Title

On the Construction of a Novel Mean Field Platform and Broadly Applicable Variational Principle Methods for Electronically Excited States

Author

Shea, Jacqueline

Publication Date

2021

Peer reviewed|Thesis/dissertation

On the Construction of a Novel Mean Field Platform and Broadly Applicable Variational Principle Methods for Electronically Excited States

by

Jacqueline R. Shea

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Eric Neuscamman, Chair
Professor Eran Rabani
Professor Kranthi Mandadapu

Spring 2021

On the Construction of a Novel Mean Field Platform and Broadly Applicable Variational Principle Methods for Electronically Excited States

Abstract

On the Construction of a Novel Mean Field Platform and Broadly Applicable Variational Principle Methods for Electronically Excited States

by

Jacqueline R. Shea

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Eric Neuscamman, Chair

Electronic structure theory is an evolving field with abounding potential applications to a multitude of interesting systems, including atoms and molecules in electronically excited states. Central to this dissertation are two critical components of excited state electronic structure methods – the ansatzes that describe a system's electronic configuration and the algorithms used to optimize them. In this thesis, advancements on both fronts are presented. To contextualize this new work, several prominent excited state ansatzes and optimization methods are reviewed, and their performances in applications to DNA photophysics and organic photovoltaics are examined. The strengths and weakness of the existing ansatzes inspired the construction of a flexible, computationally affordable excited state mean field wave function that is analogous to ground state mean field theory. This qualitatively accurate, state-specific ansatz provides the foundation upon which higher accuracy correlation methods are built that rival the accuracy of existing excited state theories at lower overall cost scaling. Further developments detailed in this thesis occur on the optimization side of electronic structure methods and relate to excited state variational principles. First, a novel analysis on the lack of size consistency within a class of excited state variational principles widely used in stochastic quantum chemistry methods is presented. A unique algorithm that transforms between variational principles on the fly while rigorously maintaining size consistency and state specificity is shown to eliminate this optimization-induced source of error. Finally, a generalized variational principle that guarantees state specificity via a unique global minimum and identifies electronic states based on a list of user-specified properties is defined, and its ability to resolve even energetically degenerate states in dense excitation manifolds is realized. From the construction of a novel mean field excited state wave function, to the analysis and restoration of size consistency in a class of excited state variational principles, to the development of a generalized variational principle, the research herein constitutes significant steps towards more robust, efficient, and accurate modelling of electronically excited states.

Dedicated to the memory of Bryan Shea

I couldn't have done it without you, Dad.

Contents

List of Figures

List of Tables

Acknowledgments

A great mentor sees our hidden talents and abilities and inspires us to be better than we know how to be, and I am incredibly fortunate to have had the best possible mentors a young scientist could ask for. First, I would like to express immense gratitude to Professor Eric Neuscamman. A teacher, an advocate, and a cheerleader, Eric truly encompasses what it means to be a great mentor, and his expertise, guidance, and understanding were invaluable at every step of the way through graduate school and very difficult moments in my personal life. From him, I have learned to be an independent researcher, a creative scientist, and open to new ideas. Eric gave me the freedom to explore the science that excited me, and what all started as a small proof-of-principle project blossomed into the research within this thesis. I am beyond thankful for all of his time, effort, and patience over the past five years and want to recognize the absolutely critical role he played in my success and future as a scientist. Second, I would like to thank Professor Steven Wheeler, who took a chance on me as an undergraduate and went above and beyond what is necessary for a mentor. I vividly remember my first day of research, where Steven sat down with me for hours and walked me through how to open a terminal, write code, and use a supercomputer. On that day, he inspired a frankly clueless undergraduate to pursue computational sciences and instilled in me an enthusiasm for this field that only continues to grow.

I also want to thank the amazing and empowering women in my life who inspire me every day. Mom, from helping me with homework in elementary school, to taking me to the crafts store once again for another box of rhinestones that were *absolutely necessary* for that English poster in ninth grade, to listening to every one of my rants about the tribulations of graduate school, you have always been a pillar of strength for me and supported me through my academic journey. I always believed in myself because I knew you believed in me, too. To my sister Laura, and my best friends Dr. Analise Doney and Dr. Katherine Oosterbaan, thank you for answering all of my late night texts and phone calls and for your emotional support. I could not ask for better friends. To my group members Rachel Clune, Becky Hanscam, Tarini Hardikar, and Connie Robinson, thank you for the constant laughs and your never-ending encouragement. Of everything I have accomplished during graduate school, I am truly the most proud of your growth as researchers and whatever part of that I was so lucky to have played.

To my roommate and dear friend Addison Schile, thank you for all of your support during graduate school, especially while I was writing my thesis. I promise I will eventually get around to washing those dishes.

Finally, to my father Bryan Shea – I struggle to find the words to express how deeply I wish you were still here. You were always so proud of me, telling all of your friends about my getting into Berkeley and each of my research papers, and I just know you would have printed this thesis out and carried it around to show to every single person you ever met. This one is for you, Dad.

Chapter 1

Introduction

1.1 Vertical Excitations in Quantum Chemistry

Methods for computationally studying excited states have many potential uses such as interpreting spectroscopic data, predicting experimental yields of photochemical processes, understanding photocatalysis and how the shapes of excited state orbitals can influence reaction rates, and designing materials with specifically tailored band gaps. A theoretical study is able to provide detailed information about the involved processes which may not be accessible through experimental investigations. Turning towards electronic structure theory is particularly enticing in scenarios where a current chemical system is inaccessible to experimentalists, whether that is due to cost, difficulty of synthesis, or system toxicity. Furthermore, electronic structure theory can expedite molecular design – for example, rather than synthesizing entire manifolds of novel photocatalysts in a laboratory and studying each catalyst's performance spectroscopically, theoretical chemistry can be used to identify the motifs that are most likely to be successful, thereby filtering out ineffective contending systems and lessening the work for synthetic chemists and spectroscopists. As theoretical research requires mainly a computer, paper, and several cups of coffee, it can also be viewed as a green alternative to laboratory research. The intent of this section is to briefly define vertical excitations in quantum chemistry and overview two applications (although there are many more) for which excited state theoretical chemistry can offer innovative solutions beyond what experiment is capable of.

Vertical excitations

A vertical excitation is a transition between electronic states, such as the ground state and the first excited state, without displacements in the molecular geometry. As shown in Fig. 1.1, relative to the adiabatic excitation energy, in which the energy difference is taken at fully relaxed geometries for both electronic states, the vertical excitation energy is larger. The vertical excitation energy is more representative of the instantaneous change in the system's energy upon excitation and is most relevant when the Born-Oppenheimer approximation is

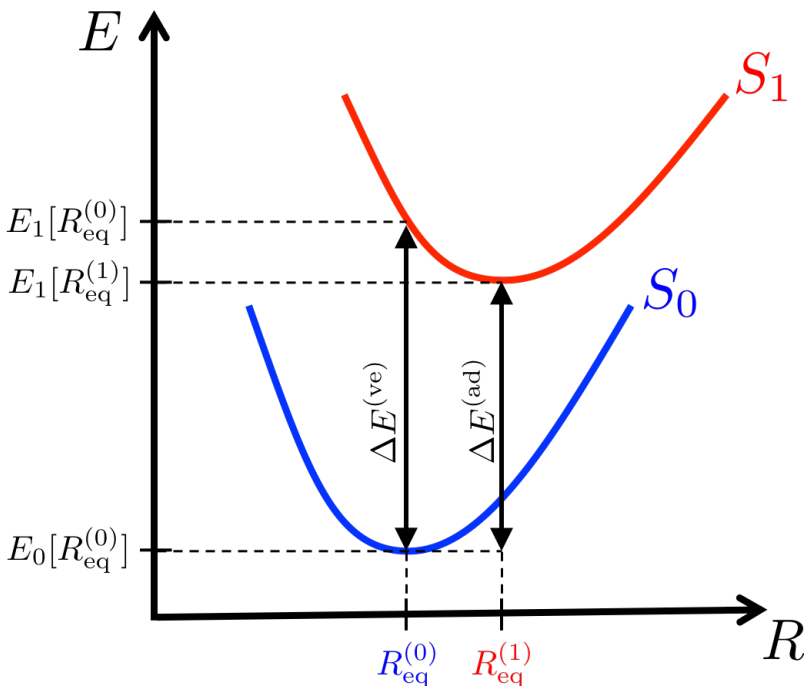


Figure 1.1: Definitions of the vertical excitation energy $\Delta E^{(\mathrm{ve})}$ and adiabatic excitation energy $\Delta E^{(\mathrm{ad})}$ with respect to geometry coordinate R for an $S_0 \rightarrow S_1$ excitation

valid. This approximation decouples the motion of the nuclei from the motion of the electrons under the assumption that due to the relative masses of nuclei and electrons, the time scale of nuclear motion versus the time scale of electron motion is such that the electrons effectively experience nuclei that are fixed in space.[1–3] The Born-Oppenheimer approximation is an appropriate simplification for a significant portion of electronic structure theory, though there are systems in which it breaks down. For example, the Born-Oppenheimer approximation is not valid in multi-dimensional systems that have intersecting ground and excited state potential energy surfaces, that is, a conical intersection. In geometries near and at the conical intersection, the electron-nuclear coupling is nontrivial and cannot be neglected.[4] However, this thesis focuses on systems and geometries for which the Born-Oppenheimer is applicable and thus vertical excitation energies are relevant. It is prudent to note that even in some systems for which the Born-Oppenheimer approximation is valid, such as the dye-sensitized solar cells (DSCs) discussed later within this introduction, ignoring the change in energy due to the relaxation of the molecule in the excited state can lead to consequential errors when considering dynamics of the system as this reorganization energy can be appreciable.[5–7] In fact, such approximations could lead to significant errors in the rates of charge transfer reactions and limit our understanding of these systems, our ability to simulate experimental observations, and our predictions for novel DSC components.[5–7]

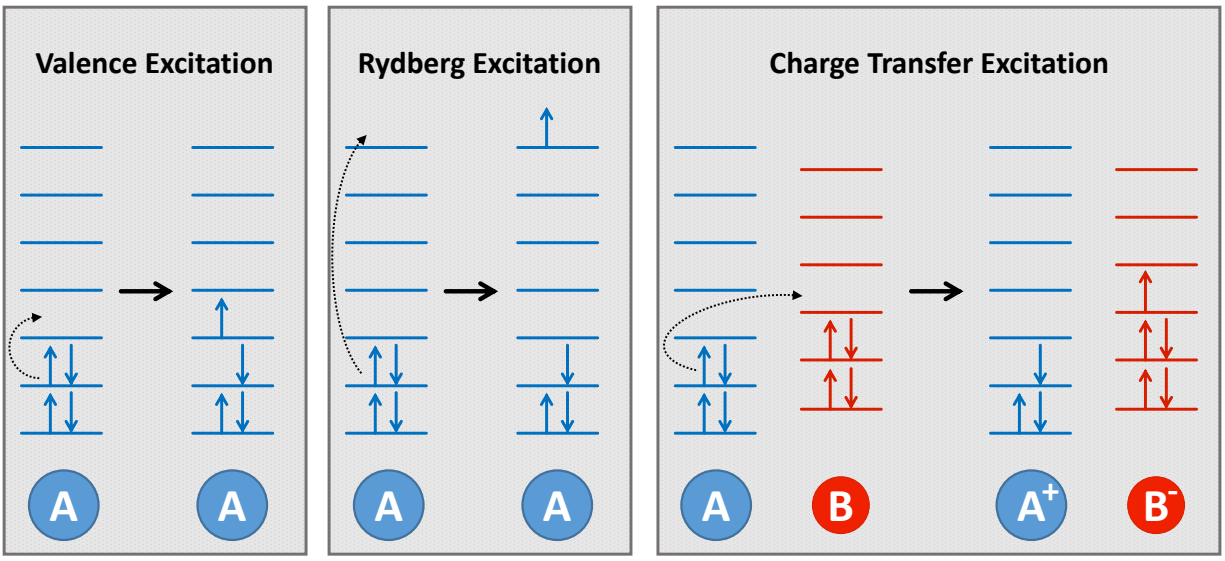


Figure 1.2: Definitions of valence, Rydberg, and charge transfer excitations

In such cases, computing valence excitation energies accurately and affordably is still critical and will interplay with further studies, such as excited state geometry optimization, that would address the reorganization energy.

Three classes of vertical excitations will be discussed in this work – valence excitations, Rydberg excitations, and charge transfer (CT) excitations – which are depicted in Fig. 1.2. Valence excited states are excitations from a molecule's highest (energy) occupied molecular orbitals, or HOMOs, to its lowest (energy) unoccupied molecular orbitals, or LUMOs. The first singlet excitation in many organic molecules is dominated by the HOMO→LUMO valence excitation. A Rydberg excitation occurs when an electron is excited into a high energy, spatially diffuse unoccupied molecular orbital. As the electron occupies a region further from the nucleus, the wave function and potential energy surface (PES) of a Rydberg state more closely resembles those of the cationic ground state than those of a valence excited state. Rydberg excitation energies typically start around 50,000 cm^{-1}, or about 6.2 eV.[8] However, Rydberg states do mix with valence excitations, and Rydberg-valence interactions are present in both closed- and open-shell molecules.[8–14] These states and their interactions with the valence states have an important effect on PESs,[11, 13], spectroscopy, and dynamics [9, 11–13]. These interactions typically describe the more general mixing of states with different characters such as the superposition of covalent and zwitterionic states.[15] Like Rydberg excitations, there is also a significant change in electron density and thus orbital relaxation in charge transfer (CT) excitations. However, these excitations involve an excitation between two subsystems, creating a separation of charge. As a positive and negative charge are created, the energy of a charge transfer excitation scales as $\frac{1}{R}$, where R is the distance between the positive and negative ions, as they are electrostatically attracted to

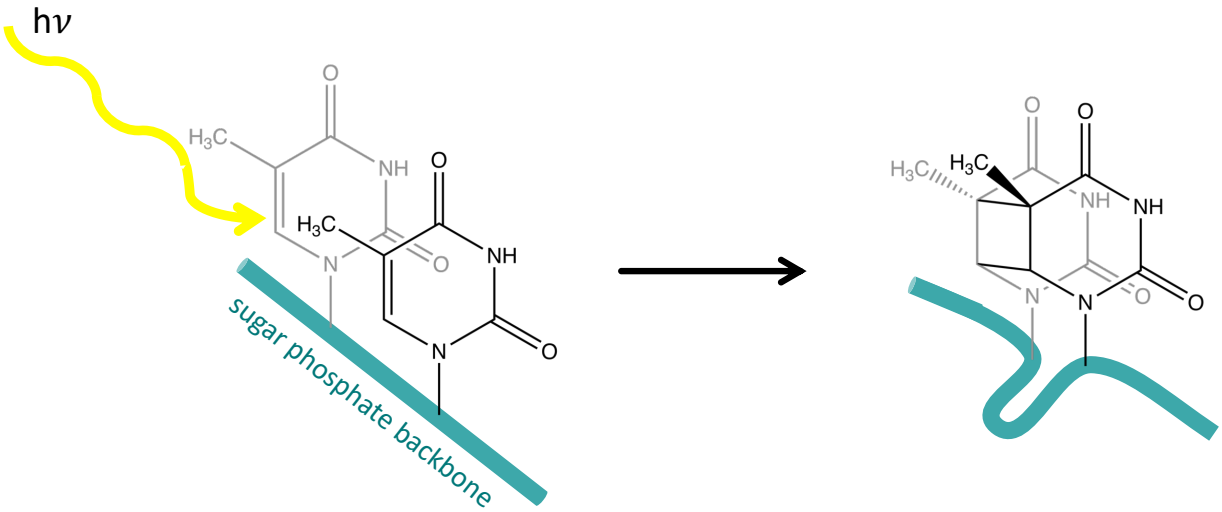


Figure 1.3: Thymine dimerization in DNA upon photoexcitation

each other. Charge transfer excitations are of particular interest in energy transfer, as the movement of electrons can be used to generate a current.

Having defined the excitations of interest to this thesis, let us now strengthen the argument for improving theoretical methods by briefly considering two fields where excited state chemistry is essential and where electronic structure theory should be able to provide significant insights.

DNA Photophysics

Understanding the interaction of UV light and DNA as well as its building blocks can provide us with answers to questions like "What damage does light cause in our genetic code?" and "How can we intervene before this damage leads to cancer?".[16] The CDC summarized several pieces of research that are particularly unnerving. First, approximately 40% of Americans get sunburned each year. This fact becomes particularly insidious when paired with the following statistics. Skin cancer is the most common form of cancer in America, affecting almost 10% of the population, and incidence rates are increasing. Most frighteningly, 90% of melanoma are caused by skin cell damage from UV radiation.[17, 18] (If there is only one takeaway from this thesis, the author hopes it is the importance of sunscreen and protective clothing during periods of sun exposure.) These alarming statistics motivate research in the field of DNA photophysics as it may lead to new understanding of the disease, more effective treatments for it, and maybe even the discovery of alternative preventative measures.

The adverse effects of UV radiation can be traced down to the molecular level – the photoexcitation of nucleotides. UV irradiation of thymine and cytosine excites them to the singlet $\pi \rightarrow \pi^*$ state and opens up their carbon-carbon double bond, allowing them to

dimerize with a neighboring cytosine or thymine on the same sugar backbone.[19–23] These dimers form a stiff kink in the DNA backbone as shown in Fig. 1.3, which disrupts the hydrogen bonds at the site of the dimer and at adjacent base pairs[19] and can lead to DNA polymerase mistranslating the genetic code, causing a mistake in DNA translation.[24] A single point mutation in the genome can prevent DNA replication and transcription, leading to cell death.[19] Moreover, if this mutation occurs in a gene that controls cell growth, such as the p53 tumor suppression gene, it can lead to cancer.[25–28] While the probability of this reaction is low because the bases are rarely in the correct reactive orientation, pump-probe spectroscopy indicates that when the nucleobases are positioned in the reactive orientation, the dimerization is near spontaneous, occurring in less than a picosecond.[29, 30]

The importance of understanding all photoexcited nucleobase mechanisms is self-evident. However, probing the photophysics of DNA nucleotides via experiment is no easy task. All nucleotides have a bright singlet $\pi \rightarrow \pi^*$ state that is overwhelmingly populated when a nucleotide is excited from the singlet ground state and two dark states – a singlet $n \rightarrow \pi^*$ and a triplet $\pi \rightarrow \pi^*$ state. Analysis of the dark states requires significantly more complicated spectroscopic techniques and data analysis since they are reached via internal conversion (IC).[23, 31, 32] The dark states cannot be neglected either, as they have very interesting effects on the photophysics. For example, the dark singlet $n \rightarrow \pi^*$ state in pyrimidine bases is considered a long-lived "trap" state, as it traps about half of the population of pyrimidine bases, preventing them from fully relaxing to the ground state for even more than 100 ps, while the remaining population decays on an ultrafast timescale.[23, 33] Regarding thymine dimerization specifically, there are competing theories regarding the multiplicity of the excited state that leads to the dimer; however, the previously described singlet state mechanism is the predominant viewpoint.[23, 29, 30, 34] As different experimental techniques seem to deliver different results, these dark states are the cause of many debates including the de-excitation mechanism from the dark states to the ground state[35–41] and the effect of the sugar backbone on the lifetimes of excited nucleobases.[23, 33]

In scenarios where experiments disagree or where observations are difficult to make due to the ultrafast and sometimes metastable character of the reactions, theory is often needed to interpret the spectra and provide comprehensive insight into the system.[16, 23, 42–45] However, the level of theory used to characterize the ground and excited state potential energy surfaces and other electronic properties also leads to different conclusions – different conclusions from theoretical studies,[36, 46, 47] and different conclusions from experiments.[39–41] Correctly describing DNA interactions via electronic structure theory is quite challenging, as the level of theory must be able to balance several dichotomous phenomena – bond breaking and bond formation,[19, 48] hydrogen bonds[48, 49] and π-stacking interactions,[48, 50–52] and local (valence, Rydberg)[48, 52, 53] and charge transfer excitations.[45, 48, 53] Furthermore, realistic simulations of DNA photophysics also demand the inclusion of environmental effects from adjacent nucleobases, the sugar backbone, and the solvent.[33, 39, 48, 54, 55]

Organic Photovoltaics

The field of organic photovoltaics (OVPs) is also a significant area of research where theory has brought and can bring further insight. OVPs are low cost, easily fabricated[56] solar cells that are not only flexible, such that they could eventually be incorporated into consumer goods (think a backpack that charges your cell phone with solar power) but are also semi-transparent, and progress is being made towards making them clear, such that they could even be used over windows.[57] With recent groundbreaking advancements, the record power conversion rate for an organic solar cell is over 12%.[58] Some drawbacks, however, include a limited lifetime,[57, 59] particularly since they degrade when exposed to the elements.[60–63] Improving the lifetime of the production materials would be revolutionary, as an efficiency of 10-15% and a production material lifetime of over 10 years would make OPV-captured solar energy the cheapest energy source, and green, renewable energy would be economical for the public.[64, 65]

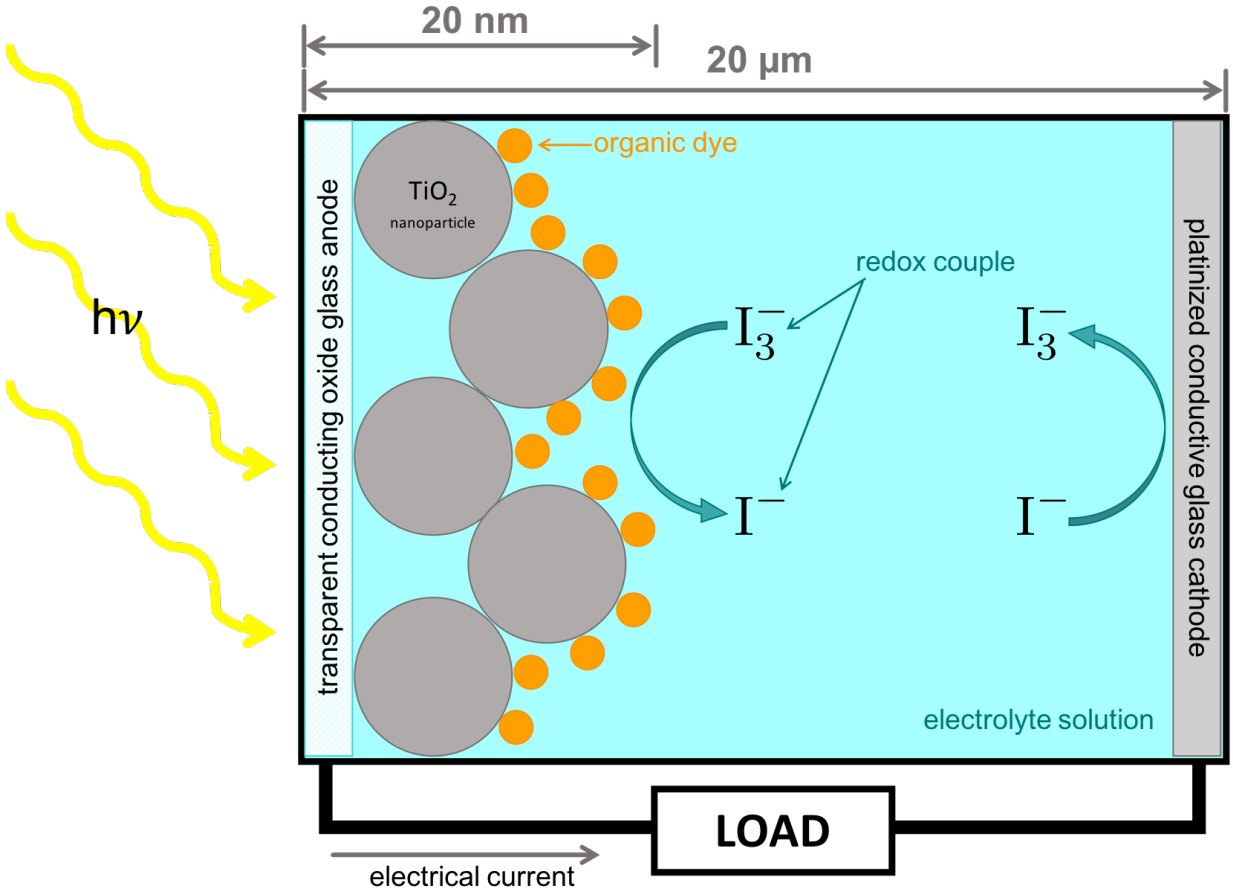


Figure 1.4: Schematic of a dye-sensitized solar cell

There are two main types of OVPs – dye-sensitized solar cells (DSCs)[66–69] and heterojunction solar cells (HJCs).[70–72] Let us consider the former, while acknowledging that there are certainly opportunities for electronic structure theory in HJC applications as evidenced by the Harvard Clean Energy Project, a large-scale computational project dedicated to screening and designing new organic photovoltaic materials.[72] As shown in Fig. 1.4, DSCs work by layering first a transparent, conductive oxide glass, then a semi-porous layer of (typically TiO_2) semi-conducting nanoparticles. A thin monolayer of an organic dye is then deposited onto the nanolayer surface, forming covalent bonds between the two layers. Between the glass-nanoparticle-dye photoanode and a counter electrode composed of platinum coated conductive glass is an electrolyte solution containing I^- and I_3^-. When sunlight shines on the photoanode, it travels through the glass and semi-porous nanoparticle layer and excites an electron on the organic dye. The excited dye electron then is injected into the nanoparticle's conduction band, where it flows flows through the semi-conductor to the photoanode, then through a wire to power the external load, and the circuit is completed as the wire connects with the platinized cathode. The charges are rebalanced by the electrolytic solution as I_3^- reduces the cationic dye molecule, and the positive charge is transferred to the cathode by I^-.[66–69]

How can theory assist in developing more efficient DSCs? Electronic structure theory can be used, for example, to investigate individual dye molecules to identify dyes that strongly absorb light from the visible range of the spectrum (400-750 nm, or a photon with energy between 3.11 and 1.65 eV) and for which the electron transfer from the dye to the nanoparticle is energetically favorable.[73] Not only must this theory get valence excitations right within a few tenths of an eV, it must also treat charge transfer excitations with equitable accuracy. While not two formally discrete systems, in these intramolecular charge transfer excitations, the electron moves from one area of the molecule to a distinctly separate area of the molecule, causing a significant change in the electron density. Since the most efficient dyes[74] undergo metal-ligand or ligand-ligand intramolecular charge transfer excitations upon excitation by the photon,[75–81] the method must be able to accurately capture the significant resulting orbital relaxations and ideally compare these dyes to others where valence excitations dominate. Moreover, these excitations contain many different electron configurations,[75, 77–82] so the theory needs to be able capture this superposition as well. One such example is the N3 ruthenium dye, which has achieved 10% efficiency in DSCs[83, 84] and is considered a benchmark to compare new dyes against.[73] The excitation in N3 is dominated by multiple single valence excitations between its four HOMOs which, as shown in Fig. 1.5 at the HF/LANL2DZ[85–87] level of theory, are localized on the ruthenium center and the NCS^{-1} groups, and its four LUMOs which are localized on the bipyridine ligands, leading to a significant redistribution of charge upon excitation.[75, 77–81] Finally, current theory tends to focus on the dye's ground state LUMO and its energy relative to the energy of the nanoparticle's conduction band, suggesting this difference determines whether the electron injection is thermodynamically favorable.[72, 73, 88, 89] However, the author suggests here that now that the LUMO has an electron in it, its energy and shape can relax significantly, and the energy of this excited orbital is more relevant to the kinetics of the electron injection.

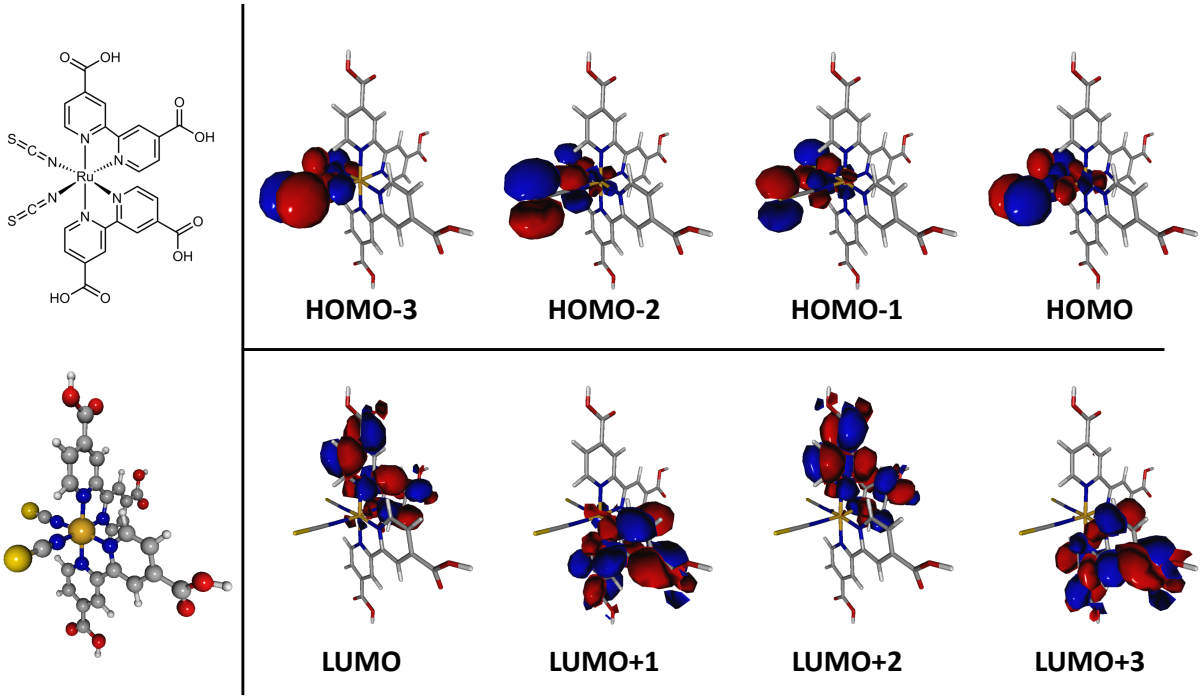


Figure 1.5: The N3 ruthenium dye's Lewis structure, 3D structure, four highest occupied molecular orbitals and four lowest unoccupied molecular orbitals computed at the HF/LANL2DZ level of theory

Accurate comparisons of the excited dye's orbital energies to the nanoparticle's conduction band requires fully relaxed excited state orbitals, another goal for the theoretical method used to study DSCs.

Keeping these two applications in mind, the remaining structure of the introduction is as follows. First, a brief overview of some fundamentals of modern electronic structure theory will contextualize the discussion of the author's viewpoints on the ideal properties of an excited state method. The next two sections will then review the most prominent ansatzes and optimization methods used in excited state electronic structure theory, with the final section outlining the remaining chapters in this thesis.

1.2 An Overview of Modern Electronic Structure Theory

The time-independent Schrödinger equation

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

where $\hat{H}$ is the Hamiltonian operator that describes the physics governing a system, E is the energy of a system, and $|\psi\rangle$ is a stationary state of the system, is fundamental to the majority of modern electronic structure theory. As we are operating under the Born-Oppenheimer approximation (see Sec. 1.1) and approximate wave functions that can be expressed by a finite set of N_{bas} orbitals $\{\phi\}$ within this thesis, Eq. 1.1 involves the electronic Hamiltonian, given by[90]

$$\hat{H} = \hat{h} + \hat{V} \tag{1.2}$$

$$\hat{h} = \sum_{pq}^{N_{\text{bas}}} \sum_{\sigma\in\{\alpha,\beta\}} h_{pq}\hat{a}^{\dagger}_{p\sigma}\hat{a}_{q\sigma} \tag{1.3}$$

$$\hat{V} = \frac{1}{2}\sum_{pqrs}^{N_{\text{bas}}} \sum_{\sigma,\tau\in\{\alpha,\beta\}} (pq|rs)\hat{a}^{\dagger}_{p\sigma}\hat{a}^{\dagger}_{r\tau}\hat{a}_{s\tau}\hat{a}_{q\sigma}. \tag{1.4}$$

where indices p, q, r, and s run over all orbitals in the basis, $\hat{a}^{\dagger}_{p\sigma}$ is the creation operator that puts an electron of spin σ in orbital ϕ_p, and $\hat{a}_{q\tau}$ is the annihilation operator that removes an electron with spin τ from orbital ϕ_q. The one-electron operator $\hat{h}$ in Eq. 1.3 contains two pieces – the kinetic energy operator for each electron and the Coulomb operator that accounts for the attraction between each electron and each of the N_{nuc} nuclei with charges $\{Z\}$ – and has corresponding one-electron integrals h_{pq} defined in atomic units by[91]

$$h_{pq} = \int \phi^*_p(\vec{r}_1)\bigg(-\frac{1}{2}\nabla^2_{\vec{x}_1} - \sum_{M}^{N_{\text{nuc}}} \frac{Z_M}{|\vec{x}_1 - \vec{x}_M|}\bigg)\phi_q(\vec{r}_1)\, d\vec{r}_1 \tag{1.5}$$

where the i-th electron's position $\vec{x}_i \in \mathbb{R}^3$ and spin $\sigma_i \in \{\alpha, \beta\}$ are represented by the composite variable $\vec{r}_i = \{\vec{x}_i, \sigma_i\}$. The two-body operator $\hat{V}$ in Eq. 1.4 is given by the Coulombic electron-electron repulsion operator and has corresponding two-electron integrals $(pq|rs)$ defined by[91]

$$(pq|rs) = \int\int \phi^*_p(\vec{r}_1)\phi_q(\vec{r}_1)\, \frac{1}{|\vec{x}_1 - \vec{x}_2|}\, \phi^*_r(\vec{r}_2)\phi_s(\vec{r}_2)\, d\vec{r}_1\, d\vec{r}_2. \tag{1.6}$$

The total energy of a system is thus given by the sum of the electronic energy E and the classical Coulombic nuclear repulsion energy E_{nn}.[2] The eigensolution of the Schrödinger equation with the lowest corresponding energy eigenvalue is the ground state of a system, while higher energy stationary states of the Hamiltonian are excited states. In the Hilbert space given by the span of orbitals $\{\phi\}$, the ground state is a global minimum and excited states are saddle points.[92–94]

Mean Field Theory

Deceptively trivial to write down, the Schrödinger equation is impossible to solve exactly for systems with more than one electron. The distance-dependence of the strong Coulombic repulsion between the negatively charged electrons leads to the infamous many-body problem,

in which there are too many interactions to consider simultaneously and thus no analytic solution to Eq. 1.1 exists for multi-electron systems.[95] Though it seems like we have been stopped before we even started, don't fret! One way forward is through mean field theory, an approximation that reduces the many-body problem down to a one-body problem by substituting all interacting terms with an average, or effective, potential. Let's consider the simplest mean field theory in electronic structure.

Hartree Theory

Taking the simplest non-interacting, N-body Hamiltonian as a sum of currently undefined one-body operators, $\hat{h}$, which act on the i-th particle,

$$\hat{H}_0 = \sum_i^N \hat{h}(i) \tag{1.7}$$

one can derive that the eigenstates of $\hat{H}_0$ are the product of the eigenstates

$$\hat{h}(i)|\phi_j(\vec{r}_i)\rangle = \epsilon_j|\phi_j(\vec{r}_i)\rangle \tag{1.8}$$

since the individual electrons are uncoupled. Therefore, the eigensolution to Eq. 1.7 is

$$|\Psi_{\mathrm{H}}(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N)\rangle = \phi_1(\vec{r}_1)\phi_2(\vec{r}_2)...\phi_N(\vec{r}_N) \tag{1.9}$$

$$E = \sum_i^N \epsilon_i. \tag{1.10}$$

This is, in fact, a notable property of traditional mean field theories – since the particles are not interacting, the total representation of the system (be this a partition function, wave function, etc.) is product-factorizable with respect to each particle, and the total energy is the sum of the one-particle energies.[96, 97]

There is, however, a predicament regarding the Hartree wave function. Upon swapping two electrons,

$$|\Psi_{\mathrm{H}}(\vec{r}_2, \vec{r}_1, ..., \vec{r}_N)\rangle = \phi_1(\vec{r}_2)\phi_2(\vec{r}_1)...\phi_N(\vec{r}_N), \tag{1.11}$$

we see that the Hartree wave function fails to represent the indistinguishability of electrons or the fermionic anstisymmetry principle which requires that for a system of N indistinguishable fermions, exchanging two particles yields an anti-symmetric wave function, i.e. $\Psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N) = -\Psi(\vec{r}_2, \vec{r}_1, ..., \vec{r}_N)$.[2] As this approach fails to take in even the most basic physics of fermionic systems, let us consider a slightly more involved wave function.

Hartree-Fock Theory

Unlike the product-separable Hartree wave function, the Hartree-Fock (HF) wave function is antisymmetrized in accordance with fermionic physics. Mathematically, the normalized

N-particle HF wave function can be expressed compactly via the Slater determinant

$$|\Psi_{\text{HF}}\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{r}_1) & \phi_2(\vec{r}_1) & \dots & \phi_N(\vec{r}_1) \\ \phi_1(\vec{r}_2) & \phi_2(\vec{r}_2) & \dots & \phi_N(\vec{r}_2) \\ \vdots & \vdots & & \vdots \\ \phi_1(\vec{r}_N) & \phi_2(\vec{r}_N) & \dots & \phi_N(\vec{r}_N) \end{vmatrix} \tag{1.12}$$

which contains $N!$ unique terms that differ by permutations of the electron coordinates and is commonly expressed in shorthand by $|\Psi_{\text{HF}}\rangle = |\phi_1\phi_2...\phi_N\rangle$, where the molecular orbital ϕ_i is defined as a linear combination of N_{bas} atomic orbitals (or basis functions) $\{\chi_j\}$.

$$\phi_i(\vec{r}) = \sum_j^{N_{\text{bas}}} c_{ji}\chi_j(\vec{r}) \tag{1.13}$$

From the determinant expression, it is clear how fermionic statistics is maintained – swapping two columns or rows returns an identical determinant but introduces a negative sign. Due to the antisymmetrization, the HF wave function does not fit the traditional definition of a mean field theory where all particles are totally non-interacting, have product-factorizable wave functions, and have energies that sum to the total system energy, which makes sense as the HF wave function now includes Pauli correlation. Pauli correlation, based on the Pauli exclusion principle, prohibits fermions with the same spin from occupying the same orbital. However, as the Hartree-Fock wave function contains the minimum correlation required to describe the ground state of an N-electron fermionic system, it is considered the central mean field wave function in atomic and molecular electronic structure theory and, conveniently, is an upper bound for the exact ground state energy.[2]

The original Hartree-Fock method used Lagrangian minimization to find the energy of $|\Psi_{\text{HF}}\rangle$ via the ground state variational principle (further discussed in Section 1.5) under the constraint that the final set of molecular orbitals are orthonormal, i.e. the following function is minimized with respect to the wave function parameters C defined in Eq. 1.13.

$$L = \frac{\langle\Psi_{\text{HF}}|\hat{H}|\Psi_{\text{HF}}\rangle}{\langle\Psi_{\text{HF}}|\Psi_{\text{HF}}\rangle} - \sum_{ij}\epsilon_{ij}\big(\langle\phi_i|\phi_j\rangle - \delta_{ij}\big) \tag{1.14}$$

The optimized Lagrange multipliers $\{\epsilon\}$ take on physical meaning, where ϵ_{ii} is the energy of the i-th molecular orbital.

The generalized eigenvalue equation is the foundation of the self consistent field (SCF) method used in most Hartree-Fock calculations to optimize the Hartree-Fock wave function. Given by

$$FC = SC\epsilon \tag{1.15}$$

where C is the molecular orbital coefficient matrix that defines the HF orbitals relative to the basis functions as defined in Eq. 1.13, S is the basis function overlap matrix with matrix

elements $S_{ij} = \langle\chi_i|\chi_j\rangle$, ϵ is a diagonal matrix of the orbital energies, and F is the matrix representation of the Fock operator $\hat{F}$, a one-particle Hamiltonian with an average of the electron repulsions felt by an electron, i.e.

$$\hat{F} = \hat{h} + \hat{V}_{\mathrm{HF}} \tag{1.16}$$

$$\hat{V}_{\mathrm{HF}} = \sum_{pq}^{N_{\mathrm{bas}}} \sum_{i}^{N_{\mathrm{occ}}} \sum_{\sigma\in\{\alpha,\beta\}} \big((pq|ii) - (pi|iq)\big)\hat{a}^{\dagger}_{p\sigma}\hat{a}_{q\sigma} \tag{1.17}$$

where $\hat{h}$ is defined in Eq. 1.3, $\hat{V}_{\mathrm{HF}}$ is expressed in the molecular orbital basis, and the index i runs over the N_{occ} occupied orbitals (note that, for efficiency, most quantum chemistry codes work in the atomic orbital basis). Referred to as the self consistent field (SCF) method, the generalized eigenvalue equation is iteratively solved with respect to C and ϵ until both matrices stop changing. In ground state HF theory, as the Fock matrix is diagonalized, the occupied and virtual orbitals are assigned based on the Aufbau principle, that is, the occupied orbitals are the N_{occ} molecular orbitals with the lowest orbital energies and the remaining N_{vir} orbitals are virtual orbitals.[2, 98]

Due to the $\mathcal{O}(N^4)$[2] scaling of HF theory and its abilities to recover approximately 99% of the total system energy and describe orbitals with qualitative accuracy for much of atomic and molecular chemistry,[98] either the HF or the closely related hybrid Kohn-Sham density functional theory (KS-DFT), which works on very similar principles,[99, 100] are used as a starting point for the majority of atomic and molecular computations, serving as platforms upon which one can build higher-order methods that will recover correlation lost in the mean field approximation.[98] For the rest of this thesis, the reference wave function of a higher order method will be referred to as $|\Phi_0\rangle$ which is synonymous with $|\Phi_{\mathrm{HF}}\rangle$ unless otherwise specified.

Perturbation Theory

One such method is second-order Møller-Plesset perturbation theory (MP2), which utilizes Rayleigh-Schrödinger perturbation theory[2, 101] to recover correlation originating from connections between the HF wave function and doubly-excited determinants. In this elegant solution, which requires partitioning the true Hamiltonian given in Eq. 1.2 into a zeroth order Hamiltonian given by the one-electron Fock operator (Eq. 1.16) and a first-order perturbative Hamiltonian representing the difference between the true two-electron Coulomb potential (Eq. 1.4) and the effective one-electron Fock potential (Eq. 1.17), the zeroth-order energy and first-order energy correction sum to the HF energy, and the second-order perturbative energy correction is given by

$$\Delta E^{(2)}_{\mathrm{MP2}} = -\frac{1}{4}\sum_{ij}^{N_{\mathrm{occ}}}\sum_{ab}^{N_{\mathrm{vir}}} \frac{\left|\langle\Phi_{\mathrm{HF}}|\hat{V}|\Phi^{ab}_{ij}\rangle\right|^2}{E^{(0)} - E^{(0)}_{ijab}} = -\frac{1}{4}\sum_{ij}^{N_{\mathrm{occ}}}\sum_{ab}^{N_{\mathrm{vir}}} \frac{\left|(ia|jb) - (ib|ja)\right|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{1.18}$$

where $|\Phi_{ij}^{ab}\rangle$ is a doubly-excited determinant defined as $|\Phi_{ij}^{ab}\rangle = \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i |\Phi_{\text{HF}}\rangle$.[2, 101]

Relatively quite effective for molecules and atomic systems already well-defined by the HF wave function, such as non-degenerate molecules with large HOMO-LUMO gaps that are without significant strong electron correlation and at their equilibrium geometry, MP2 is a highly successful, $\mathcal{O}(N^5)$ scaling method that recovers a large quantity of the remaining electron correlation energy, defined as $E_{\text{corr}} = E - E_{\text{HF}}$. However, since it depends on the qualitatively accurate description provided by the single-determinant ground HF wave function, it is unsuitable for applications to excited states.[2, 101]

1.3 Desirable Properties in a Method

A computational method consists of two parts – the ansatz that describes the electronic configuration of a system, be that the wave function or the electron density, which contains optimizable parameters and the optimization algorithm used to vary those parameters such that the ansatz is a Hamiltonian eigenstate. An ultimate goal when developing a method is to give it predictive power, i.e. the ability to draw reliable and trustworthy conclusions about materials or molecular systems, so that it can be used with confidence to study novel systems. The author considers three criteria to be most desirable in a computational method in order to give it predictive power and broad utility: physical accuracy, computational affordability, and systematic improvability.

Physical accuracy. In contrast to chemical accuracy, which describes the quantitative accuracy of measurable observables versus experiment, by physical accuracy the author refers to the qualitative accuracy of a method's results. For modelling excited states, important properties to recover include, but are not limited to, correct asymptotic behavior of the energy as subsystems are separated, correct scaling behavior as system size increases, reasonable descriptions of orbitals or electron density, correct excitation character descriptions (i.e. the contribution of single, double, etc. excitations to the electronic configuration), and spin purity.

Size consistency and size intensivity are two properties of methods that exhibit proper asymptotic behavior as subsystems are separated and that have proper energy scaling with respect to system size.[102–105] Consider a system with two non-interacting molecules, A and B. As depicted in Figure 1.6, in a size consistent method, the resulting energies of treating each subsystem in independent calculations, E_A and E_B, must sum to the total energy of a calculation that includes both systems, E_{AB}, and the resulting wave function from the combined calculation ψ_{AB} must product-factorize into wave functions of each subsystem, ψ_A and ψ_B. In a size intensive method, intensive properties of each subsystem, such as excitation energies, remain unchanged as non-interacting subsystems are added to the calculation. That is, if subsystem A has an excitation energy of ΔE_A, the addition of non-interacting subsystem B to the calculation does not affect the value of ΔE_A. Size consistency and size intensivity are important properties as the size of a system increases, as errors induced by a lack of either property are not unique to non-interacting subsystems – they would still be present

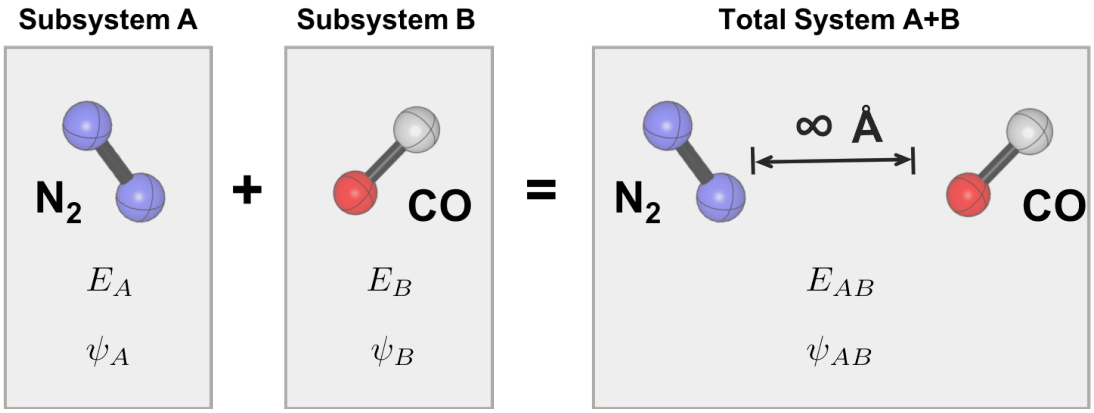


Figure 1.6: Size consistency within two non-interacting subsystems

even in computations of large, contiguous systems (such as a DNA fragment or large organic dye), and the accuracy of the calculation would degrade as the system size increased.

Touched upon in Sec. 1.1, the shapes and energies of excited state orbitals differ from those of the ground state. In order to fully understand the photochemistry governing excited systems such as the photoexcited mechanism of thymine dimerization and the rate of electron injection in DSCs, accurate representations of orbital overlaps and relative donor-acceptor orbital energies are critical. Thus, theories that can return fully relaxed, excited-state-specific orbitals are especially enticing.

Computational affordability. Highlighted by Nobel laureate John Pople, the practicality of a method is related to its ability to be applied to systems of interest "in reasonable times and at reasonable cost",[106] which will be referred to as a method's computational affordability. The phrase computational affordability is deliberately ambiguous, as computing resources range from a single-core laptop to highly-parallelized supercomputers.[107] However, even with these powerful machines, some methods such as Full Configuration Interaction (FCI) become infeasible even for relatively small molecules and thus have extremely limited applicability. This is due to the relative scaling of the method, or how the overall time and computer cost of the method changes with respect to the number of electrons in a system. FCI, for example, scales combinatorially with respect to the system size[108] and is rarely used, whereas time dependent density functional theory (TD-DFT) scales as $\mathcal{O}(N^4)$,[109] a significantly more tractable cost as system size increases, and is the most widely used method for excited state chemistry. However, reducing the computational cost of a method often goes hand-in-hand with the introduction of approximations to the physics governing the system and a reduction of the method's accuracy and predictive power.[110] While it is necessary to concede some loss of accuracy in order to produce a more computationally affordable method, there are still highly effective excited state methods that perform excellently within the domain of systems for which they were designed, meaning the cost

versus accuracy dilemma is exploitable. This wide range of excited state methods exist in effectively two categories.[109, 111] With members such as FCI,[108] complete active space self-consistent field (CASSCF),[112] variational Monte Carlo,[113] density matrix renormalization group (DMRG),[114] and equation-of-motion coupled cluster,[15] the first category is highly accurate yet only affordable in relatively small systems (while the definition of small is dependent on computational resources and the implementation of the method, think tens of atoms). In contrast, the second category consists of more affordable methods with more aggressive approximations but can be applied to significantly larger systems (think hundreds of atoms). Within this category are methods such as configuration interaction singles (CIS),[109, 115] time-dependent density functional theory,[109] Δ-SCF,[109] σ-SCF,[116] and excited state mean field theory.[117] Of course, the holy grail of electronic structure theory methods is both incredibly fast and incredibly accurate, and in the quest to find it, it is essential to determine which approximations are acceptable to describe a system, and thus which physics are most important to preserve, when designing computationally affordable methods.

Systematic improvability. In order to balance the physical accuracy and computational affordability of a method, the method should be systematically improvable such that it is very clear how a calculation can be pushed in order to recover more accurate results. Demonstrated on the water molecule for configuration interaction (CI) theory, Møller-Plesset (MP) perturbation theory, and coupled cluster (CC) theory, one can decrease the order of a relative approximation and recover further electron correlation, however often at a diminishing return. For example, CISD, which contains contributions from all singly and doubly excited determinants, recovers 94.5% of the correlation energy missing from RHF. Including triples in the truncated CI wave function recovers only an additional 1.5% of the correlation energy, while including all excitations through the quadruply excited excitation manifold recovers a further 4% of electron correlation energy and is nearly exact. Beyond quadruple excitations, maximally only 8 meV of correlation energy can be recovered.[118] This trend is also seen in higher order applications of MP and CC theories, where the lowest order expansions recover the most significant physics, and higher order expansions systematically converge to the exact energy.[118] This systematic improvability allows us to understand how to balance our approximations versus our accuracy, giving us insight in how we can improve our ansatz in order to achieve better understanding of the system. In contrast, methods based on density functional theory (DFT) have functional dependence, and there is not always a well-defined approach to improving the quality of the results. While there is extensive research into answering the question of "Which functional should I choose?" and there are certainly functionals that are more reliable in specific contexts,[119, 120] this is not necessarily the same as systematic improvability, which is particularly valuable when the correct description of a system is not yet clear or well-understood.

1.4 Standard Ansatzes for Electronically Excited States

In this section, the author reviews the most widely used ansatzes within excited state electronic structure theory. Generally, the most prominent ansatzes fit into three categories – configuration interaction wave functions based on hierarchical excitations of the HF reference wave function, linear response theories, and non-Aufbau filled single determinant wave functions. The overall cost and accuracy of each method is contextualized with studies of systems described in Sec. 1.1 which quite uniformly highlight the same few themes – the importance of orbital relaxation, the need for equitable treatment of valence, Rydberg, and charge transfer excitations, and the computational demands of realistic systems.

All of these methods have individual strengths and weaknesses, and together they highlight potential goals for newly developed ansatzes, including (*i*) state-specific, fully relaxed excited state orbitals, (*ii*) compact representations of correlation in excited state wave functions that still retain the necessary physics of an excited state, such as size consistency and spin purity, and (*iii*) systematic improvability.

