ACCURATELY COMPUTING EXCITED STATES AND LATTICE DEPENDENT TIGHT-BINDING MODELS FROM FIRST PRINCIPLES SIMULATIONS

BY

SHIVESH ANANDA PATHAK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Associate Professor Andre Schleife, Chair
Assistant Professor Lucas Wagner, Director of Research
Professor Karin Dahmen
Professor Brian DeMarco

Abstract

The core problem of condensed matter physics is the understanding of how diverse macroscopic quantum phenomenon can emerge from a basic set of constituent particles and their interactions. Fortunately, quantum mechanics provides a universal bridge between the microscopic electrons and macroscopic quantum phenomenon via the *ab initio* Hamiltonian and Hilbert space. The eigenstates and eigenenergies of the *ab initio* Hamiltonian exactly correspond to the excited states and energies of a given material, and can be used to compute physical observables like conductivity, optical gaps, and magnetization.

However, computing the eigenstates and eigenspectrum of the *ab initio* Hamiltonian through brute force methods is generally computationally intractable, except in special cases of small molecules like H_2. Instead, approximate methods have been developed and used to great success in accurately computing the eigenspectrum and eigenstates of the *ab initio* Hamiltonian, falling under the categories of first principles methods and effective model Hamiltonians. While both first principles methods and effective model Hamiltonians have been used to accurately compute properties of real materials, significant avenues of research still remain. The biggest avenue for discovery is first principles excited states methods, with accurate ground state techniques like quantum Monte Carlo lacking a mature excited state counterpart. Accompanying this large avenue for change is the constant need for adaptation and development of methods to keep up with the rapid rate of novel materials discoveries. The construction of interacting effective Hamiltonians from *ab initio* calculations is the predominant challenge in the effective model approach.

To this end, my thesis has been oriented around advancing the state of the art in first principles excited state computation and effective models with lattice effects. First, I present my work on investigating a new trial wave function for use in quantum Monte Carlo (QMC). The new non-orthogonal determinant wave function expands the possibilities of accurate QMC calculations, as the quality of the trial wave function is a primarily limiting factor of accuracy in QMC calculations. Next, I present my work on developing a stable statistical estimate for gradients used in QMC wave function optimization. This efficient method is simple to integrate into existing QMC codes, and improves the efficiency of QMC wave function optimization, an integral component of QMC calculations. Following that, I present my work on creating a novel method for computing excited states in QMC. The novel method addresses short comings of state-of-the-art QMC methods by allowing for state specific optimization with high accuracy and computational efficiency. I conclude by demonstrating my work building a tight-binding model for twisted bilayer graphene with lattice interactions from density functional theory (DFT). The work provided a pipeline for developing accurate DFT models with electron-lattice interaction, and demonstrated the importance of these interactions in the quantitative and qualitative description of the flat bands in magic angle TBLG. My work demonstrates the power of *ab initio* techniques in accurately computing excitations and eigenstates of complex quantum systems, and provides a concrete stepping stone towards the advancement of accurate and efficient first principles methods and effective model Hamiltonians.

To my family, my community, and my love

Acknowledgements

I would like to begin by thanking those in the University of Illinois Physics Department who have helped me tremendously in my research. This would include primarily S. Lance Cooper and Wendy Wimmer, without whose help I would likely be lost in a mar of bureaucracy.

Next is my advisor, Lucas Wagner, who has been an amazing mentor throughout my PhD. Even though we have clashed at times, I would wish for no other mentor during my time as a graduate student. Lucas' passion for research and research methodology has fundamentally shaped the way I do science, and all for the better. I thank him for his effort, and wish him nothing less than the best, healthiest, and happiest career.

Lastly, all those in my research group who have pushed me and inspired me with their ideas and work. These would include Alex Munoz, Yueqing Chang, William Wheeler, Brian Busemeyer, Kiel Williams.

Next are my friends and community. I'd like to extend a thanks specifically to the UIUC GEO, UIUC Native American House, and the Wesley Food Pantry. These communities and groups have solidified my notion of communal effort and progress as an integral part of my life. I could not have gotten through my PhD without their help and fervor.

I'd like to thank my two closest friends Karan Kolwalkar and Timshawn Luh for being supportive and fun for over a decade at this point. Even though we cannot see each other in person much, they are a constant and welcome presence in my life.

Lastly, those that are closest to me. I would like to thank my mom, my dad, and my younger sister. Above all, my family has been the biggest inspiration and drive for me to complete my work. My parents have made tremendous sacrifice for my freedom to choose my life, and I cannot

thank them enough. While we may not see eye-to-eye on everything, my love for my parents will transcend anything else, and I cannot thank them enough for their support. A special thanks to my sister for being a constant source of positivity through tough times. Finally, I would like to thank Yang Hu for her invaluable support, insight, and love. She has been an integral part of my work, my life, and my happiness for the last three years, and I cannot thank her enough for everything she has done for me.

Now for an extremely sharp cut to funding acknowledgments. This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (Award Nos. OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Superconducting Applications. This work was funded by Grant No. DOE FG02-12ER46875 (SciDAC). This study is based on the work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Computational Materials Sciences program under Award No. DE-SC-0020177.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The many-body problem of condensed matter physics

The core problem of condensed matter physics is the understanding of how diverse macroscopic quantum phenomenon can emerge from a basic set of constituent particles and their interactions. The particles are the electrons and nuclei which compose a particular material, and the interactions are the bare Coulomb interaction between these charged particles. While all condensed matter systems are composed of these same constituents, the variety of macroscopic properties of materials is tremendous. For example, systems can be metallic or insulating, with additional qualifiers like band metal, bad metal, strange metal, semimetal, band insulator, Mott insulator, and charge transfer insulator providing more detailed descriptions of the properties of these materials. More exotic magnetic, superconducting, and topological materials also exist, which too have additional refinement based on the details of the system like antiferromagnets, ferromagnets, paramagnets, conventional superconductors, unconventional superconductors and heavy fermion superconductors.

Fortunately, quantum mechanics provides a universal bridge between the microscopic electrons and macroscopic quantum phenomenon via the *ab initio* Hamiltonian and Hilbert space [7]. The non-relativistic *ab initio* Hamiltonian takes the following form for all systems

$$\hat{H}_{ab} = \hat{T} + \hat{V} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i<j=1}^{N}\frac{1}{|\vec{r}_i - \vec{r}_j|} + \sum_{I<J=1}^{N_{ion}}\frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} - \sum_{i=1}^{N}\sum_{J=1}^{N_{ion}}\frac{Z_J}{|\vec{r}_i - \vec{R}_J|} \tag{1.1}$$

where $\vec{r}_i$ refer to the electron coordinates, $\vec{R}_I$ the ionic coordinates, Z_J the ionic charge of ion J, and N, N_{ion} are the number of electrons and ions in the system. The fundamental constants

$\hbar = m_e = e = 4\pi\epsilon_0 = 1$ — reduced Planck's constant, electron rest mass, elementary charge, and electric constant — are set to unity, with Hartree (Ha) as the unit of energy. This Hamiltonian contains explicitly the microscopic degrees of freedom for the electrons and ions and the $1/r$ Coulomb interactions between them. The *ab initio* Hilbert space, $\mathcal{H}$, consists of all anti-symmetric N particle wave functions. Solving the eigenvalue problem $\hat{H}_{ab}\Psi = E\Psi$ over $\mathcal{H}$ will yield many-body eigenstates and spectra which can then be used to compute ground and excited state properties system such as gaps, magnetic order, and electrical conductivity.

However, computing the eigenstates and eigenspectrum of $\hat{H}_{ab}$ through brute force methods is generally computationally intractable, except in special cases of small molecules like H_2. This is best understood by looking at the size of $\mathcal{H}$, which is estimated to be 10^{42} [8] for accurate computation of spectra for the copper-oxide molecule - one copper and one oxygen atom. The brute force approach of exact diagonalization (ED), where the entire $10^{42} \times 10^{42}$ matrix is constructed and explicitly diagonalized, is a task beyond what any modern computer can accomplish. To make matters worse, $\mathcal{H}$ scales exponentially with system size, so ED for solids is even more intractable than the already incredulous task for the copper-oxide molecule. Approximate methods, however, have been developed and used to great success in accurately computing the eigenspectrum and eigenstates of $\hat{H}_{ab}$. Following is a discussion of two powerful classes of approximate methods: first principles methods and model Hamiltonians.

1.2 First principles methods and model Hamiltonians

1.2.1 State-of-the art

First principles methods are a class of techniques which approximately solve for the eigenstates of Eq 1.1, trading off accuracy for computational efficiency. Density functional theory (DFT) is a commonly used first principles method which uses a mean-field approximation to simplify the many-body Coulomb interactions in Eq 1.1. This approach is extremely efficient, being able to compute ground state properties for simulations with tens of thousands of atoms [9, 10, 11, 12, 13], but has poor accuracy when dealing with transition metal, magnetic, and superconducting systems

[8, 14, 15]. Quantum Monte Carlo (QMC) is a more accurate, and more expensive, approach than DFT which does not make a mean-field approximation, instead opting to work with explicit many-body *ansatzes* for the ground and excited state wave functions. As such, QMC methods can accurately describe the ground states of transition metal, magnetic, and superconducting systems [16, 17, 18, 19, 20, 21] up to a few hundred atoms. Filling out the rest of the accuracy/cost spectrum are a host of other first principles methods from the inexpensive Hartree-Fock, to the near exact but computationally demanding full- and semi-stochastic heat-bath configuration interaction methods (FCI, SHCI).

The effective model approach to the computational challenge of computing the eigenstates and spectra of Eq 1.1 works by constructing a simpler effective Hamiltonian and Hilbert space that replace $\hat{H}_{ab}, \mathcal{H}$ and can be more easily solved. A canonical example is the DFT band theory approach, where the *ab initio* Hamiltonian is replaced by an effective non-interacting Hamiltonian computed using DFT. This approach simplifies computation as the non-interacting model can be diagonalized with a simple Fourier transformation and yields accurate results for simple metals, semiconductors, and band insulating systems [22, 23, 24, 25]. Methods like the constrainted random phase approximation (cRPA) can be used to introduce effective interactions in model Hamiltonians, and have been successful in describing correlated systems [26, 27].

1.2.2 Research Avenues

First principles methods, while a mature field, are constantly evolving and developing. The biggest avenue for discovery is first principles excited states methods. While methods like time-dependent DFT [28, 29, 30, 31, 32] have been developed and applied extensively to molecules and solids, excited state analogues for the more accurate QMC and SHCI methods are still under heavy development [33, 34, 35] with application generally restricted to molecules. An exciting prospect is the development of first principles excited states methods across the accuracy/cost spectrum which can be used to accurately and efficiently compute the eigenstates and eigenenergies from small molecules up to solids.

Accompanying this large avenue for change is the constant need for adaptation and development of methods to keep up with the rapid rate of novel materials discoveries. A good case study is the application of DFT to newly discovered layered van Der Waals materials like h-BN and WS_2. DFT computation on layered materials using state-of-the-art functionals with no van Der Waals corrections yield extremely inaccurate results, with 20 - 30% overestimation of ground state inter-layer spacings and upwards of 90% underestimation of between-layer bonding energies [36]. The inaccuracy of state-of-the-art functionals spurred on the development of various density funtionals with non-local van Der Waals corrections, such as vDW-DF [37] and vDW-DF2 [38], which describe the inter-layer spacing and bonding energies within 10% [36]. This pattern of material discovery and method development yields a ubiquitous arena for discovery and advancement in first principles techniques.

The primary research direction for effective models is the accurate computation of effective interactions like electron-electron and electron-lattice interactions. Effective interactions are fundamental to understanding real materials as many systems cannot be described accurately without them. Systems which require effective interactions to accurately describe include, but are not limited to, magnetic systems, superconducting systems, and systems with metal-insulating or structural phase transitions. The aforementioned cRPA method can be used to compute effective interactions, but has the drawback of being a piecewise approach: the effective interactions are treated on a separate footing to the effective hopping terms, and must be carefully stitched together. A newer method named density matrix downfolding (DMD) [39] approaches effective model building statistically and can treat effective interactions and hoppings on equal footing, but has had sparse application to real systems. As such, two avenues of research are evident: first, the continued development of new methods for computing effective interactions which overcome the systematic drawbacks of current techniques like cRPA, and second, the usage of new techniques on relevant and novel materials to demonstrate their capabilities in accurately computing effective interactions.

1.3 Thesis outline

My thesis work pushes forward our ability to accurately compute ground and excited states as well as effective model Hamiltonians with interactions from first principles calculations. In particular, I have worked on improving the accuracy and efficiency of QMC ground and excited state calculations, and using DFT calculations to develop a tight-binding effective Hamiltonian with lattice interactions.

In Chapters 2 and 3, I present a review of the density functional theory (DFT) and quantum Monte Carlo (QMC) methods which I have used extensively in this thesis. For DFT, the formal theory of Hohenberg and Kohn is presented, as well as the practical formulation of most DFT codes via the Kohn-Sham equations. For QMC, the variational and diffusion Monte Carlo methods are presented both formally and with the practical formulation present in most QMC codes, with a discussion of necessary error estimation and extrapolation. A discussion of the pros and cons of these methods in practical calculations are also discussed.

In Chapter 4, I present my work on investigating a new trial wave function for use in quantum Monte Carlo [40]. It is well known that the accuracy of QMC calculations is dependent on the accuracy of the trial wave function *ansatz* used to represent the many-body ground state. A long standing challenge has been the development of a trial wave function *ansatz* which can accurately describe orbital relaxation, particularly in transition metal systems [41]. In this work I investigate the benefit of using a multi-determinant trial wave function without requiring the determinants to be orthogonal, a move away from the traditional approach where orthogonal determinants are used, as a compact *ansatz* for describing orbital relaxation. I implemented and benchmarked this new trial wave function on the C_2 molecule, and found significant improvement over the orthogonal multi-determinant wave functions.

In Chapter 5, I present my work on developing a stable statistical estimate for gradients used in QMC wave function optimization. A key step in QMC calculations is optimizing the parameters in a trial wave function *ansatz* to drive it as close to the exact ground state as possible. The current methods for optimization are very efficient unless orbital optimization is required, as in the recurring case of transition metal systems. When orbital optimization is needed, current methods for optimization require computing a gradient which has an infinite statistical error, yielding inefficient optimization. I present the new stable gradient formalization, and compare it to state of the art stable estimation techniques, finding that the new estimation technique has smaller statistical errors than the state of the art, while requiring less computational baggage. As such, it stands as a light-weight statistical estimate which is easy to implement in modern QMC codes. Our method has even motivated further investigation and development of regularized estimation, as seen in a recent study by van Rhijn *et al.* [42].

In Chapter 6, I present my work on creating a novel method for computing excited states in quantum Monte Carlo. While quantum Monte Carlo methods are mature for computing ground states, systematic excited state methods are an active area of development. In this work, I present a new method which extends the simple ground state optimization approach in QMC to an constrained optimization for excited states. Benchmarking on the benzene molecule shows that the new method is highly efficient and accurate, with our excitation energies agreeing with those measured in experiment to 0.2 eV across twelve excited states. The study provides a simple, accurate, and efficient method for computing excited states in QMC.

In Chapter 7, I demonstrate my work building a tight-binding model for twisted bilayer graphene with lattice interactions from density functional theory. I used DFT calculations for various distorted geometries of bilayer graphene to build a local, environment dependent, hopping model for bilayer graphene. The model predicts the hoppings in bilayer graphene more accurately than the state of the art methods, and also yields band structures which match DFT better than the state of the art. From this model we discerned that a previously stated claim - that geometry relaxation is required to describe band flattening in twisted bilayer graphene - was incorrect, and that with our accurate tight binding model band flattening occurs with even rigid geometries. By using first principles methods to inform the model, the model I fit more accurately describes the lattice dependence of hoppings than the state of the art *ansatzes* in use.

In Chapter 8, I summarize the results of the thesis and provide an outlook on their impact on the field and future avenues of work. A condensed version of the thesis is present in Table 1.1.

Chapter	Goal	My Contribution	Publication
4	Improvement of trial wave function quality for QMC calculations.	Non-orthogonal determinant multi-Slater Jastrow wave functions.	**S. Pathak** and L. K. Wagner , J. Chem. Phys. (2018) [40]
5	Improvement of QMC wave function optimization efficiency and stability.	A stable gradient regularization technique for wave function optimization.	**S. Pathak** and L. K. Wagner, AIP Adv. (2020) [43]
6	Construct an accurate, efficient, QMC excited state technique.	A novel penalty-based excited state method for QMC.	**S. Pathak**, B. Busemeyer, J. N.B. Rodrigues, L. K. Wagner, J. Chem. Phys. (2021) [44]
7	Construct an accurate lattice dependent tight-binding model for bilayer graphene.	A lattice dependent tight binding model for bilayer graphene using DFT.	Manuscript complete

Table 1.1: Summary of Thesis. The "Goal" column refers to general questions that motivated the projects in this thesis. "My Contribution" refer to the contributions I have made towards each goal in this Thesis, and the corresponding publications in "Publication."

