{G}_{\vec{n},j}\, \mu_j \\ &\quad = E \sum_{\vec{n}\in\Omega} \sum_{j} \frac{|\langle\vec{n}|\Psi^0\rangle|^2}{|\langle\vec{n}|\Phi\rangle|^2}\, \mathcal{D}^*_{\vec{n},i}\, \mathcal{D}_{\vec{n},j}\, \mu_j \end{aligned} \tag{3.17}$$

Thus, so long as a reasonable guiding function Φ is known and the ratios $\mathcal{D}_{\vec{n},j}$ and $\mathcal{G}_{\vec{n},j}$ can be evaluated efficiently, as is possible for the JAGP [61], then EOM estimates of the excitation energies may be evaluated for a cost similar to a ground state LM calculation.

However, while the ground state LM often makes the choice $|\Phi\rangle = |\Psi^0\rangle$, such a choice can be pathological in EOM-VMC due to a ground state sampling bias. To make this issue clear, consider the following simple model. Suppose we have a three-level system with $|1\rangle$, $|2\rangle$, and $|3\rangle$ being its exact eigenstates and we take as our ansatz the full configuration interaction (FCI) wave function, an arbitrary linear combination of all three states. Next assume that we have already optimized the ground state perfectly, so $|\Psi^0\rangle = |1\rangle$. Choosing $|\Phi\rangle = |\Psi^0\rangle$ would in this case prevent us from sampling the excited states at all, and so our stochastic estimate for the eigenvalue equation,

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} = E \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix}, \tag{3.18}$$

would be useless. While in practice our ground state estimate $|\Psi^0\rangle$ is unlikely to be exact, using it as the guiding function will be statistically inefficient, especially in cases when an excited state has a different symmetry than the ground state.

In future work it may be profitable to test general solutions to this problem, and indeed it has been addressed in other work by knowing the basic structure of the excited states of interest and using a guiding function that is a linear combination of these states [83]. In the present study, we have avoided sampling pathologies by adding random noise to the ground state JAGP's pairing matrix. We find that noise distributed uniformly between -0.1 and 0.1 (when the largest pairing matrix element is 1) is effective.

JAGP's Linear Response Space

The JAGP ansatz is defined in Equation 2.17. While it has produced highly accurate results in a number of difficult molecules, especially upon optimization of its orbital basis [77, 78], such results have all been obtained through non-linear parameter optimizations targeted at individual eigenstates. In the present chapter, we seek to determine its efficacy when such individual state optimizations are eschewed in favor of an EOM approach. To better understand what capabilities and limitations to expect in this new use of JAGP, we will discuss the nature of its linear response space and make formal comparisons to other EOM methods, namely CIS and EOM-CCSD.

The accuracy of any LR based methods depend on both the number and nature of the ansatz's first derivatives. In considering the nature of the JAGP's wave function derivatives, we separate them into those for the pairing matrix (F) and Jastrow factor (J) variables.

The AGP by itself is able to create closed-shell and open-shell configurations. Consider the simple H_2 molecule in a minimal basis, noting that rotations of the one-particle basis will not change the span of the AGP's first derivatives and that we can analyze its properties under any rotation that is convenient. If we work in molecular orbital basis, for example, and label the bonding orbital as 1 and anti-bonding orbital as 2, then the RHF solution,

encoded in an AGP pairing matrix, is:

$$F_{\text{closed-shell}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{3.19}$$

in which the nonzero matrix element creates a pair of electrons in the bonding orbital. Similarly, the HOMO→ LUMO singlet open-shell configuration is encoded as:

$$F_{\text{open-shell}} = \begin{pmatrix} 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 \end{pmatrix} \tag{3.20}$$

Clearly, the open-shell pairing matrix can be written as a sum of derivatives of the closed-shell matrix with respect to its elements, and so we would expect EOM-AGP, like CIS, to be capable of describing this type of simple excitation.

The derivative of the JAGP with respect to a pairing matrix element is:

$$\frac{\partial \left|\Psi_{\text{JAGP}}\right\rangle}{\partial F_{p\bar{q}}} = (N/2) \exp\left(\hat{J}\right) a_p^\dagger a_{\bar{q}}^\dagger \left(\sum_{r\bar{s}} F_{r\bar{s}} a_r^\dagger a_{\bar{s}}^\dagger\right)^{(N/2-1)} \tag{3.21}$$

which, although more complicated than the simple H_2 example, will have a similar physical meaning when the ground state is dominated by the RHF determinant. In this discussion we will limit our analysis to this single-reference case, although it would be interesting in future to investigate how the derivatives change in more multi-configurational settings.

If p and $\overline{q}$ are both occupied in the ground state, the derivative above will essentially give the ground state wave function back and little information will be gained about excited states. If p is occupied but $\overline{q}$ is empty in the ground state, this derivative will create a $p \to \overline{q}$ single excitation. If both p and $\overline{q}$ are empty in the ground state, this derivative will create a double excitation.

Derivatives with respect to Jastrow factor variables,

$$\frac{\partial \left|\Psi_{\text{JAGP}}\right\rangle}{\partial J_{p\bar{q}}^{\alpha\beta}} = \hat{n}_p \hat{n}_{\bar{q}} \left|\Psi_{\text{JAGP}}\right\rangle \tag{3.22}$$

although easy to evaluate, are not so easily analyzed as those for the pairing matrix, in part because their character is strongly dependent on the one-particle basis chosen for the number operators. In this work, as in other studies of the Hilbert-space JAGP, this basis is chosen to be local. Thus Jastrow derivatives produce projections of the ground state wave function in which two particular local orbitals are guaranteed to be occupied. While many such projections are no doubt components of excited states, predicting their significance in EOM-JAGP is not so straightforward. At best, the coupled cluster interpretation [77] of the Jastrow factor would suggest that such derivatives provide a limited subset of the excitations present in EOM-CCSD. However, this subset will have been optimized for the purposes of lowering the ground state energy, and as it is only a small subset, it would be a surprise if it could reproduce the highly flexible linear response space provided by the coupled cluster doubles operator.

Comparison with CIS

CIS, equivalent to EOM-RHF, has a first derivative subspace of size $N_{occ}N_{vir}$, consisting exclusively of single excitations out of the RHF determinant. Thus both EOM-JAGP and CIS have a first derivative space of size $\mathcal{O}\left(N^2\right)$, with EOM-JAGP's being larger by a constant prefactor. Given that it has a larger EOM subspace, contains RHF as a special case, and has some potential for treating double excitations, one might expect EOM-JAGP to be strictly superior to CIS in terms of accuracy in excitation energies. While this appears to be true in our results for relative excitation energies, it is not always true for absolute excitation energies due to JAGP's much stronger ground state bias (see Section 3.2).

Comparison with EOM-CCSD

Like EOM-CCSD, EOM-JAGP has the potential to treat both single and double excitations. Further, given the presence of double excitations in its EOM subspace, one might expect JAGP to benefit from the tendency, common in EOM-CCSD, of these excitations to act to relax the wave function's orbitals in the presence of a single excitation. However, the double excitations in EOM-JAGP are much less flexible than in EOM-CCSD, a reality made clear by a close look at the the CCSD wave function:

$$\begin{aligned} |\Psi_{\text{CCSD}}\rangle &= \exp\left(\hat{T}_1 + \hat{T}_2\right)|\text{RHF}\rangle \\ \hat{T}_1 &= \sum_{ia} t_i^a a_i a_a^\dagger \\ \hat{T}_2 &= \sum_{i>j,a>b} t_{ij}^{ab} a_i a_a^\dagger a_j a_b^\dagger \end{aligned} \tag{3.23}$$

From this expression, we can see that derivatives with respect to the cluster amplitudes will produce an EOM subspace containing $N_{occ}N_{vir}$ single excitations and $\mathcal{O}\left(N_{occ}^2 N_{vir}^2\right)$ double excitations. Contrast this with EOM-JAGP, where we find only $\mathcal{O}\left(N_{vir}^2\right)$ double excitations, suggesting a great disparity in flexibility with respect to doubles. Beyond sheer number, the occupied-orbital indexation of the CCSD doubles gives EOM-CCSD direct control of which occupied orbitals a double excitation is promoted from, whereas the double excitations in EOM-JAGP are in effect indexed only by their virtual orbitals, creating what we think of as the "uncontrolled hole" problem in which EOM-JAGP has difficulty ensuring that a double excitation be promoted from physically reasonable occupied orbitals. In practice, we will see that these disparities prevent EOM-JAGP from achieving the high accuracy typical of EOM-CCSD, presumably because they rob it of the ability to carefully relax orbital shapes in the presence of an excitation, although interestingly their performance is more similar (although not particularly good) for double excitations.

Ground State Bias

All EOM-based methods should be expected to suffer from a bias in favor of the ground state, and therefore too-high excitation energies, as the initial variational parameters $\vec{x}_0$ have been optimized for this state in a nonlinear fashion that takes into account interactions between the effects of different parameters. Linear response methods, by their very nature, cannot achieve this degree of tailoring for the excited states. For example, in CIS, the EOM subspace contains the freedom to shape the excitation's orbital, but cannot achieve the second-order effect of relaxing the shapes of other orbitals in the presence of the excitation. This lack of orbital relaxation is perhaps in practice the most common and important source of ground state bias, showing up also in EOM-CCSD in the case of doubly excited states (for which EOM-CCSD has no triples to use to couple in relaxation in the way it can via its doubles for singly excited states).

In EOM-JAGP we find that ground state bias can be particularly severe, because while it's EOM subspace is closer in its flexibility to that of CIS than to that of EOM-CCSD, the correlation included by and thus the accuracy of ground state JAGP is closer to that of CCSD. From the EOM perspective, JAGP is in a sense too clever for its own good: by capturing a large amount of ground state correlation energy using an ansatz with a small number of parameters and thus a relatively inflexible EOM subspace, it is virtually guaranteed to have a sizable bias in favor of the ground state. As we will see, this bias is severe enough that it tends to overestimate excitation energies even more so than CIS, despite having a somewhat more flexible EOM space. Fortunately, as the ground state bias affects all excitation energies roughly equally, relative energies between different excitations should be little affected and should be expected to show improvement over methods (like CIS) that have less flexible EOM spaces.

While it is easy to confirm the presence of ground state bias by comparing absolute excitation energies to those of a benchmark method, analyzing the accuracy of relative excitation energies is less straightforward. The approach we take rests on the idea that two methods, both with exact relative energies between excited states but with different ground state biases, can be made the same by applying a constant shift, and thus we seek a measure of relative accuracy that will automatically account for any such constant shift. To this effect we will use a root-mean-square relative deviation (RMSRD) metric

$$\mathrm{RMSRD}_N \equiv \left(\frac{1}{N} \sum_{i}^{N} \left[\Delta_{i,\mathrm{Method}} - \Delta_{i,\mathrm{Benchmark}} \right]^2 \right)^{1/2} \tag{3.24}$$

where $\Delta_i = \hbar(\omega_i - \bar{\omega})$ is the deviation of a method's ith excitation energy ($\hbar\omega_i$) from the mean ($\hbar\bar{\omega}$) of that method's first N excitation energies. RMSRD_N thus measures how closely a method's excitation energies' deviations from their own mean match the corresponding deviations in a benchmark method, and so the RMSRD_N for a method with exact relative energies between excited states but a large ground state bias would be zero, while that for a method with no ground state bias overall but large errors in relative excitation energies

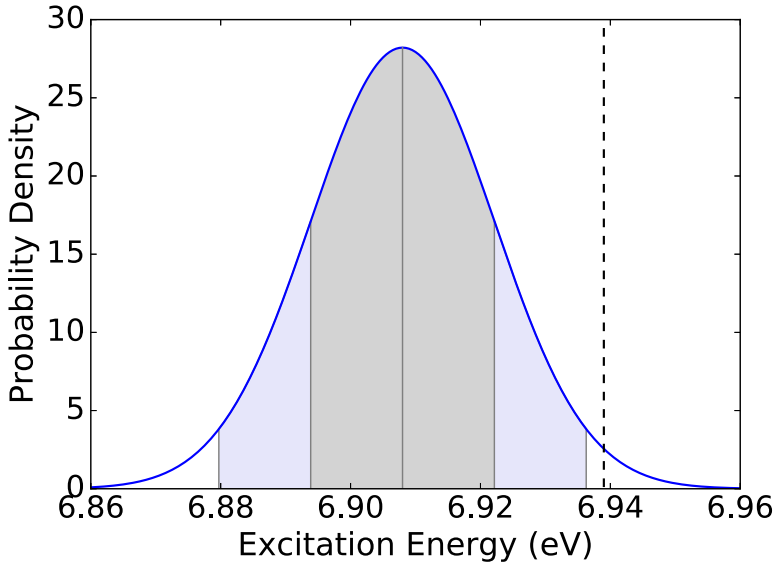

Figure 3.1: Plot of a Gaussian representing the statistical uncertainty of the EOM-JAGP 1st excitation energy for H_2O in a 6-31G basis after 100 independent Monte Carlo calculations. Shaded regions represent the 1-σ and 2-σ confidence intervals, while the dashed vertical line shows the brute-force, non-stochastic result. See Section 3.2 for details.

would be large. Thus, by analyzing both absolute excitation energies and excitation energies' deviations from their own mean, we will attempt to distinguish the effect of ground state bias from other sources of error.

Statistical Uncertainty and Bias

Unlike ground state optimization, the evaluation of statistical effects for eigenvalues of the Hamiltonian matrix is less straightforward, as the eigenvalues are related in a nonlinear way to the sampled local energies and derivatives. To achieve a simple estimate of our excitation energies' statistical uncertainty, we have therefore run 10 independent calculations for each system with different random seeds and evaluated the mean and statistical uncertainty of each excitation energy directly. All the numbers we report in this paper are the mean value of independent runs with statistical uncertainties estimated to be below 0.05eV. This should be compared with the statistical uncertainties we observe for the systems' ground states, all of which are below 0.01eV, which suggests that some combination of the guiding functions we employ and the nonlinearity of the eigenvalues is acting to increase the statistical uncertainty of excitation energies relative to ground state energies.

In addition to statistical uncertainty, we must be mindful of the potential for bias due to the nonlinear relationship between the eigenvalues that give our excitation energies and the

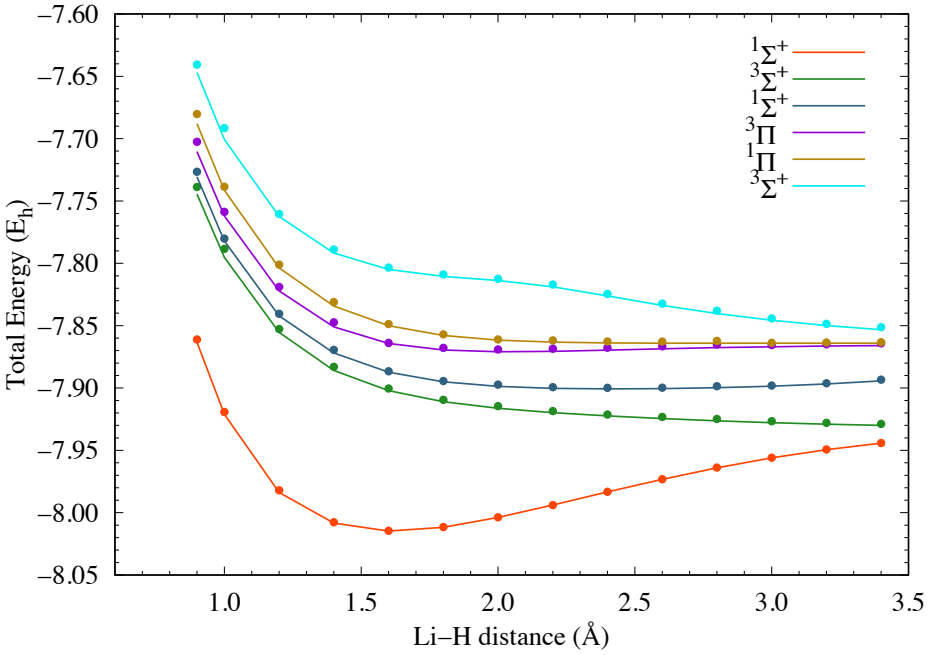

Figure 3.2: Potential energy curves for the lowest excited states of lithium hydride in the cc-pVDZ basis set. Solid lines are FCI while dots are EOM-JAGP calculation. EOM-JAGP correctly finds the lowest three triplet and two singlet states.

Hamiltonian and overlap matrix elements, as these elements are all evaluated using the same Markov chain. To quantify the magnitude of such a bias, we performed 100 independent runs for H_2O in a 6-31G basis and analyzed the resulting distribution of its first excitation energy, which a standard normality test revealed to be almost perfectly normally distributed. In Figure 3.1, we compare the resulting uncertainty in the average excitation energy from these 100 runs to that obtained by a non-stochastic evaluation of the excitation energy, which we get by taking a "sample" in which every configuration is visited once and given a weight of $|\langle \overrightarrow{n}|\Psi\rangle|^2$. We see that our statistical resolution from our 100 runs is not sufficient to give a very precise value for the bias, but we can conclude that it is on the order of 0.05 eV or less.

Table 3.1: Vertical excitation energies and $RMSRD_5$ values (both in eV) for the water molecule at equilibrium in a 6-31G basis set.

	Excitation Energy (eV)			
State	EOM-JAGP	CIS	EOM-CCSD	FCI
$^3B_2$	6.95	6.64	6.03	6.10
$^1B_1$	7.89	7.71	6.79	6.85
$^3A_1$	8.39	7.82	8.02	8.07
$^3A_2$	9.04	8.51	8.17	8.17
$^1A_1$	9.73	9.31	8.74	8.75
$RMSRD_5$	0.26	0.37	0.02	0.00

3.3 Results

Case Study 1: LiH in cc-pVDZ

We begin our results with a simple example, the LiH molecule in a cc-pVDZ basis [84]. This system has only two valence electrons, and thus we expect the two-electron nature of the AGP to result in high accuracy for EOM-JAGP. Due to the two-electron nature of AGP and the fact that all low lying excited states are singly excited states, we expect nearly exact results. Figure 3.2 shows our results for the lowest five excited states at 14 bond lengths between 0.9 $\mathring{A}$ and 3.4 $\mathring{A}$. As expected, EOM-JAGP is quite accurate, with an overall average error from FCI of just 2 milliHartree.

Case Study 2: H_2O in 6-31G

The water molecule provides an excellent illustration of how the different EOM methods we discuss are expected to perform for single excitations in a single reference system. As is well known, EOM-CCSD proves highly accurate in this setting. EOM-JAGP and CIS are less accurate, but in different ways. Table 3.1 and Figure 3.3 shows that EOM-JAGP tends to overestimate water's excitation energies, even more so than does CIS, in keeping with the expectation of a more severe ground state bias. Also as expected, the $RMSRD_5$ values and Figure 3.4 show that EOM-JAGP produces more accurate relative energies between excitations than CIS, but less accurate relative energies than EOM-CCSD.

Case Study 3: C_2 in 6-31G

Our last case study is C_2 molecule in 6-31G [85] basis set, which presents a major challenge for EOM methods due to the presence of strong correlation in the ground state and low-lying doubly excited states. Indeed, the very poor quality of RHF for the ground state of C_2 leads CIS to predict the first five excited states to lie *below* the ground state in energy, and its

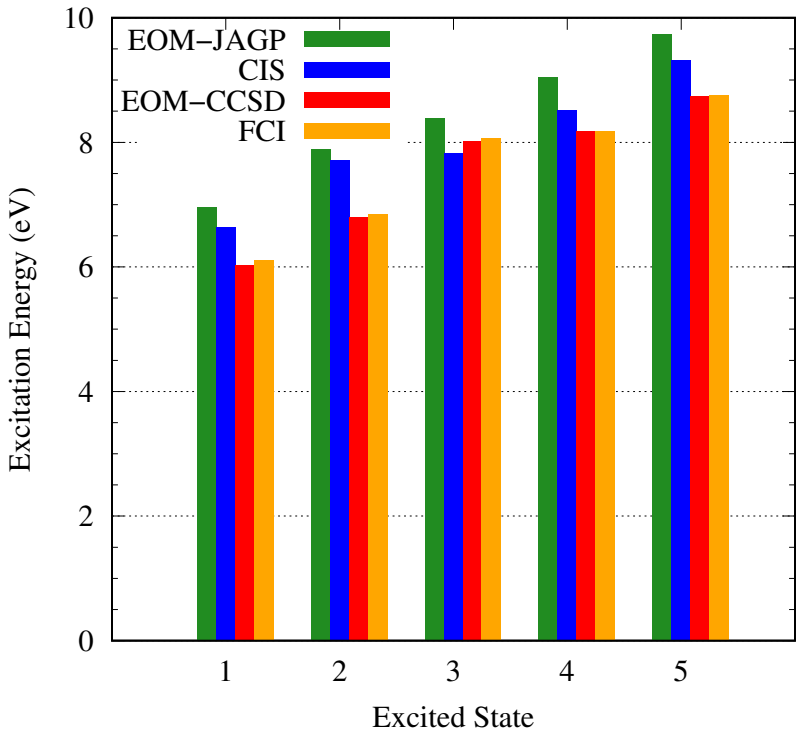

Figure 3.3: Vertical excitation energy for H_2O at equilibrium in a 6-31G basis set.

Table 3.2: Vertical excitation energies and $RMSRD_5$ values (both in eV) for C_2 at equilibrium in a 6-31G basis set.