Configuration Interaction

The full configuration interaction (FCI) wave function is a linear combination of the reference determinant, most often the Hartree-Fock determinant, and all possible excitations from the reference determinant

$$\begin{aligned} |\Psi_{\text{FCI}}\rangle = c_0|\Phi_0\rangle + \sum_{i}^{N_{\text{occ}}}\sum_{a}^{N_{\text{vir}}} c_i^a \hat{a}_a^\dagger \hat{a}_i |\Phi_0\rangle + \sum_{ij}^{N_{\text{occ}}}\sum_{ab}^{N_{\text{vir}}} c_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i |\Phi_0\rangle \\ + \ldots + \sum_{ij\ldots k}^{N_{\text{occ}}}\sum_{ab\ldots c}^{N_{\text{vir}}} c_{ij\ldots k}^{ab\ldots c} \hat{a}_a^\dagger \hat{a}_b^\dagger \ldots \hat{a}_c^\dagger \hat{a}_k \ldots \hat{a}_j \hat{a}_i |\Phi_0\rangle \end{aligned} \tag{1.19}$$

where the first summation is the singles manifold, i.e. the set of all determinants constructed by exciting one electron in the reference wave function, the second summation is the doubles manifold, and the excitation order increases by one until terminating at the final summation, which is the manifold of determinants in which all electrons in the reference wave function are excited. The FCI wave function is most often solved iteratively to locate only the first several states. The orbitals of the reference wave function are not optimized in FCI, leaving just the linear coefficient expansion to solve via iterative eigensystem-solving methods as diagonalizing the full Hamiltonian is prohibitively expensive except for the smallest of systems.[108]

Though exact within the Born-Oppenheimer approximation and the selected basis set, as previously discussed, the FCI wave function scales combinatorially with the size of the system,[108] resulting in its limited applicability, with the largest converged FCI calculation to date consisting of 22 electrons in 22 orbitals, or approximately 500 *billion* determinants.[121] Therefore, applications of FCI to even the smallest systems discussed in Sec. 1.1 – a single nucleotide or chromophore – are infeasible with modern computing resources.

Truncated Configuration Interaction

One route towards making the FCI wave function more tractable is through hierarchical truncations of excitation manifolds. For example, configuration interaction singles (CIS) includes only the reference determinant and all possible singly excited determinants and scales as $\mathcal{O}(N^4)$,[109] while configuration interaction singles and doubles (CISD) includes the entirety of the CIS wave function as well as the manifold of double excitations and scales as $\mathcal{O}(N^6)$.[108] As in FCI, the orbitals are not optimized in truncated CI, leaving just the linear coefficient expansion to solve iteratively. Despite the significant reduction of terms compared to FCI due to the truncation, these CI wave functions are still not particularly compact, leading to relatively slow convergence although reasonable scaling.[108, 122] With advancements in the use of graphical processing units, CIS calculations on substantial π-conjugated systems and solvated chromophores with hundreds (and nearly one thousand) second-row atoms, including oligothiophene dendrimers, the photoactive yellow protein (PYP), and the PYP chromophore solvated with 900 quantum mechanical water molecules, have been achieved.[123]

These truncated CI wave functions are systematically improvable in that as one increases the order of excitation manifold included in the wave function, more electron correlation is recovered.[122] However, that means that the truncated CI wave functions with the most approximations relative to FCI also have the most degradation in their physical accuracy. CIS, for example, fails to relax excited state orbitals, resulting in excitation energies that are typically 0.5-2 eV too large[109, 115, 124–126] (as such, it is interesting to note that a CIS energy is an upper bound to the exact singly-excited state's energy within the basis set).[122] Moreover, CIS does not have equitable performance between valence and CT excitations, as the latter often involves significant changes in electron density and orbitals, resulting in CT excitation energy errors that are more than an eV larger than errors in valence excitations.[127] Thus, despite the incredible system size that can be treated with CIS, the shortcomings of this method prevent it from accurately treating systems which need balanced descriptions of these excitations, including the DNA nucleobase dimers and the organometallic dyes discussed in Sec. 1.1. Higher-order truncated CI methods will also not be able to reliably model such systems, but for a different reason – these methods lack the scaling properties of size consistency and size intensivity. Accordingly, the descriptions of large molecules by CISD and all further truncated CI wave functions will systematically deteriorate as the number of electrons increases.[108]

Orbital Optimized Configuration Interaction Singles

An interesting method demonstrates how even partially reoptimizing the reference orbitals in CIS can restore balanced descriptions of CT and non-CT excitation energies at only twice the cost of a CIS calculation. Orbital optimized CIS (oo-CIS),[128] corrects the CIS energy and wave function by optimizing the reference orbitals with a single Newton-Raphson optimization step. To do this, the developers first perform a standard CIS calculation,

computing E_{CIS} and $|\Psi_{\text{CIS}}\rangle$. They then consider how unitary rotations of the reference wave function's orbitals would affect the CIS energy. As E_{HF} and thus E_{CIS} are invariant to occupied-occupied and virtual-virtual orbital rotations,[108, 122] expanding E_{CIS} with respect to occupied-virtual orbital rotation parameters $\{\theta\}$ and solving for the optimal orbital rotation parameters that minimize $E_{\text{CIS}}(\theta)$ reveals their dependence on the inverse of the Hessian $\mathbf{H}_{\text{CIS}}$, where

$$(\mathbf{H}_{\text{CIS}})_{ia,jb} = \frac{\partial^2 E_{\text{CIS}}}{\partial\theta_{ia}\partial\theta_{jb}} \tag{1.20}$$

which is not only expensive to compute, store, and invert, it also is not necessarily positive definite and thus not necessarily invertible. Avoiding the complications created by H_{CIS} and keeping the cost-scaling of oo-CIS low, the developers chose to replace the CIS Hessian with the HF Hessian, resulting in the following expressions for the oo-CIS energy.

$$E_{\text{oo-CIS}} = E_{\text{CIS}} - \frac{1}{2}\sum_{ijab} \frac{\partial E_{\text{CIS}}}{\partial\theta_{ia}}\Big|_{\theta=0} \left(\frac{\partial^2 E_{\text{HF}}}{\partial\theta_{ia}\partial\theta_{jb}}\Big|_{\theta=0}\right)^{-1} \frac{\partial E_{\text{CIS}}}{\partial\theta_{jb}}\Big|_{\theta=0} \tag{1.21}$$

Although the one-shot Newton-Raphson optimization step is incomplete and the orbital relaxation is partially dependent on ideal rotations for the ground state, this straightforward one-step orbital optimization led to meaningful corrections to CIS excitation states, with oo-CIS excitation energies rivaling those of CIS with a perturbative doubles correction (CIS(D)).[129] Moreover, the accuracy of oo-CIS in charge transfer excitations was equitable to its accuracy in valence excitations.[128] These results were particularly intriguing as the oo-CIS method's cost scaling allows for the method's treatment of systems with several hundreds of second-row atoms. Additionally, with its balanced description of valence, Rydberg, and CT excitations, applications to real systems like large chromophoric dyes and solvated DNA fragments are within reach, although perhaps the overall accuracy still leaves a bit to be desired. This demonstration that meaningful orbital relaxations can be achieved at $\mathcal{O}(N^4)$ cost inspired the excited state mean field (ESMF) wave function (to be discussed extensively in Chapter 3) which fully relaxes the excited state orbitals of a CIS-like wave function.

Complete Active Space Self Consistent Field

The final configuration interaction ansatz to be discussed in this introduction is the complete active space (CAS) wave function, which takes a different approach to FCI truncation.[112, 130, 131] Rather than selecting only certain orders of excitation to include in the wave function, the CAS wave function partitions the orbitals in the reference state into three categories: core orbitals, active orbitals, and virtual orbitals. As depicted in Fig. 1.7, the core orbitals are low energy, doubly occupied orbitals that are not expected to be significantly involved in a system's excited states. In contrast, the active orbitals are the mid-energy orbitals that are most likely to participate in the system's excitations and will include the highest energy occupied molecular orbitals and the lowest energy unoccupied orbitals from the reference

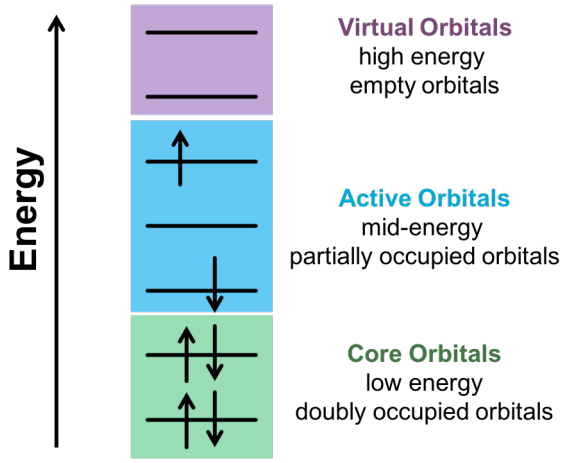


Figure 1.7: CASSCF wave function

space. The CAS wave function is a linear combination of the reference determinant and all possible excitations within the active orbitals. Last and certainly least, the virtual orbitals are high energy orbitals that were unoccupied in the reference wave function and remain as such within the CAS wave function.

Unlike in FCI and the hierarchically truncated CI methods, the CAS self consistent field optimization (CASSCF) optimizes the excitation configuration weights *in addition to* the molecular orbitals within the reference wave function to obtain the total electronic wave function with the lowest possible energy. The CAS wave function is given by

$$|\Psi_{\text{CAS}}\rangle = e^{\hat{X}} \sum_{\mu} c_{\mu} |\Psi_{\mu}\rangle \tag{1.22}$$

where $\hat{X}$ is an anti-Hermitian operator and $e^{\hat{X}}$ thus performs unitary rotations on the molecular orbitals, and where the sum runs over the manifold of all possible determinants within the active space.[112, 130]

Due to the inclusion of orbital optimization, the resulting CAS wave function is considerably more flexible within the active space than truncated CI methods and, particularly for strongly correlated systems with degenerate or nearly-degenerate electronic configurations, is considered a qualitatively correct reference wave function upon which higher order multi-reference methods can be applied to recover the dynamical correlation lacking in the CAS wave function.[130] One particular limitation of the orbital rotations, however, is that if multiple CAS wave functions are optimized simultaneously, the orbitals are most often optimized such that a weighted average of the sum of the energy of all states is minimized, meaning the orbitals are not ideal for any specific state and can create imbalances if the changes leading to the optimal orbitals of one state contradict the changes needed for others. While ameliorated by the recently developed state-specific CASSCF optimization method,[132, 133] state-averaging is still the most common approach to CASSCF optimizations.

As it still includes the entire excitation manifold within the active space, CAS scales combinatorially with system size, but the wave function is significantly more compact than the FCI wave function.[112, 130] Even in the example given in Fig. 1.7, the FCI wave function consists of 1,225 determinants while the CAS wave function has only 9 determinants. However, defining the active space actually poses a problem for CASSCF methods as the end result will vary with respect to active space choice.[134] This is because in the limit where the active space includes all orbitals and electrons, CASSCF becomes equivalent to FCI, and in the limit where the active space includes no electrons, CASSCF is equivalent to the reference method (normally HF). It is thus important to choose the "best" active space that will include the most important orbitals involved in the excited states and still be computationally affordable, a choice that is not always apparent.[130]

CASSCF is routinely utilized as the quantum mechanical component of many molecular dynamics studies of nucleobases. It has been used in QM/MM[135] simulations to investigate the de-excitation pathways of individual,[136, 137] stacked,[51] and hydrogen-bonded[49, 138] nucleobases as well as to provide insight into the excited reaction mechanism of thymine dimerization.[139] Also employed in mixed quantum-classical surface-hopping methods,[140] CASSCF is able to bring insight into the relaxation dynamics[141, 142] and intersystem crossing mechanisms[143] of single nucleobases. Additionally, CASSCF and its higher order correlation methods CASPT2[144–147] and MRCI[148–150] were used in conjunction with nonadiabatic trajectory dynamics to determine that there is no single relaxation mechanism in any of the five DNA and RNA nucleobases, but rather a "pool of reaction pathways."[151] Consider, however, that the largest fully-optimized CASSCF calculations to date have a 20 electron, 20 orbital active space and exploit the highest abelian point group symmetry (D_{2h}) for symmetry simplifications, leading to wave functions containing more than 4 billion determinants.[121] While an incredible feat of software engineering and massive parallelization, this active space limit is still often inadequate to treat systems with transition metals, which require at least the valence s, p, and d orbitals (9 total orbitals) to be in the active space to properly describe static correlation,[152] if there are important contributing ligands or other metals. While orbital localization procedures can reduce the number of orbitals needed to capture the static correlation in such systems,[153, 154] the restrictive computational scaling of CASSCF and the involved nature of the CASPT2/CASSCF approach[155] demand that the theorist make nontrivial choices about the number of states and their relative weights in state-averaged CASSCF calculations and what areas of Hilbert space are most important to include in the active space,[36] as these may lead to different results. For example, a CASSCF and its higher order correlation method NEVPT2[156–158] study of the electronic spectra of the N3 organometallic dye (depicted in Fig. 1.5) found that with an active space of 14 electrons and 12 orbitals, NEVPT2 calculations on wave functions originating from 15-state state-averaged CASSCF calculations were significantly different from NEVPT2 calculations on wave functions originating from 20-state state-averaged CASSCF calculations, determining that including more roots with different electronic distributions in the state-averaged CASSCF optimization deteriorated the quality of the zeroth-order wavefunction and, as a result, the reliability of the NEVPT2 description.[47] Additionally, in a study of

cytosine's singlet excited state decay paths, varying the size of the active space led to significant differences in the character of the lowest singlet $n \rightarrow \pi^*$ excitation, with a 12 electron, 12 orbital active space exciting an electron from the oxygen lone pair, while a 12 electron, 11 orbital active space excited an electron from the nitrogen lone pair.[36] Moreover, choices regarding active space and state-averaging could potentially explain theoretical results that lead to conclusions counter to what experiment has demonstrated, such as in the disagreement between theory and experiment regarding which singlet excited state of cytosine and uracil undergoes intersystem crossing to generate the triplet $\pi \rightarrow \pi^*$ state.[23, 39–41] In short, while CASSCF has been successful in treating systems such as those described in Sec. 1.1, the choice of the active space and the consequences of state-averaging can significantly alter final results and conclusions.

Linear Response

In parallel with laboratory experiments where a molecular system's response to external fields like electromagnetic fields is probed to better understand properties of the system, linear response theory utilizes small perturbative potentials to nudge a molecule out of ground state equilibrium, allowing the wave function to evolve into a superposition of many electronic states such that properties of excited states can be measured in the virtual environment.

A general prescription for linear response theory that is applicable in both wave function-based and density-based methods begins with identifying a time-dependent observable $\langle A \rangle$ and expressing its evolution through time in response to a frequency-dependent external perturbation $V(\omega)$. A critical assumption in response theory is that the external potential is significantly weaker than the molecule's internal potential, allowing the applied potential to be treated perturbatively. Comparing the hydrogen atom, which has a ground state energy of ≈ 0.5 Ha and where the most probable distance between the electron and the proton is given by the Bohr radius, $a_0 \approx 0.5$ Å, and thus an internal potential on the order of 10^{11} eV/m, to a typical experimental electric field set up of 10^3 eV/m,[159] one sees that a perturbative treatment of the external potential is reasonable. Therefore, the first order expansion of the time evolution of $\langle A \rangle$ is given by

$$\langle A \rangle = \langle A \rangle^{(0)} + \langle A \rangle^{(1)} \tag{1.23}$$

$$\langle A \rangle^{(0)} = \langle \Psi_{t=0} | A | \Psi_{t=0} \rangle \tag{1.24}$$

$$\langle A \rangle^{(1)} = \int_{-\infty}^{\infty} \langle\langle A; V(\omega) \rangle\rangle_{\omega} e^{-i\omega t} d\omega \tag{1.25}$$

$$\langle\langle A; V(\omega) \rangle\rangle_{\omega} = \frac{1}{\hbar} \sum_{k \neq 0} \left(\frac{\langle \Psi_{t=0} | A | k \rangle \langle k | V(\omega) | \Psi_{t=0} \rangle}{\omega - \omega_{k0}} - \frac{\langle \Psi_{t=0} | V(\omega) | k \rangle \langle k | A | \Psi_{t=0} \rangle}{\omega + \omega_{k0}} \right) \tag{1.26}$$

where in the final equation, which is the linear response function and is typically referred to as the sum-over-states expression or the spectral resolution, ω_{k0} is related to the excitation energy from the ground state to the k-th excited state via $\hbar\omega_{k0} = E_k - E_0$, where $H_{t=0}|k\rangle =$

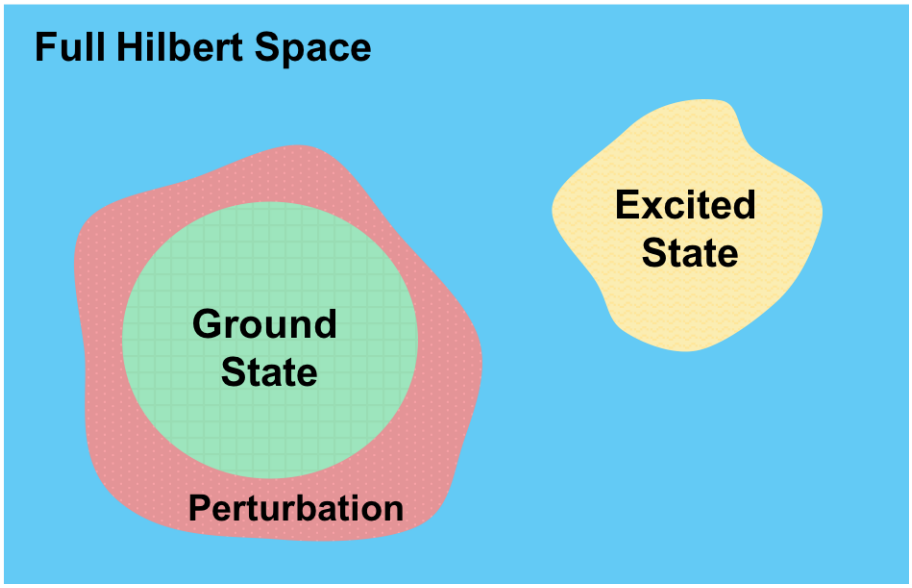


Figure 1.8: Breakdown of linear response theory

$E_k|k\rangle$. Since the idea of linear response theory was to not directly compute the eigenvectors and eigenvalues of the time-independent Hamiltonian, we are missing information for how to compute Eq. 1.26. While one could scan over frequencies ω to locate the linear response function's poles, i.e. where $\omega = \pm\omega_{k0}$, an alternative route exists that utilizes the linear response *equation*, which has the general form of

$$(H - \omega S)\,\lambda^{(1)}(\omega) = -\mathcal{V}^{(1)}(\omega) \tag{1.27}$$

where $\lambda^{(1)}(\omega)$ is a vector containing the first-order time-dependent parameters of the wave function (or density), $\mathcal{V}^{(1)}(\omega)$ is a vector dependent on $\lambda^{(1)}(\omega)$ and $V(\omega)$, ω is the perturbation frequency, and H and S are the ground state Hamiltonian and overlap matrices, respectively.[160]

Naively, one could attempt to directly solve Eq. 1.27 for $\lambda^{(1)}(\omega)$ by inverting $(H - \omega S)$, a task that quickly grows computationally infeasible as the number of elements in each ground state matrix is equal to the square of the number of parameters in the wave function (or density) and algebraically impossible at values of ω for which $(H - \omega S)$ is singular. Alternatively, one could use an iterative procedure to solve Eq. 1.27 without matrix inversions, retrieving the excited states parameters that way and then using that information to solve for observables as in Eqs. 1.23-1.26.[160]

When linear response is combined with an approximate ansatz whose tangent space (into which linear response reaches) is limited by the ansatz flexibility, there is often an implicit assumption of similarity between the excited states and the ground state that is related to the limitations of this tangent space. However, this is not necessarily a valid assumption.

For example, excitations of interacting systems generally involve mixtures of multiply excited determinants, and states with significant multiply excited character could exist outside the space of singly excited determinants that comprise the ground state's tangent space (see Fig. 1.8), and linear response theory would break down.

Equation-of-Motion Coupled Cluster Singles and Doubles

A widely adopted method for computing electronically excited states, equation-of-motion coupled cluster singles and doubles (EOM-CCSD) is an $\mathcal{O}(N^6)$-scaling, well-benchmarked, linear response method in which one can simultaneously compute many excited states.[15, 161–165] Despite the existence of linear response coupled cluster (LR-CC) methods which are derived in accordance with traditional response theory, the author chooses to focus on the EOM-CC formulation as the energies of LR-CC methods are identical to the energies of the corresponding EOM-CC method,[166] and EOM-CC methods are more widely used than LR-CC methods.

The ground state CCSD wave function is defined as the following

$$|\Psi_{\text{CCSD}}\rangle = e^{\hat{T}_1+\hat{T}_2}|\Psi_0\rangle \tag{1.28}$$

$$\hat{T}_1 = \sum_{i}^{N_{\text{occ}}}\sum_{a}^{N_{\text{vir}}} t_i^a \hat{a}_a^\dagger \hat{a}_i \tag{1.29}$$

$$\hat{T}_2 = \frac{1}{4}\sum_{ij}^{N_{\text{occ}}}\sum_{ab}^{N_{\text{vir}}} t_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i \tag{1.30}$$

where the doubles cluster operator $\hat{T}_2$ describes important electron pair interactions and the singles cluster operator $\hat{T}_1$ relaxes the orbitals in response to the field generated by said interactions.

The EOM-CCSD wave function is given by the follow parameterization.

$$|\Psi_{\text{EOM-CCSD}}\rangle = \hat{\mathcal{R}}\ \left(e^{\hat{T}_1+\hat{T}_2}|\Psi_0\rangle\right) = e^{\hat{T}_1+\hat{T}_2}\ \left(\hat{\mathcal{R}}|\Psi_0\rangle\right) \tag{1.31}$$

$$\hat{\mathcal{R}} = 1 + \sum_{i}^{N_{\text{occ}}}\sum_{a}^{N_{\text{vir}}} r_i^a \hat{a}_a^\dagger \hat{a}_i + \frac{1}{4}\sum_{ij}^{N_{\text{occ}}}\sum_{ab}^{N_{\text{vir}}} r_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i \tag{1.32}$$

In this formulation, $\mathcal{R}$ is a linear excitation operator that commutes with $e^{\hat{T}}$, and thus can be considered to either act on the CCSD ground state like the first definition in Eq. 1.31 or to excite the reference state like the second definition in Eq. 1.31. From the second interpretation of the EOM-CCSD wave function, two potential limitations of the EOM-CCSD wave function are elucidated. First, since the cluster operator $\exp[\hat{T}]$ was optimized for the ground state, it provides a sub-optimal description of electron correlation for the excited states generated by $\hat{\mathcal{R}}|\Psi_0\rangle$.[167] Second, while changing the reference orbitals within

$|\Psi_0\rangle$ may not significantly alter the ground state CCSD energy and wave function, it may significantly alter EOM-CCSD results. As the linear combination of ground and excited wave functions generated by $\hat{\mathcal{R}}|\Psi_0\rangle$ is dependent on the many-electron basis functions within $|\Psi_0\rangle$, the ability of $\hat{\mathcal{R}}|\Psi_0\rangle$ to represent the excited state in question is also dependent on $|\Psi_0\rangle$. Therefore, the choice of the reference wave function is in and of itself a parameter of the EOM-CCSD wave function.[168, 169] This reference dependence was highlighted in a study individual nucleobase monomers where comparing EOM-CCSD's performance when using a diffuse versus a non-diffuse basis set led to distinct, nonsystematic errors in the treatment of valence and Rydberg states.[53]

Furthermore, EOM-CCSD is unable to relax orbitals following a double excitation. Thus, it overestimates the interaction energy between those doubly occupied orbitals, suppressing and in some cases drastically underestimating the doubly excited character in a wave function. This error is mitigated by including the effects of triple excitations.[161, 165] In a series of benchmarks that investigated individual nucleobases both solvated and in vacuo, hydrogen bonded guanine-thymine and guanine-cytosine pairs, and stacked adenine-thymine and guanine-cytosine pairs, comparisons between EOM-CCSD and higher order (and higher scaling) EOM-CC methods that include effects of triple excitations demonstrated that the perturbatively-corrected, $\mathcal{O}(N^7)$-scaling EOM-CCSD(T) method lowered EOM-CCSD excitation energies by 0.1-0.3 eV to match experiment within 0.1 eV,[53, 55, 161–165, 170] with the largest adjustments in CT states.[53] In some cases, including the effects of triple excitations changed the ordering of CT excitations.[53] It seems, therefore, that for quantitatively accurate descriptions of nucleotide systems, triples contribution is unavoidable.[53, 161, 165] While quite accurate, these triple corrections are rather expensive[53, 165] with their range of applicability limited to about 20 second-row atoms.[53] While there have been studies applying EOM-CCSD to relatively large DNA fragments, such as a DNA fragment consisting of 5 adenine-thymine pairs and no sugar backbone (95 second row atoms),[52] there are some studies[171, 172] yet few thorough benchmarks of EOM-CCSD applied to the types of dyes relevant to DSCs due to the need for many virtual orbitals to properly describe valence and Rydberg states and the relative cost of the method. Based on the previously discussed benchmarks, one would expect similar results in these systems – a systematic error of a few tenths of an eV higher than experiment, with a slight bias that artificially raises CT excitation energies due to the lack of second-order orbital relaxation.

Time-Dependent Density Functional Theory

Doubtlessly the most widely used method for studying singly excited states, time-dependent density functional theory, or TD-DFT, is a size consistent, spin pure, $\mathcal{O}(N^4)$ scaling method that calculates excitation energies by analyzing the linear time-dependent response of the time-independent ground state electron density to a time-dependent external electric field. For an excellent derivation of TD-DFT formalism, the author refers readers to Ref. [109], a very thorough and comprehensive review of CIS, TD-HF, and TD-DFT by Dreuw and

Head-Gordon. The critical takeaway is the canonical TD-DFT eigenvalue equation

$$\begin{bmatrix} A & B \\ B^* & A^* \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \omega \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \tag{1.33}$$

where X and Y are excitation and de-excitation amplitudes, respectively, ω is the excitation energy, and elements of A and B are given by

$$A_{ia,jb} = \delta_{ij}\delta_{ab}\left(\epsilon_a - \epsilon_i\right) + (ia|jb) - c_{\mathrm{HF}}(ij|ab) + (1 - c_{\mathrm{HF}})\left(ia|f_{\mathrm{xc}}|jb\right) \tag{1.34}$$

$$B_{ia,jb} = (ia|bj) - c_{\mathrm{HF}}(ib|aj) + (1 - c_{\mathrm{HF}})\left(ia|f_{\mathrm{xc}}|bj\right) \tag{1.35}$$

where c_{HF} is the amount of HF exchange in the chosen exchange potential, and the terms involving f_{xc} are the response of said exchange potential.[109]

One common approximation to TD-DFT (and TD-HF) is the Tamm-Dancoff Approximation (TDA), in which the matrix elements of B are zeroed out, and Eq. 1.33 becomes simply $AX = \omega X$. Interestingly, TDA-TD-HF is equivalent to CIS, and TDA-TD-DFT and TDA-TD-HF differ only in choice of exchange potential. Further connections between TD-DFT, TD-HF, and CIS are apparent upon analysis of Eqs. 1.34 and 1.35. Despite being derived specifically for a hybrid density functional, i.e. one that contains a non-zero percentage of HF exchange, the limiting case of $c_{\mathrm{HF}} = 0$ returns the expressions of A and B when TD-DFT is used with a pure density functional, and the limiting case of $c_{\mathrm{HF}} = 1$ returns the TD-HF expressions for A and B.[109]

While theoretically exact, TD-DFT parallels ground state Kohn-Sham DFT in that the exact time-dependent *exchange-correlation action functional* (also referred to as the exchange kernel or the density functional) is unknown and must be approximated. Rather than developing density functionals specifically for excited states, it is most common to assume that the adiabatic local density approximation (ALDA), which states that the electron density varies slowly in time, is valid. This allows the exact, non-local in time, time-dependent density functional to be replaced with a local, time-independent density functional – including those developed for ground state DFT. One effect caused by adopting the ALDA is TD-DFT's notable sensitivity to choice of density functional. This is because problems with ground state exchange-correlation functionals, such as issues describing hydrogen bonding and dispersion interactions[173] and bond breaking and bond formation,[174] carry over into the excited state method. Therefore, studies that involve these complicated covalent interactions and bond changes, such as that with TD-DFT/M052X[175] which aimed to elucidate photochemical and photophysical decay routes of the thymine deoxy-dinucleotide[176], are further complicated as effects of the exchange-correlation functional in both ground and excited states must be thoroughly understood in order to trust the conclusions. A specific example highlighting disagreements between density functionals occurs in studies regarding the ultrafast de-excitation of singlet adenine, in which a new deactivation pathway was identified at the TD-DFT/B3LYP level of theory,[37] but TD-DFT/ωB97X-D[177] in conjunction with surface-hopping dynamics completely failed to predict the ultrafast de-excitation.[46] The

sensitivity of TD-DFT to density functionals is such that experts suggest always comparing its results to wave function-based benchmark calculations when available and performing multiple calculations with differing density functionals to check for functional sensitivity.[109]

Despite this clear drawback of TD-DFT, the method is incredibly accurate for valence excited states, with errors typically in the range of 0.1-0.5 eV when using a sufficiently large basis set. This accuracy is due to the leading term of Eq. 1.34, $(\epsilon_a - \epsilon_i)$, which is typically a good approximation to a valence excitation energy when using Kohn-Sham (KS) orbitals. However, the accuracy of TD-DFT breaks down severely, with errors of several eV, in a number of scenarios.[109] Let us first discuss Rydberg states. Since standard density functionals decay faster than $\frac{1}{r}$, where r is the electron-nucleus distance, the total energy of the system as well as the core and valence KS orbital energies are all upshifted quite significantly relative to exact KS theory. While the very high-energy, diffuse, Rydberg orbitals are also upshifted relative to their exact KS counterparts, the shift is significantly smaller than that experienced by the core and valence orbitals. Thus, orbital energy differences between orbitals within or below the valence range are still good approximations to excitation energies, while valence-Rydberg orbital energy differences are too small. Since orbital energy differences dominate in Eq. 1.34, especially when orbital overlaps are negligible, the resulting Rydberg excitation energies are often severely underestimated.[178–181] Additionally, consider how Rydberg state orbitals must change relative to ground state orbitals. The electron density around the nucleus changes drastically in Rydberg excitations compared to how it changes in valence excitations, meaning the orbitals in Rydberg states are more different from the ground state orbitals than valence-excited state orbitals are from the ground state orbitals. Therefore, it can be concluded that the assumptions of linear response theory are more likely to fail for Rydberg states. The second scenario in which TD-DFT fails is in excitations with significant doubly excited character. While the exact linear response formalism of TD-DFT allows for multiplet excitations, ALDA restricts TD-DFT to single excitations exclusively, and doubly excited character cannot be recovered by TD-DFT unless nonadiabatic corrections are included.[109, 182] Another notable motif of excitations in which TD-DFT fails is in charge transfer excitations.[183] Consider a CT excitation between two hydrogen atoms, well-separated at distance R. Assuming a pure density functional and that the hydrogen atoms' orbitals do not overlap, Eq. 1.34 reduces to $A_{ia\sigma,jb\tau} = \delta_{\sigma\tau}\delta_{ij}\delta_{ab}\left(\epsilon_a - \epsilon_i\right)$ and Eq. 1.35 reduces to 0. Notably, any non-zero terms in A completely lack dependence on the electrons' positions in the excited state. Therefore, the asymptotic behavior of the energy with respect to R is not $\frac{1}{R}$ as expected for the electrostatically attracted hydrogen cation and hydrogen anion, but is instead a constant and evidence of self-interaction error. While range-separated density functionals, which replace the DFT exchange potential with HF exchange when electrons are well-separated,[184–192] have partially ameliorated the self-interaction error in CT excitations, there is still dependence on the parameterization of the range-separation, and the methods are not systematically improvable.[44, 73, 193–197] Moreover, range-separated hybrid functionals do not even fully alleviate the problems TD-DFT experiences when treating CT excitations. Due to the ALDA, while the particle and hole orbitals are relaxed by TD-DFT, the other orbitals are not. This is because the ALDA, as previously discussed,

cannot capture double excitations. Therefore, as the leading order term of the Taylor expansion of a fully orbital-relaxed singly excited state is a linear combination of doubly excited determinants, when using a local-in-time exchange-correlation functional in TD-DFT, the substantial changes in orbital shape with respect to the ground state that are necessary for accurate descriptions of CT are absent.[198–200]

Several benchmark comparisons of TD-DFT within nucleobase systems, which as mentioned in Sec. 1.1 are characterized by valence, Rydberg, and CT excitations, found that TD-DFT errors tended to be nonsystematic and significantly larger than those of CASPT2 and EOM-CCSD, often resulting in different orderings of excited states.[53, 55, 170, 201, 202] For example, within both stacked guanine-cytosine and adenine-thymine nucleobase pairs, TD-DFT/M06-HF[203] and TD-DFT/CAM-B3LYP[204] overestimate most transition energies, while others are nonsystematically underestimated, resulting in state orderings that contradict results from high-level EOM-CC methods accepted as benchmarks.[53, 205] When other functionals, such as TD-DFT/LC-PBE0,[44] give quite good energetic results for CT excitations of nucleobase monomers, all other excitation energies are too high,[170, 201, 206, 207], with TD-DFT/B3LYP[208, 209] "not... good enough to predict even the observable UV spectrum of the [individual] nucleobases".[170] Similar results were found in dye-sensitized solar cell benchmarks.[89, 171, 210–212] The performance of TD-DFT/MPW1K[213] in five different organic dyes deteriorated as CT excitation character increased[210] and "erratic lowering of the excitation energy" was seen in ruthenium organometallic dyes, which are characterized by strong charge transfer excited states.[89] Furthermore, no functionals were able to equitably treat all types of excitations in the trans-(Cl)-Ru(bpy)$Cl_2(CO)_2$ dye,[212] and within a survey of functionals used with TD-DFT to study the N3 dye on a $(TiO_2)_{82}$ slab ($\sim$5000 electrons!), no functionals were able to simultaneously reproduce the optical absorption spectrum and correctly predict the favorability of the electron injection pathway.[89]

Due to the failures of TD-DFT in these scenarios, it is not as broadly applicable as other methods – a serious frustration for theoretical and computational chemists as TD-DFT has particularly amicable cost scaling and is able to treat systems of truly impressive size. Through these benchmarks of systems characterized by charge transfer excitations, it is evident that fully relaxed excited state Kohn-Sham orbitals are necessary to increase the dependability of TD-DFT methods.

Non-Aufbau Filled Single Determinant Wave Functions

The general idea governing non-Aufbau filled single Slater determinant wave functions is to take advantage of the compact, easily interpretable, computationally inexpensive ansatz and existing code architecture for HF theory or DFT. To model excited states, one takes the ground state single determinant wave function then singly-occupies the valence orbitals that participate in the excitation and optimizes the orbitals in this non-Aufbau filled determinant.[93, 198, 214]

Δ-SCF

Δ-SCF is such a single-determinant excited state method that scales as $\mathcal{O}(N^{3-4})$ depending on the inclusion of HF exchange,[185, 215] and is a similarly reliable predictor of excitation energies in organic chromophores as TD-DFT.[198, 216] Unlike TD-DFT, Δ-SCF solutions are time-independent, static excited states. The procedure for HF based Δ-SCF (Δ-SCF-HF) is identical to that for DFT (Δ-SCF-DFT) as the HF and DFT SCF equations are the same except in their expression of the Fock matrix F which includes effects from the density functional in the latter. A general Δ-SCF excited state method begins, like most, with a ground state HF or DFT SCF optimization. The occupied and virtual orbitals involved in the desired excitation are then swapped, the resulting non-Aufbau filled determinant then goes through the HF/DFT SCF cycle, the molecular orbitals are optimized, and the energy is minimized. Since the Δ-SCF wave function consists of only a single determinant, the molecular orbitals are not spin-restricted, nor is the wave function spin pure. If properties of the spin pure excitations such as, for example, the first open-shell singlet excitation are of interest, two Δ-SCF states must be computed – the "mixed-spin" state and one triplet state. Since the singlet wave function

$$|\Psi_S\rangle = \frac{1}{\sqrt{2}}\left(\hat{a}_a^\dagger \hat{a}_i + \hat{a}_{\bar{a}}^\dagger \hat{a}_{\bar{i}}\right)|\phi_i\phi_{\bar{i}}\ldots\rangle = \frac{1}{\sqrt{2}}\left(|\phi_a\phi_{\bar{i}}\ldots\rangle - |\phi_{\bar{a}}\phi_i\ldots\rangle\right) \tag{1.36}$$

is a linear combination of one of the single-determinant "mixed-spin" states

$$|\Psi_{M,\uparrow\downarrow}\rangle = \hat{a}_a^\dagger \hat{a}_i |\phi_i\phi_{\bar{i}}\ldots\rangle = \quad |\phi_a\phi_{\bar{i}}\ldots\rangle \tag{1.37}$$

$$|\Psi_{M,\downarrow\uparrow}\rangle = \hat{a}_{\bar{a}}^\dagger \hat{a}_{\bar{i}} |\phi_i\phi_{\bar{i}}\ldots\rangle = -|\phi_{\bar{a}}\phi_i\ldots\rangle \tag{1.38}$$

and the two-determinant $m_s = 0$ triplet wave function

$$|\Psi_T\rangle = \frac{1}{\sqrt{2}}\left(\hat{a}_a^\dagger \hat{a}_i - \hat{a}_{\bar{a}}^\dagger \hat{a}_{\bar{i}}\right)|\phi_i\phi_{\bar{i}}\ldots\rangle = \frac{1}{\sqrt{2}}\left(|\phi_a\phi_{\bar{i}}\ldots\rangle + |\phi_{\bar{a}}\phi_i\ldots\rangle\right) \tag{1.39}$$

which is degenerate with both of the single-determinant $m_s = \pm 1$ triplet states,

$$|\Psi_{T,\uparrow\uparrow}\rangle = \hat{a}_a^\dagger \hat{a}_{\bar{i}} |\phi_i\phi_{\bar{i}}\ldots\rangle = -|\phi_a\phi_i\ldots\rangle \tag{1.40}$$

$$|\Psi_{T,\downarrow\downarrow}\rangle = \hat{a}_{\bar{a}}^\dagger \hat{a}_i |\phi_i\phi_{\bar{i}}\ldots\rangle = |\phi_{\bar{a}}\phi_{\bar{i}}\ldots\rangle \tag{1.41}$$

the energy of the Δ-SCF singlet excited state can be expressed as

$$E_S^{\Delta-\mathrm{SCF}} \approx 2E_M[\{\phi_i^M\}] - E_T[\{\phi_i^T\}]. \tag{1.42}$$

This spin purification process is only approximate within Kohn-Sham DFT because the triplet determinant that generated E_T was assumed to be degenerate to the $m_s = 0$ triplet state so was not obtained exactly via spin raising or lowering operators.[216–218] Furthermore, in Δ-SCF-DFT, as the density is optimized, the spin purity of the ansatz is lost. As

shown in a Δ-SCF-DFT study of 16 large organic chromophores, the value of $\langle \hat{S}^2 \rangle$ for the mixed state and for the triplet states deviated from spin purity with an RMSD of 0.032 and 0.038 a.u., respectively.[217] Since the mixed and triplet states are not spin pure, the spin purification procedure in Eq. 1.42 is approximate and will not return the excitation energy of a spin pure singlet.

In large chromophoric dyes such as those that could be used in organic photovoltaics, while Δ-SCF-DFT and TD-DFT have *statistically* similar accuracy for the singlet states, the methods disagreed significantly in some excitations, with discrepancies in excitation energies of up to 0.6 eV.[217] These fluctuations can be traced back to the limited single-determinant Δ-SCF-DFT ansatz. Relative to TD-DFT, Δ-SCF-DFT is less flexible in its descriptions of excitations as the ansatz can only be generated from single orbital replacements.[219] In excitations strongly dominated by only one excitation however, Δ-SCF-DFT rivals the accuracy of TD-DFT, with Δ-SCF-DFT modelling triplet excitations in small organic chromophores with an RMSD about 0.05 eV closer to the benchmark results than TD-DFT.[198] This accuracy is explained by Δ-SCF-DFT's ability to fully relax the orbitals for the excited state of interest.[198] For this reason, Δ-SCF-DFT also successfully models charge transfer excitations and, unlike TD-DFT, does properly capture the $\frac{1}{R}$ dependence of CT excitation energies.[220]

With its ability to fully relax excited state orbitals for systems with hundreds of second row atoms, Δ-SCF-DFT is a competitive excited state ansatz. However, its dependence on density functionals, limited flexibility in the face of systems with multiple contributing single excitations, and frustrating optimization problems (to be discussed further in Sec. 1.5) leave a lot to be desired.

Restricted Open-Shell Kohn-Sham Theory

Designed for multiplet states that cannot be represented by a single Slater determinant, Restricted Open-Shell Kohn-Sham Theory (ROKS) provides a formalism for modelling such excited spin eigenstates within the DFT formalism without explicitly introducing symmetry dependence into density functionals or introducing time-dependence.[221, 222] The multiplet wave function is a linear combination of Slater determinants that share a spin-restricted basis set and that have doubly occupied core orbitals and fixed-occupation valence orbitals. Since minimizing the multiplet energy with respect to the density is equivalent to identifying the fixed-occupancy molecular orbitals that minimize the energy, the ROKS target function for a singlet state

$$E_S^{\mathrm{ROKS}} = 2E_M[\{\phi_i\}] - E_T[\{\phi_i\}] \tag{1.43}$$

has a very similar form as the Δ-SCF spin purification formula in Eq. 1.42, with the main difference being that since the single-determinant "mixed-spin" and triplet states are optimized simultaneously during the ROKS SCF procedure, they share an orbital basis. The E_S^{ROKS}-minimization procedure is also formally an approximation because a Slater determinant is not constructed in ROKS theory – that is, since the KS orbitals are associated with a non-interacting reference system rather than the true wave function, the evaluation of the

total spin angular momentum operator (a two-body operator) is not strictly possible in this one-body picture without further approximations.[223, 224]

One pervasive problem in ROKS theory occurs when seeking the lowest singlet excited state of the same symmetry as the ground state – orbital mixing between the two open shell orbitals can result in a singlet excitation that strongly overlaps with the ground state reference configuration.[216, 222, 223] The result of such mixing is a significantly red-shifted excitation energy – which has been seen in studies of highly-conjugated polyenes[225], organic chromophores, and individual nucleobases – and can result in different orderings of the excited states relative to accepted benchmarks.[216] In these studies, ROKS achieves multiplet states with similar accuracy as those in Δ-SCF-DFT without experiencing variational collapse. However, another dilemma innate to density functional theories is once again highlighted – the quality of ROKS and Δ-SCF-DFT singlet excitation energies are still dependent on the chosen density functional.[223] Finally, Van Voorhis and co. highlighted a strong dependence of ROKS on the basis set, finding that more diffuse basis sets substantially lowered excitation energies for certain transitions, a behavior that was not corrected by using different classes of density functionals.[216]

With the ability to treat significant systems, including zinc phthalocyanine (40 second row atoms, one metal atom)[216] due to its competitive $\mathcal{O}(N^{3-4})$ (depending on the presence of HF exchange in the density functional) scaling[185, 215] and friendlier optimization than Δ-SCF-DFT, ROKS is a competitive excited state ansatz, though its dependence on non-systematically improvable density functionals and its artificial suppression of first singlet excited state energies do leave room for improvement on the methodology.

1.5 Optimization Methods

A wave function or density ansatz is rendered impotent if it cannot be converged to an approximate Hamiltonian eigenstate. Therefore, there are significant efforts being made to design robust optimization methods that can overcome the complications associated with finding excited states. Excited states are saddle points on the energy surface in the Hilbert space,[92–94] and are inherently more complicated to locate with numerical optimization techniques[92] and more susceptible to variational collapse. This problem is not assisted by the fact that the Hilbert space hypersurface is extremely difficult to fully map out.[226] Despite the inherent difficulty of excited state optimization, an assortment of optimization methods have been developed for excited states. From orbital tracking, to modifications of the Hilbert space energy surface, to state-targeting variational principles, the approaches to excited state optimization run the gamut, and each optimization strategy has its own advantages and drawbacks. For example, while a method need not be so easy to use that it is considered a "blackbox", large benchmarks or potential energy surface mappings would benefit from procedures that are relatively controllable and predictable – not to mention computationally affordable. Furthermore, strong initial guess dependence could hinder investigations into novel systems for which specific details about the electronic configuration

of an excited state are unknown. Finally, locating a specific excited state amid a dense manifold of potentially degenerate states remains a challenge for many of these optimization techniques. Within this section, the author discusses several optimization methods, many of which were taken into consideration in the development of the generalized variational principle, which is discussed in Chapter 4.

Self Consistent Field Methods

As higher energy solutions to the SCF equations are increasingly being interpreted as mean field approximations to excited states,[220, 226–231] optimization methods that utilize the computationally efficient HF and DFT SCF cycles and seek such solutions are prevalent in excited state optimizations. Two such SCF optimization methods are discussed here.

Δ-SCF and the Maximum Overlap Method

While the previously described Δ-SCF ansatz showed promisingly accurate excited state energetics, the SCF optimization procedure is much easier described than computed. Unlike in the ground state where orbitals are filled according to the Aufbau principle,[2, 98] the protocol for selecting which orbitals are occupied in the excited state is less clear, particularly when orbitals swap in energetic ordering or significantly change shape during the orbital relaxation.[93, 198, 216, 219, 232–234] This uncertainty can force the user to make choices about orbital occupations that may lead the optimization down different paths that end at disparate states, and it is often unclear ahead of time how such decisions will change the final outcome of the calculation. Additionally, one may experience an optimization phenomenon known as *occupancy sloshing*, where the occupancy of orbitals flip in a continuous cycle without convergence.[235] Thus, converging the Δ-SCF ansatz to a specific state can be quite frustrating, and historically, use of the method declined due to the troublesome optimization.

Interest in the method was revitalized by a rather elegant solution for systematically tracking non-Aufbau electronic occupations through the SCF procedure. This maximum overlap method, or MOM, proceeds as follows. Consider an SCF procedure on a system of N_{ele} electrons and N_{bas} basis functions. At the beginning of each SCF iteration, the current molecular orbital coefficients, C_{old}, are used to construct the Fock matrix F, and the generalized eigenvalue problem

$$F[C_{\text{old}}]C_{\text{new}} = SC_{\text{new}}\epsilon \tag{1.44}$$

is then solved for the new molecular orbital coefficients C_{new} and their energies ϵ. Traditionally, the SCF algorithm would select the N_{ele} orbitals with lowest orbital energies, i.e. the Aufbau configuration, as the new molecular orbitals. In MOM, however, orbital overlap matrix O is constructed according to

$$O = C_{\text{old}}^{\dagger} S C_{\text{new}}, \tag{1.45}$$

where O_{ij} quantifies the overlap between the i-th old orbital and the j-th new orbital. The orbital overlaps are then used to compute projections, p, of the new orbitals onto the old occupied orbital space via

$$p_j = \sum_{i}^{N_{\text{ele}}} O_{ij} = \sum_{\nu}^{N_{\text{bas}}} \sum_{\mu}^{N_{\text{bas}}} \sum_{i}^{N_{\text{ele}}} \left(C_{\text{old}}\right)_{i\mu} S_{\mu\nu} \left(C_{\text{new}}\right)_{\nu j} . \tag{1.46}$$

At a cost scaling of only $\mathcal{O}(N_{\text{bas}}^2)$, all N_{bas} new orbital projection values $\{p\}$ are computed, and the N_{ele} new orbitals with the largest projection values are selected for the next round of the SCF cycle. Thus, MOM has returned the new occupied orbitals that overlap the most with the span of the old occupied orbitals.[228] More than its history, MOM is not exclusively a Δ-SCF-DFT method and can be applied to many types of SCF ansatzes, including Δ-SCF-HF,[231] CCSD(T),[236] and CASSCF,[132] and can be integrated into direct-optimization schemes.[237]

Though it has been successfully used in tandem with Δ-SCF-DFT to identify atomic Rydberg states,[220] core excitations of small molecules,[228] intermolecular CT excitations,[220] and the mixed spin and triplet spin excited states of a set of large chromophores that would be of interest in organic photovoltaics such as β-carotene and phthalocyanine (40 second row atoms, each),[217] there are systems in which MOM fails. As observed in the boron atom, during the SCF cycle MOM can allow orbitals to drift through Hilbert space further and further away from the initial configuration. When this happens, it is possible for the final state to have very little overlap with the initial guess despite the projection metrics (Eq. 1.46) between each SCF cycle reflecting a very healthy, stable optimization.[227] Furthermore, while MOM reduces the frequency of variational collapse and occupancy sloshing, both were observed in CASSCF/MOM studies of lithium hydride, ozone, and formaldehyde.[132] Thus, while MOM is applicable to many types of methods and provides a systematic way to track orbitals through optimizations, it could be improved by incorporating it with another method that can target specific excited states.

SCF Metadynamics

Another approach to identifying higher energy HF solutions is SCF metadynamics.[238] In this method, as new RHF solutions are identified via iterative energy minimizations, the energy is modified with a Gaussian biasing potential according to

$$\tilde{E} = E + \sum_{n} N_n e^{-\lambda_n d_{0n}^2} \tag{1.47}$$

where $d_{0n}^2 = \langle \Psi_0 | \rho_0 - \rho_n | \Psi_0 \rangle$ and serves as a distance metric to describe the relative differences in density between the present state 0 and the n-th previously located state. The biasing potential increases the energy in the region surrounding the previously located roots, effectively "filling-in" the minimum and removing the solution from the energy surface. Further SCF optimizations on this modified energy surface should then avoid the roots already

located and identify a new SCF solution. Designed to increase a user's confidence that they have identified the global, rather than local, minimum of the SCF equations, SCF metadynamics can, in theory, be used to exhaustively obtain all solutions of the HF and DFT SCF equations.[238] While SCF metadynamics has been used in tandem with HF theory to create a reference wave function for density matrix embedding theory[239–241] and a basis for nonorthogonal configuration interaction theory,[230, 242] there are perhaps some drawbacks to this method. First, multiple SCF cycles are typically run in order to identify the desired state.[238] Second, since SCF metadynamics can locate all SCF solutions, there can be ambiguity in assigning states, in contrast with optimization methods which are optimized specifically for one excited state.[241] Though no studies applying SCF metadynamics to the applications in Sec. 1.1 were found, one might predict its usefulness in organometallic dyes where there are many degenerate or nearly-degenerate excited states due to the presence of one or more transition metals.[238] However, for systems with a very large number of excited states, such as a DNA fragment, this method may not be the optimal choice due to the rugged nature of the energy landscape.

Variational Principles

The extremization of a target function $\Omega(\Psi(\vec{\nu}))$, which is dependent on the wave function's variable parameters $\vec{\nu}$, is the cornerstone of variational principle-based optimization methods. Once the target function is identified, one typically can construct either a matrix equation ($Ax = b$) or an eigenvalue problem ($Ax = \lambda x$) based on Ω or its derivatives that ensure the final parameterization of $\Psi(\vec{\nu})$ occurs at a critical point of Ω, i.e. where $\nabla_{\vec{\nu}}\Omega = \vec{0}$. Typically, these equations can then be solved efficiently via iterative methods including, but certainly not limited to, Newton and Quasi-Newton methods[243] or Krylov subspace methods[244, 245] such as the Davidson method.[246]

Energy Minimization

The most widely used variational principle is the ground state variational principle,

$$E_0 \leq \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{1.48}$$

$$\Omega_E(\Psi) = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{1.49}$$

where the current wave function is $|\Psi\rangle$ and E_0 is the exact ground state energy. Based on the idea that the wave function $|\Psi\rangle$ can be described as a linear combination of Hamiltonian eigenstates, $|\Psi\rangle = \sum_i c_i|\phi_i\rangle$; $\hat{H}|\phi_i\rangle = E_i|\phi_i\rangle$, where the exact ground state energy of the system E_0 is strictly less than or equal to the remaining eigenvalues, one can show that the current energy of $|\Psi\rangle$ must be greater than or equal to the exact ground state energy.