Chapter 2

Density functional theory

2.1 Hohenberg-Kohn theorems

DFT is an *ab initio* method which solves Eq 1.1 efficiently by mapping the eigenvalue problem for a many-body wavefunction to a functional minimization problem with respect to the simpler electron density. This is accomplished by inverting the following relationship between external potential $v_{ext}(\vec{r})$, ground state wave function Ψ_0, and ground state density $\rho_0(\vec{r})$

$$N, v_{ext}(\vec{r}) \rightarrow \hat{H}_{ab} \rightarrow \Psi_0(\vec{r}_1, ..., \vec{r}_N) \rightarrow \rho_0(\vec{r}), E_0. \tag{2.1}$$

The definitions and justification for this relationship follow. We define the external potential v_{ext} by rearranging Eq 1.1 into its canonical form used in DFT

$$\hat{H}_{ab} = \Big[-\frac{1}{2}\sum_i \nabla_i^2\Big] + \Big[\sum_{i<j}\frac{1}{|\vec{r}_i - \vec{r}_j|}\Big] + \Big[\sum_{I<J}\frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} - \sum_{iJ}\frac{Z_J}{|\vec{r}_i - \vec{R}_J|}\Big] = \hat{T} + \hat{V}_{ee} + \hat{V}_{ext}$$

where $\hat{T}$ is the electron kinetic energy, $\hat{V}_{ee}$ is the electron-electron Coulomb interaction, and $\hat{V}_{ext}$ is the ionic contribution to the Coulomb energy. Noting that $\hat{V}_{ext}$ is a sum of identical terms for each electron i, we rewrite the ab initio Hamiltonian as such

$$\hat{H}_{ab} = (\hat{T} + \hat{V}_{ee}) + \sum_{i=1}^{N} v_{ext}(\vec{r}_i).$$

Now from the form of $\hat{T} + \hat{V}_{ee}$, it is clear that these two terms only depend on the number of electrons in the system N. As such, the entire $\hat{H}_{ab}$ depends only on two quantities: N, the total number of electrons in the system and $v_{ext}(\vec{r})$, the single-electron external potential. Going a step further, any derived quantities from $\hat{H}_{ab}$, such as the ground state wave function $\Psi_0(\vec{r}_1, ..., \vec{r}_N)$, the ground state energy E_0 and ground state electron density ρ_0

$$E_0 = \langle \Psi_0 | \hat{H}_{ab} | \Psi_0 \rangle, \; \rho_0 = N \int d\vec{r}_2, ..., d\vec{r}_N \Psi_0^*(\vec{r}, \vec{r}_2, ..., \vec{r}_N) \Psi_0(\vec{r}, \vec{r}_2, ..., \vec{r}_N),$$

also only depend on N and v_{ext}. Therefore, the relationship in Eq 2.1 follows.

The inverse of Eq 2.1 is the backbone of DFT and was first proven by Hohenberg and Kohn [45] through two theorems.

Theorem 1. *The external potential $v_{ext}(\vec{r})$ is uniquely determined, up to an additive constant, by the ground state electron density $\rho_0(\vec{r})$.*

Proof. We carry out a proof by contradiction. Assume that there are two Hamiltonians corresponding to a system with N electrons, with different external potentials v and v', $v - v' \neq const.$ (and corresponding operators $\hat{V}, \hat{V}'$) which have identical ground state densities $\rho_0(\vec{r})$. Since the Hamiltonians $\hat{H}, \hat{H}'$ are not identical, the ground state wave functions must also be different, so $\Psi_0 \neq \Psi_0'$. By the variational principle of quantum mechanics

$$E_0 = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle < \langle \Psi' | \hat{H} | \Psi' \rangle = \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle = E_0' + \int d\vec{r} \rho_0(\vec{r})(v(\vec{r}) - v'(\vec{r})).$$

Inverting the primed and unprimed quantities and adding the resultant equation to the above, we find the following final relationship $E_0 + E_0' < E_0 + E_0'$, and our proof by contradiction is complete. □

It should be noted that this is a proof of uniqueness and not existence, meaning that for a given arbitrary ρ it is not necessarily true that a corresponding v exists. However, if such a v were to exist, that v would be unique and no other v' (up to an added constant) could yield such a ρ. We call the set of all ρ for which v exists "v-representable."

Theorem 2. *There is a density dependent energy functional for v-representable densities and the exact ground state density is the global minimum of this functional.*

Proof. Since v is a functional of ρ_0, the ground state density, so is the ground state wave function Ψ_0 and any quantities that correspond to it, such as the energy. More explicitly, we see that

$$E_0[\rho_0] = \langle \Psi_0[\rho_0] | \hat{T} + \hat{V}_{ee} + \hat{V} | \Psi_0[\rho_0] \rangle = F_{HK}[\rho_0] + \int d\vec{r} v[\rho_0](\vec{r}) \rho_0(\vec{r})$$

where we have defined a universal functional F_{HK} that does not depend on v_{ext} and can be used for all different electronic systems. This is a subtle functional because if we compute $F_{HK}[\rho]$ for an arbitrary v-representable density, we will actually be computing $\langle \Psi_0[\rho] | \hat{T} + \hat{V}_{ee} | \Psi_0[\rho] \rangle$, which is the expectation value of the kinetic and electron-electron terms with ground state wave function of the unique Hamiltonian which has ground state density ρ. We can generalize this functional by allowing v to be a fixed potential which does not depend on the input density ρ, leading to the Hohenberg-Kohn functional

$$E_v[\rho] = F_{HK}[\rho] + \int d\vec{r} v(\vec{r}) \rho(\vec{r}), \tag{2.2}$$

The Hohenberg-Kohn functional has a variational principle with respect to the density. Consider a potential v and corresponding ground state density ρ_0, as well as a second ground state density ρ such that $\rho \neq rho_0$. If we compute $E_v[\rho]$, we will find that $E_v[\rho] = \langle \Psi_0[\rho] | \hat{T} + \hat{V}_{ee} + \hat{V} | \Psi_0[\rho] \rangle$. Since we know that $\rho \neq \rho_0$, then the corresponding $\Psi_0[\rho] \neq \Psi_0[\rho_0]$, and therefore by the standard variational principle of quantum mechanics, $E_v[\rho] \geq E_v[\rho_0]$. □

The theorems of Hohenberg and Kohn allows us to recast the variational principle of standard quantum mechanics, where the energy is variationally optimized with respect to a high dimensional wave function that takes $3N$ inputs, to a simpler one where the energy is optimized with respect to an electronic density which only takes 3 inputs. Further, they provide an explicit form for the functional that one needs to optimize, the Hohenberg-Kohn functional in Eq 2.2 which has a universal functional F_{HK} that can be used for any system, and a simple system dependent term which accounts for the external potential v. Unfortunately, while the form of E_v is provided, the form for the universal functional F_{HK} is still unknown to this day, and the bulk of research in density functional theory calculations has been the accurate and efficient approximation of this functional.

2.2 Kohn-Sham equations

To address not knowing F_{HK}, Kohn and Sham cleverly separated the terms in the F_{HK} such that the complex portions which account for electronic correlations and exchange were treated as corrections to a simpler functional [46]. The separation they made was as follows

$$F_{HK}[\rho] = T_s[\rho] + \frac{1}{2}\int d\vec{r}d\vec{r'}\,\frac{\rho(\vec{r})\rho(\vec{r'})}{|\vec{r}-\vec{r'}|} + E_{xc}[\rho] \tag{2.3}$$

where T_s is the kinetic energy of a non-interacting electron gas with the given density, the second term is the classical electrostatic energy, and the third term known as the "exchange-correlation" energy $E_{xc} = F_{HK} - T_s - \frac{1}{2}\int d\vec{r}d\vec{r'}\,\frac{\rho(\vec{r})\rho(\vec{r'})}{|\vec{r}-\vec{r'}|}$ is unknown and serves as a correction to two larger terms. While this separation seems arbitrary, extensive benchmarking shows that indeed the bulk of the electronic energy can be captured with just the first two terms and the exact exchange [47]. However, this does not mean that the interesting low-energy behavior is well described by just these terms, as inaccurate treatement of the exchange correlation functional is one of the main reasons DFT fails to describe strongly correlated systems well, as I discuss later.

With this separation in hand, one still must make a choice for an approximation to E_{xc} before any practical calculation can be done. The field of developing, testing, and benchmarking exchange-correlation functionals is extensive, but for now I will introduce just two categories that are generally used. The simplest form of E_{xc} is the local density approximation (LDA) where E_{xc} is defined as a local integral of the density

$$E_{xc}^{LDA} = \int d\vec{r}\epsilon_{xc}[n(\vec{r})]n(\vec{r}).$$

This correlation functional yields reasonable results for solid state systems [45], but fails to describe chemical properties like molecular binding energies accurately, with large errors up to 30kcal/mol

[48, 49, 50]. To address this, the generalized gradient approximation (GGA) was developed where E_{xc} now includes the gradient of the density as well, allowing for more descriptive power

$$E_{xc}^{GGA} = \int d\vec{r}\epsilon_{xc}[n(\vec{r}), \nabla n(\vec{r})]n(\vec{r}).$$

The GGA functionals are efficient and address the shortcomings of LDA functionals, such as accurately describing bond dissociations energies where the LDA functionals could not [49]. Even more complex functionals do exist and are regularly used, such as Meta-GGA and hybrid functionals, with the prior depending on the Kohn-Sham kinetic energy density, and the latter mixing in the exact exchange operator with the GGA exchange correlation term.

The minimization procedure in DFT is simple and follows the same framework as the self-consistent field theory (SCF). Beginning with the Kohn-Sham separated energy functional, we can analytically minimize with respect to the density to get the Kohn-Sham equations

$$\frac{\delta E_v}{\delta \rho(\vec{r})} = 0 \rightarrow (-\frac{1}{2}\nabla^2 + v_{ext}(\vec{r}) + \int d\vec{r}' \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{\delta E_{xc}[\rho]}{\delta \rho(\vec{r})})\phi(\vec{r}) = \epsilon\phi(\vec{r}). \tag{2.4}$$

The solution to this problem is a single Slater determinant wave function composed of single particle orbitals ϕ. The density ρ is the electronic density corresponding to this single Slater determinant. The computation of the single particle orbitals can be done efficiently using the SCF technique where an initial guess for the orbitals ϕ is decided, and the equation solved with the computed densities. The solutions are then updated and the procedure repeated until the solutions have converged in both energy ϵ and orbital ϕ. While nominally an N^3 procedure in computational cost with respect to the number of electrons, efficient approximations can yield linear scaling computational cost with system size [9, 10, 11, 12, 13].

Practically, DFT is a highly efficient method which describes weakly correlated systems accurately, but fails to describe strong correlations well. For example, recent DFT calculations have exceeded 10,000 atoms [9, 10, 11, 12, 13] on calculations of large biological molecules and condensed matter systems. On the other hand, DFT calculations still continue to fail when describing strongly correlated electronic systems. For example, in a recent study Williams *et al.* demonstrated quantitiatively the performance of DFT calculations when describing the total energies of transition metal atoms and transition metal oxide molecules [8]. They found that across various DFT functionals, the typical DFT calculation yielded anywhere between 50 - 90% of the exact correlation energy across these transition metal systems. The study shows that DFT calculations on transition metal systems are generally inaccurate, but more importantly are very inconsistent, with huge variations in accuracy across a set of similar materials. Compare this to a correlated method like diffusion Monte Carlo, where the average accuracy was 90% of the total correlation energy with a 5% variation across materials, and the difference is night and day.

Chapter 3

Quantum Monte Carlo

3.1 Variational Monte Carlo (VMC)

VMC is a numerical technique for approximating the ground state energy of a system which makes use of the variational principle of quantum mechanics. Consider a given system with N particles and *ab initio* Hamiltonian $\hat{H}_{ab}$. The variational principle states that for any given N particle (unnormalized) trial wavefunction, $\Psi_T(\vec{r}_1, ..., \vec{r}_N)$, an upper bound to the ground state energy can be computed

$$\frac{\langle \Psi_T | \hat{H}_{ab} | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} \geq E_0. \tag{3.1}$$

As the approximate wave function Ψ_T gets closer to the exact ground state wave function Ψ_0, the upper bound becomes tighter and a better approximation to the ground state energy. Therefore, a simple procedure for approximating the ground state energy of a system is to generate an accurate trial wave functions and compute the energy expectation value. Unlike DFT, however, the variation principle can work with explicitly correlated wave functions, leading to the high accuracy ceiling of the method. The key premise of the VMC algorithm is the efficient computation of the expectation values and trial wave functions necessary to make this theoretical framework a practically useful tool.

To compute the energy expectation value in Eq 3.1 efficiently, VMC makes use of Monte Carlo integration. The reason for this choice is the high dimensionality, $3N$, of the expectation value integral. If one were to compute this integral using Simpson's rule for quadrature using M mesh points, the error in the computed integral would scale as $M^{-\frac{4}{3N}}$, meaning that resolving errors for larger systems becomes exponentially more challenging [16]. Instead, one can take a statistical approach by rewriting the integral in the following form

$$\frac{\langle\Psi_T|\hat{H}_{ab}|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} = \int d\vec{R}\frac{|\Psi_T(\vec{R})|^2}{\langle\Psi_T|\Psi_T\rangle}\frac{\hat{H}_{ab}\Psi_T}{\Psi_T}(\vec{R}) \simeq \frac{1}{M}\sum_{\vec{R}_i\sim\rho_T, i=1}^{M}\frac{\hat{H}_{ab}\Psi_T}{\Psi_T}(\vec{R}_i), \tag{3.2}$$

and statistically estimating the integral via the mean where $\vec{R} = (\vec{r}_1, ..., \vec{r}_N)$ and M samples $\vec{R}_i$ are drawn from the $3N$ dimensional probability distribution $\rho_T = |\Psi_T(\vec{R})|^2/\langle\Psi_T|\Psi_T\rangle$. This estimate has a *statistical* error which scales as $1/\sqrt{M}$ independent of the dimensionality N, allowing for an efficient, scalable method for computing expectation values at the cost of a statistical error bar.

To compute an accurate trial wave function, one must begin with a parameterization of the many-body trial wave function which can be used to approximate the true ground state. The simplest form of the trial wave function, for fermionic systems, is the Slater determinant:

$$\Psi_S(\vec{R}) = \mathbf{A}(\phi_1(\vec{r}_1), ..., \phi_N(\vec{r}_N)) \tag{3.3}$$

where $\mathbf{A}$ is the N-particle antisymmetrizing operator, and ϕ_i are single particle orbitals with spatial and spin components, typically constructed from DFT or Hartree-Fock. This trial wave function has a major flaw, however, which is that electrons of the opposite spin can have a high probability of being arbitrarily close to each other, increasing the Coulomb repulsion energy between them drastically, a feature which is disallowed for same spin electrons due to the explicit antisymmetry. To address this, a Jastrow factor J is introduced [16, 51, 52], leading to the Slater-Jastrow wave function

$$\Psi_{SJ}(\vec{R}) = e^{J(\vec{r}_1,...,\vec{r}_N)}\Psi_S(\vec{r}_1, ..., \vec{r}_N),\ J = \sum_{i=1}^{N}\chi(\vec{r}_i) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}u(\vec{r}_i, \vec{r}_j). \tag{3.4}$$

The repulsive u term in the Jastrow factor is to explicitly reduce the probability of electrons being near each other and the χ term ensures the charge density distributions around nuclei is not too disturbed by the introduction of u [16]. Other popular wave functions are basic extensions of the Slater-Jastrow form and include the multi-Slater-Jastrow wavefunctions [53, 54, 55, 40, 43, 44]

— where a multi-determinant wave function replaces the single slater determinant in Eq 3.4 — and backflow wavefunctions [56, 57] — where the coordinates $\vec{r}_i$ are transformed to the so called quasiparticle coordinates.

Next, the parameters in the chosen trial wave function must be optimized. Typical parameters optimized in VMC calculations are orbital parameters, the parameters that define ϕ_i from Eq 3.3, the Jastrow parameters, the parameters that define χ, u in Eq 3.4, and determinant coefficients and backflow parameters for the multi-Slater-Jastrow and backflow wavefunctions. Two broad classes of methods exist for optimizing the parameters: energy minimization and variance minimization. The energy minimization approach directly attempts to minimize the total energy expectation value of the trial wave function with respect to the parameters, employing some form of gradient descent using the energy gradients $\partial E/\partial p_i$ with respect to the parameters p_i computed using Monte Carlo integration. Methods that use the energy minimization approach include stochastic reconfiguration [58], Newton's method and the linear method [59]. The other class of techniques are variance minimization techniques, which attempt to minimize the variance of the energy expectation value with respect to the parameters, and are generally much faster than energy minimization techniques but suffer from numerical stability issues [60] and are less accurate than wave functions optimized using energy minimization [61]. With the optimized wave function on hand, one can approximate the ground state energy via Eq 3.2.

3.2 Diffusion Monte Carlo (DMC)

The ultimate goal of DMC is to efficiently compute the exact ground state energy via a mixed estimation

$$\frac{\langle \Psi_0 | \hat{H}_{ab} | \Psi_T \rangle}{\langle \Psi_0 | \Psi_T \rangle} = E_0. \tag{3.5}$$

Here Ψ_T is a trial wave function, typically taken from an optimized VMC calculation, Ψ_0 is the exact ground state wave function of $\hat{H}_{ab}$ and E_0 is the corresponding exact ground state energy. This is a seemingly similar expectation value to Eq 3.1 used in VMC, with two key differences: Ψ_0, the exact ground state of $\hat{H}_{ab}$ is involved, and the mixed estimate on the left exactly equals the ground state energy E_0 independent of Ψ_T, rather than being an upper bound like in VMC. To understand the statistical interpretation of the mixed estimator, we recast it following Eq 3.2

$$\frac{\langle\Psi_0|\hat{H}_{ab}|\Psi_T\rangle}{\langle\Psi_0|\Psi_T\rangle} = \int d\vec{R}\frac{\Psi_T(\vec{R})\Psi_0(\vec{R})}{\langle\Psi_T|\Psi_0\rangle}\frac{\hat{H}_{ab}\Psi_T(\vec{R})}{\Psi_T(\vec{R})} \simeq \frac{1}{M}\sum_{\vec{R}_i\sim\rho_{DMC},i=1}^{M}\frac{\hat{H}_{ab}\Psi_T}{\Psi_T}(\vec{R}_i), \tag{3.6}$$

where $\rho_{DMC} \equiv \frac{\Psi_T\Psi_0}{\langle\Psi_T|\Psi_0\rangle}$. Now we understand a DMC computation as sampling M positions from the mixed distribution ρ_{DMC} to estimate Eq 3.5 using the statistical form in Eq 3.6. Just like in VMC, the resultant estimate of the integral has a statistical error which scales efficiently as $1/\sqrt{M}$. The only hangup in this process is access to ρ_{DMC}, as this distribution requires knowledge of the exact ground state wave function Ψ_0, which is generally unknown.