	Excitation Energy			
State	EOM-JAGP	CIS	EOM-CCSD	FCI
$^3\Sigma_g^-$	0.61	-1.46	0.59	0.60
$^3\Pi_u$	1.12	-2.14	0.88	0.89
$^3\Pi_u$	1.12	-2.14	0.88	0.89
$^1\Pi_u$	2.41	-0.98	2.36	2.17
$^1\Pi_u$	2.41	-0.98	2.36	2.17
$^1\Sigma_g^+$	4.30	N/A	4.39	3.28
$RMSRD_5$	0.09	0.42	0.09	0.00

limitation to single excitations causes CIS to miss the doubly-excited sixth excited state entirely. The presence of strong correlation is not nearly so problematic for EOM-JAGP, which performs if anything better than expected, especially in absolute energies for single excitations (the first five excitations) in which it displays very little ground state bias. The most likely explanation for this lack of bias lies in the ground state JAGP's inability to capture as high a fraction of the correlation energy as in H_2O (although it is still vastly

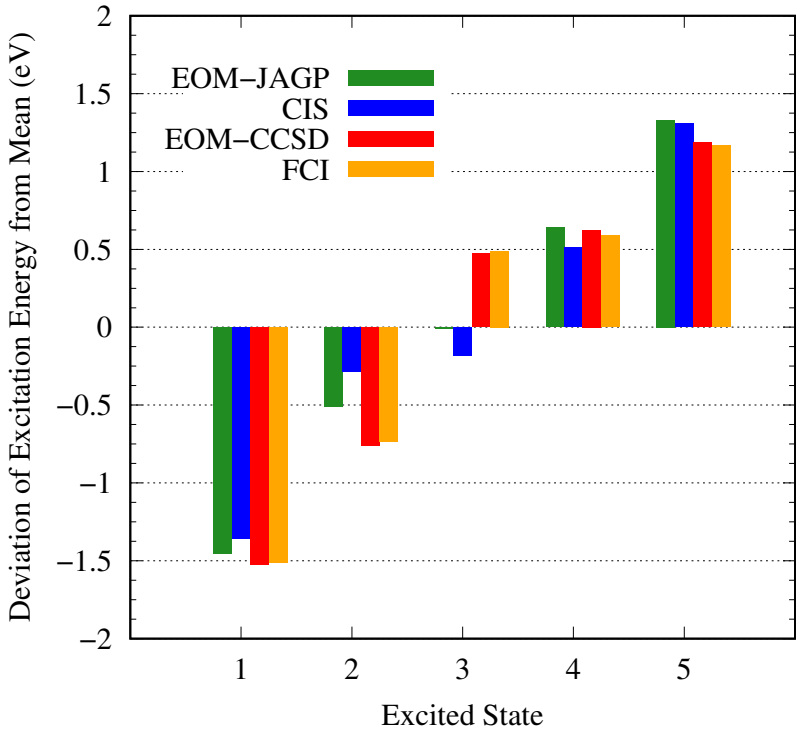

Figure 3.4: For each method, the deviation of each excitation energy from that method's mean excitation energy for H_2O in a 6-31G basis.

superior to RHF), and so some cancellation of error appears to be at work. Overall, the EOM-JAGP results are similar to those of EOM-CCSD, being accurate for single excitations but having a much too high energy for the doubly excited state due to a lack of triples excitations in their EOM subspaces.

Other Benchmarking Calculations

To further test the performance of EOM-JAGP, we performed calculations of vertical excitation energies and RMSRD values for a number of diatomic (He_2, Li_2, N_2 and CO) and polyatomic (HCN, CH_2O, and C_2H_2) systems, with results shown in Figure 3.7 and Tables 3.3 and 3.4. As one might expect, EOM-JAGP delivers nearly exact results for He_2 and Li_2 as these consist of very weakly interacting pairs of electrons, an ideal situation for a pairing theory. In CH_2, where we report excitation energies relative to the lowest singlet, we see both EOM-JAGP's large ground state bias (large enough that it fails to predict a triplet ground state) and its difficulty in handling double excitations (the error is larger than 1eV for the $^1A_1$ $(HOMO)^2 \rightarrow (LUMO)^2$ excitation). The latter failure should be put into context, however, as EOM-CCSD has essentially the same difficulty. CH_2's double excitation is thus a good example of how EOM methods' quality typically degrades as an excited state

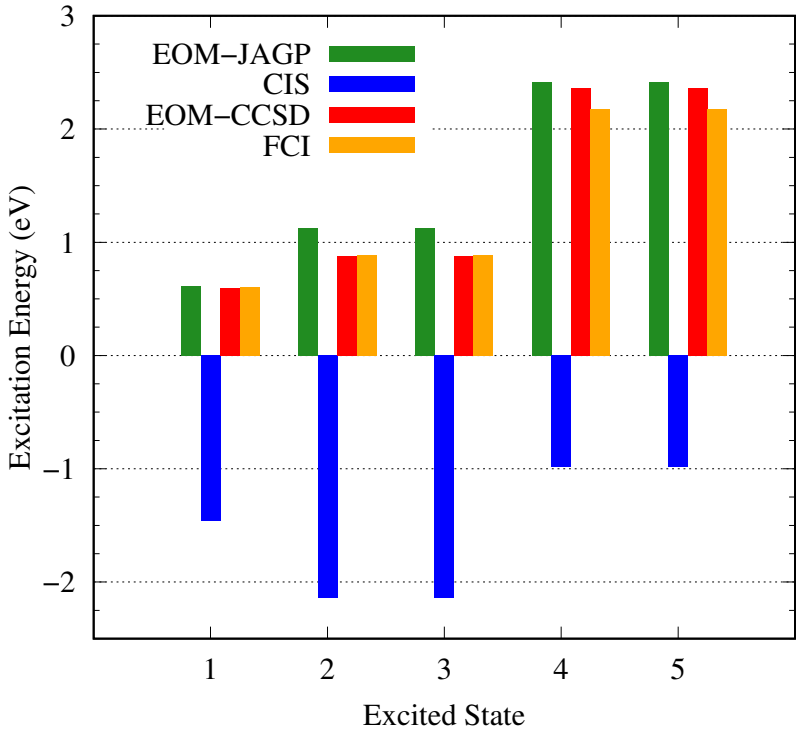

Figure 3.5: Vertical excitation energy for C_2 in a 6-31G basis set.

becomes more and more different from the ground state, as occurs when increasing numbers of electrons are excited.

In N_2, CO, HCN, C_2H_2 and CH_2O, EOM-JAGP displays less of a ground state bias. As we have not in this study optimized the orbital basis for the Jastrow factor [77], the ground state JAGP is hampered in its recovery of dynamic correlation, which appears to manifest more strongly in these multiply-bonded systems in which the correlation between electron pairs is expected to be more important. We expect that the resulting raising of JAGP's ground state energy is responsible for the observed reduction in ground state bias. In terms of relative energies (as measured by RMSRD), EOM-JAGP's performance in these molecules is intermediate between CIS and EOM-CCSD, as one would expect from a simple examination of the size of the methods' derivative subspaces.

3.4 Conclusions

We have discussed the equation of motion (EOM) linear response (LR) VMC method in this chapter. As an initial example, we pair the formalism with the Jastrow antisymmetric geminal power (JAGP) ansatz, whose EOM subspace we find to be intermediate in flexibility between that of configuration interaction singles (CIS) and coupled cluster singles

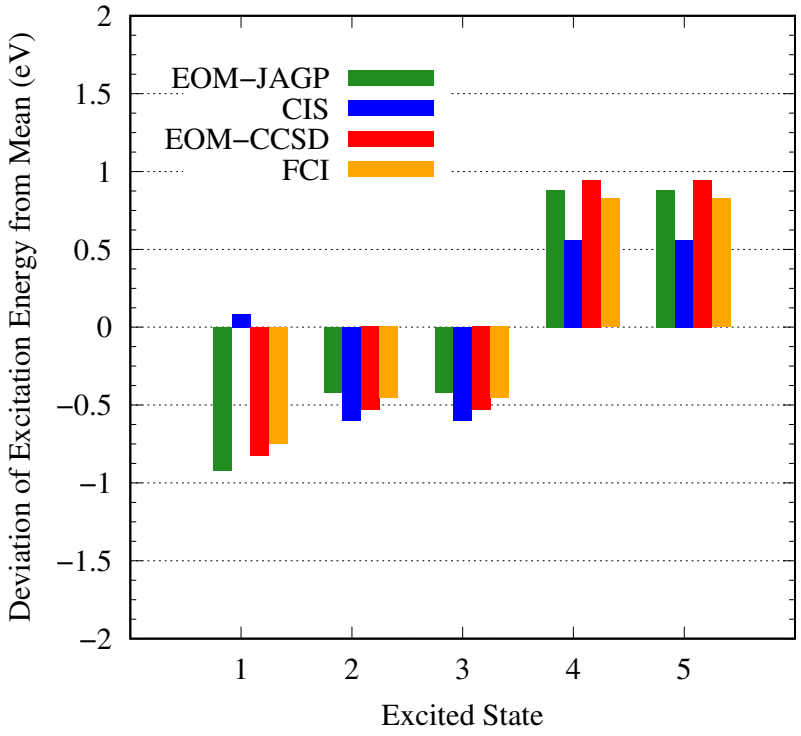

Figure 3.6: For each method, the deviation of each excitation energy from that method's mean excitation energy for C_2 in a 6-31G basis.

and doubles (CCSD). Somewhat counterintuitively, the unusually compact (compared to its high accuracy) nature of the JAGP for ground states leads to a general overestimation of excited state energies. Nonetheless, we find that in terms of relative energies between excited states, EOM-JAGP is as expected intermediate in accuracy between CIS and EOM-CCSD in single-reference systems, while performing much more reliably than CIS in the more multi-reference setting of the carbon dimer and displaying a similarly poor accuracy as EOM-CCSD for double excitations. This is in contrast to highly accurate predictions to the doubly excited states of JAGP using the new excited state variational principle in Chapter 2. As discussed, such inaccurate results are mainly due to the limited flexibility of the LR space of the JAGP ansatz, showing the necessity of using non-perturbative approaches to model excited states.

Even though it is difficult to recommend EOM-JAGP or EOM-VMC in general for quantitative predictions of excited states, the fact that it could yield qualitatively correct results makes it an excellent starting point for more powerful excited state methods, such as the excited states variational principle. For example, the recent development of variation-after-response (VAR) methods[87, 88] optimize individual excited state with EOM-VMC predictions as optimization starting point. These methods have been used to generate highly accurate descriptions to excited states with charge transfer characters, an important but chal-

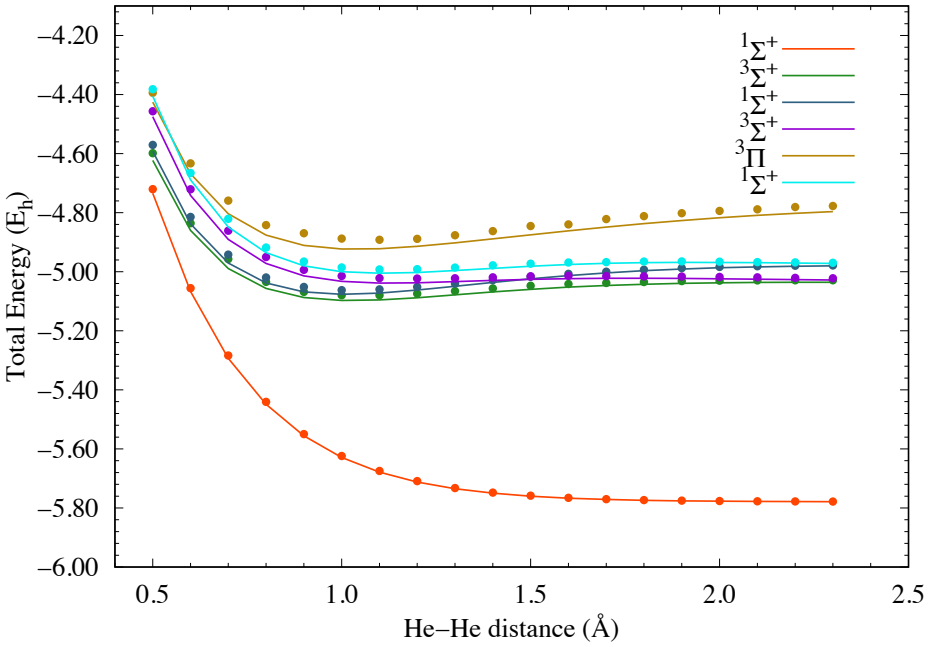

Figure 3.7: Potential energy curves for the lowest excited states of the helium dimer in the aug-cc-pVDZ [86] basis set. Solid lines correspond to FCI, dots to EOM-JAGP. The EOM-JAGP correctly finds the lowest three triplet and two singlet states.

lenging type of excitations for the widely used TDDFT methods. Furthermore, as shown by a recent study[89], the optimized wave functions produced by VAR method could also be used to reduce the fixed node error of DMC's predictions to excited states.

Table 3.3: Vertical excitation energies and RMSRD_N values (both in eV) for molecules at equilibrium in a 6-31G basis.

		Excitation Energy			
Molecule	State	EOM-JAGP	CIS	EOM-CCSD	FCI
Li_2	$^3\Sigma_u^+$	1.20	0.63	1.18	1.17
	$^3\Pi_g$	1.45	0.86	1.42	1.42
	$^1\Pi_g$	1.90	2.32	1.86	1.86
	$^3\Sigma_u^+$	2.27	1.67	2.23	2.23
	$^1\Sigma_u^+$	2.90	3.21	2.87	2.87
	RMSRD_5	0.00	0.47	0.00	0.00
N_2	$^3\Sigma_u^-$	7.89	6.10	7.97	7.87
	$^3\Pi_g$	8.81	7.31	8.01	7.94
	$^3\Sigma_g^+$	9.45	7.70	9.58	9.36
	RMSRD_3	0.39	0.52	0.06	0.00
CO	$^3\Pi$	6.57	5.66	6.16	6.01
	$^3\Sigma^-$	8.75	7.63	8.53	8.61
	$^1\Pi$	9.32	8.74	8.63	8.67
	RMSRD_3	0.22	0.43	0.10	0.00
CH_2	$^3B_1$	0.26	-0.74	-0.38	-0.38
	$^1A_1$	5.84	N/A	5.84	4.81
	RMSRD_2	0.20	N/A	0.52	0.00

Table 3.4: Vertical excitation energies and RMSRD_N values (both in eV) for molecules at equilibrium in a 6-31G basis.

		Excitation Energy			
Molecule	State	EOM-JAGP	CIS	EOM-CCSD	MRCI+Q
HCN	$^3\Sigma^-$	6.57	5.09	6.57	6.55
	$^3\Sigma^-$	8.20	6.26	8.18	8.02
	$^3\Sigma^-$	8.83	7.22	8.92	8.65
	$^1\Sigma^-$	9.42	7.22	9.45	9.22
	RMSRD_4	0.07	0.23	0.09	0.00
C_2H_2	$^3\Sigma_u^-$	5.74	4.60	5.71	5.72
	$^3\Sigma_u^-$	7.28	5.73	7.28	7.16
	$^1\Sigma_u^-$	8.37	6.61	8.47	8.30
	$^1\Delta_u$	8.70	7.08	8.81	8.64
	RMSRD_4	0.03	0.21	0.07	0.00
CH_2O	$^3B_2$	3.90	3.51	3.56	3.60
	$^1B_2$	4.55	4.32	3.96	3.95
	$^3A_1$	6.17	4.55	6.05	6.10
	$^3A_2$	9.18	9.75	8.54	8.42
	$^1B_1$	10.47	9.48	9.38	9.24
	RMSRD_5	0.40	0.93	0.08	0.00

Chapter 4

Variational Optical Gaps of Solids

4.1 Introduction

In this chapter we turn our attention to electronic excitations in solids, which remains a central topic in condensed matter theory due to their importance in the spectroscopic characterization of materials and in technological applications such as light harvesting. In the language of condensed matter physics, the electronic excitations are usually described by electron transitions from filled quasi-particle states to unfilled ones, and the band gap is computed as the energy difference between the quasi-particle states (VBM and CBM) involved in the transition. The exact quasi-particle states could be obtained by solving the quasi-particle equation[30, 90, 28],

$$h_0(\mathbf{r})\psi_s(\mathbf{r}) + \int \Sigma(\mathbf{r}, \mathbf{r}', \epsilon)\psi_s(\mathbf{r}')d\mathbf{r}' = \epsilon_s\psi_s(\mathbf{r}) \tag{4.1}$$

in which $h_0(\mathbf{r})$ is the one-body potential of the quasi-particle. $\Sigma(\mathbf{r}, \mathbf{r}', \epsilon)$ is the self-energy and contains non-trivial many-body effects of the quasi-particle. s is the index for quasi-particle states.

In practice, the exact self-energy is not known and approximate ones need to be applied. The KS-DFT's exchange-correlation (xc) potential is usually used as the simplest approximations to the self energy. The KS equations are[8],

$$(h_0(\mathbf{r}) + \upsilon_{xc}(\mathbf{r}))\psi_s^{(KS)}(\mathbf{r}) = \epsilon_s\psi_s^{(KS)}(\mathbf{r}) \tag{4.2}$$

comparing to Equation 4.1, we find the self-energy is approximated as,

$$\Sigma(\mathbf{r}, \mathbf{r}', \epsilon) \approx \upsilon_{xc}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}') \tag{4.3}$$

and the exact quasi-particle states are approximated as the KS one-electron states. However, because of the self-interaction error of KS-DFT, the VBM energy is overestimated and the predicted band gaps are usually underestimated even for simple, weakly correlated materials.

In order to fix the band gap problem of KS-DFT, one has to go beyond the xc-potential approximation to the self-energy. One way to do so is to apply the G_0W_0 approximation[28, 30, 91],

$$\Sigma(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int G_0(\mathbf{r}, \mathbf{r}', \omega + \omega') W_0(\mathbf{r}, \mathbf{r}', \omega') e^{i\omega'\eta} d\omega' \tag{4.4}$$

in which G_0 is the non-interacting single-particle Green's function[30],

$$G_0(\mathbf{r}, \mathbf{r}', \omega) = \sum_{n\mathbf{k}} \frac{\psi^{*(KS)}_{n\mathbf{k}}(\mathbf{r}')\psi^{(KS)}_{n\mathbf{k}}(\mathbf{r})}{\omega - \epsilon^{KS}_{n\mathbf{k}}} \tag{4.5}$$

in which n and $\mathbf{k}$ represent band index and reciprocal lattice vector, and W_0 is the dynamically screened Coulomb potential[30, 92],

$$W_0(\mathbf{r}, \mathbf{r}', \omega) = \int \epsilon^{-1}(\mathbf{r}, \mathbf{r}'', \omega) v_C(\mathbf{r}'', \mathbf{r}') d\mathbf{r}'' \tag{4.6}$$

in which v_C is the Coulomb potential and ϵ^{-1} is the inverse of the dielectric function computed using the random-phase approximation (RPA)[28, 92],

$$\epsilon(\mathbf{r}, \mathbf{r}', \omega) = 1 - \int v_C(\mathbf{r}, \mathbf{r}'') P_0(\mathbf{r}'', \mathbf{r}', \omega) d\mathbf{r}'' \tag{4.7}$$

in which $P_0(\mathbf{r}'', \mathbf{r}', \omega)$ is the polarizability within RPA,

$$P_0(\mathbf{r}'', \mathbf{r}', \omega) = -\frac{i}{2\pi} \int G_0(\mathbf{r}'', \mathbf{r}', \omega + \omega') G_0(\mathbf{r}', \mathbf{r}, \omega') d\omega' \tag{4.8}$$

The major advantage of the G_0W_0 approximation to self-energy comparing to the xc-potential approximation is that in G_0W_0 screened Coulomb interaction is used to describe the interactions between quasi-particles. Capturing the screening effect is of great importance. Since the quasi-particle is surrounded by particle cloud carrying the opposite charges, their interaction with other particles should be "screened" by the surrounding charge cloud.

The approximate self-energy defined in Equation 4.4, is called the G_0W_0 approximation. The quasi-particle energies could be obtained by performing first order perturbations to the KS-DFT one-particle energies[30, 91],

$$\epsilon^{QP}_{n\mathbf{k}} = \epsilon^{KS}_{n\mathbf{k}} + \left\langle \psi^{KS}_{n\mathbf{k}} | [Re\Sigma_{n\mathbf{k}}(\epsilon^{KS}_{n\mathbf{k}}) - v^{xc}_{n\mathbf{k}}] | \psi^{KS}_{n\mathbf{k}} \right\rangle \tag{4.9}$$

The quasi-particle energy difference between VBM and CBM defines the quasi-particle band gap: the excited electron and hole left behind do not interact with each other. For neutral excitations, one needs to solve BSE equations on top of the quasi-particle states to capture the interaction between particle and hole (exciton binding effect). Especially for simple semiconductors, such as diamond and silicon, G_0W_0 methods and BSE have been quite successful in terms of the predictions of band structure and optical spectrum.

As one can see from Equation 4.9, MBPT method in the form G_0W_0 is a first-order perturbation theory starting from the one-particle states of KS-DFT. Although BSE method is not an explicit perturbation approach, it takes the screened Coulomb interaction and quasi-particle energies predicted by G_0W_0 as its inputs. To be more specific, G_0W_0 and BSE assume a zeroth order picture in which the excitation is described by a single particle-hole transition within a given zeorth order quasi-particle basis. For the perturbation theory to work, such a zeroth order basis should qualitatively resemble the true quasi-particle states so that accurate results could be obtained by making small corrections to them with MBPT.