Therefore, locating the global minimum of target function $\Omega_E(\Psi)$ will return the ground state energy and wave function within the current ansatz.[2]

While powerful in identifying ground states in HF theory and the lowest few eigenvalues in CI theories, attempts to use the ground state variational principle to locate higher-energy excited states are plagued with variational collapse, an occurrence in an optimization in which the the wave function deviates significantly from its initial state and optimizes to a different energy stationary state. Consider the energy surface in Fig. 1.9, in which a wave function dependent on variational parameters ν_1 and ν_2 undergoes energy minimization. Despite the initial guess's location at the yellow star and its proximity to excited state c, there are many possible and more probable routes (depicted by yellow arrows) the optimization could take that would lead the wave function into states a or b rather than c. As the excited state cannot be the global minimum of $\Omega_E(\Psi)$, variational collapse is a very realistic obstacle to identifying the desired excited state that is quite likely to occur unless the initial guess for the wave function is already very close to that of the targeted state. Thus, simply minimizing the ground state variational principle without any other considerations in the optimization is a particularly poor idea for excited states.

Variance Minimization

An alternative to energy minimization popular in Quantum Monte Carlo (QMC) methods, variance minimization uses the statistical uncertainty of the energy as a metric to optimize the wave function. Defined as

$$\Omega_{\sigma^2}(\Psi) = \sigma^2(\Psi) = \frac{\langle\Psi|\big(E-\hat{H}\big)^2|\Psi\rangle}{\langle\Psi|\Psi\rangle} = \frac{\langle\Psi|\hat{H}^2|\Psi\rangle}{\langle\Psi|\Psi\rangle} - E^2 \tag{1.50}$$

σ^2 is equal to zero when $|\Psi\rangle$ is an exact eigenstate of the Hamiltonian. Therefore, minimizing $\Omega_{\sigma^2}(\Psi)$ with respect to $|\Psi\rangle$ provides optimization pathways for both ground and excited states. However, σ^2 is at its global minimum of zero for *every* eigenstate, and this lack of uniqueness can also result in variational collapse if the initial wave function is not sufficiently close to the desired state. Although variance minimization has proven to be successful in excited state optimizations, its utility decreases in scenarios where there are degeneracies or near-degeneracies in the spectrum or when a "close-enough" initial wave function guess is not available.[247–252]

Excited State Variational Principles

As a solution to the optimization issues experienced with energy and variance minimization, excited state variational principles introduce the concept of state targeting or state specificity to excited state optimizations. These variational principles are designed such that their global minimum occurs for only one eigenstate which is selected via an identifying energetic target provided by the user.

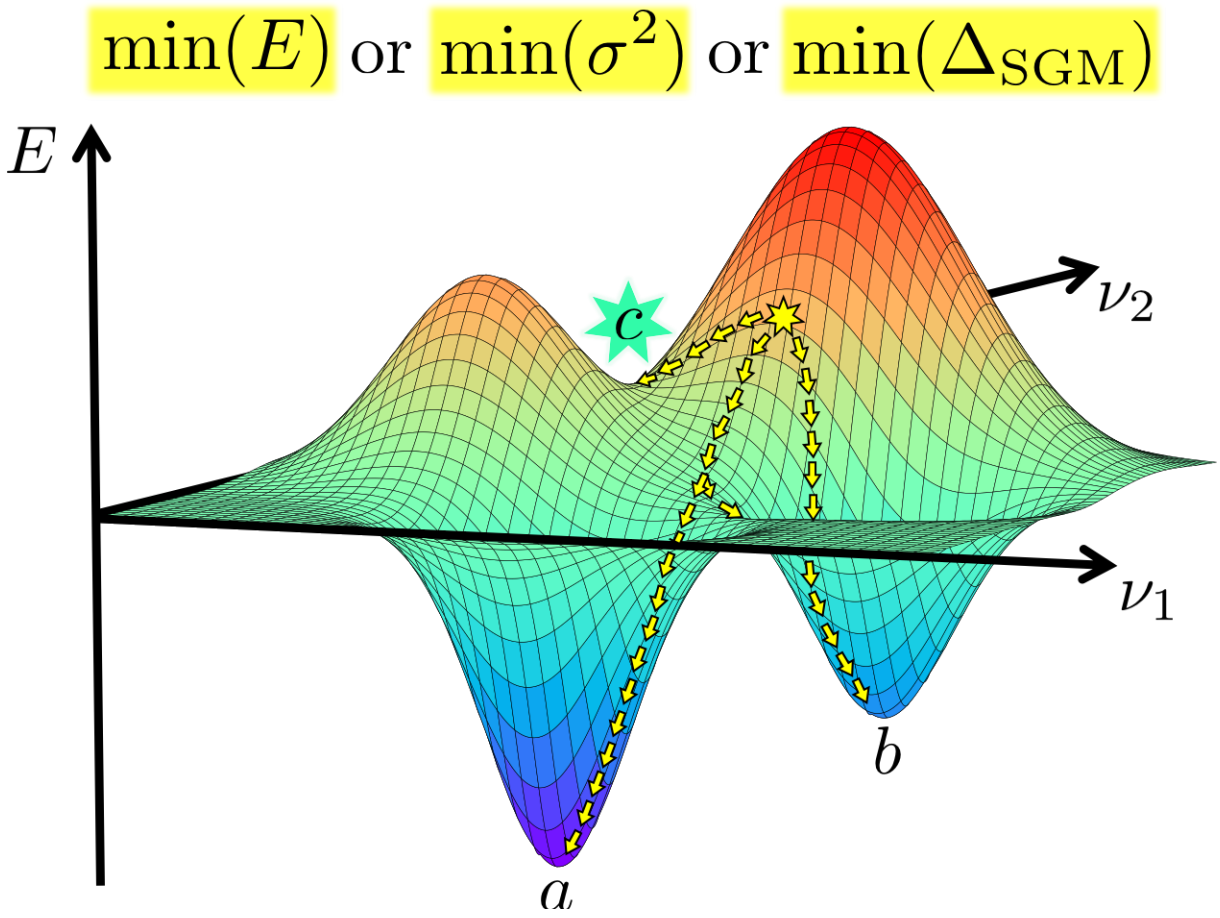


Figure 1.9: Possible routes for variational collapse of E, σ^2, and Δ_{SGM} when targeting excited state c

The following two excited state variational principles will be overviewed here.

$$\hat{W}_1(\Psi) = \langle\Psi|\big(\omega - \hat{H}\big)^2|\Psi\rangle \tag{1.51}$$

$$\hat{W}_2(\Psi) = \frac{\langle\Psi|\omega - \hat{H}|\Psi\rangle}{\langle\Psi|\big(\omega - \hat{H}\big)^2|\Psi\rangle}. \tag{1.52}$$

Using energy targeting parameter ω, which is set to an approximate value of the desired eigenstate's energy, both W_1 and W_2 essentially quantify the distance between the current energy of the wave function and the targeted energy. The global minimum of $\hat{W}_1(\Psi)$ occurs when the energy of $|\Psi\rangle$ is nearest ω and is unique except in cases of energetic degeneracy or incredibly rare coincidences between the value of ω and the energy of the eigenstates above and below it.[116, 253–255] In slight contrast, ω is a lower bound for the energy of the targeted eigenstate when using $\hat{W}_2$, for which the global minimum occurs at the eigenstate

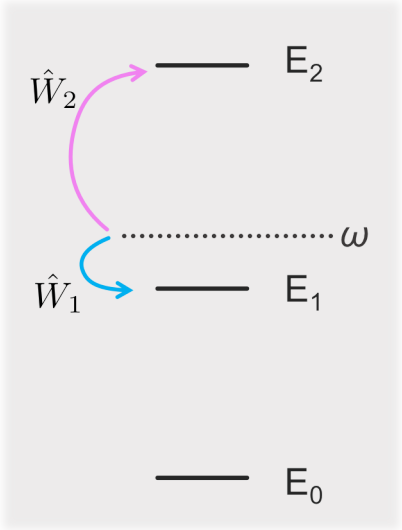


Figure 1.10: Excited state variational principle targeting behavior

with energy nearest *but greater than* ω.[256] The differing behaviors of $\hat{W}_1$ and $\hat{W}_2$ with respect to ω are depicted in Fig. 1.10. In this scenario, ω is closer to the energy of $|\Psi_1\rangle$ than that of $|\Psi_2\rangle$. Therefore, the global minimum of $\hat{W}_1$ occurs at $\{E_1, |\Psi_1\rangle\}$ while the global minimum of $\hat{W}_2$ occurs at $\{E_2, |\Psi_2\rangle\}$.

While unappealing in deterministic quantum chemistry methods due to the expense of computing $\langle \hat{H}^2 \rangle$, excited state variational principles that include the square of the Hamiltonian are not cost-prohibitive in Quantum Monte Carlo (QMC) algorithms as these can avoid explicitly squaring $\hat{H}$ by resolving the identity $\mathbb{1} = \sum_m |m\rangle\langle m|$ between factors of $\hat{H}$.

$$\langle \Psi | \hat{H}^2 | \Psi \rangle = \sum_m \langle \Psi | \hat{H} | m \rangle \langle m | \hat{H} | \Psi \rangle = \sum_m \left| \frac{\langle m | \hat{H} | \Psi \rangle}{\langle m | \Psi \rangle} \right|^2 \left| \langle m | \Psi \rangle \right|^2 \tag{1.53}$$

By drawing samples ξ from the probability distribution $\left|\langle m|\Psi\rangle\right|^2$ via a Metropolis walk,[257] the variational principle $\hat{W}_1$, for example, can be evaluated approximately to a controllable statistical uncertainty using Monte Carlo integration through the following expression.[249, 255, 256]

$$\hat{W}_1 \approx \sum_{m \in \{\xi\}} \left| \frac{\langle m | \omega - \hat{H} | \Psi \rangle}{\langle m | \Psi \rangle} \right|^2 \tag{1.54}$$

Despite their state-targeting capabilities, another drawback of excited state variational principles that include $\hat{H}^2$ and a static targeting parameter is their lack of size consistency. Traditionally thought of as a property of wave function ansatzes, it is demonstrated in Chapter 2 that an optimization method can also be a source of size consistency errors. In the

upcoming chapter, the author presents an optimization algorithm that combines the ideas of excited state variance principle minimization with variance minimization in order to restore size consistency while rigorously maintaining state specificity. If not for the large prefactor expense associated with QMC methods,[258] this algorithm would allow for QMC studies of specific excited states within large complexes such as DNA fragments and organometallic dyes. The algorithm does not, however, enable these excited state variational principles to resolve degenerate eigenstates. As $\hat{W}_1$ and $\hat{W}_2$ are only dependent on energy-related properties of the wave function, optimized solutions in the presence of a degeneracy can be superpositions of states, depending of course on the initial guess and symmetry constraints. To be discussed further in Chapter 4, these obstacles are completely hurdled by a novel, broadly applicable, size consistent, generalized variational principle that avoids computing $\langle \hat{H}^2 \rangle$ altogether, and instead minimizes a property deviation vector.

Energy Gradient Minimization

In a similar vein as variational principal minimization, one can directly search for stationary points of the energy by minimizing functions dependent on the gradient of the energy such that at convergence $\nabla_{\vec{\nu}} E = \vec{0}$.

This approach was used in the initial implementation of the ESMF method[117] as discussed in Chapter 3. Through the formalism of Lagrangian optimization, the overall goal of the method was to target a specific state by minimizing an approximation to the excited state variational principle $\hat{W}_1$ under the constraint that critical points of the Lagrangian target function are also critical points of the energy, i.e.

$$L(\Psi, \vec{\lambda}) = \langle \Psi | \big(\omega - \hat{H}\big)^2 | \Psi \rangle + \vec{\lambda} \cdot \nabla_{\vec{\nu}} E \tag{1.55}$$

where $\vec{\lambda}$ are Lagrange multipliers that are optimized in tandem with wave function parameters $\vec{\nu}$. As previously mentioned, computing $\langle \hat{H}^2 \rangle$ is costly in deterministic quantum chemistry methods, so Eq. 1.55 was modified with an approximation to $\hat{W}_1$,

$$L(\Psi, \vec{\lambda}) \approx \big(\omega - E\big)^2 + \vec{\lambda} \cdot \nabla_{\vec{\nu}} E \tag{1.56}$$

for which derivatives with respect to $\vec{\nu}$ and $\vec{\lambda}$ are as follows,

$$\nabla_{\vec{\nu}} L(\Psi, \vec{\lambda}) = -2\big(\omega - E\big)\nabla_{\vec{\nu}} E + \mathbf{H}_E \cdot \vec{\lambda} \tag{1.57}$$

$$\nabla_{\vec{\lambda}} L(\Psi, \vec{\lambda}) = \nabla_{\vec{\nu}} E \tag{1.58}$$

where $\mathbf{H}_E$ is the Hessian of the ESMF energy. Note, however, that excited state solutions of $L(\Psi, \vec{\lambda})$ are neither minima of $L(\Psi, \vec{\lambda})$, as it is unbounded from above and below due to the constraint term, nor zeros of $L(\Psi, \vec{\lambda})$, as the first and second terms in Eq. 1.56 could cancel each other when $\nabla_{\vec{\nu}} E \neq \vec{0}$, but rather high-index saddle points where $\nabla L = \vec{0}$. As finding

saddle points requires necessarily more complicated optimization procedures than finding minima, we instead elected to minimize the target function

$$\Omega(\Psi, \vec{\lambda}) = \left|\nabla L(\Psi, \vec{\lambda})\right|^2 \tag{1.59}$$

which has a unique global minimum at the Hamiltonian eigenstate with energy closest to ω and is more amenable to Quasi-Newton optimization algorithms such as BFGS.[259–262] Incorporating state targeting and energy gradient minimization, this approach provides a general framework for ensuring that the correct stationary point is found even if a high-quality initial wave function guess is not available, and further work in which analytic gradients of the ESMF energy and $\Omega(\Psi, \vec{\lambda})$ were derived resulted in a more efficient implementation of the optimizer than was realized in our initial pilot code.[263]

Very recently, additional innovations in energy gradient optimization methods have been developed. The square gradient minimization (SGM) method, which seeks to minimize Δ_{SGM}, the squared norm of the gradient of variational principle $f(\Psi)$, with respect to the wave function parameters $\vec{\nu}$

$$\Delta_{\text{SGM}}(\Psi) = \left|\nabla f(\Psi)\right|^2 \tag{1.60}$$

extends the efficient framework of the geometric direct minimization (GDM) method[264] via a finite difference formulation that is numerically efficient and applicable to many wave function ansatzes.[265] In this study, the system energy was used as f, although the authors posit that the overall optimization algorithm is independent of the choice of the variational principle of interest.[265] Since $\nabla E = \vec{0}$ when the wave function is a eigenstate of the Hamiltonian, it follows that Δ_{SGM} is strictly greater than or equal to zero, with equality occurring for every stationary state of the variational principle.

Just as with the excited state variational principles in Eqs. 1.51 and 1.52, one drawback of energy gradient minimization methods that are only functions of E, ∇E, and its higher-order derivatives is the inability of these methods to resolve energetically-degenerate eigenstates. We will see in Chapter 4 how incorporating other state-targeting properties in the generalized variational principle remedies this problem.

1.6 Outline

Chapter 2

Publication: Jacqueline A. R. Shea and Eric Neuscamman. "Size Consistent Excited States via Algorithmic Transformations between Variational Principles." *J. Chem. Theory Comput.* **13 (12)**, 6078-6088 (2017).

Diving first into optimization methods, the properties of of a prominent class of state targeting excited state variational principles are explored, and the lack of size consistency in this subset is proven formally and numerically. The author then details a novel approach that transforms between variational principles on the fly in a manner that guarantees size

consistency while also rigorously targeting a specific excited state. Now implemented in a leading QMC software package,[266] the efficacy of this approach is demonstrated in a system of non-interacting small molecules as well as in a hydrogen-bonded formaldehyde-water complex.

Chapter 3

Publication: Jacqueline A. R. Shea and Eric Neuscamman. "Communication: A mean field platform for excited state quantum chemistry." *J. Chem. Phys.* **149**, 081101 (2018).

Moving from stochastic quantum chemistry to deterministic quantum chemistry, the excited state mean field (ESMF) ansatz is presented. With the flexibility to describe states with multiple contributing excitations, ESMF delivers fully relaxed excited state orbitals at only $\mathcal{O}(N^4)$ cost. In analogy to ground state mean field theory, the ESMF wave function is a qualitatively correct ansatz that serves as a platform upon which higher order correlation methods can be built, as demonstrated by the development of excited state Møller-Plesset perturbation theory (ESMP2).

Chapter 4

Publication: Jacqueline A. R. Shea, Elise Gwin, and Eric Neuscamman. "A Generalized Variational Principle with Applications to Excited State Mean Field Theory." *J. Chem. Theory Comput.* **16 (3)**, 1526-1540 (2020).

Returning to optimization methods, a generalization of the variational principle that can target specific excited states, even when states are energetically degenerate or near-degenerate, is described. Applied to ESMF theory, the optimization efficiency of the generalized variational principle relative to the previously used Lagrangian target function scheme is discussed, and the accuracy of ESMP2 theory is tested on a much wider range of small organic molecules and charge transfer systems than previously accessible. Furthermore, spin-targeting, partial charge-targeting, and energy-targeting allow for degenerate CT states in stretched lithium hydride to be resolved, demonstrating the broad applicability of this optimization strategy.

Chapter 5

Summarizing the novel developments and implications of the discussed works, the author then looks forward to future methodological developments and demonstrations of the ESMF as a novel platform for excited state quantum chemistry.

Chapter 2

Size Consistent Excited States via Algorithmic Transformations between Variational Principles

2.1 Abstract

We demonstrate that a broad class of excited state variational principles is not size consistent. In light of this difficulty, we develop and test an approach to excited state optimization that transforms between variational principles in order to achieve state selectivity, size consistency, and compatibility with quantum Monte Carlo. To complement our formal analysis, we provide numerical examples that confirm these properties and demonstrate how they contribute to a more black box approach to excited states in quantum Monte Carlo.

2.2 Introduction

In a large range of chemical and materials applications, including homogeneous catalysis, liquid-phase light harvesting, and band gap engineering, it is essential that the predictions made by theoretical methods retain their accuracy as the system size is varied. For example, a method whose accuracy was highly dependent on the number of solvent molecules included in a simulation is less useful than a method whose accuracy is not. Likewise, when performing solid-state calculations on a series of increasingly large simulation cells in order to get a grip on finite-size effects [267], it is important that a method's accuracy not be dependent on the system size, or else it will be difficult to separate real finite size effects from other methodological artifacts. At the very least, it is desirable that methods used in these contexts satisfy size consistency, which demands that two separated systems produce the same results whether they are simulated independently or together.

Famously, not all wave function approximations satisfy size consistency. While the coupled cluster ansatz does, truncated configuration interaction does not [2, 268–271]. Like-

wise, the single Slater determinant of standard Hartree-Fock theory is size consistent, but symmetry-projected Hartree-Fock theory is not [2, 272, 273]. The antisymmetric geminal power (AGP) [274–279] is not size consistent when used alone, but becomes size consistent when paired with the right type of Jastrow factor (JAGP) [280, 281]. Indeed, when designing new or improved wave function ansatzes, an important theoretical test is to check whether or not size consistency is retained.

Like wave function ansatzes, variational principles come in both size consistent and size inconsistent varieties. The most famous and widely used variational principle, the ground state energy, is of course size consistent, but others, including some used for the direct optimization of excited states [255, 256], are not. Indeed, even when paired with a size consistent wave function (e.g. one that product factorizes) such size inconsistent variational principles can lead to size inconsistent results. Thus, when designing variational principles and methods based on them, it is important to consider the consequences that different choices will have on size consistency.

Of course, many other properties, not least of which is affordability, must be considered when designing principles and algorithms for use in the optimization of wave functions. For example, one can imagine incorporating higher powers of the Hamiltonian operator when constructing a new variational principle, although in practice it is quite rare to see powers higher than two for the simple reason that higher Hamiltonian powers tend to lead to higher evaluation costs. As accurate electronic structure methods are already quite computationally intensive, it is not appealing to raise costs further.

Unfortunately, there is a strong formal problem that arises for excited states when limiting the functional form of a variational principle to include only the first and second power of the Hamiltonian. As we prove in this paper, such variational principles cannot simultaneously target an individual excited state and remain size consistent. In light of this challenge, we advocate that in practice a wave function optimization method intended for use with excited states may be best served by amalgamating multiple variational principles. For example, as was achieved recently by the σ-SCF method, [116] an optimizer might begin by minimizing a size inconsistent but state specific variational principle, but upon approaching convergence gradually transition to minimizing a size consistent but state nonspecific variational principle. The idea is for the first variational principle to get the optimization close enough to the desired eigenstate so that the lack of state specificity in the final variational principle is no longer an issue.

Following our formal proof, we will present one such amalgamation that works in the context of wave function optimization via quantum Monte Carlo (QMC).[258] Crucially, QMC can work with many excited state variational principles for a cost similar to its ground state cost, [256] and ground state QMC can reach scales up to hundreds of atoms thanks to its low scaling and easy parallelization.[267, 282] The realization of a state specific and size consistent excited state optimizer in QMC thus marks an important step towards achieving more reliable predictions of excited states and spectral properties in complicated molecules and materials.

2.3 Theory

Variational Principles

For the purposes of this paper, let us define a state selective variational principle as a smooth function of a wave function ansatz's variables with the following property: if the ansatz is capable of exactly describing the individual Hamiltonian eigenstate of interest, then the function will have its unique global minimum at the variable values corresponding to that exact eigenstate. If the state being targeted is the ground state, as occurs for the function

$$E(\Psi) = \langle \hat{H} \rangle = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \tag{2.1}$$

then we will call the function a ground state variational principle. An excited state variational principle is, therefore, a state selective variational principle for which the global minimum corresponds to an excited state.

Note that the energy variance

$$\sigma^2(\Psi) = \frac{\langle \Psi | (\hat{H} - E)^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2, \tag{2.2}$$

can be employed as a variational principle, [247–252] but that it is not state selective, as its global minimum is not unique. Indeed, any Hamiltonian eigenstate gives the equally low value of $\sigma^2 = 0$. As we will discuss further in Section 2.4, this lack of state selectivity can make optimization to the desired eigenstate more difficult than one would prefer.

To be practical, a variational principle must be paired with an efficient method for its evaluation and minimization. This requirement more or less explains why the energy-based ground state variational principle has been more successful than variational principles for excited states. Note that the functional form of E requires an expectation value of only the first power of the Hamiltonian, in contrast to σ^2 whose evaluation requires expectation values of both $\hat{H}$ and $\hat{H}^2$, the latter of which is in most circumstances more computationally demanding. Indeed, the construction of excited state variational principles that work by measuring a wave function's "energetic distance" from a desired position ω in the spectrum, such as [116, 253–255]

$$W(\Psi) = \frac{\langle \Psi | (\omega - \hat{H})^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} = (\omega - E)^2 + \sigma^2 \tag{2.3}$$

or [256]

$$\Omega(\Psi) = \frac{\langle \Psi | (\omega - \hat{H}) | \Psi \rangle}{\langle \Psi | (\omega - \hat{H})^2 | \Psi \rangle} = \frac{\omega - E}{(\omega - E)^2 + \sigma^2}, \tag{2.4}$$

tend to also require $\hat{H}^2$, because computing a distance typically involves taking a square. In this respect, variational Monte Carlo (VMC) offers the advantage that $\hat{H}^2$ expectation values can be evaluated via Monte Carlo integration [251, 256] of the integral

$$\langle \hat{H}^2 \rangle = \frac{\langle \Psi | \hat{H}^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \int \frac{|\langle \Psi | \vec{r} \rangle|^2}{\langle \Psi | \Psi \rangle} \left| \frac{\langle \vec{r} | \hat{H} | \Psi \rangle}{\langle \vec{r} | \Psi \rangle} \right|^2 d\vec{r}. \tag{2.5}$$

This approach avoids having to explicitly square the Hamiltonian operator and is thus similar in difficulty to a VMC evaluation of the energy E. In principle, variational principles that depend on cubic or higher powers of $\hat{H}$ could be constructed, but these are likely to be even less practical, and so we will for this study limit our attention to quadratic and lower powers of $\hat{H}$.

The $V_{1,2}$ set

To begin, let us define $V_{1,2}$ as the set of all state selective variational principles that have the following three properties. First, in the interest of generality and affordable evaluation, we require the functional form of any $\Gamma \in V_{1,2}$ to depend on the wave function variables only through the expectation values $\langle \hat{H} \rangle$ and $\langle \hat{H}^2 \rangle$. Using Eqs. (2.1) and (2.2), we see that this is the same as requiring that Γ depend on the wave function only through E and σ^2,

$$\Gamma(\Psi) = \Gamma(E(\Psi), \sigma^2(\Psi)). \tag{2.6}$$

Second, we require that Γ have a unique global minimum corresponding to a particular interior and nondegenerate eigenstate of $\hat{H}$, thus limiting the analysis to nondegenerate excited states and excluding both $\Gamma = E$ and $\Gamma = \sigma^2$ as possibilities. Finally, we require Γ to be real analytic (i.e. real valued and equal to its Taylor series) in a contiguous, open region around the global minimum $(E, \sigma^2) = (E_t, 0)$, where E_t is the energy of the targeted eigenstate. Upon close inspection, one finds that both $\Omega \in V_{1,2}$ and $W \in V_{1,2}$ when ω is close to but below E_t.

No Size Consistency in $V_{1,2}$

Interestingly, one can show that variational principles within $V_{1,2}$ are not size consistent when paired with an approximate wave function ansatz, even if that ansatz is product factorizable. Of course, with an exact ansatz, results would be exact, and with a non-factorizable ansatz, they would contain size inconsistency errors regardless of the variational principle employed. It is thus the fact that members of the $V_{1,2}$ set are size inconsistent for factorizable but approximate ansatzes that makes them fundamentally different from the ground state variational principle or variance minimization.

Although the formal proof of size inconsistency is somewhat involved (see next section), it can be motivated by showing that any Γ in $V_{1,2}$ must have a nonlinear form. First, note that

$\Gamma(E, \sigma^2)$ can be written as a Taylor series around the energy E_t of the targeted eigenstate.

$$\Gamma(E, \sigma^2) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} (\sigma^2)^m (E - E_t)^n. \tag{2.7}$$

In order to satisfy properties of the $V_{1,2}$ set, this Taylor expansion is subject to a number of restrictions, the most noteworthy of which is that it must contain nonlinear terms. Consider if instead Γ took the particularly simple form in which all coefficients are zero except for a_{10},

$$\Gamma = a_{10} \sigma^2 \tag{2.8}$$

so that it was proportional to Eq. (2.2). As every Hamiltonian eigenstate has $\sigma^2 = 0$, this form does not have a unique global minimum and is thus not within the $V_{1,2}$ set.

Next, imagine that the only nonzero term in the series was

$$\Gamma = a_{01}(E - E_t). \tag{2.9}$$

In this case, Γ is linear with respect to E, and like the ground state variational principle (Eq. (2.1)), is not a member of the $V_{1,2}$ set as it does not target an interior eigenstate of $\hat{H}$. The only other way for Γ to remain linear in its variables would be for it to be a linear combination of σ^2 and $(E - E_t)$,

$$\Gamma(E, \sigma^2) = a_{01}(E - E_t) + a_{10}\sigma^2. \tag{2.10}$$

However, as $\sigma^2 = 0$ for any Hamiltonian eigenstate, we find that this form also fails to target an interior eigenstate, as it will be minimal for either the lowest or highest eigenstate (depending on the sign of a_{01}).

Concluding that Γ must be nonlinear if it is to be an element of $V_{1,2}$, we now consider the effects that nonlinearity will have on size consistency when working with an approximate ansatz. Of course, we should not expect results to be size consistent if the ansatz does not product factorize, and so we assume that we are working with two completely separate subsystems A and B and that our overall ansatz can be written as a product of separate ansatzes for the two subsystems, $\Psi_{AB} = \Psi_A \Psi_B$. In this case, both the energy and variance will be additive:

$$E_{AB} = E_A + E_B, \tag{2.11}$$
$$\sigma^2_{AB} = \sigma^2_A + \sigma^2_B, \tag{2.12}$$

where $E_{AB} = E(\Psi_{AB})$ is the energy of the system when evaluated as a combined whole and $E_A = E(\Psi_A)$ is the energy of A when treated alone. Now, if Γ were linear in E and σ^2, then it would also be additively separable into $\Gamma_{AB} = \Gamma_A + \Gamma_B$ and size consistency would follow. However, the fact that Γ must be nonlinear in order to target a unique interior eigenstate implies that it will contain cross terms that, in general, can be expected to prevent additive

separability and thus violate size consistency. Consider the variational principle W, Eq. (2.3), which when treating systems A and B together will take on the form

$$\begin{aligned} W_{AB} &= (\omega_A + \omega_B - E_A - E_B)^2 + \sigma_A^2 + \sigma_B^2 \\ &= W_A + W_B + 2(\omega_A - E_A)(\omega_B - E_B). \end{aligned} \tag{2.13}$$

The presence of the cross term means that the minimum of W_{AB} will be different from that of $W_A + W_B$, which is a size consistency violation that will cause the optimal parameters in our ansatz to differ when treating A and B separately or together. As nonlinear functions will in general have such cross terms, it is therefore reasonable to expect that variational principles in $V_{1,2}$ will violate size consistency. While we do provide a more rigorous proof of this flaw in the next section, readers who are satisfied with the argument presented above may wish to skip ahead to Section 2.3, where we present an optimization method that mixes different variational principles in order to target interior eigenstates while maintaining size consistency.

Proof of No Size Consistency

We now proceed to formally show that any variational principle within $V_{1,2}$ is not size consistent. First, note some general properties that the Taylor series of Γ, Eq. (2.7), must satisfy if it is to be a member of $V_{1,2}$. To start, we note that there must be a nonzero coefficient among the terms with $m = 0$ and $n > 1$. If there were not, then either Γ would not be state specific or it would not target an interior eigenstate. Similarly, there must be a nonzero coefficient among the terms with $n = 0$ and $m > 0$, or else any state with E_t as its energy expectation value would give the same value for Γ as the targeted eigenstate. Finally, for ease of analysis and without loss of generality, we will set $a_{00} = 0$, as this does not alter the nature of the global minimum. With these restrictions and defining $\Delta = E - E_t$, we can write the Taylor series as

$$\Gamma = \sum_{m=p}^{\infty} a_{m0}(\sigma^2)^m + \sum_{m=q}^{\infty}\sum_{n=r}^{\infty} a_{mn}(\sigma^2)^m\Delta^n + \sum_{n=s}^{\infty} a_{0n}\Delta^n \tag{2.14}$$

in which p, q r, and s are positive integers, $a_{p0} \neq 0$, $a_{0s} \neq 0$, and it is understood that there exist $1 < \tilde{s} < \infty$ and $0 < \tilde{p} < \infty$ such that $a_{0\tilde{s}} \neq 0$ and $a_{\tilde{p}0} \neq 0$. Note that it is possible for all the elements in the middle sum to have zero coefficients, but they can in general be nonzero. In the latter case, $a_{qr} \neq 0$.

With these qualities of Γ in mind, consider the stationary condition

$$0 = \frac{\partial\Gamma}{\partial\Delta}\frac{\partial E}{\partial x} + \frac{\partial\Gamma}{\partial\sigma^2}\frac{\partial\sigma^2}{\partial x} \tag{2.15}$$

for minimizing Γ when system A is treated alone with an ansatz depending on a single variable x. For any choice of $\Gamma \in V_{1,2}$, we show in Appendix A that there exist system/ansatz pairs

for which neither E nor σ^2 are stationary at the global minimum of Γ. Choosing this type of system/ansatz pair for system A and defining the analytic functions

$$\mu(\Delta, \sigma^2) \equiv \frac{\partial \Gamma}{\partial \Delta} = \sum_{m=q}^{\infty} \sum_{n=r}^{\infty} n\, a_{mn} (\sigma^2)^m (\Delta)^{n-1} + \sum_{n=s}^{\infty} n\, a_{0n} (\Delta)^{n-1} \tag{2.16}$$

and

$$\nu(\Delta, \sigma^2) \equiv \frac{\partial \Gamma}{\partial \sigma^2} = \sum_{m=p}^{\infty} m\, a_{m0} (\sigma^2)^{m-1} + \sum_{m=q}^{\infty} \sum_{n=r}^{\infty} m\, a_{mn} (\sigma^2)^{m-1} (\Delta)^n \tag{2.17}$$

we can rewrite the stationary condition as

$$0 = \mu(\Delta, \sigma^2) \frac{\partial E}{\partial x} + \nu(\Delta, \sigma^2) \frac{\partial \sigma^2}{\partial x}. \tag{2.18}$$

When system A is alone, this condition is satisfied for some $x = x_A$ at which $\Delta = \Delta_A$, $\sigma^2 = \sigma_A^2$, $\partial E/\partial x \neq 0$ and $\partial \sigma^2/\partial x \neq 0$. Now imagine if we added a system Q that is completely separated from system A such that $\hat{H} = \hat{H}_A + \hat{H}_Q$. Choosing the overall wave function ansatz to be a product of the ansatzes from A and Q such that Eqs. (2.11) and (2.12) apply, we see that Δ and σ^2 will be the only parts of Eq. (2.18) affected by the addition of Q so long as x is held fixed at $x = x_A$. Crucially, note that $\partial E/\partial x$ and $\partial \sigma^2/\partial x$ are not affected.

We now separate $V_{1,2}$ into two subsets and show that size consistency is violated in both. First, take the subset in which the middle sum of Eq. (2.14) is absent, $p = 1$, and $a_{m0} = 0$ for $m > 1$. In this case, the right hand side of Eq. (2.18) will not be a function of σ^2. If we were to hold $x = x_A$ fixed, then it would be a nonconstant and analytic function of Δ, and thus by the principle of permanence its root at $\Delta = \Delta_A$ would be isolated. This implies that upon adding a system Q so that $\Delta \to \Delta_A + \Delta_Q$, the stationary condition would for small but nonzero Δ_Q no longer be satisfied unless the value of x were adjusted. Thus, in this subset of $V_{1,2}$, the addition of system Q would change the optimal wave function in system A, despite the overall wave function product factorizing and the two subsystems not interacting.

Second, take the subset that contains all $\Gamma \in V_{1,2}$ not in the previous subset. In this case, if we were to hold $x = x_A$ fixed, the right hand side of Eq. (2.18) will be an analytic function that depends on both Δ and σ^2. By now adding one subsystem of type B and one of type C (see Appendix B) such that none of the three subsystems interact and the overall wave function is a product of the three subsystem wave functions, we will have

$$\Delta = \Delta_A + \Delta_B + \Delta_C \tag{2.19}$$

$$\sigma^2 = \sigma_A^2 + \sigma_B^2 + \sigma_C^2 \tag{2.20}$$

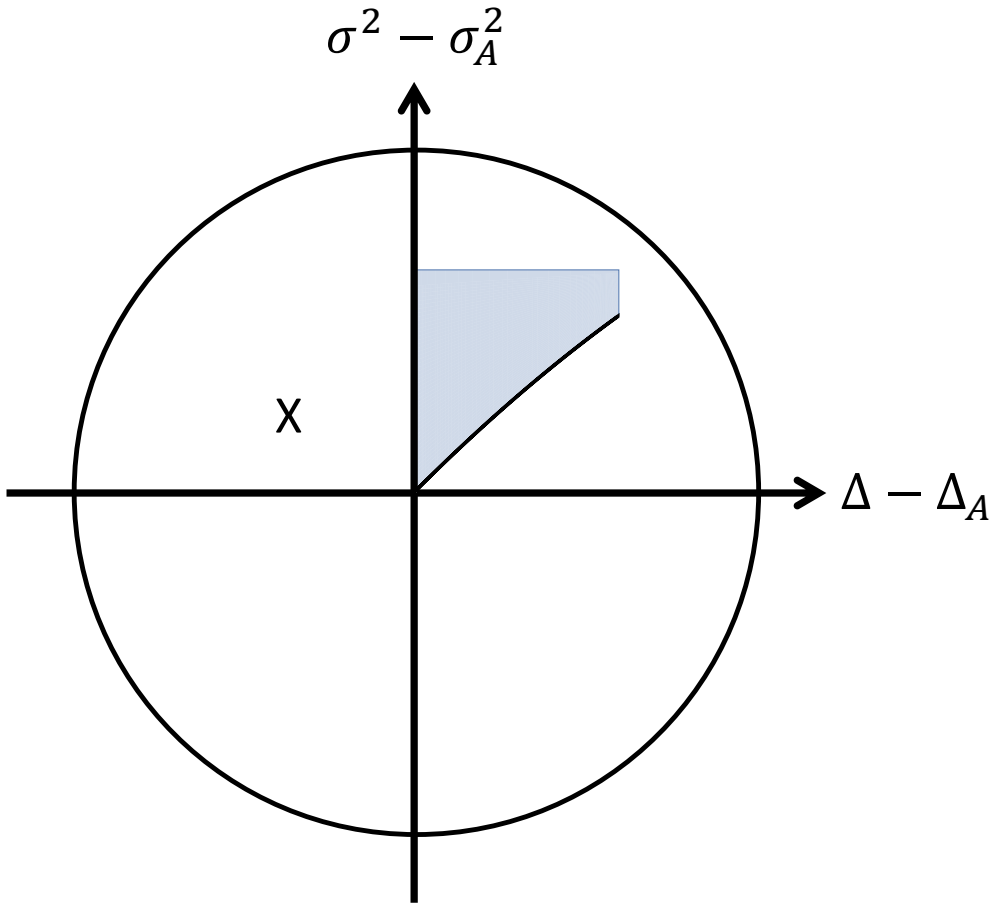


Figure 2.1: A schematic showing values of Δ and σ^2 that can be reached by adding systems of types B and C. The line shows values accessible by the addition of systems of type B. The shaded region shows values accessible by the addition of systems of types B and C together. The circle and the X give examples for Γ's analytic region and global minimum, respectively.

in which

$$\Delta_B = \frac{\alpha^2}{1+\alpha^2}, \tag{2.21}$$

$$\sigma_B^2 = \frac{\alpha^2}{(1+\alpha^2)^2}, \tag{2.22}$$

$$\Delta_C = 0, \tag{2.23}$$

$$\sigma_C^2 = \frac{\beta^2}{1+\beta^2}, \tag{2.24}$$

where α and β are real numbers. By choosing different systems B and C, we may vary α and β to map out a contiguous two-dimensional patch within the region on which Γ is analytic, as shown in Figure 2.1. As we can choose the system/ansatz pair in A such that its stationary point (Δ_A, σ_A^2) is arbitrarily close to the global minimum, we may assume without loss of generality that this patch is inside the region within which Γ is analytic. If the stationary condition in Eq. (2.18) were satisfied at all points in the patch, then by repeated use of the principle of permanence, we see that it would also be satisfied at all points in an open region encompassing the global minimum. As this would violate our assumption of a unique global minimum, we must conclude that at the vast majority of points in the mapped-out patch, i.e. for most choices of systems B and C with small α and β, Eq. (2.18) will not be satisfied when $x = x_A$. In other words, the addition of these completely separate subsystems changes the optimal wave function in system A. As this will in turn change the energy, we see that size consistency is violated.

To summarize, we have found that for any $\Gamma \in V_{1,2}$, it is possible to construct a product separable ansatz for completely separate subsystems in such a way that the optimal wave function on one subsystem is changed by the presence of other subsystems. As a result, the total energy will be different if we treat the systems separately instead of together. We must thereby conclude that there are no size consistent variational principles in $V_{1,2}$.

Transformations Between Variational Principles

While individual members of $V_{1,2}$ are not size consistent, it is nonetheless possible to employ them as part of an overall optimization scheme that is both state selective and size consistent. As was achieved for Slater determinants in the σ-SCF method [116], the general strategy is to begin the optimization with a state selective variational principle in order to ensure the correct state is targeted. Once the wave function was "close" to the desired state, σ-SCF prescribed a transition to state nonspecific variance minimization, which, among other benefits, ensures size consistency. Here, we present an evolution of this general strategy that both makes it compatible with VMC and guarantees that state selectivity is maintained throughout the optimization, even in the final stage in which size consistency is achieved through variance minimization.

The key to our strategy is to recognize that special choices for ω can make either W or Ω (and likely many other members of $V_{1,2}$) become akin to variance minimization. For example, one sees that

$$W(\Psi)\Big|_{\omega \to E} = \sigma^2 \tag{2.25}$$

and

$$\Omega(\Psi)\Big|_{\omega \to E-\sigma} = -\frac{1}{2\sigma}. \tag{2.26}$$

If, in the final stage of the optimization, we ensure that ω is chosen appropriately and self consistently, then minimizing W or Ω for a particular state will produce the same result as

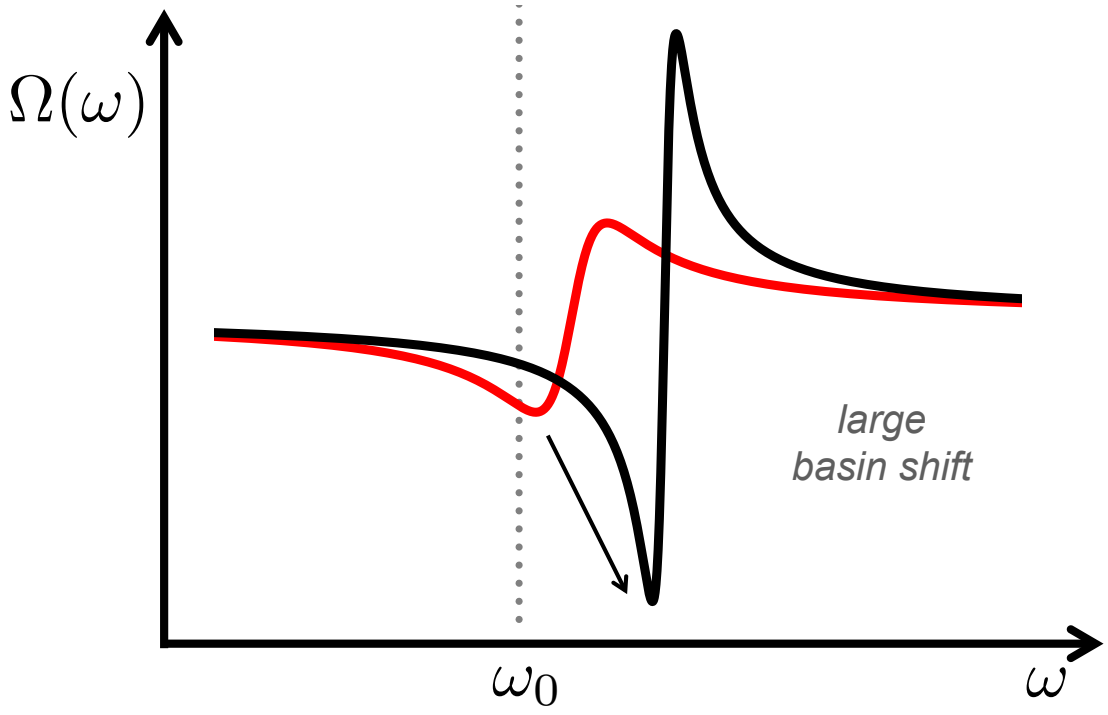


Figure 2.2: Example of a large basin shift for Ω that may occur if we skip the transitional-ω stage of the optimization.

if variance minimization had been achieved for that state. Crucially, we adopt a strategy in which Ψ and ω are updated *separately* in a "tick-tock" fashion, which ensures that ω is fixed during an update step for Ψ. This choice guarantees that the desired state is targeted, as it remains the global minimum of W or Ω during the Ψ update step. If we instead simply switched to variance minimization, we would in general have to rely on the state in question being a stable local minimum of the variance, which does not offer the same convergence guarantees as a state selective approach in which the desired state is the global minimum. Between each Ψ update, we adjust ω to its special value (e.g. E or $E-\sigma$) so that at convergence the result is equivalent to variance minimization and thus size consistent. Although we have chosen to test this strategy using Ω as the variational principle and the VMC linear method [256, 283–287] as the wave function update method, we expect it to be effective for other variational principles and updated methods as well.

In practice, one must take care in handling the transition between the initial "fixed-ω" and the final "adaptive-ω" stages of the optimization. When ω is changed, the nonlinear function being optimized is altered. One can imagine that a large alteration made abruptly at the end of the fixed-ω stage (Figure 2.2) could move the function's minimum far enough that the

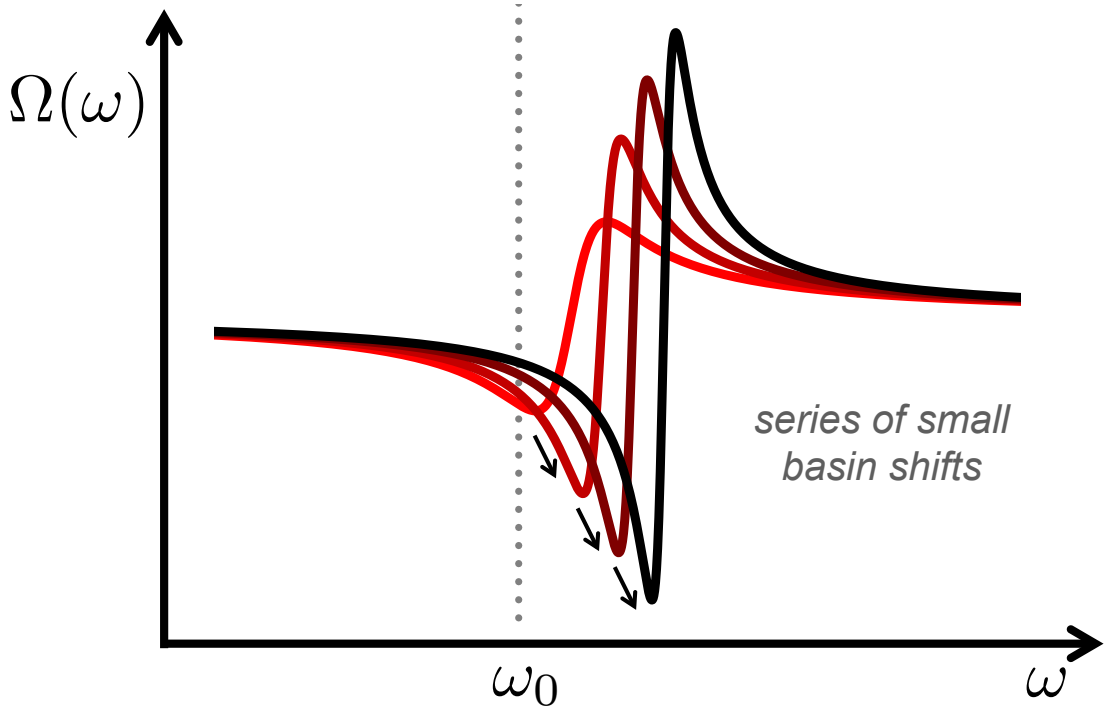


Figure 2.3: Example of a series of small basin shifts that occur during the transitional-ω stage of the optimization.

wave function variables were no longer within the basin of convergence for the chosen update method. Indeed, we have observed exactly this behavior in some tests involving Ω and the linear method update scheme. To avoid such pathologies, we instead add a "transitional-ω" stage to our optimization, in which ω is gradually interpolated between its initial fixed value and the value required to achieve variance minimization. In this way, the variational principle's minimum is moved only small steps at a time (Figure 2.3) to ensure that the wave function remains within its basin of convergence. In this study, we use the interpolation

$$\omega_j = \alpha_j \, \omega_0 + (1 - \alpha_j)(E_{j-1} - \sigma_{j-1}) \tag{2.27}$$

$$\alpha_j = \begin{cases} 1 & j \leq N_F \\ \frac{1}{N_T}(N_F + N_T - j) & N_F < j \leq N_F + N_T \\ 0 & j > N_F + N_T \end{cases} \tag{2.28}$$

in which j is the linear method iteration number and N_F and N_T are the number of iterations in the fixed-ω and transitional-ω stages, respectively. In the cases tested here, we find that N_F between 5 and 20 and N_T between 10 and 20 are effective choices.

In addition to ensuring both state selectivity and size consistency, this strategy improves the practical usability of excited state variational principles in VMC. Although the final wave function's energy is often not very sensitive to the choice of ω [256], there are likely to be cases where the user's choice of ω has a meaningful effect on the results. In previous work, ω has sometimes been adjusted by hand in order to minimize Ω. Although this does make the choice of ω unique, the process is tedious and prevents the overall methodology from achieving black box operation. With the adaptive approach described here, a user need only specify the initial value ω_0 so as to target the desired state.

2.4 Results

Computational Details

In the next few sections, we will present numerical results that complement our formal analysis. Results for CO and N_2 were obtained with our own Hilbert space VMC software in a STO-3G basis [288], with integrals imported from PySCF[289]. Bond distances were fixed at 1.19 Å and 1.18 Å for CO and N_2, respectively. For the formaldehyde-water system, the geometry was optimized to a local minimum (see Figure 2.6) using the ω-B97X-D density functional[177] and a 6-311G basis[290] set within QChem[291]. VMC results for this system were obtained with a development version of QMCPACK[282] with molecular orbitals and configuration interaction singles (CIS)[109] initial guesses imported from GAMESS.[292] Equation of motion coupled cluster with singles and doubles (EOM-CCSD) [15] results were obtained with MOLPRO.[293, 294] The VMC orbitals as well as the CIS and EOM-CCSD results for the formaldehyde-water system used the pseudopotentials of Burkatzki et al.[295], replacing core electrons for C and O atoms, and the corresponding valence double zeta (VDZ) basis set.

Comparison to Variance Minimization

In this section, we use the JAGP ansatz in Hilbert space to compare the results of our optimization scheme to those that are obtained by a simple minimization of σ^2. To begin, we apply both optimizations (with $\omega_0 = -109.00$ Hartrees, $N_F = 8$, and $N_T = 10$ for the adaptive-ω method) to the ground state of N_2, using the restricted Hartree-Fock (RHF) determinant with slightly randomized orbital coefficients for the JAGP initial guess. As seen in Figures 2.4a and 2.4b, the two methods produce the same values for E and σ^2 upon convergence. We also apply both optimization methods (with $\omega_0 = -111.30$ Hartrees, $N_F = 15$, and $N_T = 10$ for the adaptive-ω method) to the first excited singlet of CO. In this case, the initial guess for the JAGP pairing matrix was constructed by adding a HOMO-LUMO promotion and slight orbital coefficient randomization to the RHF ground state to produce a crude open-shell singlet representation. Figures 2.4c and 2.4d show that, starting

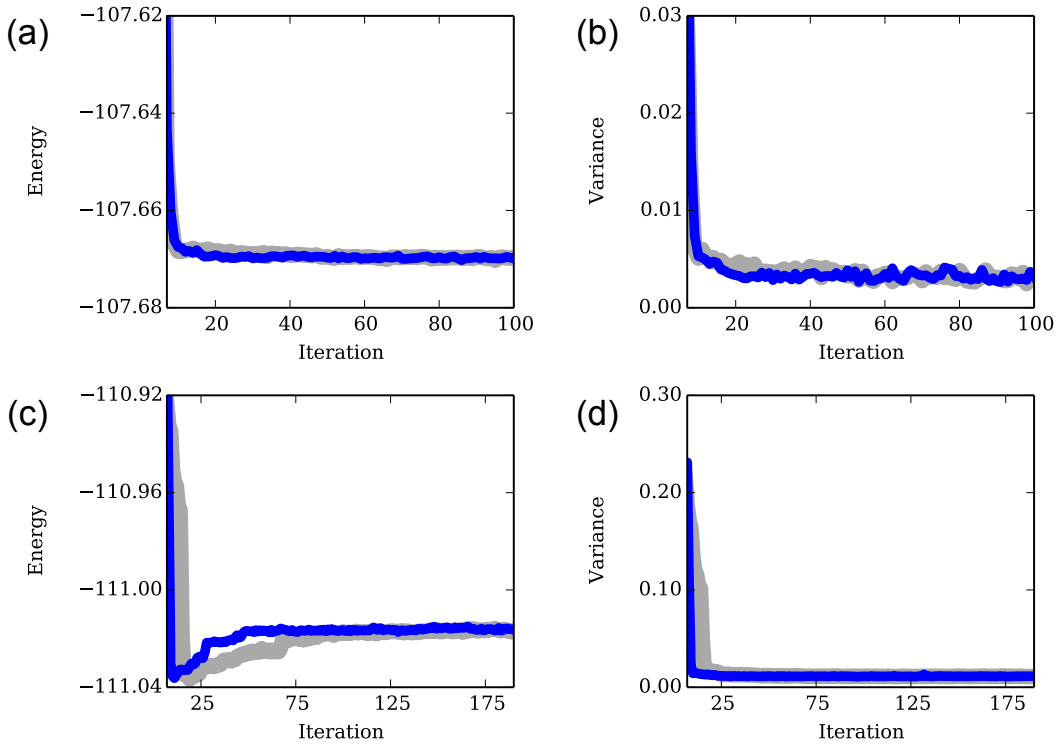


Figure 2.4: Ground state N_2 (a) energy and (b) variance and first excited singlet of CO (c) energy and (d) variance in a.u. with respect to the optimization step of σ^2 minimization (gray) and our adaptive-ω method (blue).

from this guess, both simple variance minimization and our Ω-based, adaptive-ω method converge to the same result, as expected.