We use the method of projection to compute Ψ_0, and subsequently ρ_{DMC}. Beginning with Ψ_T, we apply the following projection to get the exact ground state

$$|\Psi_0\rangle \propto \lim_{\tau\to\infty} e^{-\hat{H}_{ab}\tau}|\Psi_T\rangle. \tag{3.7}$$

This expression is exact, and will yield the exact ground state Ψ_0 of the Hamiltonian $\hat{H}_{ab}$ if the trial wave function Ψ_T has a non-zero overlap with Ψ_0. The proof follows

Proof. Begin by expanding Ψ_T into the exact eigenstates of $\hat{H}_{ab}$ labelled Ψ_i and corresponding eigenvalues E_i: $|\Psi_T\rangle = \sum_i \langle\Psi_i|\Psi_T\rangle|\Psi_i\rangle$. Apply the projection operator to this expansion

$$\lim_{\tau\to 0} e^{-H_{ab}}|\Psi_T\rangle = \lim_{\tau\to 0} e^{-E_0\tau}\sum_i e^{-(E_i-E_0)\tau}\langle\Psi_i|\Psi_T\rangle|\Psi_i\rangle \propto |\Psi_0\rangle \text{ if } \langle\Psi_0|\Psi_T\rangle \neq 0$$

with the proportionality appearing as all terms other than Ψ_0 in the summation are exponentially supressed to zero as $(E_i - E_0) > 0\forall\ i > 0$. □

A small adjustment to the projection operation in Eq 3.7 provides a similar projective method for computing ρ_{DMC}. We start by recasting Eq 3.7 to integral form

$$\Psi_0(\vec{R}) \propto \lim_{\tau\to\infty}\int d\vec{R}' g(\vec{R}\leftarrow\vec{R}',\tau)\Psi_T(\vec{R}'),\ g(\vec{R}\leftarrow\vec{R}',\tau) \equiv \langle\vec{R}|e^{-\tau\hat{H}_{ab}}|\vec{R}'\rangle,$$

where g is a Green function corresponding to the exponential projection oprator. Multiplying the left and right hand sides by $\Psi_T(\vec{R})$ and inserting a unity factor of $\Psi_T(\vec{R}')/\Psi_T(\vec{R}')$ inside the integral, we get the following projection equation for ρ_{DMC}

$$\rho_{DMC} \propto \lim_{\tau\to\infty} \int d\vec{R}' G(\vec{R} \leftarrow \vec{R}', \tau)|\Psi_T(\vec{R}')|^2, \; G(\vec{R} \leftarrow \vec{R}', \tau) \equiv \Big[\frac{\Psi_T(\vec{R})}{\Psi_T(\vec{R}')} g(\vec{R} \leftarrow \vec{R}', \tau)\Big]. \tag{3.8}$$

As such, we compute ρ_{DMC} from the initial distribution $|\Psi_T|^2$ by evolving it according to the Green function G for a long time $\tau \gg 1$.

DMC relies on a stochastic representation to efficiently compute the projection Eq 3.8. I will begin by discussing the DMC procedure for bosonic solutions to Eq 1.1, and then generalize to the fermionic problem that ultimately needs to be solved. First, the distribution $|\Psi_T|^2$ is represented as a sum of Dirac delta functions centered on points sampled from the normalized distribution $\rho_T = |\Psi_T|^2/\langle\Psi_T|\Psi_T\rangle$, referred to as walkers

$$|\Psi_T|^2 \sim \sum_{\vec{R}_i \sim \rho_T} \delta(\vec{R} - \vec{R}_i).$$

This representation is passed into the projection operation Eq 3.8 to get the following expression

$$\rho_{DMC}(\vec{R}) \sim \sum_{\vec{R}_i \sim \rho_T} G(\vec{R} \leftarrow \vec{R}', \tau). \tag{3.9}$$

Following the framework of stochastic processes [62], we understand the projection operation in three steps: M walkers are sampled from $|\Psi_T|^2$, they are moved forwards using the transition probability Green function G for a large value of τ, and the final M walkers represent the distribution ρ_{DMC}.

The Green function G, however, has no analytical form for the Hamiltonian Eq 1.1, and must be treated approximately. This is due to the presence of g in the definition of G. We begin by introducing the Trotter-Suzuki decomposition which allows for a simpler sequential form of g at the cost of a cubic time step error that depends on τ:

$$g(\vec{R} \leftarrow \vec{R}', \tau) = \langle \vec{R} | e^{-\tau \hat{H}_{ab}} | \vec{R}' \rangle \simeq \langle \vec{R} | e^{-\tau \hat{V}/2} e^{-\tau \hat{T}} e^{-\tau \hat{V}/2} | \vec{R}' \rangle + O(\tau^3) \tag{3.10}$$

where $\hat{T}, \hat{V}$ are the electron kinetic energy and total potential energy operators from Eq 1.1. The approximate decomposed form in Eq 3.10 has an analytical expression for small values of τ, with an additional τ^3 error [16].

$$\langle \vec{R} | e^{-\tau \hat{V}/2} e^{-\tau \hat{T}} e^{-\tau \hat{V}/2} | \vec{R}' \rangle \simeq (2\pi\tau)^{\frac{-3N}{2}} e^{\left[-\frac{(\vec{R}-\vec{R}')^2}{2\tau}\right]} e^{[-\tau \frac{V(\vec{R})+V(\vec{R}')}{2}]} + O(\tau^3) \equiv \tilde{g}(\vec{R} \leftarrow \vec{R}_i, \tau) + O(\tau^3). \tag{3.11}$$

The pre-factor and first exponential term represent a diffusion process with a corresponding time scale proportional to τ. The second exponential term is a reweighting factor which depends on the change in potential energy before and after moves [16, 17], or a birth/death process in modern codes. As such, the moves dictated by $\tilde{g}$ correspond to a weighted diffusion process for the initial distribution of walkers which eventually converges on ρ_{DMC}.

To ensure that the error in the approximations Eq 3.11 are manageable, we break up the Green function g into $n \gg 1$ sequential projections, leading to the final form of the projection operator used in DMC codes

$$g(\vec{R}_n \leftarrow \vec{R}_0, \tau) = \prod_{i=0}^{n-1} g(\vec{R}_{i+1} \leftarrow \vec{R}_i, \tau/n) \simeq \prod_{i=0}^{n-1} \tilde{g}(\vec{R}_{i+1} \leftarrow \vec{R}_i, \tau/n) + O((\frac{\tau}{n})^3)$$

By extension, we also have an approximate form for G and can carry out the stochastic process in Eq 3.9

$$G(\vec{R}_n \leftarrow \vec{R}_0, \tau) = \prod_{i=0}^{n-1} G(\vec{R}_{i+1} \leftarrow \vec{R}_i, \tau/n) \simeq \prod_{i=0}^{n-1} \frac{\Psi_T(\vec{R}_{i+1})}{\Psi_T(\vec{R}_i)} \tilde{g}(\vec{R}_{i+1} \leftarrow \vec{R}_i, \tau/n) + O((\frac{\tau}{n})^3) \tag{3.12}$$

The remaining cubic time step error can be removed by an extrapolation procedure where the same calculation is done with sequentially smaller τ [63]. The final walker positions after undergoing the stochastic process defined by Eq 3.9 using the approximate form Eq 3.12 can be used in the mixed estimator Eq 3.5 to compute E_0.

Unfortunately, generalization of this simple DMC implementation for fermionic wave functions leads to a fermion sign problem. The main issue is that during the projection steps, ρ_{DMC} can take negative sign and does not represent a true probability distribution. A simple generalization, where the walkers also carry the sign of the distribution and the sign is included in averaging for expectation values, while formally exact, leads to an exponential noise to signal ratio [16, 53] in the statistical estimate for E_0. This occurs as we are computing an average with highly oscillatory samples which flip between negative and positive signs in a very high dimensional integration space [64].

A simple fix to the sign problem is the fixed-node approximation which clamps the wave function nodes during the projection operation to those of the initial trial wave function. This method involves altering the projecting Hamiltonian $\hat{H}_{ab}$ to include an infinite potential barrier at the nodes of Ψ_T, namely where $\Psi_T = 0$. This altered projection ensures that walkers always stay within pockets of wave function with the same sign, effectively making the fixed-node procedure a collection of bosonic projections in each nodal pocket of the fermionic wave function. The tradeoff is that fixed-node DMC (FN-DMC) is not exact, and will only return the exact fermionic ground state energy if Ψ_T has the same nodal surface as the exact ground state Ψ_0. If Ψ_T does not have the same nodal surface as Ψ_0, the FN-DMC energy will be higher than the exact ground state energy, providing a variational principle in FN-DMC depending on the nodal surface of the trial wave function. Reducing the fixed-node error is an open area of research currently and is usually carried out with sophisticated tehcniques like release-node DMC [65] or through systematic improvement of the trial wave function quality [66].

A typical workflow that I use for a complete FN-DMC calculations is below, including the resolution of the various systematic errors listed above.

1. Carry out a reference DFT or multi-reference SCF calculations

2. Optimize a sequence of trial wave functions of increasing complexity using VMC and the reference SCF calculations, $\Psi_{T,1}, ..., \Psi_{T,k}$

3. For each trial wavefunction
    (a) Compute FN-DMC energies using a sequence of decreasing $\tau < 0.1 Ha^{-1}$
    (b) Carry out time-step extrapolation to get extrapolated FN-DMC energies
4. Deduce the fixed-node error from the FN-DMC results
5. Carry out release-node DMC if the fixed-node error is significant

Practically, VMC and FN-DMC are efficient methods which can describe strongly correlated systems accurately from small systems like molecules to moderately large solids. For example, while highly computationally efficient methods like DFT can simulate 10,000 atom systems [9, 10, 11, 12, 13], they fail to describe strong correlation effects like magnetism and strongly correlated transition metal systems accurately [47, 48]. The other end of the spectrum are highly expensive methods like semi-stochastic heat-bath configuration interaction [67] which are the state of the art methods for accuracy even for strongly correlated systems, but are prohibitively expensive on large systems like solids [8]. QMC methods like VMC and FN-DMC address these concerns, and perform extremely well in describing magnetic states [18, 21], transition metal systems [8, 17], and even more exotic systems like unconventional superconductors [19, 20], on simulations up to about 500 atoms [68]. The efficiency of QMC comes from their polynomial scaling with the VMC algorithm scaling as N^3 and FN-DMC, in its simplest form, scaling as N^7 [16]. QMC methods are thus well equipped primarily for calculations on large systems that require strong electronic correlation to describe accurately.

Chapter 4

Non-orthogonal determinant trial wave functions in quantum Monte Carlo

This chapter is based off of my published work titled "Non-orthogonal determinants in multi-Slater-Jastrow trial wave functions for fixed-node diffusion Monte Carlo" [40].

4.1 Introduction

A major theoretical challenge in the FN-DMC computation of ground state properties of many-body quantum systems is the accuracy of the trial wave function nodal surface. As discussed in the previous chapter on QMC, the quality of the trial wave function appears in the fixed-node variational principle present in FN-DMC. Reiterating this principle, if the nodes of the trial wave function matches those of the exact ground state, the FN-DMC calculation will yield the exact ground state energy. For a trial wave function where the nodes deviate from those of the ground state, the FN-DMC energy will over estimate the ground state energy, and the error is called the fixed-node error arising from the fixed-node variational principle. As such, the accuracy of the FN-DMC calculation is determined by the nodal surface of the VMC trial wave function fed in to the calculation.

Constructing trial wave functions for FN-DMC that accurately reproduce the ground state nodal structure is extremely challenging due to the high dimensionality of the many-body Hilbert space. Compact parameterizations of many-body states can be used to generate trial wave functions that accurately, but still approximately, reproduce the ground state nodal structure. A common choice of the trial wave function for *ab initio* FN-DMC calculations is the multi-Slater-Jastrow (MSJ) form. [69] Since the nodal structure is solely determined by the multi-Slater part of a MSJ trial wave function, improved nodal surfaces can be achieved by optimizing the single particle orbitals within the multiple Slater determinants.

There have been a number of ways of improving the trial wave function for QMC in the literature. Among these, Wagner and Mitas [41] show that the accuracy of calculations for ground state energies and dipole moments of transition metal monoxides in FN-DMC can be increased by using single particle orbitals from a DFT calculation with the B3LYP functional over those from a ROHF calculation. The orbitals from the DFT calculation can be thought of as relaxed, or optimized, versions of the ROHF orbitals. Toulouse and Umrigar describe an efficient parameterization for orbital optimization using orbital rotations[70, 71, 72] and demonstrate the increased accuracy of DMC calculations for the ground state energy of a C_2 molecule when using an MSJ trial wave function with optimized orbitals. Both of these approaches to optimization maintain the orthogonality of determinants in the multi-Slater expansion.

On the other hand, results from multi-configuration calculations on some test atoms and molecules indicate that full-CI calculation accuracy can be achieved using wave functions with fewer optimized non-orthogonal determinants than if the determinants are kept orthogonal. [73, 74, 75] Optimized non-orthogonal Slater determinant expansions without a Jastrow have been used to accurately calculate bonding properties of molecules in valence bond theory [76], ground state correlations of small molecules using Hartree-Fock with symmetry-projected wave functions [77, 78, 79], and correlation energies on the 1-d Hubbard model using resonanting Hartree-Fock with spin-projected wave functions. [80, 81] However, so far non-orthogonal determinants with a Jastrow have not been extensively explored as trial wave functions for FN-DMC calculations.

In this paper, we investigate MSJ trial wave functions with non-orthogonal determinants for FN-DMC calculations, in which identical orbitals in different determinants are optimized independently. We conduct FN-DMC calculations for the ground state energy of a C_2 molecule on the first-principles Hamiltonian with electronic core potentials (ECPs) using three kinds of trial wave functions: MSJ states, MSJ states with optimized orthogonal determinants (MSJ+O), and MSJ states with optimized non-orthogonal determinants (MSJ+NO). We used the QWalk[82] package for the QMC calculations, in which we implemented orthogonal and non-orthogonal optimization

Table 4.1: Wave function *ansatzes* and their corresponding variational parameters.

Name	Form	Parameters
MSJ	$\psi = e^{J(\vec{\alpha})} \sum_I C_I \vert D_I(\{\phi\})\rangle$	$\vec{\alpha}, \vec{C}$
MSJ+O	$\psi = e^{J(\vec{\alpha})} e^{\Theta(\vec{\theta})} \sum_I C_I \vert D_I(\{\phi\})\rangle$	$\vec{\alpha}, \vec{C}, \vec{\theta}$
MSJ+NO	$\psi = e^{J(\vec{\alpha})} \sum_I e^{\Theta(\vec{\theta_I})} C_I \vert D_I(\{\phi\})\rangle$	$\vec{\alpha}, \vec{C}, \{\vec{\theta_I}\}$

for this project. We find that the MSJ+NO trial wave function yields improvements to the ground state FN-DMC energies and single particle properties in addition to the improvements from using the MSJ+O trial wave function.

4.2 Methods

As for all calculations referred to in this thesis, we will be solving the *ab-initio* Hamiltonian Eq 1.1. Electronic core potentials (ECPs) from Burkatski *et al* [83, 84] were used to eliminate the core electrons. We considered three different wave function *ansatzes*, also called trial wave functions, summarized in Table 4.1.

The simplest trial wave functions used are of MSJ form with optimized Jastrow and determinant coefficients:

$$\psi_{MSJ}(\vec{\alpha}, \vec{C}) = e^{J(\vec{\alpha})} \sum_I C_I |D_I(\{\phi\})\rangle. \tag{4.1}$$

Here e^J represents the Jastrow factor which depends on Jastrow parameters α followed by a multi-Slater determinant expansion with coefficients C composed of single particle orbitals ϕ, in this case from a restricted open Hartree-Fock calculation. We constructed fifteen different MSJ states by attaching three-body Jastrow factors given in Wagner and Mitas [82] to distinct multi-Slater states calculated in PySCF. [85] These determinant expansions were constructed using a complete active space method with four active electrons per carbon atom and an active space N_{cas} of 6,8, and 10. Determinants were selected with coefficient weights $|C_I|$ greater than 0.05, 0.025, 0.0175, 0.01, and 0.0075, after which we saw little improvement in the total energy. This procedure generated 15 starting MSJ states, whose energies are reported in Fig 4.1 as MSJ.

The MSJ+O trial wave functions build upon the MSJ trial wave functions by including orthogonal orbital rotations, the simplest form of orbital relaxation we will consider. To each of the fifteen MSJ states we now additionally apply an orbital rotation operator following Toulouse *et al*[72],

$$\psi_{MSJ+O}(\vec{\alpha},\vec{C},\vec{\theta}) = e^{J(\vec{\alpha})}e^{\Theta(\vec{\theta})}\sum_I C_I|D_I(\{\phi\})\rangle. \tag{4.2}$$

The rotation operator $e^{\Theta(\vec{\theta})}$ acts on occupied orbitals as a rotation in orbital space

$$\phi_i \rightarrow \sum_j [e^{\Theta(\vec{\theta})}]_{ij}\phi_j \tag{4.3}$$

where the set $\{\phi_j\}$ is composed of select single particle orbitals that have the same symmetry as the occupied orbital ϕ_i. We included in the set $\{\phi_j\}$ all occupied orbitals with the same symmetry as orbital ϕ_i and the lowest energy unoccupied orbital with the same symmetry. The rotation operator acts on different spin channels independently, therefore the optimized state may be spin contaminated: it may not be an eigenstate of the total S^2 operator. Since the rotation operator acts identically on each determinant in a given spin channel, the determinants will remain orthogonal after optimization, so we refer to such a state as an MSJ state with optimized orthogonal determinants (MSJ+O state).