Although there is no requirement that MBPT should start from the one-particle states of DFT, in practice KS one-particle states are used in nearly all G_0W_0 applications. Consequently, MBPT would break down if the KS states differ significantly from the exact quasi-particle states. Zinc oxide represents a particularly difficult example for MBPT. It has been shown that its band gap is extremely hard to converge and also highly sensitive to the DFT exchange-correlation functionals. Such a problem is not unique in ZnO. G_0W_0 also fails to obtain accurate band gaps for strongly correlated transition metal oxides. Considering the importance of these materials in light harvesting and functional materials applications, having a non-perturbative method that yields quantitatively accurate predictions in cases where MBPT performs poorly is highly desired.

One may argue that the sensitivity of G_0W_0 with respect to different DFT functionals could be removed by solving the GW equations self-consistently[90]. Admittedly, self-consistent GW does carry a weaker dependence on the starting DFT states, but it does not usually yield improved predictions over G_0W_0. In addition, the method "self-consistent GW" is not as well-defined as G_0W_0. One could choose to self-consistently update the Green's function, the screened Coulomb interaction, or both, and the corresponding results will not be necessarily better than G_0W_0. Therefore, it is far too early to claim that self-consistent GW provides us with a universally better way of removing the drawbacks of G_0W_0, and alternative methods other than GW are still worth pursuing.

In this chapter, we will introduce a variational formalism that enables accurate and systematically improvable predictions of a material's lowest excited state wave function and the corresponding optical gap, which can be used as a stand-alone predictive tool and as a window into the relationship between density functionals and the assumed zeroth-order picture of MBPT. This approach combines the new excited state variational principle with the MSJ wave function ansatz, which is suitable for both the ground state and the VBM→CBM states. Crucially, the ansatz can describe both nontrivial BSE-like superpositions of particle-hole excitations and the dynamic polarizations of the electron cloud found in the vicinity of an exciton. Gap comparisons aside, the fact that the method yields an explicit wave function for the VBM→CBM excitation allows us to directly inspect how well a given density functional satisfies MBPT's zeroth-order picture and thus how likely it is that accurate predictions will result.

We will first discuss how the MSJ wave functions for CBM and VBM are constructed, along with various techniques used to maintain balanced descriptions between VBM and CBM. We will also present method to address the finite size effect. Numerical results are

presented for a variety of solids ranging from small gap semiconductors to large gap insulators. In the end, we will take the challenging case of ZnO as an example to provide insights into the zeroth order picture of MBPT.

4.2 Method and Computational Details

Construction of the Exciton Wave Function

We use the MSJ ansatz with the excited state variational principle. The MSJ ansatz and the excited variational principle are defined in Equation 2.20 and 2.3. For the ground state, we include the closed shell Kohn-Sham determinant for the basic ground state structure plus all single particle-hole excitations, which represent the leading order terms in a Taylor expansion of the orbital rotation that would transform the Kohn-Sham determinant into whichever determinant minimizes Ω in the presence of the correlation factor. For the excited state, we would like to include all single particle-hole excitations as in the BSE approach as well as the closed shell determinant and all double particle-hole excitations. A schematic plot of the single excitations are shown in Figure 4.1. This would again allow us to capture the leading order effects of an orbital rotation [93, 94] that would in this case accommodate repolarizations of the electron cloud in the vicinity of the exciton. However, as it is prohibitively expensive to include all double excitations in real materials, we approximate orbital relaxations by first minimizing Ω for singles and the closed shell term and then adding only those doubles that contain a singles component with coefficient larger than 0.1. A schematic plot of the double excitations is shown in Figure 4.2.

Variance Matching

In order to maintain balance between CBM and VBM so as to maximize cancellations of error, we also follow the recent approach of adjusting the flexibility of one of the wave functions in order to ensure that, as measured by σ^2, the ground and excited wave functions are of equal quality. While this variance matching could be achieved by limiting the flexibility of either the ground or the excited state, we have done so in this study by withholding enough high energy singles from the ground state such that its variance matches that of the excited state.

Addressing Finite Size Effects

Since QMC simulations of extended systems are performed in finite simulation cells, proper care needs to be taken to account for the interactions that come from electrons and ions outside the supercell. In order to address the finite size effects, we perform VMC calculations for LiH and LiF in simulation cells containing 2, 4, 8, 16, and 24 atoms, after which $1/N$ extrapolations were used to predict the band gap in the bulk limit. The same approach was

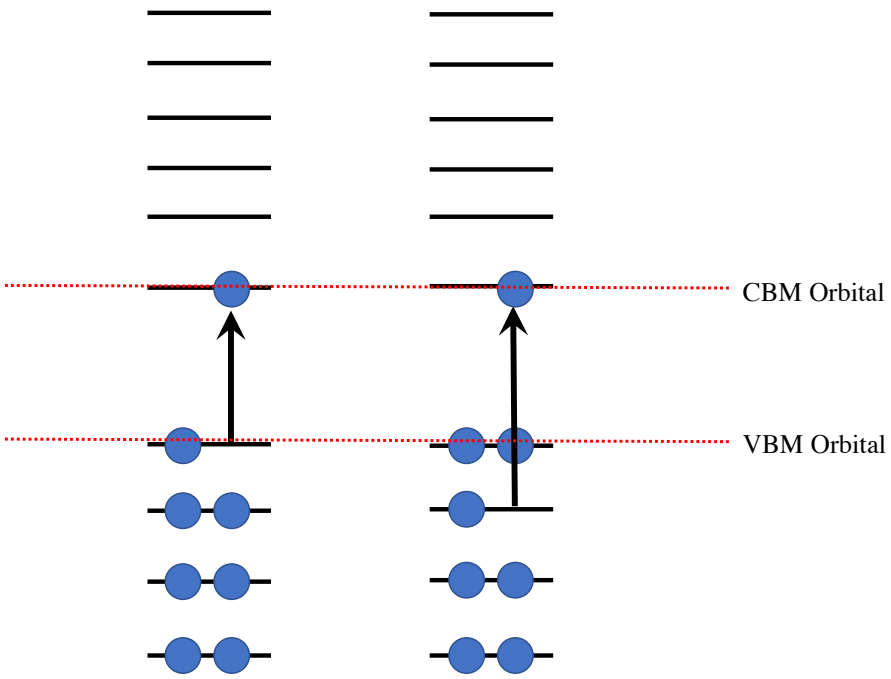

Figure 4.1: A schematic representation of single excitations used to construct the exciton wave function.

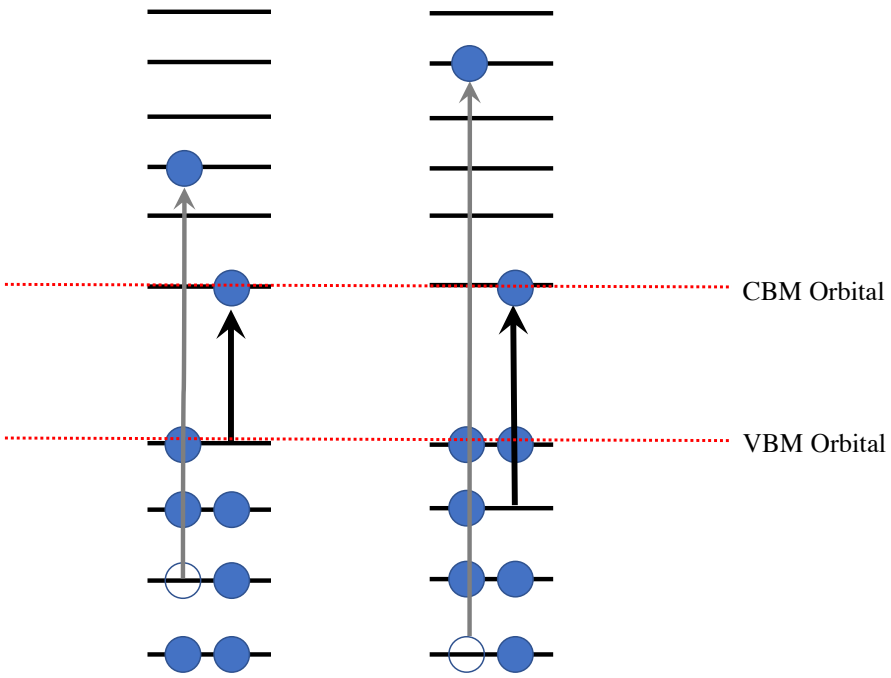

Figure 4.2: A schematic representation of double excitations used to construct the exciton wave function.

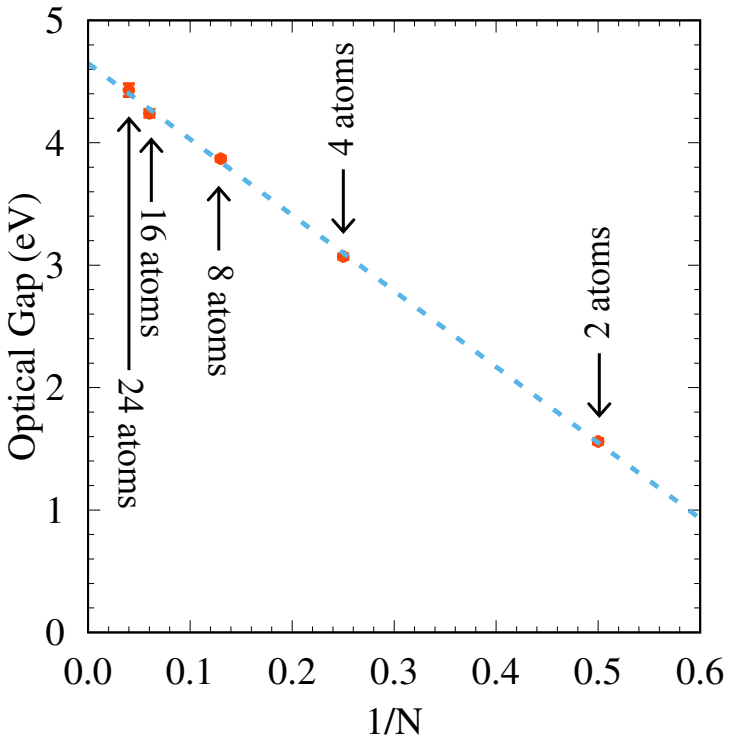

Figure 4.3: Extrapolation to the bulk limit in LiH. Points are our VMC data while the dashed line is a linear fit against the inverse of the number of atoms N in the simulation cell.

used for C diamond and Si, but with 8, 16 and 24 atom simulations cells. See Figure 4.3 for an example extrapolation of LiH. The extrapolation of Si, C, and LiF could be found in the Appendix.

Due to the high cost of simulating the semi-core electrons of Zn, which was necessary to produce physically reasonable results, we were limited by our current software implementation to a maximum of 8 atoms in our simulation cell for ZnO, which did not permit us to perform the same type of finite size correction as for the systems above. Instead, we have derived a simple finite size correction based on previous diffusion Monte Carlo (DMC) work[95] in which nodal surfaces for both the CBM and VBM were constructed using a simple single-Slater model. The previous study[96] reports results for a 48 atom simulation cell, and so we have performed the equivalent single-Slater DMC calculations for our 8 atom cell and used the difference in the DMC gap at these two cell sizes to provide an approximate finite size correction for our 8 atom VMC gap. Note that this approach has no effect on our conclusions with regard to either the nature of the first excited state under different density functionals or the insensitivity of our VMC gap predictions to the choice of functional, as these properties are entirely determined within our 8 atom VMC evaluations.

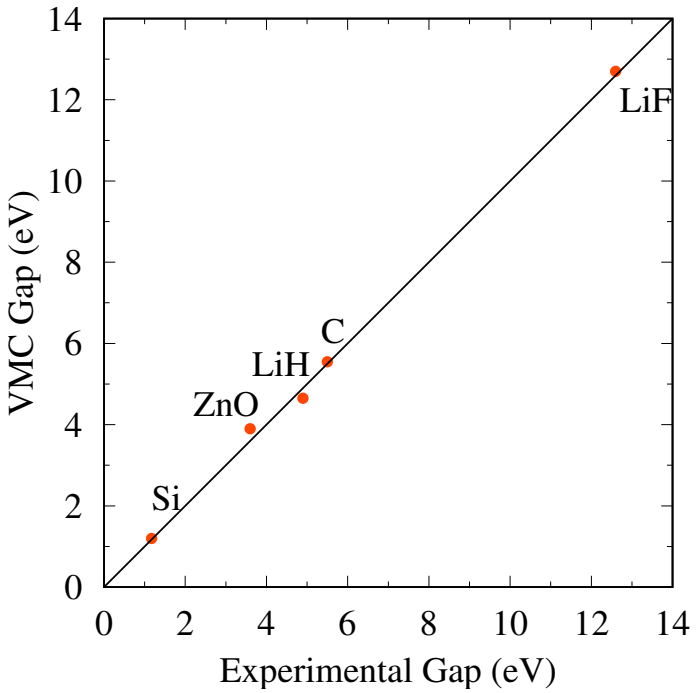

Figure 4.4: VMC-CISD optical gap predictions plotted against experimental results. See Table 4.1 for more details.

4.3 Results

Optical Gaps

As seen in Figure 4.4 and Tables 4.1 and 4.2, the approach in which we include both singly and doubly excited configurations in the excited state (VMC-CISD) is quite effective for predicting optical gaps in small (Si), medium (C, LiH, ZnO), and large (LiF) band gap materials. Its mean-absolute-deviation (MAD) from experimental values across these five systems is just 3.5%, compared to MADs more than twice this large for the optical gaps obtained by subtracting the known exciton binding energies from G_0W_0 and self-consistent GW gaps. Of course, MBPT is highly effective in Si, C, and LiH, and so we expect that in these cases the zeroth order DFT wave function is sound. The analysis in Figure 4.5 confirms this expectation by showing that over 90% of the VMC-CISD wave function is accounted for by LDA's VBM→CBM transition. Thus, in these three cases, LDA provides good zeroth order wave functions and we can confirm that the accuracy of MBPT derives from the appropriateness of its approximation.

The story is quite different in LiF and ZnO, where Figure 4.5 reveals that LDA's zeroth order picture accounts for less than 80% of the high-level wave function. At a minimum, this implies that LDA's VBM and CBM orbitals are not the correct shape for the real exciton's particle and hole, a point we will return to in our discussion of ZnO. Figure 4.5 also reveals

Table 4.1: Band gaps in eV. The quasiparticle gaps of DFT and the *GW* methods should be reduced by the EBE when comparing to the VMC and experimental optical gaps.

	C	Si	LiH	LiF
LDA	3.93	0.47	2.68	8.60
G_0W_0	5.50[35]	1.12[35]	4.64[97]	13.27[35]
GW	5.99 [35]	1.28 [35]	4.75 [98]	15.10 [35]
VMC-CIS	5.68(6)	1.41(6)	5.01(6)	14.6(1)
VMC-CISD	5.55(6)	1.20(6)	4.65(6)	12.7(1)
Experiment	5.50 [99]	1.17 [99]	4.90 [100]	12.6 [101]
EBE	0.07 [102]	0.015 [103]	0.1 [104]	1.6 [101]

Table 4.2: ZnO band gaps and EBE in eV.

LDA	0.83
PBE0	3.03 [105]
G_0W_0-LDA	2.14 [35]
GW-LDA	3.20 [35]
G_0W_0-PBE0	3.24 [105]
VMC-CIS(LDA)	3.9(2)
VMC-CIS(PBE0)	4.6(2)
VMC-CISD(LDA)	3.9(2)
VMC-CISD(PBE0)	3.9(2)
Experiment	3.6 [106]
EBE	0.06 [107]

that in these two systems, the fraction of exact exchange can have a significant effect on how closely DFT's zeroth order wave function matches the VMC prediction. Although there are also the orbital energies to consider (see ZnO discussion below), these findings help explain why MBPT can be so sensitive to the choices made in its practical application. [36, 108, 109] Work by Sommer et al [110] reveals that these issues can carry over to the BSE approach, which fails to provide a satisfactory correction to GW in LiF, although vertex-corrected solutions to Hedin's equations can help in that case. [111, 112] Note that these issues do not necessarily imply a failure of one-particle theory in these systems, as there may exist a one particle basis in which the true exciton really does look like the simple VBM→CBM transition. Indeed, in ZnO, to which we will now turn our attention, we will provide an analysis showing that such a basis does indeed exist. Thus, while Figure 4.5 makes plain that commonly used density functionals struggle to meet the needs of MBPT in both ZnO and LiF, the insights gleaned from systematically improvable wave function methods should help resolve this difficulty in future.

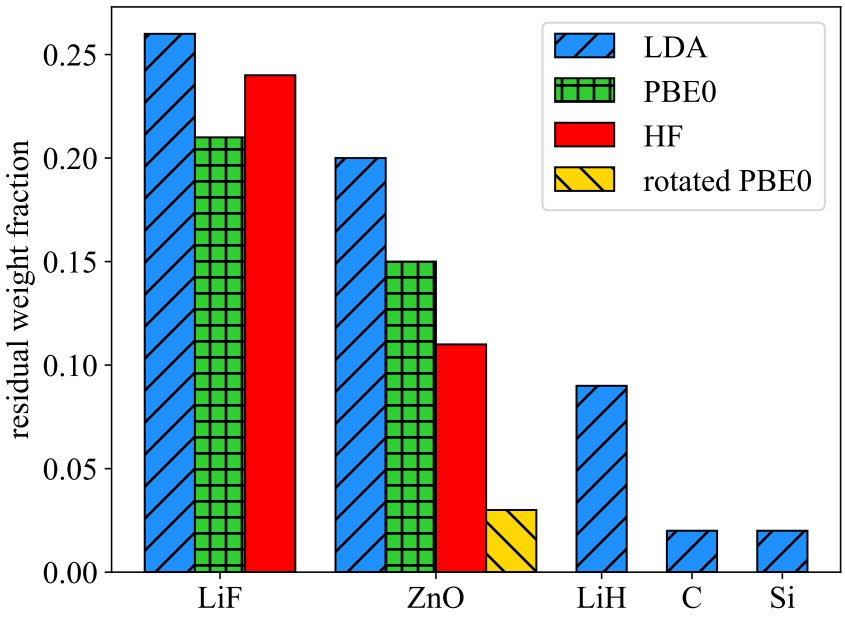

Figure 4.5: Here we investigate the appropriateness of various one-particle orbital sets for MBPT by plotting VMC-CISD's residual weight fraction, which we define as the sum of squared CI coefficients on all configurations other than the primary VBM→CBM transition when working in a particular orbital basis. In cases where degeneracy in the VBM leads to multiple equal-energy VBM→CBM configurations, the sum excludes all such configurations.

Wave Function Analysis of ZnO

ZnO represents a particularly difficult case for MBPT, especially when considering its low-order and highly-efficient G_0W_0 variant. [35, 36] The left hand side of Figure 4.6 makes clear that the accuracy of this low order perturbative treatment is highly sensitive to the inclusion of exact exchange. In contrast, we see that the VMC-CISD results are insensitive to whether we employ the LDA, PBE0, or even the Hartree Fock (HF) one-particle basis sets. The reasons for this success are two-fold. First, the wave function was designed so as to be able to approximate an orbital rotation in order to counteract shortcomings in the starting DFT orbitals. Indeed, if we remove this ability by removing the doubles excitations from the excited state and the singles from the ground state, the resulting VMC-CIS results are more sensitive and less accurate overall, as seen in Tables 4.1 and 4.2. Second, VMC takes the issue of the DFT orbital energies off the table entirely, as it directly evaluates the energy expectation value of its wave function using the full ab initio Hamiltonian so that the only dependence on DFT is via the shapes of the one-particle orbitals.