To compare the state selectivity of our method with that of variance minimization, we have also performed a series of optimizations for CO in which the initial guess for the wave function was interpolated between a ground state guess and an excited state guess. Specifically, we used pairing matrix guesses of the form

$$\mathcal{M} = (1-\mu)\mathcal{M}^{(0)} + \mu\,\mathcal{M}^{(1)} \tag{2.29}$$

where $\mathcal{M}^{(0)}$ is the pairing matrix corresponding to the RHF ground state, $\mathcal{M}^{(1)}$ is the open-shell singlet pairing matrix resulting from a HOMO-LUMO promotion, and $\mu \in [0,1]$. As shown in Figure 2.5, simple variance minimization converged to the ground state when $\mu \leq 0.3$. In contrast, our Ω-based optimization (with $\omega_0 = -111.30$ Hartrees, $N_F = 20$, and $N_T = 10$) converged to the (targeted) excited state for all cases except $\mu = 0$. Indeed, we found that as little as 0.5% excited state character (i.e. $\mu = 0.005$) in the initial guess was

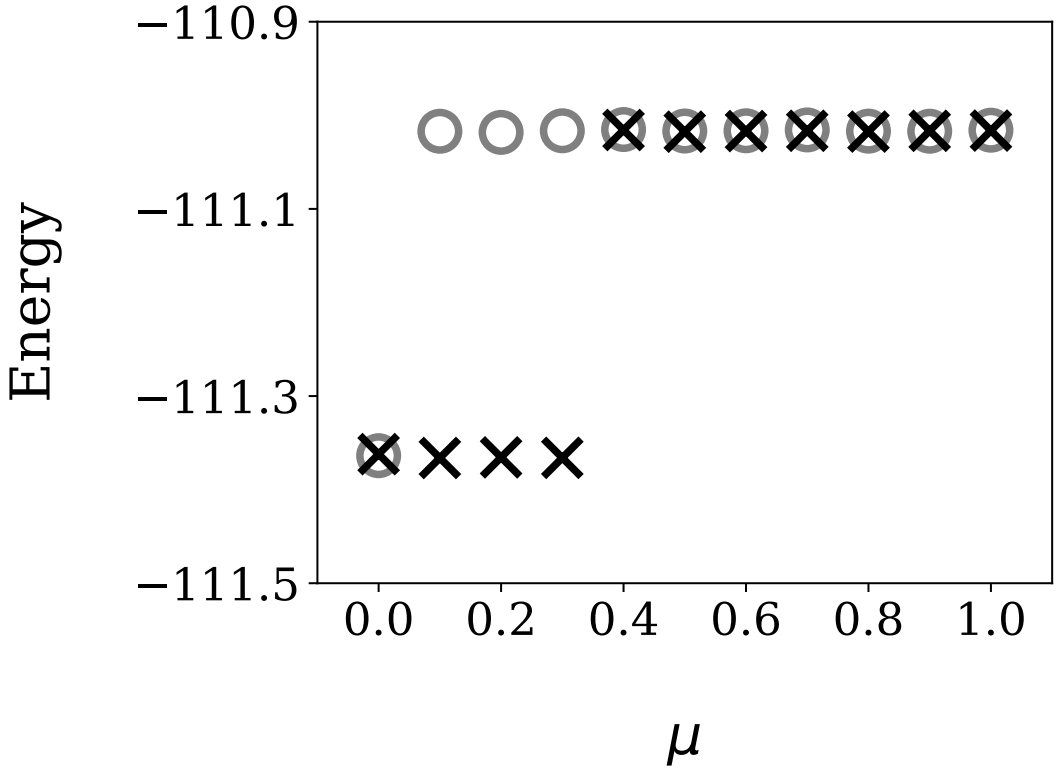


Figure 2.5: Energies of optimized JAGP wave functions for CO when starting from a range of different initial guesses that interpolate between a pure ground state guess ($\mu = 0$) and a pure excited state guess ($\mu = 1$). Results are shown for both σ^2 minimization (black X) and our adaptive-ω method (gray circles).

sufficient for our method to converge to the excited state, providing a clear example of the advantage offered by a state selective approach.

Size Consistency

As shown in Section 2.3, optimizing Ω with a particular, fixed value of ω can lead to size consistency issues when working with approximate wave functions. Of course, with exact wave functions, results will be ω-independent and size consistency will be achieved because in this limit, minimizing Ω will produce exact Hamiltonian eigenstates.[256] While it is not always easy to tell how far from this limit one is, one indication may be how sensitive the optimized wave function's energy is to the precise choice of ω. In the systems at hand, we find that N_2's ground state energy is quite insensitive to ω, varying by less than 10^{-3} Hartrees for

Table 2.1: Size consistency errors $|E_{CO+N_2} - E_{CO} - E_{N_2}|$ and their statistical uncertainties for the first singlet excited state of CO when combined with a far away nitrogen molecule.

Method	Error (mE_h)
fixed-ω	2.6 $\pm$ 0.2
adaptive-ω	0.04 $\pm$ 0.2

fixed-ω optimizations in which ω is set anywhere between -107.7 to -108.6 Hartrees. The first excited state of CO is more sensitive, with fixed-ω optimizations producing energies that change by as much as 0.01 Hartrees as ω is varied between -111.10 and -111.35 Hartrees.

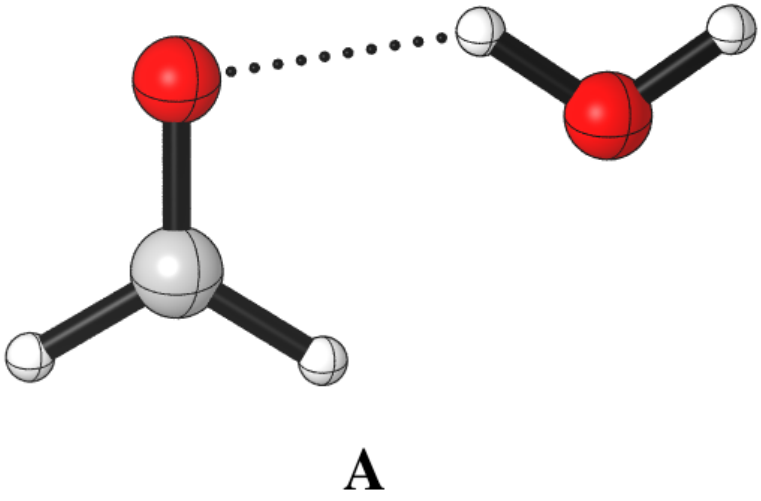
Given that the JAGP wave function approximation does produce nontrivial sensitivity to ω in at least one of these molecules, it is an interesting case in which to investigate size consistency. We should stress that, although it is an approximate ansatz, the Hilbert-space JAGP product factorizes and so will produce size consistent energies when paired with a size consistent variational principle [280]. Thus, any size consistency violation in its use can be linked to the variational principle. The test we perform is to optimize N_2's ground state and CO's first excited state, both separately and when the two molecules are treated together at a distance of 20 Å. This test provides a simple case in which we may ask whether CO's excited state is affected by the presence of a far away molecule.

When we minimize Ω with a fixed value of ω chosen in Hartrees as $\omega = \tilde{E} - 0.26$, with $\tilde{E}$ being the expected energy based on variance minimization results for the separate molecules, we find that the size consistency error, $E_{CO+N_2} - E_{CO} - E_{N_2}$, is over 2 milliHartrees. When instead we employ our adaptive-ω method, the results are size consistent to within our statistical uncertainty, as shown in Table 2.1. Thus, the ability to gradually transform the variational principle so that it is equivalent to variance minimization at convergence allows size consistent results to be achieved.

Formaldehyde and Water

To show a slightly more realistic example where having a size consistent optimization method matters, we turn to a hydrogen-bonded complex between formaldehyde and water, shown as complex **A** in Figure 2.6. To evaluate the first singlet excitation energy on the formaldehyde, we employ the recently-developed variation after response (VAR) approach [296, 297] as implemented in a development version of QMCPACK for a Slater determinant in real-space. In this context, VAR uses a finite-difference scheme to allow orbital optimization and a Jastrow factor to be applied variationally to a CIS-like linear response expansion [297]. By combining the adaptive-ω optimization we've presented here with VAR's ability to start from the output of a CIS calculation, we intend to show how VMC may deal with an excited state in a relatively black box fashion.

For the Jastrow-Slater ground state optimization, we chose $\omega_0 = -44.175$ to be several



A

Figure 2.6: Hydrogen-bonded formaldehyde-water complex. [298]

Hartrees below the RHF energy. As shown in Figure 2.7, this resulted in the energy increasing slightly during the transitional-ω phase of the optimization, as is to be expected when converting from a ground-state-energy-like variational principle to something more akin to variance minimization. For the VAR excited state optimization, we chose $\omega_0 = -41.175$ Hartrees, which is in between the values -40.8747 and -42.2812 that result for $E_{CIS} - \sigma$ when σ is taken from the optimized Jastrow-Slater ground state or the CIS wave function, respectively. For both states, we set $N_F = 10$ and $N_T = 20$.

As shown in Table 2.2, the VMC-based excitation energy agrees more closely with that of EOM-CCSD than with that of CIS. Presumably, this is due to the excited state orbital optimization lowering the excited state energy as compared to CIS, whereas even in RHF the ground state already enjoyed state specific orbital optimization and so its energy was lowered less by VMC. Although complex **A** is used here for the purposes of demonstration, the ability of VMC to produce a relatively accurate, nearly black box result for an excited state in the presence of a hydrogen-bonded solvent molecule is promising. Given QMC's low scaling compared to EOM-CCSD (N^4 versus N^6), it will be interesting to explore its prospects in larger and more technologically relevant examples of solvated photo-absorbers.

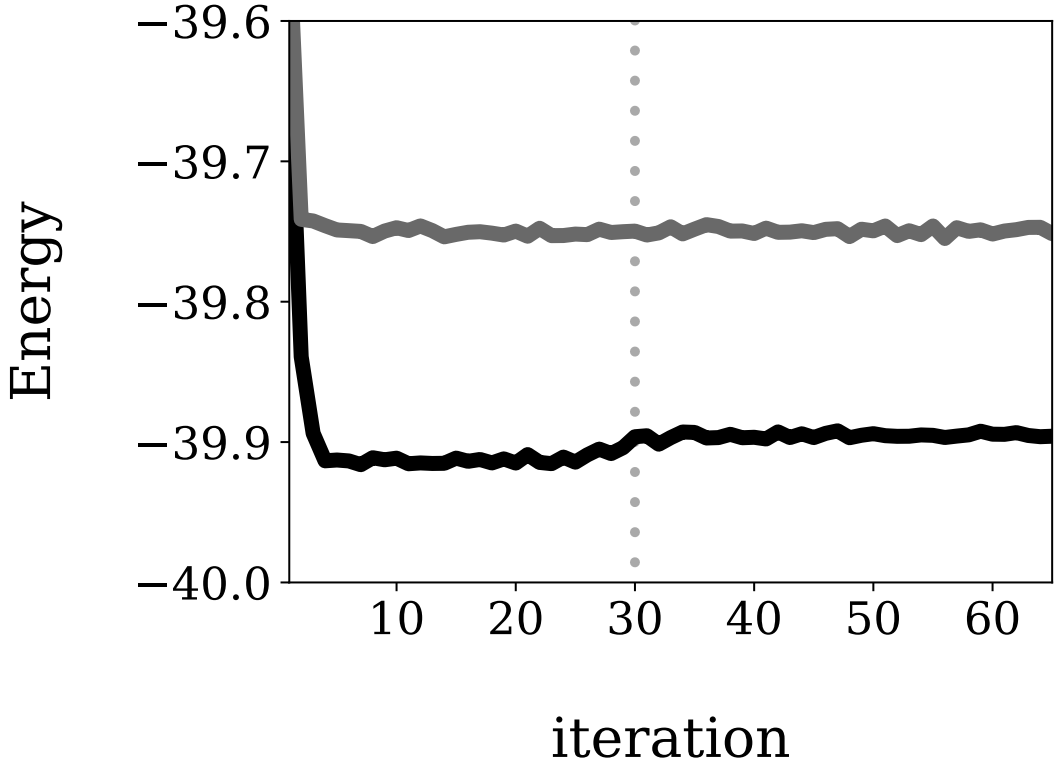


Figure 2.7: Energy versus iteration for the ground state (solid black) and excited state (solid gray) of **A**. The dotted line marks the end of the transitional-ω phase.

Table 2.2: Excitation energies in E_h for complex **A**.

Method	Excitation Energy
CIS	0.1718
EOM-CCSD	0.1511
VMC	0.1439 ± 0.0006

2.5 Conclusions

We have shown that size consistency is lacking in interior state selective variational principles that are analytic around their global minima and based on at most the second power of the Hamiltonian, a set we have denoted as $V_{1,2}$. In contrast, the well-established approach of variance minimization is known to be size consistent but not state selective. To achieve the best of both worlds, we have proposed a general optimization strategy that amalgamates variance minimization with a state selective variational principle from $V_{1,2}$. The approach is size consistent at convergence and maintains rigorous state selectivity at all stages. In an initial exploration with the Ω variational principle, we find that it is important that the transformation of the variational principle proceed gradually, lest the basin of convergence be moved away from the current wave function. We note that the overall strategy is applicable to a wide range of variational principles, including the W function employed recently by σ-SCF, and is readily compatible with the leading wave function optimizers in variational Monte Carlo.

Having demonstrated a viable path to variational, size consistent excited states, it is worth considering where such methods may be most useful in future. Thanks to QMC's ability to work with either open or periodic boundary conditions, the methodology should be equally applicable in both solids and molecules. Thus, in addition to its implications for modeling charge transfer excitations in the presence of explicit solvent, the approach may also prove useful in modeling defect-centered excitons. Regardless of the specific application, the removal of ω as a free parameter should make QMC-based excited state investigations more straightforward. In conjunction with recent variation-after-response developments that can build atop simple excited state quantum chemistry methods, the optimization approach presented here presages a more black box route to high-accuracy QMC results in a wide variety of excited state applications.

2.6 Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, as part of the Computational Materials Sciences Program and Center for Predictive Simulation of Functional Materials. Calculations were performed using the Berkeley Research Computing Savio cluster.

2.7 Appendix A

Here we show that for any state specific target function $\Gamma(E, \sigma^2)$ that is real analytic (i.e. can be written as a convergent power series with real coefficients) in a region around its global minimum, there exist system/ansatz pairs for which neither E nor σ^2 is stationary at the Γ minimum. We will do so by constructing a particular counterexample, although we suspect that other counterexamples exist. Consider a system in which three of the Hamiltonian

eigenvalues are $-b$, 1, and 2, with normalized eigenvectors

$$\hat{H}|\Phi_b\rangle = -b|\Phi_b\rangle, \quad \hat{H}|\Phi_1\rangle = |\Phi_1\rangle, \quad \text{and} \quad \hat{H}|\Phi_2\rangle = 2|\Phi_2\rangle. \tag{2.30}$$

For our approximate ansatz we choose the single-variable wave function

$$|\Psi(x)\rangle = x|\Phi_b\rangle + |\Phi_1\rangle + (b^2 + x)|\Phi_2\rangle, \tag{2.31}$$

in which x is allowed to take on real values. This defines a set of system/ansatz pairs in which we can control through b how closely the ansatz can come to an exact representation of the $|\Phi_1\rangle$ eigenstate, which we will take to be the state targeted by Γ.

The energy and variance of this ansatz can be written as

$$E = \frac{P}{D}, \tag{2.32}$$

$$\sigma^2 = \frac{QD - P^2}{D^2}, \tag{2.33}$$

in which we have used the three polynomials

$$P = 1 - bx^2 + 2(b^2 + x)^2, \tag{2.34}$$

$$Q = 1 + b^2x^2 + 4(b^2 + x)^2, \tag{2.35}$$

$$D = 1 + x^2 + (b^2 + x)^2. \tag{2.36}$$

Using the properties of the geometric series and the fact that D cannot be zero when $b > 0$, we note that both E and σ^2 are analytic functions of x and b so long as the point (x, b) is sufficiently close to $(0, 0)$. By inspecting the stationary points of E and σ^2, we will find that we can always choose b positive but small enough that these stationary points are distinct from the Γ minimum.

Begin with the stationary point for the energy, at which

$$\frac{\partial E}{\partial x} = 0, \tag{2.37}$$

which may be rearranged as

$$b\Big((2b + b^2)x^2 + (1 + 2b^3 + b^4)x - b \Big) = 0, \tag{2.38}$$

from which we see that the energy is always stationary when $b = 0$. When b is small but positive, we will have two roots, but only one of them,

$$x_E = b - 2b^3 + \mathcal{O}(b^4), \tag{2.39}$$

will occur near the origin. The other root,

$$x_F = -\frac{1}{2b} + \frac{1}{4} + \mathcal{O}(b), \tag{2.40}$$

will be far from the origin.

Moving on to the variance, we find that its stationary condition,

$$\frac{\partial \sigma^2}{\partial x} = 0 \tag{2.41}$$

can be rearranged into a cubic polynomial in x,

$$\begin{aligned} 0 =& 2b^2(1-b^4) + 2\big(2 + b(2 + b + 6b^4 + 2b^5 + 4b^7 + 4b^8 + b^9)\big)x \\ &+ 6b^2(1+b)^3\big(2 + b(b^2 + b - 2)\big)x^2 + 4\big(2 + 2b + b^4(2+b)^2\big)x^3. \end{aligned} \tag{2.42}$$

At small values of b, one can show that the discriminant of this polynomial is negative, implying that it has one real and two complex roots. As our ansatz does not admit complex values for x, the variance will have only one stationary point. Using the cubic formula and assuming b is small, this root can be found to be

$$x_V = -\frac{b^2}{2} + \frac{b^3}{2} - \frac{b^4}{4} + \mathcal{O}(b^6). \tag{2.43}$$

Having found the energy and variance stationary points, we now consider the target function as given in Eq. (2.14). For our particular system/ansatz choice, we can see that Γ will be a real analytic function of x and b when b is small and x is close to the global minimum, which we know occurs at $x = 0$ when b is chosen to be 0. This is because both E and σ^2 are analytic in x and b in this region, and by assumption Γ is real analytic near its global minimum. Furthermore, for small b and x, we have

$$\Delta = b\,(\,b^3 + 2bx - x^2 + \text{higher order terms}\,), \tag{2.44}$$

$$\sigma^2 = b^4 + 2b^2x + (2 + 2b + b^2)x^2 + \text{higher order terms}, \tag{2.45}$$

and so at $b = 0$, the leading order term in Γ will be an even power of x with degree 2 or higher. This makes sense as the minimum could not be $x = 0$ if the leading term were odd. Let us now express the stationary condition for Γ as

$$\frac{\partial \Gamma}{\partial x} = Z(x, b) = 0. \tag{2.46}$$

By the differentiability of analytic functions, Z will be real analytic and thus smooth in the region surrounding Γ's minimum. When $b = 0$, we know that Z will have an odd power of x with degree 1 or greater as its leading order term, and so the stationary point that is the global minimum will occur at $x = 0$ as expected. As Z is a smooth function of both x and b, this implies that when $b > 0$, the value of x that minimizes Γ can be forced arbitrarily close to zero by making b sufficiently small. Finally, note that $Z(x, 0)$ is a nonconstant function of x in the region of small x, as this is required for the global minimum of Γ to be unique. As Z is smooth in b, this implies that if we hold b fixed at a positive but small value, Z will still be a nonconstant function of x. By the principle of permanence, this function will have

isolated roots, and so there will be only one stationary point of Γ that approaches $x = 0$ as b becomes small. It now remains to show that b can be chosen positive but small enough so as to prevent this stationary point coinciding with either x_E or x_V, which are the only stationary points of the energy and variance that approach 0 as b becomes small.

First consider x_E. When b is small and $x = x_E$, we find that

$$\Delta = b^3 + b^4 + \mathcal{O}(b^5) \quad \text{and} \quad \sigma^2 = 2b^2 + 4b^3 + \mathcal{O}(b^4). \tag{2.47}$$

This implies that for $x = x_E$ and small but nonzero b, σ^2 will be small and nonzero. Furthermore, because x_E and x_V are distinct for nonzero b, σ^2 will not be stationary here. We will now deal with two cases for Γ separately. First, if the middle sum in Eq. (2.14) is absent, we have

$$\frac{\partial \Gamma}{\partial x} = \sum_{m=p}^{\infty} m\, a_{m0}\, (\sigma^2)^{m-1} \frac{\partial \sigma^2}{\partial x}, \tag{2.48}$$

which, based on what we know about σ^2, shows that Γ will not be stationary at $x = x_E$ if b is positive and sufficiently small. For the second case, in which the middle sum is present, we have to be more careful. Using the fact that σ^2 is not stationary at $x = x_E$, the stationary condition $\partial\Gamma/\partial x = 0$ can for this second case be written as

$$0 = \sum_{m=p}^{\infty} m\, a_{m0}\, (\sigma^2)^{m-1} + \sum_{m=q}^{\infty} \sum_{n=r}^{\infty} m\, a_{mn}\, (\sigma^2)^{m-1} \Delta^n. \tag{2.49}$$

If $p \leq q$, this condition cannot be satisfied at small but nonzero b as all the terms in the right hand sum will be higher order in b than the first term in the left hand sum, because $r > 0$ and Δ is order b^3. If instead $p > q$, we may divide through by $(\sigma^2)^{q-1}$ to obtain

$$0 = \sum_{m=p}^{\infty} m\, a_{m0}\, (\sigma^2)^{m-q} + \sum_{m=0}^{\infty} \sum_{n=r}^{\infty} (m+q) a_{(m+q)n}\, (\sigma^2)^{m} \Delta^n. \tag{2.50}$$

This equation is analytic in b and has a solution at $b = 0$. As nonconstant analytic functions have isolated roots, Eq. (2.50) cannot also have a solution at arbitrarily small but positive b unless the right hand side is zero for all b. We can show that this is not so by inspecting the leading order terms in b. Using Eq. (2.47), we see that if the lowest order terms from the two sums are to be the same order, we must have

$$2(p - q) = 3r, \tag{2.51}$$

which implies that there is a positive integer z such that

$$p - q = 3z \quad \text{and} \quad r = 2z. \tag{2.52}$$

If so, the leading order terms from Eq. (2.50)'s left hand sum will be proportional to

$$b^{6z} + 6zb^{6z+1} \tag{2.53}$$

while the leading order terms from its right hand sum will be proportional to

$$b^{6z} + 2zb^{6z+1}. \tag{2.54}$$

Thus, regardless of the values of p, q, and r, these two sums cannot cancel exactly and so Γ will not be stationary at $x = x_E$ when b is chosen to be positive and small.

We follow a similar analysis to show that Γ will not be stationary at $x = x_V$, where

$$\Delta = \frac{3b^5}{4} - \frac{b^7}{2} + \mathcal{O}(b^8) \quad \text{and} \quad \sigma^2 = \frac{b^4}{2} + \frac{b^5}{2} + \mathcal{O}(b^6). \tag{2.55}$$

As the energy is not stationary at $x = x_V$, the stationary condition $\partial\Gamma/\partial x = 0$ can be written as

$$0 = \sum_{m=q}^{\infty}\sum_{n=r}^{\infty} n\, a_{mn}\, (\sigma^2)^m \Delta^{n-1} + \sum_{n=s}^{\infty} n\, a_{0n}\, \Delta^{n-1}. \tag{2.56}$$

If the qr sum is not present, then clearly Γ will not be stationary at $x = x_V$ when b is positive and small. Otherwise, following the same logic we used for x_E, we now need to show that the right hand side of Eq. (2.56) is not zero for all b. If $s \leq r$, the leading order term in the right hand sum will be of a different order in b than that of the left hand sum, and so the right hand side will be nonzero for b small and positive. If instead $s > r$, we may divide through by Δ^{r-1} to obtain

$$0 = \sum_{m=q}^{\infty}\sum_{n=0}^{\infty} (n+r)\, a_{m(n+r)}\, (\sigma^2)^m \Delta^{n} + \sum_{n=s}^{\infty} n\, a_{0n}\, \Delta^{n-r}. \tag{2.57}$$

If the two sums are to cancel, so must their leading order terms. Using Eq. (2.55), we see that this can only occur if

$$4q = 5(s - r), \tag{2.58}$$

which implies that there is a positive integer y such that

$$q = 5y \quad \text{and} \quad s - r = 4y. \tag{2.59}$$

If q, r, and s have this relationship, we find that the left hand sum is proportional to

$$b^{20y} + 5yb^{20y+1} + \mathcal{O}(b^{20y+2}) \tag{2.60}$$

while the right hand sum is proportional to

$$b^{20y} - \frac{8y}{3} b^{20y+2} + \mathcal{O}(b^{20y+3}). \tag{2.61}$$

Thus, regardless of the values of q, r, and s, we see that Γ will not be stationary at $x = x_V$ when b is positive and small.

To conclude, we find that in this system/ansatz pairing, b can be chosen to be positive but small enough such that the minimum of Γ, which will approach $x = 0$ as b gets small, does not coincide with either the lone energy stationary point near $x = 0$ nor the lone variance stationary point near $x = 0$. We therefore conclude that for any $\Gamma \in V_{1,2}$ there exist system/ansatz pairs in which neither the energy nor the variance is stationary at the Γ minimum.

2.8 Appendix B

Here we provide two classes of system that are used in our proof of no size consistency. First, we define systems of type B in which we target the $E = 0$ eigenstate. Let this system have among its eigenstates the two states

$$\hat{H}|B_0\rangle = 0, \quad \hat{H}|B_1\rangle = |B_1\rangle. \tag{2.62}$$

Let the approximate wave function be

$$|\Psi\rangle = |B_0\rangle + \alpha|B_1\rangle \tag{2.63}$$

in which α is a nonzero real number. We then find that

$$\Delta_B = E - 0 = \frac{\alpha^2}{1+\alpha^2} \quad \text{and} \quad \sigma_B^2 = \frac{\alpha^2}{(1+\alpha^2)^2} > 0. \tag{2.64}$$

Second, we define systems of type C, for which $\Delta = 0$ and $\sigma^2 > 0$ when targeting its $E = 1$ eigenstate. Let this system have among its eigenstates the three states

$$\hat{H}|C_0\rangle = 0, \quad \hat{H}|C_1\rangle = |C_1\rangle, \quad \hat{H}|C_2\rangle = 2|C_2\rangle. \tag{2.65}$$

Let the approximate wave function be

$$|\Psi\rangle = \frac{\beta}{\sqrt{2}}|C_0\rangle + |C_1\rangle + \frac{\beta}{\sqrt{2}}|C_2\rangle, \tag{2.66}$$

in which β is a nonzero real number. We then find that

$$\Delta_C = E - 1 = 0 \quad \text{and} \quad \sigma_C^2 = \frac{\beta^2}{1+\beta^2} > 0. \tag{2.67}$$

Chapter 3

Excited State Mean Field Theory

3.1 Abstract

We present a mean field theory for excited states that is broadly analogous to ground state Hartree-Fock theory. Like Hartree-Fock, our approach is deterministic, state-specific, applies a variational principle to a minimally correlated ansatz, produces energy stationary points, relaxes the orbital basis, has a Fock-build cost-scaling, and can serve as the foundation for correlation methods such as perturbation theory and coupled cluster theory. To emphasize this last point, we pair our mean field approach with an excited state analogue of second order Møller-Plesset theory and demonstrate that in water, formaldehyde, neon, and stretched lithium fluoride, the resulting accuracy far exceeds that of configuration interaction singles and rivals that of equation of motion coupled cluster.

3.2 Introduction

In a nutshell, Hartree-Fock (HF) theory [2] applies the ground state variational principle to a wave function ansatz that includes only those correlations that are absolutely necessary to produce a qualitatively correct description of the electrons in a simple molecular ground state. Indeed, HF theory's Slater determinant hews as closely as possible to a classical mean field state, in which the particles would be completely uncorrelated, while accommodating the Pauli correlations that must be included when describing electrons. This simplicity keeps HF theory (relatively) affordable as it proceeds to minimize its variational principle and make the energy stationary with respect to changes in the molecular orbital basis. That HF theory has for decades been the central platform on which high-accuracy weak correlation treatments are built [268, 299] is a reminder of how valuable a qualitatively correct, minimally correlated wave function with relaxed orbitals can be. For strongly correlated ground states in which an unambiguous minimal active space exists, the story is much the same, but with complete active space self consistent field (CASSCF) [300–303] and its descendants [147, 304–310] replacing HF theory as the basic platform upon which weak correlation methods are built.

[156, 311–324] In this paper, we develop an affordable mean field platform for simple excited states while also providing an initial weak correlation treatment and a discussion of how the approach can be generalized to strongly correlated excitations.

3.3 Theory

Unlike the situation for ground states, even zeroth order descriptions of most excited states require more than Pauli correlations. For example, consider the open shell singlet of a simple HOMO→LUMO excitation. In this state, two opposite-spin electrons are strongly correlated with each other so as not to occupy the same orbital at the same time, implying that a minimally correlated excited state ansatz must incorporate correlations not present in a Slater determinant. An obvious candidate for this job is a configuration interaction singles (CIS) ansatz [325] which, to stay in line with ground state mean field methods like HF and CASSCF, has had its energy made stationary with respect to relaxations of the molecular orbital basis. While this direction has of course been explored before, previous approaches have tended to approximate the orbital relaxation in order to suppress cost. [128, 326–328] For example, Subotnik's OO-CIS approach [128] achieves its efficiency by employing an incomplete and HF-approximated Newton-Raphson optimization. Inspired by the need for fully-relaxed excited state orbitals in situations like charge transfer and core spectroscopy, we seek here to achieve them at the same cost-scaling as CIS by combining recent progress in excited state variational principles [116, 256, 329, 330] with a compound application of automatic differentiation. To demonstrate the efficacy of this approach as a mean field platform on which to build excited state correlation methods, we use it as the basis for an excited state analogue of second order Møller-Plesset (MP2) theory. [299]

As we are looking to describe an excited state, minimizing the energy

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{3.1}$$

in pursuit of the ground state variational principle will not guarantee convergence to the desired state. Instead, we will seek to minimize the Lagrangian

$$L = W + \vec{\mu} \cdot \frac{\partial E}{\partial \vec{\nu}} \tag{3.2}$$

in which the excited state variational principle [255]

$$W = \frac{\langle \Psi | (\omega - \hat{H})^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{3.3}$$

guarantees that a sufficiently flexible wave function will converge to the exact (excited) energy eigenstate with energy closest to ω while the Lagrange multipliers $\vec{\mu}$ guarantee that an energy stationary point with respect to the wave function variables $\vec{\nu}$ will be achieved even when

working with an approximate ansatz. While we could have done without W if we merely wished to ensure an energy stationary point was reached, its presence guarantees that in the limit of a sufficiently flexible wave function, the global minimum of L will be the stationary point corresponding to the desired excited state. Although we do not pursue it here, one can imagine augmenting W with functions of the dipole moment or other observables in order to differentiate between states that are energetically degenerate.

To make the minimization of this Lagrangian affordable, we must deal both with the difficulty of the $\hat{H}^2$ term and with the fact that derivatives of L with respect to the variational parameters $\vec{\nu}$ (e.g. the CI coefficients and the orbital rotation variables) lead to second derivatives of the energy. To avoid $\hat{H}^2$, we can resolve an identity in the basis of Slater determinants

$$W = \frac{1}{\langle\Psi|\Psi\rangle} \sum_I \langle\Psi|(\omega - \hat{H})|I\rangle\langle I|(\omega - \hat{H})|\Psi\rangle \tag{3.4}$$

and, following the approach of coupled cluster (CC) theory, [268] make an approximation in which we restrict our attention (and the range of the sum) to the most chemically relevant corner of Hilbert space, in this case the span of the closed shell "ground state" determinant and the singles excitations. Note that W need not be evaluated exactly for the theory to operate correctly, as its role is merely to guide the optimization into the correct stationary point. So long as W provides a sufficiently strong nudge to get us close, the Lagrange multiplier term in Eq. (3.2) will ensure convergence to the stationary point. This type of "nudged" convergence to a stationary point has been seen before in the application of full configuration interaction quantum Monte Carlo to excited states, [331] where a heavily approximated projection operator that could only serve to nudge the state propagation away from the lower states was sufficient to converge the imaginary time evolution onto excited states. When an approximate nudge is insufficient, our formulation allows for a systematic retreat to safety via increasingly accurate evaluations of W, but we stress that in our initial testing this has yet to prove necessary. Indeed, in all cases tested so far, the even more aggressive approximation

$$W \approx (\omega - E)^2 \tag{3.5}$$

leads the minimization of L to converge on to the same stationary point as when truncating the sum in Eq. (3.4). If this equivalence is maintained after more extensive testing in the future, there would be a strong simplicity argument in favor of employing Eq. (3.5).

Unlike the $\hat{H}^2$ terms, the challenge of second energy derivatives can be overcome by a compound use of automatic differentiation (AD), [332] which ensures that all of the derivatives of a many-input-single-output function can be evaluated for a small constant multiple of the cost of evaluating the function itself. As AD's use in quantum chemistry is still in its infancy, [281, 333, 334] let us briefly explain the principles so that its usefulness for our purposes is clear. First, consider that many complicated functions, such as the CIS energy, can be written as a graph in which each node is one of the four basic binary operations

Table 3.1: Expressions for the derivatives of the four basic arithmetic functions in terms of the (presumably stored) values of their inputs a and b and output f.

$f(a,b)$	$a+b$	$a-b$	ab	a/b
$\partial f/\partial a$	1	1	b	$1/b$
$\partial f/\partial b$	1	-1	a	$-f/b$

shown in Table 3.1 in which two input quantities a and b go in and the output quantity f that comes out may then become one of the inputs for one or more of the other nodes in the graph. One can compute the overall function value $g(\vec{x})$ by traversing the graph, starting from the dangling edges that are the inputs $\vec{x}$ and moving forward through all the nodes until the final output $g(\vec{x})$ is reached. Now, if one can afford to store the outputs of each node in the graph, it becomes possible to evaluate all of g's first derivatives with respect to the elements of $\vec{x}$ for a cost that is a small constant multiple of the evaluation cost of g via a sort of reverse traversal of the graph. [332] Crucially, if one considers a given node and assumes that they already know the partial derivative of g with respect to the output f of that node, then the chain rule and the simple derivative formulas in Table 3.1 ensure that the partial derivatives of g with respect to the node's inputs a and b can be evaluated via three or fewer binary arithmetic operations. Starting at the final node and working backwards, one finds that the number of operations required to get all the derivatives of g with respect to all the intermediates and all the elements of $\vec{x}$ is not worse than four times the number of nodes in the graph, and so the cost to get all the derivatives $\partial g/\partial \vec{x}$ is a small constant multiple of the cost to evaluate g itself.

This approach, known as reverse accumulation, provides a straightforward if tedious recipe for constructing a low cost implementation of analytic derivatives, and we may for example apply it to the CIS energy E to obtain an efficient function for the energy first derivatives that appear in Eq. (3.2). Folding this logic over on itself, we recognize that thanks to the dot product of these efficient derivatives with $\vec{\mu}$, L itself can be implemented as a many-input-single-output function of the variables $\vec{\nu}$ and $\vec{\mu}$ whose cost is a small constant multiple of that of E. By a second, compound application of reverse accumulation we may thus arrive at an implementation that delivers analytic derivatives of L with respect to both $\vec{\nu}$ and $\vec{\mu}$ for a cost that is also a small constant multiple of that of a CIS energy evaluation and with an additional memory requirement for intermediate storage that scales only as the square of the system size. Note that, in practice, setting up reverse accumulation is tedious work that can and has been automated in many software packages. In the present study, we have leveraged the machine learning community's rapid progress in reverse accumulation software by employing the TensorFlow [335] framework to evaluate our Lagrangian's derivative vector, the norm of which we minimize by a quasi-Newton approach. [336] As the cost of a CIS

energy evaluation scales as the cost of a Fock matrix build, so does evaluation of L and the necessary derivatives, leading our excited state mean field approach to have the same cost scaling as CIS.

While we have motivated this Lagrangian approach with the prospects of relaxing the orbitals in CIS, the logic supporting its construction is much more general, and indeed L can in principle be employed efficiently with any ansatz for which E and a reasonable approximation to W can be efficiently evaluated. L could, for example, be used as a more rigorous alternative to maximum overlap methods [220] when optimizing a CASSCF wave function for an individual excited state. Although we are quite curious about this possibility, we do not pursue strongly correlated excited state treatments in this study. Instead, we focus on delivering a fully orbital-relaxed CIS wave function and testing its ability to act as a platform for excited state correlation treatments in the same way HF theory does for ground states.

To this end, we employ the CIS-like ansatz

$$|\Phi\rangle = e^{\hat{X}} \left(c_0|0\rangle + \sum_{ia} c_{ia}|_i^a\rangle + \sum_{\bar{i}\bar{a}} c_{\bar{i}\bar{a}}|_{\bar{i}}^{\bar{a}}\rangle \right), \tag{3.6}$$

in which excitations are labeled by alpha (i,a) or beta ($\bar{i}$,$\bar{a}$) indices and the closed shell determinant $|0\rangle$ is included to help the orbital-relaxed excited state better maintain orthogonality to the RHF ground state. The vector of variables $\vec{\nu}$ that we optimize via L includes the coefficients c_0, c_{ia}, and $c_{\bar{i}\bar{a}}$, as well as the elements of the matrix X that defines the orbital rotation operator,

$$\hat{X} = \sum_{p<q} X_{pq} \left(\hat{a}_p^+ \hat{a}_q - \hat{a}_q^+ \hat{a}_p \right), \tag{3.7}$$

which for the present study we constrain so as to keep the orbitals spin-restricted. Although we do not explore the possibility in the present study, this approach could be generalized to work with a CASSCF wave function by replacing the CIS expansion in Eq. (3.6) with the CASSCF CI expansion. Given the much greater size of such CI expansions, it may in that case be more effective to use L only for the orbital rotation optimization and to instead rely on modern CI solvers to keep the energy stationary with respect to the CI coefficients.

While the excited state mean field (ESMF) ansatz $|\Phi\rangle$ is more flexible than CIS and might therefore be expected to be more accurate, one should remember that, due to the significant effect of weak correlation on energetics, HF theory itself is quite poor quantitatively even when it is qualitatively a good zeroth order wave function. By the same reasoning, total energies and energy differences from ESMF wave functions (which revert to RHF when ω targets the ground state) are not likely to be competitive in accuracy with methods that incorporate correlation effects, such as equation of motion coupled cluster singles and doubles (EOM-CCSD).[15]

As in the ground state, we are interested in the ESMF wave function not as a destination in and of itself, but rather as a reliable platform upon which to construct correlation treatments that can reasonably hope to achieve more quantitative accuracy. Considering first the

prospects for a coupled cluster theory, we note that due to the natural termination of the Baker-Campbell-Hausdorff (BCH) expansion in traditional, similarity-transformed CC singles and doubles (CCSD), [268] the usual approach of projecting the CC eigenvalue equation

$$(e^{-\hat{T}}\hat{H}e^{\hat{T}} - E)|\Phi\rangle = 0 \tag{3.8}$$

into the space of low order (in this case internally contracted) excitations from the reference will, as in the ground state theory, lead to a polynomially complex system of equations for the cluster amplitudes. However, due to the fact that bare triples excitations would be present within the internally contracted doubles, the cost scaling of such an approach would, although still polynomial, be substantially higher than in ground state CCSD. More enticing is the prospect of leveraging the fact that the potentially long-range nature of the excitation and the orbital relaxations it induces should already be accounted for in the ESMF reference state, allowing lower-cost local CC approaches [337–340] to focus on what they do best: treating short-ranged weak correlation.

As intriguing as CC methods may be, the principle of Occam's razor suggests that an analogue of the much simpler MP2 theory would be a wiser starting point for investigations into post-ESMF correlation methods. We can arrive at just such a method, which we will denote as ESMP2, by applying standard Rayleigh-Schrödinger perturbation theory to the zeroth order Hamiltonian

$$\hat{H}_0 = \hat{R}(\hat{F} - \hat{H})\hat{R} + \hat{P}\hat{H}\hat{P} + \hat{Q}\hat{F}\hat{Q} \tag{3.9}$$

where $\hat{F}$ is the Fock operator with respect to the ESMF one-body density matrix, $\hat{R} = |\Phi\rangle\langle\Phi|$, $\hat{P}$ is the projector to the span of $|0\rangle$ and the singles excitations in the ESMF orbital basis, and $\hat{Q} = 1 - \hat{P}$. Note that $\hat{P}|\Phi\rangle = |\Phi\rangle$ and $\hat{Q}|\Phi\rangle = 0$. This choice of $\hat{H}_0$ leads to the MP2-like zeroth order relationship

$$\hat{H}_0|\Phi\rangle = E_0|\Phi\rangle = \langle\Phi|\hat{F}|\Phi\rangle|\Phi\rangle \tag{3.10}$$

and indeed we see that when we set ω so as to target the ground state, $|\Phi\rangle$ becomes the RHF state and ESMP2 simplifies to MP2. In the excited state case, as in traditional MP2, the first order wave function contains no determinants with fewer than two excitations, which is a consequence of including the $\hat{P}\hat{H}\hat{P}$ term in $\hat{H}_0$. We also maintain the relationship $E_{\text{ESMF}} = E_0 + E_1$ in direct analogy to the MP2 relationship $E_{\text{HF}} = E_0 + E_1$. As the Fock operator is not diagonal for excited states, the first order amplitudes on doubles and triples are found by inverting $\hat{F} - E_0$ via the minimal residual Krylov subspace method, for which the MP2-style denominators are an excellent preconditioner. Note that this approach produces a fully excited-state-specific first order wave function, as opposed to the CIS(D) method where the triples are a product of the CIS coefficients and the ground state MP2 amplitudes. [129, 341]

While we have, for the sake of simplifying development and testing, written our pilot ESMP2 implementation in a fully uncontracted form whose triples part has an N^7 cost

Table 3.2: Comparisons for singlet excitations. For EOM-CC(2,3) we report excitation energies in eV, with other methods' results reported as excitation energy errors in eV relative to EOM-CC(2,3) and summarized in terms of mean unsigned error (UE) and maximum UE. For stretched LiF, transitions are relative to the closed shell ionic state and are labeled by the F $\rightarrow$ Li orbitals involved (the bond is aligned along the z axis). For other cases, transitions are relative to the ground state. Traditional methods were evaluated with Molpro [294] and QChem. [291]

State	EOM CC(2,3)	CIS	OO-CIS	CIS(D)	ESMF	ESMP2	EOM CCSD
H_2O, cc-pVDZ, $r = 0.9614$ Å, $a = 104.4°$							
$n \rightarrow \sigma^*$	8.22	0.96	-0.35	-0.18	-0.74	-0.01	-0.08
$n \rightarrow \pi^*$	10.25	0.70	-0.32	-0.11	-0.77	0.00	-0.06
$\sigma \rightarrow \sigma^*$	10.86	0.94	-0.26	-0.17	-0.73	-0.05	-0.06
$\sigma \rightarrow \pi^*$	12.93	0.65	-0.26	-0.10	-0.83	-0.03	-0.04
$\pi \rightarrow \sigma^*$	14.82	0.18	-0.47	-0.06	-0.82	-0.04	-0.01
CH_2O, cc-pVDZ, geometry: B3LYP/cc-pVTZ							
$n \rightarrow \pi^*$	4.22	0.40	-0.14	0.01	-0.98	-0.09	-0.05
$\pi \rightarrow \pi^*$	10.02	0.26	-0.30	0.46	-1.12	-0.06	0.12
$n \rightarrow \sigma^*$	8.70	1.68	0.50	-0.46	-0.71	0.09	-0.08
LiF, cc-pVDZ, $r = 8.0$Å							
$2p_x \rightarrow 2s$	-2.68	1.49	-3.03	-1.44	-1.02	0.10	-0.22
$2p_z \rightarrow 2s$	-2.68	1.44	-1.24	-1.31	-1.02	0.10	-0.22
$2p_z \rightarrow 2p_z$	-0.84	1.48	-3.20	-1.48	-1.02	0.10	-0.22
Ne, cc-pVTZ							
$2s \rightarrow 3p$	64.32	2.66	1.74	0.41	1.36	0.34	0.06
Mean UE		1.07	0.98	0.51	0.93	0.08	0.10
Max UE		2.66	3.20	1.48	1.36	0.34	0.22

scaling, multiple avenues exist for recovering the N^5 scaling of traditional MP2. On the one hand, we could exploit sparsity in the ESMF coefficient matrix, which we observe to have only a handful of elements that are not small. Indeed, we have tested this idea by setting all but three of the elements to zero when solving the ESMP2 linear equation and evaluating its second order energy for the molecules discussed below and found that excitation energy predictions are not strongly affected. On the other hand, we could follow the internally

contracted approach of Evangelista and coworkers. [342]

3.4 Results

Table 3.2 reveals that, like its HF cousin in the ground state, the ESMF approximation does not confer quantitative accuracy. This behavior can be understood as a direct consequence of our design goal of hewing as closely as possible to a classical mean field theory. By including only those correlations that are absolutely necessary to realize a fermionic excited state (namely Pauli correlations and the open shell correlation), ESMF is missing all weak correlations and so, like HF, does not produce quantitative energies. The fact that ESMF tends to underestimate excitation energies can also be understood in terms of what correlations are included. Indeed, in creating their open-shell arrangements, ESMF gives each excited state roughly one pair's worth of electron correlation, and so these states' energies are biased low compared to that of the closed shell ground state. Thus, in direct analogy to HF theory, the simplicity and mean field nature of ESMF prevents it from delivering accurate energetics on its own, but this of course was not the intention. What is more important is the question of whether ESMF can match HF theory's ability to act as a foundation for correlation methods, a question that our early results appear to answer strongly in the affirmative.

The data reveal that in water, formaldehyde, stretched lithium fluoride, and neon, ESMP2 rivals EOM-CCSD in accuracy. Its errors are typically at least a factor of five smaller than CIS, and it substantially outperforms CIS(D) in the charge transfer states of stretched LiF. Our primary explanation for this success is that the reference wave function's mean-field quality of having fully relaxed orbitals places it at a similar "distance" from the correct wave function as for HF in the ground state, with the subtle effects of weak electron correlation being all that is missing. In contrast, CIS(D), which is also inspired by and closely entwined with MP2 theory, is asking its perturbation to capture both correlation effects and orbital relaxation. While the former are typically small in systems that are not strongly correlated, the size of the latter is much more system dependent. Table 3.2 shows that in the low lying transitions of water, in which the overall spatial distribution of electrons is not greatly changed, both CIS(D) and ESMP2 are highly accurate. In stretched LiF, however, the transitions convert between ions and neutral atoms, and as one would expect these large charge density changes lead to strong orbital relaxation effects, as revealed by comparing the CIS and ESMF energies. These relaxations are much more difficult to treat perturbatively, and, to make matters worse, cause the closed-shell-state MP2 amplitudes that all CIS(D) states rely on to be less appropriate for the open-shell states. Looked at from this perspective, it is not surprising that ESMP2, thanks to its orbital-relaxed reference and fully state-specific perturbation, delivers more uniform accuracy across charge-transfer, valence, and Rydberg states alike. Indeed, when compared to the high-level benchmark of EOM-CC(2,3), [343] the maximum unsigned error of ESMP2 is less than the mean unsigned error for CIS(D).

We can also compare our results with more recent attempts to provide orbital relaxations for CIS. The OO-CIS method, for example, provides relaxations via a single Newton-Raphson

orbital optimization step in which the HF Hessian is used as an approximation to the CIS orbital Hessian. [128] As OO-CIS lacks a treatment for weak correlation, it is not surprising that in H_2O, CH_2O, and Ne, its accuracy is for most states better than CIS (which has neither orbital relaxation nor a weak correlation treatment) but worse than CIS(D), EOM-CCSD, and ESMP2, as shown in Table 3.2. The large errors that OO-CIS makes in LiF can be understood as a consequence of its Hessian approximation, as LiF's smallest HF Hessian eigenvalues are about an order of magnitude smaller than those in the other molecules, and so when this Hessian is inverted to get the OO-CIS orbital relaxation, the resulting Newton-Raphson step is much too large. In this system at least, it appears that the HF Hessian is not an effective surrogate for the CIS Hessian. Unlike OO-CIS and our approach, the variationally orbital-adapted CIS (VOA-CIS) method seeks to provide orbital relaxations through specially chosen expansions of the configuration interaction space. [327] However, there is at present no single prescription for defining the expansion (the developers explore at least eight possibilities in their initial paper [327]) and so instead of making an extensive and hard-to-interpret direct comparison, we will point out the developers' conclusion that "most of the time, VOA-CIS closely follows CIS(D)" and their data that shows that it is not unusual for VOA-CIS to be in error by between 0.5 eV and 1 eV for single excitation energies. [327] In summary, our preliminary testing shows ESMP2, with its inclusion of both orbital relaxation and state-specific correlation, to be closer in its behavior to EOM-CCSD than to previous attempts at augmenting the CIS wave function.

3.5 Conclusions

We have presented an excited state mean field theory and investigated its potential as a platform on which to build excited state correlation treatments. Like HF theory, the method relies on making a minimally correlated wave function's energy stationary with respect to orbital relaxations. While HF does this for a Slater determinant, we do so for a CIS-like wave function in order to accommodate the basic structure of simple excitations, and so the cost of our mean field optimization has the same scaling with system size as CIS. Unlike HF theory, our approach incorporates an excited state variational principle into its Lagrangian so that ground and excited states are treated equally. In exploring the method's potential as a platform for correlation treatments, we have constructed an excited state analogue of MP2 theory and found that, in initial tests, its accuracy rivals that of EOM-CCSD.

Looking to the future, there are a number of important questions to consider about this excited state mean field approach. First, how will it fare when applied to larger and more complicated charge transfer systems, core excitations, and Rydberg states? Second, will its advantages be maintained when paired with more complex and expensive methods such as coupled cluster and the complete active space self-consistent field? Finally, can this minimally-correlated, excited-state-specific wave function be usefully employed as a replacement for the Slater determinant in an excited-state-specific generalization of the Kohn-Sham approximation? We look forward to exploring these and other questions that will

undoubtedly arise in the context of excited state mean field theory.

3.6 Acknowledgements

We thank Joe Subotnik for many helpful discussions, and we acknowledge funding through the Early Career Research Program of the Office of Science, Office of Basic Energy Sciences, the U.S. Department of Energy, grant No. DE-SC0017869. Calculations were performed both on our own desktop computers and at the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Chapter 4

A Generalized Variational Principle

4.1 Abstract

We present a generalization of the variational principle that is compatible with any Hamiltonian eigenstate that can be specified uniquely by a list of properties. This variational principle appears to be compatible with a wide range of electronic structure methods, including mean-field theory, density functional theory, multi-reference theory, and quantum Monte Carlo. Like the standard variational principle, this generalized variational principle amounts to the optimization of a nonlinear function that, in the limit of an arbitrarily flexible wave function, has the desired Hamiltonian eigenstate as its global minimum. Unlike the standard variational principle, it can target excited states and select individual states in cases of degeneracy or near-degeneracy. As an initial demonstration of how this approach can be useful in practice, we employ it to improve the optimization efficiency of excited state mean field theory by an order of magnitude. With this improved optimization, we are able to demonstrate that the accuracy of the corresponding second-order perturbation theory rivals that of singles-and-doubles equation-of-motion coupled cluster in a substantially broader set of molecules than could be explored by our previous optimization methodology.