The MSJ+NO trial wave functions build upon the MSJ+O trial wave functions, allowing for non-orthogonal orbital rotations. We relax the orthogonality of determinants by moving the orbital rotation operator within the sum, such that

$$\psi_{\mathrm{MSJ+NO}}(\vec{\alpha},\vec{C},\{\vec{\theta_I}\}) = e^{J(\vec{\alpha})}\sum_I e^{\Theta(\vec{\theta_I})}C_I|D_I(\{\phi\})\rangle. \tag{4.4}$$

Unlike the MSJ+O case, MSJ+NO *ansatz* allows orbitals in different determinants to be optimized independently, potentially allowing for a more compact wave function.

For each trial wave function, we optimized the parameters listed in Table 4.1 using the linear method described in Toulouse *et al.*[70] Filippi *et al* describe an efficient method for calculating, in quantum Monte Carlo, the parameter derivatives required for the optimization. [86] These optimized trial wave functions were then used in FN-DMC calculations. We performed FN-DMC using T-moves[87] with a timestep of $\tau = 0.01$. This choice of time step leads to a time step error of approximately 1mHa, which is an order of magnitude smaller than the differences in FN-DMC energies we are concerned with.

Alongside energies, we also calculated VMC and FN-DMC single particle densities. In FN-DMC there is a systematic mixed-estimator error that affects the calculation of all quantities which do not commute with H, such as the charge density. In order to account for the first order of this mixed-estimator bias we extrapolate the FN-DMC single particle densities using the formula $\rho_{extrap} = 2\rho_{FN-DMC} - \rho_{VMC}$. [69] This bias does not appear in VMC.

4.3 Results

Fig. 4.1 shows the variational Monte Carlo (VMC) and FN-DMC energies for the multiple trial wave functions described earlier. The VMC energies were calculated after the optimization of the variational parameters in the trial wave functions. The method for calculation and $\mathrm{N}_{\mathrm{cas}}$ are shown in the subtitles. The average improvement in variational energy when using optimized orthogonal determinants is $\langle E_{\mathrm{MSJ}} - E_{\mathrm{MSJ+O}} \rangle$ =0.32 eV with a standard deviation of 0.08 eV, where the average $\langle\rangle$ is over states at all $\mathrm{N}_{\mathrm{cas}}$, $\mathrm{N}_{\mathrm{det}}$. The FN-DMC average improvement is smaller, at 0.14 eV, with a standard deviation of 0.03 eV. The significant improvement in FN-DMC energy resulting from orthogonally optimizing determinants in trial wave functions indicates the importance of orbital relaxation for accurate QMC calculations.

MSJ+NO trial wave functions provide additional improvements to the ground state energies. The average improvement in VMC energy is $\langle E_{\mathrm{MSJ+O}} - E_{\mathrm{MSJ+NO}} \rangle = 0.031$eV (standard deviation 0.022 eV) and 0.032 eV (standard deviation 0.019 eV) for the FN-DMC energy. We believe the large standard deviation may arise from two main sources. First, different trial functions will have different changes in energy due to the non-orthogonal orbital rotation. For example, closer to a complete multi-determinant expansion of the ground state, ($\mathrm{N}_{\mathrm{det}}, \mathrm{N}_{\mathrm{cas}} \to \infty$), the decrease in energy

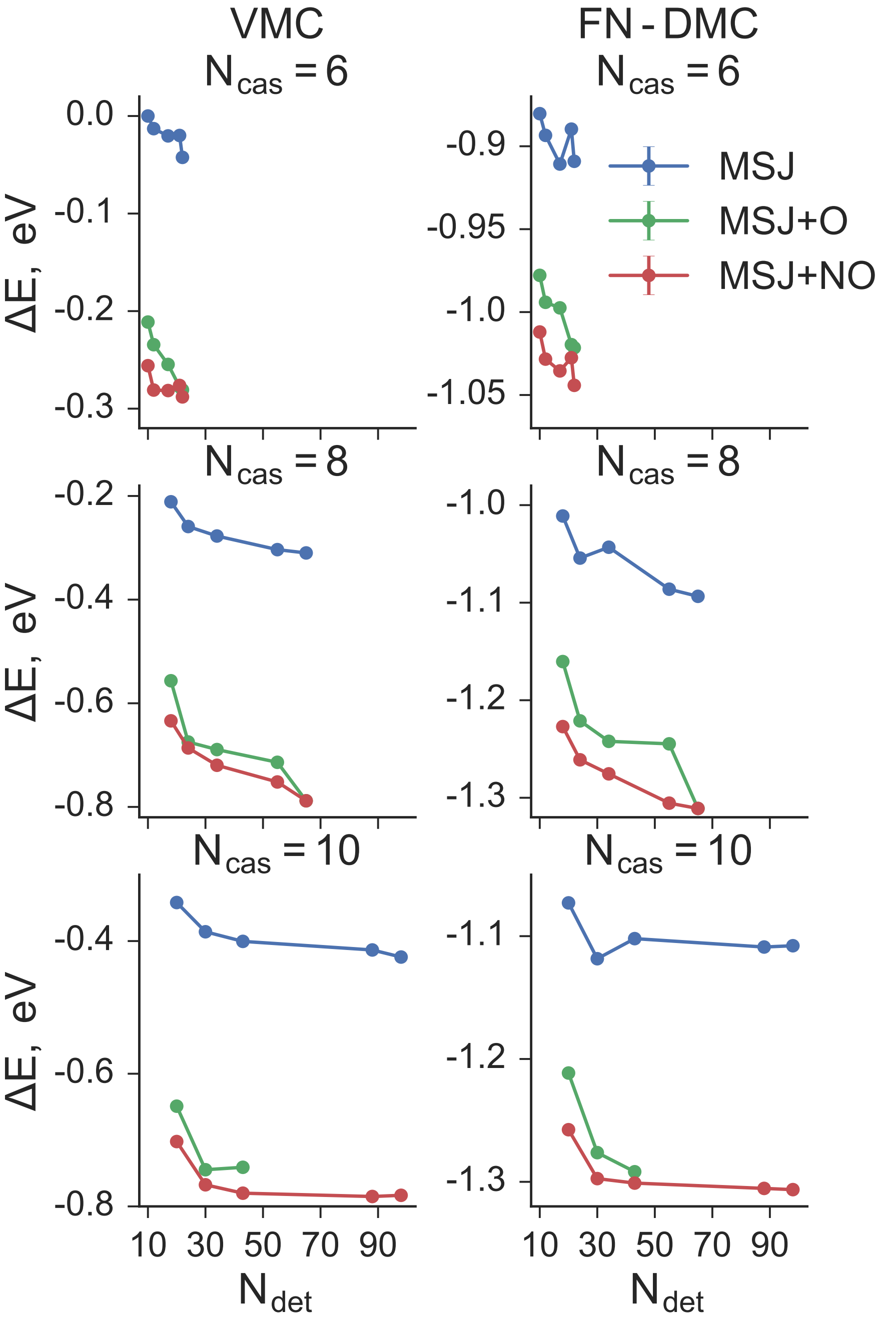


Figure 4.1: Optimized VMC energies and FN-DMC energies for trial wave functions with differing active space sizes (N_{cas}) and number of determinants (N_{det}). All energies are presented relative to the optimized VMC energy of the MSJ state with N_{cas}=6, N_{det}=10; $\Delta E = E - E^{VMC}_{MSJ}[N_{cas}=6, N_{det}=10]$. For each choice of N_{cas}, N_{det} there are three trial functions, MSJ, MSJ+O and MSJ+NO with sequentially lower VMC/FN-DMC energies.

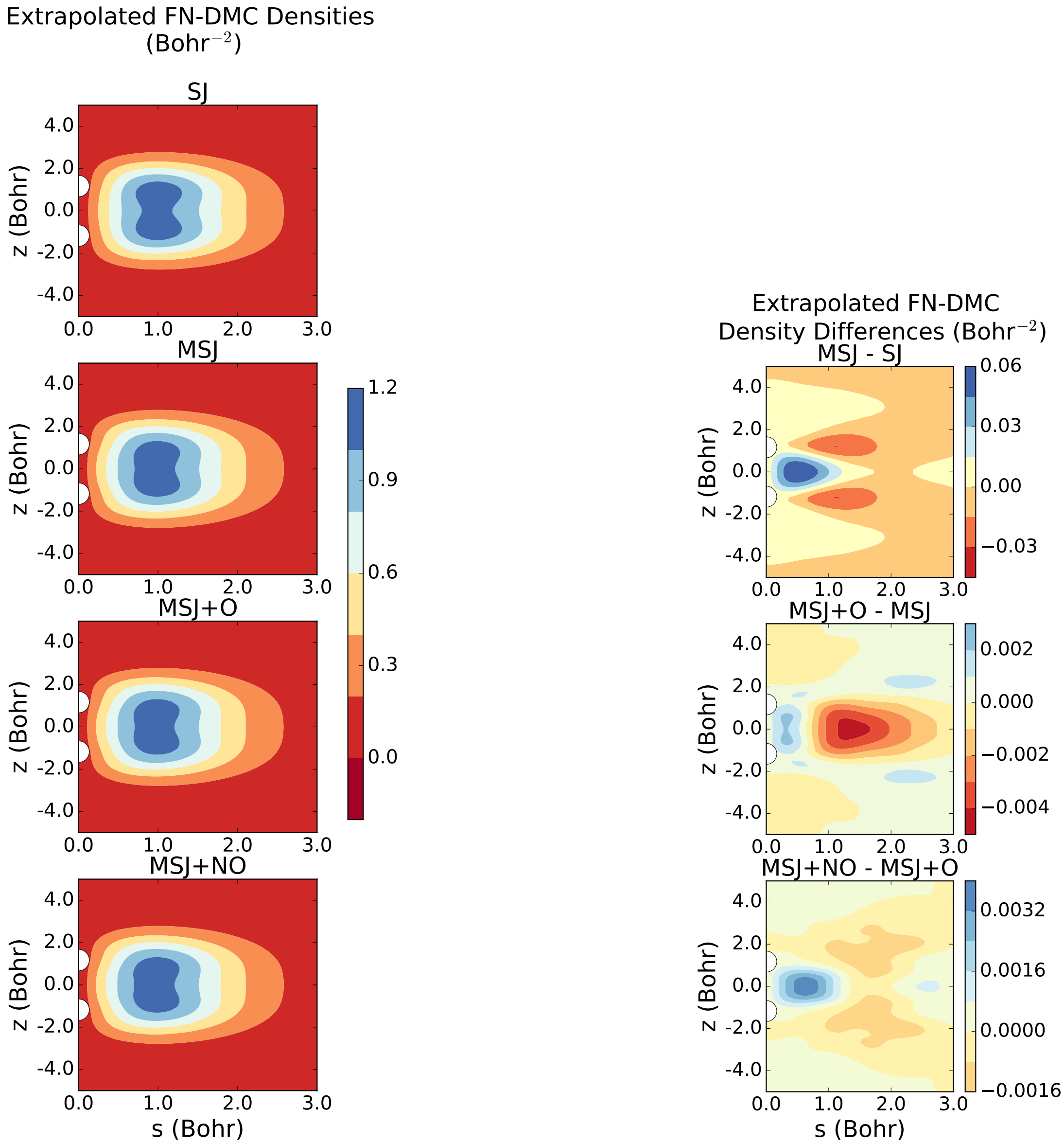


Figure 4.2: The left column shows cylindrically averaged ($\int d\phi\, \rho(s,\phi,z)\, 2\pi s$) FN-DMC charge densities for various optimized trial wave functions, with all multi-Slater trial functions built with $\mathrm{N_{cas}}$=10, $\mathrm{N_{det}}$=43. Here s is the radial coordinate, z is the z coordinate, and the cylindrical averaging was done over the azimuthal angle ϕ. The right column shows differences in the charge densities in the left column.

due to non-orthogonal orbital optimization will approach zero. On the other hand, some finite multi-determinant expansion will have a non-zero decrease in energy due to this optimization. Second, since we are now allowing for independent rotation parameters for each determinant, the variational space becomes $\mathrm{N}_{\mathrm{det}}$ times more complex than the space for orthogonal orbital optimization. A more complex variational space means that the linear optimization may more frequently get caught in local minima. This effect does not appear in our calculations given the nearly monotonic decrease in VMC and FN-DMC energies as we increase $\mathrm{N}_{\mathrm{det}}$ and $\mathrm{N}_{\mathrm{cas}}$.

Further, the MSJ+NO trial wave functions can achieve similar FN-DMC energies with fewer determinants when compared with MSJ+O trial wave functions. For example, the FN-DMC energy when using an MSJ+NO trial wave function with 24 determinants is lower than when using an MSJ+O trial wave function with 55 determinants with active space size $\mathrm{N}_{\mathrm{cas}}$=8. The introduction of optimized non-orthogonal determinants can therefore increase the compactness of QMC trial wave functions without sacrificing accuracy.

To understand the effect of the improved trial functions, we compared the one particle density for each wave function type. The first column of plots in Fig 4.2 shows the cylindrically averaged charge densities calculated in FN-DMC with mixed-estimator extrapolation for a Slater-Jastrow and MSJ, MSJ+O and MSJ+NO trial functions with $\mathrm{N}_{\mathrm{cas}}$=10, $\mathrm{N}_{\mathrm{det}}$=43. The second column in Fig 4.2 presents differences in these charge densities. The white semicircles represent the C atoms.

Going from from the SJ to MSJ trial function, the extrapolated charge density increases between the two carbon atoms. This increase in bonding character makes sense since the ground state of C_2 is multi-reference in character. [88] Orbital optimization continues this trend; however, the subsequent increase in charge density in the bonding region is an order of magnitude smaller than the increase between SJ and MSJ. This result is reasonable, since introducing correlation into trial wave functions allows for electrons to avoid each other while still occupying the bonding region. The extrapolated spin densities exhibit spin contamination for the MSJ, MSJ+O and MSJ+NO trial functions, but the magnitude of the spin contamination is $\mathrm{O}(10^{-4})$ Bohr^{-2}, an order of magnitude smaller than the charge redistribution.

4.4 Conclusion

We find that using MSJ+NO trial wave functions yield improvements to the ground state FN-DMC energy and single particle properties of a C_2 molecule in addition to the improvements from using MSJ+O trial wave functions. For example, the FN-DMC energy calculated using an MSJ+NO trial function with only 24 determinants is lower than the FN-DMC energy using an MSJ+O trial function with 55 determinants. Compared to an average decrease in the FN-DMC energy of 0.14 eV (standard deviation 0.03 eV) when using MSJ+O trial wave functions, using MSJ+NO trial wave functions provide an additional average reduction of 0.032 eV (standard deviation 0.019 eV). Using trial wave functions with either orthogonally or non-orthogonally optimized determinants increases the bonding character of the FN-DMC charge density when compared to the density using a bare MSJ trial wave function. Our results indicate that using non-orthogonal determinants in multi-Slater expansions may lead to more compact multi-Slater-Jastrow trial wave functions for small molecules and can help reduce the fixed-node error in FN-DMC by improving the ground state nodal surface.

Chapter 5

A light-weight regularization for wavefunction parameter gradients in quantum Monte Carlo

This chapter is based on my published work titled ”A light weight regularization for wave function parameter gradients in quantum Monte Carlo” [43].

5.1 Introduction

The optimization of the trial wavefunction parameters is an integral task in accurate VMC calculations. In energy optimization techniques, the parameters $\vec{p}$ in a wavefunction $\Psi(\vec{R}, \vec{p})$ are optimized to minimize the energy expectation value $E(\vec{p}) \equiv \langle\Psi(\vec{p})|\hat{H}|\Psi(\vec{p})\rangle / \langle\Psi(\vec{p})|\Psi(\vec{p})\rangle$. This is accomplished through iterative approaches where a stochastic estimate of the gradient $\partial E(\vec{p})/\partial p$ is used in updating the parameter p at each iteration [89, 90, 70, 91, 71, 72].

The simplest estimator for $\partial E/\partial p$, Eqn 5.1, comes from the derivative of the energy expectation value. This estimator is unbiased and has a zero-variance principle for any wave function parameter p if $\Psi(\vec{R}, \vec{p})$ is an eigenstate of $\hat{H}$. However, the variance of this estimator diverges when Ψ is not an exact eigenstate and p is a parameter which affects the nodes of the wave function, as occurs in the case of orbital or determinant coefficients[1, 4].

Previous to this work, the infinite variance has been removed in several ways. One method is to sample from distributions other than the trial wave function squared.[1, 2, 3] Alternatively, one can modify the estimator using an auxillary wave function.[92, 93, 94] However, both of these approaches require a choice of guiding or auxillary functions to implement the algorithm.

In this paper, we derive and test a simple regularized estimator for $\partial E/\partial p$ which has finite variance and can be efficiently extrapolated to zero bias. Instead of guiding wave functions, the divergent variance is removed by regularizing the naïve estimator within a distance ϵ of the nodes of $\Psi(\vec{p}, \vec{R})$. We show that the bias scales as ϵ^3 and the variance scales as $1/\epsilon$. For realistic systems, we test that the predicted scaling is seen in a real calculation, show that extrapolation to zero bias is straightforward and requires no extra computation, and conduct an energy optimization calculation on a multi-Slater-Jastrow trial wave function.