Using our DFT-insensitive VMC methodology as a guide, one can investigate how commonly used density functionals' zeroth order pictures deviate from reality in ZnO and whether

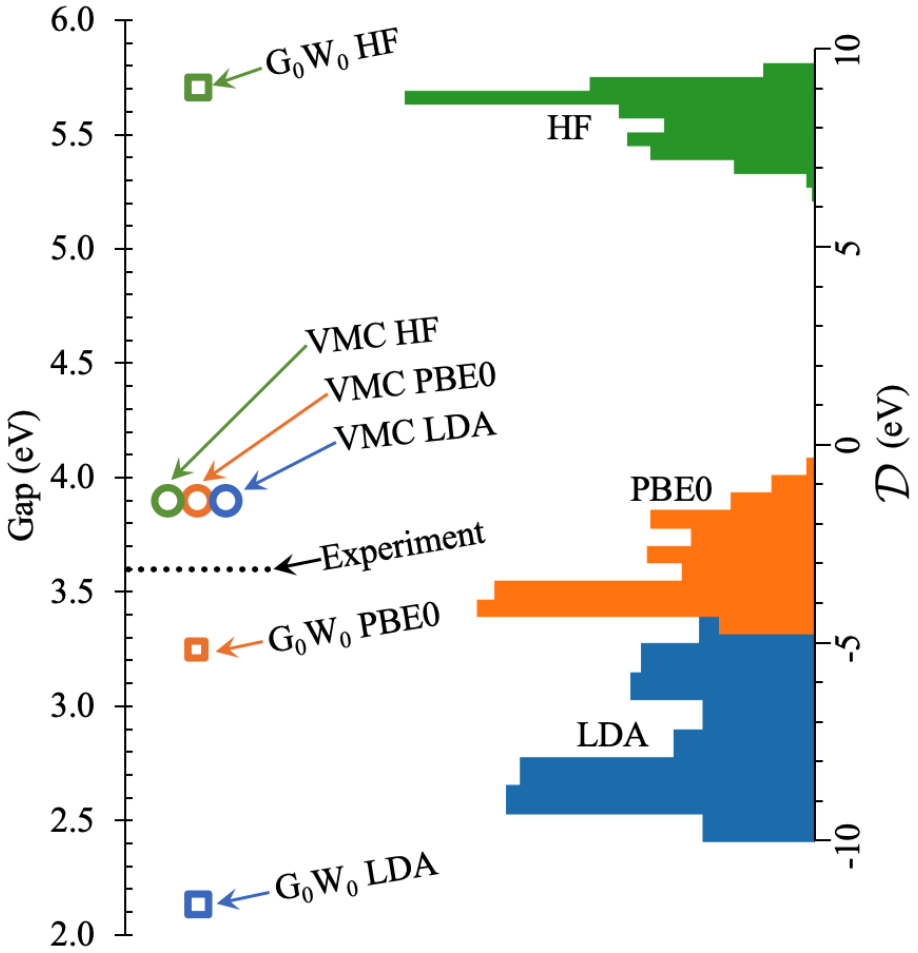

Figure 4.6: Optical gap and single-particle transition energy data for ZnO. On the left, we compare G_0W_0 fundamental gaps using one-particle starting points that employ different fractions of exact exchange with our VMC-CISD optical gaps based on the same starting points. For the various $i \to a$ transitions, we plot on the right histograms of the differences $\mathcal{D}_{ia} = \Delta_{ia}^{\mathrm{DFT}} - \Delta_{ia}^{\mathrm{VMC}}$ between the DFT estimates (i.e. the orbital energy differences $\Delta_{ia}^{\mathrm{DFT}} = \epsilon_a - \epsilon_i$) for the energy cost of promoting an electron from orbital i to orbital a and the analogous quantities $\Delta_{ia}^{\mathrm{VMC}}$, which are the VMC energy differences between the $i \to a$ excited and the ground state Jastrow-modified Slater determinants. G_0W_0 data from Fuchs. [105] Experimental result from Lauck. [106]

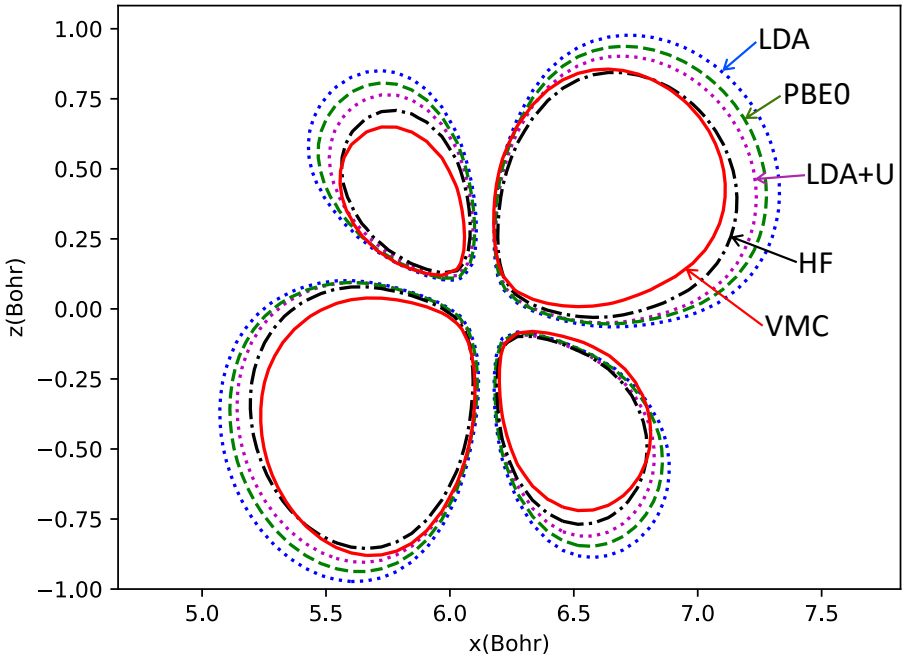

Figure 4.7: A cut along ZnO's $(\bar{1}2\bar{1}0)$ plane in which we investigate the lowest energy excitation's hole density in the vicinity of the Zn atom. For each method, we plot the contour along which the number of holes per $Å^3$ is equal to 1.2.

it is even possible to construct a one particle picture upon which MBPT should be reliable in this material. First, we stress that although Figure 4.5 revealed that G_0W_0's sensitivity to exact exchange is likely due in part to the varying quality of the zeroth order wave functions, the right hand side of Figure 4.6 emphasizes the importance of the zeroth order transition energies and how they are also quite sensitive to exact exchange. By considering zeroth order wave functions and transition energies together, we gain an appreciation for how challenging this system is for density functional theory. Indeed, HF theory with its 100% exact exchange gives better orbitals for the purpose of describing the first excited state, but its transition energies are grossly too high, whereas PBE0 has better transition energies but worse orbitals. Among the three options of LDA, PBE0, and HF, PBE0 clearly makes for the best compromise between wave function and transition energy accuracies, but our results suggest that both its energetics and orbitals would be improved in ZnO with a higher fraction of exact exchange.

With an explicit high-level wave function in hand, we can ask highly detailed questions about the exciton, such as to what degree the O 2p and Zn 3d orbitals hybridize in the hole state. Indeed, metal-oxide over-hybridization has been pointed out as a key deficiency in LDA and other pure functionals. [36] We approach this question by performing a density matrix difference analysis [113] in which the difference between the one-body density matrices of our VMC-CISD excited and ground state is diagonalized. As occurs for any excited

state consisting of a single one-particle transition of the type assumed in MBPT's zeroth order picture, [113] the resulting eigenvalues are all close to zero (absolute values less than 0.1) except for one with a value near 1 and one with a value near -1. The eigenvectors corresponding to these two large eigenvalues are the attachment and detachment orbitals, respectively, and represent the particle and hole orbitals that most closely represent the transition between a correlated many-body ground state and excited state. By plotting the hole density from this detachment orbital in the vicinity of the Zn atom alongside the hole densities predicted by the VBM of different density functionals, Figure 4.7 makes clear that, compared to our high-level VMC results, LDA does indeed include too much Zn character in the VBM through over-hybridization. More surprisingly, we see that although LDA+U [114] with the U value used previously [36] does decrease the degree of hybridization, our detachment density is even less hybridized, with LDA+U bringing us only about halfway in between the LDA and VMC extremes. Another important point that the detachment orbital reveals is that some hybridization is definitely present, just not so much as common density functionals, even those specifically designed to address this issue, predict.

Although it is frustrating that current functionals face the various difficulties discussed above, the fact that the VMC density difference analysis strongly resembles a simple single-particle transition suggests that it should be possible to design a functional that delivers an excellent zeroth order starting point for MBPT. To make this idea more concrete, we can test whether such an orbital basis exists by applying an orbital rotation to our wave function (starting with the optimized VMC-CISD state in the PBE0 orbitals) in order to minimize the residual weight fraction of the exciton. As seen in Figure 4.5, this rotated PBE0 one-particle basis matches the assumptions of MBPT in ZnO almost as well as the LDA basis does for Si or diamond. This finding also serves to reassure us that the error we do see in VMC-CISD's optical gap prediction (and its moderate disagreement with previous projector Monte Carlo estimates [115, 116, 96]) is most likely due to the imperfect nature of our finite size correction rather than to the appropriateness of our wave function approach, as it validates the assumption that the excitonic state is dominated by single particle-hole transitions with the doubles only contributing small corrections. While a good one-particle basis is just a start (density functionals must also produce reasonable zeroth order transition energies) the insights we now have from VMC paint a bright picture for the prospects of increasing the accuracy and reliability of MBPT in cases like ZnO.

4.4 Conclusions

We have shown that the new excited state variational principle can be combined with simple, physically-motivated wave function approximations to evaluate optical band gaps in a way that is both insensitive to the DFT starting point and informative about the assumptions of MPBT. Given the dominant role that MBPT plays in the theoretical interpretation of materials spectroscopy, a method that is able to improve its predictive power has the potential to be highly impactful. Even in cases where exciton-induced repolarization effects are

large and it is not possible to identify a density functional that yields a one-particle basis appropriate for describing both the ground and the low-lying conduction band states, the ability to provide variational predictions of band-edge energies, perhaps even in a k-point-by-k-point fashion, would create the possibility of developing first-principles-based scissors corrections for the BSE Hamiltonian, a practice that at present can be quite effective when performed empirically. [107, 117] In molecular excitations, variational excited states [41, 67, 75, 118] and MBPT [119] have so far been explored separately, but the same potential for strong synergies is present. In both molecules and solids, our approach also provides a reasonably black-box route to producing high-quality nodal surfaces for excited states in diffusion Monte Carlo, which even with less sophisticated VMC preparations has already shown promise in evaluating band gaps. [120, 121, 122, 95, 123, 96] The prospects for increased accuracy and scalability in this area are especially bright in light of recent progress in VMC methods for optimizing the one-particle basis [124, 125] and achieving compact representations of excited states, [88, 89] not to mention the rapid progress in selective CI methods that synergize strongly with multi-Slater VMC. [126, 127, 128, 129, 130, 131, 132, 133] With this wide range of promising connections, we look forward to further exploring the role that variational approaches can play in deciphering and designing molecular and materials spectra.

Chapter 5

Blocked Linear Method for the Optimization of Large Parameter Sets

5.1 Introduction

The excited state variational principle, as an ansatz-based approach to excited states, needs to optimize the ansatz parameters for a particular system. The capabilities of the method used to optimize the ansatz are every bit as important as the flexibility of the ansatz itself. For example, both the coupled cluster [134] and matrix product state [64] ansatzes would be much less useful if we lacked the projected Schrödinger equation and DMRG methods that allow us to optimize them efficiently. To address unsolved problems in electronic structure — such as catalytic cycles in which many bonds are simultaneously rearranged [135], double excitations in large π-conjugated molecules[136], and high-temperature superconductivity [137, 138] — it is therefore essential that improvements to optimization methods be made alongside innovations in ansatz design.

The excited state variational principle relys on QMC for its evaluation and optimization, an area in which the need for improved optimization methods is especially pressing. Until very recently, optimization methods in this area were limited to a few thousand variational parameters when using a fully *ab initio* Hamiltonian, a constraint that holds back progress in a wide variety of areas. In fixed-node projector Monte Carlo methods such as diffusion Monte Carlo (DMC) [7, 139, 140], the inability to systematically converge the trial function's nodal surface due to insufficiently flexible ansatzes is responsible for both the fixed node error and the pseudopotential locality error, the latter of which becomes acutely problematic in 3rd-row and heavier elements where the nonlocal part of the pseudopotential cannot be ignored. Even in variational Monte Carlo [7, 140] (VMC) itself, recent innovations in ansatz design create a pressing need for expanding the number of variational parameters that can be treated. Examples in this category include the variation after response approach to excited states [141], efficient methods for large multi-Slater Jastrow (MSJ) expansions [73, 56, 124], variational analogues of coupled cluster theory [142], and wave function stenciling approaches

[143, 78, 144] that tightly couple the optimization of correlation factors and molecular orbitals. For all of these reasons, and indeed for the simple reason of enabling systematic improvability within a given ansatz, improvements in VMC optimization capabilities are sorely needed.

The linear method [42, 43, 44, 45] (LM) developed by Umrigar and coworkers is currently the most effective VMC optimizer for cases in which the number of variables is a few thousand or less. By solving a projected Schrödinger equation in the vector space spanned by the current wave function and its first parameter derivatives, a space we will refer to as the self-plus-tangent space, the LM produces update steps that account for second order couplings between variables and in practice often out-perform Newton-Raphson steps, a success due in no small part to the fact that these updates satisfy a strong zero variance principle [44, 45]. However, the standard LM's need to explicitly construct the Hamiltonian and overlap matrices in the self-plus-tangent space becomes cumbersome when the number of variational parameters exceeds a few thousand due to the large amounts of memory required to store these matrices. This issue becomes especially fraught when trying to match the LM to modern supercomputing resources, as each parallel Markov chain must make space for its own copies of these matrices (a tall order given typical per-core memory restrictions), which must then be communicated and combined prior to diagonalization. While one could use Krylov subspace methods to solve the eigenproblem without explicitly constructing the matrices, as was done for the related stochastic reconfiguration (SR) method [145], our experience in practice has taught us that finding a preconditioning scheme capable of reducing the condition numbers of the LM matrices to manageable levels is not trivial. As far as we are aware, these various issues have prevented the LM from being used in regimes beyond about 16,000 variables, which occurred in the context of a ground state MSJ expansion for the water molecule [73].

Very recently, Booth and coworkers introduced an alternative VMC optimization method that takes advantage of optimal descent theory and a stochastic gradient evaluation in order to produce robust energy minimizations despite avoiding second-derivatives entirely [144]. Impressively, this method appears capable of handling more than 60,000 variational parameters for ansatzes that support efficient inner products with the Slater determinant basis functions of Fock space. Recently, similar accelerated gradient descent method has been implemented in real-space VMC[47]. Although stochastic gradient descent is very general and requires significantly less amount of memory than LM, they are also less effective in optimizing highly non-linear wave function parameters, such as orbital rotation parameters and number counting Jastrow factors[48]. It is shown that the optimization results depend strongly on the hyper-parameters of the algorithm such as the step size[47]. Therefore, promising new directions in VMC optimization are a welcome development.

In the present chapter, we seek to retain the advantages of the standard LM — which include Fock space and real space compatibility, robust convergence in a small number of iterations, and access to excited states through our recently introduced [41] excited state variational principle — while reducing its memory footprint so as to facilitate larger variable sets and better compatibility with modern parallel computers. Our strategy will be

to separate the variable space into blocks, within each of which we estimate a small number of important update directions that can then be used to construct a relatively small LM eigenproblem in the overall basis of important directions. We will demonstrate that this approach drastically reduces memory requirements without significantly affecting the accuracy of the optimization. In addition to tests on small molecules using our in-house Hilbert space software, we will use the implementation that we recently contributed to the open-source QMCPACK software package [146, 147] to demonstrate this method's excited state capabilities in the context of a hydrogen ring's Mott-like metal-insulator transition. By evaluating the optical gap for a series of increasingly flexible MSJ expansions, the largest of which contains over 25,000 variational parameters, this study points the way towards a systematically convergent and non-perturbative approach to predicting optical gaps in the Mott-insulating regimes of real materials.

5.2 Theory

The Linear Method

The standard LM works by repeatedly solving the Schrödinger equation in the self-plus-tangent subspace of the full Hilbert space, defined by the span of the wave function and its first derivatives with respect to its variational parameters. As the derivatives are not necessarily orthogonal to each other, this approach leads to a generalized eigenvalue problem

$$\sum_{j\in\{0,1,\ldots\}} \left\langle \Psi^i | H | \Psi^j \right\rangle a_j = \lambda \sum_{j\in\{0,1,\ldots\}} \left\langle \Psi^i | \Psi^j \right\rangle a_j \tag{5.1}$$

where $|\Psi^i\rangle$ and $|\Psi^j\rangle$ are the derivatives of $|\Psi\rangle$ with respect to the ith and jth wave function parameters μ_i and μ_j, respectively, and $|\Psi^0\rangle \equiv |\Psi\rangle$. After solving this eigenvalue problem for $\vec{c}$, one updates the parameters by

$$\mu_i \rightarrow \mu_i + a_i/a_0 \quad \forall \quad x \in \{1, 2, ...\} \tag{5.2}$$

after which the updated $|\Psi\rangle$ will be a good approximation for the subspace eigenfunction $\sum_j c_j |\Psi^j\rangle$ so long as the updates a_i/a_0 are sufficiently small in magnitude. This requirement can be ensured by applying a diagonal shift to the Hamiltonian matrix [45], which plays the same role as a trust radius would in a Newton-Raphson optimization. The updated ansatz in hand, a new self-plus-tangent space may be constructed and the procedure repeated until convergence is reached.

In practice, the Hamiltonian and overlap matrix elements are estimated via Monte Carlo

sampling,

$$\begin{aligned}&\sum_{\vec{n}\in\xi}\sum_{j\in\{0,1,\ldots\}}\frac{|\langle\vec{n}|\Psi\rangle|^2}{\mathcal{P}(\vec{n})}\frac{\langle\Psi^i|\vec{n}\rangle}{\langle\Psi|\vec{n}\rangle}\frac{\langle\vec{n}|H|\Psi^j\rangle}{\langle\vec{n}|\Psi\rangle}a_j\\&=\lambda\sum_{\vec{n}\in\xi}\sum_{j\in\{0,1,\ldots\}}\frac{|\langle\vec{n}|\Psi\rangle|^2}{\mathcal{P}(\vec{n})}\frac{\langle\Psi^i|\vec{n}\rangle}{\langle\Psi|\vec{n}\rangle}\frac{\langle\vec{n}|\Psi^j\rangle}{\langle\vec{n}|\Psi\rangle}a_j\end{aligned}\tag{5.3}$$

where ξ is a set of samples drawn from the probability distribution $\mathcal{P}(\vec{n})$ (which is typically chosen as $|\langle\vec{n}|\Psi\rangle|^2$) using Markov chain Monte Carlo. Note that although we have depicted the sampling as running over occupation-number-vector-labeled determinants in Fock space, the LM is equally viable if instead the sampling is carried out in real space, where $\mathcal{P}(\vec{r})$ is typically chosen to be $|\Psi(\vec{r})|^2$. The LM will thus be efficient (i.e. polynomial cost) for ansatzes that support the efficient evaluation of the derivative ratios $\langle\vec{n}|\Psi^i\rangle/\langle\vec{n}|\Psi\rangle$ and $\langle\vec{n}|H|\Psi^i\rangle/\langle\vec{n}|\Psi\rangle$, examples of which include MSJ expansions [43, 73, 56], the Jastrow antisymmetric geminal power [57, 60, 61] (JAGP), and amplitude determinant coupled cluster with pairwise doubles [142].

While the cost scaling may be polynomial with system size, the memory required to store the Hamiltonian and overlap matrices in the self-plus-tangent space can be a serious impediment to practical computation. For example, when using 8-byte floating point numbers and an ansatz with 30,000 variational parameters, the standard LM requires 14.4 gigabytes of memory per Markov chain. Such storage requirements create problems with the typical parallelization scheme of running one Markov chain per core, as modern supercomputers typically have closer to 2 gigabytes of memory available per core.

One approach to circumventing matrix storage difficulties would be to use a Krylov subspace method to solve for $\vec{c}$ without constructing the matrices explicitly. While this strategy has shown promise in the related SR method, where it succeeded in working with an ansatz containing half a million variables [145], Krylov subspace methods are only efficient if the condition numbers of the matrices involved (the ratio of the magnitudes of their largest and smallest magnitude eigenvectors) can be brought close to unity through preconditioning. Usually, the matrices encountered in the LM are ill-conditioned since most ansatzes used in QMC contain non-linear parameters, such as Jastrow factors, orbital rotation parameters and pairing matrix elements in JAGP, and the wave function first derivative vectors with respect to these parameters are not necessarily orthogonal therefore could contain linearly dependencies. Although we have made some ad-hoc investigations into this area, such as applying the so-called "Subspace Projected Approximate Matrix (SPAM)" modification to the Davidson Method[148] and normalizing LM derivative vectors before any matrix operations, we have not found preconditioners that can reliably reduce the condition numbers involved below about 10^{10}. This reality mattered less in SR, as in that case each Krylov subspace iteration requires only effecting an overlap matrix multiplication and so does not involve the Hamiltonian operator [145]. In the LM, both overlap and Hamiltonian matrix multiplications must be effected for each Krylov iteration, which for a fully ab initio Hamiltonian greatly increases the cost per Krylov iteration. Thus, in practice, the cost of the

Krylov approach in the LM appears to be more sensitive to condition number than for SR. While this does not preclude the existence of an effective preconditioning scheme, it does prompt us to investigate approaches, like the one in the next section, that remain effective even in the face of highly ill-conditioned matrices.

The Blocked Linear Method

Ultimately, the goal of the LM is to find the best update direction and step length within the tangent space of the wave function. Imagine instead holding half the variables fixed and inspecting the tangent space for the other half. The diagonalization of the linear method eigenproblem within this self-plus-half-tangent space will produce a set of update directions that can be ordered by importance, as measured by their eigenvalues, which inform us as to how much a move along an eigen-direction would decrease or increase the energy. Noting that the optimal direction $\vec{c}_{\text{opt}}$ in the full tangent space, whose dimension is the total number of variational parameters N_V, can be written as a linear combination of $N_V/2$ orthogonal directions within one half-tangent space and $N_V/2$ orthogonal directions from the other half-tangent space, it seems intuitive that a very bad update direction in one of the half-tangent spaces is unlikely to be an important component of $\vec{c}_{\text{opt}}$. Taken further, this logic suggests that it may be possible to construct a close approximation to $\vec{c}_{\text{opt}}$ using a linear combination of only a few update directions from each half-tangent space. In essence, the blocked linear method (BLM) is an attempt to systematically exploit this structure by (a) dividing the variable space into a number of blocks, (b) making intelligent estimates for which directions within those blocks will be most important for constructing $\vec{c}_{\text{opt}}$, and (c) estimating $\vec{c}_{\text{opt}}$ by solving a smaller, more memory-efficient eigenproblem in the basis of these supposedly important block-wise directions.