4.2 Introduction

While the ground state variational principle has acted as the cornerstone of electronic structure theory for decades, its usefulness is limited by its focus on the lowest Hamiltonian eigenstates. Certainly this reality has not prevented the development of powerful excited state methods based on other principles, such as linear response methods, or even methods based on the variational principle itself, such as state-averaging methods. However, these methods rely on making additional approximations beyond those required for the ground state theories from which they are derived. Linear response of course assumes that the excited states are in some sense close to the ground state in state space (specifically, it assumes that they live in the ground state's tangent space), [325] whereas state-averaging assumes

that important wave function traits such as the molecular orbitals are shared by all states. [131, 132, 344] In a huge variety of applications, these approaches have been successful. However, there remain important areas — such as charge transfer, core excitation, and doubly excited states — where these additional layers of approximation continue to impair predictive power and where it would be desirable to construct excited state methods that do not require them. [15, 165, 296, 345]

One route to doing so is to work with excited state variational principles, which can fully tailor the flexibility of an approximate wave function ansatz to the needs of an individual excited state. Typically based on functional forms that involve squaring the Hamiltonian operator, [116, 249, 255, 256] these approaches must either accept a higher computational scaling than their ground state counterparts [116] or resort to statistical evaluation [249, 256] or approximations to their functional forms. [117] These challenges in mind, it would be interesting if a class of exact excited state variational principles could be formulated without the need to square this difficult operator. In this paper, we present one such class, discuss its prospects for wide utility, and show that it can be used to improve the efficiency of excited state mean field (ESMF) theory. [117]

One seemingly inescapable difficulty with excited states and degenerate states is that they are harder to specify uniquely than non-degenerate ground states. Indeed, the latter can simply be specified by demanding the state of lowest energy, a prescription that is both straightforward and widely applicable. For excited states, defining the Hamiltonian eigenstate that one wants is much less straightforward. At the very least, one must say something more specific about it, such as where it is in the state ordering or what its properties are like. This specification may be relatively simple, such as specifying that one is interested in the Hamiltonian eigenstate with energy closest to a given value, but clearly must become more involved in cases with degeneracy or near-degeneracy. Here, we will take the perspective that Hamiltonian eigenstates whose unique specification requires making more precise statements about their properties be accommodated by crafting a generalized variational principle in which these more precise specifications can be encoded. For example, when dealing with degeneracy, uniquely specifying the desired stationary state might be accomplished by specifying desired values for both the energy and dipole moment. Even in cases that are not strictly degenerate, optimization may be easier if one can make statements about properties other than the energy that help differentiate the state from other energetically similar states. Crucially, however, these statements should do no more than identify the state, and so we will insist that the overall approach produce the same optimized wave function regardless of the details of what properties were used to uniquely identify it.

Although we will argue below that this generalized variational principle (GVP) can be employed in many areas of electronic structure and will point out parallels to recent work in density functional theory [200] and multi-reference theory, [132] we will in this study use ESMF theory as an example in which the approach offers clear practical benefits. In our previous study of ESMF, we coupled an approximate excited state variational principle to a Lagrangian constraint that ensured the optimizations produce energy stationary points. [117] While this approach allowed us to verify that ESMF theory could act as a powerful platform

for excited state correlation methods and helped to inspire the GVP that we introduce here, it possessed a number of difficulties. First, the method of Lagrange multipliers is, strictly speaking, a saddle-point method, [346] and so complications arise when coupling it to standard unconstrained quasi-Newton methods. Second, we have found that, in practice, the approach suffers from poor numerical conditioning and can take many thousands of iterations to converge, which offsets the advantages of Hartree-Fock cost scaling with an unusually high prefactor. Third, the original formulation was entirely based on the energy and thus not appropriate for cases with degeneracy. As we discuss below, all three of these difficulties can be addressed by optimizing the ESMF wave function with a GVP. The result is an order-of-magnitude speedup for ESMF wave function optimization, moving the method firmly into the regime where subsequent correlation methods, rather than the mean field starting point itself, dominate the cost of making predictions.

The proposed GVP appears to create a number of opportunities in other areas of electronic structure. For example, many other approaches exist for optimizing the orbitals of weakly correlated excited states, such as Δ-SCF [229] and the more recently introduced σ-SCF. [116, 347] In the case of Δ-SCF, the GVP's specification of desired properties is similar in spirit to the maximum-overlap approach, and in practice may be able to compliment it by helping avoid variational collapse. Using similar logic, it appears likely that the GVP will be immediately compatible with a very recent excited-state-specific variant of CASSCF that at present [132] employs the same type of Lagrange multiplier approach as in our original formulation of ESMF. Likewise, so long as the properties used to specify the desired state can be statistically estimated alongside the energy, it should be possible to realize a GVP approach within variational Monte Carlo. Finally, as we will discuss briefly below, the proposed GVP can be used to define a constrained search procedure similar to the ground state constrained search of Levy, [348] allowing formally exact density functionals to be defined for excited states that can be specified uniquely by a list of properties. This approach should allow a recent density functional extension of ESMF theory [200] to be reformulated and placed on firmer theoretical foundations.

This paper is organized as follows. First, we introduce the GVP, discuss why it should be applicable to a wide range of wave function approximations, and show that it may also have uses in density functional theory via a constrained search procedure. We then review the ESMF wave function ansatz before discussing both our original optimization approach and a more efficient approach based on a simple version of the GVP. We then present results showing the advantages of the newer optimization approach, use the improved optimizer to converge ESMF for a wider selection of molecules than was previously possible, and analyze the accuracy of the corresponding [117] perturbation theory (ESMP2). Finally, we provide a proof-of-principle example of how the use of properties other than the energy can assist in optimization and in the face of energetic degeneracy. We conclude with a summary of our findings and some thoughts on future directions.

4.3 Theory

A Generalized Variational Principle

Although rigorous excited state variational principles (i.e. functions whose global minimums are exact excited states) can be constructed by squaring the Hamiltonian operator, [255] we will take a different approach here as working with $\hat{H}^2$ is typically more difficult than working with expressions involving only a single power of $\hat{H}$. We begin by taking note of the fact that, when the ansatz is chosen as the exact (within the orbital basis) full configuration interaction (FCI) wave function, all of the Hamiltonian's eigenstates are energy stationary points and all of the energy stationary points are Hamiltonian eigenstates. This reality is made plain by simply constructing the FCI energy gradient with respect to the coefficient vector $\vec{c}$,

$$\nabla E = \frac{\partial E}{\partial \vec{c}} = \frac{2}{|\vec{c}|^2}\big(H - E\big)\vec{c}, \tag{4.1}$$

and recognizing that it is zero if and only if $\vec{c}$ satisfies the FCI eigenvalue equation. Thus, when attempting to construct a variational principle that yields an exact excited state in the limit of an exact (FCI) ansatz, it is sufficient to take an approach that searches only among energy stationary points and that, in the exact limit, is guaranteed to produce the specific desired stationary point. Of course, the investigator must know something about the state they want in order to ask for it specifically, and so the approach will need a formal mechanism for defining which stationary point (i.e. energy eigenstate) is being sought. Crucially, while working with energy stationary points does require differentiating the energy, and minimizing a function that is itself defined in terms of the energy gradient may require some second derivative information, the necessary derivatives do not require squaring the Hamiltonian and can usually be evaluated without increasing the cost scaling beyond what is already necessary for evaluating the energy in the first place. [117] Recognizing these formal advantages, we thus seek to define a generalized variational principle for ground, excited, and even degenerate states in which the energy gradient is the central player and a very general mechanism is provided for specifying which Hamiltonian eigenstate is being sought.

To begin, let us make an exact formal construction, after which we will discuss how this construction may be converted into a practical tool. First, take the wave function to be a linear combination (with coefficient vector $\vec{c}$) of all the N-electron Slater determinants that can be formed from the (finite) set of one-electron kinetic energy eigenstates that (a) satisfy the particle-in-a-box boundary conditions for a large box whose edges are length L and (b) have kinetic energy less than QL, where Q is a fixed positive constant. As the large box gets larger, this wave function ansatz will eventually be able to describe any normalizable Hamiltonian eigenstate to an arbitrarily high accuracy. Next, choose a set of operators $\hat{B}_i$ and their desired expectation values b_i and define a vector $\vec{d}$ of property deviations.

$$\vec{d} = \left\{ \left\langle \hat{B}_1 \right\rangle - b_1, \quad \left\langle \hat{B}_2 \right\rangle - b_2, \quad \dots \right\} \tag{4.2}$$

Now, if this vector uniquely specifies an exact Hamiltonian eigenstate, by which we mean that one such eigenstate produces a lower norm for this vector than any other Hamiltonian eigenstate, then that eigenstate will be the result of the following limit, which forms our generalized variational principle (GVP).

$$\lim_{L\to\infty}\lim_{\mu\to 0}\min_{\vec{c}}\left(\mu\left|\vec{d}\right|^2+(1-\mu)\left|\nabla E\right|^2\right) \tag{4.3}$$

Note the order of limits, in which we take the limit in $\mu > 0$ for each value of L as L is made progressively larger. For any finite L, the properties of the system will be finite regardless of the choice of $\vec{c}$, and so the largest possible norm for the vector $\vec{d}$ will also be finite. This implies that, as μ becomes small, the only states that stand a chance of being the minimum are energy stationary states. As the box gets bigger and bigger, the wave function approximation becomes exact, and the energy stationary states tend towards the exact Hamiltonian eigenstates. By assumption, one of these eigenstates has a lower value for the norm of $\vec{d}$ than the others, and so that eigenstate results from the minimization, as desired. While a non-degenerate ground state can be found via a property vector containing only the energy by setting the target energy to a very large negative number, one can instead seek excited states by choosing other target energies and furthermore can address degeneracies by adding additional properties.

Although this formal definition works in principle, let us now turn our attention to how it can be made useful in practice. First, we replace the large-box FCI wave function with an approximate wave function ansatz. This removes the limit in L, which was in any case merely a way of defining an explicit systematic approach towards an exact wave function. In its place, we now have the idea of systematic improvability that usually gets associated with variational principles: as the approximate ansatz becomes more and more flexible, we are guaranteed to recover the exact eigenstate eventually. Of course, as with the traditional variational principle, there will be systems for which the approximate wave functions we can afford to work with will not approach the exact limit closely enough to be useful. For example, the Hartree-Fock Slater determinant is known to break symmetry in unphysical ways in many situations upon energy minimization, [349] and we see no reason that similar qualitative failures should not occur for too-approximate wave functions when using a GVP. That said, the data provided below for ESMF theory suggest that there are many cases where even relatively simple wave functions are flexible enough to make the approach useful.

In terms of affordability, notice that we have not relied on squaring the Hamiltonian operator but have instead employed the norm of the energy gradient with respect to the wave function's variational parameters. If our wave function approximation allows for an affordable energy evaluation, then automatic differentiation guarantees that the energy gradient, as well as the gradient of the norm of the gradient needed to perform the minimization, [117] can be evaluated at a constant multiple of the cost for the energy. In the same way that we apply this approach to ESMF theory below, we expect that a related recent approach [132] to excited-state-specific CASSCF can also be reformulated in terms of this GVP. Of course, in practice, there may be faster ways of evaluating the necessary gradients than resorting to

automatic differentiation, but we at least have that option in principle and so the worst-case scenario for cost scaling should not be worse than the parent method. As one additional comment on practical minimization, note that the limit on μ will have to be discretized, but as we discuss below in our application to ESMF theory, this does not appear to create any significant difficulty.

At first glance, one might worry that, since the same eigenstate can be specified by many different property deviation vectors, the approach may give different results for different users who make different choices. However, due to the limit on μ, only energy stationary points will result from the minimization. So as long as the different property deviation vectors all specify the same stationary point, they will produce the same results (under the usual assumptions of nonlinear minimization methods not getting trapped in local minima). Of course, there will be cases where it is not clear what to specify, but in these cases it is not obvious that state-specific variational methods should be used at all, for how does one design an optimization to find a state about which nothing is known? If something is known, and if other states that the desired state could get confused with have been identified, then projections (even approximate ones) against those states can be included in the property deviation vector in an effort to uniquely specify the elusive state. Of course, such an approach has its limits, and in systems with very dense spectra, such as excitations inside a band of states in a solid, state-specific methods are hard to recommend.

Before moving on to combining this general approach with ESMF theory, let us make a short statement about density functional theory. For a very large box (take a box as large as you need to make the following as exact as you would like, e.g. set L to one kilometer and Q to one megajoule per kilometer) we define the density functional

$$G_\mu[\rho] = \min_{\vec{c}\to\rho} \left(\mu \left|\vec{d}\right|^2 + (1-\mu) \left|\nabla E\right|^2 \right) \tag{4.4}$$

in which we use Levy's constrained search [348] approach of searching over only those coefficient vectors $\vec{c}$ whose wave functions have density ρ. Minimizing this functional in the small-μ limit,

$$\lim_{\mu\to 0} \min_{\rho} \left(G_\mu[\rho] \right), \tag{4.5}$$

is then guaranteed to produce the exact density for the Hamiltonian eigenstate specified by the supplied property deviation vector. Although minimizing G by varying ρ is an entirely impractical way of finding the prescribed state's exact density, it is one way to do it, and we therefore see that exact density functionals (meaning functionals of the density that when minimized give the exact density) exist for states that can be specified uniquely by property deviation vectors. We make no attempt here to put precise bounds on how broad a class of states this is, but we expect that it contains the vast majority of molecular excited states that chemists have questions about. Certainly, any non-degenerate molecular bound state falls in this category as such states can be specified uniquely by their energy.

The ESMF Ansatz

Turning now to the construction of a practical optimization method for ESMF theory based on the above GVP, let us begin by reminding the reader how ESMF theory defines its wave function approximation.

$$|\Psi\rangle = e^{\hat{X}} \left(c_0|\Phi\rangle + \sum_{ia} \sigma_{ia} \hat{a}^\dagger_{a\uparrow} \hat{a}_{i\uparrow} |\Phi\rangle + \tau_{ia} \hat{a}^\dagger_{a\downarrow} \hat{a}_{i\downarrow} |\Phi\rangle \right) \tag{4.6}$$

Here $|\Phi\rangle$ is the Restricted Hartree Fock (RHF) solution and is included to help maintain orthogonality between the excited state and the reference ground state, and the coefficients σ_{ia} and τ_{ia} correspond to excitations of an up-spin and down-spin electron, respectively, from the i-th occupied orbital to the a-th virtual orbital. In a finite basis set of N_{bas} spatial orbitals, the operator $\hat{X}$ is given by

$$\hat{X} = \sum_{p<q}^{N_{\text{bas}}} X_{pq} \left(\hat{a}^\dagger_p \hat{a}_q - \hat{a}^\dagger_q \hat{a}_p \right), \tag{4.7}$$

in which $\hat{X}$ is real and restricted to be the same for up- and down-spin spin-orbitals. These choices ensure that the orbital relaxation operator given by $e^{\hat{X}}$ is unitary[350] and spin-restricted. As only the elements above the diagonal of $\hat{X}$ matter here, the number of variational parameters is reduced from N^2_{bas} as in Fig. 4.1(a) to $N_{\text{bas}}(N_{\text{bas}}-1)/2$ as in Fig. 4.1(b) . Furthermore, as noted by Van Voorhis and Head-Gordon, the energy of this type of singly excited wave function is invariant to rotations between occupied orbitals and to rotations between virtual orbitals.[264] Therefore, these energy-invariant rotation parameters lead to redundancy in the parameters such that two wave functions could have different values of $\vec{x}$, c_0, $\vec{\sigma}$, and $\vec{\tau}$, but have the same energy. As this would further complicate any numerical optimization strategy, we only allow for rotations between occupied and virtual orbitals, making X an off-diagonal block matrix (see Fig. 4.1(c)), reducing the number of variational rotation parameters to $N_{\text{occ}}N_{\text{vir}}$, and redefining $\hat{X}$

$$\hat{X} = \sum_{i}^{N_{\text{occ}}} \sum_{a}^{N_{\text{vir}}} X_{ia} \left(\hat{a}^\dagger_a \hat{a}_i - \hat{a}^\dagger_i \hat{a}_a \right). \tag{4.8}$$

Accounting for all of the variables in the ESMF wave function, for a system with a closed-shell ground state with N_{elec} electrons in a basis of N_{bas} orbitals, there are $N_{\text{occ}} = N_{\text{elec}}/2$ occupied orbitals, and $N_{\text{vir}} = N_{\text{bas}} - N_{\text{occ}}$ virtual orbitals. The CIS-like coefficient vector $\vec{c} = \{c_0, \vec{\sigma}, \vec{\tau}\}$ has $(1 + 2N_{\text{occ}}N_{\text{vir}})$ elements; the orbital rotation parameter vector $\vec{x}$ has $N_{\text{occ}}N_{\text{vir}}$ elements; thus, the ESMF wave function $\Psi(\vec{\nu})$, where $\vec{\nu} = \{\vec{c}, \vec{x}\}$, has a total of $3N_{\text{occ}}N_{\text{vir}} + 1$ variables.

Despite being composed of multiple determinants, we assert that the ESMF wave function hews closely to the principle of a mean-field theory, as it retains only the minimum

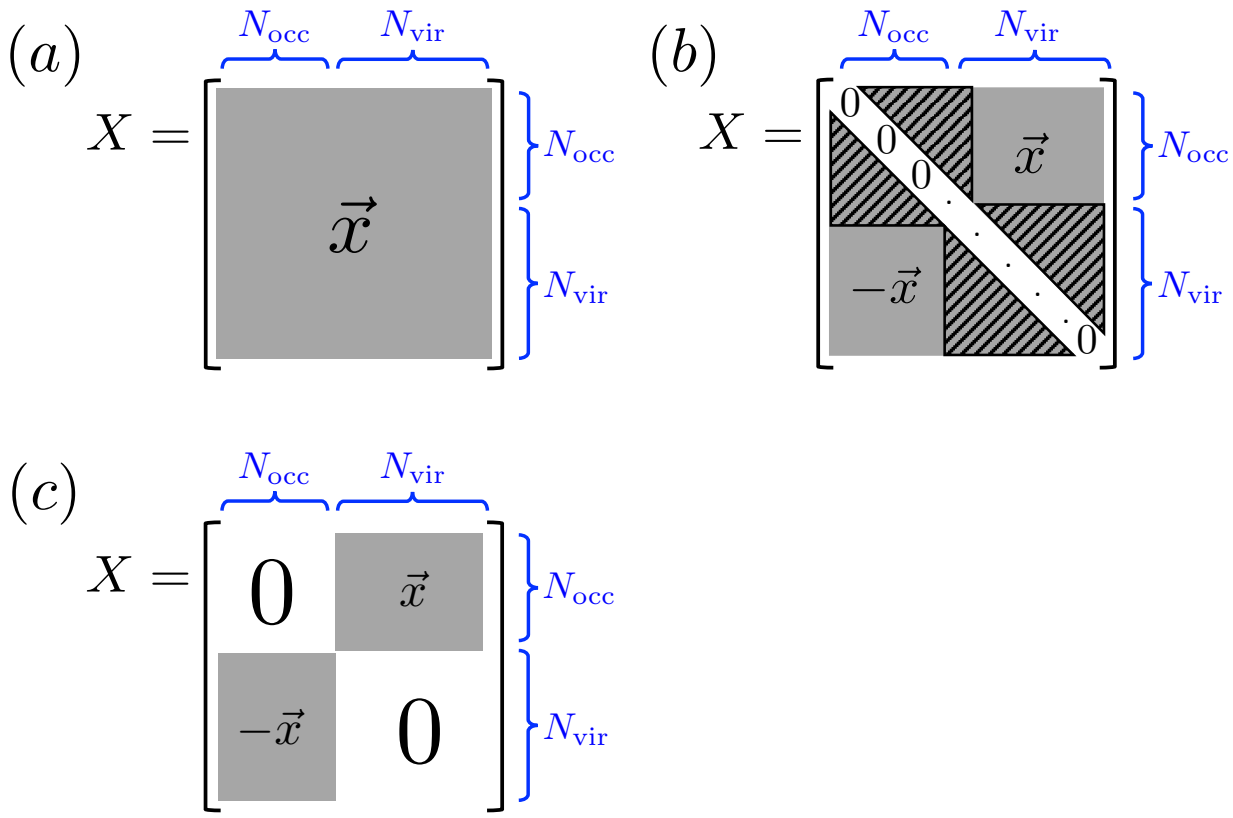


Figure 4.1: Gray areas represent non-zero variational parameters in the orbital rotation space. The black-striped triangles in (b) identify energy-invariant parameters. The structure of the rotation coefficient matrix with no assumptions is shown in (a), using the property of anti-hermiticity in (b), and using the property of invariance under occupied-occupied and virtual-virtual rotations in (c).

correlations necessary to describe an open shell excited state of a fermionic system. Like the mean field Hartree Fock wave function, the ESMF wave function captures Pauli exclusion via the antisymmetry of Slater determinants. However, more correlation is needed to describe an open shell excited state. Not only can the electrons in the open-shell arrangement not occupy the same *spin* orbital, these opposite-spin electrons cannot occupy the same *spatial* orbital, which is a strong correlation not present in a closed-shell ground state that we insist on capturing here, which immediately requires at least two Slater determinants. As including the entire set of single excitations keeps the approach general by ensuring that any such open-shell correlation can be captured and does not increase the cost scaling compared to using a single pair of determinants, we opt for the wave function above as our ansatz. We do note, however, that a two-determinant ansatz would likely be effective in many cases, although making that simplification has the potential to complicate the corresponding per-

turbation theory, a point we will return to in discussing our results. Either way, double and triple excitations — where most of the weak correlation effects that would move us squarely away from the mean-field spirit are to be found — are not included.

More accurate than CIS due to the added flexibility from the orbital rotation operator, [325] yet less accurate than methods like EOM-CCSD that incorporate more electron correlation, [15] ESMF is best seen as a gateway towards quantitatively accurate descriptions of excited states rather than an accurate method in its own right. Like Hartree Fock theory, [268, 351] its ultimate utility is intended as a platform upon which useful correlation methods can be built. So far, three such methods have been investigated: 1) a DFT-inspired extension to ESMF whose preliminary testing [200] reveals a valence-excitation accuracy similar to that of TD-DFT but also the promise of significant advantages in charge transfer states, 2) an excited state analog of Moller-Plesset perturbation theory, ESMP2,[117] which we show below to be highly competitive in accuracy with EOM-CCSD, and 3) a state-specific complete active space self-consistent field (SS-CASSCF) approach whose orbital optimization mirrors that of ESMF and whose root-tracking approach is similar in spirit to the GVP discussed here. [132]

While our original strategy for optimizing the ESMF ansatz achieved the same $\mathcal{O}(N^4)$ scaling as ground state mean-field theory, the actual cost of the optimization was unacceptably high. In many cases, it was more expensive than working with our fully-uncontracted version of ESMP2, whose cost scaling goes as $\mathcal{O}(N^7)$. In searching for more practical optimization methods, we have considered the option of deriving Roothaan-like equations, but so far this approach has not yielded a practical optimization strategy. Instead, we have found two other approaches to be more effective, at least for now. First, one can replace the BFGS [259–262] minimization of our original Lagrangian with an efficient Newton-Raphson (NR) approach, which handles the strong couplings between ansatz variables and Lagrange multipliers more effectively. Second, one can use the GVP introduced above to redefine the optimization target function in a way that avoids Lagrange multipliers entirely, at which point both BFGS and NR are considerably accelerated. In the next section, we will discuss the old and new target functions, after which we turn our attention to comparing the relative merits of BFGS and NR and how ESMF admits a useful finite-difference approach to the latter.

Target Functions

Lagrange Multiplier Formalism

While Hartree-Fock theory uses Lagrange multipliers to enforce orthonormality between the orbitals,[2] our original target function for optimizing the ESMF ansatz used Lagrange multipliers to ensure that the optimization ended on an energy stationary point even when we approximated an $\hat{H}^2$-based variational principle to keep it affordable. By minimizing

$$L_{\vec{\lambda}} = W + \vec{\lambda} \cdot \nabla E, \tag{4.9}$$

in which W is an approximated excited state variational principle and $\vec{\lambda}$ are the Lagrange multipliers, we guarantee that the optimization preserves the useful properties of energy stationary points, such as size-consistency for a product-factorizable ansatz. Note that, even if W is not approximated, such properties can be violated in the absence of the Lagrangian constraint. [330] Note also that, since we rotate from already-orthonormal Hartree Fock orbitals, we do not need to add additional Lagrange multipliers for orthonormality. In practice, we approximated the excited state variational principle

$$W = \frac{\langle\Psi|(\omega - \hat{H})^2|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{4.10}$$

whose global minimum is the exact (excited) energy eigenstate with energy closest to ω.[247, 255] Somewhat surprisingly, we found that even the aggressive approximation

$$W \approx (\omega - E)^2 \tag{4.11}$$

was sufficient in practice. [117]

While this formalism allows us to produce a number of successful optimizations in small molecules and achieves the desired cost scaling, it does create multiple complications. In particular, this Lagrangian is unbounded from below with respect to the Lagrange multipliers, so simple descent methods like standard BFGS cannot be applied directly. Instead, we proceeded by minimizing the squared norm of the gradient of $L_{\vec{\lambda}}$, i.e. $|\nabla L_{\vec{\lambda}}|^2$, and although this does not increase the cost scaling, it leads to an additional layer of automatic differentiation that increases the cost prefactor. Worse yet, as we will make clear below, this strategy was very poorly numerically conditioned, causing BFGS to require a large number of iterations to converge. In contrast, a NR algorithm can directly search for and locate saddle points of this Lagrangian target function, which are the solutions we actually seek, thus avoiding the need for an extra layer of derivatives.[346] Furthermore, NR helps with the speed of convergence due to its more robust handling of second-derivative couplings. However, whether working with NR or BFGS, we find it even more effective to avoid Lagrange multipliers entirely by reformulating the target function using the GVP.

Generalized Variational Principle Approach

Consider instead an optimization target function that can be switched between the energy itself and a simple version of the GVP in which the energy is the only property in the deviation vector.

$$L_{\mu\chi} = \chi\Big(\mu(\omega - E)^2 + (1-\mu)|\nabla E|^2\Big) + (1-\chi)E \tag{4.12}$$

If we first consider the case where we set $\chi = 1$, we see that we have a simple version of Eq. (4.3) in which the limit on L has been replaced by using the approximate ESMF ansatz. This target function is bounded from below, and so a series of optimizations in which μ is

made progressively smaller and then set to zero can immediately employ either BFGS or NR. Once we are close to convergence, we can in the case of NR then switch χ to zero and rely on NR's ability to hone in on an energy saddle point. Of course, in practice, it may be system specific when and if switching χ to zero is advantageous. If one does not switch χ to zero, then it is important to realize that it is possible for an optimization to end at a point where $\nabla|\nabla E|^2 = 0$ but $|\nabla E|^2 \neq 0$, so the user must be careful to verify at the end of the optimization that the energy gradient is indeed zero as expected. In the results below, we did not encounter any optimizations converging to a stationary point of $|\nabla E|^2$ that was not a stationary point of E, but such points clearly exist and so this simple check should be standard procedure.

While the use of $(\omega - E)^2$ was seen in our previous approach as an approximation to an excited state variational principle, the GVP approach helps us see that this is not where the approximation lies. Indeed, for non-degenerate states, this simple choice for the deviation vector will give exact results when used with an exact ansatz. From this perspective, it is clear that the approximation being made is instead an ansatz approximation. To be precise, the assumption is that the stationary points of the ESMF ansatz are, for the states we seek, similar to those of FCI, which is the same assumption that is made when formulating the Roothaan equations to find an energy stationary point of the Slater determinant in Hartree-Fock theory. Thus, as in the ground state case, the central assumption is that the relevant energy stationary point of the mean-field ansatz is a good approximation for the exact Hamiltonian eigenstate. Energy minimization (for the ground state) or the use of the GVP (for any state) are simply means of arriving at the relevant stationary point.

To give a concrete idea of how this target function may be used in practice, let us sketch out how we might use it to find a particular stationary point (point $\boldsymbol{c}$) in Fig. 4.2. Initially, we set $\chi = 1$ and $\mu = \frac{1}{2}$ and select an ω value near to where we expect the stationary point's energy to be. We then take a series of NR steps seeking to minimize $L_{\mu\chi}$, but in between each step we decrease μ by some small amount, say $\frac{1}{10}$, until we reach $\mu = 0$. At this point, the GVP is assumed to have done its job (locating the neighborhood of the correct stationary point) and we switch χ to 0, thus converting the optimization into a standard saddle point search (if this proves unstable then χ can be held at 1 instead). The potential advantage of this $\chi = 0$ stage is that it can employ well-known preconditioning methods when solving the NR linear equation, such as the Hartree-Fock energy hessian approximation

$$M_{ia,jb} = \frac{\delta_{ij}\delta_{ab}}{2(F_{aa} - F_{ii} + \Delta E)} \tag{4.13}$$

that is often used to accelerate ground state optimizations. [264, 352–356] Here F is the Fock matrix, and ΔE is the change in energy between the last two optimization steps. In summary, what starts as a minimization method guided to the desired stationary point by the GVP ends with a straightforward energy saddle point search. Compared to the Lagrange multiplier approach, this approach has the advantage that there are no poorly-preconditioned Lagrange multipliers and the objective function is bounded from below. Thus, most standard minimization methods can be used without the need for an additional layer of derivatives,

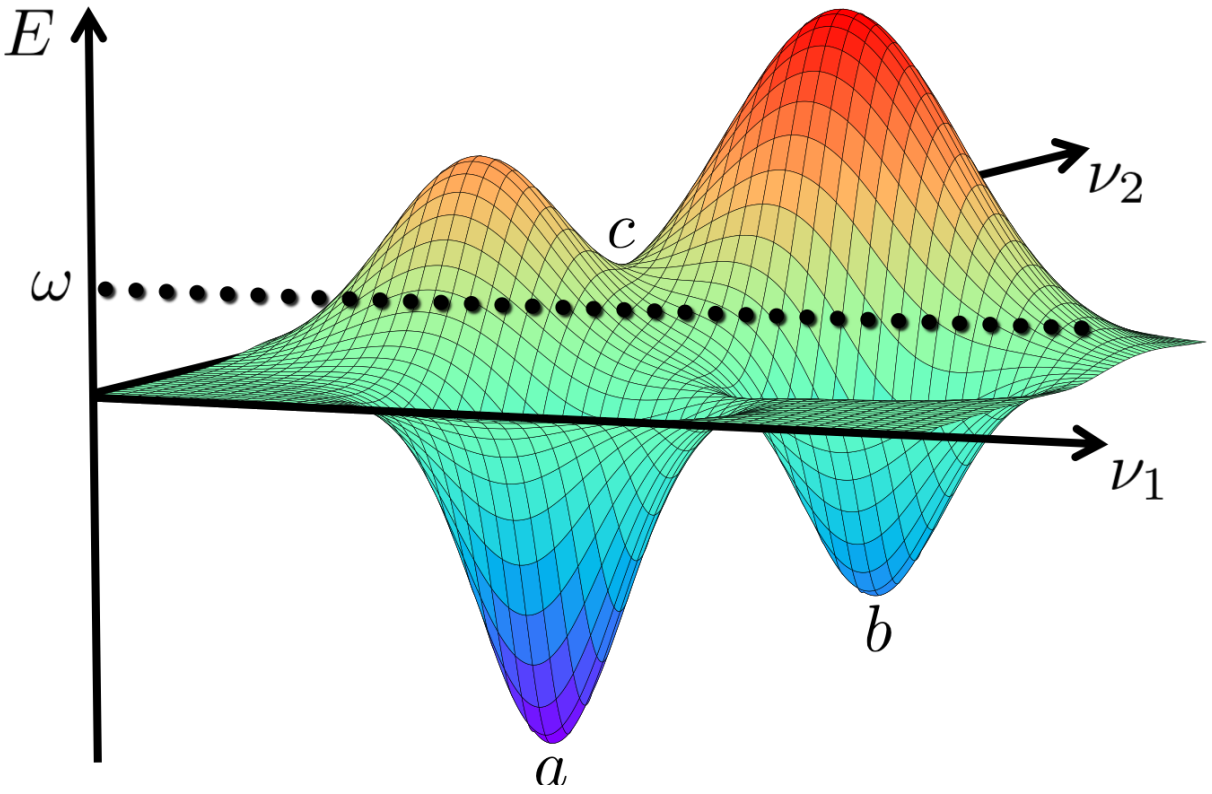


Figure 4.2: An idealized representation of the relationship between two variables, ν_1 and ν_2, in the ESMF wave function and the ESMF energy in a model system. The global minimum at **a** is the lowest energy ground state of the system, while the two stationary points at **b** and **c** correspond to excited states of the system, and can be individually resolved using the state-targeting parameter ω which is shown by the dotted line.

and the target function can convert into a form that is easily preconditioned near the end of the optimization.

Finite-Difference NR

With the ability to evaluate gradients of both $L_{\vec{\lambda}}$ and $L_{\mu\chi}$ comes the option of employing a finite-difference approximation [357] to the NR method (FDNR) that allows us to avoid the expensive construction of Hessian matrices. To begin, let us briefly review the standard NR method in order to orient the reader and set notation. In attempting to optimize L with respect to the variables $\vec{\nu}$, the NR approach approximates changes in L to second order in

the yet-to-be-determined update to the current variable vector $\Delta\vec{\nu}$.

$$L(\vec{\nu}+\Delta\vec{\nu}) \approx L(\vec{\nu}) + \nabla L \cdot \Delta\vec{\nu} + \frac{1}{2}\Delta\vec{\nu}^T \cdot \mathbf{H} \cdot \Delta\vec{\nu} \tag{4.14}$$

Here, the Hessian is $\mathbf{H}_{ij} \equiv \partial^2 L/\partial\nu_i\partial\nu_j$. Setting the gradient of this approximation (w.r.t. $\Delta\vec{\nu}$) to zero and solving for $\Delta\vec{\nu}$ gives an estimate for the variable change that would lead to L being stationary.

$$\mathbf{H} \cdot \Delta\vec{\nu} \approx -\nabla L \tag{4.15}$$

If $\mathbf{H}$ can be explicitly constructed and inverted, then one simply does so in order to solve for the update, but in many situations (including ours) the explicit construction of the Hessian is prohibitively expensive.

Instead, we follow Pearlmutter [357] and solve Eq. (4.15) via a Krylov subspace method (we use GMRES[358]) in which the matrix-vector product is formed efficiently via a finite-difference of gradients. Noting that the key matrix-vector product is

$$[\,\mathbf{H} \cdot \Delta\vec{\nu}\,]_i = \sum_j \frac{\partial^2 L}{\partial\nu_i\partial\nu_j}\Delta\nu_j \tag{4.16}$$

we compare to the differentiation of a first-order Taylor expansion of L.

$$L(\vec{\nu}+\Delta\vec{\nu}) \approx L(\vec{\nu}) + \sum_j \frac{\partial L(\vec{\nu})}{\partial\nu_j}\Delta\nu_j \tag{4.17}$$

$$\frac{\partial L(\vec{\nu}+\Delta\vec{\nu})}{\partial\nu_i} \approx \frac{\partial L(\vec{\nu})}{\partial\nu_i} + \sum_j \frac{\partial^2 L(\vec{\nu})}{\partial\nu_i\partial\nu_j}\Delta\nu_j \tag{4.18}$$

Combining Eqs. (4.16) and (4.18), one arrives at a simple approximation for the matrix-vector product in the form of a gradient difference.

$$[\,\mathbf{H} \cdot \Delta\vec{\nu}\,]_i \approx \frac{\partial L(\vec{\nu}+\Delta\vec{\nu})}{\partial\nu_i} - \frac{\partial L(\vec{\nu})}{\partial\nu_i} \tag{4.19}$$

Note that, if $\Delta\vec{\nu}$ is not small enough to justify the Taylor expansion, we can exploit the linearity of the matrix-vector product to make the finite-difference more accurate by scaling the vector down and then scaling the resulting matrix-vector product back up.

$$\mathbf{H} \cdot \Delta\vec{\nu} \approx \frac{1}{\epsilon}\Big(\nabla L(\vec{\nu}+\epsilon\Delta\vec{\nu}) - \nabla L(\vec{\nu})\Big) \tag{4.20}$$

In a given FDNR iteration, $\nabla L(\vec{\nu})$ is evaluated once and stored, so that each additional finite-difference estimate of a matrix-vector product requires only a single additional gradient evaluation. Although in principle an even more accurate finite-difference can be achieved at the cost of two gradients per matrix-vector multiply via the symmetric finite-difference formula, we have not found this to be advantageous. As our results below demonstrate, the simpler one-gradient approach is already a very accurate approximation.

4.4 Computational Details

Calculations in this work include timing comparisons between the previously discussed ESMF optimization methods and vertical excitation benchmarks of ESMF and ESMP2 against a range of single-reference excited state methods. Generally, generating initial guesses for an ESMF calculation is straightforward and does not necessitate any post-HF calculations. An initial RHF calculation computes the orthonormal HF orbitals which are used unrotated as the initial ESMF orbitals, i.e. $X = 0$, and therefore the initial unitary transformation matrix is simply the identity matrix. Often, the dominant configuration state function (CSF) in an excitation is sufficient as the initial guess for $\vec{c}$. For example, using the notation from Eq. (4.6), the initial guess for a singlet excitation from the i-th to the a-th orbital is simply $|\Psi\rangle = \frac{1}{\sqrt{2}}\mathbf{I}(\hat{a}^\dagger_{a\uparrow}\hat{a}_{i\uparrow}|\Phi\rangle + \hat{a}^\dagger_{a\downarrow}\hat{a}_{i\downarrow}|\Phi\rangle)$. Choosing ω is system dependent and requires some intuition, but even a rough approximation to the excitation energy from experiment, TD-DFT, or a CIS calculation is often sufficient because the energy stationary point criteria ensures that we will recover a size-consistent solution regardless of our choice of ω. If a rough approximation of the excited state energy is not feasible, one can slowly increase ω over the course of several optimizations and identify an entire spectrum of excited states. In this specific survey, the former approximation process was used to generate the majority of initial ESMF guesses. For a few systems, however, the initial guesses were generated from the most important single excitation coefficients of the EOM-CCSD wave function, and ω was calculated using

$$\omega = E_{RHF} + \Delta E_{EOM-CCSD} - \frac{0.5}{27.211} \tag{4.21}$$

where E_{RHF} is the RHF ground state energy and $\Delta E_{EOM-CCSD}$ is the EOM-CCSD vertical excitation energy. The energy correction is included since ESMF tends to underpredict excitation energies by approximately 0.5 eV relative to high level benchmarks. These few initial guesses built from EOM-CCSD results were only necessary to confirm that both methods were describing the same excitation and ensure fair comparisons and evaluations between them. The ground state references for ESMF and ESMP2 excitation energies are RHF and MP2, respectively.

Timing data reported in this work was produced on one 24-core node of the Berkeley Research Computing Savio cluster. Note that this timing data is meant to compare between different optimization methods for ESMF that use the same Python-based Fock build code, and are not intended to represent production level timings. Work is underway on a low-level implementation that exploits a faster Fock build, but this is not the focus of the current study. For FDNR-$L_{\mu\chi}$ timing data reported in Sec. 4.5, μ was set to $\frac{1}{2}$ in the first iteration, $\frac{1}{4}$ in the second iteration, and 0 in all subsequent iterations, and χ was switched from 1 to 0 after the first 10 iterations. All calculations were completed under the frozen core approximation and most in the cc-pVDZ basis set;[359, 360] exceptions to the latter include the Hessian data reported in Fig. 4.7 and the LiH system in Sec. 4.5 which used the minimal STO-3G basis set,[288, 361] and CIS(D) benchmark data which employed

Table 4.1: For different molecules and stages in a NR optimization in the cc-pVDZ basis, we show the number of orbitals in the basis set N_{bas}, the time it takes to construct $\mathbf{H}_{\text{ex}}$ and then apply it to $\nabla L_{\mu\chi}$ one hundred times, the time it takes to estimate the same one hundred matrix-vector products via Eq. (4.20), and the average relative error associated with Eq. (4.20).

Molecule	N_{bas}	NR iter.	$\mathbf{H}_{\text{ex}}$ time (min)	$\mathbf{H}_{fd}$ time (min)	relative error
Water	23	0	0.40	0.39	2.4×10^{-6}
		5	0.40	0.39	6.0×10^{-6}
		10	0.40	0.39	6.0×10^{-6}
Ammonia	28	0	0.97	0.79	2.4×10^{-6}
		10	1.00	0.82	6.1×10^{-6}
		30	1.03	0.93	6.1×10^{-6}
Formaldehyde	36	0	4.52	2.08	2.8×10^{-6}
		10	4.53	2.08	6.5×10^{-6}
		18	4.53	2.08	6.5×10^{-6}
Methanimine	41	0	9.02	3.62	2.6×10^{-6}
		10	9.22	3.81	6.5×10^{-6}
		30	9.38	3.77	6.5×10^{-6}

the rimp2-cc-pVDZ auxiliary basis set.[362] With MOLPRO version 2019.1,[294, 363] we optimized the geometries of a set of small organic molecules at the B3LYP/6-31G* level of theory. [209, 364–366] Explicit geometry coordinates and the main CSFs contributing to each excitation are provided in the Supporting Information (SI). Moving to the cc-PVDZ basis set, we then performed ground state Restricted Hartree-Fock calculations to compute the initial orbitals used in ESMF, and CIS, MP2, and EOM-CCSD calculations for later benchmarking. CIS(D), TD-DFT/B3LYP, and TD-DFT/ωB97X-V [367] excitation energies were computed with Q-Chem version 5.2.0[368], and δ-CR-EOM-CC(2,3)[165] calculations were performed with GAMESS.[292] As some theories may lead to different orbital energies and thus orbital orderings, for each theory that used a different molecular orbital basis, we used Molden[369, 370] to plot and compare the main orbitals involved in each excitation to ensure we were comparing the same state between theories.

4.5 Results

Assessing the Finite-Difference Approximation

While the numerical efficiency offered by the FDNR approach is welcome, its overall applicability depends on the accuracy of its finite-difference approximation. To quantify its accuracy we computed the exact Hessian of $L_{\mu\chi}$, $\mathbf{H}_{\text{ex}}$, with Tensorflow's automatic differentiation software [335] for water, ammonia, formaldehyde, and methanimine, at the beginning, middle, and end of the optimization. We then constructed a "finite-difference Hessian" of $L_{\mu\chi}$, $\mathbf{H}_{\text{fd}}$, by applying Eq. (4.20) to the columns of the identity matrix. In Table 4.1, we report 1) the number of orbitals in these systems with the cc-pVDZ basis set, 2) the time in minutes required to build $\mathbf{H}_{\text{ex}}$ once and compute the matrix-vector product $\mathbf{H}_{\text{ex}}\nabla L_{\mu\chi}$ one hundred times, 3) the time in minutes required to compute $\mathbf{H}_{\text{fd}}\nabla L_{\mu\chi}$ using Eq. (4.20) one hundred times, and 4) the relative error between the results of applying either $\mathbf{H}_{\text{fd}}$ or $\mathbf{H}_{\text{ex}}$ to $L_{\mu\chi}$, i.e. $\left|\mathbf{H}_{\text{fd}}\nabla L - \mathbf{H}_{\text{ex}}\nabla L\right|/|\mathbf{H}_{\text{ex}}\nabla L|$. The data demonstrate that the finite-difference approach to applying the Hessian matrix has much more favorable scaling than building and applying the exact Hessian. In fact, the cost of building the exact Hessian scales so rapidly that, for cyclopropene (59 orbitals in the cc-pVDZ basis), $\mathbf{H}_{\text{ex}}$ could not be computed in under two hours on a NERSC Cori Haswell node. Additionally, the relative errors assure us that the finite-difference approximation is highly accurate, and so we have used it in all of the iterative FDNR optimizations discussed below.

Comparing Optimization Strategies

We now turn to the question of which strategy is most efficient when optimizing the ESMF wave function. Our key findings here are that the GVP-based $L_{\mu\chi}$ objective function leads to faster optimization than does $L_{\vec{\lambda}}$ and that, once using $L_{\mu\chi}$, the efficiencies of the BFGS and FDNR methods become system-dependent but similar. Figures 4.3-4.6 show four examples of this trend, which we have observed across all of the systems we have tested. In each of these four examples, roughly one order of magnitude in speed is gained by moving to the $L_{\mu\chi}$ objective function.

To understand why the $L_{\vec{\lambda}}$ objective function is less efficient, it is useful to analyze the Lagrange multipliers. At convergence, the energy will be stationary, which in turn implies that the Lagrange multiplier values will all be zero.

$$\frac{\partial L_{\vec{\lambda}}}{\partial \vec{\nu}} = \frac{\partial}{\partial \vec{\nu}}\left(W + \vec{\lambda}\cdot\frac{\partial E}{\partial \vec{\nu}}\right) = \frac{\partial W}{\partial \vec{\nu}} + \mathbf{H}_E\vec{\lambda} = 0 \tag{4.22}$$

$$\Rightarrow \quad \mathbf{H}_E\vec{\lambda} = 2(\omega - E)\frac{\partial E}{\partial \vec{\nu}} = 0 \tag{4.23}$$

$$\Rightarrow \quad \vec{\lambda} = \vec{0} \tag{4.24}$$

Thus, we have guessed $\vec{\lambda} = 0$ in our optimizations. However, as can be seen in the lower panels of Figures 4.3-4.6, the optimization moves the Lagrange multipliers significantly away

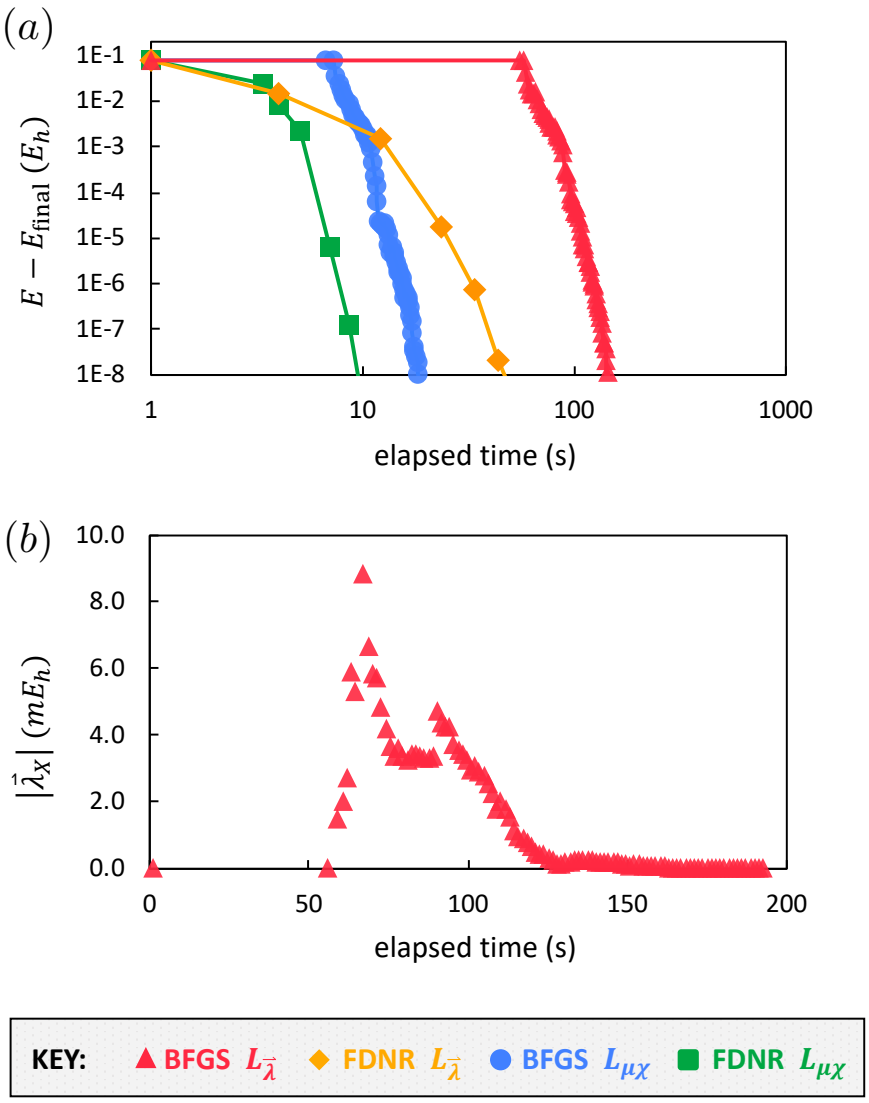


Figure 4.3: Two plots are shown for different optimizations of the HOMO-LUMO singlet excitation in cc-pVDZ water. In (a), the difference between the current and converged excited state energy in Hartree is compared to the elapsed optimization time for the BFGS-$L_{\vec{\lambda}}$ (red triangles), FDNR-$L_{\vec{\lambda}}$ (orange diamonds), BFGS-$L_{\mu\chi}$ (blue circles), and FDNR-$L_{\mu\chi}$ (green squares) optimization strategies. In (b), the norm (in millihartrees) of the Lagrange multipliers associated with the orbital rotation parameters is shown over the course of an BFGS-$L_{\vec{\lambda}}$ optimization. Note that the elapsed time is plotted on a log scale in (a) but on a linear scale in (b) and that each marker represents one iteration in the associated algorithm.

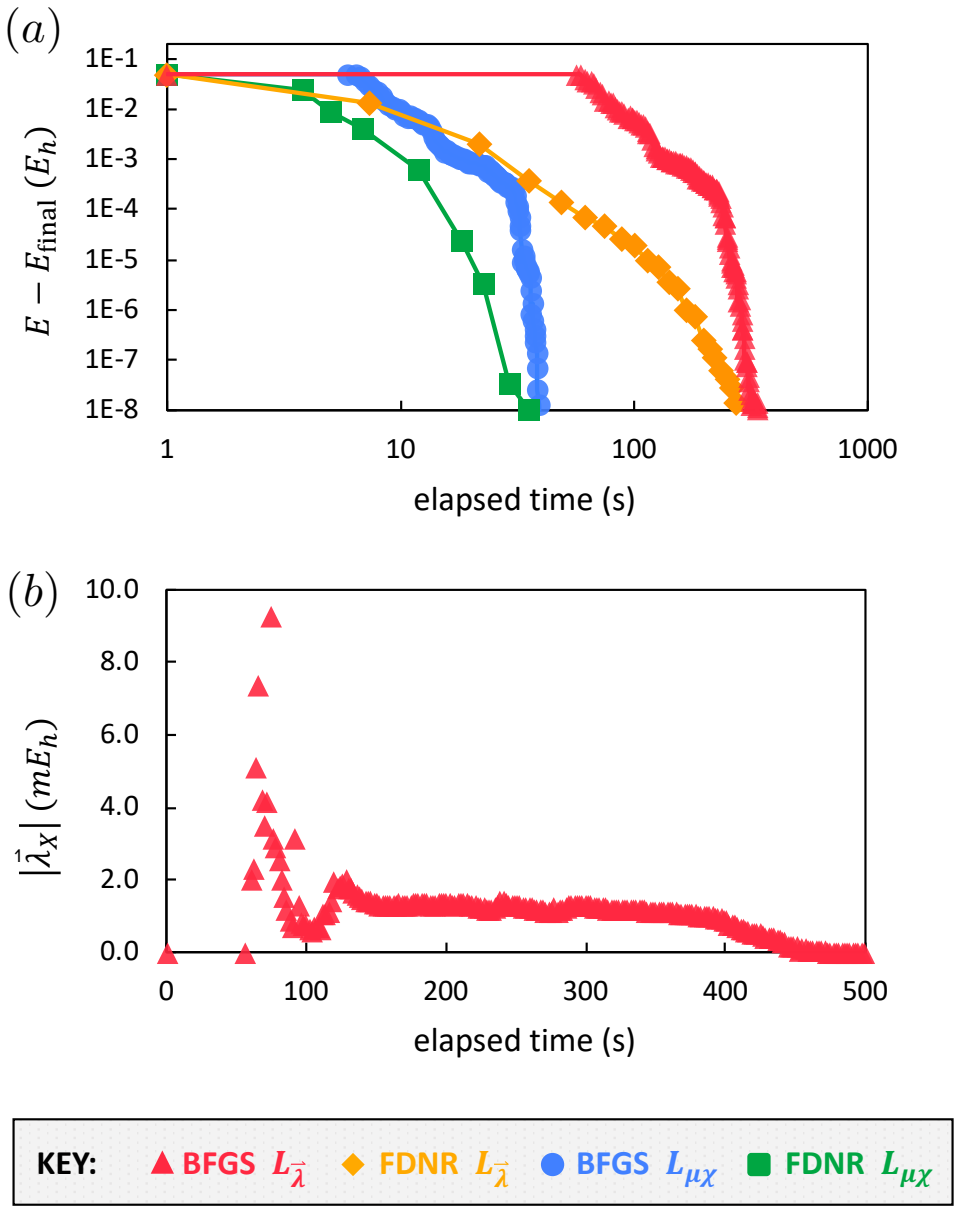


Figure 4.4: For the HOMO-LUMO singlet excitation in cc-pVDZ carbon monoxide, the difference between the current and converged excited state energy in Hartree is compared across optimization strategies in (a) and the norm (in millihartrees) of the Lagrange multipliers associated with the orbital rotation parameters is shown over the course of an BFGS-$L_{\vec{\lambda}}$ optimization in (b).