5.2 Regularized estimator

The naïve estimator for $\partial E/\partial p$ evaluated on a wave function $\Psi(\vec{R}, \vec{p})$ is the random variable $\hat{\theta}(\vec{R}, \vec{p})$

$$\hat{\theta}(\vec{R}, \vec{p}) \equiv 2\,\mathrm{Re}\left(\left(E_L(\vec{R}) - \left\langle E_L(\vec{R})\right\rangle\right)\frac{\partial_p \Psi(\vec{R}, \vec{p})}{\Psi(\vec{R}, \vec{p})}\right), \tag{5.1}$$

where $E_L = (\hat{H}\Psi)/\Psi$ is the local energy with Hamiltonian operator $\hat{H}$ and the brackets $\langle\ \rangle$ denote a quantum mechanical expectation value over Ψ. This estimator has zero bias for any wave function Ψ and parameter p over the distribution $|\Psi|^2$, but can acquire a divergent variance. The divergent contribution to the variance arises from the evaluation of the integral

$$\int \left| E_L \frac{\partial_p \Psi}{\Psi}\right|^2 |\Psi|^2 dR. \tag{5.2}$$

when Ψ is not an eigenstate of $\hat{H}$ and the parameter p affects the nodes of Ψ, such as orbital or determinantal coefficients. In this case, to lowest order in distance $|\vec{R} - \vec{N}|$ from a nodal point $\vec{N}$ of Ψ, $\hat{H}\Psi$ and $\partial_p\Psi \simeq const$ and $\Psi \simeq \nabla\Psi(\vec{N}) \cdot (\vec{R} - \vec{N})$, leading to the integrand in Eqn 5.2 behaving as $1/|\vec{R} - \vec{N}|^2$ near the node. Since the integration domain includes the nodal surface of Ψ, the $1/|\vec{R} - \vec{N}|^2$ behavior of the integrand as $|\vec{R} - \vec{N}| \to 0$ leads to a divergent integral.

We obtain a finite variance estimator for $\partial E/\partial p$ by regularizing the naïve estimator of Eqn 5.1 by a function f_ϵ:

$$\hat{\theta}_\epsilon(\vec{R}, \vec{p}) \equiv 2\,\mathrm{Re}\left(\left(E_L(\vec{R}) - \left\langle E_L(\vec{R})\right\rangle\right)\frac{\partial_p \Psi(\vec{R}, \vec{p})}{\Psi(\vec{R}, \vec{p})}\right) f_\epsilon(\vec{R}), \tag{5.3}$$

where

$$
\begin{aligned}
f_\epsilon(\vec{R}) &= \begin{cases} 7|\frac{\vec{x}}{\epsilon}|^6 - 20|\frac{\vec{x}}{\epsilon}|^4 + 9|\frac{\vec{x}}{\epsilon}|^2 & |\frac{\vec{x}}{\epsilon}| < 1 \\ 1 & |\frac{\vec{x}}{\epsilon}| \geq 1 \end{cases} \\
\vec{x} &\equiv \frac{\Psi(\vec{R},\vec{p})^* \nabla\Psi(\vec{R},\vec{p})}{|\nabla\Psi(\vec{R},\vec{p})|^2}.
\end{aligned} \tag{5.4}
$$

The quantity $|\vec{x}|$ is the normal distance between $\vec{R}$ and nearest nodal point of Ψ, ϵ is a parameter which defines how far away from the node the regularization should be carried out, and ∇ is the many-body gradient. Details on the construction of f_ϵ can be found in the Appendix.

In evaluating the variance of Eqn 5.3, the divergent integral Eqn 5.2 is replaced by a regularized integral

$$
\int \left| E_L \frac{\partial_p \Psi}{\Psi} \right|^2 |\Psi|^2 f_\epsilon^2 dR. \tag{5.5}
$$

Unlike the integrand of Eqn 5.2, the regularized integrand does not diverge near the nodes of Ψ. This is seen by Taylor expanding the integrand to lowest non-vanishing order in $|\vec{R} - \vec{N}|$. Substituting the limiting relationships for $\hat{H}\Psi$, $\partial_p\Psi$ and Ψ from above and noting

$$
\vec{x} \simeq \left[\nabla\Psi(\vec{N})^* \cdot (\vec{R} - \vec{N})\right] \frac{\nabla\Psi(\vec{N})}{|\nabla\Psi(\vec{N})|^2} \tag{5.6}
$$

as $|\vec{R} - \vec{N}| \to 0$, we find the limiting behavior of the regularized integrand near the nodes of Ψ:

$$
\lim_{|\vec{R}-\vec{N}|\to 0} \left| E_L \frac{\partial_p \Psi}{\Psi} \right|^2 f_\epsilon^2 |\Psi|^2 \propto \frac{|\nabla\Psi(\vec{N}) \cdot (\vec{R} - \vec{N})|^2}{|\nabla\Psi(\vec{N})|^4 \epsilon^4} \to 0. \tag{5.7}
$$

The regularized estimator thus removes the divergent contribution to the variance present in Eqn 5.2 and results in a finite variance estimation of $\partial E/\partial p$ for any $\epsilon > 0$.

To lowest order in ϵ, the variance of Eqn 5.3 decreases as $1/\epsilon$. This can be seen by carrying out the integral Eqn 5.5 as $\epsilon \to 0$. Since the regularization is only present for $|\vec{x}/\epsilon| < 1$, the lowest order scaling with ϵ appears in the integration domain $|\vec{x}/\epsilon| < 1$. Within this domain, the limit $\epsilon \to 0$ is equivalent to $|\vec{R} - \vec{N}| \to 0$, and the limiting integrand of Eqn 5.7 can be used. Making the change of variables $\vec{R} \to (l, \vec{N})$ where l is the normal coordinate to the nodal surface and $\vec{N}$ is the coordinate of the nodal surface, Eqn 5.5 to lowest order in ϵ is

$$\propto \int dN \frac{1}{|\nabla\Psi(\vec{N})|^2} \int_{-\epsilon}^{\epsilon} dl \frac{l^2}{\epsilon^4} \propto \frac{1}{\epsilon}. \tag{5.8}$$

The suppressed variance of the estimate comes at a cost, in the form of cubic bias $O(\epsilon^3)$ to lowest order in ϵ. The bias of the regularized estimator is

$$\int_{|\vec{x}/\epsilon|<1} (\hat{H}\Psi)(\partial_p\Psi)(f_\epsilon - 1)dR. \tag{5.9}$$

Following the analysis of the previous paragraph, we evaluate the bias to lowest non-vanishing order in ϵ by Taylor expanding the integrand in powers of $|\vec{R} - \vec{N}|$ and calculating the bias at each order in the expansion. The constant term in the Taylor expansion does not contribute to the bias by parity of the integrand. Further, the choice of coefficients for f_ϵ in Eqn 5.4 ensures the first order term vanishes, as proven in the appendix. The lowest non-vanishing contribution to the bias is due to term in the Taylor expansion $\propto l^2$, yielding an integral which scales as ϵ^3.

A zero-bias estimation of $\partial E/\partial p$ using the reguralized estimator Eqn 5.3 can be carried out using VMC in four steps

1. Conduct a standard VMC calculation to collect M configurations $\{\vec{R}_i\}_{i=1}^M$ from $|\Psi|^2$.
2. Evaluate $|\vec{x}_i|$ from Eqn 5.4 for each configuration.
3. Calculate the average of the estimator Eqn 5.3 over the collected configurations for a sequence of $\epsilon \to 0$.
4. Fit a function $a\epsilon^3 + b$ to the estimates; the intercept b is the zero bias, finite variance estimate for $\partial E/\partial p$.

This procedure can be carried out very efficiently as only a single set of VMC configurations must be drawn for the entire extrapolation, resulting in an inexpensive, finite-variance, zero bias estimate of $\partial E/\partial p$.

5.3 Application to LiH, CuO molecules

We verify the predicted mathematical behavior of $\hat{\theta}_\epsilon$ by computing $\partial E/\partial p$ for a determinantal coefficient of an unoptimized multi-Slater Jastrow (MSJ) wave function for the LiH molecule:

$$\Psi(\vec{c}, \vec{\alpha}) = e^{J(\vec{\alpha})} \sum_i c_i |D_i\rangle. \tag{5.10}$$

The determinants $|D_i\rangle$ and coefficients c_i were taken from a full configuration interaction (CI) expansion over the Li 1s, 2s, 2p and H 1s orbitals. The orbitals were constructed from a restricted open-shell Hartree Fock (ROHF) calculation using a correlation consistent quadruple-zeta valence basis set[95]. A 2-body Jastrow factor $J(\vec{\alpha})$ of the form in[82] was used with $\vec{\alpha} = 0$ except for electron-electron cusp conditions. The ROHF and CI calculations were done using the PySCF package[96] and all QMC calculations were carried out using `pyqmc`[97].

We begin by verifying that the regularized integral in Eqn 5.5 has finite value across a node for $\epsilon > 0$. To do so, we evaluate the regularized integrand for various ϵ along a path which passes through a node $\vec{N}$,

$$\vec{R}(x) = \vec{N} + l\nabla\Psi(\vec{N})/|\nabla\Psi(\vec{N})|, \tag{5.11}$$

with $l \in [-0.1, 0.1]$ Bohr. The results are shown in the lower plot of Figure 5.1. For all values of ϵ, the value of the integrand is pushed to zero at the node, removing the divergence in the integral. The sharp increase in the integrand near $l = 0$ as ϵ decreases is indicative of the increase in the value of the integral, and hence estimator variance, as $\epsilon \to 0$. Shown in the upper plot is the first term in Eqn 5.3 across the node, exhibiting no divergences.

By integrating across the normal coordinate l, the predicted $O(1/\epsilon)$, $O(\epsilon^3)$ scalings of the variance and bias are shown Figure 5.2. The integration for both the bias and variance are carried out for the path in Figure 5.1 from $l = -0.1$ to 0.1 Bohr and normalized by a factor $\int |\Psi|^2 dR$ along

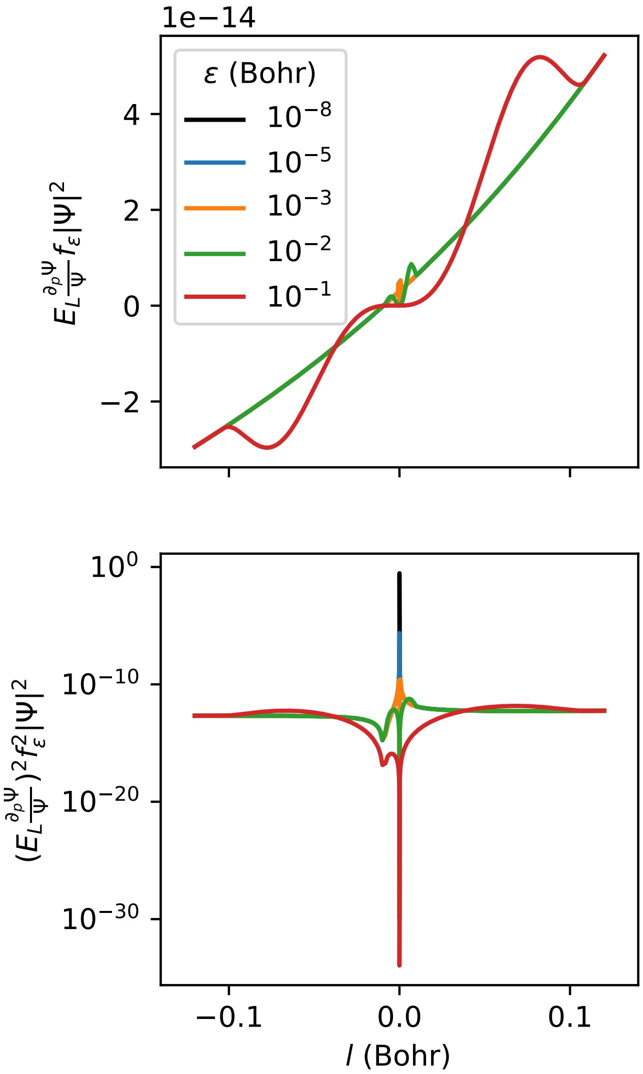


Figure 5.1: $E_L \frac{\partial_p \Psi}{\Psi} f_\epsilon$ and logarithm $(E_L \frac{\partial_p \Psi}{\Psi} f_\epsilon)^2$, plotted against the normal coordinate l from a node of Ψ. Curve colors correspond to different values of ϵ ranging from 10^{-1} to 10^{-8} Bohr.

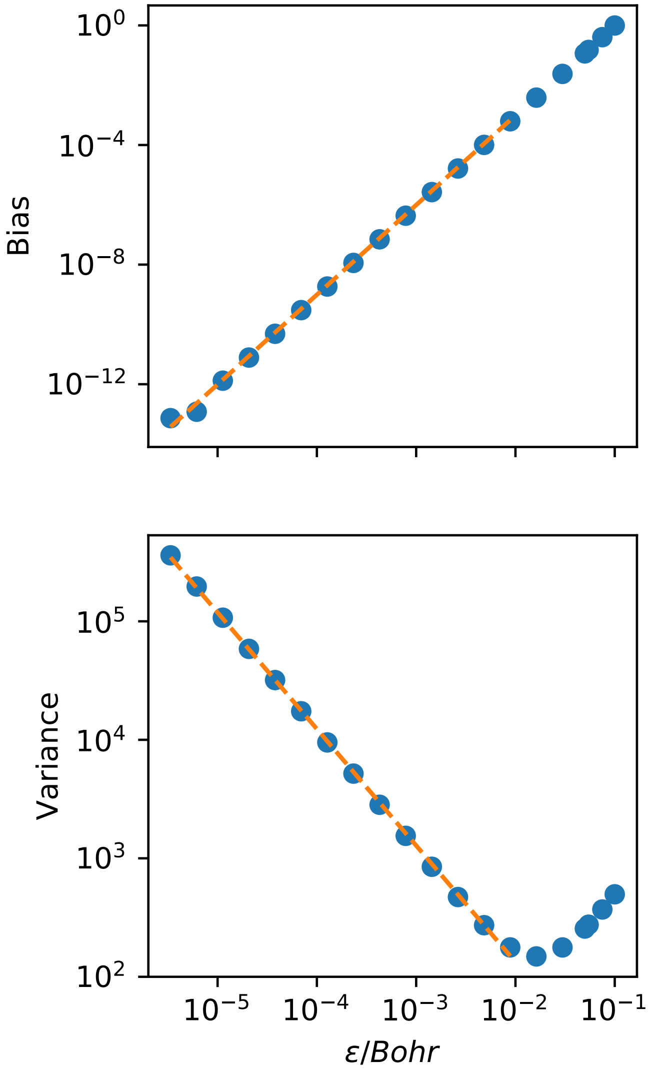


Figure 5.2: Scaled bias and variance of $E_L \frac{\partial_p \Psi}{\Psi} f_\epsilon$ evaluated by numerical integration from $l = -0.1$ to $l = 0.1$ across the node in Figure 5.1. The blue dots are the numerically integrated values and the orange curves indicate best fits to the functions $a\epsilon^3$ and $b + c/\epsilon$ for the bias and variance, respectively.