Rather than the standard LM's expansion of the wave function in its self-plus-tangent space, consider instead the "one-block" expansion

$$|\Phi_b\rangle = \alpha_b|\Psi\rangle + \sum_{i=1}^{M_b} \beta_{bi}|\Psi^{i,b}\rangle + \sum_{j=1}^{N_O}\sum_{\substack{k=1\\k\neq b}}^{N_B} \gamma_{bjk}|\Theta_{jk}\rangle. \tag{5.4}$$

In the first two terms, we have a linear expansion of the wave function with respect to the variables belonging to the bth block, with α_b and β_{bi} the expansion coefficients, M_b the number of variables in the block, and $|\Psi^{i,b}\rangle$ defined as the wave function derivative with respect to the ith variable of the bth block. If we drop the third term for now (i.e. set $\gamma_{bjk} = 0$), we have a wave function whose energy minimization

$$\min_{\alpha,\beta} \ \langle\Phi_b|\hat{H}|\Phi_b\rangle/\langle\Phi_b|\Phi_b\rangle \tag{5.5}$$

leads to a generalized eigenvalue problem in the same form as for the standard LM, Eq. (5.1), the only difference being that we are now holding the variables outside the chosen block fixed. (Note that while we will develop the discussion here in terms of energy minimization, the

BLM is equally applicable to the target function used in the direct, variational targeting of excited states [41] and has been implemented and tested for both cases). Each eigenvector will have its own values for the α_b and β_{bi} coefficients and will correspond to an eigenvalue that gives an estimate for what the energy of our original wave function would be if we were to update this block's variables according to $\mu_{i,b} \rightarrow \mu_{i,b} + \beta_{bi}/\alpha_b$. Thus, the eigenvalues of this block's eigenproblem inform us as to which directions in its variable space are expected to be "good" update directions (those with the lowest eigenvalues) and which are expected to be "bad" directions (those with the highest eigenvalues).

Having performed this diagonalization within each of our blocks, we are now in a position to construct an approximation to the wave function in its full self-plus-tangent space by retaining from each variable block only a small number of what are expected to be the best update directions. By organizing the best N_K update directions from the bth block into the rows of a matrix $\boldsymbol{B}^{(b)}$, this self-plus-tangent space approximation can be written as

$$|\Upsilon(\alpha, \boldsymbol{A})\rangle = \alpha|\Psi\rangle + \sum_{b=1}^{N_B}\sum_{j=1}^{N_K} A_{bj} \sum_{i=1}^{M_b} B_{ji}^{(b)} |\Psi^{i,b}\rangle. \tag{5.6}$$

As the elements of the $\boldsymbol{B}$ matrices are now held fixed, this expansion is not as flexible as that of the standard LM, but we hope the fact that it is built out of a linear combination of the best update directions from each block will give it the correct flexibility to closely approximate the optimal update direction in the full tangent space. This direction is now estimated via

$$\min_{\alpha,\boldsymbol{A}} \ \langle\Upsilon|\hat{H}|\Upsilon\rangle / \langle\Upsilon|\Upsilon\rangle \tag{5.7}$$

which again produces a generalized eigenvalue problem, this time of dimension $1 + N_B N_K$, whose lowest energy eigenvector corresponds to the overall BLM update,

$$\mu_{i,b} \rightarrow \mu_{i,b} + \frac{[\boldsymbol{A}\boldsymbol{B}^{(b)}]_{bi}}{\alpha}. \tag{5.8}$$

Crucially, the Hamiltonian and overlap matrix elements involved in the eigenvalue problems that stem from Eqs. (5.5) and (5.7) can be estimated using the same information as in the standard LM, namely the derivative ratios $\langle\vec{n}|\Psi^x\rangle \,/\, \langle\vec{n}|\Psi\rangle$ and $\langle\vec{n}|H|\Psi^x\rangle \,/\, \langle\vec{n}|\Psi\rangle$, at each sampled configuration $\vec{n}$ (or position $\vec{r}$ in real space). While the most efficient way to construct these matrices now that the $\boldsymbol{B}^{(b)}$ coefficients are known appears to be to re-run the same sample that was used to construct the block-specific matrices, we feel that this second sampling is a price worth paying in order to remove the standard LM's memory bottleneck.

So far, we have ignored the fact that inter-block variable couplings will affect which directions in a block are optimal for use in constructing an overall update direction. Accounting for such couplings is the purpose of the third term in Eq. (5.4), in which

$$|\Theta_{jk}\rangle = \sum_{l=1}^{M_k} D_{jkl} |\Psi^{l,k}\rangle \tag{5.9}$$

is a linear combination of wave function derivatives from the kth block that is presumed to correspond to a good update direction for that block. By including a small number N_O of these directions from each other block in the wave function expansion $|\Phi_b\rangle$ for the current block, we hope to provide the minimization

$$\min_{\alpha,\beta,\gamma} \langle\Phi_b|\hat{H}|\Phi_b\rangle/\langle\Phi_b|\Phi_b\rangle, \tag{5.10}$$

which replaces that of Eq. (5.5) in the overall method outlined above, with the coupling information necessary so that the directions it contributes to $\boldsymbol{B}^{(b)}$ are optimal with respect to both intra-block and inter-block variable couplings. While there are many possible choices for the linear combinations $|\Theta_{jk}\rangle$, we thought it natural to derive them from previous iterations' BLM updates, following the idea that using previous update directions to inform the current direction is a common theme in numerical minimization, occurring for example in both the BFGS [149] and accelerated descent [150] methods. Specifically, for the nth iteration of the BLM, we take $|\Theta_{jk}\rangle$ as the kth block's component of the $(n-j)$th iteration's overall update, with $j \in \{1, 2, \ldots, N_O\}$. As our results will demonstrate, even relatively short history lengths N_O can be beneficial in accounting for inter-block variable couplings and thereby recovering the performance of the standard LM.

To understand the reduced memory footprint of the BLM, it is helpful to consult a visual guide to the structure of the Hamiltonian and overlap matrices resulting from Eq. (5.10). Figure 5.2 shows this structure for the Hamiltonian; the overlap matrix has an analogous structure. Noting that the different blocks' eigenproblems can be solved independently, we can see that only one block's matrices need to be fully constructed at a time, which greatly reduces memory requirements by allowing us to store one copy, rather than N_B copies, of the blue elements in Figure 5.2. For the green elements, however, we must store N_B copies simultaneously, so that each sampled configuration $\vec{n}$ or $\vec{r}$ can efficiently add its unique contribution to each of them. Nonetheless, storage requirements are much lower than in the standard LM, whose Hamiltonian matrix contains $(1 + N_V)^2$ elements. Although the precise formula for the BLM's Hamiltonian storage requirement is more longwinded, the terms that dominate, N_V^2/N_B and $2N_V N_O(N_B - 1)$, are much smaller than the dominant N_V^2 term in the LM. Thus, if no previous updates are being used (i.e. $N_O = 0$), the BLM reduces memory requirement by a factor of N_B, and although the use of $N_O > 0$ increases the BLM's memory requirement somewhat, the savings remain substantial. For example, when using 8-byte floats and 30,000 variational parameters, the standard LM requires 14.4 gigabytes of memory per process, while the BLM with $N_B = 100$ and $N_O = 5$ requires only 0.5 gigabytes per process.

5.3 Results

N_2 and H_2O with JAGP

We begin our numerical tests with the ground states of two small molecules, N_2 and H_2O, choosing the JAGP for our ansatz and performing VMC sampling in the second-quantized

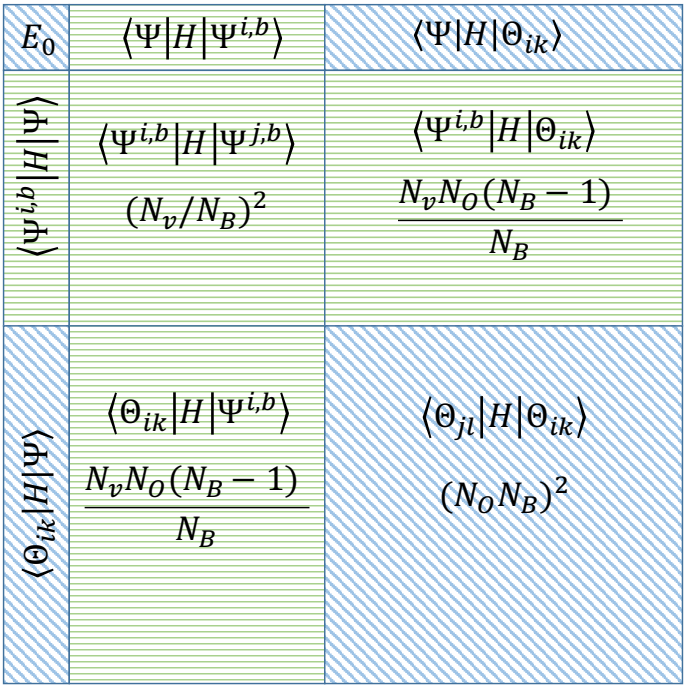

Figure 5.1: Structure of the BLM Hamiltonian matrix for the bth block, with each section of the matrix displaying the type of matrix element it contains. Green-shaded sections contain elements that are unique to each block; for the larger among these, we print the number of elements that must be stored per block. Blue-shaded sections contain elements shared by all blocks; for the larger among these, we print the total storage requirement across all blocks. Total memory consumption can then be evaluated as blue + $N_B \times$green.

Hilbert-space of the 6-31G [85] orbital basis. These choices give us 408 and 273 nonlinear parameters to optimize in N_2 and H_2O, respectively, which are few enough so as to make direct comparisons to the standard LM straightforward. Tables 5.1 and 5.2 show the results for various combinations of the number of blocks N_B, previous update vectors N_O, and retained block directions N_K. The reported optimization error is defined as the difference in energy between the minimums found by the BLM and the standard LM, the latter of which is denoted by $N_B = 1$ in the tables.

The first observation to be made is that although small, errors with respect to the standard LM are not zero. The likely explanation for this fact is that the BLM update direction, like that of the standard LM, is a nonlinear function of the random variables drawn by our Markov chains. Unlike linear functions of random variables that have statistical uncertainty but no systematic bias, nonlinear functions produce a systematic bias, albeit one that can in principle be mitigated by increasing the sample size. We suspect that our 2-step process of first diagonalizing N_B block-wise eigenproblems before constructing and diagonalizing one overall eigenproblem, which we note uses the same VMC sample for both steps, is essentially more nonlinear than the standard method's 1-step process. In other words, both the BLM

and LM should be expected to converge to a point in variable space that is slightly off-center from the true minimum due to systematic bias, but we expect the BLM to be more off-center due to its additional nonlinearities. Indeed, we have verified that the two methods converge to the same minimum in the limit of infinite sampling, and as can be seen in the results, differences for finite sample lengths are modest and decrease as we retain more directions N_K from each block.

The second observation is that the error behaves as expected for different values of N_B, N_K, and N_O. Increasing the number of blocks N_B, which makes it harder to account for second-order couplings between variables when choosing update directions, increases the deviation from the standard LM. Also as expected, increasing N_K and N_O tends to decrease the deviation. As hypothesized in the motivation for the BLM, only modest values of N_K and N_O are required to produce close approximations to the optimal update direction, and so mitigating deviations from the standard LM is not difficult. Finally, we note that although the BLM typically requires more iterations to converge, the convergence speed remains similar to the standard LM, especially when taking advantage of both multiple directions N_K per block and some number N_O of previous update directions.

C_2 with MSJ

We next switch from sampling in Fock space to sampling in real space, with Table 5.3 giving results for the ground state of C_2 as modeled by a MSJ ansatz containing 1,100 CSFs and 30 spline-based Jastrow variables. To construct our CSF expansion, we began with a GAMESS optimization of an (8,8) complete active space self-consistent field (CASSCF) ansatz in the cc-pVTZ basis [84]. The 1,100 largest-coefficient CSFs were then selected from a single-reference configuration interaction calculation including up to quadruples (CISDTQ) performed in the optimized CASSCF orbital basis. As before, we see that increasing the number of blocks eventually results in a significant deviation from the standard LM energy, which is then reduced by increasing the number of old updates used and the number of directions retained from each block. Again, while larger, the number of iterations required

Table 5.1: Comparison of the LM ($N_B = 1$) and BLM for the ground state of N_2 using the JAGP ansatz with Hilbert-space sampling in the 6-31G basis.

N_B	N_O	N_K	Energy (a.u.)	Error (10^{-3}a.u.)	Number of Iterations
1	N/A	N/A	-109.089	0.0	18
4	1	1	-109.088	1.5	19
4	5	1	-109.088	1.5	21
8	1	1	-109.087	3.0	29
8	5	1	-109.087	3.0	19
16	1	1	-109.086	3.3	38
16	5	1	-109.086	3.3	24

Table 5.2: Comparison of the LM ($N_B = 1$) and BLM for the ground state of H_2O using the JAGP ansatz with Hilbert-space sampling in the 6-31G basis.

N_B	N_O	N_K	Energy (a.u.)	Error (10^{-3}a.u.)	Number of Iterations
1	N/A	N/A	-76.109	0.0	8
2	5	1	-76.108	1.5	10
4	1	1	-76.106	3.0	11
4	3	1	-76.106	3.3	11
4	5	1	-76.106	3.0	11
8	1	1	-76.103	5.9	9
8	1	2	-76.104	4.8	10
8	1	4	-76.107	2.2	13
8	3	1	-76.104	4.8	12
8	3	2	-76.106	2.9	12
8	3	4	-76.106	2.9	11
8	5	1	-76.107	2.6	12
8	5	2	-76.106	2.9	13
8	5	4	-76.108	1.5	12

to converge the BLM was similar to that for the standard LM.

The H_{16} Hydrogen Ring

Having tested our method in settings where it can be easily checked against the standard LM, we now turn our attention to the metal-insulator transition in a 16-atom hydrogen ring, where we will use the BLM in conjunction with our excited state targeting method [41] to systematically converge the post-transition optical gap via a series of increasingly large MSJ

Table 5.3: Comparison of the LM ($N_B = 1$) and BLM for the ground state of C_2 using a MSJ expansion with real-space sampling.

N_B	N_O	N_K	Energy (a.u.)	Error (10^{-3}a.u.)	Number of Iterations
1	N/A	N/A	-75.834	0.0	8
4	1	1	-75.834	0.4	8
8	1	1	-75.833	1.1	10
16	1	1	-75.833	0.4	11
50	1	1	-75.832	1.5	10
100	1	1	-75.827	6.6	12
100	5	1	-75.831	2.9	11
100	5	5	-75.832	1.5	10

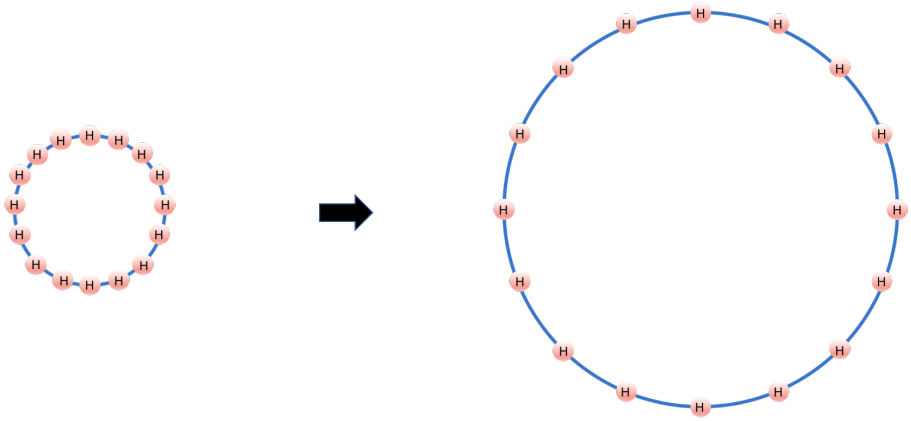

Figure 5.2: A schematic representation of the squeezed (left) and stretched (right) 16-atom hydrogen ring.

expansions. Closely related hydrogen chains have been the subject of much attention [151, 152, 153] due to the Mott-like behavior of the metal-insulator transition that occurs as one enlarges the interatomic distance a. As a surpasses a certain critical distance a_c, a large number of natural orbitals become degenerate as the electrons transition out of the weakly correlated metallic state and into the strongly correlated and more localized Mott-insulator state.

Using JAGP approximations for the ground state of the 1D chain, Sorella and coworkers [153] located a_c by evaluating the complex polarization function [154]

$$z = \langle \Psi | \exp\left(\frac{2\pi i}{L} \sum_k r_k^{\parallel}\right) | \Psi \rangle, \tag{5.11}$$

where $r_k^{\parallel}$ is the component of $\vec{r}_k$ parallel to the chain axis. The modulus of z can be thought of as a measurement of insulating behavior: $|z| \to 1$ as electrons localize about the nuclei, as occurs in the insulating phase, while $|z| \to 0$ as the electrons become fully delocalized, as occurs in the metallic phase. As we are studying a hydrogen ring instead of a periodic chain, we find it appropriate to instead define the complex polarization function as

$$z = \langle \Psi | \exp\left(i \sum_k \theta_k\right) | \Psi \rangle, \tag{5.12}$$

where θ_k is the angle around the ring for the kth electron's position. As for the chain, fully localized versus delocalized behavior in the ring will lead to the $|z| \to 1$ and $|z| \to 0$ limits, respectively.

In addition to probing the locality of its physics, theoretical methods can also offer predictions about an insulator's optical gap. Although this gap was not accessible in the ground-state work of Sorella, the BLM can directly target an excited state by minimizing the

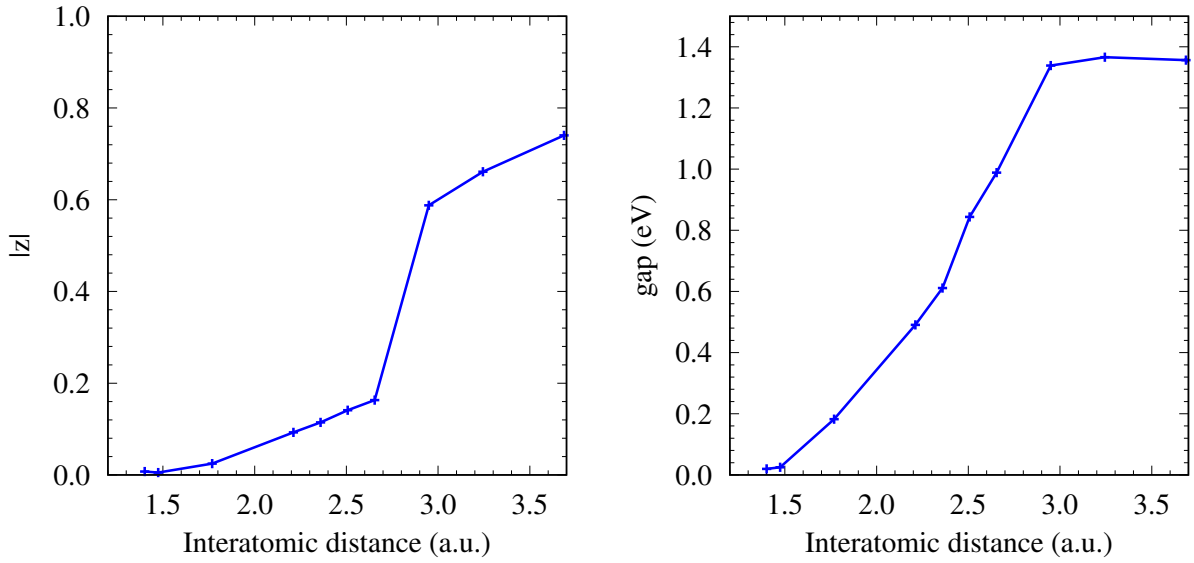

Figure 5.3: The complex polarization $|z|$ and optical gap of the H_{16} ring as a function of the interatomic distance, evaluated using a MSJ ansatz containing all CISDTQ configurations with coefficients above 0.01.

function $\Omega = \langle\Psi|(\omega - \hat{H})|\Psi\rangle / \langle\Psi|(\omega - \hat{H})^2|\Psi\rangle$, which, when the energy shift ω is placed inside the gap, will have the first excited state as its global minimum [41]. As this excited state approximates the state at the bottom of the infinite ring's conduction band, this approach would represent a direct, many-body, non-perturbative, and systematically improvable route to estimating the optical gap of a solid. In this study, we will explore a simple prototype of this idea by converging the gap for the H_{16} ring by systematically increasing the number of CSFs included in a MSJ expansion. Although linear combinations of CSFs are not natural fits for the strongly correlated physics of a Mott transition and will thus require a large number of CSFs be employed, they do offer straightforward systematic improvability and anyways allows us to demonstrate that the BLM can handle the correspondingly large number of variational parameters.

To construct our MSJ expansion, we begin by using GAMESS to optimize a (6,6) state-average CASSCF ansatz in the cc-pVDZ basis [84]. We then perform a single-reference CISDTQ for each state, after which we truncate this expansion at different coefficient thresholds to produce a series of increasingly large CSF expansions. By combining these with QMCPACK's standard spline-based, cusp-inducing e-e and e-n two-body Jastrow factors, we produce two sets of MSJ expansions, on each for the ground and excited state. Finally, choosing the value of ω that is appropriate for each state by adjusting it to find the overall minimum of the target function Ω [41], we optimize both the CSF coefficients and Jastrow variables simultaneously using the BLM.