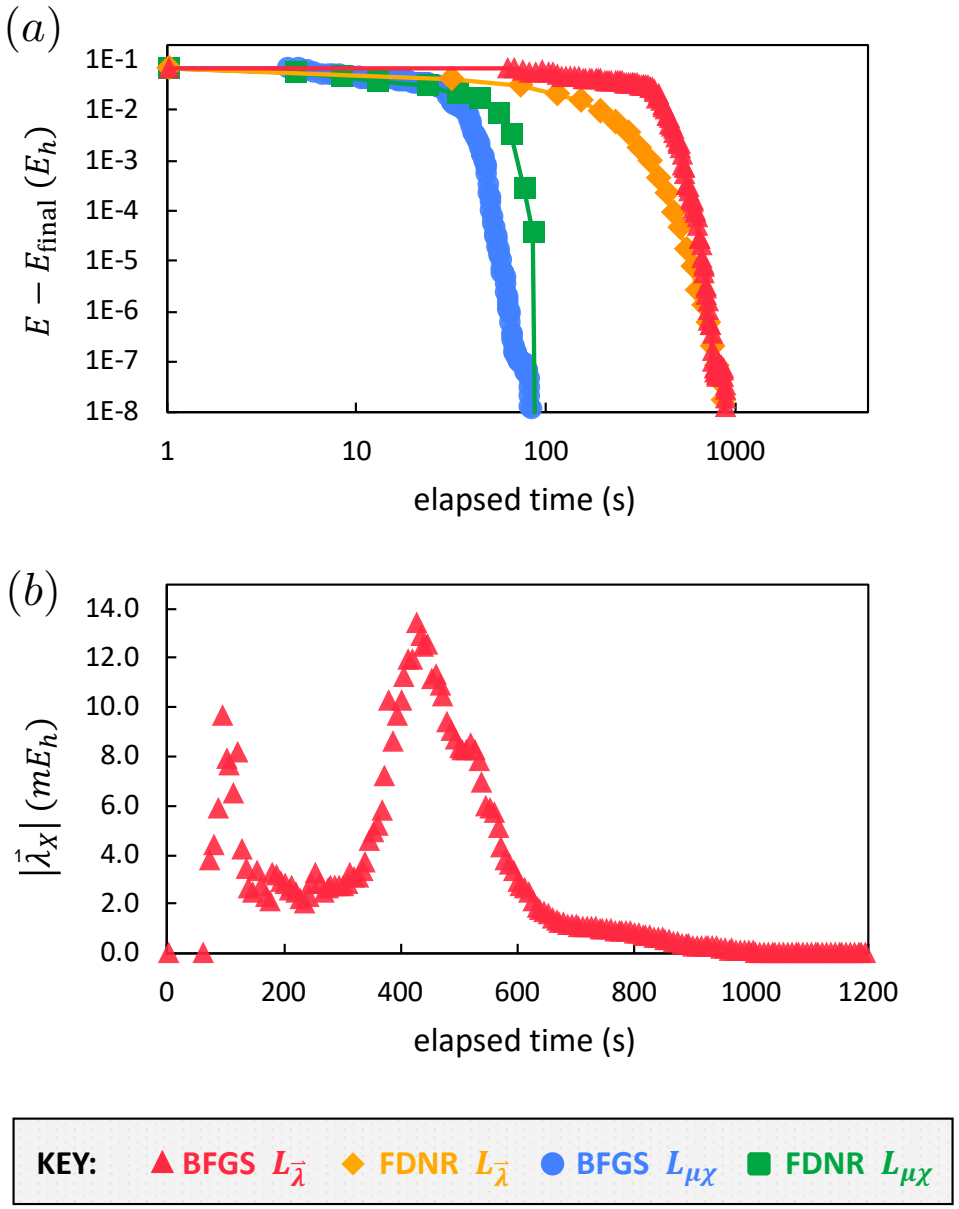


Figure 4.5: For the HOMO-LUMO singlet excitation in cc-pVDZ formaldehyde, the difference between the current and converged excited state energy in Hartree is compared across optimization strategies in (a) and the norm (in millihartrees) of the Lagrange multipliers associated with the orbital rotation parameters is shown over the course of an BFGS-$L_{\vec{\lambda}}$ optimization in (b).

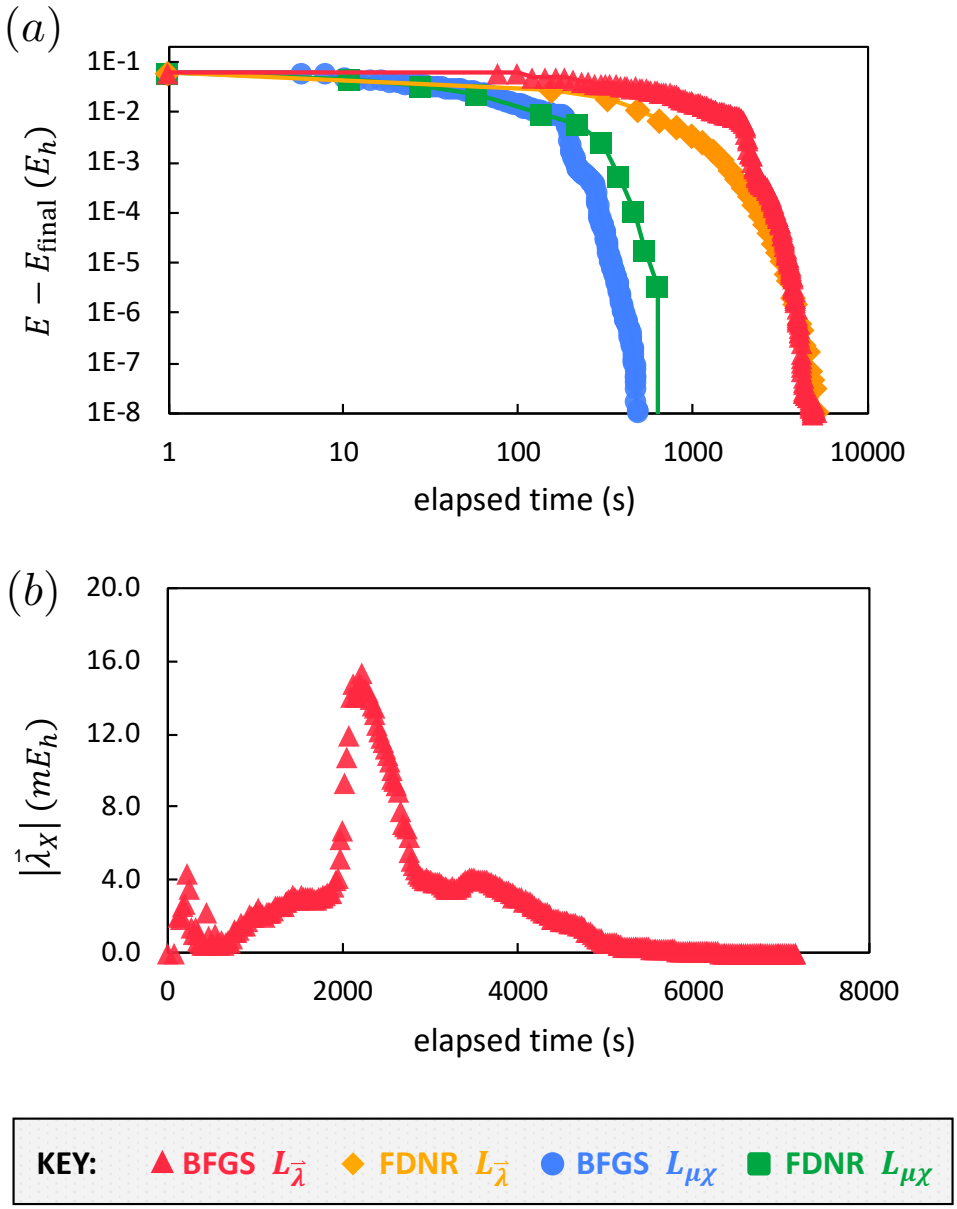


Figure 4.6: For the HOMO-LUMO singlet excitation in cc-pVDZ diazomethane, the difference between the current and converged excited state energy in Hartree is compared across optimization strategies in (a) and the norm (in millihartrees) of the Lagrange multipliers associated with the orbital rotation parameters is shown over the course of an BFGS-$L_{\vec{\lambda}}$ optimization in (b).

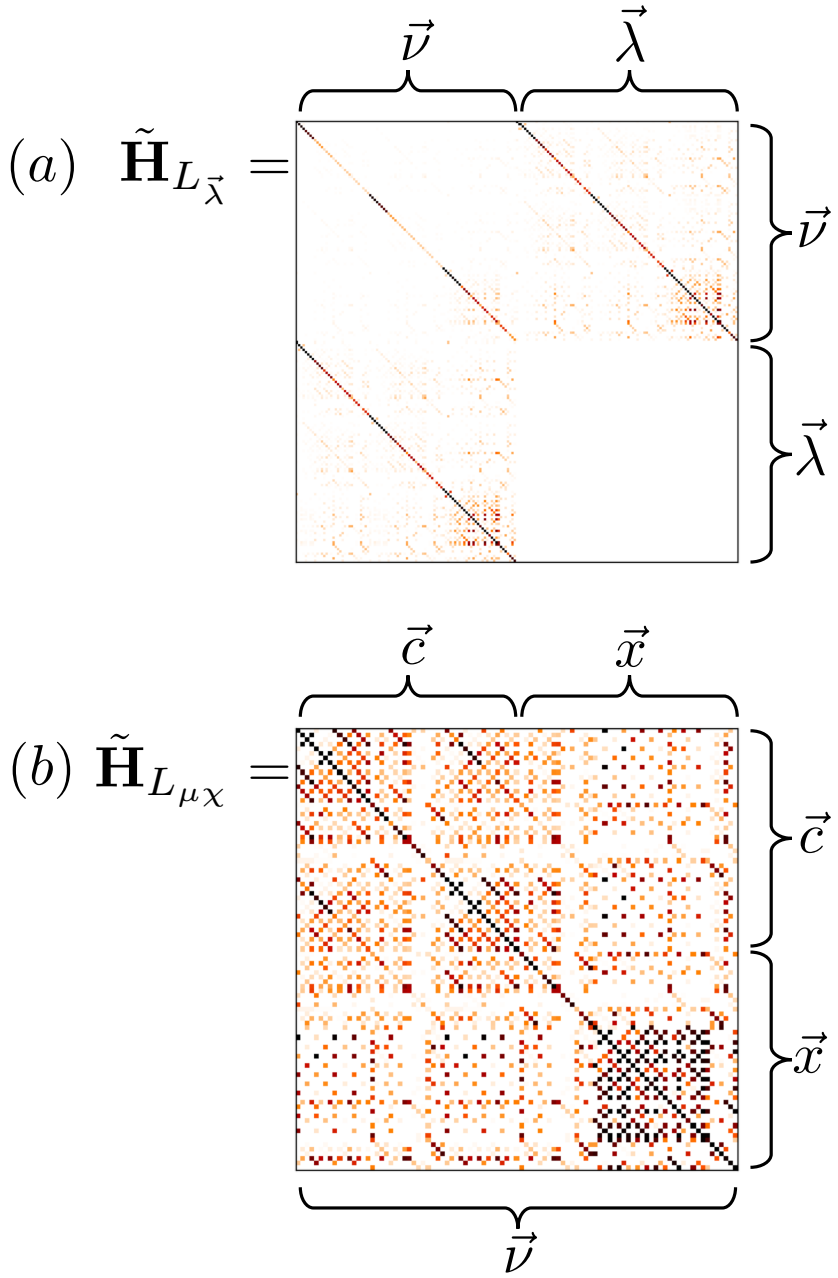


Figure 4.7: Heatmaps of the initial Hessian matrix for (a) the $L_{\vec{\lambda}}$ target function and (b) the $L_{\mu\chi}$ target function where $\mu = \frac{1}{2}$ and $\chi = 1$ for the HOMO-LUMO excitation in formaldehyde in the STO-3G basis. The values of the matrix are scaled such that the matrix elements equal to zero are white and the elements darken as they increase in magnitude. Note that in order to emphasize detail, the Hessians are scaled according to $\tilde{\mathbf{H}}_{\mathbf{ij}} = (1 - \exp[-|\mathbf{H}_{\mathbf{ij}}|])$ and that $\tilde{\mathbf{H}}_{L_{\mu\chi}}$ is enlarged with respect to $\tilde{\mathbf{H}}_{L_{\vec{\lambda}}}$.

from zero before returning them there. At the very least, this suggests that while our initial guesses for the wave function parameters and Lagrange multipliers are reasonable, they are not particularly well paired, in that the optimization of the wave function variables drives the multipliers away from their optimal values during a large fraction of the optimization. This issue is simply not present when using the $L_{\mu\chi}$ objective function as no multipliers are present and thus no guess for them is required.

To gain additional insight into the difficulties created through the Lagrange multipliers, we can look at the Hessian matrices produced by the two different objective functions, for which examples are shown in Fig. 4.7. As one might expect, the $L_{\vec{\lambda}}$ Hessian has a blocked structure, with the multiplier-multiplier block being zero trivially and the other three blocks being diagonally dominant. This structure is quite far from the identity-matrix guess of standard BFGS, and although it may be possible to construct a better estimate for the initial inverse Hessian this would require evaluating at least some of the second derivatives of the objective function individually, which is not guaranteed to have the same cost scaling as evaluating the energy. Although good estimates may be achievable at low cost, we have not in this study made any attempt at improving the initial BFGS Hessian guess for either objective function, and have simply used the identity matrix in both cases. As Fig. 4.7 shows, this very simple guess is a better fit for the single-diagonal diagonally dominant Hessian of the GVP-based objective function. As we move towards a production-level implementation of the ESMF wavefunction and GVP, we hope to further improve the optimization algorithms. As the overall computational cost of ESMF is dominated by the number of Fock builds, we anticipate significant speedups through Hessian preconditioning, integral screening,[371] resolution of the identity approaches,[372, 373] and, as our objective function is invariant to some orbital rotations, geometric descent minimization methods.[264]

Table 4.2: The maximum and mean unsigned errors for singlet excitation energies from seven methods vs δ-CR-EOM-CC(2,3)D in eV for several small organic molecules.

Unsigned Errors	**CIS**	**CIS(D)**	**EOM-CCSD**	**B3LYP**	**ωB97X-V**	**ESMF**	**ESMP2**
All Molecules							
Max	2.38	0.82	0.60	6.91	2.69	1.49	0.42
Mean	0.87	0.34	0.26	0.99	0.44	0.58	0.13
No CT							
Max	1.55	0.82	0.45	1.08	1.07	1.39	0.42
Mean	0.73	0.32	0.22	0.27	0.22	0.51	0.12

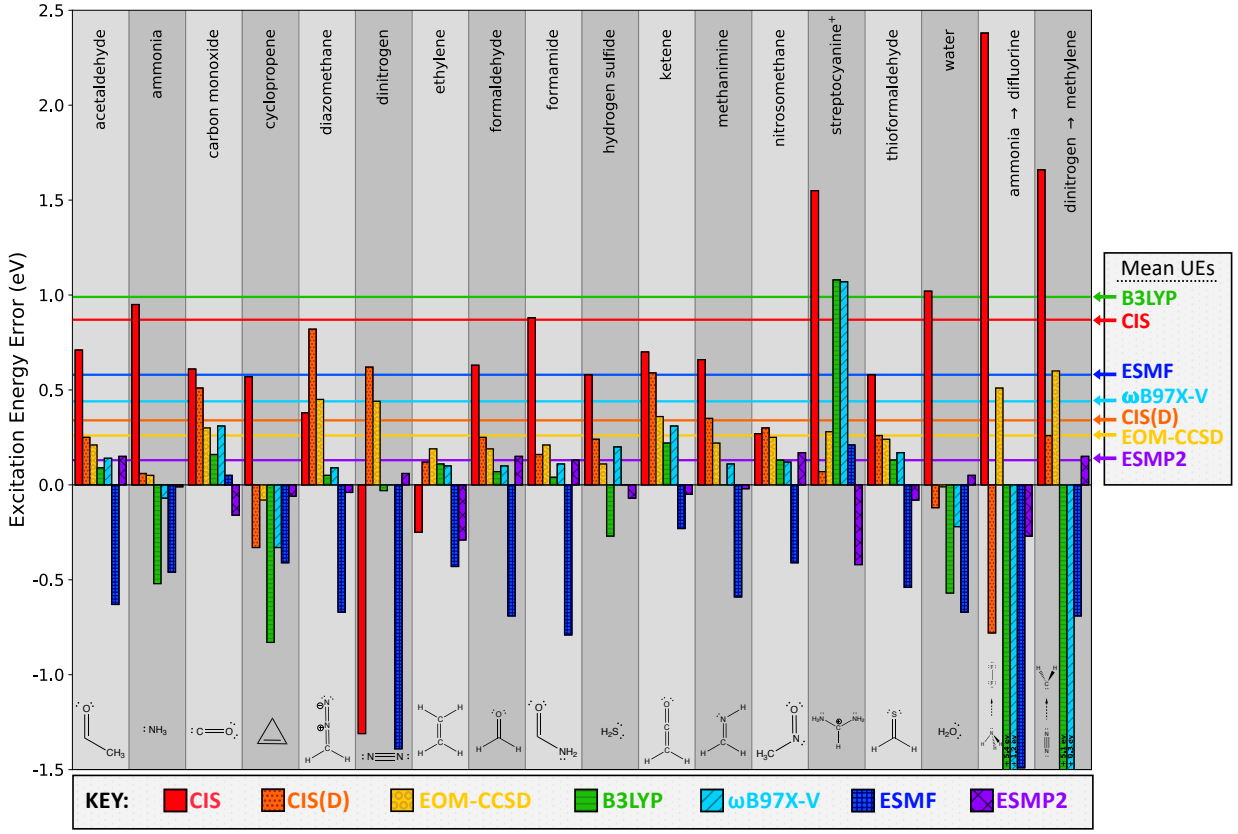


Figure 4.8: Singlet excitation energy errors from seven methods vs δ-CR-EOM-CC(2,3)D in eV for several small organic molecules.

Benchmarking Excitation Energies

We compiled a modest test set of small organic molecules that allows us to compare the accuracy of ESMF and ESMP2 against that of a range of single reference, weakly correlated excited state wave function methods. These methods include CIS, CIS(D), and notably EOM-CCSD and δ-CR-EOM-CC(2,3),D, the latter of which scales as $\mathcal{O}(N^7)$ and is used as a high level benchmark.[165] To contextualize the accuracy of ESMF and ESMP2 theories within the wider realm of excited state methods, we also present TD-DFT benchmarks against δ-CR-EOM-CC(2,3),D for both the B3LYP functional and the ωB97X-V functional – two popular hybrid GGA functionals.[209, 364, 367] For reference, the formal scaling of all methods used here is summarized in Table 4.3. While these scalings can in some cases be reduced via sparse linear algebra or integral screening, [374–376] we compare to canonical scalings here as our ESMF implementation does not yet take advantage of such approaches.

Our test set includes a number of intramolecular HOMO-LUMO singlet excitations as well as two long-range charge transfer excitations, $NH_3(n) \rightarrow F_2(\sigma^*)$ with a 6 Å separa-

Table 4.3: Formal cost-scaling for methods used in this work. These scalings are with respect to the system size N and are for canonical versions of the methods, i.e. without accelerations from two-electron integral screening or factorization.

Method	Formal Scaling
RHF[2]	N^4
CIS[325]	N^4
TD-DFT [325]	N^4
ESMF	N^4
CIS(D)[129]	N^5
EOM-CCSD[165]	N^6
δ-CR-EOM-CC(2,3),D[165]	N^7
ESMP2	N^7

tion, and $N_2(\pi) \rightarrow CH_2(2p)$ with a 10.4 Å separation. These types of CT excitations are known to cause difficulties for linear response theories; for example, CIS fails to capture how the shapes and sizes of the donor's and acceptor's orbitals change following the excitation. [127] In contrast, EOM-CCSD (through its doubles response operator) and ESMP2 (through ESMF's variational optimization) do capture the relaxation effects, which helps them achieve significantly better (although not perfect) energetics in these CT cases. [15, 296, 345, 377–380] Although the analysis for TD-DFT is less straightforward, even modern range-separated functionals do not account properly for all orbital relaxation effects, which continue to produce difficulties in charge transfer excitations[200, 325, 381] despite the clear improvements [120, 196, 382, 383] that range-separation offers.

The results for this survey are shown in Fig. 4.8 and tabulated in the SI. Overall, we see that EOM-CCSD and ESMP2 are most accurate in this test set, which we attribute to their ability to provide fully excited-state-specific orbital relaxations. In contrast, CIS, which lacks proper orbital relaxation, has the largest mean unsigned error (MUE) and maximum error of the wavefunction methods, performing especially poorly in the two charge transfer systems. Note that, although CIS can shape the orbitals for the electron and hole involved in the excitation via superpositions of different singles, it leaves the remaining occupied orbitals unrelaxed, which is notably inappropriate in long range CT where the large changes in local electron densities should lead all nearby valence orbitals to relax significantly.

CIS(D), an excited state analog of ground state MP2, recovers much of the electron correlation effects missing in CIS, halving the CIS MUE. Employing perturbation theory in the space of double excitations and using products of CIS single excitation amplitudes and ground state MP2 double excitation amplitudes to account for triple excitations from the ground state wave function, CIS(D) captures weak correlation and can improve the

CIS excitation energies for only $\mathcal{O}(N^5)$ cost.[129] The effect of the perturbative doubles and approximated triples on the excitation energies is certainly noticeable as the accuracy improves; however, CIS(D)'s use of ground state MP2 amplitudes leaves it lacking as from first principles, the electron-electron correlation in the ground state should differ from that in the excited state for any electrons involved in or near the excitation. We would thus expect a fully excited-state-specific perturbation theory to outperform CIS(D), and indeed this is what we find.

Within this test set of systems, we unsurprisingly observed that the accuracy of TD-DFT is both system and functional dependent. As summarized in Fig. 4.8, the accuracy of the B3LYP and ωB97X-V functionals across the intramolecular valence-excitations rivals that of EOM-CCSD. Indeed, it is difficult to motivate moving away from the remarkably low computational cost TD-DFT methods for such excitations.[120] However, the TD-DFT results are not as accurate in the charge transfer systems. The B3LYP functional performs particularly poorly, and while the accuracy of TD-DFT with the ωB97X-V functional, which uses HF exchange at long-ranges, is not quite so catastrophic, it drastically underestimates the excitation energy by multiple eV.

In this survey, ESMF consistently underestimates the excitation energy, and when the unsigned errors are compared, was more accurate than CIS, yet not as accurate as CIS(D). The underestimation can be understood by recognizing that, in the excited state, ESMF captures the pair-correlation energy between the two electrons in open-shell orbitals, whereas no correlation at all is preset in the RHF ground state (apart from Pauli correlation that of course ESMF also has). In addition, ESMF does in fact recover some weak correlation between different configurations of singly-excited determinants, i.e. for some configurations of i, j, a, and b, $\langle \Phi_i^a | e^{-\hat{X}} \hat{H} e^{\hat{X}} | \Phi_j^b \rangle \neq 0$. We are thus not surprised that ESMF excitation energies tend to be underestimates due to this capture of some correlation. Note that, as this is a very incomplete accounting of correlation effects (doubles and triples are missing), it is also not surprising that the overall accuracy of ESMF is inferior to that of CIS(D), which provides at least an approximate estimate of what the second order correction for the doubles and triples should be. Another notable point about the accuracy of ESMF is that it is not significantly different in the CT systems as compared to the other systems, suggesting that it has successfully captured the larger orbital relaxations present in CT.

Although its stand-alone accuracy leaves something to be desired, the ESMF wave function does provide an excellent starting point for post-mean-field correlation theories, as evidenced by the excellent performance of ESMP2. Thanks to its orbital-relaxed starting point and excited-state-specific determination of the doubles and triples, ESMP2 delivers the highest overall accuracy when compared to the δ-CR-EOM(2,3)D benchmark. In both intra- and inter-molecular excitations, ESMP2 is significantly more accurate than CIS, CIS(D), or ESMF, and slightly more accurate than EOM-CCSD. Of particular interest to note is that ESMP2 maintains its accuracy across both intramolecular valence excitations and long-range charge transfer excitations, and while this test set is too limited and the basis set too small to make strong recommendations, this data suggests that ESMP2 may in some circumstances be preferable to EOM-CCSD as well as TD-DFT in both intra- and inter-molecular excita-

tions. Certainly the data motivate work on versions of ESMP2 that avoid using the full set of uncontracted triples in the first order interaction space, which should lower its cost-scaling.

A Single-CSF Ansatz

As many excitations are dominated by a single open-shell CSF, one might wonder whether in these cases the full CIS-like CI expansion within ESMF is strictly necessary. Although the effect of the remaining singly-excited CSFs is not negligible, one could argue that their small weights put them firmly in the category of weak correlation effects that should be handled by the perturbation theory. For now, we have chosen not to pursue this direction in ESMP2 for two reasons. First, it would limit the theory to single-CSF-dominated excitations. Second, many of the other single excitations are much closer in energy to the reference wave function than the doubles excitations are, thus significantly increasing the risk of encountering intruder states. That said, we have used our present implementation to test how much the absence of these terms in ESMP2 matters if we restrict the reference function to be a single CSF with optimized orbitals, which we will refer to here as the oo-CSF ansatz.

$$|\text{oo-CSF}\rangle = e^{\hat{X}} \left(\hat{a}^\dagger_{a\uparrow} \hat{a}_{i\uparrow} |\Phi\rangle + \eta\, \hat{a}^\dagger_{a\downarrow} \hat{a}_{i\downarrow} |\Phi\rangle \right). \tag{4.25}$$

As in Eq. 4.6, $|\Phi\rangle$ denotes the RHF solution. However, the definition of $\hat{X}$ is slightly different. The oo-CSF ansatz is invariant to occupied-occupied and virtual-virtual orbital rotations that do not involve orbitals i or a, but such rotations that do involve these orbitals now matter, and so we have enabled these portions of the X matrix in addition to the ESMF occupied-virtual block shown in Fig. 4.1(b). Finally, note that η is not a variable and is simply set to 1 if we wish to work with the spin singlet and -1 for the triplet.

As seen in Table 4.4, we tested oo-CSF as a reference for ESMP2 in three systems where the structure of the optimized ESMF wave function suggested that oo-CSF had a good chance of being effective and one in which it did not appear appropriate. For water, the ESMF wave function is already dominated by a single CSF. For formaldehyde and methanimine, the additional subset of occupied-occupied rotations that we enabled for oo-CSF allow the primary components of the excitation to be converted into a single CSF by mixing the ESMF HOMO with the other occupied orbitals. While this simplification is certainly not always possible (N_2 is a good counterexample, having two large components involving completely separate sets of molecular orbitals) our results suggest that when it is, the absence of the other singles excitations in our ESMP2 method may not be of much consequence. In the future, the efficacy of oo-CSF for single-CSF-dominated states could perhaps be exploited in a couple of different ways. On the one hand, it is a simpler ansatz and so may prove easier to optimize than ESMF, which even in systems where secondary CSFs were not negligible could be useful if it provides a low-cost, high-quality initial guess for the ESMF optimization. On the other hand, its simpler structure could prove useful in simplifying the implementation of ESMP2.

Table 4.4: Excitation energy errors in eV relative to δ-CR-EOM-CC(2,3),D for the HOMO→LUMO singlet excitations of water, formaldehyde, methanimine, and dinitrogen. Below the name of each molecule, we report the CSF coefficients in the ESMF wave function with amplitudes larger than 0.1.

		ESMF	oo-CSF	ESMP2 w/ESMF	ESMP2 w/oo-CSF
Water		-0.67	-0.66	0.05	0.06
HOMO → LUMO	0.70				
Formaldehyde		-0.69	-0.66	0.15	0.18
HOMO → LUMO	0.66				
HOMO-3 → LUMO	0.22				
Methanimine		-0.59	-0.55	-0.02	0.02
HOMO → LUMO	0.67				
HOMO-2→ LUMO	-0.17				
Dinitrogen		-1.39	-1.13	0.06	0.52
HOMO-1 → LUMO	0.49				
HOMO→ LUMO+1	0.49				

Targeting with Other Properties

So far, we have focused on how an energy-targeting GVP can improve ESMF optimizations. We now turn our attention to the use of other properties to improve the robustness of optimization in the face of poor initial guesses, energetic degeneracy, and poor energy targeting. To investigate these aspects of the GVP, we study stretched LiH (bond distance 7 Å) in the STO-3G basis, whose low-lying states can be seen in Fig. 4.9. The idea is to optimize to the $|\Psi^{CT}_{s=0}\rangle$ state despite the challenges of (a) initial guesses that contain varying mixtures of $|\Psi^{EX}_{s=0}\rangle$ and $|\Psi^{CT}_{s=0}\rangle$ character, (b) setting the energy targeting to aim at the wrong state, namely setting ω to the ESMF energy for $|\Psi^{EX}_{s=0}\rangle$, and (c) the presence of $|\Psi^{CT}_{s=1}\rangle$, which is energetically degenerate with $|\Psi^{CT}_{s=0}\rangle$ at this bond distance. While the latter difficulty could be resolved by constraining our CI coefficients to produce only singlet states, we intentionally leave our CI coefficients unconstrained. Instead, we will investigate the efficacy of overcoming the challenges of degeneracy, poor ω choice, and poor initial guesses by including additional properties in the GVP's deviation vector $\vec{d}$.

As we wish to arrive at the neutral $|\Psi^{CT}_{s=0}\rangle$ state while avoiding the corresponding triplet

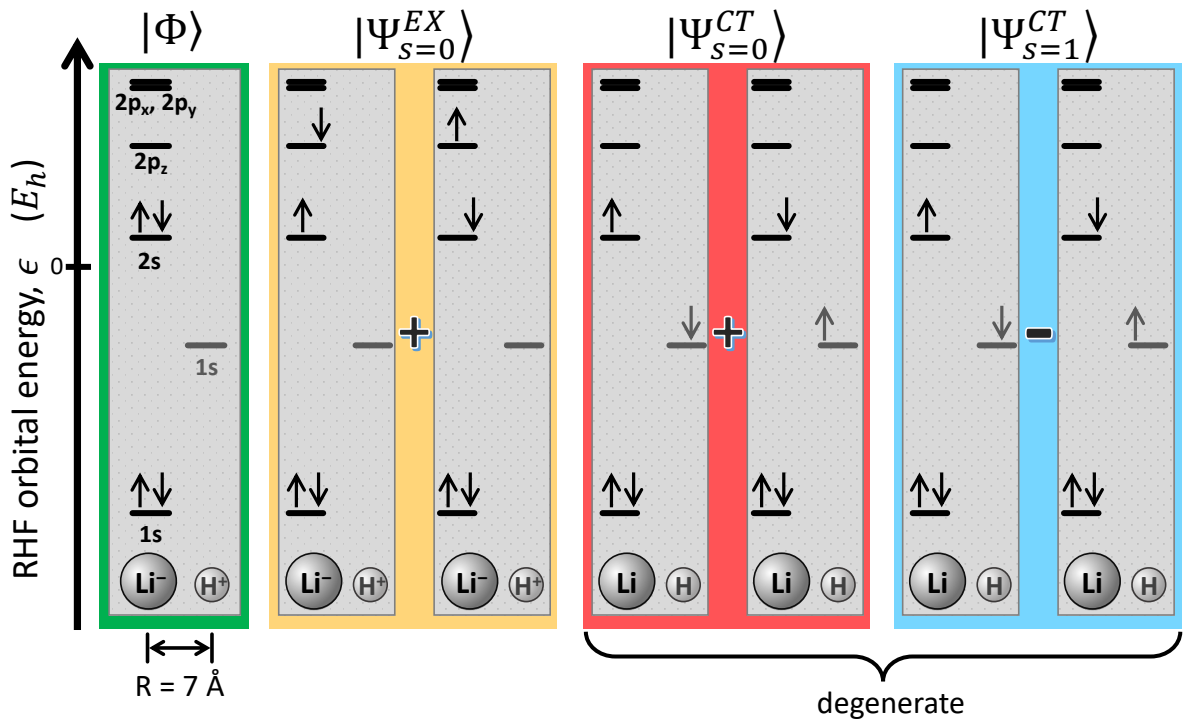


Figure 4.9: Molecular orbital diagrams for the key determinants in the four relevant states in our LiH test. From left to right, we have the ionic RHF wave function $|\Phi\rangle$, and the main CSFs that contribute to the singlet state $|\Psi^{EX}_{s=0}\rangle$ that maintains the ionic character, and that contribute to the degenerate singlet and triplet states $|\Psi^{CT}_{s=0}\rangle$ and $|\Psi^{CT}_{s=1}\rangle$ in which neutrality has been restored by an Li→H charge transfer. Note that while the molecular orbitals are arranged based on their RHF orbital energies, the energy gaps are not to scale.

state and being resilient to an initial guess contaminated by the ionic $|\Psi^{EX}_{s=0}\rangle$ state, the total spin and the Mulliken charges [2] of the atoms are obvious candidates for additional properties that should help uniquely identify our target state. We therefore chose our property deviation vector as

$$\vec{d} = \left\{ \langle E\rangle - \omega,\ \langle Q_{Li}\rangle - \eta,\ \sqrt{\langle S^2\rangle} - \zeta \right\} \tag{4.26}$$

where $\langle Q_{Li}\rangle$ is the Mulliken charge on the Li atom and we set $\eta = 0$ and $\zeta = 0$ so as to target a neutral singlet. Happily, both the values and the derivatives of the Mulliken charges and the total spin

$$\langle S^2\rangle = \frac{\sum_{ia}(\sigma_{ia} - \tau_{ia})^2}{\sum_{ia}\sigma_{ia}^2 + \tau_{ia}^2} \tag{4.27}$$

are easily evaluated for a CIS-like wavefunction like ESMF, and so the use of these properties does not change the cost-scaling of the method. In order to conveniently study the effects of

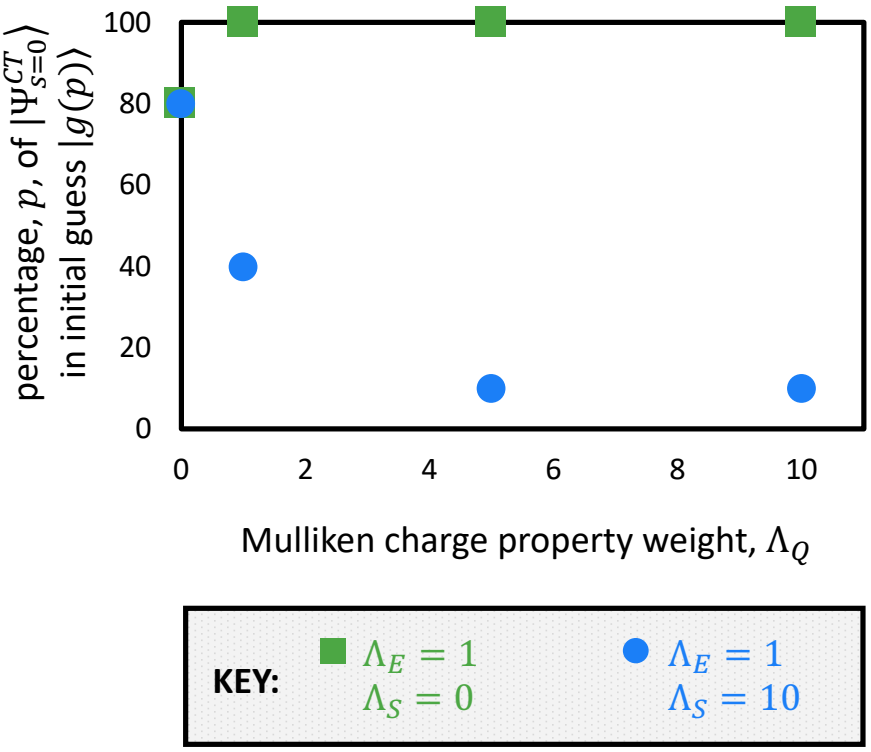


Figure 4.10: For various choices of the Λ matrix weights, we report the minimum value of p required for our initial guess $|g(p)\rangle$ to successfully optimize to the correct energy for the targeted $|\Psi_{s=0}^{CT}\rangle$ state.

putting different amounts of emphasis on different property deviations, we modify the GVP objective function to take the following form.

$$L_\Lambda = \mu\ \vec{d}^{\,T}\Lambda\ \vec{d} + (1-\mu)|\nabla E|^2 \tag{4.28}$$

$$\Lambda = \begin{bmatrix} \Lambda_E & 0 & 0 \\ 0 & \Lambda_Q & 0 \\ 0 & 0 & \Lambda_S \end{bmatrix} \tag{4.29}$$

Of course, this is equivalent to setting the semi-positive-definite matrix Λ to unity and scaling the definitions of the different properties, but we find the above form more convenient for presenting the different relative weightings that we placed on our three different property deviations.

For our initial guess $|g(p)\rangle$, we have set the orbital basis to the the RHF orbitals and have used varying mixtures of $|\phi^{EX}\rangle$ and $|\phi^{CT}\rangle$, which are the CIS wave functions corresponding

to $|\Psi^{EX}_{s=0}\rangle$ and $|\Psi^{CT}_{s=0}\rangle$, respectively.

$$|\mathrm{g}(p)\rangle = \sqrt{\frac{p}{100}}\ \left|\phi^{CT}\right\rangle + \sqrt{\frac{100-p}{100}}\ \left|\phi^{EX}\right\rangle \tag{4.30}$$

For each choice of the property weights in the Λ matrix, we tested whether $|\mathrm{g}(p)\rangle$ would successfully converge to the $|\Psi^{CT}_{s=0}\rangle$ state for the cases $p = 0$, 10, 20, ..., 100. Each optimization was performed via FDNR minimization of $L = \chi L_\Lambda + (1-\chi)E$, with μ stepped down from one to zero by intervals of 0.1 and χ switched from one to zero on the twentieth FDNR iteration. A value of -7.146 E_h, the ESMF energy for $|\Psi^{EX}_{s=0}\rangle$, was used for ω throughout.

As seen in Fig. 4.10, placing significant weights on both the charge and spin deviations allows for successful optimizations even with very poor initial guesses and our intentionally off-center value for ω. With $\Lambda_S = 10$ and $\Lambda_Q > 4$, we find that having as little as 10% of the correct CIS wavefunction in the initial guess leads to a successful optimization. When we do not include the spin targeting (i.e. when we set $\Lambda_S = 0$), we find that the charge targeting is much less effective, with no optimizations succeeding when less than 80% of the correct CIS wavefunction is in the guess, regardless of the value of Λ_Q. This result was somewhat unexpected, given that our guess is a pure spin singlet. We had expected that by giving the optimization a strong preference for neutral states, we would have converged to a linear combination of $|\Psi^{CT}_{s=0}\rangle$ and $|\Psi^{CT}_{s=1}\rangle$ that, while perhaps displaying some spin contamination, at least had the correct energy. Instead, we find that using spin to break the optimization degeneracy (by setting $\Lambda_S = 10$) is essential for robust convergence.

4.6 Conclusions

We have presented a generalization of the variational principle based on the energy gradient and the idea of constructing a flexible system for optimizing a state that can be specified uniquely by a list of properties. This approach is formally exact while avoiding the difficulties associated with squaring the Hamiltonian operator. Instead, it demands that a limited amount of energy second derivative information be evaluated, but, and this point is crucial, the required derivatives do not lead to an increase in cost scaling compared to the traditional ground state variational principle. So long as the properties used to identify the desired state do not themselves lead to an increase in cost scaling, the approach is therefore expected to maintain the scaling of its ground state counterpart.

Combining these ideas with excited state mean field theory, we have shown that the latter's optimization can be carried out without the need for the Lagrange multipliers that were present in its original formulation. We find that this approach leads to substantial efficiencies in the optimization thanks to both a simpler Hessian and an objective function that is bounded from below and thus easier to use straightforwardly with quasi-Newton optimization methods. We have also shown that a full Newton-Raphson approach can be

realized efficiently and without Hessian matrix construction by formulating Hessian matrix-vector products approximately via a finite-difference of gradients. Although it is not yet clear whether quasi-Newton methods or this full Newton approach will ultimately be faster for excited state mean field optimizations, what is clear is that the objective function based on the generalized variational principle is strongly preferable to the original objective function that relied on Lagrange multipliers.

With the ability to converge excited state mean field calculations in a larger set of molecules than was previously possible, we compared the corresponding second order perturbation theory to other commonly-used single-reference excited state methods and found its accuracy to be highly competitive. This success motivates both work on an internally contracted version of this perturbation theory in order to reduce its cost scaling and on fully excited-state-specific coupled cluster methods, which, if the history of ground state investigations is any guide, should be even more reliable than the perturbation theory.

More broadly, the generalized variational principle appears to offer new opportunities in many different areas of electronic structure theory. The ability to use a property vector to define which state is being sought without changing the final converged wave function should be especially useful in multi-reference investigations, where root flipping often prevents excited-state-specific calculations. By combining the energy with other properties, we demonstrated that the GVP could be used to resolve an individual state even in the presence of degeneracy, poor initial guesses, and poor energy targeting. There are of course many properties one could explore, but some that come immediately to mind are the dipole moment, changes in bond order from the ground state, and the degree of overlap with wave function estimates from other methods such as state averaging. In addition to multi-reference theory, the generalized variational principle appears to offer a route to defining exact density functionals for excited states so long as those states can be specified uniquely by a list of properties. Combined with promising preliminary data from a density functional extension of excited state mean field theory, [200] this formal foundation may allow for interesting new directions in density functional development.

4.7 Acknowledgements

We thank Piotr Piecuch for helpful discussions. This work was supported by the National Science Foundation's CAREER program under Award Number 1848012. Development, testing, and timing calculations were performed using the Berkeley Research Computing Savio cluster. The more extensive calculations required for the energy comparison between methods were performed using the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

J.A.R.S. acknowledges that this material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1752814. Any opinions, findings, and conclusions or recommendations expressed in this material are

those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

4.8 Supporting Information

The supporting information includes technical details about the FDNR numerical optimization, descriptions of the excitations in the test set, tabulated results for excitation energy errors, and molecular geometries for the systems tested in this work. This information is available free of charge via the Internet at http://pubs.acs.org.

Chapter 5

Conclusions

Three works were presented in this dissertation. Within Chapter 2, it was demonstrated that size consistency is lacking in interior state selective variational principles that are analytic around their global minima and based on at most the second power of the Hamiltonian. Additionally, a general optimization strategy that gradually merges a state selective variational principle from this set with variance minimization achieves size consistency at convergence and maintains rigorous state selectivity at all stages was presented. Widely applicable to a range of variational principles and compatible with wave function optimizers in VMC methods, the algorithm removes solution dependence on the energy targeting parameter ω, making QMC-based excited state optimization more blackbox. Implemented in a leading QMC software package,[266] this approach has been applied by others in the QMC community to study core excitations,[384] optical gaps in real solids,[385] and strongly correlated and frustrated quantum systems.[386] Moreover, the transformation algorithm opens doors for high-accuracy QMC studies of excited systems such as charge transfer excitations in solvated systems, perhaps those of interest in DNA photophysics, and intermolecular excitations between molecules and solids like those between organic dyes and nanoparticles in dye-sensitized solar cells.

Moving from stochastic to deterministic electronic structure theory, in Chapter 3 a minimally-correlated, computationally affordable, wave function-based excited state mean field theory that gives state specific, fully relaxed orbitals was described. Capable of modelling excitations with more than one significant contributing single excitation, this CIS-like wave function achieves qualitative descriptions of excited states at only $\mathcal{O}(N^4)$ cost in parallel with ground state Hartree Fock theory. Additionally, its potential as a platform on which to build excited state correlation treatments was realized through the development of an excited state analog to second-order Møller-Plesset perturbation theory, which rivals the accuracy of EOM-CCSD in valence, Rydberg, and CT excitations and has recently been implemented at $\mathcal{O}(N^5)$ cost.[387] These particularly promising initial results were bolstered by a broader study of organic molecules presented in Chapter 4, and within this set of valence and CT excitations, ESMP2 had a mean unsigned error with respect to a high-level benchmark half that of EOM-CCSD. The accuracy of ESMP2 in excitations dominated by

one or more single excitations is further evidenced by a currently unpublished benchmark of the method within the Thiel set of small and medium sized organic chromophores, which includes the nucleotides uracil, cytosine, thymine, and adenine.[162] The success of ESMF as a reference wave function for ESMP2 theory has encouraged the development of an $\mathcal{O}(N^6)$ scaling excited state analog to the coupled cluster singles and doubles method which is currently under construction.

To optimize the ESMF wave function, a novel energy gradient minimization method that uses Lagrangian optimization techniques to minimize an approximate excited state variational principle and treats ground and excited states on equal footing was introduced in Chapter 3. This optimization was improved upon by the research reported in Chapter 4. With the development of the generalized variational principle, which is based on the energy gradient and the idea of constructing a flexible system for optimizing a state that can be specified uniquely by a list of properties, the difficulties associated with squaring the Hamiltonian operator and with optimizing the Lagrange multipliers are avoided. Applying the GVP to the ESMF wave function demonstrated that this approach leads to substantial efficiencies in the optimization due to both a simpler Hessian and an objective function that is bounded from below and thus easier to use straightforwardly with quasi-Newton optimization methods, which were implemented efficiently based on a finite-difference approximation to Hessian matrix-vector multiplication. While the GVP requires some energy second derivatives to be computed, they do not lead to an increase in cost scaling compared to the traditional ground state variational principle. Similarly, as long as other state targeting properties do not increase the relative cost scaling, the scaling of the GVP will remain the same. Moreover, the ability to use a property vector to define which state is being sought offers new opportunities in many different areas of electronic structure theory. The property vector also allows for a natural translation between experiment and theory – for example, an experimentalist inquiring about a dark $\pi \rightarrow \pi^*$ state at 250 nm provides the theorist with targeting information regarding relative dipole, bond order, and energy targets which could be incorporated in the GVP. Additionally, targets related to the degree of overlap with different wave functions could assist theoretical investigations in which root-flipping and occupancy sloshing complicate optimizations. Finally, since the resulting optimized wave function is independent of the targeting parameters, the GVP provides an effective route to resolving energetically dense spectra without empirical parameterization. With its broad applicability to a wide variety of existing ansatzes and ongoing efforts towards its implementation in quantum Monte Carlo methods, the GVP is an exciting innovation in electronic structure optimization methods.

Bibliography

[1] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln". In: *Annalen der Physik* 389.20 (1927), pp. 457–484. DOI: https://doi.org/10.1002/andp.19273892002. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19273892002. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19273892002.

[2] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Mineola, N.Y.: Dover Publications, 1996, pp. 43–45.

[3] Aleksandr V. Marenich et al. "Practical computation of electronic excitation in solution: vertical excitation model". In: *Chem. Sci.* 2 (11 2011), pp. 2143–2161. DOI: 10.1039/C1SC00313E. URL: http://dx.doi.org/10.1039/C1SC00313E.

[4] Graham A. Worth and Lorenz S. Cederbaum. "BEYOND BORN-OPPENHEIMER: Molecular Dynamics Through a Conical Intersection". In: *Annual Review of Physical Chemistry* 55.1 (2004). PMID: 15117250, pp. 127–158. DOI: 10.1146/annurev.physchem.55.091602.094335. eprint: https://doi.org/10.1146/annurev.physchem.55.091602.094335. URL: https://doi.org/10.1146/annurev.physchem.55.091602.094335.

[5] Kristian B. Ørnsø et al. "Importance of the Reorganization Energy Barrier in Computational Design of Porphyrin-Based Solar Cells with Cobalt-Based Redox Mediators". In: *The Journal of Physical Chemistry C* 119.23 (2015), pp. 12792–12800. DOI: 10.1021/jp512627e. eprint: https://doi.org/10.1021/jp512627e. URL: https://doi.org/10.1021/jp512627e.

[6] Valérie Vaissier et al. "Influence of polar medium on the reorganization energy of charge transfer between dyes in a dye sensitized film". In: *Phys. Chem. Chem. Phys.* 15 (13 2013), pp. 4804–4814. DOI: 10.1039/C3CP44562C. URL: http://dx.doi.org/10.1039/C3CP44562C.

[7] Davide Moia et al. "The reorganization energy of intermolecular hole hopping between dyes anchored to surfaces". In: *Chem. Sci.* 5 (1 2014), pp. 281–290. DOI: 10.1039/C3SC52359D. URL: http://dx.doi.org/10.1039/C3SC52359D.

[8] Melvin B. Robin. *Higher Excited States of Polyatomic Molecules: Volume II*. New York, New York: Academic Press, Inc., 1975.

[9] Igor Fedorov et al. "Theoretical and Experimental Investigations of the Electronic Rydberg States of Diazomethane: Assignments and State Interactions". In: *The Journal of Physical Chemistry A* 111.21 (2007), pp. 4557–4566. DOI: 10.1021/jp071590r. eprint: https://doi.org/10.1021/jp071590r. URL: https://doi.org/10.1021/jp071590r.

[10] Sergey V. Levchenko and Anna I. Krylov. "Electronic structure of halogen-substituted methyl radicals: Excited states of CH2Cl and CH2F". In: *The Journal of Chemical Physics* 115.16 (2001), pp. 7485–7494. DOI: 10.1063/1.1400143. eprint: https://doi.org/10.1063/1.1400143. URL: https://doi.org/10.1063/1.1400143.

[11] Sergey V. Levchenko et al. "Rydberg–valence interactions in $CH_2Cl \rightarrow CH_2+Cl$ photodissociation: Dependence of absorption probability on ground state vibrational excitation". In: *The Journal of Chemical Physics* 118.20 (2003), pp. 9233–9240. DOI: 10.1063/1.1568076. eprint: https://doi.org/10.1063/1.1568076. URL: https://doi.org/10.1063/1.1568076.

[12] O. Geßner et al. "Femtosecond Multidimensional Imaging of a Molecular Dissociation". In: *Science* 311.5758 (2006), pp. 219–222. ISSN: 0036-8075. DOI: 10.1126/science.1120779. eprint: https://science.sciencemag.org/content/311/5758/219.full.pdf. URL: https://science.sciencemag.org/content/311/5758/219.

[13] Sergey V. Levchenko et al. "Photodissociation dynamics of the NO dimer. I. Theoretical overview of the ultraviolet singlet excited states". In: *The Journal of Chemical Physics* 125.8 (2006), p. 084301. DOI: 10.1063/1.2222355. eprint: https://doi.org/10.1063/1.2222355. URL: https://doi.org/10.1063/1.2222355.

[14] Lucas Koziol, Sergey V. Levchenko, and Anna I. Krylov. "Beyond Vinyl: Electronic Structure of Unsaturated Propen-1-yl, Propen-2-yl, 1-Buten-2-yl, and trans-2-Buten-2-yl Hydrocarbon Radicals". In: *The Journal of Physical Chemistry A* 110.8 (2006), pp. 2746–2758. DOI: 10.1021/jp055375r. eprint: https://doi.org/10.1021/jp055375r. URL: https://doi.org/10.1021/jp055375r.

[15] Anna I. Krylov. "Equation-of-Motion Coupled-Cluster Methods for Open-Shell and Electronically Excited Species: The Hitchhiker's Guide to Fock Space". In: *Annual Review of Physical Chemistry* 59.1 (2008), pp. 433–462. DOI: 10.1146/annurev.physchem.59.032607.093602. eprint: https://doi.org/10.1146/annurev.physchem.59.032607.093602. URL: https://doi.org/10.1146/annurev.physchem.59.032607.093602.