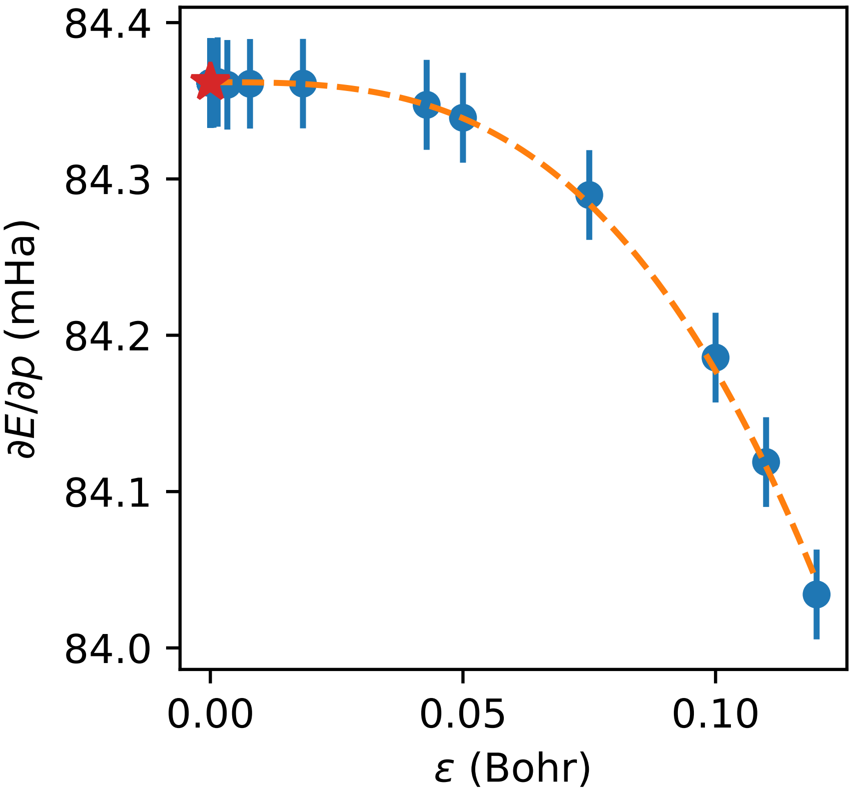
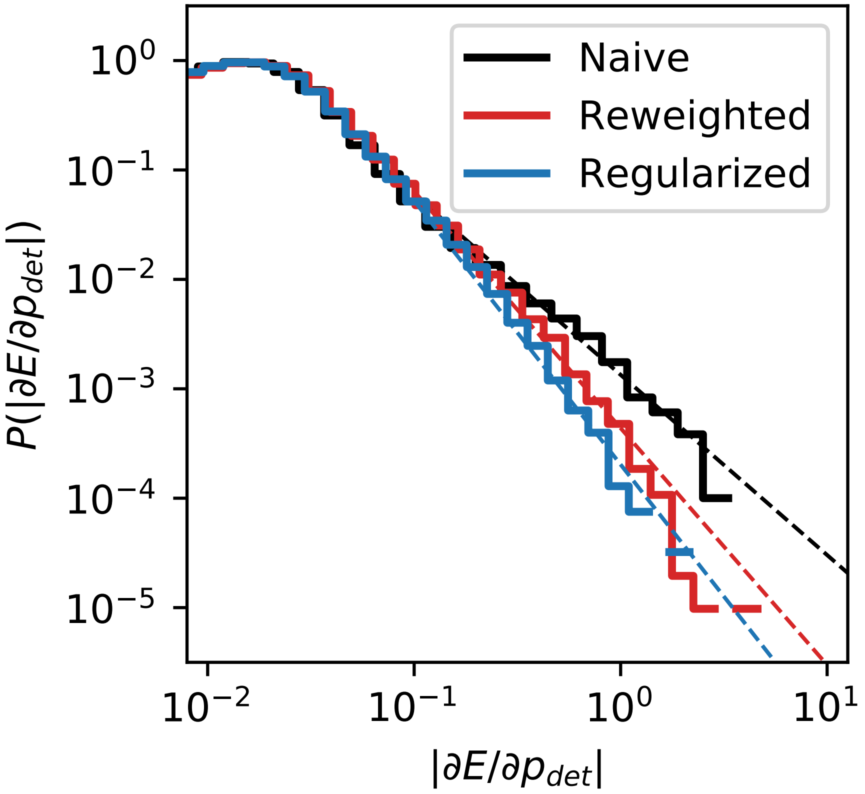
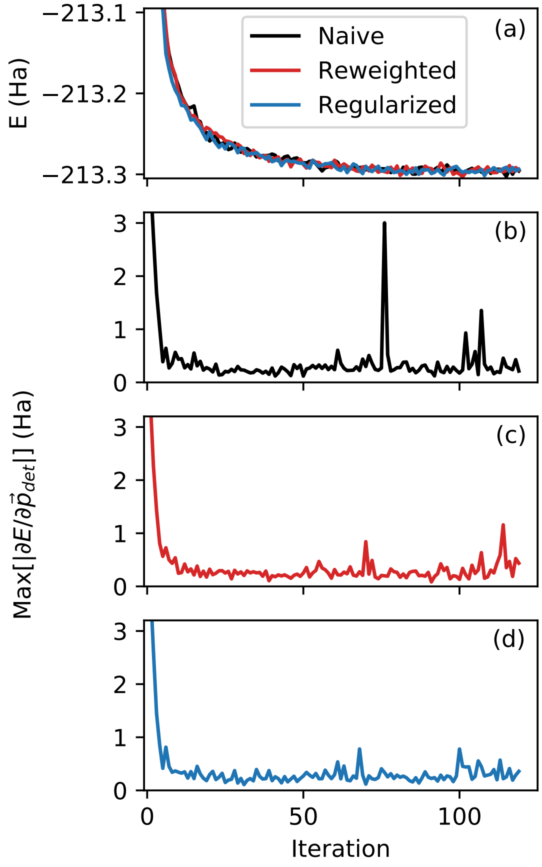


Figure 5.3: Zero bias, finite variance extrapolation for $\partial E/\partial p$ using the regularized estimator. The blue points are evaluated using VMC, the orange curve is a fit to $a + b\epsilon^3$, and the red star denotes the extrapolated estimation.



Estimator	**Tail exponent**
Naïve	$-1.65(5)$
Reweighted	$-2.16(5)$
Regularized	$-2.41(6)$

Figure 5.4: Histogram of the gradient magnitudes evaluated using VMC with respect to determinantal parameters over the last twenty iterations of the energy optimization calculation on the CuO molecule. Dashed lines are constructed by linear regression to the power-law tails of the distributions. The fit tail exponents are reported in the table.



Estimator	**Energy (Ha)**
Naïve	$-213.296(3)$
Reweighted	$-213.295(3)$
Regularized	$-213.295(2)$

Figure 5.5: Results from optimization of a 10^4 determinant multi-Slater-Jastrow wave function for the CuO molecule using three different estimators: the naïve estimator Eqn 5.1, regularized estimator Eqn 5.3, and reweighted estimator[1, 2, 3] using Umrigar's guiding wave function[4]. Each curve is an average over five independent calculations. Panel (a) shows the energy versus iteration curve, while panels (b)–(d) present the maximum absolute value component of the estimated gradient vector $\partial E/\partial \vec{p}_{det}$ over determinantal coefficients. The table shows the average energy over the last 20 iterations of the optimization for the three different estimators.

that path. The predicted cubic bias is observed for six decades in ϵ while the $O(1/\epsilon)$ decrease in variance stands for four decades. The increase in the variance after $\epsilon = 10^{-2}$ Bohr occurs due to the breakdown of the assumption $\Psi(\vec{R}) \simeq \nabla\Psi(\vec{N}) \cdot (\vec{R} - \vec{N})$, resulting in a linear increase in variance with ϵ. This change is not mirrored in the bias since the sub-leading order scaling with ϵ appears when the assumptions of $\hat{H}\Psi$ and $\partial_p \Psi \simeq const$ break down.

Next, we carry out the four step, zero-bias, finite variance extrapolation for $\partial E/\partial p$ proposed in the previous section, shown in Figure 5.3. First, a standard VMC calculation with 200,000 steps and 2,000 configurations per step was carried out. Then, $\partial E/\partial p$ was estimated using Eqn 5.3 for ϵ between 1.2×10^{-1} Bohr and 10^{-5} Bohr using the VMC configurations, shown in the data points. Since the same configurations were used for each evaluation, the estimates have a strong statistical correlation and the predicted ϵ^3 bias is clearly present, shown by the orange fit curve. The intercept of the fit curve is shown by the red star and is the zero-bias estimate of $\partial E/\partial p$. The variance can be deduced by the error bar of the blue data points, in this case 0.025 mHa. Since the bias is zero within statistical errors for small values of ϵ, most practical calculations can be carried out for a fixed value of ϵ between 10^{-5} and 10^{-2} Bohr without extrapolation.

We conclude by comparing the effectiveness of the regularized estimator to the commonly used reweighting technique[1, 2, 3] in the optimization of a 10^4 determinant multi-Slater-Jastrow trial wave function for the CuO molecule. The trial wave function takes the form Eqn 5.10, with a 2-body Jastrow factor as in Ref [82]. Orbitals were constructed from a ROHF calculation using the Trail-Needs triple-zeta basis and electronic core potentials[98]. The determinants were selected by taking the 10^4 highest weighted determinants from a 920,205 determinant semi-stochastic heat-bath configuration interaction (SHCI) calculation [99] over an active space of four core and thirteen active orbitals. The ROHF and SHCI calculations were done using the PySCF[96] and Arrow packages[99, 100, 101].

The optimization was performed using the stochastic reconfiguration method (SR)[102] for VMC energy minimization. The time step was determined at each iteration by a line search along the update direction using correlated sampling, referred to as a stabilization by Toulouse and Umrigar [70]. To accelerate convergence, the parameters were optimized in batches. In the first 50 iterations only the 10^2 largest determinant coefficients from the SHCI calculation were updated, the next 50 iterations included the top 10^3 determinants, and the last 20 iterations included all 10^4 determinants. The energy optimization was carried out in `pyqmc`[97]. This identical procedure was carried out for all gradient estimators.

Additionally, the finite variance estimators were applied to the SR overlap matrix to ensure the estimation of the matrix had finite variance. The SR overlap matrix has elements $S_{ij} = \langle \partial_{p_i}(\Psi/\sqrt{\langle\Psi|\Psi\rangle}) | \partial_{p_j}(\Psi/\sqrt{\langle\Psi|\Psi\rangle})\rangle$ where ∂_{p_i} is the gradient of the wave function with respect to parameter p_i. These matrix elements exhibit an identical divergence in the variance present in Eqn 5.2 with the replacement of E_L with $\partial_{p_i}\Psi/\Psi$. To remove this divergent variance in optimization calculations using the regularized estimator, S_{ij} was estimated using the finite variance estimator with $\epsilon = 10^{-2}$

$$\left(\Big(\frac{\partial_{p_i}\Psi(\vec{R},\vec{p})}{\Psi(\vec{R},\vec{p})} - \left\langle \frac{\partial_{p_i}\Psi(\vec{R},\vec{p})}{\Psi(\vec{R},\vec{p})} \right\rangle\Big) \frac{\partial_{p_j}\Psi(\vec{R},\vec{p})}{\Psi(\vec{R},\vec{p})}\right) f_\epsilon(\vec{R}). \tag{5.12}$$

In optimization calculations using the reweighted estimator, finite variance estimation of S_{ij} was carried out using Umrigar's guiding wave function [4] with $\epsilon = 10^{-2}$. Only in the optimization calculations using the naïve estimator did the estimation of S_{ij} have infinite variance.

Figure 5.5 shows the results of the energy optimization using the naïve estimator Eqn 5.1, the regularized estimator Eqn 5.3 with $\epsilon = 10^{-2}$, and the reweighted estimator using Umrigar's guiding wave function[4] with $\epsilon = 10^{-2}$. Each optimization step was performed with 16,000 walkers and 100 Monte Carlo steps. Each curve is an average over five independent optimization calculations. Panel (a) shows the energy versus iteration, and the table below shows the average and standard deviation of the energy over the last 20 iterations. The final energies for the three different estimators are equal within error bars, indicating that the choice of gradient estimator does not affect the final energy of the optimized wave function. We attribute this similar long-timestep performance due to the line minimization step, which stabilizes occasional large fluctuations in the estimation of the gradient. However, as one can verify from close inspection of the energy graph, some optimization steps are wasted due to fluctuations in the naïve estimator.

Now we turn to the gradient. At each step of the optimization, the full gradient vector $\partial E/\partial\vec{p}$

was estimated using 16,000 walkers and 100 Monte Carlo steps for both Jastrow and determinantal coefficients. Since the infinite variance is limited to parameters which can change the nodal surface of $\Psi(\vec{R}, \vec{p})$, we restrict our analysis to $\partial E/\partial \vec{p}_{det}$ which only includes gradients with respect to determinantal coefficients. To visualize fluctuations in the gradient estimation, we plot the maximum absolute value component of $\partial E/\partial \vec{p}_{det}$ estimated at each iteration in Figure 5.5(b)–(d). The regularized and reweighted estimator have smaller fluctuations in the gradient maximum than the bare estimation. However, small spikes remain in both the regularized and reweighted estimator, indicating that the distribution of gradients may have heavy tails.

In Figure 5.4 we present histograms of the absolute value of the components of $\partial E/\partial \vec{p}_{det}$ estimated during the last twenty iterations of optimization. Each distribution is constructed from $5\times20\times10{,}000$ gradient estimates, corresponding to the number of independent optimization calculations, iterations per calculation, and gradient components evaluated per iteration. The distributions for all three techniques have power-law tails, reflective of the many small spikes seen in Figure 5.5. The naïve estimator has an infinite variance tail with a slope of $-1.65(5) > -2$. The regularized and reweighted estimators both have finite variance tails with exponent < -2. The regularized estimator provides a lower variance estimation than the reweighted estimator as evidenced by the tail exponents $-2.41(6)$ and $-2.16(5)$.

Although the reweighted and regularized estimates provide finite-variance estimates for $\partial E/\partial p$, the convergence of the distribution of sample means to a normal distribution is much slower than the convergence for E. Compare this to the local energy $E_L = \frac{H\Psi}{\Psi}$, where the distribution of sample means are Gaussian with just 10^3 configurations[103]. Therefore, even after regularization or reweighting, there are non-Gaussian fluctuations in the mean of $\partial E/\partial p$ for typical numbers of samples, as seen in Fig 5.5. Even when the gradients are regularized, it is thus advantageous to use adaptive updates, such as a line search, to control the effects of these fluctuations.

5.4 Conclusion

In this work, we derived and tested a simple regularized estimator for $\partial E/\partial p$ which has finite variance and can be extrapolated to zero bias. The divergent variance present in the naïve estimator is suppressed via multiplication by a polynomial function within a distance ϵ of the nodes of $\Psi(\vec{p}, \vec{R})$. The regularized estimator has a finite variance by incurring a cubic bias which can be efficiently extrapolated to zero bias. We demonstrate this by carrying out a finite variance, zero-bias estimate of $\partial E/\partial p$ for a determinantal coefficient in a trial wave function for the LiH molecule. Additionally, we carried out energy optimization calculations on the CuO molecule for a multi-Slater-Jastrow wave function using the bare, regularized, and commonly used reweighted estimators[1, 2, 3]. We found that the regularized and reweighted estimators lead to finite-variance, heavy-tailed distributions of $\partial E/\partial p$ with the regularized estimator having a more negative power-law exponent than the reweighted estimator for the same cutoff value.

The regularized estimator is a simple alternative to known finite-variance estimation techniques for $\partial E/\partial p$. These popular techniques require guiding[1, 2, 3] or auxiliary wave functions[92, 93, 94] in order to estimate $\partial E/\partial p$ for the trial wave function $\Psi(\vec{R}, \vec{p})$. The regularized estimator does not require these additional wave functions, leading to a simple algorithm for the finite variance, zero-bias, estimation of $\partial E/\partial p$ using only $\Psi(\vec{R}, \vec{p})$.

Chapter 6

A novel penalty method for excited states in quantum Monte Carlo

This chapter is based off my published work titled ”Excited states in variation Monte Carlo using a penalty method” [44].

6.1 Introduction

An important task in condensed matter physics is the accurate *ab initio* computation of excited states, particularly using the moderately scaling and accurate QMC techniques. Generally, this is accomplished through extensions of VMC and DMC from just ground state to excited state calculation. In the earliest case, Bernu and Ceperley proposed an algorithm based on diffusion Monte Carlo,[104] which has not been applied to many practical cases, since it has a sign problem which leads to exponential scaling in the system size. Blunt et al.[105] proposed an algorithm using full configuration interaction, which is also exponentially scaling.

On the low-scaling side, Filippi and coworkers[106, 107, 108] implemented a method similar to state averaged CASSCF in VMC, and demonstrated the technique on impressively large wave function expansions. However, state averaging is not optimal when the optimal ground and excited orbitals are very different,[109] as often occurs in strongly correlated systems. Neuscamman and coworkers[110, 111, 112, 113] have proposed a low-scaling method that instead uses alternate objective functions to optimize excited states. While this technique does not suffer from the state averaging problem, it so far has only been applied to very few excited states, and can experience difficulties converging to the correct state.[114]

In this manuscript, we implement and demonstrate a simple penalty method based on orthogonalizing to lower energy states to compute excited states using variational Monte Carlo. Similar techniques are commonly used in density matrix renormalization group calculations,[115] but to our knowledge have not been applied in the variational Monte Carlo context. This technique obtains excited states one by one by enforcing orthogonality to lower energy states, and can optimize general wave function parameters, including orbital parameters as shown here. The scaling of the technique is $\mathcal{O}(N_{ex}N_e^M) + c\mathcal{O}(N_{ex}^2 N_e^M)$, where N_e is the number of electrons, N_{ex} is the number of excited states computed, c is a small constant, and M is dependent on the wave function, 3 for a Slater-Jastrow wave function. The orthogonalization based technique allows for access to multiple excited states, and does not require state averaging. The technique is implemented in the `pyqmc` package, available online.[97] The method is applied to benzene with a $\sim$10,000 parameter wave function, showing high accuracy compared to experiment and coupled cluster calculations on 12 excited states. The state-specific orbital optimization made possible by the new technique provides an efficient and stable technique for accurate excited states using QMC.

6.2 Penalty-based optimization using variational Monte Carlo

The method solves for each energy eigenstate one at a time, by following the procedure outlined here:

1. First, the stochastic reconfiguration method[116] is used to find the VMC approximation to the ground state, $|\Psi_0\rangle$.
2. The first excited state $|\Psi_1\rangle$ is found by optimizing the objective function Eqn 6.7 with $\vec{S}^* = [0]$ and λ set larger than the expected $E_1 - E_0$. The algorithm is not very sensitive to the value of λ, so we typically use λ of order 1 Hartree, substantially higher than excitation energies in the systems considered here.
3. The second excited state is found by optimizing the objective function Eqn 6.7 with $\vec{S}^* = [0, 0]$ and anchor states $|\Psi_0\rangle$ and $|\Psi_1\rangle$.

4. Further excited states are found in the same way, by orthogonalizing to the ones found in the previous steps.

6.2.1 Objective function

As diagrammed in Fig 6.1, it is straightforward to show that if $|\Phi_0\rangle$ is the ground state of the Hamiltonian, then so long as $\lambda > E_1 - E_0$ the function

$$\arg\min_{\Psi}(E[\Psi] + \lambda N_\Psi^2 N_{\Phi_0}^2 | \langle \Psi|\Phi_0\rangle |^2) \tag{6.1}$$

is equal to the first excited state $|\Phi_1\rangle$, where $E[\Psi]$ is the expectation value of the energy and $N_\Psi = 1/\sqrt{|\langle\Psi|\Psi\rangle|}$.