Figure 5.3 shows the norm of the complex polarization function as well as the optical gap

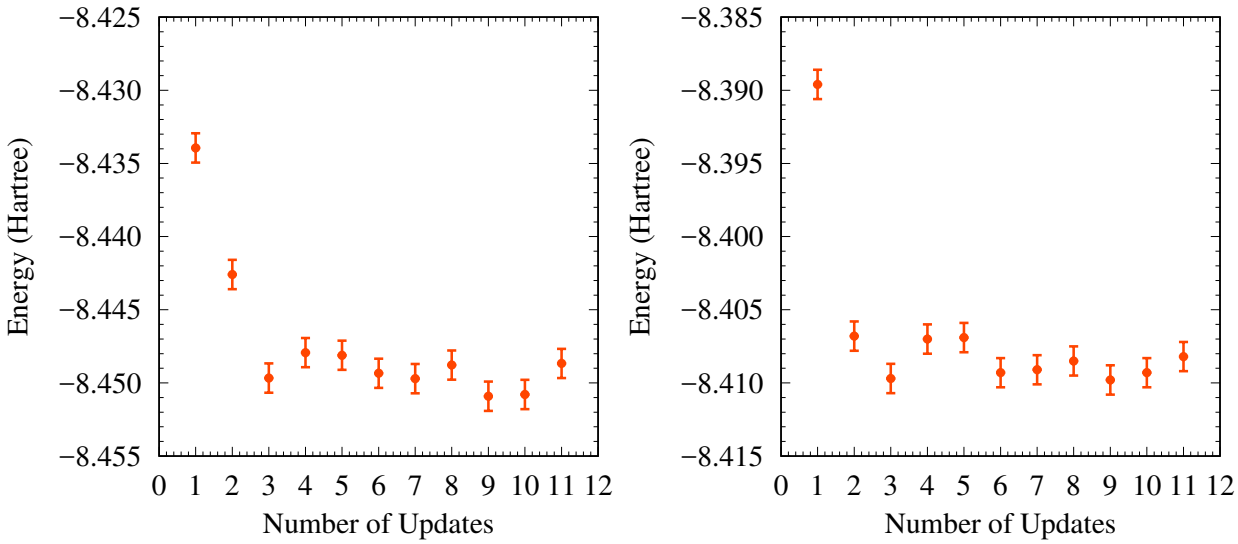

Figure 5.4: BLM convergence for the hydrogen ring's MSJ energy in the ground state (left, 21,401 parameters) and first excited state (right, 25,297 parameters).

estimate (defined as the difference between excited and ground state energies) as functions of interatomic distance a for a coefficient truncation threshold of 0.01. As expected, both $|z|$ and the gap are zero for small a, where previous studies have found hydrogen chains to be metallic. As a increases, we see an abrupt change in $|z|$ that suggests that by $a = 3.0a.u.$, the ring has transitioned into an insulating state. Being a finite system, the energy gap does not open discontinuously, and we see instead a rapid rise in the gap until it reaches a plateau beyond $a = 3.0a.u.$, thus agreeing with $|z|$ as to the location of the transition.

To ensure we have accurately converged the size of the gap in the insulating plateau region, we have performed our analysis of systematically increasing CSF expansion sizes at $a = 2.95a.u.$, where we transition from $N_B = 1$ (the standard LM) to $(N_B = 100, N_O = 5, N_K = 3)$ when the number of variables surpasses 5,000. Figure 5.4 shows the convergence behavior for the optimization of the largest MSJ expansions for both the ground and excited states, which involved 21,401 and 25,297 variational parameters, respectively. Note that, as is typical for the standard LM, the BLM converges in a handful of iterations. It is also important to point out that the total computational cost for evaluating all of the data points in Figure 5.4 amounted to 8,000 core-hours using the 2.3 GHz Intel Xeon 12-core Haswell processors of Berkeley's Savio computing cluster. Although this cost is not trivial, it is modest on the scale of modern parallel computation, giving ample room for this approach to be scaled up both to larger systems and larger variational parameter sets. Finally, in Figure 5.5, we show the convergence of the energy gap as the variational flexibility of the ansatz is increased, seeing clearly that, to within our statistical uncertainty, the gap has converged with respect to the addition of further CSFs into the wave function. Thus, by

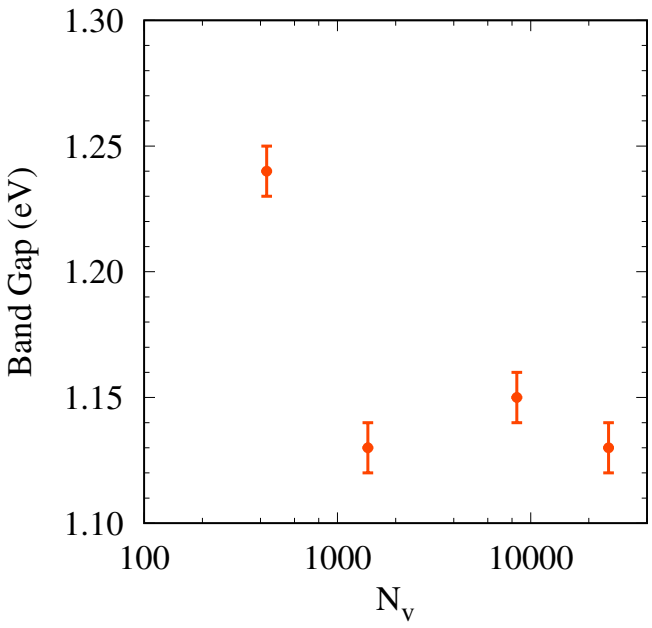

Figure 5.5: Convergence of the hydrogen ring's optical gap with respect to increasing variational flexibility, with N_V the number of variational parameters in the excited state ansatz.

combining the direct optimization of ansatzes for the ground and conduction edge states with the ability to optimize the large number of parameters inherent to a systematic expansion of ansatz flexibility, we provide an example of how the optical gap of a Mott insulator may be converged with respect to the effects of strong, many-body correlations.

To show the insensitivity of gap with respect to the parameters N_O and N_K of BLM, we compute the gap with difference choices of these two parameters. Results are shown in Table 5.4. 100 blocks are used in order to maintain consistency with previous calculations. Results show that the calculated gap is highly insensitive to difference choices of N_O and N_K.

To analyze the size of error introduced by BLM, we compute the gap withe the difference

Table 5.4: Optical gap of the hydrogen ring with different choices of N_O and N_K using 100 blocks.

N_O	N_K	Gap (eV)
5	3	1.14(2)
5	1	1.13(2)
3	3	1.15(2)
3	1	1.16(2)

choices of N_B. Since as N_B gets smaller, BLM should gradually converge to standard LM. Results are shown in Table 5.5. Thanks to the linearity of CSF coefficients and sparsity of the parameter space, the gap is highly insensitive to N_B so that the error introduced by BLM is negligible in this case.

Table 5.5: Optical gap of the hydrogen ring with different choices of N_B using $N_O = 3$ and $N_K = 1$.

N_B	Gap (eV)
100	1.16(2)
50	1.16(2)
25	1.14(2)

5.4 Conclusions

We have presented the blocked linear method, a wave function optimization method for variational Monte Carlo that addresses a crucial memory bottleneck in the highly successful standard linear method. By dividing ansatz variables into blocks, finding important update directions in each block, and then combining these directions to find an overall update for the current wave function, our method minimizes either the energy or a function suitable for targeting excited states while avoiding both the construction of overly large matrices and any requirement that such matrices be well conditioned. In small molecule tests that employed multiple ansatz types and involved both real space and Hilbert space sampling, we showed that the method reproduces the results of the standard linear method to a very good approximation.

In a demonstration of the method's ability to optimize large variable sets, we showed that the optical gap of a Mott-insulating hydrogen ring could be systematically converged with respect to increasing flexibility in the ansatzes for the ground and conduction band edge states. While many of the challenges inherent to simulating real Mott insulators are not present in this example, the method's success here raises interesting questions about the role it might play in the solid state. For example, it may be possible to pursue the systematic convergence of conduction band states' nodal structures for use in DMC. Furthermore, given the importance of addressing finite size effects by examining larger and larger simulation cells, the ability to optimize a large number of variables could prove especially useful in the solid state. We thus look forward to future work exploring applications of the blocked linear method in both solids and larger molecules.

In recent years, we have seen a rapid growth in different types of first order accelerated gradient descent optimization methods and their applications in machine learning. These first order optimization methods are usually much more memory friendly comparing to second order ones. However, the price one needs to pay for the memory efficiency is their slow

convergence behavior. The recent developed hybrid optimization algorithm[47] combines accelerated gradient descent and BLM. The resulting algorithm appears to work better than using either class of method on its own. The superiority of the hybrid method opens new research directions in terms of further investigations aimed at improving the methodology itself.

Chapter 6

Conclusions

We have introduced a new excited state variational principle, and discussed its applications in Fock and real space, in open and extended systems, and with different approximate wave function ansatzes. We compared it with EOM/LR based excited state methods and discussed the merits and drawbacks of both approaches. We also made improvements on the underlying optimization algorithm to address a fatal memory bottleneck that limits the capabilities of the method. In its current form, this new excited state variational principle offers accurate and systematically improvable predictions to excitation energies and optical gaps, both in weakly and strongly correlated systems.

As a non-perturbative excited state method, the new excited state variational principle is capable of tackling problems that are difficult to solve by perturbative methods. In doubly excited states, the new excited state variational principle yields highly accurate excitation energies that show order-of-magnitude of improvements over the predictions of conventional EOM/LR based methods. In solid ZnO, while MBPT in the form of G_0W_0 becomes inaccurate and sensitive to the underlying one-particle basis, the new excited state variational principle remains accurate and is insensitive to the DFT starting point.

In molecules, the ability to yield accurate excitation energies and excited state wave functions makes this new excited state variational principle an useful tool to study charge-transfer (CT) and doubly excited states. CT and doubly excited states play vital roles in biological systems[155], photocatalysts[11], and photodynamics[27], but they are generally hard to model with LR/EOM based methods. In order to better understand experimental results and develop new experimental directions, the predictive power of this new excited state method on these two kinds of excited states would be extremely helpful.

In materials, the new excited state variational principle offers a new way to obtain accurate and systematically improvable optical gaps, which could in principle guide the design of solar cells and help to predict and interpret spectroscopic observations. Furthermore, with accurate exciton wave function, one could address the debate about the nature of excitations for materials of technological importance. For example, the origin of the insulating behavior (charge-transfer v.s. Mott) of NiO is still under debate[156], and having an accurate exciton wave function would shine new insight into and even resolve this or other similar problems.

Obtaining correct dynamical behavior of excited states is as important as predicting static properties such as excitation energy. The time evolution of excited states has led to countless important effects, such as singlet fission[157], decoherence[155], and long persistent luminescence[158]. Quantities that are necessary for dynamics simulation, such as electronic coupling between excited states and ground state, non-adiabatic coupling, and nuclear forces could not be computed correctly without accurate excited state wave functions. In this sense, this new excited state variational principle also has potential applicability in dynamic modeling.

Looking forward, many questions remain unanswered for this new excited state variational principle. The most pressing one would be whether it is possible to develop efficient deterministic algorithms to evaluate and optimize it. In this way the high computational prefactor of QMC could be avoided and large-scale systems in condensed matter can be studied. Besides efficiency, the fact that the excited state variational principle can only target one excited state at a time makes it less competitive to use for computing spectra compared to LR/EOM and MBPT approaches. Therefore, it would be important to know whether the accuracy of the excited state variational principle in single excited states could be transferred to accurate predictions of spectra. In addition, the continuing development of robust optimization techniques and accurate wave function ansatzes would also benefit the method.

Bibliography

[1] T. Helgaker, P. Jøgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000.

[2] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Mineola, N.Y.: Dover Publications, 1996.

[3] P.J. Knowles and N.C. Handy. "A new determinant-based full configuration interaction method". In: *Chem. Phys. Lett.* 111 (1984), p. 315.

[4] P.J. Knowles and N.C. Handy. "A determinant based full configuration interaction program". In: *Comp. Phys. Commun.* 54 (1989), p. 75.

[5] W. Kutselnigg. "Present and Future Trends in Quantum Chemical Calculations". In: *J. Mol. Struct(THEOCHEM)* 181 (1988), p. 33.

[6] Ulrich Schollwöck. "The density-matrix renormalization group". In: *Rev. Mod. Phys.* 77 (2005), p. 259.

[7] W. M. C. Foulkes et al. "Quantum Monte Carlo simulations of solids". In: *Rev. Mod. Phys.* 73 (2001), p. 33.

[8] Robert G. Parr and Weitao Yang. *Density-functional theory of atoms and molecules*. New York: Oxford University Press, 1989.

[9] A. Ma et al. "Scheme for adding electron–nucleus cusps to Gaussian orbitals". In: 122 (2005), p. 224322.

[10] U Bach et al. "Solid-state dye-sensitized mesoporous TiO2 solar cells with high photon-to-electron conversion efficiencies". In: *Nature* 395.6702 (1998), pp. 583–585.

[11] Jacek K. Stolarczyk et al. "Challenges and Prospects in Solar Water Splitting and CO2 Reduction with Inorganic and Hybrid Nanostructures". In: *ACS Catal* 8 (2018), pp. 3602–3635.

[12] John C. Tully. "Perspective: Nonadiabatic dynamics theory". In: *J. Chem. Phys.* 137 (2012), 22A301.

[13] John C. Tully. "Nonadiabatic Processes in Molecular Collisions". In: *Dynamics of Molecular Collisions: Part B*. Ed. by William H. Miller. Boston, MA: Springer US, 1976, pp. 217–267. ISBN: 978-1-4757-0644-4. DOI: 10.1007/978-1-4757-0644-4_5.

[14] Wolfgang Domcke and David R. Yarkony. "Role of Conical Intersections in Molecular Spectroscopy and Photoinduced Chemical Dynamics". In: *Annu. Rev. Phys. Chem.* (2012), 63:325–52.

[15] J. M. Zhang and R. X. Dong. "Exact diagonalization: the Bose–Hubbard model as an example". In: *Eur. J. Phys.* 31 (2010), pp. 591–602.

[16] Yann Garniron et al. "Quantum Package 2.0: An Open-Source Determinant-Driven Suite of Programs". In: *J. Chem. Theory Comput.* 15 (2019), pp. 3591–3609.

[17] Robert J. Buenker, Sigrid D. Peyerimhoff, and Werner Butscher. "Applicability of the multi-reference double-excitation CI(MRD-CI) method to the calculation of electronic wavefunctions and comparison with related techniques". In: *Mol. Phys.* 35 (1978), pp. 771–791.

[18] N. S. Blunt, Ali Alavi, and George H. Booth. "Krylov-Projected Quantum Monte Carlo Method". In: *Phys. Rev. Lett.* 115 (2015), p. 050603.

[19] N. S. Blunt et al. "An excited-state approach within full configuration interaction quantum Monte Carlo". In: *J. Chem. Phys.* 143 (2015), p. 134117.

[20] Debashree Ghosh et al. "Orbital optimization in the density matrix renormalization group, with applications to polyenes and beta-carotene". In: *J. Chem. Phys.* 128 (2008), p. 144117.

[21] Weifeng Hu and Garnet Kin-Lin Chan. "Excited-State Geometry Optimization with the Density Matrix Renormalization Group, as Applied to Polyenes". In: *J. Chem. Theory. Comput.* 11 (2015), p. 3000.

[22] Andreas Dreuw and M. Head-Gordon. "Single-Reference ab Initio Methods for the Calculations of Excited States of Large Molecules". In: *Chem. Rev.* 105 (2005), p. 4009.

[23] A. Krylov. "Equation-of-Motion Coupled-Cluster Methods for Open-Shell and Electronically Excited Species: The Hitchhiker's Guide to Fock Space". In: *Annu. Rev. Phys. Chem.* 59 (2008), p. 433.

[24] Naoki Nakatani et al. "Linear response theory for the density matrix renormalization group: Efficient algorithms for strongly correlated excited states". In: *J. Chem. Phys.* 140 (2014), p. 024108.

[25] Luning Zhao and Eric Neuscamman. "Equation of Motion Theory for Excited States in Variational Monte Carlo and the Jastrow Antisymmetric Geminal Power in Hilbert Space". In: *J. Chem. Theory Comput.* 12 (2016), pp. 3719–3726.

[26] Mark A. Watson and Garnet Kin-Lic Chan. "Excited States of Butadiene to Chemical Accuracy: Reconciling Theory and Experiment". In: *J. Chem. Theory Comput.* 8 (2012), pp. 4013–4018.

[27] Yinan Shu and Donald G. Truhlar. "Doubly Excited Character or Static Correlation of the Reference State in the Controversial ^{2}1A$_g$ State of trans-Butadiene?" In: *J. Am. Chem. Soc.* 139 (2017), pp. 13770–13778.

[28] Lars Hedin. "New method for calculating the one-particle Green's function with application to the electron-gas problem". In: *Phys. Rev.* 139.3A (1965), A796.

[29] G. Strinati. "Effects of dynamical screening on resonances at inner-shell thresholds in semiconductors". In: *Phys. Rev. B* 29.10 (1984).

[30] Giovanni Onida, Lucia Reining, and Angel Rubio. "Electronic excitations: density-functional versus many-body Green's-function approaches". In: *Rev. Mod. Phys.* 74.2 (2002), p. 601.

[31] Suvadip Das, John E. Coulter, and Efstratios Manousakis. "Convergence of quasiparticle self-consistent GW calculations of transition-metal monoxides". In: *Phys. Rev. B* 91 (2015), p. 115105.

[32] John P Perdew. "Density functional theory and the band gap problem". In: *Int. J. Quantum Chem.* 28.S19 (2009), pp. 497–523. DOI: 10.1002/qua.560280846.

[33] J. P. Perdew and Alex Zunger. "Self-interaction correction to density-functional approximations for many-electron systems". In: *Phys. Rev. B* 23.10 (1981), pp. 5048–5079. DOI: 10.1103/PhysRevB.23.5048.

[34] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. "Generalized gradient approximation made simple". In: *Phys. Rev. Lett.* 77.18 (1996), pp. 3865–3868. DOI: 10.1103/PhysRevLett.77.3865.

[35] M Shishkin and G Kresse. "Self-consistent G W calculations for semiconductors and insulators". In: *Phys. Rev. B* 75.23 (2007), p. 235102.

[36] Bi-Ching Shih et al. "Quasiparticle band gap of ZnO: High accuracy from the conventional G_0W_0 approach". In: *Phys. Rev. Lett.* 105.14 (2010), p. 146401.

[37] Carlo Adamo and Vincenzo Barone. "Toward reliable density functional methods without adjustable parameters: The PBE0 model". In: *J. Chem. Phys.* 110.6158 (1999).

[38] Jochen Heyd and Gustavo E. Scuseria. "Hybrid functionals based on a screened Coulomb potential". In: *J. Chem. Phys.* 118.8207 (2003).

[39] Jong H. Choi, Charles F. Lebeda, and Richard P. Messmer. "Variational principle for excited states: exact formulation and other extensions". In: *Chem. Phys. Lett.* 5 (1970), p. 503.

[40] Jonathan J. Dorando, Johannes Hachmann, and Garnet Kin-Lic Chan. "Targeted excited state algorithms". In: *J. Chem. Phys.* 127 (2007), p. 084109.

[41] Luning Zhao and Eric Neuscamman. "An efficient variational principle for the direct optimization of excited states". In: *arXiv* (2016), p. 1508.06683.

[42] M. P. Nightingale and V. Melik-Alaverdian. In: *Phys. Rev. Lett.* 87 (2001), p. 043401.

[43] C. J. Umrigar et al. "Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions". In: *Phys. Rev. Lett.* 98 (2007), p. 110201.

[44] J. Toulouse and C. J. Umrigar. In: *J. Chem. Phys.* 126 (2007), p. 084102.

[45] J. Toulouse and C. J. Umrigar. In: *J. Chem. Phys.* 128 (2008), p. 174101.

[46] Lauretta R. Schwarz, A. Alavi, and George H. Booth. "Projector Quantum Monte Carlo Method for Nonlinear Wave Functions Lauretta". In: *Phys. Rev. Lett.* 118 (2017), p. 176403.

[47] Leon Otis and Eric Neuscamman. "Complementary first and second derivative methods for ansatz optimization in variational Monte Carlo". In: *Phys. Chem. Chem. Phys.* 21 (2019), pp. 14491–14510.

[48] Brett Van Der Goetz, Leon Otis, and Eric Neuscamman. "Clean and Convenient Tessellations for Number Counting Jastrow Factors". In: *J. Chem. Theory Comput.* 15 (2019), pp. 1102–1121.

[49] Luning Zhao and Eric Neuscamman. "A Blocked Linear Method for Optimizing Large Parameter Sets in Variational Monte Carlo". In: *J. Chem. Theory Comput.* 13(6) (2017), pp. 2604–2611.

[50] Hong-Zhou Ye et al. "σ-SCF: A direct energy-targeting method to mean-field excited states". In: *J. Chem. Phys.* 147 (2017), p. 214104.

[51] G. Kotliar et al. "Electronic structure calculations with dynamical mean-field theory". In: *Rev. Mod. Phys.* 78 (2006), p. 865.

[52] Tsz S. Chwee and Emily A. Carter. "Valence Excited States in Large Molecules via Local Multireference Singles and Doubles Configuration Interaction". In: *J. Chem. Theory Comput.* 7 (2011), p. 103.

[53] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins. "Optimized trial wave functions for quantum Monte Carlo calculations". In: *Phys. Rev. Lett.* 60 (1988), p. 1719.

[54] Andreas Dreuw and M. Head-Gordon. "Single-Reference ab Initio Methods for the Calculations of Excited States of Large Molecules". In: *Chem. Rev.* 105 (2005), p. 4009.

[55] Myint Tin. "Comparison of Some Ratio Estimators". In: *JASA* 60 (1965), p. 294. DOI: 10.1080/01621459.1965.10480792.

[56] Miguel A. Morales et al. "Multideterminant Wave Functions in Quantum Monte Carlo". In: *J. Chem. Theory Comput.* 8 (2012), p. 2181.

[57] Michele Casula and Sandro Sorella. "Geminal wave functions with Jastrow correlation: A first application to atoms". In: *J. Chem. Phys.* 119 (2003), p. 6500.