[16] Philipp Marquetand et al. "Challenges in Simulating Light-Induced Processes in DNA". In: *Molecules* 22.1 (2017). ISSN: 1420-3049. DOI: 10.3390/molecules22010049. URL: https://www.mdpi.com/1420-3049/22/1/49.

[17] Jr. Guy Gery P. et al. "Vital Signs : Melanoma Incidence and Mortality Trends and Projections — United States, 1982–2030." In: *Morbidity and Mortality Weekly Report* 64.21 (2015), p. 591. ISSN: 01492195.

[18] U.S. Cancer Statistics Working Group. *U.S. Cancer Statistics Data Visualizations Tool, based on November 2018 submission data (1999-2016)*. 2019. URL: %5Curl%7Bwww.cdc.gov/cancer/dataviz%7D.

[19] M. H. Patrick and R. O. Rahn. *Photochemistry and Photobiology of Nucleic Acids Volume 2*. New York, New York: Academic Press, Inc., 1976, pp. 35–95.

[20] Brant E. Billinghurst, Ralph Yeung, and Glen R. Loppnow. "Excited-State Structural Dynamics of 5-Fluorouracil". In: *The Journal of Physical Chemistry A* 110.19 (2006). PMID: 16686451, pp. 6185–6191. DOI: 10.1021/jp0609333. eprint: https://doi.org/10.1021/jp0609333. URL: https://doi.org/10.1021/jp0609333.

[21] Brant E. Billinghurst and Glen R. Loppnow. "Excited-State Structural Dynamics of Cytosine from Resonance Raman Spectroscopy". In: *The Journal of Physical Chemistry A* 110.7 (2006). PMID: 16480294, pp. 2353–2359. DOI: 10.1021/jp0561571. eprint: https://doi.org/10.1021/jp0561571. URL: https://doi.org/10.1021/jp0561571.

[22] Soujanya Yarasi, Philip Brost, and Glen R. Loppnow. "Initial Excited-State Structural Dynamics of Thymine Are Coincident with the Expected Photochemical Dynamics". In: *The Journal of Physical Chemistry A* 111.24 (2007). PMID: 17530833, pp. 5130–5135. DOI: 10.1021/jp071443t. eprint: https://doi.org/10.1021/jp071443t. URL: https://doi.org/10.1021/jp071443t.

[23] Chris T. Middleton et al. "DNA Excited-State Dynamics: From Single Bases to the Double Helix". In: *Ann. Rev. Phys. Chem.* 60.1 (2009). PMID: 19012538, pp. 217–239. DOI: 10.1146/annurev.physchem.59.032607.093719. eprint: https://doi.org/10.1146/annurev.physchem.59.032607.093719. URL: https://doi.org/10.1146/annurev.physchem.59.032607.093719.

[24] Philip C. Hanawalt et al. "DNA Repair in Bacteria and Mammalian Cells". In: *Annual Review of Biochemistry* 48.1 (1979). PMID: 382997, pp. 783–836. DOI: 10.1146/annurev.bi.48.070179.004031. eprint: https://doi.org/10.1146/annurev.bi.48.070179.004031. URL: https://doi.org/10.1146/annurev.bi.48.070179.004031.

[25] DL Mitchell and RS Nairn. "The (6-4) photoproduct and human skin cancer". In: *Photodermatology* 5.2 (1988), pp. 61–64. ISSN: 0108-9684. URL: http://europepmc.org/abstract/MED/3399436.

[26] DAVID L. MITCHELL et al. "Photoreactivation Of Cyclobutane Dimers and (6–4) Photoproducts in the Epidermis of the Marsupial, Monodelphis domestica". In: *Photochemistry and Photobiology* 51.6 (1990), pp. 653–658. DOI: https://doi.org/10.1111/php.1990.51.6.653. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/php.1990.51.6.653. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/php.1990.51.6.653.

[27] Leena Latonen and Marikki Laiho. "Cellular UV damage responses—Functions of tumor suppressor p53". In: *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1755.2 (2005), pp. 71–89. ISSN: 0304-419X. DOI: https://doi.org/10.1016/j.bbcan.2005.04.003. URL: https://www.sciencedirect.com/science/article/pii/S0304419X05000107.

[28] Patricia A.J. Muller. "Mutant p53 in Cancer: New Functions and Therapeutic Opportunities". In: *Cancer Cell* 25.3 (2014), pp. 304–317. DOI: 10.1016/j.ccr.2014.01.021. URL: https://doi.org/10.1016/j.ccr.2014.01.021.

[29] Wolfgang J. Schreier et al. "Thymine Dimerization in DNA Is an Ultrafast Photoreaction". In: *Science* 315.5812 (2007), pp. 625–629. ISSN: 0036-8075. DOI: 10.1126/science.1135428. eprint: https://science.sciencemag.org/content/315/5812/625.full.pdf. URL: https://science.sciencemag.org/content/315/5812/625.

[30] Carlos E. Crespo-Hernández, Boiko Cohen, and Bern Kohler. "Base stacking controls excited-state dynamics in A·T DNA". In: *Nature* 436 (2005), pp. 1141–1144. DOI: 10.1038/nature03933. URL: https://doi.org/10.1038/nature03933.

[31] Carlos E. Crespo-Hernández et al. "Ultrafast Excited-State Dynamics in Nucleic Acids". In: *Chemical Reviews* 104.4 (2004). PMID: 15080719, pp. 1977–2020. DOI: 10.1021/cr0206770. eprint: https://doi.org/10.1021/cr0206770. URL: https://doi.org/10.1021/cr0206770.

[32] P R Callis. "Electronic States and Luminescence of Nucleic Acid Systems". In: *Annual Review of Physical Chemistry* 34.1 (1983), pp. 329–357. DOI: 10.1146/annurev.pc.34.100183.001553. eprint: https://doi.org/10.1146/annurev.pc.34.100183.001553. URL: https://doi.org/10.1146/annurev.pc.34.100183.001553.

[33] Patrick M. Hare, Carlos E. Crespo-Hernández, and Bern Kohler. "Internal conversion to the electronic ground state occurs via two distinct pathways for pyrimidine bases in aqueous solution". In: *Proc. Natl. Acad. Sci.* 104.2 (2007), pp. 435–440. ISSN: 0027-8424. DOI: 10.1073/pnas.0608055104.

[34] Wai-Ming Kwok, Chensheng Ma, and David Lee Phillips. "A Doorway State Leads to Photostability or Triplet Photodamage in Thymine DNA". In: *Journal of the American Chemical Society* 130.15 (2008). PMID: 18335986, pp. 5131–5139. DOI: 10.1021/ja077831q. eprint: https://doi.org/10.1021/ja077831q. URL: https://doi.org/10.1021/ja077831q.

[35] Clélia Canuel et al. "Excited states dynamics of DNA and RNA bases: Characterization of a stepwise deactivation pathway in the gas phase". In: *The Journal of Chemical Physics* 122.7 (2005), p. 074316. DOI: 10.1063/1.1850469. eprint: https://doi.org/10.1063/1.1850469. URL: https://doi.org/10.1063/1.1850469.

[36] Lluís Blancafort. "Energetics of Cytosine Singlet Excited-State Decay Paths—A Difficult Case for CASSCF and CASPT2†". In: *Photochemistry and Photobiology* 83.3 (2007), pp. 603–610. DOI: https://doi.org/10.1562/2006-05-29-RA-903. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1562/2006-05-29-RA-903. URL: https://onlinelibrary.wiley.com/doi/abs/10.1562/2006-05-29-RA-903.

[37] Christel M. Marian. "A new pathway for the rapid decay of electronically excited adenine". In: *The Journal of Chemical Physics* 122.10 (2005), p. 104314. DOI: 10.1063/1.1861452. eprint: https://doi.org/10.1063/1.1861452. URL: https://doi.org/10.1063/1.1861452.

[38] Helmut Satzger et al. "Primary processes underlying the photostability of isolated DNA bases: Adenine". In: *Proceedings of the National Academy of Sciences* 103.27 (2006), pp. 10196–10201. ISSN: 0027-8424. DOI: 10.1073/pnas.0602663103. eprint: https://www.pnas.org/content/103/27/10196.full.pdf. URL: https://www.pnas.org/content/103/27/10196.

[39] Patrick M. Hare, Carlos E. Crespo-Hernández, and Bern Kohler. "Solvent-Dependent Photophysics of 1-Cyclohexyluracil: Ultrafast Branching in the Initial Bright State Leads Nonradiatively to the Electronic Ground State and a Long-Lived $^1n \rightarrow \pi^*$ State". In: *The Journal of Physical Chemistry B* 110.37 (2006). PMID: 16970494, pp. 18641–18650. DOI: 10.1021/jp064714t. eprint: https://doi.org/10.1021/jp064714t. URL: https://doi.org/10.1021/jp064714t.

[40] Manuela Merchán et al. "Triplet-State Formation along the Ultrafast Decay of Excited Singlet Cytosine". In: *Journal of the American Chemical Society* 127.6 (2005). PMID: 15701017, pp. 1820–1825. DOI: 10.1021/ja044371h. eprint: https://doi.org/10.1021/ja044371h. URL: https://doi.org/10.1021/ja044371h.

[41] Teresa Climent et al. "On the intrinsic population of the lowest triplet state of uracil". In: *Chemical Physics Letters* 441.4 (2007), pp. 327–331. ISSN: 0009-2614. DOI: https://doi.org/10.1016/j.cplett.2007.05.040. URL: https://www.sciencedirect.com/science/article/pii/S0009261407006276.

[42] Carlos E. Crespo-Hernández and Bern Kohler. "Influence of Secondary Structure on Electronic Energy Relaxation in Adenine Homopolymers". In: *J. Phys. Chem. B* 108.30 (2004), pp. 11182–11188. DOI: 10.1021/jp0496046. eprint: https://doi.org/10.1021/jp0496046. URL: https://doi.org/10.1021/jp0496046.

[43] Carlos E. Crespo-Hernández, Boiko Cohen, and Bern Kohler. "Base stacking controls excited-state dynamics in A·T DNA". In: *Nature* 436 (2005), pp. 1141–1144. DOI: 10.1038/nature03933. URL: https://doi.org/10.1038/nature03933.

[44] Adrian W. Lange, Mary A. Rohrdanz, and John M. Herbert. "Charge-Transfer Excited States in a π-Stacked Adenine Dimer, As Predicted Using Long-Range-Corrected Time-Dependent Density Functional Theory". In: *The Journal of Physical Chemistry B* 112.20 (2008). PMID: 18438995, pp. 6304–6308. DOI: 10.1021/jp802058k. eprint: https://doi.org/10.1021/jp802058k. URL: https://doi.org/10.1021/jp802058k.

[45] Vincent A. Spata, Wook Lee, and Spiridoula Matsika. "Excimers and Exciplexes in Photoinitiated Processes of Oligonucleotides". In: *The Journal of Physical Chemistry Letters* 7.6 (2016). PMID: 26911276, pp. 976–984. DOI: 10.1021/acs.jpclett.5b02756. eprint: https://doi.org/10.1021/acs.jpclett.5b02756. URL: https://doi.org/10.1021/acs.jpclett.5b02756.

[46] Felix Plasser et al. "Surface Hopping Dynamics with Correlated Single-Reference Methods: 9H-Adenine as a Case Study". In: *Journal of Chemical Theory and Computation* 10.4 (2014), pp. 1395–1405. DOI: 10.1021/ct4011079. eprint: https://doi.org/10.1021/ct4011079. URL: https://doi.org/10.1021/ct4011079.

[47] Mariachiara Pastore, Filippo De Angelis, and Celestino Angeli. "Optical absorption spectrum of the N3 solar cell sensitizer by second-order multireference perturbation theory". In: *Theoretical Chemistry Accounts* 135 (2016), p. 108. DOI: 10.1007/s00214-016-1868-y. URL: https://doi.org/10.1007/s00214-016-1868-y.

[48] You Lu, Zhenggang Lan, and Walter Thiel. "Computational Modeling of Photoexcitation in DNA Single and Double Strands". In: *Photoinduced Phenomena in Nucleic Acids II: DNA Fragments and Phenomenological Aspects*. Ed. by Mario Barbatti, Antonio Carlos Borin, and Susanne Ullrich. Cham: Springer International Publishing, 2015, pp. 89–122. ISBN: 978-3-319-13272-3. DOI: 10.1007/128_2014_533. URL: https://doi.org/10.1007/128_2014_533.

[49] Tomáš Zelený et al. "Strikingly Different Effects of Hydrogen Bonding on the Photodynamics of Individual Nucleobases in DNA: Comparison of Guanine and Cytosine". In: *Journal of the American Chemical Society* 134.33 (2012). PMID: 22845192, pp. 13662–13669. DOI: 10.1021/ja3028845. eprint: https://doi.org/10.1021/ja3028845. URL: https://doi.org/10.1021/ja3028845.

[50] Jiří Šponer, Kevin E. Riley, and Pavel Hobza. "Nature and magnitude of aromatic stacking of nucleic acid bases". In: *Phys. Chem. Chem. Phys.* 10 (19 2008), pp. 2595–2610. DOI: 10.1039/B719370J. URL: http://dx.doi.org/10.1039/B719370J.

[51] Irene Conti et al. "Excited state evolution of DNA stacked adenines resolved at the CASPT2//CASSCF/Amber level: from the bright to the excimer state and back". In: *Phys. Chem. Chem. Phys.* 17 (11 2015), pp. 7291–7302. DOI: 10.1039/C4CP05546B. URL: http://dx.doi.org/10.1039/C4CP05546B.

[52] Zsuzsanna Benda and Péter G. Szalay. "Characterization of the excited states of DNA building blocks: a coupled cluster computational study". In: *Phys. Chem. Chem. Phys.* 18 (34 2016), pp. 23596–23606. DOI: 10.1039/C6CP02969H. URL: http://dx.doi.org/10.1039/C6CP02969H.

[53] Péter G. Szalay et al. "Benchmark Studies on the Building Blocks of DNA. 3. Watson–Crick and Stacked Base Pairs". In: *The Journal of Physical Chemistry A* 117.15 (2013), pp. 3149–3157. DOI: 10.1021/jp3100975. eprint: https://doi.org/10.1021/jp3100975. URL: https://doi.org/10.1021/jp3100975.

[54] Chris T. Middleton, Boiko Cohen, and Bern Kohler. "Solvent and Solvent Isotope Effects on the Vibrational Cooling Dynamics of a DNA Base Derivative". In: *The Journal of Physical Chemistry A* 111.42 (2007). PMID: 17910423, pp. 10460–10467. DOI: 10.1021/jp0740595. eprint: https://doi.org/10.1021/jp0740595. URL: https://doi.org/10.1021/jp0740595.

[55] Péter G. Szalay et al. "Benchmark Studies on the Building Blocks of DNA. 2. Effect of Biological Environment on the Electronic Excitation Spectrum of Nucleobases". In: *The Journal of Physical Chemistry A* 116.35 (2012). PMID: 22845884, pp. 8851–8860. DOI: 10.1021/jp305130q. eprint: https://doi.org/10.1021/jp305130q. URL: https://doi.org/10.1021/jp305130q.

[56] Martin A Green. "Third generation photovoltaics: solar cells for 2020 and beyond". In: *Physica E: Low-dimensional Systems and Nanostructures* 14.1 (2002), pp. 65–70. ISSN: 1386-9477. DOI: https://doi.org/10.1016/S1386-9477(02)00361-2. URL: https://www.sciencedirect.com/science/article/pii/S1386947702003612.

[57] Vishal Shrotriya. "Polymer power". In: *Nature Photonics* 3 (2009), pp. 447–449. DOI: 10.1038/nphoton.2009.130. eprint: https://doi.org/10.1038/nphoton.2009.130. URL: https://doi.org/10.1038/nphoton.2009.130.

[58] Baobing Fan et al. "Fine-tuning of the chemical structure of photoactive materials for highly efficient organic photovoltaics". In: *Nature Energy* 3.12 (2018), pp. 1051–1058. DOI: 10.1038/s41560-018-0263-4. URL: https://doi.org/10.1038/s41560-018-0263-4.

[59] Craig H. Peters et al. "High Efficiency Polymer Solar Cells with Long Operating Lifetimes". In: *Advanced Energy Materials* 1.4 (2011), pp. 491–494. DOI: https://doi.org/10.1002/aenm.201100138. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.201100138. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/aenm.201100138.

[60] Agnès Rivaton et al. "Light-induced degradation of the active layer of polymer-based solar cells". In: *Polymer Degradation and Stability* 95.3 (2010). Special Issue: MoDeSt 2008, pp. 278–284. ISSN: 0141-3910. DOI: https://doi.org/10.1016/j.polymdegradstab.2009.11.021. URL: https://www.sciencedirect.com/science/article/pii/S0141391009003711.

[61] Matthieu Manceau et al. "Light-induced degradation of the P3HT-based solar cells active layer". In: *Solar Energy Materials and Solar Cells* 95.5 (2011). Special Issue : 3rd International Summit on OPV Stability, pp. 1315–1325. ISSN: 0927-0248. DOI: https://doi.org/10.1016/j.solmat.2010.09.021. URL: https://www.sciencedirect.com/science/article/pii/S0927024810005453.

[62] Sabine Bertho et al. "Influence of thermal ageing on the stability of polymer bulk heterojunction solar cells". In: *Solar Energy Materials and Solar Cells* 91.5 (2007). Selected Papers from the European Conference on Hybrid and Organic Solar Cells – ECHOS '06, pp. 385–389. ISSN: 0927-0248. DOI: https://doi.org/10.1016/j.solmat.2006.10.008. URL: https://www.sciencedirect.com/science/article/pii/S0927024806004089.

[63] Sabine Bertho et al. "Effect of temperature on the morphological and photovoltaic stability of bulk heterojunction polymer:fullerene solar cells". In: *Solar Energy Materials and Solar Cells* 92.7 (2008). Degradation and Stability of Polymer and Organic Solar Cells, pp. 753–760. ISSN: 0927-0248. DOI: https://doi.org/10.1016/j.solmat.2008.01.006. URL: https://www.sciencedirect.com/science/article/pii/S0927024808000081.

[64] Russell Gaudiana and Christoph Brabec. "Fantastic plastic". In: *Nature Photonics* 2.5 (2008), pp. 287–289. DOI: 10.1038/nphoton.2008.69. URL: https://doi.org/10.1038/nphoton.2008.69.

[65] Sean E. Shaheen, David S. Ginley, and Ghassan E. Jabbour. "Organic-Based Photovoltaics: Toward Low-Cost Power Generation". In: *MRS Bulletin* 30.1 (2005), pp. 10–19. DOI: 10.1557/mrs2005.2. URL: https://doi.org/10.1557/mrs2005.2.

[66] Brian O'Regan and Michael Grätzel. "A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films". In: *Nature* 353 (6346 1991), pp. 737–740. DOI: 10.1038/353737a0. URL: https://doi.org/10.1038/353737a0.

[67] Brian E. Hardin, Henry J. Snaith, and Michael D. McGehee. "The renaissance of dye-sensitized solar cells". In: *Nature Photonics* 6.3 (2012), pp. 162–169. DOI: 10.1038/nphoton.2012.22. URL: https://doi.org/10.1038/nphoton.2012.22.

[68] Chin-Li Wang et al. "Enhanced light harvesting with π-conjugated cyclic aromatic hydrocarbons for porphyrin-sensitized solar cells". In: *Energy & Environmental Science* 4 (5 2011), pp. 1788–1795. DOI: 10.1039/C0EE00767F. URL: http://dx.doi.org/10.1039/C0EE00767F.

[69] Jiawei Gong et al. "Review on dye-sensitized solar cells (DSSCs): Advanced techniques and research trends". In: *Renewable and Sustainable Energy Reviews* 68 (2017), pp. 234–246. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2016.09.097. URL: https://www.sciencedirect.com/science/article/pii/S1364032116305883.

[70] M Riede et al. "Small-molecule solar cells—status and perspectives". In: *Nanotechnology* 19.42 (2008), p. 424001. DOI: 10.1088/0957-4484/19/42/424001. URL: https://doi.org/10.1088/0957-4484/19/42/424001.

[71] Serap Günes, Helmut Neugebauer, and Niyazi Serdar Sariciftci. "Conjugated Polymer-Based Organic Solar Cells". In: *Chemical Reviews* 107.4 (2007). PMID: 17428026, pp. 1324–1338. DOI: 10.1021/cr050149z. eprint: https://doi.org/10.1021/cr050149z. URL: https://doi.org/10.1021/cr050149z.

[72] Johannes Hachmann et al. "The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid". In: *The Journal of Physical Chemistry Letters* 2.17 (2011), pp. 2241–2251. DOI: 10.1021/jz200866s. eprint: https://doi.org/10.1021/jz200866s. URL: https://doi.org/10.1021/jz200866s.

[73] Natalia Martsinovich and Alessandro Troisi. "Theoretical studies of dye-sensitised solar cells: from electronic structure to elementary processes". In: *Energy Environ. Sci.* 4 (11 2011), pp. 4473–4495. DOI: 10.1039/C1EE01906F. URL: http://dx.doi.org/10.1039/C1EE01906F.

[74] Farah Matar et al. "A new ruthenium polypyridyl dye, TG6, whose performance in dye-sensitized solar cells is surprisingly close to that of N719, the 'dye to beat' for 17 years". In: *J. Mater. Chem.* 18 (36 2008), pp. 4246–4253. DOI: 10.1039/B808255C. URL: http://dx.doi.org/10.1039/B808255C.

[75] Jeremy E. Monat, Jorge H. Rodriguez, and James K. McCusker. "Ground- and Excited-State Electronic Structures of the Solar Cell Sensitizer Bis(4,4'-dicarboxylato-2,2'-bipyridine)bis(isothiocyanato)ruthenium(II)". In: *The Journal of Physical Chemistry A* 106.32 (2002), pp. 7399–7406. DOI: 10.1021/jp020927g. eprint: https://doi.org/10.1021/jp020927g. URL: https://doi.org/10.1021/jp020927g.

[76] K. Srikanth, V. R. Marathe, and Manoj K. Mishra. "Role of electronic structure of ruthenium polypyridyl dyes in the photoconversion efficiency of dye-sensitized solar cells: Semiempirical investigation". In: *International Journal of Quantum Chemistry* 89.6 (2002), pp. 535–549. DOI: https://doi.org/10.1002/qua.10321. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.10321. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.10321.

[77] Jean-François Guillemoles et al. "A Theoretical Investigation of the Ground and Excited States of Selected Ru and Os Polypyridyl Molecular Dyes". In: *The Journal of Physical Chemistry A* 106.46 (2002), pp. 11354–11360. DOI: 10.1021/jp021517v. eprint: https://doi.org/10.1021/jp021517v. URL: https://doi.org/10.1021/jp021517v.

[78] Simona Fantacci, Filippo De Angelis, and Annabella Selloni. "Absorption Spectrum and Solvatochromism of the [Ru(4,4'-COOH-2,2'-bpy)2(NCS)2] Molecular Dye by Time Dependent Density Functional Theory". In: *Journal of the American Chemical Society* 125.14 (2003). PMID: 12670263, pp. 4381–4387. DOI: 10.1021/ja0207910. eprint: https://doi.org/10.1021/ja0207910. URL: https://doi.org/10.1021/ja0207910.

[79] Fumihiko Aiga and Tsukasa Tada. "Molecular and electronic structures of black dye; an efficient sensitizing dye for nanocrystalline TiO2 solar cells". In: *Journal of Molecular Structure* 658.1 (2003), pp. 25–32. ISSN: 0022-2860. DOI: https://doi.org/10.1016/S0022-2860(03)00408-3. URL: https://www.sciencedirect.com/science/article/pii/S0022286003004083.

[80] Sutapa Ghosh et al. "Electronic Structures and Absorption Spectra of Linkage Isomers of Trithiocyanato (4,4',4"-Tricarboxy-2,2':6,2"-terpyridine) Ruthenium(II) Complexes: A DFT Study". In: *Inorganic Chemistry* 45.19 (2006). PMID: 16961351, pp. 7600–7611. DOI: 10.1021/ic051851g. eprint: https://doi.org/10.1021/ic051851g. URL: https://doi.org/10.1021/ic051851g.

[81] Ming-Xia Li et al. "Theoretical Studies of the Electronic Structure and Spectroscopic Properties of $[Ru(Htcterpy)(NCS)_3]^{3-}$". In: *European Journal of Inorganic Chemistry* 2007.15 (2007), pp. 2171–2180. DOI: https://doi.org/10.1002/ejic.200601052. eprint: https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/ejic.200601052. URL: https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/ejic.200601052.

[82] Ming-Xia Li et al. "Theoretical Studies on Structures and Spectroscopic Properties of Photoelectrochemical Cell Ruthenium Sensitizers, $[Ru(H_m tcterpy)(NCS)_3]^{n-}$ (m=0,1,2, and 3; n=4,3,2, and 1)". In: *Inorganic Chemistry* 47.7 (2008). PMID: 18311890, pp. 2312–2324. DOI: 10.1021/ic070139b. eprint: https://doi.org/10.1021/ic070139b. URL: https://doi.org/10.1021/ic070139b.

[83] M. K. Nazeeruddin et al. "Conversion of light to electricity by cis-X2bis(2,2'-bipyridyl-4,4'-dicarboxylate)ruthenium(II) charge-transfer sensitizers (X = Cl-, Br-, I-, CN-, and SCN-) on nanocrystalline titanium dioxide electrodes". In: *Journal of the American Chemical Society* 115.14 (1993), pp. 6382–6390. DOI: 10.1021/ja00067a063. eprint: https://doi.org/10.1021/ja00067a063. URL: https://doi.org/10.1021/ja00067a063.

[84] Karim T. Mukaddem et al. "Dye-Anchoring Modes at the Dye...TiO2 Interface of N3- and N749-Sensitized Solar Cells Revealed by Glancing-Angle Pair Distribution Function Analysis". In: *The Journal of Physical Chemistry C* 124.22 (2020), pp. 11935–11945. DOI: 10.1021/acs.jpcc.0c02314. eprint: https://doi.org/10.1021/acs.jpcc.0c02314. URL: https://doi.org/10.1021/acs.jpcc.0c02314.

[85] P. Jeffrey Hay and Willard R. Wadt. "Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg". In: *The Journal of Chemical Physics* 82.1 (1985), pp. 270–283. DOI: 10.1063/1.448799. eprint: https://doi.org/10.1063/1.448799. URL: https://doi.org/10.1063/1.448799.

[86] Willard R. Wadt and P. Jeffrey Hay. "Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi". In: *The Journal of Chemical Physics* 82.1 (1985), pp. 284–298. DOI: 10.1063/1.448800. eprint: https://doi.org/10.1063/1.448800. URL: https://doi.org/10.1063/1.448800.

[87] P. Jeffrey Hay and Willard R. Wadt. "Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals". In: *The Journal of Chemical Physics* 82.1 (1985), pp. 299–310. DOI: 10.1063/1.448975. eprint: https://doi.org/10.1063/1.448975. URL: https://doi.org/10.1063/1.448975.

[88] W. Taouali et al. "Theoretical design of new small molecules with a low band-gap for organic solar cell applications: DFT and TD-DFT study". In: *Computational Materials Science* 150 (2018), pp. 54–61. ISSN: 0927-0256. DOI: https://doi.org/10.1016/j.commatsci.2018.03.038. URL: https://www.sciencedirect.com/science/article/pii/S0927025618301915.

[89] Mariachiara Pastore, Simona Fantacci, and Filippo De Angelis. "Modeling Excited States and Alignment of Energy Levels in Dye-Sensitized Solar Cells: Successes, Failures, and Challenges". In: *The Journal of Physical Chemistry C* 117.8 (2013), pp. 3685–3700. DOI: 10.1021/jp3095227. eprint: https://doi.org/10.1021/jp3095227. URL: https://doi.org/10.1021/jp3095227.

[90] T. Helgaker, P. Jörgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 143–146.

[91] T. Helgaker, P. Jörgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 35–38.

[92] Danny L. Yeager et al. "Newton-Raphson approaches and generalizations in multiconfigurational self-consistent field calculations". In: *The Journal of Physical Chemistry* 86.12 (1982), pp. 2140–2153. DOI: 10.1021/j100209a006. eprint: https://doi.org/10.1021/j100209a006. URL: https://doi.org/10.1021/j100209a006.

[93] Mary McCourt and James W. McIver Jr. "On the SCF calculation of excited states: Singlet states in the two-electron problem". In: *Journal of Computational Chemistry* 8.4 (1987), pp. 454–458. DOI: https://doi.org/10.1002/jcc.540080424. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.540080424. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540080424.

[94] M. Lewin. "On the computation of excited states with MCSCF methods". In: *Journal of Mathematical Chemistry volume* 44 (2008), pp. 967–980. DOI: https://doi.org/10.1007/s10910-008-9355-x. URL: https://doi.org/10.1007/s10910-008-9355-x.

[95] M D Segall et al. "First-principles simulation: ideas, illustrations and the CASTEP code". In: *Journal of Physics: Condensed Matter* 14.11 (2002), pp. 2717–2744. DOI: 10.1088/0953-8984/14/11/301. URL: https://doi.org/10.1088/0953-8984/14/11/301.

[96] Leo P. Kadanoff. "More is the Same; Phase Transitions and Mean Field Theories". In: *Journal of Statistical Physics* 137.11 (2009), pp. 777–797. DOI: 10.1007/s10955-009-9814-1. URL: https://doi.org/10.1007/s10955-009-9814-1.

[97] David Chandler. *Introduction to modern statistical mechanics*. New York: Oxford University Press, 1987.

[98] T. Helgaker, P. Jörgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 433–522.

[99] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Phys. Rev.* 140 (4A 1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133. URL: https://link.aps.org/doi/10.1103/PhysRev.140.A1133.

[100] Kieron Burke. "Perspective on density functional theory". In: *The Journal of Chemical Physics* 136.15 (2012), p. 150901. DOI: 10.1063/1.4704546. eprint: https://doi.org/10.1063/1.4704546. URL: https://doi.org/10.1063/1.4704546.

[101] T. Helgaker, P. Jörgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 192–196.

[102] John A. Pople, J. Stephen Binkley, and Rolf Seeger. "Theoretical models incorporating electron correlation". In: *International Journal of Quantum Chemistry* 10.S10 (1976), pp. 1–19. DOI: https://doi.org/10.1002/qua.560100802. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560100802. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560100802.

[103] R J Bartlett. "Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules". In: *Annual Review of Physical Chemistry* 32.1 (1981), pp. 359–401. DOI: 10.1146/annurev.pc.32.100181.002043. eprint: https://doi.org/10.1146/annurev.pc.32.100181.002043. URL: https://doi.org/10.1146/annurev.pc.32.100181.002043.

[104] Rodney J. Bartlett and George D. Purvis. "Many-body perturbation theory, coupled-pair many-electron theory, and the importance of quadruple excitations for the correlation problem". In: *International Journal of Quantum Chemistry* 14.5 (1978), pp. 561–581. DOI: https://doi.org/10.1002/qua.560140504. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560140504. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560140504.

[105] Marcel Nooijen, K. R. Shamasundar, and Debashis Mukherjee. "Reflections on size-extensivity, size-consistency and generalized extensivity in many-body theory". In: *Molecular Physics* 103.15-16 (2005), pp. 2277–2298. DOI: 10.1080/00268970500083952. eprint: https://doi.org/10.1080/00268970500083952. URL: https://doi.org/10.1080/00268970500083952.

[106] John A. Pople. "Nobel Lecture: Quantum chemical models". In: *Rev. Mod. Phys.* 71 (5 1999), pp. 1267–1274. DOI: 10.1103/RevModPhys.71.1267. URL: https://link.aps.org/doi/10.1103/RevModPhys.71.1267.

[107] Mark S. Gordon and Theresa L. Windus. "Editorial: Modern Architectures and Their Impact on Electronic Structure Theory". In: *Chemical Reviews* 120.17 (2020). PMID: 32900196, pp. 9015–9020. DOI: 10.1021/acs.chemrev.0c00700. eprint: https://doi.org/10.1021/acs.chemrev.0c00700. URL: https://doi.org/10.1021/acs.chemrev.0c00700.

[108] T. Helgaker, P. Jörgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 523–550.

[109] Andreas Dreuw and Martin Head-Gordon. "Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules". In: *Chemical Reviews* 105.11 (2005). PMID: 16277369, pp. 4009–4037. DOI: 10.1021/cr0505627. eprint: https://doi.org/10.1021/cr0505627. URL: https://doi.org/10.1021/cr0505627.

[110] C. David Sherrill. "Frontiers in electronic structure theory". In: *The Journal of Chemical Physics* 132.11 (2010), p. 110902. DOI: 10.1063/1.3369628. eprint: https://doi.org/10.1063/1.3369628. URL: https://doi.org/10.1063/1.3369628.

[111] Carlos A. Jiménez-Hoyos, R. Rodríguez-Guzmán, and Gustavo E. Scuseria. "Excited electronic states from a variational approach based on symmetry-projected Hartree–Fock configurations". In: *The Journal of Chemical Physics* 139.22 (2013), p. 224110. DOI: 10.1063/1.4840097. eprint: https://doi.org/10.1063/1.4840097. URL: https://doi.org/10.1063/1.4840097.

[112] Björn O. Roos, Peter R. Taylor, and Per E.M. Sigbahn. "A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach". In: *Chemical Physics* 48.2 (1980), pp. 157–173. ISSN: 0301-0104. DOI: https://doi.org/10.1016/0301-0104(80)80045-0. URL: https://www.sciencedirect.com/science/article/pii/0301010480800450.

[113] Julien Toulouse, Roland Assaraf, and Cyrus J Umrigar. "Introduction to the variational and diffusion Monte Carlo methods". In: *Adv. Quantum Chem.* 73 (2016), pp. 285–314.

[114] U. Schollwöck. "The density-matrix renormalization group". In: *Rev. Mod. Phys.* 77 (1 2005), pp. 259–315. DOI: 10.1103/RevModPhys.77.259. URL: https://link.aps.org/doi/10.1103/RevModPhys.77.259.

[115] James B. Foresman et al. "Toward a systematic molecular orbital theory for excited states". In: *The Journal of Physical Chemistry* 96.1 (1992), pp. 135–149. DOI: 10.1021/j100180a030. eprint: https://doi.org/10.1021/j100180a030. URL: https://doi.org/10.1021/j100180a030.

[116] 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. DOI: 10.1063/1.5001262.

[117] Jacqueline A. R. Shea and Eric Neuscamman. "Communication: A mean field platform for excited state quantum chemistry". In: *J. Chem. Phys.* 149 (2018), p. 081101. DOI: 10.1063/1.5045056.

[118] T. Helgaker, P. Jörgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 183–195.

[119] Dimitrios A. Pantazis Maylis Orio and Frank Neese. "Density functional theory". In: *Photosynthesis Research* 102 (2009), pp. 443–453. DOI: 10.1007/s11120-009-9404-8. URL: https://doi.org/10.1007/s11120-009-9404-8.

[120] Narbe Mardirossian and Martin Head-Gordon. "Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals". In: *Mol. Phys.* 115.19 (2017), pp. 2315–2372. DOI: 10.1080/00268976.2017.1333644. eprint: https://doi.org/10.1080/00268976.2017.1333644. URL: https://doi.org/10.1080/00268976.2017.1333644.

[121] Konstantinos D. Vogiatzis et al. "Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations". In: *The Journal of Chemical Physics* 147.18 (2017), p. 184111. DOI: 10.1063/1.4989858. eprint: https://doi.org/10.1063/1.4989858. URL: https://doi.org/10.1063/1.4989858.

[122] T. Helgaker, P. Jörgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 181–186.

[123] Christine M. Isborn et al. "Excited-State Electronic Structure with Configuration Interaction Singles and Tamm–Dancoff Time-Dependent Density Functional Theory on Graphical Processing Units". In: *Journal of Chemical Theory and Computation* 7.6 (2011). PMID: 21687784, pp. 1814–1823. DOI: 10.1021/ct200030k. eprint: https://doi.org/10.1021/ct200030k. URL: https://doi.org/10.1021/ct200030k.

[124] So Hirata, Martin Head-Gordon, and Rodney J. Bartlett. "Configuration interaction singles, time-dependent Hartree–Fock, and time-dependent density functional theory for the electronic excited states of extended systems". In: *The Journal of Chemical Physics* 111.24 (1999), pp. 10774–10786. DOI: 10.1063/1.480443. eprint: https://doi.org/10.1063/1.480443. URL: https://doi.org/10.1063/1.480443.

[125] John F. Stanton et al. "A comparison of single reference methods for characterizing stationary points of excited state potential energy surfaces". In: *The Journal of Chemical Physics* 103.10 (1995), pp. 4160–4174. DOI: 10.1063/1.469601. eprint: https://doi.org/10.1063/1.469601. URL: https://doi.org/10.1063/1.469601.

[126] Lars Goerigk and Stefan Grimme. "Assessment of TD-DFT methods and of various spin scaled CIS(D) and CC2 versions for the treatment of low-lying valence excitations of large organic dyes". In: *The Journal of Chemical Physics* 132.18 (2010), p. 184103. DOI: 10.1063/1.3418614. eprint: https://doi.org/10.1063/1.3418614. URL: https://doi.org/10.1063/1.3418614.

[127] 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.

[128] Xinle Liu et al. "Communication: Adjusting charge transfer state energies for configuration interaction singles: Without any parameterization and with minimal cost". In: *The Journal of Chemical Physics* 136.16 (2012), p. 161101. DOI: 10.1063/1.4705757. eprint: https://doi.org/10.1063/1.4705757. URL: https://doi.org/10.1063/1.4705757.

[129] 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.* 219.1-2 (1994), pp. 21–29.

[130] T. Helgaker, P. Jörgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 176–180.

[131] Hans-Joachim Werner and Wilfried Meyer. "A quadratically convergent MCSCF method for the simultaneous optimization of several states". In: *J. Chem. Phys.* 74.10 (1981), pp. 5794–5801. DOI: 10.1063/1.440892.

[132] Lan Nguyen Tran, Jacqueline A. R. Shea, and Eric Neuscamman. "Tracking Excited States in Wave Function Optimization Using Density Matrices and Variational Principles". In: *Journal of Chemical Theory and Computation* 15.9 (2019). PMID: 31393725, pp. 4790–4803. DOI: 10.1021/acs.jctc.9b00351. eprint: https://doi.org/10.1021/acs.jctc.9b00351. URL: https://doi.org/10.1021/acs.jctc.9b00351.

[133] Lan Nguyen Tran and Eric Neuscamman. "Improving Excited-State Potential Energy Surfaces via Optimal Orbital Shapes". In: *The Journal of Physical Chemistry A* 124.40 (2020). PMID: 32885970, pp. 8273–8279. DOI: 10.1021/acs.jpca.0c07593. eprint: https://doi.org/10.1021/acs.jpca.0c07593. URL: https://doi.org/10.1021/acs.jpca.0c07593.

[134] Sebastian Keller et al. "Selection of active spaces for multiconfigurational wavefunctions". In: *The Journal of Chemical Physics* 142.24 (2015), p. 244104. DOI: 10.1063/1.4922352. eprint: https://doi.org/10.1063/1.4922352. URL: https://doi.org/10.1063/1.4922352.

[135] A. Warshel and M. Levitt. "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme". In: *Journal of Molecular Biology* 103.2 (1976), pp. 227–249. ISSN: 0022-2836. DOI: https://doi.org/10.1016/0022-2836(76)90311-9. URL: https://www.sciencedirect.com/science/article/pii/0022283676903119.

[136] Hanneli R. Hudock et al. "Ab Initio Molecular Dynamics and Time-Resolved Photoelectron Spectroscopy of Electronically Excited Uracil and Thymine". In: *The Journal of Physical Chemistry A* 111.34 (2007). PMID: 17685594, pp. 8500–8508. DOI: 10.1021/jp0723665. eprint: https://doi.org/10.1021/jp0723665. URL: https://doi.org/10.1021/jp0723665.

[137] Mario Barbatti et al. "Photodynamical simulations of cytosine: characterization of the ultrafast bi-exponential UV deactivation". In: *Phys. Chem. Chem. Phys.* 13 (13 2011), pp. 6145–6155. DOI: 10.1039/C0CP01327G. URL: http://dx.doi.org/10.1039/C0CP01327G.

[138] Gerrit Groenhof et al. "Ultrafast Deactivation of an Excited Cytosine-Guanine Base Pair in DNA". In: *Journal of the American Chemical Society* 129.21 (2007). PMID: 17488008, pp. 6812–6819. DOI: 10.1021/ja069176c. eprint: https://doi.org/10.1021/ja069176c. URL: https://doi.org/10.1021/ja069176c.

[139] Clemens Rauer et al. "Cyclobutane Thymine Photodimerization Mechanism Revealed by Nonadiabatic Molecular Dynamics". In: *Journal of the American Chemical Society* 138.49 (2016). PMID: 27682199, pp. 15911–15916. DOI: 10.1021/jacs.6b06701. eprint: https://doi.org/10.1021/jacs.6b06701. URL: https://doi.org/10.1021/jacs.6b06701.

[140] Mario Barbatti. "Nonadiabatic dynamics with trajectory surface hopping method". In: *WIREs Computational Molecular Science* 1.4 (2011), pp. 620–633. DOI: https://doi.org/10.1002/wcms.64. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.64. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.64.

[141] Martin Richter et al. "Femtosecond Intersystem Crossing in the DNA Nucleobase Cytosine". In: *The Journal of Physical Chemistry Letters* 3.21 (2012). PMID: 26296011, pp. 3090–3095. DOI: 10.1021/jz301312h. eprint: https://doi.org/10.1021/jz301312h. URL: https://doi.org/10.1021/jz301312h.

[142] Martin Richter et al. "Ultrafast intersystem crossing dynamics in uracil unravelled by ab initio molecular dynamics". In: *Phys. Chem. Chem. Phys.* 16 (44 2014), pp. 24423–24436. DOI: 10.1039/C4CP04158E. URL: http://dx.doi.org/10.1039/C4CP04158E.

[143] Sebastian Mai et al. "The DNA nucleobase thymine in motion – Intersystem crossing simulated with surface hopping". In: *Chemical Physics* 482 (2017). Electrons and nuclei in motion - correlation and dynamics in molecules (on the occasion of the 70th birthday of Lorenz S. Cederbaum), pp. 9–15. ISSN: 0301-0104. DOI: https://doi.org/10.1016/j.chemphys.2016.10.003. URL: https://www.sciencedirect.com/science/article/pii/S0301010416305092.

[144] Kerstin Andersson et al. "Second-order perturbation theory with a CASSCF reference function". In: *The Journal of Physical Chemistry* 94.14 (1990), pp. 5483–5488. DOI: 10.1021/j100377a012. eprint: https://doi.org/10.1021/j100377a012. URL: https://doi.org/10.1021/j100377a012.

[145] Kerstin Andersson, Per-Åke Malmqvist, and Björn O. Roos. "Second-order perturbation theory with a complete active space self-consistent field reference function". In: *The Journal of Chemical Physics* 96.2 (1992), pp. 1218–1226. DOI: 10.1063/1.462209. eprint: https://doi.org/10.1063/1.462209. URL: https://doi.org/10.1063/1.462209.

[146] James Finley et al. "The multi-state CASPT2 method". In: *Chemical Physics Letters* 288.2 (1998), pp. 299–306. ISSN: 0009-2614. DOI: https://doi.org/10.1016/S0009-2614(98)00252-8. URL: https://www.sciencedirect.com/science/article/pii/S0009261498002528.

[147] Per-Åke Malmqvist et al. "The restricted active space followed by second-order perturbation theory method: Theory and application to the study of CuO2 and Cu2O2 systems". In: *The Journal of Chemical Physics* 128.20 (2008), p. 204109. DOI: 10.1063/1.2920188. eprint: https://doi.org/10.1063/1.2920188. URL: https://doi.org/10.1063/1.2920188.

[148] Robert J. Buenker and Sigrid D. Peyerimhoff. "Individualized configuration selection in CI calculations with subsequent energy extrapolation". In: *Theoretica chimica acta* 35 (1974), pp. 33–58. DOI: https://doi-org.libproxy.berkeley.edu/10.1007/BF02394557.

[149] Hans-Joachim Werner and Peter J. Knowles. "An efficient internally contracted multiconfiguration-reference configuration interaction method". In: *The Journal of Chemical Physics* 89.9 (1988), pp. 5803–5814. DOI: 10.1063/1.455556. eprint: https://doi.org/10.1063/1.455556. URL: https://doi.org/10.1063/1.455556.

[150] Peter J. Knowles and Hans-Joachim Werner. "Internally contracted multiconfiguration-reference configuration interaction calculations for excited states". In: *Theoretica chimica acta* 84 (1992), pp. 95–103. DOI: https://doi-org.libproxy.berkeley.edu/10.1007/BF01117405.

[151] Mario Barbatti et al. "Relaxation mechanisms of UV-photoexcited DNA and RNA nucleobases". In: *Proceedings of the National Academy of Sciences* 107.50 (2010), pp. 21453–21458. ISSN: 0027-8424. DOI: 10.1073/pnas.1014982107. eprint: https://www.pnas.org/content/107/50/21453.full.pdf. URL: https://www.pnas.org/content/107/50/21453.

[152] Cristopher Camacho, Shigeyoshi Yamamoto, and Henryk A. Witek. "Choosing a proper complete active space in calculations for transition metal dimers: ground state of Mn2 revisited". In: *Phys. Chem. Chem. Phys.* 10 (33 2008), pp. 5128–5134. DOI: 10.1039/B805125A. URL: http://dx.doi.org/10.1039/B805125A.

[153] Bruno Senjean et al. "Generalization of Intrinsic Orbitals to Kramers-Paired Quaternion Spinors, Molecular Fragments, and Valence Virtual Spinors". In: *Journal of Chemical Theory and Computation* 17.3 (2021). PMID: 33555866, pp. 1337–1354. DOI: 10.1021/acs.jctc.0c00964. eprint: https://doi.org/10.1021/acs.jctc.0c00964. URL: https://doi.org/10.1021/acs.jctc.0c00964.

[154] Tommaso Giovannini and Henrik Koch. "Energy-Based Molecular Orbital Localization in a Specific Spatial Region". In: *Journal of Chemical Theory and Computation* 17.1 (2021). PMID: 33337150, pp. 139–150. DOI: 10.1021/acs.jctc.0c00737. eprint: https://doi.org/10.1021/acs.jctc.0c00737. URL: https://doi.org/10.1021/acs.jctc.0c00737.

[155] Luis Serrano-Andrés, Manuela Merchán, and Roland Lindh. "Computation of conical intersections by using perturbation techniques". In: *The Journal of Chemical Physics* 122.10 (2005), p. 104107. DOI: 10.1063/1.1866096. eprint: https://doi.org/10.1063/1.1866096. URL: https://doi.org/10.1063/1.1866096.

[156] C. Angeli, M. Pastore, and R. Cimiraglia. "New perspectives in multireference perturbation theory: the n-electron valence state approach". In: *Theoretical Chemistry Accounts* 117 (2007), pp. 743–754. DOI: 10.1007/s00214-006-0207-0. URL: https://doi.org/10.1007/s00214-006-0207-0.

[157] C. Angeli et al. "Introduction of n-electron valence states for multireference perturbation theory". In: *The Journal of Chemical Physics* 114.23 (2001), pp. 10252–10264. DOI: 10.1063/1.1361246. eprint: https://doi.org/10.1063/1.1361246. URL: https://doi.org/10.1063/1.1361246.

[158] Celestino Angeli, Renzo Cimiraglia, and Jean-Paul Malrieu. "N-electron valence state perturbation theory: a fast implementation of the strongly contracted variant". In: *Chemical Physics Letters* 350.3 (2001), pp. 297–305. ISSN: 0009-2614. DOI: https://doi.org/10.1016/S0009-2614(01)01303-3. URL: https://www.sciencedirect.com/science/article/pii/S0009261401013033.

[159] E. Goulielmakis et al. "Direct Measurement of Light Waves". In: *Science* 305.5688 (2004), pp. 1267–1269. ISSN: 0036-8075. DOI: 10.1126/science.1100866. eprint: https://science.sciencemag.org/content/305/5688/1267.full.pdf. URL: https://science.sciencemag.org/content/305/5688/1267.

[160] Thomas Bondo Pedersen. "Introduction to Response Theory". In: *Handbook of Computational Chemistry*. Ed. by Jerzy Leszczynski. Dordrecht: Springer Netherlands, 2012, pp. 135–156. ISBN: 978-94-007-0711-5. DOI: 10.1007/978-94-007-0711-5_5. URL: https://doi.org/10.1007/978-94-007-0711-5_5.

[161] Varun Rishi et al. "Excited states from modified coupled cluster methods: Are they any better than EOM CCSD?" In: *The Journal of Chemical Physics* 146.14 (2017), p. 144104. DOI: 10.1063/1.4979078. eprint: https://doi.org/10.1063/1.4979078. URL: https://doi.org/10.1063/1.4979078.

[162] Marko Schreiber et al. "Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3". In: *The Journal of Chemical Physics* 128.13 (2008), p. 134110. DOI: 10.1063/1.2889385. eprint: https://doi.org/10.1063/1.2889385. URL: https://doi.org/10.1063/1.2889385.

[163] J. Sous, P. Goel, and M. Nooijen. "Similarity transformed equation of motion coupled cluster theory revisited: a benchmark study of valence excited states". In: *Molecular Physics* 112.5-6 (2014), pp. 616–638. DOI: 10.1080/00268976.2013.847216. eprint: https://doi.org/10.1080/00268976.2013.847216. URL: https://doi.org/10.1080/00268976.2013.847216.

[164] Nina O. C. Winter et al. "Benchmarks for 0–0 transitions of aromatic organic molecules: DFT/B3LYP, ADC(2), CC2, SOS-CC2 and SCS-CC2 compared to high-resolution gas-phase data". In: *Phys. Chem. Chem. Phys.* 15 (18 2013), pp. 6623–6630. DOI: 10.1039/C2CP42694C. URL: http://dx.doi.org/10.1039/C2CP42694C.

[165] Piotr Piecuch, Jared A. Hansen, and Adeayo O. Ajala. "Benchmarking the completely renormalised equation-of-motion coupled-cluster approaches for vertical excitation energies". In: *Molecular Physics* 113.19-20 (2015), pp. 3085–3127. DOI: 10.1080/00268976.2015.1076901. eprint: https://doi.org/10.1080/00268976.2015.1076901. URL: https://doi.org/10.1080/00268976.2015.1076901.

[166] Henrik Koch et al. "Calculation of size-intensive transition moments from the coupled cluster singles and doubles linear response function". In: *The Journal of Chemical Physics* 100.6 (1994), pp. 4393–4400. DOI: 10.1063/1.466321. eprint: https://doi.org/10.1063/1.466321. URL: https://doi.org/10.1063/1.466321.

[167] T. Helgaker, P. Jörgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, p. 678.

[168] Sergey V. Levchenko. "Equation-of-Motion Coupled Cluster Model with Single and Double Substitutions: Theory and Applications". PhD thesis. Los Angeles, CA: University of Southern California, 2005.

[169] Jesse J. Lutz et al. "Reference dependence of the two-determinant coupled-cluster method for triplet and open-shell singlet states of biradical molecules". In: *The Journal of Chemical Physics* 148.16 (2018), p. 164102. DOI: 10.1063/1.5025170. eprint: https://doi.org/10.1063/1.5025170. URL: https://doi.org/10.1063/1.5025170.

[170] Péter G. Szalay et al. "Benchmark Studies on the Building Blocks of DNA. 1. Superiority of Coupled Cluster Methods in Describing the Excited States of Nucleobases in the Franck-Condon Region". In: *The Journal of Physical Chemistry A* 116.25 (2012). PMID: 22587574, pp. 6702–6710. DOI: 10.1021/jp300977a. eprint: https://doi.org/10.1021/jp300977a. URL: https://doi.org/10.1021/jp300977a.

[171] Giulio Pepe et al. "Molecular engineering of cyanine dyes to design a panchromatic response in co-sensitized dye-sensitized solar cells". In: *Mol. Syst. Des. Eng.* 1 (1 2016), pp. 86–98. DOI: 10.1039/C6ME00014B. URL: http://dx.doi.org/10.1039/C6ME00014B.