For completeness, we show this here. Consider the objective functional

$$O[\Psi] = E[\Psi] + N_\Psi^2 N_{\Phi_0}^2 \lambda_i |\langle\Psi|\Phi_0\rangle|^2 \tag{6.2}$$

Then the functional derivative of each term is given by

$$\begin{aligned}\frac{\delta N_\Psi^2 N_{\Phi_0}^2 | \langle\Psi|\Phi_0\rangle |^2}{\delta\Psi^*} &= N_\Psi^2 (N_{\Phi_0}^2 \langle\Phi_0|\Psi\rangle|\Phi_0\rangle \\ &\quad - N_\Psi^2 N_{\Phi_0}^2 | \langle\Psi|\Phi_0\rangle |^2 |\Psi\rangle)\end{aligned} \tag{6.3}$$

and

$$\frac{\delta E[\Psi]}{\delta\Psi^*} = N_\Psi^2 (H - E[\Psi])|\Psi\rangle. \tag{6.4}$$

We have highlighted the unpaired kets in blue for clarity. Combining the two, and setting $\frac{\delta O}{\delta \Psi^*}$ to zero, we obtain

$$\begin{aligned}&(H - E[\Psi] - \lambda N_\Psi^2 N_{\Phi_0}^2 | \langle\Psi|\Phi_0\rangle |^2)|\Psi\rangle \\ &\qquad + \lambda N_{\Phi_0}^2 \langle\Phi_0|\Psi\rangle|\Phi_0\rangle = 0\end{aligned} \tag{6.5}$$

The $H - E[\Psi]$ term ensures that this equation can only be satisfied by an an energy eigenstate $|\Phi_i\rangle$. For $i \neq 0$, $|\Psi\rangle = |\Phi_i\rangle$ is a solution because $\langle\Phi_i|\Phi_0\rangle = 0$. $|\Phi_0\rangle$ is a solution since $\langle\Phi_0|\Phi_0\rangle = 1$, which allows the λ terms to cancel. The value of the functional $O[\Phi_i] = E_i + \lambda\delta_{i0}$. Therefore, if $\lambda > E_1 - E_0$, the global minimum is at $|\Phi_1\rangle$. Because of the structure demonstrated in Fig 6.1, there are no local minima in the complete Hilbert space. Similarly, the N'th excited state is given by

$$|\Phi_N\rangle = \arg\min_{\Psi} \left(E[\Psi] + \sum_{i}^{N-1} \lambda_i N_\Psi^2 N_{\Phi_0}^2 |\langle\Psi|\Phi_i\rangle|^2 \right). \tag{6.6}$$

as long as $\lambda_i > E_i - E_0$.

We write the algorithm in terms of *anchor states* $|\Psi_i\rangle$, where $i = 0, \ldots, N-1$. These states are fixed during the optimization, and only the parameters of a single wave function $|\Psi\rangle$ are optimized. While ideally the anchor states are energy eigenstates, in general they will be best approximations to them. We also find it useful to consider the objective function

$$O[\Psi] = E[\Psi] + \sum_i \lambda_i |\vec{S}_i - \vec{S}_i^{\,*}|^2, \tag{6.7}$$

where

$$S_i = \frac{\langle\Psi|\Psi_i\rangle}{\sqrt{\langle\Psi|\Psi\rangle\langle\Psi_i|\Psi_i\rangle}}, \tag{6.8}$$

and $\vec{S}^*$ is a set of *target* overlaps. For example, to obtain the N'th excited state one would use $N-1$ anchor states each set to the best approximation to the $N-1$ lowest energy eigenstates and set $\vec{S}^*$ equal to a zero vector of $N-1$ length.

6.2.2 Computation of the objective function and its derivatives using variational Monte Carlo

In this section, we will explain how to evaluate the quantities needed using standard variational Monte Carlo techniques.[117] In this implementation, we sample a different distribution for each anchor state:

$$\rho_i(\mathbf{R}) = |\Psi_i(\mathbf{R})|^2 + |\Psi(\mathbf{R})|^2. \tag{6.9}$$

Table 6.1: Implementation of the penalty-based optimization for excited states.

```
1. Choose target overlaps $\vec{S}^*$
2. Initialize $\vec{p_0}$
3. for i in range(nsteps) do
    (a) Compute $N_i$, $\vec{S}$, $\nabla_p N_0$, $\nabla_p \vec{S}$, $E$, $\nabla_p E$
    (b) if abs($N_0$-0.5) > threshold, normalize $\Psi_N$
    (c) Objective function is $O = E + \vec{\lambda} \cdot (\vec{S} - \vec{S}^*)^2$
    (d) Construct gradient $\nabla_p O = \nabla_p E + \nabla_p(\vec{\lambda} \cdot (\vec{S} - \vec{S}^*)^2)$
    (e) $\nabla_p N \leftarrow R^{-1} \nabla_p N$
    (f) $\nabla_p O \leftarrow R^{-1} \nabla_p O$
    (g) $\nabla_p O \leftarrow \nabla_p O - \frac{(\nabla_p O)\cdot(\nabla_p N)}{|\nabla_p N|^2} \nabla_p N$
    (h) $\vec{p}(\tau) \leftarrow \vec{p}_{i-1} - \tau \nabla_p O$
    (i) $\vec{p}_i \leftarrow \arg\min_p (O(\vec{p}(\tau)), \tau)$
```

Then the unnormalized overlap is estimated in Monte Carlo

$$\langle \Psi_j | \Psi_k \rangle_i \simeq \left\langle \frac{\Psi_j^*(\mathbf{R}) \Psi_k(\mathbf{R})}{\rho_i(\mathbf{R})} \right\rangle_{\mathbf{R} \sim \rho_i}, \tag{6.10}$$

where $\mathbf{R}$ is the many-electron coordinate and $\mathbf{R} \sim \rho_i$ means that $\mathbf{R}$ is sampled from the normalized distribution $\frac{\rho_i(\mathbf{R})}{\int \rho_i(\mathbf{R})}$. Here the subscript i indicates that the overlap was estimated using ρ_i.

The relative normalization of the wave function $|\Psi\rangle$ is thus

$$N_i = \langle \Psi | \Psi \rangle_i \,, \tag{6.11}$$

which is evaluated the same way as Eqn 6.10. For a parameter p of $|\Psi\rangle$,

$$\partial_p N_i = 2\mathrm{Re} \left\langle \frac{\partial_p \Psi^*(\mathbf{R}) \Psi(\mathbf{R})}{\rho_i(\mathbf{R})} \right\rangle_{\mathbf{R} \sim \rho_i} \tag{6.12}$$

The overlap is given by

$$S_i = \frac{\langle \Psi | \Psi_i \rangle_i}{A_i}, \tag{6.13}$$

where $A_i = \sqrt{\langle \Psi | \Psi \rangle_i \langle \Psi_i | \Psi_i \rangle_i}$. The derivative of the unnormalized overlap is

$$\langle \partial_p \Psi | \Psi_i \rangle = \left\langle \frac{\partial_p \Psi^*(\mathbf{R}) \Psi_i(\mathbf{R})}{\rho_i(\mathbf{R})} \right\rangle_{\mathbf{R} \sim \rho_i}. \tag{6.14}$$

The derivative of the normalized overlap is computed using the above components as follows

$$\partial_p S_i = \frac{\langle \partial_p \Psi | \Psi_i \rangle_i}{A_i} - \frac{1}{2} \frac{\langle \Psi | \Psi_i \rangle_i}{A_i} \frac{\partial_p N_i}{N_i}. \tag{6.15}$$

Thus, all derivatives can be computed using only the wave function parameter derivatives.

We also regularize all derivatives using the stochastic reconfiguration[116] step to compute

$$R_{pq} = \langle \partial_p \Psi | \partial_q \Psi \rangle , \tag{6.16}$$

for parameter indices p and q.

6.2.3 Practical details

The algorithm is outlined in Table 6.1. We provide comments on steps that have some nontrivial considerations. In step **2**, the parameters are initialized. We typically initialize the parameters using an approximate excited state, typically either from an orbital promotion in a single determinant, or from a small quantum chemistry calculation. We have checked that it is possible to optimize starting from the ground state, but such a strategy is unnecessarily expensive, in particular because the objective function of Eqn 6.7 is a saddle point at the ground state.

In step **3(a)**, all the quantities in Sec 6.2.2 are computed. A Monte Carlo sampling of ρ_i is done for each anchor wave function. The algorithm thus scales mildly with the number of anchor wave functions. The energy and its derivatives are averaged among all samples, so the most costly component of the calculation does not increase much with the number of anchor wave functions.

It is important for the normalization of the wave functions to be similar; otherwise the density in Eqn 6.9 is unbalanced. One could adjust weights in Eqn 6.9, but we find it more convenient to normalize all wave functions. We ensure that all anchor states have the same normalization, and use the first anchor state as a reference. Before performing an optimization move, we first check whether N_0 is too far from $\frac{1}{2}$. The threshold is typically about 0.1. If it is too far, the parameters are rescaled to normalize the wave function and VMC is repeated with the renormalized wave function parameters. This is performed in step **3(b)** in Table 6.1.

In steps **3(e)** and **3(f)**, we regularize the gradients of the normalization and the objective function. It is necessary to regularize both prior to the projection that will come in the next step.

To prevent moves that change the normalization, we project out the derivative of the normalization from the objective function. Otherwise, the moves diverge from the equal normalization manifold, and it becomes difficult to evaluate the overlaps accurately. This is performed in step **3(g)** in Table 6.1.[1]

We find that line minimization, performed in step **3(i)** in Table 6.1, improves the performance of the algorithm significantly. We use correlated sampling to compute the objective function for various values of τ, and fit to a quadratic. We also reject moves that change the relative normalization by more than 0.3.

6.3 Demonstration of excited state optimization using VMC

To demonstrate the technique, we apply it to two cases; H_2 at varying bond lengths to check for correctness versus an exact solution, and the excited states of benzene to demonstrate that it is capable of optimizing about 10,000 parameters on a system with 30 electrons in the calculation.

---

[1] Since we are using multi-Slater-Jastrow wave functions in this work, it was more convenient to control the normalization of the wave function in this way. Another option, not tried here, is to parameterize the wave function in a norm-conserving manner, which could increase the efficiency of the algorithm. On the other hand, for emerging wave functions, it may be more convenient to allow the normalization to vary.

6.3.1 Application to H_2

For application to H_2, the trial wave function was a simple complete active space (CASCI) wave function with 2 electrons and 2 orbitals taken from restricted Hartree-Fock. We take the lowest three excited states from this calculation, labeled $\Psi_{\text{CASCI,i}}$, where i runs from 0 to 2. This wave function was modified using a 2-body Jastrow as parameterized in previous work[118] to obtain

$$\Psi_{\text{CASCI-J,0}} = \Psi_{\text{CASCI,0}} e^{U_0} \tag{6.17}$$

We then optimized the determinant, orbital, and Jastrow parameters using a modified version of stochastic reconfiguration implemented in `pyqmc` to obtain $\Psi_{\text{CASCI-J,0}}$.

We then considered two Jastrow-based approximations to the excited states. The first, which we denote “Fixed,” is commonly used in the literature. It is given by

$$\Psi_{\text{Fixed CASCI-J, i}} = \Psi_{\text{CASCI,i}} e^{U_0}. \tag{6.18}$$

In this fixed wave function, no parameters are optimized at all; that is, the Jastrow is the same as the ground state and the determinant and orbital coefficients are kept fixed. The second, we denote “Optimized,” begins with Eqn 6.18, and uses the penalty method to optimize determinant, orbital, and Jastrow parameters while ensuring orthogonality to lower states. In the Supplementary Information, a Snakemake workflow[119] is provided that performs the calculations shown here in `pyqmc` and `pyscf`.[120] We use a λ of 2 Hartrees for these calculations, which were converged.

For reference values, we used full configuration interaction (Full CI) to compute exact energies in a finite basis. We found that at the triple ζ level, the energies were fairly well converged. This is slightly earlier than most materials, due to the fact that H_2 is very simple.

In Fig 6.2, we demonstrate the targeting capability of this technique. Each point is a wave function generated in the Dunning cc-pvdz[5] basis as described above, with S^* set to various points on the Bloch sphere connecting the first 3 excited states. As expected, the superpositions of low energy wave functions fall on a plane as sketched in Fig 6.1, a critical check that the calculation is creating the desired wave functions.

Comparisons between the CASCI, QMC, and full CI results are presented in Fig 6.3. With a rather compact wave function, the optimized CASCI-J wave function obtains close agreement with the exact calculation, while optimizing significantly from the starting fixed CASCI-J wave function. In the case of the first excited state, simply adding a Jastrow factor to an existing CASCI wave function does not improve the energies at all; optimization is required to obtain accurate results.

6.3.2 Application to benzene excited states

As a demonstration on a larger system, we computed the full π space spectrum of benzene using our excited state optimization method. This set of thirteen states contains rich physical structure, with two different spin channels, single and double electron excitations, and states with ionic bonding in contrast to the covalently bonded ground state.[121] Some of these states, such as the ${}^1\mathrm{E}_{1u}$ state, have been considered strongly correlated by previous authors.[121] As such, these excitations are a standard benchmark set for excited state methods and have been used to validate and compare electronic structure techniques for some time.[122, 123]

To represent the wave functions, we used a multi-Slater-Jastrow wave function parametrization of the form

$$|\Psi(\vec{\alpha}, \vec{c}, \vec{\beta})\rangle = e^{J(\vec{\alpha})} \sum_i c_i |D_i(\vec{\beta})\rangle. \tag{6.19}$$

The Jastrow factor $J(\vec{\alpha})$ is a 2-body Jastrow[118] factor. The determinants $|D_i\rangle$ in the multi-Slater expansion were selected from a minimal CASCI calculation over the six π electrons and six π orbitals in benzene. The single particle orbitals used in the CASCI were computed using density functional theory (DFT) with the B3LYP functional, BFD triple-ζ basis, and BFD pseudo-potential.[124] The CASCI and DFT calculations were carried out using `pyscf`.[120] The parameters in the wave function are: 108 Jastrow parameters $\vec{\alpha}$, 400 determinant coefficients $\vec{c}$, and 9,288 orbital parameters $\vec{\beta}$.

We used the parameterization of Eqn 6.19 to compute the benzene spectra using three different methods, with increasing cost. The first method, denoted fixed, is a standard QMC excited state

Table 6.2: Comparison of theoretically computed excitation energies to experimental values. All values are in eV. Maximum indicates the transition of maximum intensity. The adiabatic and ZPVE corrections are estimated using TDDFT with the PBE functional and 6-31g basis.

	Spectroscopy[122, 125, 126, 127, 128]		**Corrections**		**Vertical excitation**				
State	Maximum	E_{00}	Adiabatic	ZPVE	Experiment	**DMC** $\vec{\alpha}, \vec{c}, \vec{\beta}$	**CC3**[123]	**CASPT2**[121]	**TDDFT-PBE0**[129]
$^1B_{2u}$	4.90	4.72	-0.19	-0.18	5.09	5.15(3)	5.08	4.7	5.39
$^1B_{1u}$	6.20	6.03	-0.19	-0.33	6.55	6.62(4)	6.54	6.1	6.05
$^1E_{1u}$	6.94	6.87	-0.24	-0.32	7.43	7.72(4)	7.13	7.06	7.21
$^1E_{1u}$	6.94	6.87	-0.24	-0.32	7.43	7.63(3)	7.13	7.06	7.21
$^1E_{2g}$	7.80(20)	7.81	-0.21	-0.45	8.47	8.38(3)	8.41	7.77	7.52
$^1E_{2g}$	7.80(20)	7.81	-0.21	-0.45	8.47	8.34(3)	8.41	7.77	7.52
$^3B_{1u}$	3.94	3.65	-0.55	-0.19	4.39	4.15(3)	4.15	3.94	3.82
$^3E_{1u}$	4.76	4.63				4.89(3)	4.86	4.5	4.7
$^3E_{1u}$	4.76	4.63				4.96(4)	4.86	4.5	4.7
$^3B_{2u}$	5.60	5.58				6.08(4)	5.88	5.44	5.05
$^3E_{2g}$	7.49(25)	7.49(25)				7.74(4)	7.51	7.03	7.18
$^3E_{2g}$	7.49(25)	7.49(25)				7.60(4)	7.51	7.03	7.18

technique, where the coefficients $\vec{\alpha}, \vec{c}$ are first optimized on the ground state CASCI root with frozen orbital coefficients $\vec{\beta}$, then the optimized Jastrow is multiplied with higher energy CASCI roots, and finally these trial wave functions are used in VMC to compute excited state energies. This method does not allow for state-specific optimization of any of the parameters $\vec{\alpha}, \vec{c}$, or $\vec{\beta}$.

To understand the effects of orbital optimization in this system, we consider two parameter sets using the penalty technique. In the first, we fix the orbital parameters $\vec{\beta}$ to the DFT ground state orbital coefficients, but allow the other parameters $\vec{\alpha}, \vec{c}$ to be optimized; we denote the wave functions as VMC $\vec{\alpha}, \vec{c}$. Finally, we optimize all coefficients in Eqn 6.19, denoting those wave functions $\vec{\alpha}, \vec{c}, \vec{\beta}$. All QMC calculations were carried out in `pyqmc`.[97]

The results of our excited state computations at the VMC level are shown in Fig 6.4a. We see a consistent 0.2 eV decrease in total energy across all twelve excited states going from the fixed parameter QMC method to the $\vec{\alpha}, \vec{c}$ method using our new optimization technique. We find that optimizing $\vec{\alpha}, \vec{c}, \vec{\beta}$ yields up to a 0.5 eV reduction in total energy relative to the fixed technique across nearly all of the excited states, a gain of around 0.3 eV due to orbital optimization. However, the differences between the excited states and the ground state are very similar between the frozen and optimized orbital calculations, which demonstrates the extent to which standard excited state techniques depend heavily on error cancellation.