[58] Michele Casula, Claudio Attaccalite, and Sandro Sorella. "Correlated geminal wave function for molecules: An efficient resonating valence bond approach". In: *J. Chem. Phys.* 121 (2004), p. 7110.

[59] Sandro Sorella, Michele Casula, and Dario Rocca. "Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods". In: *J. Chem. Phys.* 127 (2007), p. 014105.

[60] Mariapia Marchi et al. "Resonating valence bond wave function with molecular orbitals: Application to first-row molecules". In: *J. Chem. Phys.* 131 (2009), p. 154116.

[61] Eric Neuscamman. "The Jastrow antisymmetric geminal power in Hilbert space: theory, benchmarking, and application to a novel transition state". In: *J. Chem. Phys.* 139 (2013), p. 194105.

[62] Eric Neuscamman. "Size Consistency Error in the Antisymmetric Geminal PowerWave Function can be Completely Removed". In: *Phys. Rev. Lett.* 109 (2012), p. 203001.

[63] Garnet Kin-Lic Chan and Sandeep Sharma. "The Density Matrix Renormalization Group in Quantum Chemistry". In: *Annu. Rev. Phys. Chem.* 62 (2011), p. 465.

[64] Ulrich Schollwöck. "The density-matrix renormalization group in the age of matrix product states". In: *Ann. Phys.* 326.1 (2011), pp. 96–192.

[65] J. R. Trail. "Heavy-tailed random error in quantum Monte Carlo". In: *Phys. Rev. E* 77 (2008), p. 016703. DOI: 10.1103/PhysRevE.77.016703.

[66] Sergio D. Pineda Flores and Eric Neuscamman. "Excited State Specific Multi-Slater Jastrow Wave Functions". In: *J. Phys. Chem. A.* 123 (2019), pp. 1487–1497.

[67] Paul. J. Robinson, Sergio D. Pineda Flores, and Eric Neuscamman. "Excitation variance matching with limited configuration interaction expansions in variational Monte Carlo". In: *J. Chem. Phys.* 147 (2017), p. 164114.

[68] Jeongnim Kim et al. "QMCPACK : An open source ab initio Quantum Monte Carlo package for the electronic structure of atoms, molecules, and solids". In: *J. Phys. Condens. Matter* 30 (2018), p. 195901. DOI: 10.1088/1361-648X/aab9c3.

[69] Alan R. Tackett and Massimiliano Di Ventra. "Targeting specific eigenvectors and eigenvalues of a given Hamiltonian using arbitrary selection criteria". In: *Phys. Rev. B* 66 (2002), p. 245104.

[70] Gerald Jordan et al. "Fast iterative interior eigensolver for millions of atoms". In: *J. Comput. Phys.* 231 (2012), p. 4836.

[71] Ronald B. Morgan. "Computing interior eigenvalues of large matrices". In: *Linear Algebra Appl.* 154-156 (1991), p. 289.

[72] Gerard L. G. Sleijpen and Henk A. Van Der Vorst. "A Jacobi-Davidson Iteration Method For Linear Eigenvalue Problems". In: *SIAM J. Matrix Anal. Appl.* 17 (1996), p. 401.

[73] Bryan K. Clark et al. "Computing the energy of a water molecule using multideterminants: A simple, efficient algorithm". In: *J. Chem. Phys.* 135 (2011), p. 244105.

[74] Miguel A. Morales et al. "Multideterminant Wave Functions in Quantum Monte Carlo". In: *J. Chem. Theory Comput.* 8 (2012), p. 2181.

[75] Jacqueline A R Shea and Eric Neuscamman. "Size consistent excited states via algorithmic transformations between variational principles". In: *J. Chem. Theory Comput.* 13.12 (2017), pp. 6078–6088.

[76] C. J. Umrigar and Claudia Filippi. "Energy and Variance Optimization of Many-Body Wave Functions". In: *Phys. Rev. Lett* 94 (2005), p. 150201.

[77] Eric Neuscamman. "Communication: A Jastrow factor coupled cluster theory for weak and strong electron correlation". In: *J. Chem. Phys* 139 (2013), p. 181101.

[78] Eric Neuscamman. "Improved Optimization for the Cluster Jastrow Antisymmetric Geminal Power and Tests on Triple-Bond Dissociations". In: *arXiv* (2016), p. 1603.06605.

[79] Trygve Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley & Sons, Ltd., 2000.

[80] Sandeep Sharma. "A general non-Abelian density matrix renormalization group algorithm with application to the C2 dimer". In: *J. Chem. Theory Comput.* 142 (2015), p. 024107.

[81] N. S. Blunt et al. "An excited-state approach within full configuration interaction quantum Monte Carlo". In: *J. Chem. Theory Comput.* 143 (2015), p. 134117.

[82] Wirawan Purwanto, Shiwei Zhang, and Henry Krakauer. "Excited state calculations using phaseless auxiliary-field quantum Monte Carlo: Potential energy curves of low-lying C2 singlet states". In: *J. Chem. Phys.* 130 (2009), p. 094107. DOI: 10.1063/1.3077920.

[83] Claudia Filippi, Maurizio Zaccheddu, and Francesco Buda. "Absorption Spectrum of the Green Fluorescent Protein Chromophore: A Difficult Case for ab Initio Methods?" In: *J. Chem. Theory. Comput.* 5 (2009), p. 2074.

[84] T. Dunning. "Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen". In: *J. Chem. Phys* 90 (1989), p. 1007.

[85] W. J. Hehre, R. Ditchfield, and J. A. Pople. "Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules". In: *J. Chem. Phys* 56 (1972), p. 2257.

[86] T. Dunning and D. Woon. "Gaussian Basis Sets for Use in Correlated Molecular Calculations. IV. Calculation of static electrical response properties". In: *J. Chem. Phys* 100 (1994), p. 2975.

[87] Eric Neuscamman. "Communication: Variation after response in quantum Monte Carlo". In: *J. Chem. Phys.* 145 (2016), p. 081103.

[88] N. S. Blunt and Eric Neuscamman. "Charge-transfer excited states: Seeking a balanced and efficient wavefunction ansatz in variational Monte Carlo". In: *J. Chem. Phys.* 147 (2017), p. 194101.

[89] N. S. Blunt and Eric Neuscamman. "Excited-state diffusion Monte Carlo calculations: a simple and efficient two-determinant ansatz". In: *J. Chem. Theory Comput.* (2018). DOI: 10.1021/acs.jctc.8b00879.

[90] M. Shishkin and G. Kresse. "Self-consistent GW calculations for semiconductors and insulators". In: *Phys. Rev. B* 75 (2007), p. 235102.

[91] Xin-Zheng Li et al. "Impact of widely used approximations to the G_0W_0 method: an all-electron perspective". In: *New J. Phys.* 14 (2012), p. 023006.

[92] M. S. Hybertsen and S. G. Louie. "First-Principles Theory of Quasiparticles: Calculations of Band Gaps in Semiconductors and Insulators". In: *Phys. Rev. Lett.* 55 (1985), p. 1418.

[93] Martin Head-Gordon et al. "A doubles correction to electronic excited states from configuration interaction in the space of single substitutions". In: *Chem. Phys. Lett.* 219 (1994), p. 21.

[94] J E Subotnik. "Communication: configuration interaction singles has a large systematic bias against charge-transfer states." In: *J. Chem. Phys.* 135.7 (2011), p. 071104.

[95] Elif Ertekin, Lucas K Wagner, and Jeffrey C Grossman. "Point-defect optical transitions and thermal ionization energies from quantum Monte Carlo methods: Application to the F-center defect in MgO". In: *Phys. Rev. B* 87.15 (2013), p. 155210.

[96] Juan A Santana et al. "Structural stability and defect energetics of ZnO from diffusion quantum Monte Carlo". In: *J. Chem. Phys.* 142.16 (2015), p. 164705.

[97] S. Lebègue et al. "Pressure-induced simultaneous metal-insulator and structural-phase transitions in LiH: A quasiparticle study". In: *EPL* 63.4 (2003), p. 562.

[98] M. J. van Setten et al. "Electronic strucrue and optical properties of lightweight metal hydrides". In: *Phys. Rev. B* 75 (2007), p. 035204.

[99] K. P. O'Donnell and X. Chen. "Temperature dependence of semiconductor band gaps". In: *Appl. Phys. Lett.* 58 (1991), p. 2924.

[100] V. G. Plekhanov et al. In: *Sov. Phys. Solid State* 18 (1976), p. 2438.

[101] D. M. Roessler and W. C. Walker. "Electronic Spectrum of Crystalline Lithium Fluoride". In: *J. Phys. Chem. Solids* 28 (1967), p. 1507.

[102] S. Adachi et al. "Group IV Elements, IV-IV and III-V Compounds Part b: Electronic, Transport, Optical and Other Properties". In: *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology*. Ed. by U. Rössler. Vol. 41-A1β. ISBN 978-3540428763. Berlin: Springer, 2002.

[103] Marc Dvorak, Su-Huai Wei, and Zhigang Wu. "Origin of the Variation of Exciton Binding Energy in Semiconductors". In: *Phys. Rev. Lett.* 110 (2013), p. 016402.

[104] R. A. Kink et al. "Reflection spectra of lithium hydride crystals in 4–25 eV range at 5 K". In: *Nucl. Instr. Meth. Phys. Res. A* 261 (1987), pp. 138–139.

[105] F. Fuchs et al. "Quasiparticle band structure based on a generalized Kohn-Sham scheme". In: *Phys. Rev. B* 76.11 (2007), p. 115109.

[106] S. Tsoi et al. "Isotopic-mass dependence of the A, B, and C excitonic band gaps in ZnO at low temperatures". In: *Phys. Rev. B* 74 (2006), p. 165203.

[107] Paola Gori et al. "Optical spectra of ZnO in the far ultraviolet: First-principles calculations and ellipsometric measurements". In: *Phys. Rev. B* 81 (2010), p. 125207.

[108] Christoph Friedrich, Mathias C. Müller, and Stefan Blügel. "Band convergence and linearization error correction of all-electron *GW* calculations: The extreme case of zinc oxide". In: *Phys. Rev. B* 83 (2011), p. 081101.

[109] J. A. Berger, Lucia Reining, and Francesco Sottile. "Efficient *GW* calculations for SnO_2, ZnO, and rubrene: The effective-energy technique". In: *Phys. Rev. B* 85 (2012), p. 085126.

[110] Christoph Sommer, Peter Krüger, and Johannes Pollmann. "Optical spectra of alkali-metal fluorides Christoph". In: *Phys. Rev. B* 86 (2012), p. 155212.

[111] Andrey L Kutepov. "Electronic structure of Na, K, Si, and LiF from self-consistent solution of Hedin's equations including vertex corrections". In: *Phys. Rev. B* 94.15 (2016), p. 155101.

[112] Andrey L Kutepov. "Self-consistent solution of Hedin's equations: Semiconductors and insulators". In: *Phys. Rev. B* 95.19 (2017), p. 195120.

[113] Martin Head-Gordon et al. "Analysis of electronic transitions as the difference of electron attachment and detachment densities". In: *J. Phys. Chem.* 99.39 (1995), pp. 14261–14270. DOI: 10.1021/j100039a012.

[114] Burak Himmetoglu et al. "Hubbard-Corrected DFT Energy Functionals: The LDA+U Description of Correlated Systems". In: *Int. J. Quantum Chem.* 114 (2014), p. 14. DOI: 10.1002/qua.24521.

[115] Fengjie Ma, Shiwei Zhang, and Henry Krakauer. "Excited state calculations in solids by auxiliary-field quantum Monte Carlo". In: *New J. Phys.* 15.9 (2013), p. 093017.

[116] Jaehyung Yu, Lucas K Wagner, and Elif Ertekin. "Towards a systematic assessment of errors in diffusion Monte Carlo calculations of semiconductors: Case study of zinc selenide and zinc oxide". In: *J. Chem. Phys.* 143.22 (2015), p. 224707.

[117] A. Schleife et al. "Optical and energy-loss spectra of MgO, ZnO, and CdO from ab initio many-body calculations". In: *Phys. Rev. B* 80 (2009), p. 035112.

[118] Hong-Zhou Ye et al. "σ-SCF: A direct energy-targeting method to mean-field excited states". In: *J. Chem. Phys.* 147.21 (2017), p. 214104.

[119] Cloé Azarias et al. "Calculations of n$\rightarrow \pi^*$ Transition Energies: Comparisons Between TD-DFT, ADC, CC, CASPT2, and BSE/GW Descriptions". In: *J. Phys. Chem. A* 121.32 (2017), pp. 6122–6134.

[120] Jindřich Kolorenč and Lubos Mitas. "Quantum Monte Carlo calculations of structural properties of FeO under pressure". In: *Phys. Rev. Lett.* 101.18 (2008), p. 185502.

[121] Lubos Mitas and Jindřich Kolorenč. "Quantum Monte Carlo studies of transition metal oxides". In: *Rev. Mineral. Geochem.* 71.1 (2010), pp. 137–145.

[122] M Abbasnejad et al. "Quantum Monte Carlo study of high-pressure cubic TiO2". In: *Appl. Phys. Lett.* 100.26 (2012), p. 261902.

[123] Lucas K Wagner and Peter Abbamonte. "Effect of electron correlation on the electronic structure and spin-lattice coupling of high-T c cuprates: Quantum Monte Carlo calculations". In: *Phys. Rev. B* 90.12 (2014), p. 125129.

[124] Claudia Filippi, Roland Assaraf, and Saverio Moroni. "Simple formalism for efficient derivatives and multi-determinant expansions in quantum Monte Carlo". In: *J. Chem. Phys.* 144.19 (2016), p. 194105.

[125] Roland Assaraf, Saverio Moroni, and Claudia Filippi. "Optimizing the Energy with Quantum Monte Carlo: A Lower Numerical Scaling for Jastrow–Slater Expansions". In: *J. Chem. Theory Comput.* 13.11 (2017), pp. 5273–5281.

[126] Peter J Knowles. "Compressive sampling in configuration interaction wavefunctions". In: *Mol. Phys.* 113.13-14 (2015), pp. 1655–1660.

[127] Wenjian Liu and Mark R Hoffmann. "iCI: Iterative CI toward full CI." In: *J. Chem. Theory Comput.* 12.3 (2016), pp. 1169–1178.

[128] Jeffrey B Schriber and Francesco A Evangelista. "Communication: An adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy". In: *J. Chem. Phys.* 144 (2016), p. 161106.

[129] Norm M Tubman et al. "A deterministic alternative to the full configuration interaction quantum Monte Carlo method." In: *J. Chem. Phys.* 145.4 (2016), p. 044112.

[130] Adam A Holmes, Norm M Tubman, and C J Umrigar. "Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling." In: *J. Chem. Theory Comput.* 12.8 (2016), pp. 3674–3680.

[131] Sandeep Sharma et al. "Semistochastic Heat-bath Configuration Interaction method: selected configuration interaction with semistochastic perturbation theory". In: *J. Chem. Theory Comput.* 13.4 (2017), pp. 1595–1604.

[132] Paul M Zimmerman. "Incremental full configuration interaction". In: *J. Chem. Phys.* 146.10 (2017), p. 104102.

[133] Yuhki Ohtsuka and Jun-ya Hasegawa. "Selected configuration interaction method using sampled first-order corrections to wave functions". In: *J. Chem. Phys.* 147.3 (2017), p. 034102.

[134] Rodney J. Bartlett and Monika Musial. "Coupled-cluster theory in quantum chemistry". In: *Rev. Mod. Phys.* 79 (2007), p. 291.

[135] Yuki Kurashige, Garnet Kin-Lic Chan, and Takeshi Yanai. "Entangled quantum electronic wavefunctions of the Mn4CaO5 cluster in photosystem II". In: *Nature Chem.* 5 (2013), p. 660.

[136] Stefan Grimme and Maja Parac. "Substantial Errors from Time-Dependent Density Functional Theory for the Calculation of Excited States of Large pi Systems". In: *ChemPhysChem* 4 (2003), p. 292.

[137] Alaska Subedi et al. "Density functional study of FeS, FeSe, and FeTe: Electronic structure, magnetism, phonons, and superconductivity". In: *Phys. Rev. B* 78 (2008), p. 134514.

[138] B. Keimer et al. "From quantum matter to high-temperature superconductivity in copper oxides". In: *Nature* 518 (2015), pp. 179–186.

[139] Nicholas Metropolis and S. Ulam. "The Monte Carlo Method". In: *J. Am. Stat. Assoc.* 44 (1949), p. 335.

[140] C. J. Umrigar. "Observations on variational and projector Monte Carlo methods". In: *J. Chem. Phys.* 143 (2015), p. 164105.

[141] Eric Neuscamman. "Communication: Variation after response in quantum Monte Carlo." In: *J. Chem. Phys.* 145.8 (Aug. 2016), p. 081103.

[142] Luning Zhao and Eric Neuscamman. "Amplitude Determinant Coupled Cluster with Pairwise Doubles." In: *J. Chem. Theory Comput.* 12.12 (Dec. 2016), pp. 5841–5850.

[143] Eric Neuscamman. "Subtractive manufacturing with geminal powers:making good use of a bad wave function". In: *Mol. Phys.* 114.5 (2016), pp. 577–583.

[144] Lauretta R. Schwarz, A. Alavi, and George H. Booth. "A Projector Quantum Monte Carlo Method for Nonlinear Wave Functions". In: *Phys. Rev. Lett.* 118 (2017), p. 176403.

[145] Eric Neuscamman, C. J. Umrigar, and Garnet Kin-Lic Chan. "Optimizing large parameter sets in variational quantum Monte Carlo". In: *Phys. Rev. B* 85 (2012), p. 045103.

[146] J. Kim et al. "Hybrid algorithms in quantum Monte Carlo". In: *J. Phys.: Conf. Ser.* 402 (2012), p. 012008.

[147] K. Esler et al. "Accelerating Quantum Monte Carlo Simulations of Real Materials on GPU Clusters". In: *Comput. Sci. Eng.* 14 (2012), pp. 40–51.

[148] Ron Shepard et al. "The Subspace Projected Approximate Matrix (SPAM) Modification of the Davidson Method". In: *J. Comput. Phys.* 172 (2001), pp. 472–514.

[149] Jorge Nocedal. "Updating quasi-Newton matrices with limited storage". In: *Math. Comp.* 35 (1980), p. 773.

[150] Y. Nesterov. In: *Soviet Mathematics Doklady* 27 (1983), p. 372.

[151] Johannes Hachmann, Wim Cardoen, and Garnet Kin-Lic Chan. "Multireference correlation in long molecules with the quadratic scaling density matrix renormalization group". In: *J. Chem. Phys* 125 (2006), p. 144101.

[152] Takashi Tsuchimochi and Gustavo E. Scuseria. "Strong correlations via constrained-pairing mean-field theory". In: *J. Chem. Phys* 131 (2009), p. 121102.

[153] Lorenzo Stella et al. "Strong electronic correlation in the hydrogen chain: A variational Monte Carlo study". In: *Phys. Rev. B* 84 (2011), p. 245117.

[154] P. Umari et al. "Dielectric Response of Periodic Systems from Quantum Monte Carlo Calculations". In: *Phys. Rev. Lett.* 95 (2005), p. 207602.

[155] Shahnawaz Rafiq and Gregory D. Scholes. "From Fundamental Theories to Quantum Coherences in Electron Transfer". In: *J. Am. Chem. Soc.* 141(2) (2019), pp. 708–722.

[156] T. M. Schuler et al. "Character of the insulating state in NiO: A mixture of charge-transfer and Mott-Hubbard character". In: *Phys. Rev. B* 71 (2005), p. 115113.

[157] Diana Y Qiu, Felipe H da Jornada, and Steven G Louie. "Environmental Screening Effects in 2D Materials: Renormalization of the Bandgap, Electronic Structure, and Optical Spectra of Few-Layer Black Phosphorus". In: *Nano Lett.* 17.8 (2017), pp. 4706–4712.

[158] Ryota Kabe and chihaya Adachi. "Organic long persistent luminescence". In: *Nature* 550 (2017), p. 384.

[159] H.-J. Werner et al. *MOLPRO, version 2012.1, a package of ab initio programs.* see http://www.molpro.net. Cardiff, UK.

[160] Y. Shao et al. "Advances in methods and algorithms in a modern quantum chemistry program package". In: *Phys. Chem. Chem. Phys.* 8 (2006), p. 3172.

[161] A.I. Krylov and P.M.W. Gill. "Q-Chem: An engine for innovation". In: *WIREs Comput. Mol. Sci.* 3 (2013), p. 317.

[162] J. Kim et al. "Hybrid algorithms in quantum monte carlo". In: *JPCS* 402 (2012), p. 012008.

[163] K. P. Esler et al. "Fully accelerating quantum monte carlo simulations of real materials on gpu clusters". In: *Comput. Sci. Eng.* 14 (2012), p. 40.

[164] M. W. Schmidt et al. "General Atomic and Molecular Electronic Structure System". In: *J. Comput. Chem.* 14 (1993), p. 1347.

[165] T. Daniel Crawford et al. In: *J. Comput. Chem.* 28 (2007), p. 1610.

[166] W. J. Hehre, R. F. Stewart, and J. A. Pople. "Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals". In: *J. Chem. Phys.* 51 (1969), p. 2657.

[167] W. J. Hehre, R. Ditchfield, and J. A. Pople. "Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules". In: *J. Chem. Phys.* 56 (1972), p. 2257.

[168] Thom H. Dunning Jr. "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen". In: *J. Chem. Phys.* 90 (1989), p. 1007.

[169] Paolo Giannozzi et al. "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials". In: *J. Phys. Condens. Matter* 21.39 (2009), p. 395502. URL: `http://www.quantum-espresso.org`.