[172] Sijin Ren, Joseph Harms, and Marco Caricato. "An EOM-CCSD-PCM Benchmark for Electronic Excitation Energies of Solvated Molecules". In: *Journal of Chemical Theory and Computation* 13.1 (2017). PMID: 27973775, pp. 117–124. DOI: 10.1021/acs.jctc.6b01053. eprint: https://doi.org/10.1021/acs.jctc.6b01053. URL: https://doi.org/10.1021/acs.jctc.6b01053.

[173] Aron J. Cohen, Paula Mori-Sánchez, and Weitao Yang. "Challenges for Density Functional Theory". In: *Chemical Reviews* 112.1 (2012), pp. 289–320. DOI: 10.1021/cr200107z. eprint: https://doi.org/10.1021/cr200107z. URL: https://doi.org/10.1021/cr200107z.

[174] Gary S. Kedziora et al. "Bond breaking in stretched molecules: multi-reference methods versus density functional theory". In: *Theoretical Chemistry Accounts* 135.3 (2016), p. 79. DOI: 10.1007/s00214-016-1822-z. URL: https://doi.org/10.1007/s00214-016-1822-z.

[175] Yan Zhao, Nathan E. Schultz, and Donald G. Truhlar. "Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions". In: *Journal of Chemical Theory and Computation* 2.2 (2006). PMID: 26626525, pp. 364–382. DOI: 10.1021/ct0502763. eprint: https://doi.org/10.1021/ct0502763. URL: https://doi.org/10.1021/ct0502763.

[176] Roberto Improta. "Photophysics and Photochemistry of Thymine Deoxy-Dinucleotide in Water: A PCM/TD-DFT Quantum Mechanical Study". In: *The Journal of Physical Chemistry B* 116.49 (2012). PMID: 23134336, pp. 14261–14274. DOI: 10.1021/jp3093385. eprint: https://doi.org/10.1021/jp3093385. URL: https://doi.org/10.1021/jp3093385.

[177] Jeng-Da Chai and Martin Head-Gordon. "Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections". In: *Phys. Chem. Chem. Phys.* 10 (44 2008), pp. 6615–6620. DOI: 10.1039/B810189B. URL: http://dx.doi.org/10.1039/B810189B.

[178] Shaohong L. Li and Donald G. Truhlar. "Improving Rydberg Excitations within Time-Dependent Density Functional Theory with Generalized Gradient Approximations: The Exchange-Enhancement-for-Large-Gradient Scheme". In: *Journal of Chemical Theory and Computation* 11.7 (2015). PMID: 26575749, pp. 3123–3130. DOI: 10.1021/acs.jctc.5b00369. eprint: https://doi.org/10.1021/acs.jctc.5b00369. URL: https://doi.org/10.1021/acs.jctc.5b00369.

[179] R. van Meer, O. V. Gritsenko, and E. J. Baerends. "Physical Meaning of Virtual Kohn–Sham Orbitals and Orbital Energies: An Ideal Basis for the Description of Molecular Excitations". In: *Journal of Chemical Theory and Computation* 10.10 (2014). PMID: 26588140, pp. 4432–4441. DOI: 10.1021/ct500727c. eprint: https://doi.org/10.1021/ct500727c. URL: https://doi.org/10.1021/ct500727c.

[180] Mark E. Casida et al. "Molecular excitation energies to high-lying bound states from time-dependent density-functional response theory: Characterization and correction of the time-dependent local density approximation ionization threshold". In: *The Journal of Chemical Physics* 108.11 (1998), pp. 4439–4449. DOI: 10.1063/1.475855. eprint: https://doi.org/10.1063/1.475855. URL: https://doi.org/10.1063/1.475855.

[181] Mark E. Casida, Kim C. Casida, and Dennis R. Salahub. "Excited-state potential energy curves from time-dependent density-functional theory: A cross section of formaldehyde's $^1A_1$ manifold". In: *International Journal of Quantum Chemistry* 70.4-5 (1998), pp. 933–941. DOI: https://doi.org/10.1002/(SICI)1097-461X(1998)70:4/5<933::AID-QUA39>3.0.CO;2-Z.

[182] Robert J. Cave et al. "A dressed TDDFT treatment of the 21Ag states of butadiene and hexatriene". In: *Chemical Physics Letters* 389.1 (2004), pp. 39–42. ISSN: 0009-2614. DOI: https://doi.org/10.1016/j.cplett.2004.03.051. URL: https://www.sciencedirect.com/science/article/pii/S0009261404004142.

[183] Andreas Dreuw and Martin Head-Gordon. "Failure of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excited States: The Zincbacteriochlorin-Bacteriochlorin and Bacteriochlorophyll-Spheroidene Complexes". In: *Journal of the American Chemical Society* 126.12 (2004). PMID: 15038755, pp. 4007–4016. DOI: 10.1021/ja039556n. eprint: https://doi.org/10.1021/ja039556n. URL: https://doi.org/10.1021/ja039556n.

[184] Julien Toulouse, Fran çois Colonna, and Andreas Savin. "Long-range–short-range separation of the electron-electron interaction in density-functional theory". In: *Phys. Rev. A* 70 (6 2004), p. 062505. DOI: 10.1103/PhysRevA.70.062505. URL: https://link.aps.org/doi/10.1103/PhysRevA.70.062505.

[185] Thomas M. Henderson, Benjamin G. Janesko, and Gustavo E. Scuseria. "Range Separation and Local Hybridization in Density Functional Theory". In: *The Journal of Physical Chemistry A* 112.49 (2008). PMID: 19006280, pp. 12530–12542. DOI: 10.1021/jp806573k. eprint: https://doi.org/10.1021/jp806573k. URL: https://doi.org/10.1021/jp806573k.

[186] Andreas Savin and Heinz-Jürgen Flad. "Density functionals for the Yukawa electron-electron interaction". In: *International Journal of Quantum Chemistry* 56.4 (1995), pp. 327–332. DOI: https://doi.org/10.1002/qua.560560417. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560560417. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560560417.

[187] A. Savin. "On degeneracy, near-degeneracy and density functional theory". In: *Recent Developments and Applications of Modern Density Functional Theory*. Ed. by J.M. Seminario. Vol. 4. Theoretical and Computational Chemistry. Elsevier, 1996, pp. 327–357. DOI: https://doi.org/10.1016/S1380-7323(96)80091-4. URL: https://www.sciencedirect.com/science/article/pii/S1380732396800914.

[188] Thierry Leininger et al. "Combining long-range configuration interaction with short-range density functionals". In: *Chemical Physics Letters* 275.3 (1997), pp. 151–160. ISSN: 0009-2614. DOI: https://doi.org/10.1016/S0009-2614(97)00758-6. URL: https://www.sciencedirect.com/science/article/pii/S0009261497007586.

[189] Roi Baer, Ester Livshits, and Ulrike Salzner. "Tuned Range-Separated Hybrids in Density Functional Theory". In: *Annual Review of Physical Chemistry* 61.1 (2010). PMID: 20055678, pp. 85–109. DOI: 10.1146/annurev.physchem.012809.103321. eprint: https://doi.org/10.1146/annurev.physchem.012809.103321. URL: https://doi.org/10.1146/annurev.physchem.012809.103321.

[190] Xu Zhang et al. "Linear-Response Time-Dependent Density Functional Theory with Stochastic Range-Separated Hybrids". In: *Journal of Chemical Theory and Computation* 16.2 (2020). PMID: 31899638, pp. 1064–1072. DOI: 10.1021/acs.jctc.9b01121. eprint: https://doi.org/10.1021/acs.jctc.9b01121. URL: https://doi.org/10.1021/acs.jctc.9b01121.

[191] Daniel Neuhauser et al. "Stochastic Optimally Tuned Range-Separated Hybrid Density Functional Theory". In: *The Journal of Physical Chemistry A* 120.19 (2016). PMID: 26651840, pp. 3071–3078. DOI: 10.1021/acs.jpca.5b10573. eprint: https://doi.org/10.1021/acs.jpca.5b10573. URL: https://doi.org/10.1021/acs.jpca.5b10573.

[192] Leeor Kronik et al. "Excitation Gaps of Finite-Sized Systems from Optimally Tuned Range-Separated Hybrid Functionals". In: *Journal of Chemical Theory and Computation* 8.5 (2012). PMID: 26593646, pp. 1515–1531. DOI: 10.1021/ct2009363. eprint: https://doi.org/10.1021/ct2009363. URL: https://doi.org/10.1021/ct2009363.

[193] Éric Brémond et al. "Assessing challenging intra- and inter-molecular charge-transfer excitations energies with double-hybrid density functionals". In: *Journal of Computational Chemistry* (). DOI: https://doi.org/10.1002/jcc.26517. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.26517. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.26517.

[194] Erik Donovan Hedegård et al. "Assessment of charge-transfer excitations with time-dependent, range-separated density functional theory based on long-range MP2 and multiconfigurational self-consistent field wave functions". In: *The Journal of Chemical Physics* 139.18 (2013), p. 184308. DOI: 10.1063/1.4826533. eprint: https://doi.org/10.1063/1.4826533. URL: https://doi.org/10.1063/1.4826533.

[195] Stephan Kümmel. "Charge-Transfer Excitations: A Challenge for Time-Dependent Density Functional Theory That Has Been Met". In: *Advanced Energy Materials* 7.16 (2017), p. 1700440. DOI: https://doi.org/10.1002/aenm.201700440. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.201700440. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/aenm.201700440.

[196] Tamar Stein, Leeor Kronik, and Roi Baer. "Reliable Prediction of Charge Transfer Excitations in Molecular Complexes Using Time-Dependent Density Functional Theory". In: *Journal of the American Chemical Society* 131.8 (2009). PMID: 19239266, pp. 2818–2820. DOI: 10.1021/ja8087482. eprint: https://doi.org/10.1021/ja8087482. URL: https://doi.org/10.1021/ja8087482.

[197] Neepa T Maitra. "Charge transfer in time-dependent density functional theory". In: *Journal of Physics: Condensed Matter* 29.42 (2017), p. 423001. DOI: 10.1088/1361-648x/aa836e. URL: https://doi.org/10.1088/1361-648x/aa836e.

[198] Young Choon Park, Mykhaylo Krykunov, and Tom Ziegler. "On the relation between adiabatic time dependent density functional theory (TDDFT) and the ΔSCF-DFT method. Introducing a numerically stable ΔSCF-DFT scheme for local functionals based on constricted variational DFT". In: *Molecular Physics* 113.13-14 (2015), pp. 1636–1647. DOI: 10.1080/00268976.2014.1003260. eprint: https://doi.org/10.1080/00268976.2014.1003260. URL: https://doi.org/10.1080/00268976.2014.1003260.

[199] Tom Ziegler et al. "On the relation between time-dependent and variational density functional theory approaches for the determination of excitation energies and transition moments". In: *The Journal of Chemical Physics* 130.15 (2009), p. 154102. DOI: 10.1063/1.3114988. eprint: https://doi.org/10.1063/1.3114988. URL: https://doi.org/10.1063/1.3114988.

[200] Luning Zhao and Eric Neuscamman. "Density Functional Extension to Excited-State Mean-Field Theory". In: *Journal of Chemical Theory and Computation* 16.1 (2020). PMID: 31765142, pp. 164–178. DOI: 10.1021/acs.jctc.9b00530. eprint: https://doi.org/10.1021/acs.jctc.9b00530. URL: https://doi.org/10.1021/acs.jctc.9b00530.

[201] Lasse Jensen and Niranjan Govind. "Excited States of DNA Base Pairs Using Long-Range Corrected Time-Dependent Density Functional Theory". In: *The Journal of Physical Chemistry A* 113.36 (2009). PMID: 19678692, pp. 9761–9765. DOI: 10.1021/jp905893v. eprint: https://doi.org/10.1021/jp905893v. URL: https://doi.org/10.1021/jp905893v.

[202] Marco Caricato et al. "Oscillator Strength: How Does TDDFT Compare to EOM-CCSD?" In: *Journal of Chemical Theory and Computation* 7.2 (2011). PMID: 26596165, pp. 456–466. DOI: 10.1021/ct100662n. eprint: https://doi.org/10.1021/ct100662n. URL: https://doi.org/10.1021/ct100662n.

[203] Yan Zhao and Donald G. Truhlar. "The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals". In: *Theoretical Chemistry Accounts* 120 (2 2008), pp. 215–241. DOI: 10.1007/s00214-007-0310-x. URL: https://doi.org/10.1007/s00214-007-0310-x.

[204] Takeshi Yanai, David P. Tew, and Nicholas C. Handy. "A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP)". In: *Chemical Physics Letters* 393.1 (2004), pp. 51–57. ISSN: 0009-2614. DOI: https://doi.org/10.1016/j.cplett.2004.06.011. URL: https://www.sciencedirect.com/science/article/pii/S0009261404008620.

[205] Adélia J. A. Aquino et al. "The charge-transfer states in a stacked nucleobase dimer complex: A benchmark study". In: *Journal of Computational Chemistry* 32.7 (2011), pp. 1217–1227. DOI: https://doi.org/10.1002/jcc.21702. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21702. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21702.

[206] Alexandra E. Raeber and Bryan M. Wong. "The Importance of Short- and Long-Range Exchange on Various Excited State Properties of DNA Monomers, Stacked Complexes, and Watson-Crick Pairs". In: *Journal of Chemical Theory and Computation* 11.5 (2015). PMID: 26574420, pp. 2199–2209. DOI: 10.1021/acs.jctc.5b00105. eprint: https://doi.org/10.1021/acs.jctc.5b00105. URL: https://doi.org/10.1021/acs.jctc.5b00105.

[207] Haitao Sun et al. "Theoretical study of excited states of DNA base dimers and tetramers using optimally tuned range-separated density functional theory". In: *Journal of Computational Chemistry* 37.7 (2016), pp. 684–693. DOI: https://doi.org/10.1002/jcc.24266. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.24266. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.24266.

[208] A. D. Becke. "Density-functional exchange-energy approximation with correct asymptotic behavior". In: *Phys. Rev. A* 38 (6 1988), pp. 3098–3100. DOI: 10.1103/PhysRevA.38.3098. URL: https://link.aps.org/doi/10.1103/PhysRevA.38.3098.

[209] Chengteh Lee, Weitao Yang, and Robert G. Parr. "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density". In: *Phys. Rev. B* 37 (2 1988), pp. 785–789. DOI: 10.1103/PhysRevB.37.785. URL: https://link.aps.org/doi/10.1103/PhysRevB.37.785.

[210] Mariachiara Pastore et al. "A Computational Investigation of Organic Dyes for Dye-Sensitized Solar Cells: Benchmark, Strategies, and Open Issues". In: *The Journal of Physical Chemistry C* 114.15 (2010), pp. 7205–7212. DOI: 10.1021/jp100713r. eprint: https://doi.org/10.1021/jp100713r. URL: https://doi.org/10.1021/jp100713r.

[211] Antonín Vlček and Stanislav Záliš. "Modeling of charge-transfer transitions and excited states in d6 transition metal complexes by DFT techniques". In: *Coordination Chemistry Reviews* 251.3 (2007). A Special Issue Highlighting the Many Aspects of the Electronic Spectrosocpy of Inorganic Compounds, pp. 258–287. ISSN: 0010-8545. DOI: https://doi.org/10.1016/j.ccr.2006.05.021. URL: https://www.sciencedirect.com/science/article/pii/S0010854506001706.

[212] Daniel Escudero and Leticia González. "RASPT2/RASSCF vs Range-Separated/Hybrid DFT Methods: Assessing the Excited States of a Ru(II)bipyridyl Complex". In: *Journal of Chemical Theory and Computation* 8.1 (2012). PMID: 26592882, pp. 203–213. DOI: 10.1021/ct200640q. eprint: https://doi.org/10.1021/ct200640q. URL: https://doi.org/10.1021/ct200640q.

[213] Benjamin J. Lynch et al. "Adiabatic Connection for Kinetics". In: *The Journal of Physical Chemistry A* 104.21 (2000), pp. 4811–4815. DOI: 10.1021/jp000497z. eprint: https://doi.org/10.1021/jp000497z. URL: https://doi.org/10.1021/jp000497z.

[214] Yoshihiro Osamura et al. "Unified theoretical treatment of analytic first and second energy derivatives in open-shell Hartree—Fock theory". In: *Chemical Physics* 72.1 (1982), pp. 131–139. ISSN: 0301-0104. DOI: https://doi.org/10.1016/0301-0104(82)87073-0. URL: https://www.sciencedirect.com/science/article/pii/0301010482870730.

[215] W. Kohn, A. D. Becke, and R. G. Parr. "Density Functional Theory of Electronic Structure". In: *The Journal of Physical Chemistry* 100.31 (1996), pp. 12974–12980. DOI: 10.1021/jp9606691. eprint: https://doi.org/10.1021/jp9606691. URL: https://doi.org/10.1021/jp9606691.

[216] Tim Kowalczyk et al. "Excitation energies and Stokes shifts from a restricted open-shell Kohn-Sham approach". In: *The Journal of Chemical Physics* 138.16 (2013), p. 164101. DOI: 10.1063/1.4801790. eprint: https://doi.org/10.1063/1.4801790. URL: https://doi.org/10.1063/1.4801790.

[217] Tim Kowalczyk, Shane R. Yost, and Troy Van Voorhis. "Assessment of the Δ SCF density functional theory approach for electronic excitations in organic dyes". In: *The Journal of Chemical Physics* 134.5 (2011), p. 054128. DOI: 10.1063/1.3530801. eprint: https://doi.org/10.1063/1.3530801. URL: https://doi.org/10.1063/1.3530801.

[218] Tom Ziegler, Arvi Rauk, and Evert J. Baerends. "On the calculation of multiplet energies by the hartree-fock-slater method". In: *Theoretica chimica acta* 43 (1977), pp. 261–271. DOI: https://doi-org.libproxy.berkeley.edu/10.1007/BF00551551.

[219] Pablo Ramos and Michele Pavanello. "Low-lying excited states by constrained DFT". In: *The Journal of Chemical Physics* 148.14 (2018), p. 144103. DOI: 10.1063/1.5018615. eprint: https://doi.org/10.1063/1.5018615. URL: https://doi.org/10.1063/1.5018615.

[220] Andrew T. B. Gilbert, Nicholas A. Besley, and Peter M. W. Gill. "Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM)". In: *The Journal of Physical Chemistry A* 112.50 (2008). PMID: 18729344, pp. 13164–13171. DOI: 10.1021/jp801738f. eprint: https://doi.org/10.1021/jp801738f. URL: https://doi.org/10.1021/jp801738f.

[221] Michael Filatov and Sason Shaik. "Spin-restricted density functional approach to the open-shell problem". In: *Chemical Physics Letters* 288.5 (1998), pp. 689–697. ISSN: 0009-2614. DOI: https://doi.org/10.1016/S0009-2614(98)00364-9. URL: https://www.sciencedirect.com/science/article/pii/S0009261498003649.

[222] Michael Filatov and Sason Shaik. "Application of spin-restricted open-shell Kohn-Sham method to atomic and molecular multiplet states". In: *The Journal of Chemical Physics* 110.1 (1999), pp. 116–125. DOI: 10.1063/1.477941. eprint: https://doi.org/10.1063/1.477941. URL: https://doi.org/10.1063/1.477941.

[223] Salomon R. Billeter and Daniel Egli. "Calculation of nonadiabatic couplings with restricted open-shell Kohn-Sham density-functional theory". In: *The Journal of Chemical Physics* 125.22 (2006), p. 224103. DOI: 10.1063/1.2360261. eprint: https://doi.org/10.1063/1.2360261. URL: https://doi.org/10.1063/1.2360261.

[224] J. R. Schmidt, Neil Shenvi, and John C. Tully. "Controlling spin contamination using constrained density functional theory". In: *The Journal of Chemical Physics* 129.11 (2008), p. 114110. DOI: 10.1063/1.2978168. eprint: https://doi.org/10.1063/1.2978168. URL: https://doi.org/10.1063/1.2978168.

[225] Stephan Grimm, Christel Nonnenberg, and Irmgard Frank. "Restricted open-shell Kohn-Sham theory for π-π^* transitions. I. Polyenes, cyanines, and protonated imines". In: *The Journal of Chemical Physics* 119.22 (2003), pp. 11574–11584. DOI: 10.1063/1.1623742. eprint: https://doi.org/10.1063/1.1623742. URL: https://doi.org/10.1063/1.1623742.

[226] Hugh G. A. Burton and David J. Wales. "Energy Landscapes for Electronic Structure". In: *Journal of Chemical Theory and Computation* 17.1 (2021). PMID: 33369396, pp. 151–169. DOI: 10.1021/acs.jctc.0c00772. eprint: https://doi.org/10.1021/acs.jctc.0c00772. URL: https://doi.org/10.1021/acs.jctc.0c00772.

[227] Giuseppe M. J. Barca, Andrew T. B. Gilbert, and Peter M. W. Gill. "Simple Models for Difficult Electronic Excitations". In: *Journal of Chemical Theory and Computation* 14.3 (2018). PMID: 29444408, pp. 1501–1509. DOI: 10.1021/acs.jctc.7b00994. eprint: https://doi.org/10.1021/acs.jctc.7b00994. URL: https://doi.org/10.1021/acs.jctc.7b00994.

[228] Nicholas A. Besley, Andrew T. B. Gilbert, and Peter M. W. Gill. "Self-consistent-field calculations of core excited states". In: *The Journal of Chemical Physics* 130.12 (2009), p. 124308. DOI: 10.1063/1.3092928. eprint: https://doi.org/10.1063/1.3092928. URL: https://doi.org/10.1063/1.3092928.

[229] Jeppe Gavnholt et al. "Δ self-consistent field method to obtain potential energy surfaces of excited molecules on surfaces". In: *Phys. Rev. B* 78 (7 2008), p. 075441. DOI: 10.1103/PhysRevB.78.075441. URL: https://link.aps.org/doi/10.1103/PhysRevB.78.075441.

[230] Alex J. W. Thom and Martin Head-Gordon. "Hartree-Fock solutions as a quasidiabatic basis for nonorthogonal configuration interaction". In: *The Journal of Chemical Physics* 131.12 (2009), p. 124113. DOI: 10.1063/1.3236841. eprint: https://doi.org/10.1063/1.3236841. URL: https://doi.org/10.1063/1.3236841.

[231] Giuseppe M. J. Barca, Andrew T. B. Gilbert, and Peter M. W. Gill. "Communication: Hartree-Fock description of excited states of H_2". In: *The Journal of Chemical Physics* 141.11 (2014), p. 111104. DOI: 10.1063/1.4896182. eprint: https://doi.org/10.1063/1.4896182. URL: https://doi.org/10.1063/1.4896182.

[232] Ernest R. Davidson and Larry E. Nitzsche. "Vertical excitation energy to the lowest 1 pi pi* state of acrolein". In: *Journal of the American Chemical Society* 101.22 (1979), pp. 6524–6526. DOI: 10.1021/ja00516a008. eprint: https://doi.org/10.1021/ja00516a008. URL: https://doi.org/10.1021/ja00516a008.

[233] Ernest R. Davidson and Larilyn Zeller Stenkamp. "SCF methods for excited states". In: *International Journal of Quantum Chemistry* 10.S10 (1976), pp. 21–31. DOI: https://doi.org/10.1002/qua.560100803. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560100803. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560100803.

[234] Chiao-Lun Cheng, Qin Wu, and Troy Van Voorhis. "Rydberg energies using excited state density functional theory". In: *The Journal of Chemical Physics* 129.12 (2008), p. 124112. DOI: 10.1063/1.2977989. eprint: https://doi.org/10.1063/1.2977989. URL: https://doi.org/10.1063/1.2977989.

[235] N D Woods, M C Payne, and P J Hasnip. "Computing the self-consistent field in Kohn-Sham density functional theory". In: *Journal of Physics: Condensed Matter* 31.45 (2019), p. 453001. DOI: 10.1088/1361-648x/ab31c0. URL: https://doi.org/10.1088/1361-648x/ab31c0.

[236] Jan-M. Mewes et al. "On the molecular mechanism of non-radiative decay of nitrobenzene and the unforeseen challenges this simple molecule holds for electronic structure theory". In: *Phys. Chem. Chem. Phys.* 16 (24 2014), pp. 12393–12406. DOI: 10.1039/C4CP01232A. URL: http://dx.doi.org/10.1039/C4CP01232A.

[237] Gianluca Levi, Aleksei V. Ivanov, and Hannes Jónsson. "Variational Density Functional Calculations of Excited States via Direct Optimization". In: *Journal of Chemical Theory and Computation* 16.11 (2020). PMID: 33064484, pp. 6968–6982. DOI: 10.1021/acs.jctc.0c00597. eprint: https://doi.org/10.1021/acs.jctc.0c00597. URL: https://doi.org/10.1021/acs.jctc.0c00597.

[238] Alex J. W. Thom and Martin Head-Gordon. "Locating Multiple Self-Consistent Field Solutions: An Approach Inspired by Metadynamics". In: *Phys. Rev. Lett.* 101 (19 2008), p. 193001. DOI: 10.1103/PhysRevLett.101.193001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.101.193001.

[239] Gerald Knizia and Garnet Kin-Lic Chan. "Density Matrix Embedding: A Simple Alternative to Dynamical Mean-Field Theory". In: *Phys. Rev. Lett.* 109 (18 2012), p. 186404. DOI: 10.1103/PhysRevLett.109.186404. URL: https://link.aps.org/doi/10.1103/PhysRevLett.109.186404.

[240] Gerald Knizia and Garnet Kin-Lic Chan. "Density Matrix Embedding: A Strong-Coupling Quantum Embedding Theory". In: *Journal of Chemical Theory and Computation* 9.3 (2013). PMID: 26587604, pp. 1428–1432. DOI: 10.1021/ct301044e. eprint: https://doi.org/10.1021/ct301044e. URL: https://doi.org/10.1021/ct301044e.

[241] Henry K. Tran, Troy Van Voorhis, and Alex J. W. Thom. "Using SCF metadynamics to extend density matrix embedding theory to excited states". In: *The Journal of Chemical Physics* 151.3 (2019), p. 034112. DOI: 10.1063/1.5096177. eprint: https://doi.org/10.1063/1.5096177. URL: https://doi.org/10.1063/1.5096177.

[242] Hugh G. A. Burton and Alex J. W. Thom. "General Approach for Multireference Ground and Excited States Using Nonorthogonal Configuration Interaction". In: *Journal of Chemical Theory and Computation* 15.9 (2019). PMID: 31381337, pp. 4851–4861. DOI: 10.1021/acs.jctc.9b00441. eprint: https://doi.org/10.1021/acs.jctc.9b00441. URL: https://doi.org/10.1021/acs.jctc.9b00441.

[243] Roger Fletcher. *Practical methods of optimization*. Chichester, New York: John Wiley & Sons, 1987.

[244] Yousef Saad. *Iterative Methods for Sparse Linear Systems*. 2nd. USA: Society for Industrial and Applied Mathematics, 2003. ISBN: 0898715342.

[245] Yousef Saad. *Numerical Methods for Large Eigenvalue Problems: Revised Edition*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2011. ISBN: 9781611970739.

[246] Ernest R. Davidson. "The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices". In: *Journal of Computational Physics* 17.1 (1975), pp. 87–94. ISSN: 0021-9991. DOI: https://doi.org/10.1016/0021-9991(75)90065-0. URL: https://www.sciencedirect.com/science/article/pii/0021999175900650.

[247] Richard P Messmer. "On a variational method for determining excited state wave functions". In: *Theoret. Chim. Acta* 14 (1969), p. 319.

[248] V. V. Murakhtanov, L. N. Mazalov, and T. I. Guzhavina. "Calculation of highly excited states using the square of the Hamiltonian". In: *J. Struct. Chem.* 23 (1982), p. 189.

[249] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins. "Optimized trial wave functions for quantum Monte Carlo calculations". In: *Phys. Rev. Lett.* 60 (17 1988), pp. 1719–1722. DOI: 10.1103/PhysRevLett.60.1719. URL: https://link.aps.org/doi/10.1103/PhysRevLett.60.1719.

[250] B. L. Hammond, W. A. Lester Jr., and P. J. Reynolds. *Monte Carlo Methods in Ab Initio Quantum Chemistry*. Singapore: World Scientific, 1994, p. 70.

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

[252] Jarrod R McClean et al. "The theory of variational hybrid quantum-classical algorithms". In: *New J. Phys.* 18.2 (2016), p. 023023.

[253] DH Weinstein. "Modified Ritz method". In: *Proc. Natl. Acad. Sci. USA* 20.9 (1934), p. 529.

[254] JKL MacDonald. "On the modified Ritz variation method". In: *Phys. Rev.* 46.9 (1934), p. 828.

[255] Jong H Choi, Charles F Lebeda, and Richard P Messmer. "Variational principle for excited states: Exact formulation and other extensions". In: *Chem. Phys. Lett.* 5.8 (1970), p. 503.

[256] Luning Zhao and Eric Neuscamman. "An Efficient Variational Principle for the Direct Optimization of Excited States". In: *Journal of Chemical Theory and Computation* 12.8 (2016). PMID: 27379468, pp. 3436–3440. DOI: 10.1021/acs.jctc.6b00508. eprint: https://doi.org/10.1021/acs.jctc.6b00508. URL: https://doi.org/10.1021/acs.jctc.6b00508.

[257] Nicholas Metropolis et al. "Equation of state calculations by fast computing machines". In: *J. Chem. Phys.* 21.6 (1953), pp. 1087–1092.

[258] W. M. C. Foulkes et al. "Quantum Monte Carlo simulations of solids". In: *Rev. Mod. Phys.* 73 (1 2001), pp. 33–83. DOI: 10.1103/RevModPhys.73.33.

[259] C. G. Broyden. "The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations". In: *J. Inst. Math. Its Appl.* 6.1 (1970), pp. 76–90. DOI: 10.1093/imamat/6.1.76.

[260] R. Fletcher. "A new approach to variable metric algorithms". In: *Comput. J.* 13.3 (1970), pp. 317–322. DOI: 10.1093/comjnl/13.3.317.

[261] Donald Goldfarb. "A family of variable-metric methods derived by variational means". In: *Math. Comp.* 24.109 (1970), pp. 23–26. DOI: 10.1090/S0025-5718-1970-0258249-6.

[262] David F. Shanno. "Conditioning of quasi-Newton methods for function minimization". In: *Math. Comp.* 24.111 (1970), pp. 647–656. DOI: 10.1090/S0025-5718-1970-0274029-X.

[263] Luning Zhao and Eric Neuscamman. "Excited state mean-field theory without automatic differentiation". In: *The Journal of Chemical Physics* 152.20 (2020), p. 204112. DOI: 10.1063/5.0003438. eprint: https://doi.org/10.1063/5.0003438. URL: https://doi.org/10.1063/5.0003438.

[264] Tory Van Voorhis and Martin Head-Gordon. "A geometric approach to direct minimization". In: *Mol. Phys.* 100.11 (2002), pp. 1713–1721. DOI: 10.1080/00268970110103642.

[265] Diptarka Hait and Martin Head-Gordon. "Excited State Orbital Optimization via Minimizing the Square of the Gradient: General Approach and Application to Singly and Doubly Excited States via Density Functional Theory". In: *Journal of Chemical Theory and Computation* 16.3 (2020). PMID: 32017554, pp. 1699–1710. DOI: 10.1021/acs.jctc.9b01127. eprint: https://doi.org/10.1021/acs.jctc.9b01127. URL: https://doi.org/10.1021/acs.jctc.9b01127.

[266] Jeongnim Kim et al. "QMCPACK: an open source *ab initio* quantum Monte Carlo package for the electronic structure of atoms, molecules and solids". In: *Journal of Physics: Condensed Matter* 30.19 (2018), p. 195901. DOI: 10.1088/1361-648x/aab9c3. URL: https://doi.org/10.1088/1361-648x/aab9c3.

[267] Luke Shulenburger and Thomas R Mattsson. "Quantum Monte Carlo applied to solids". In: *Phys. Rev. B* 88.24 (2013), p. 245117.

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

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

[270] Krisnan Raghavachari et al. "A fifth-order perturbation comparison of electron correlation theories". In: *Chem. Phys. Lett.* 157 (1989), p. 479.

[271] George D. Purvis III and Rodney J. Bartlett. "A full couple-cluster singles and doubles model: The inclusion of disconnected triples". In: *J. Chem. Phys.* 76 (1982), p. 1910.

[272] Gustavo E Scuseria et al. "Projected quasiparticle theory for molecular electronic structure". In: *J. Chem. Phys.* 135.12 (2011), p. 124108.

[273] Carlos A. Jiménez-Hoyos et al. "Projected Hartree-Fock theory". In: *J. Chem. Phys.* 136 (2012), p. 164109.

[274] A. C. Hurley, John Lennard-Jones, and J. A. Pople. "The Molecular Orbital Theory of Chemical Valency. XVI. A Theory of Paired-Electrons in Polyatomic Molecules". In: *Proc. R. Soc. London, Ser. A* 220 (1953), p. 446.

[275] A John Coleman. "The AGP model for fermion systems". In: *Int. J. Quantum Chem.* 63.1 (1997), p. 23.

[276] Viktor N Staroverov and Gustavo E Scuseria. "Optimization of density matrix functionals by the Hartree–Fock–Bogoliubov method". In: *J. Chem. Phys.* 117.24 (2002), p. 11107.

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

[278] 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.* 127.1 (2007), p. 014105.

[279] Mariapia Marchi et al. "Resonating valence bond wave function with molecular orbitals: Application to first-row molecules". In: *J. Chem. Phys.* 131.15 (2009), p. 154116.

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

[281] 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.

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

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

[284] 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.

[285] Julien Toulouse and C. J. Umrigar. "Optimization of quantum Monte Carlo wave functions by energy minimization". In: *J. Chem. Phys.* 126.8 (2007), p. 084102. DOI: 10.1063/1.2437215.

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

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

[288] 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.* 51.6 (1969), pp. 2657–2664. DOI: 10.1063/1.1672392.

[289] *PySCF, a quantum chemistry package written in python.* see http://chemists.princeton.edu/chan/sof

[290] R. B. J. S. Krishnan et al. "Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions". In: *J. Chem. Phys.* 72.1 (1980), p. 650.

[291] Yihan Shao et al. "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package". In: *Mol. Phys.* 113 (2015), p. 184.

[292] Michael W. Schmidt et al. "General atomic and molecular electronic structure system". In: *J. Comput. Chem.* 14.11 (1993), p. 1347. ISSN: 1096-987X. DOI: 10.1002/jcc.540141112. URL: http://dx.doi.org/10.1002/jcc.540141112.

[293] H.-J. Werner et al. *MOLPRO, version 2015.1, a package of ab initio programs*. accessed on February 14, 2017. URL: http://www.molpro.net.

[294] H.-J. Werner et al. "Molpro: a general-purpose quantum chemistry program package". In: *WIREs Comput. Mol. Sci.* 2 (2012), pp. 242–253.

[295] M. Burkatzki, C. Filippi, and M. Dolg. "Energy-consistent pseudopotentials for quantum Monte Carlo calculations". In: *J. Chem. Phys.* 126 (2007), p. 234105.

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

[297] N. S. Blunt and E. Neuscamman. "Charge-transfer excited states: Seeking a balanced and efficient wave function ansatz in variational Monte Carlo". In: *arXiv* (2017), p. 1707.09439.

[298] C. Y. Legault. *CYLView, 1.0b*. 2009. URL: http://www.cylview.org.

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

[300] K. Ruedenberg et al. In: *Chem. Phys.* 71 (1982), p. 41.

[301] Hans-Joachim Werner and Peter J. Knowles. "A second order multiconfiguration SCF procedure with optimum convergence". In: *J. Chem. Phys.* 82 (1985), p. 5053.

[302] Peter J. Knowles and Hans-Joachim Werner. "An efficient second-order MC SCF method for long configuration expansions". In: *Chem. Phys. Lett.* 115 (1985), p. 259.

[303] B. O. Roos. In: *Adv. Chem. Phys.* 69 (1987), p. 399.

[304] Jeppe Olsen et al. "Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces". In: *J. Chem. Phys.* 89.4 (1988), pp. 2185–2192.

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

[306] Dominika Zgid and Marcel Nooijen. "The density matrix renormalization group self-consistent field method: Orbital optimization with the density matrix renormalization group method in the active space". In: *J. Chem. Phys.* 128 (2008), p. 144116.

[307] Giovanni Li Manni et al. "SplitGAS method for strong correlation and the challenging case of Cr2". In: *J. Chem. Theory Comput.* 9.8 (2013), pp. 3375–3384.

[308] Robert E. Thomas et al. "Stochastic Multiconfigurational Self-Consistent Field Theory". In: *J. Chem. Theory Comput.* 11 (2015), p. 5316.

[309] Giovanni Li Manni, Simon D Smart, and Ali Alavi. "Combining the Complete Active Space Self-Consistent Field Method and the Full Configuration Interaction Quantum Monte Carlo within a Super-CI Framework, with Application to Challenging Metal-Porphyrins". In: *J. Chem. Theory Comput.* 12.3 (2016), pp. 1245–1258.

[310] James E T Smith et al. "Cheap and near exact CASSCF with large active spaces". In: *J. Chem. Theory Comput.* 13.11 (2017), pp. 5468–5478.

[311] S. R. Langhoff and E. R. Davidson. In: *Int. J. Quantum Chem.* 8 (1974), p. 61.

[312] H.-J. Werner and E. A. Reinsch. In: *J. Chem. Phys.* 76 (1982), p. 3144.

[313] K. Andersson et al. In: *J. Phys. Chem.* 94 (1990), p. 5483.

[314] K. Andersson, P.-Å. Malmqvist, and B. O. Roos. In: *J. Chem. Phys.* 96 (1992), p. 1218.

[315] R. J. Gdanitz and R. Ahlrichs. In: *Chem. Phys. Lett.* 143 (1988), p. 413.

[316] H.-J. Werner and P. J. Knowles. In: *Theor. Chim Acta* 78 (1990), p. 175.

[317] P. G. Szalay and R. J. Bartlett. In: *Chem. Phys. Lett.* 214 (1993), p. 481.

[318] C. Angeli et al. In: *J. Chem. Phys.* 114 (2001), p. 10252.

[319] C. Angeli, R. Cimiraglia, and J.-P. Malrieu. In: *J. Chem. Phys.* 117 (2002), p. 9138.

[320] Eric Neuscamman, Takeshi Yanai, and Garnet K.-L. Chan. "A review of canonical transformation theory". In: *Int. Rev. Phys. Chem.* 29 (2010), p. 231.

[321] Giovanni Li Manni et al. "Multiconfiguration pair-density functional theory". In: *J. Chem. Theory Comput.* 10.9 (2014), pp. 3669–3680.

[322] Sandeep Sharma and Ali Alavi. "Multireference linearized coupled cluster theory for strongly correlated systems using matrix product states". In: *J. Chem. Phys.* 143.10 (2015), p. 102815.

[323] M K MacLeod and T Shiozaki. "Communication: automatic code generation enables nuclear gradient computations for fully internally contracted multireference theory." In: *J. Chem. Phys.* 142.5 (2015), p. 051103.

[324] Gergely Gidofalvi and David A Mazziotti. "Active-space two-electron reduced-density-matrix method: Complete active-space calculations without diagonalization of the N-electron Hamiltonian". In: *J. Chem. Phys.* 129.13 (2008), p. 134108.

[325] 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.

[326] Xinle Liu et al. "Communication: An inexpensive, variational, almost black-box, almost size-consistent correction to configuration interaction singles for valence excited states". In: *J. Chem. Phys.* 138 (2013), p. 221105.

[327] Xinle Liu and Joseph E Subotnik. "The Variationally Orbital-Adapted Configuration Interaction Singles (VOA-CIS) Approach to Electronically Excited States". In: *J. Chem. Theory Comput.* 10.3 (2014), pp. 1004–1020.

[328] Brad S Veldkamp et al. "Molecular Excited States: Accurate Calculation of Relative Energies and Electronic Coupling Between Charge Transfer and Non-Charge Transfer States". In: *J. Phys. Chem. A* 119.2 (2015), pp. 253–262.

[329] 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.* 147.16 (2017), p. 164114.

[330] 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. DOI: 10.1021/acs.jctc.7b00923.

[331] 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.

[332] Andreas Griewank and Andrea Walther. *Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation, Second Edition*. Philadelphia: Society for Industrial and Applied Mathematics, 2008.

[333] Sandro Sorella and Luca Capriotti. "Algorithmic differentiation and the calculation of forces by quantum Monte Carlo". In: *J. Chem. Phys.* 133.23 (2010), p. 234111.

[334] Teresa Tamayo-Mendoza et al. "Automatic Differentiation in Quantum Chemistry with Applications to Fully Variational Hartree–Fock". In: *ACS Cent. Sci.* 4.5 (2018), p. 559.

[335] https://www.tensorflow.org [accessed 2-October-2019].

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

[337] Martin Schütz and Hans-Joachim Werner. "Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD)". In: *J. Chem. Phys.* 114.2 (2001), pp. 661–681.

[338] Joseph E Subotnik, Alex Sodt, and Martin Head-Gordon. "A near linear-scaling smooth local coupled cluster algorithm for electronic structure". In: *J. Chem. Phys.* 125.7 (2006), p. 074116.

[339] Jun Yang et al. "The orbital-specific-virtual local coupled cluster singles and doubles method". In: *J. Chem. Phys.* 136.14 (2012), p. 144105.

[340] Christoph Riplinger and Frank Neese. "An efficient and near linear scaling pair natural orbital based local coupled cluster method". In: *J. Chem. Phys.* 138.3 (2013), p. 034106.

[341] Young Min Rhee and Martin Head-Gordon. "Scaled second-order perturbation corrections to configuration interaction singles: efficient and reliable excitation energy methods". In: *J. Phys. Chem. A* 111.24 (2007), pp. 5314–5326.

[342] Chenyang Li et al. "A low-cost approach to electronic excitation energies based on the driven similarity renormalization group". In: *J. Chem. Phys.* 147 (2017), p. 074107.

[343] So Hirata, Marcel Nooijen, and Rodney J Bartlett. "High-order determinantal equation-of-motion coupled-cluster calculations for electronic excited states". In: *Chem. Phys. Lett.* 326.3-4 (2000), pp. 255–262.

[344] Sergio D. Pineda Flores and Eric Neuscamman. "Excited State Specific Multi-Slater Jastrow Wave Functions". In: *J. Phys. Chem. A* 123.8 (2019), pp. 1487–1497. DOI: 10.1021/acs.jpca.8b10671.

[345] 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.* 12.8 (2016), pp. 3719–3726. DOI: 10.1021/acs.jctc.6b00480.

[346] Dan Kalman. "Leveling with Lagrange: An Alternate View of Constrained Optimization". In: *Mathematics Magazine* 82.3 (2009), pp. 186–196. ISSN: 0025570X, 19300980. URL: http://www.jstor.org/stable/27765899.

[347] Hong-Zhou Ye and Troy Van Voorhis. "Half-Projected σ Self-Consistent Field for Electronic Excited States". In: *J. Chem. Theory Comput.* 15.5 (2019), pp. 2954–2965. DOI: 10.1021/acs.jctc.8b01224.

[348] Mel Levy. "Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem". In: *Proc. Natl. Acad. Sci. USA* 76.12 (1979), pp. 6062–6065. DOI: 10.1073/pnas.76.12.6062.

[349] C. A. Coulson and I. Fischer. "XXXIV. Notes on the molecular orbital treatment of the hydrogen molecule". In: *Lond. Edinb. Phil. Mag.* 40.303 (1949), pp. 386–393. DOI: 10.1080/14786444908521726.

[350] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 80–82.

[351] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 433–433.

[352] Martin Head-Gordon and John A. Pople. "Optimization of Wave Function and Geometry in the Finite Basis Hartree-Fock Method". In: *J. Phys. Chem.* 92 (1988), pp. 3063–3069. DOI: 10.1021/j100322a012.

[353] Galina Chaban, Michael W. Schmidt, and Mark S. Gordon. "Approximate second order method for orbital optimization of SCF and MCSCF wavefunctions". In: *Theor. Chem. Acc.* 97 (1997), pp. 88–95. DOI: 10.1007/s002140050241.

[354] George B.Bacskay. "A quadratically convergent Hartree—Fock (QC-SCF) method. Application to closed shell systems". In: *Chem. Phys.* 61.3 (1981), pp. 385–404. DOI: 10.1016/0301-0104(81)85156-7.

[355] George B. Bacskay. "A quadritically convergent hartree-fock (QC-SCF) method. Application to open shell orbital optimization and coupled perturbed hartree-fock calculations". In: *Chem. Phys.* 65.3 (1982), pp. 383–396. DOI: 10.1016/0301-0104(82)85211-7.

[356] Thomas H. Fischer and Jan Almlof. "General Methods for Geometry and Wave Function Optimization". In: *J. Phys. Chem.* 96.24 (1992), pp. 9768–9774. DOI: 10.1021/j100203a036.

[357] Barak A. Pearlmutter. "Fast Exact Multiplication by the Hessian". In: *Neural Comput.* 6 (1994), pp. 147–160.

[358] Youcef Saad and Martin H. Schultz. "GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Ssystems". In: *SIAM J. Sci. Comput.* 7.3 (1986), pp. 856–869. DOI: 10.1137/0907058.

[359] 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.

[360] 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.

[361] John B. Collins et al. "Self-consistent molecular orbital methods. XVII. Geometries and binding energies of second-row molecules. A comparison of three basis sets". In: *J. Chem. Phys.* 64.12 (1976), pp. 5142–5151. DOI: 10.1063/1.432189.

[362] Francesco Aquilante et al. "Atomic Cholesky decompositions: A route to unbiased auxiliary basis sets for density fitting approximation with tunable accuracy and efficiency". In: *J. Chem. Phys.* 130.15 (2009), p. 154107. DOI: 10.1063/1.3116784.

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

[364] Axel D. Becke. "Density-functional thermochemistry. III. The role of exact exchange". In: *J. Chem. Phys.* 98.7 (1993), pp. 5648–5652. DOI: 10.1063/1.464913.

[365] G. A. Petersson et al. "A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements". In: *J. Chem. Phys.* 89.4 (1988), pp. 2193–2218. DOI: 10.1063/1.455064.

[366] G. A. Petersson and Mohammad A. Al-Laham. "A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms". In: *J. Chem. Phys.* 94.9 (1991), pp. 6081–6090. DOI: 10.1063/1.460447.

[367] Narbe Mardirossian and Martin Head-Gordon. "ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy". In: *Phys. Chem. Chem. Phys.* 16 (21 2014), pp. 9904–9924. DOI: 10.1039/C3CP54374A. URL: http://dx.doi.org/10.1039/C3CP54374A.

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

[369] G. Schaftenaar and J. H. Noordik. "Molden: a pre- and post-processing program for molecular and electronic structures". In: *J. Comput.-Aided Mol. Des.* 14.2 (2000), pp. 123–134. DOI: 10.1023/A:1008193805436.

[370] Gijs Schaftenaar, Elias Vlieg, and Gerrit Vriend. "Molden 2.0: quantum chemistry meets proteins". In: *J. Comput.-Aided Mol. Des.* 31.9 (2017), pp. 789–800. DOI: 10.1007/s10822-017-0042-5.

[371] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic Structure Theory*. West Sussex, England: John Wiley and Sons, Ltd, 2000, pp. 398–405.

[372] Frank Neese. "An improvement of the resolution of the identity approximation for the formation of the Coulomb matrix". In: *J. Comput. Chem.* 24.14 (2003), pp. 1740–1747. DOI: 10.1002/jcc.10318.

[373] Xinguo Ren et al. "Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions". In: *New J. Phys.* 14.5 (2012), p. 053020.

[374] Douglas L. Strout and Gustavo E. Scuseria. "A quantitative study of the scaling properties of the Hartree–Fock method". In: *J. Chem. Phys.* 102.21 (1995), pp. 8448–8452. DOI: 10.1063/1.468836.

[375] Dávid Mester and Mihály Kállay. "Reduced-Scaling Approach for Configuration Interaction Singles and Time-Dependent Density Functional Theory Calculations Using Hybrid Functionals". In: *J. Chem. Theory Comput.* 15.3 (2019), pp. 1690–1704. DOI: 10.1021/acs.jctc.8b01199.

[376] Stefan Goedecker. "Linear scaling electronic structure methods". In: *Rev. Mod. Phys.* 71 (4 1999), pp. 1085–1123. DOI: 10.1103/RevModPhys.71.1085.

[377] Jan Geertsen, Magnus Rittby, and Rodney J. Bartlett. "The equation-of-motion coupled-cluster method: Excitation energies of Be and CO". In: *Chem. Phys. Lett.* 164.1 (1989), pp. 57–62. ISSN: 0009-2614. DOI: https://doi.org/10.1016/0009-2614(89)85202-9. URL: http://www.sciencedirect.com/science/article/pii/0009261489852029.

[378] Donald C. Comeau and Rodney J. Bartlett. "The equation-of-motion coupled-cluster method. Applications to open- and closed-shell reference states". In: *Chem. Phys. Lett.* 207.4 (1993), pp. 414–423. ISSN: 0009-2614. DOI: https://doi.org/10.1016/0009-2614(93)89023-B. URL: http://www.sciencedirect.com/science/article/pii/000926149389023B.

[379] John F. Stanton and Rodney J. Bartlett. "The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties". In: *The Journal of Chemical Physics* 98.9 (1993), pp. 7029–7039. DOI: 10.1063/1.464746. eprint: https://doi.org/10.1063/1.464746. URL: https://doi.org/10.1063/1.464746.

[380] John D. Watts and Rodney J. Bartlett. "Iterative and non-iterative triple excitation corrections in coupled-cluster methods for excited electronic states: the EOM-CCSDT-3 and EOM-CCSD(T̃) methods". In: *Chem. Phys. Lett.* 258.5 (1996), p. 581. ISSN: 0009-2614. DOI: 10.1016/0009-2614(96)00708-7.

[381] Andreas Dreuw and Martin Head-Gordon. "Failure of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excited States: The Zincbacteriochlorin-Bacteriochlorin and Bacteriochlorophyll-Spheroidene Complexes". In: *J. Am. Chem. Soc.* 126.12 (2004). PMID: 15038755, pp. 4007–4016. DOI: 10.1021/ja039556n. eprint: https://doi.org/10.1021/ja039556n. URL: https://doi.org/10.1021/ja039556n.

[382] Andreas Dreuw, Jennifer L. Weisman, and Martin Head-Gordon. "Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange". In: *J. Chem. Phys.* 119.6 (2003), pp. 2943–2946. DOI: 10.1063/1.1590951. eprint: https://doi.org/10.1063/1.1590951. URL: https://doi.org/10.1063/1.1590951.

[383] Jeng-Da Chai and Martin Head-Gordon. "Systematic optimization of long-range corrected hybrid density functionals". In: *J. Chem. Phys.* 128.8 (2008), p. 084106. DOI: 10.1063/1.2834918. eprint: https://doi.org/10.1063/1.2834918. URL: https://doi.org/10.1063/1.2834918.

[384] Scott M. Garner and Eric Neuscamman. "A variational Monte Carlo approach for core excitations". In: *The Journal of Chemical Physics* 153.14 (2020), p. 144108. DOI: 10.1063/5.0020310. eprint: https://doi.org/10.1063/5.0020310. URL: https://doi.org/10.1063/5.0020310.

[385] Luning Zhao and Eric Neuscamman. "Variational Excitations in Real Solids: Optical Gaps and Insights into Many-Body Perturbation Theory". In: *Phys. Rev. Lett.* 123 (3 2019), p. 036402. DOI: 10.1103/PhysRevLett.123.036402. URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.036402.

[386] Tanja Đurić and Tomislav Ševa. "Efficient neural-network based variational Monte Carlo scheme for direct optimization of excited energy states in frustrated quantum systems". In: *Phys. Rev. B* 102 (8 2020), p. 085104. DOI: 10.1103/PhysRevB.102.085104. URL: https://link.aps.org/doi/10.1103/PhysRevB.102.085104.

[387] Rachel Clune, Jacqueline A. R. Shea, and Eric Neuscamman. "N5-Scaling Excited-State-Specific Perturbation Theory". In: *Journal of Chemical Theory and Computation* 16.10 (2020). PMID: 32816474, pp. 6132–6141. DOI: 10.1021/acs.jctc.0c00308. eprint: https://doi.org/10.1021/acs.jctc.0c00308. URL: https://doi.org/10.1021/acs.jctc.0c00308.