[170] P Giannozzi et al. "Advanced capabilities for materials modelling with QUANTUM ESPRESSO". In: *J. Phys. Condens. Matter* 29.46 (2017), p. 465901. URL: `http://stacks.iop.org/0953-8984/29/i=46/a=465901`.

[171] M. Burkatzki, Claudia Filippi, and M. Dolg. "Energy-consistent small-core pseudopotentials for 3d-transition metals adapted to quantum Monte Carlo calculations". In: *J. Chem. Phys.* 129 (2008), p. 164115.

[172] H. Shin et al. "Electronic properties of doped and defective NiO: A quantum Monte Carlo study". In: *Phys. Rev. Materials* 1 (2017), p. 073603.

[173] Jaron T. Krogel, Juan A. Santana, and Fernando A. Reboredo. "Pseudopotentials for quantum Monte Carlo studies of transition metal oxides". In: *Phys. Rev. B* 93 (2016), p. 075143.

[174] Qiming Sun et al. "PySCF: the Python-based simulations of chemistry framework". In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 8.1 (2018). ISSN: 17590884.

[175] M. W. Schmidt et al. "General Atomic and Molecular Electronic Structure System". In: *J. Comput. Chem.* 14 (1993), pp. 1347–1363.

Appendix A

Computational Details

A.1 Chapter 2

General Information

EOM-CCSD and FCI results were computed with MOLPRO [159], CIS results with QChem [160, 161], MSJ results with a modified version of QMCPACK [162, 163] with the CAS truncation taken from GAMESS [164], and JAGP results with our own prototype Hilbert space quantum Monte Carlo code with one- and two-electron integrals imported from Psi3 [165]. In JAGP we worked exclusively in the symmetrically orthogonalized "$S^{-1/2}$" one particle basis and froze the C 1s orbital at the HF level. All statistical uncertainties were converged to less than 0.01eV in all cases.

CH_2

For CH_2 we used a minimal STO-3G basis set [166] and shifts in Hartree of ω = -38.4, -38.3, -39.198, -38.15, -38.110, and -38.1 for excited states 1 to 6, respectively. As mentioned in the main text, this resulted in minima of Ω for the last two shifts that corresponded to

Table A.1: Absolute energies in Hartree of the CH_2 molecule in the STO-3G basis.

State	JAGP	CIS	EOM-CCSD	FCI
0	-38.434	-38.370	-38.435	-38.436
1	-38.340	-38.291	-38.342	-38.344
2	-38.201	N/A	-38.182	-39.201
3	-38.178	-38.104	-38.176	-38.185
4	-38.046	N/A	-38.025	-38.055
5	-37.891	-37.792	-37.885	-37.891
6	-37.878	-37.811	-37.871	-38.874

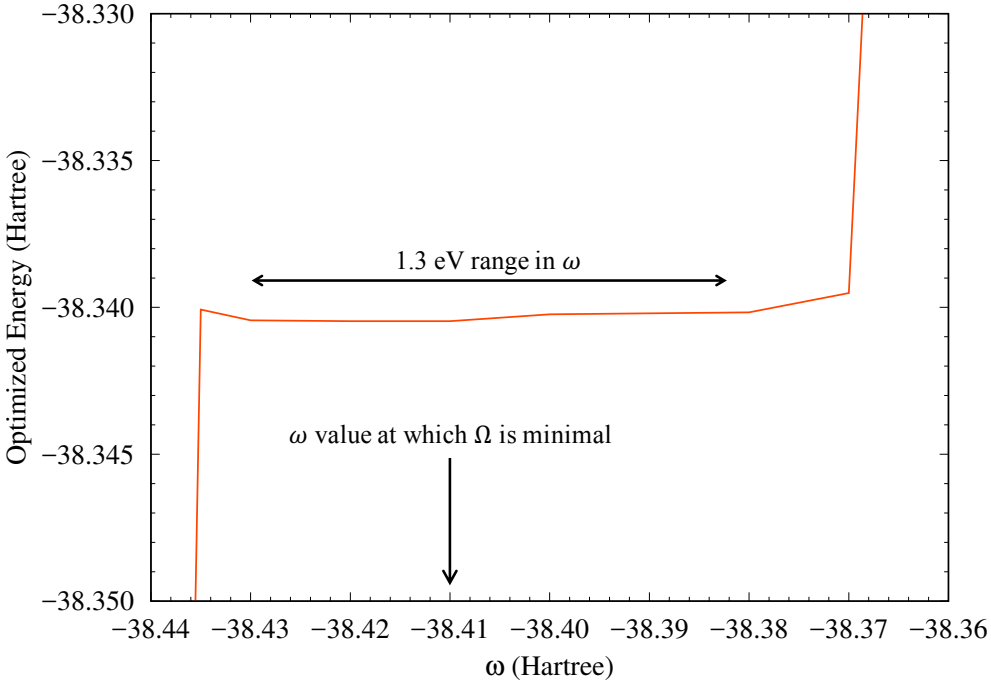

Figure A.1: Energy vs ω for the first excited state of CH_2 in a STO-3G basis.

JAGPs that were symmetry broken combinations of the 5th and 6th excited states. The broken symmetry was not imposed on the JAGPs, but instead emerged repeatedly from the optimization of Ω for a variety of different initial guesses. Essentially, the two states in question are dominated by the + and - combinations of two configurations, and Ω-optimized JAGP broke symmetry and gave each state as one or the other of these configurations in a way very reminiscent of Hartree-Fock's tendency to unrestrict and give one or the other of the two configurations of a diradical ground state. Upon performing a 2x2 diagonalization, the correct + and - combinations emerged and symmetry was restored, exactly as happens in many ground state scenarios. Note that as for Hartree Fock and a diradical, it is not necessarily possible to re-write the resulting linear combination of two JAGPs as a single JAGP with different parameters, and so it is very likely that an energetically accurate and symmetry-correct representation of these two states is simply beyond the variational freedom of JAGP. For reference, the magnitude of the overlap between the normalized JAGPs for these states was 0.13. The numbers we report for the excitation energies are those after the 2x2 diagonalization. See Table A.1 for different methods' absolute energies. In Å, the CH_2 geometry used was

C	-0.0722376285	-0.0574604043	0.0000000000
H	-0.0198102890	1.0990427214	0.0000000000
H	1.0664179823	-0.2665333714	0.0000000000

As an example of the insensitivity of the energy to the choice of ω, we have plotted in Figure A.1 the energy of CH_2's first excited state over a wide range of shifts. Over a range spanning 1.3 eV around the value of ω that minimizes Ω, we see that the predicted energy changes by less than 0.03 eV, confirming that the energy is indeed quite insensitive to the precise choice of ω so long as it is in the vicinity of this minimum. See Figure 1 of the main text for an example of what an Ω vs ω minimum looks like.

Table A.2: Absolute energies in Hartree for H_6 in a 6-31G basis.

State	JAGP	CIS	EOM-CCSD	FCI
0	-3.177	-3.043	-3.180	-3.181
1	-3.044	-2.840	-3.022	-3.051
2	-2.965	N/A	-2.897	-2.965

H_6

For H_6 we used the 6-31G basis [167] and shifts in Hartree of ω = -3.17, -3.15, and -3.15 for excited states 1 to 3, respectively. The geometry for H_6 was chosen as a regular hexagon with edge lengths (i.e. bond distances) of 1.5 Å. See Table A.2 for total energies.

Fock Space C_2

For C_2 with a Fock space random walk we used the 6-31G basis [167] and shifts in Hartree of ω = -75.85, -75.85, -75.65, -75.60, and -75.60 for excited states 1 to 5, respectively. Note that these shifts were not plotted in the main text as they are all *below* the JAGP ground state energy of -75.5915 Hartree and would not conveniently fit on the plot. As discussed in the main text, the finite variance of an approximate wave function causes the values of ω at which the Ω-minimum switches states to shift down in energy, and in C_2 this effect was large enough to push the switching energies below the ground state energy. The C_2 bond distance was 1.2425146399 Å. Total energies for different methods are shown in Table A.3.

Table A.3: Absolute energies in Hartree for C_2 in a 6-31G basis.

State	JAGP	CIS	EOM-CCSD	FCI
0	-75.591	-75.349	-75.620	-75.641
1	-75.508	-75.385	-75.534	-75.561
2	-75.508	-75.385	-75.534	-75.561
3	-75.438	N/A	-75.459	-75.520
4	-75.469	N/A	-75.458	-75.515
5	-75.469	N/A	-75.458	-75.515

Real Space C_2

For C_2 with a real space random walk we used the cc-pVTZ basis [168], both for the orbitals of the MSJ wave function and for the CIS, EOM-CCSD, CASSCF, and MRCI+Q calculations. For MSJ, these orbitals were augmented [9] to include the effects of nuclear cusps. The CAS expansion from which CSFs were taken for MSJ was the CAS space resulting from an equal weighted state averaged (8e,8o) CASSCF calculation (performed in C_1 symmetry) including the ground state and the first 5 singlet excited states. The CSF orbitals were taken as the optimized CASSCF orbitals. The Jastrow factors (one each for electron-nuclear, opposite-spin-electron, and same-spin-electron) were one dimensional functions of the magnitude of the interpartical distance, the natural logarithms of which were parameterized as a 10-section bspline with a cutoff radius of 5 Bohr for the electron nuclear and 10 Bohr for the electron-electron. Note that for the data presented for CASSCF and MRCI+Q, we instead used eight separate 4-state state-averaged CASSCF calculations, one in each D_{2h} representation, in order to reduce the computational cost of MRCI+Q.

We found that the downshifting of the switching values of ω was stronger in real space than in our Fock space tests, likely because there is more scope for inaccuracies that increase the variance when working in the complete basis of real space. We also observed that there can indeed be a significant difference in the optimized state when minimizing $E(\Psi)$ instead of $\Omega(\Psi)$, and as expected from our discussion in the main text comparing wave functions optimized with different variational forms results in biased and less accurate energy differences. To avoid this bias, which we again stress is directly analogous to the bias that is known to be introduced if optimizations of σ^2 are compared to those of E, we minimized $\Omega(\Psi)$ for all states, including the ground state, whose shift ω was chosen to lie near the point at which the minimum switched to the first excited state. As for other states, we observed that the precise choice of ω in the ground state optimization had only a very minor effect on the predicted excitation energies. The ω values used for the reported excitation energy calculations were -79.15 for the ground state and -79.00, -79.00, -78.70, -78.70, and -78.68 for excited states 1 to 5, respectively. The C_2 bond distance was 1.2425146399 Å. Table A.4 gives absolute energies for various methods.

Table A.4: Absolute energies in Hartree for C_2 in a cc-pVTZ basis, with MSJ evaluated with real space VMC.

State	MSJ(0.1)	MSJ(0.01)	CIS	CASSCF	EOM-CCSD	MRCI+Q
0	-75.758	-75.803	-75.401	-75.625	-75.773	-75.788
1	-75.722	-75.753	-75.447	-75.553	-75.724	-75.742
2	-75.722	-75.753	-75.447	-75.553	-75.724	-75.742
3	-75.686	-75.719	N/A	-75.528	-75.598	-75.709
4	-75.686	-75.719	N/A	-75.528	-75.598	-75.709
5	-75.674	-75.709	N/A	-75.525	-75.594	-75.697

A.2 Chapter 3

EOM-CCSD, Davidson-corrected MRCI (MRCI+Q) and FCI results were computed with MOLPRO [159], CIS results with QChem [161], and JAGP results with our own prototype Hilbert space quantum Monte Carlo code with one- and two-electron integrals imported from Psi3 [165]. In JAGP, we worked exclusively in the symmetrically orthogonalized $S^{-1/2}$ orbital basis and froze the C, N and O 1s orbitals at the RHF level. Unless noted otherwise, all sample lengths were 7.2×10^6.

A.3 Chapter 4

We have implemented our method within a development version of QMCPACK, [68] in which we have adapted the fast multi-Slater method [73, 74] to work with the cubic B-spline representation [68] of Kohn-Sham orbitals imported from QUANTUM ESPRESSO. [169, 170] We also customized the linear method optimizer to support complex numbers. For the correlation factor U in the main text, QMCPACK represents the one-dimensional functions V and W by 10-point cubic B-splines of the electron-nuclear (r_{ip}) and electron-electron (r_{ij}) distances. [68] Note that independent spline parameters are used for each chemical element and for same- and opposite-spin electron pairs, and that the parameters are optimized alongside the configuration interaction coefficients during the minimization of Ω.

Pseudopotentials

To avoid the unnecessary simulation of low-energy core electrons, we used Burkatzki-Filippi-Dolg (BFD) pseudopotentials [171] for Li, C, F, and Si, the norm-conserving pseudopotential of Shin et al [172] for O, and the semi-core-included pseudopotential of Krogel et al [173] for Zn.

The pseudopotential can be divided into a local and a nonlocal part[7],

$$V^{ps} = V_{loc}(\mathbf{R}) + \hat{V}_{nl} = \sum_i V^{ps}_{loc}(\mathbf{r}_i) + \sum_i \hat{V}^{ps}_{nl,i} \tag{A.1}$$

in which the nonlocal part of the pseudopotential acting on an arbitrary function of $\mathbf{r}_i$ is,

$$\hat{V}^{ps}_{nl,i} f(\mathbf{r}_i) = \sum_{l,m} V^{ps}_{nl,l}(\mathbf{r}_i) Y_{lm}(\Omega_i) \int_{4\pi} Y^*_{lm}(\Omega'_i) f(\mathbf{r}'_i) d\Omega'_i \tag{A.2}$$

in which $Y_{lm}(\Omega)$ are spherical harmonic functions.

The local part of the pseudopotential contribution to the local energy is simple,

$$\frac{V_{loc}(\mathbf{R})\Psi(\mathbf{R})}{\Psi(\mathbf{R})} = V_{loc}(\mathbf{R}) \tag{A.3}$$

as we see this part does not depend on the wave function.

The nonlocal part of the pseudopotential contribution to the local energy is a bit involved[7],

$$\begin{aligned}
\frac{\hat{V}_{nl}\Psi(\mathbf{R})}{\mathbf{R}} &= \sum_i \frac{\hat{V}^{ps}_{nl,i}\Psi(\mathbf{R})}{\Psi(\mathbf{R})} \\
&= \sum_i \sum_l V^{ps}_{nl,l}(\mathbf{r}_i) \sum_{m=-l}^{l} Y_{lm}(\Omega_{\mathbf{r}_i}) \int Y^*_{lm}(\Omega_{\mathbf{r}'_i}) \frac{\Psi(\mathbf{r}_1,...,\mathbf{r}_{i-1},\mathbf{r}'_i,\mathbf{r}_{i+1},...,\mathbf{r}_N)}{\Psi(\mathbf{r}_1,...,\mathbf{r}_{i-1},\mathbf{r}_i,\mathbf{r}_{i+1},...,\mathbf{r}_N)} d\Omega_{\mathbf{r}'_i} \\
&= \sum_i \sum_l V^{ps}_{nl,l}(\mathbf{r}_i) \frac{2l+1}{4\pi} \int P_l[\cos(\theta'_i)] \frac{\Psi(\mathbf{r}_1,...,\mathbf{r}_{i-1},\mathbf{r}'_i,\mathbf{r}_{i+1},...,\mathbf{r}_N)}{\Psi(\mathbf{r}_1,...,\mathbf{r}_{i-1},\mathbf{r}_i,\mathbf{r}_{i+1},...,\mathbf{r}_N)} d\Omega_{\mathbf{r}'_i}
\end{aligned} \tag{A.4}$$

in which the third line of the above expression is obtained by choosing the z axis along $\mathbf{r}_i$ and use the definition of spherical harmonics. P_l are Legendre polynomials.

DFT Calculations

All DFT calculations were performed with QUANTUM ESPRESSO 5.3.0 using a 350-Ry kinetic energy cutoff and a 4×4×4 k-point grid. All lattice constants were chosen based on experimental values, with an fcc lattice structure used for LiH and LiF with lattice constants of 7.716 and 7.625 Bohr, respectively. The diamond cubic structure was used for both C diamond and Si with lattice constants of 6.740 and 10.263 Bohr, respectively. The wurtzite structure was used for ZnO with lattice constants set to $a = 3.250$ Bohr and $c = 5.207$ Bohr. For ZnO, a 4-atom unit cell was used for DFT, while all other systems used a 2-atom unit cell for DFT.

Extrapolations

The extrapolations of the optical gap to bulk limit of diamond, silicon, and LiF are shown in Figure A.2, A.3, and A.4 respectively.

A.4 Chapter 5

JAGP results for N_2 and H_2O were obtained using Hilbert-space sampling via our own VMC software, which extracts one- and two-electron integrals from PySCF [174]. MSJ results for C_2 and the hydrogen ring were obtained using real-space sampling via QMCPACK [146, 147], with configuration state functions (CSFs) taken from GAMESS [175]. For JAGP, we work exclusively in the symmetrically orthogonalized "$S^{-1/2}$" one particle basis. The VMC sample size is universally chosen as 2.4×10^5, which produces statistical uncertainties whose standard deviations are less than 0.7 kcal/mol (0.03 eV) in all cases.

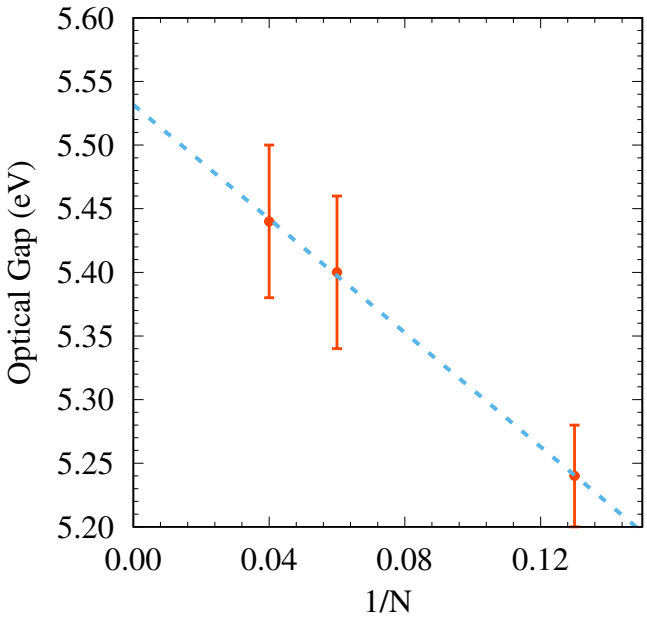

Figure A.2: Extrapolation to the bulk limit in diamond. Points are our VMC data while the dashed line is a linear fit against the inverse of the number of atoms N in the simulation cell.

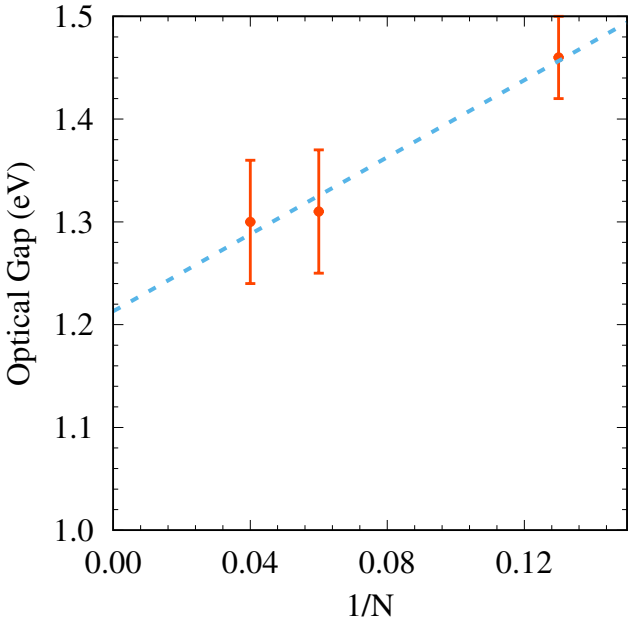

Figure A.3: Extrapolation to the bulk limit in silicon. Points are our VMC data while the dashed line is a linear fit against the inverse of the number of atoms N in the simulation cell.

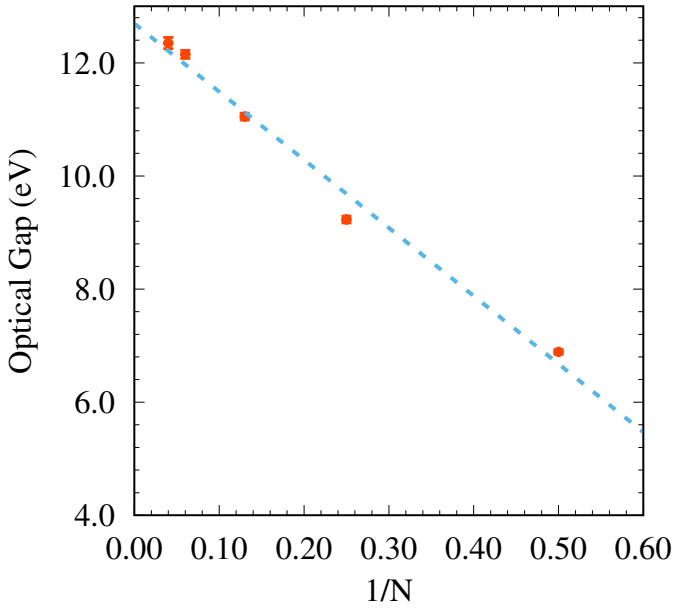

Figure A.4: Extrapolation to the bulk limit in LiF. Points are our VMC data while the dashed line is a linear fit against the inverse of the number of atoms N in the simulation cell.