We computed the time-step extrapolated DMC energies for the computed excited state wave functions, which are shown in Fig 6.4b. The additional benefit of optimizing the orbitals in DMC is

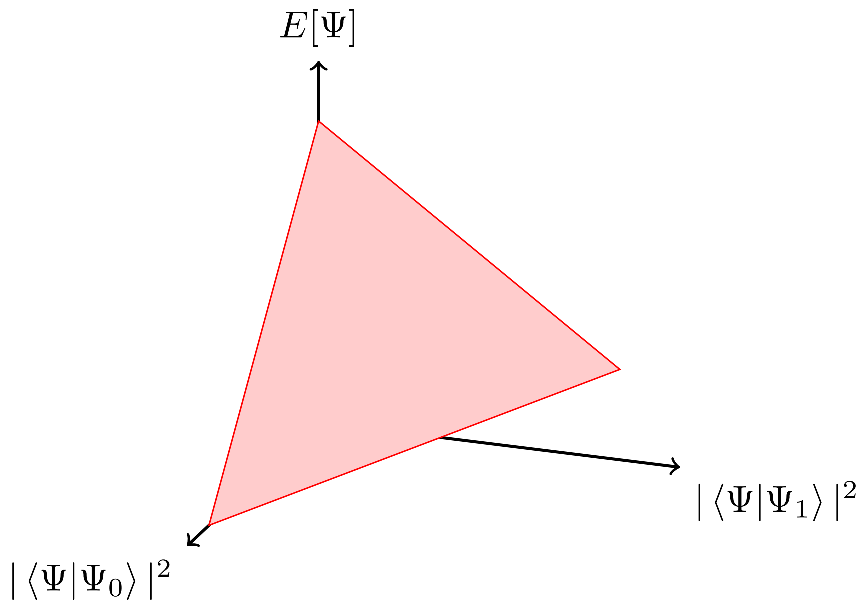


Figure 6.1: The lower bound of $E[\Psi]$ as a function of overlap with the first two eigenstates. The vertices are the first three eigenstates.

seen primarily for just three states, states 1, 6, and 11, with reductions in energy of 0.14(3), 0.18(3), 0.22(3) eV respectively. The latter two of these states are the ${}^1B_{1u}, {}^1E_{1u}$ states which have strong ionic bonding character among the π orbitals, unlike the ground state which has covalent bonding character.[121] This difference in bonding character is captured by the orbital optimization, leading to larger reductions in total energy in these states, while the other states have no reduction in total energy. The fact that only some excited states benefit from orbital optimization in DMC means that the energy differences are affected; ultimately they are improved.

Before comparing to experiment, we note that most theoretical results report the vertical excitation energy, which omits nuclear relaxation and vibrational effects. The transition from the ground vibrational level of the ground state to the ground vibrational level of the excited state is called the 0-0 excitation energy (E_{00}), and is not necessarily the maximum intensity. The vertical

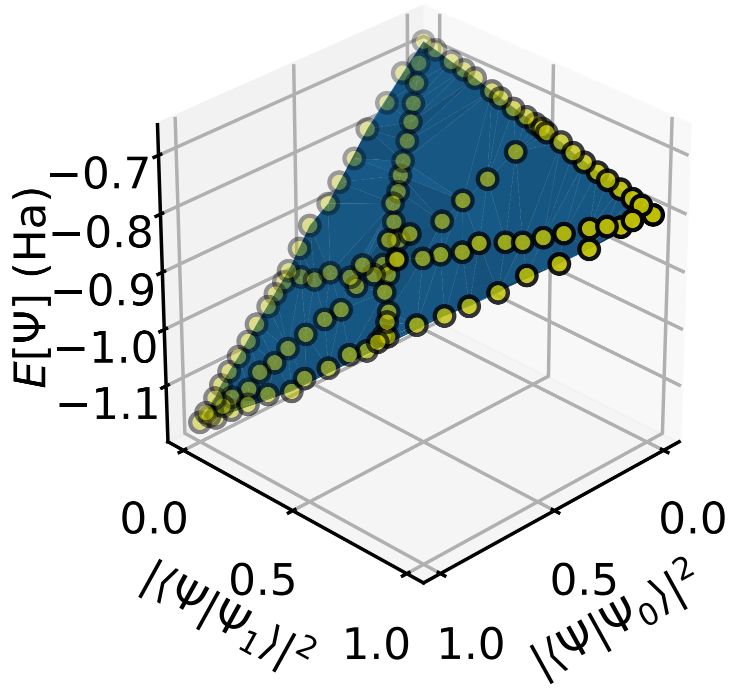


Figure 6.2: Targeted overlaps for H_2 in the space of the lowest three eigenstates. The yellow points are wave functions, using the penalty method to set S^*. The x and y coordinates are the measured overlaps after optimization, and the energy $E[\Psi]$ is the expectation value of the wave function after optimization.

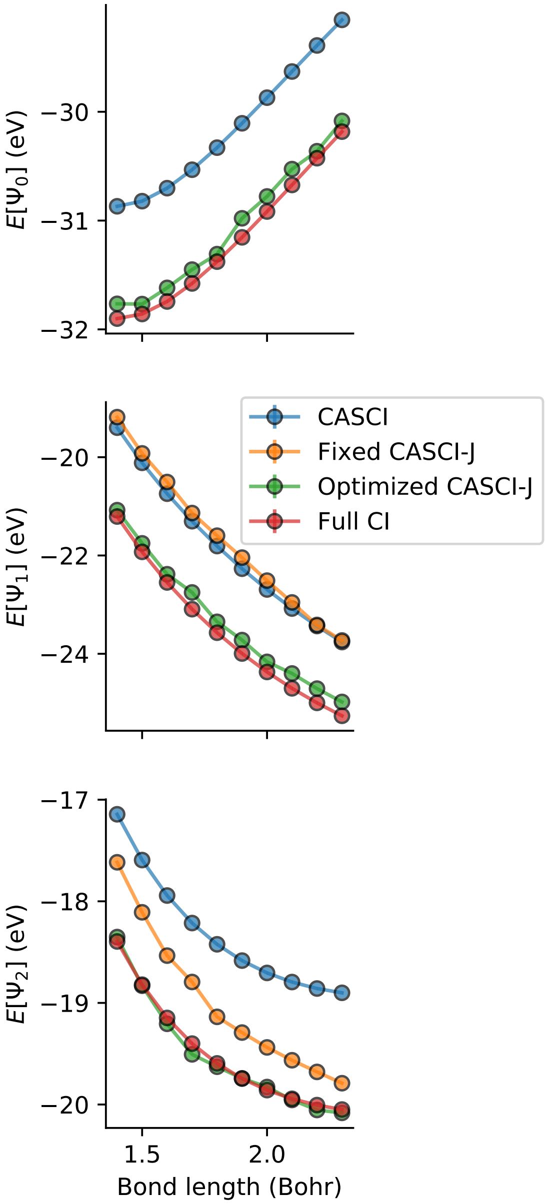


Figure 6.3: Comparison of FCI eigenvalues with energies of wave functions optimized with orthogonal optimization to target the first three eigenstates for H_2. The cc-pvtz basis of Dunning[5] was used.

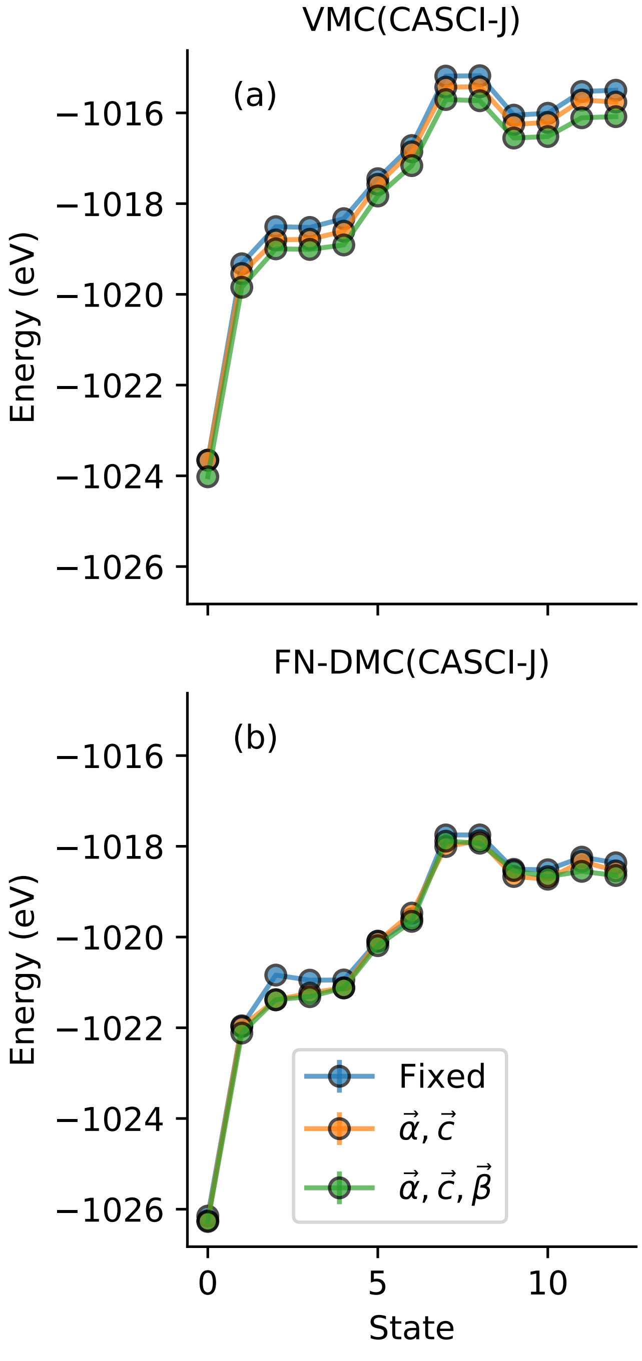


Figure 6.4: Benzene excited states computed using **(a)** VMC and **(b)** time-step extrapolated DMC for the full π-space spectrum. Different colors refer to increased parameter sets. $\vec{\alpha}$ are the Jastrow coefficients, $\vec{c}$ are determinant coefficients, and $\vec{\beta}$ are orbital coefficients, as denoted in Eqn 6.19

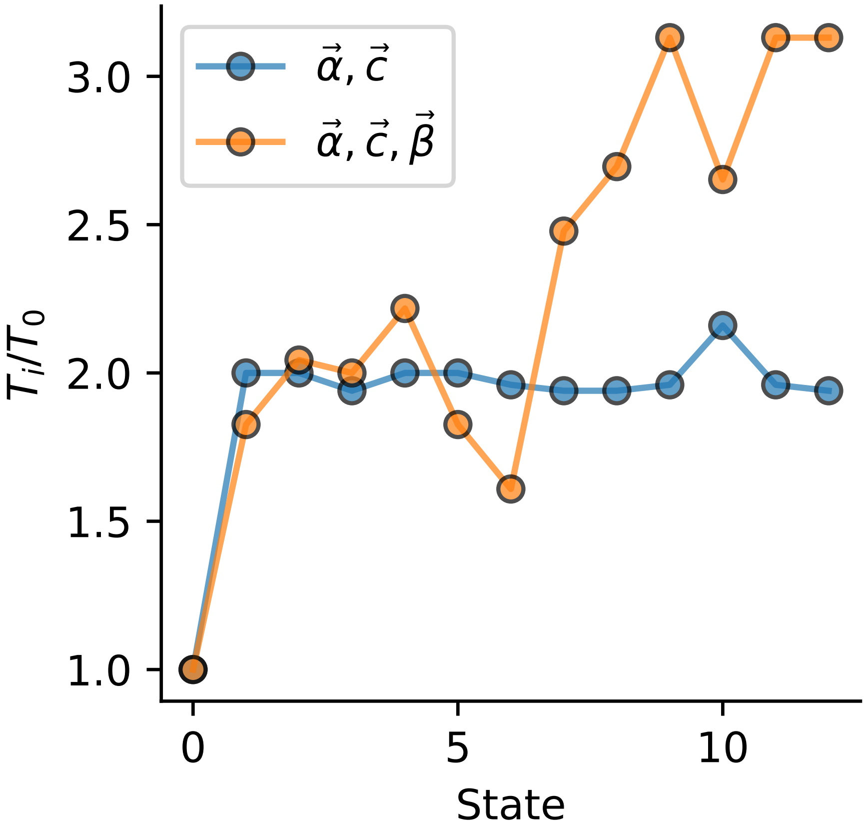


Figure 6.5: Relative cost T_i/T_0 of computing excited states in VMC for the two different parameterizations considered on the benzene molecule, where T_i is the CPU time required to perform the calculation of excited state i. T_0 for the ground state were 1.25 and 23.00 hours on a single 40 processor node for the $\vec{\alpha}, \vec{c}$ and $\vec{\alpha}, \vec{c}, \vec{\beta}$ parameterizations respectively.

excitation energy can be related to the 0-0 excitation energy as follows:

$$\begin{aligned} E_{00}(0 \to j) &= E_j(\vec{X}_0) - E_0(\vec{X}_0) \text{ (vertical)} \\ &+ E_j(\vec{X}_j) - E_j(\vec{X}_0) \text{ (adiabatic)} \\ &+ \frac{\hbar}{2}\sum_i(\omega_{j,i} - \omega_{0,i}). \text{ (ZPVE)}, \end{aligned} \tag{6.20}$$

where E_j is the Born-Oppenheimer energy of state j, X_j are the set of nuclear coordinates at the minimum energy of state j, and $\omega_{j,i}$ is the frequency of vibrational mode i for electronic state j. The first line is the fixed nuclei vertical excitation energy, the quantity which is directly computed in our work and the cited theoretical calculations, where the total energy of the jth excited state is computed at the ground state equilibrium nuclear positions $\vec{X}_0$. The second line is the adiabatic correction, which accounts for shifting of the nuclear positions in the jth excited state to the configuration $\vec{X}_j$. The last line is the zero-point vibrational energy (ZPVE) correction, and accounts for changes in the vibrational degrees of freedom between the ground and excited state. The experimental vertical excitation is obtained by subtracting the adiabatic and ZPVE energies from the 0-0 excitation energies.

In Table 6.2, we compare the vertical excitation energy computed from experiment, the fully optimized DMC results in this work, and literature results of large-basis coupled cluster (CC3)[123], complete active space perturbation theory(CASPT2)[121], and time dependent density functional theory using the PBE0 functional (TDDFT-PBE0)[129]. We also report computed ZPVE and 0-0 corrections using TDDFT with a 6-31g basis and PBE functional for the singlet states; we encountered difficulties converging the triplet states in TDDFT. We believe the TDDFT estimates to be reasonable given the agreement with computed and measured values in literature: -0.15 eV [127] and -0.17 eV [130] adiabatic and ZPVE corrections for the $^1B_{2u}$ state, -0.56 eV [131] and -0.17 eV [132] for the $^3B_{1u}$ state. For the available vertical excitations, the DMC and CC3 results both are within 0.2 eV of the experimental values, while the CASPT2 and TDDFT method have larger differences. The corrections due to adiabatic and ZPVE effects are large enough that this conclusion might be reversed without these effects included.

Table 6.3: Agreement between theoretical vertical excitation energies. Optimizing more parameters using the penalty method improves the agreement between CC3 and DMC, as seen in the mean and standard deviation of the difference between the excitation energies.

Method	Parameters	$\Delta E_j = E_j(\mathrm{m}) - E_j(\mathrm{CC3})$	
		mean (eV)	RMS (eV)
CASPT2		-0.38	0.42
TTDFT-PBE0		-0.33	0.50
VMC	None	0.34	0.46
VMC	$\vec{\alpha}, \vec{c}$	0.12	0.35
VMC	$\vec{\alpha}, \vec{c}, \vec{\beta}$	0.19	0.35
DMC	None	0.24	0.35
DMC	$\vec{\alpha}, \vec{c}$	0.18	0.31
DMC	$\vec{\alpha}, \vec{c}, \vec{\beta}$	0.15	0.24

In Table 6.3, we list the mean and standard deviation of the difference between CC3 and other methods. CC3 has been found to be very accurate for electronic excitations, including correlated doubles-like excitations for benzene.[121, 133] The consistency between accurate, explicitly correlated methods like CC3 and our new QMC technique, which make different approximations, is encouraging, as is the fact that the QMC results approach the CC3 results as the number of optimized parameters is increased.

The cost of each excited state calculation is roughly a factor of 2-3 higher than the ground state calculation, as shown in Fig 6.5. As a point of reference, the ground state optimization took 1.25 hours on a 40 core processor for the fixed orbital optimization and 23 hours on a modern 40 core processor for the orbital optimized calculation. One should keep in mind that as is typical in Monte Carlo, this cost is highly dependent on the desired uncertainty in the result. We chose very converged parameters, which results in a relatively high computational cost. The total relative cost over the ground state for all 13 states is a factor of 25 for the $\vec{\alpha}, \vec{c}$ calculation and 30 for $\vec{\alpha}, \vec{c}, \vec{\beta}$. Importantly, the relative cost does not increase rapidly with the state, since as mentioned in the introduction the overlaps are not very expensive to evaluate.

6.4 Conclusion

We have presented a scalable algorithm to compute approximate excited states of many-body systems using variational Monte Carlo. Our method is somewhat less complicated to implement than the linear method of Filippi[106, 107, 108], since the derivatives of the local energy are not required. Further, our technique is capable of optimizing complex parameters like orbital coefficients in a state-specific manner, and likely would be able to optimize parameters from wave functions such as backflow[134, 135] and neural network forms,[136, 137, 138, 139] as well as pairing functions[140], since the complexity is not much higher than normal wave function optimization. In comparison to the method of Neuscamman [110, 111, 112, 113], this technique does not require a tuned parameter ω. One positive aspect of this is that the penalty technique can access several distinct but degenerate excited states separately. Degenerate ground states would also be detected by the penalty technique.

The penalty technique is capable of optimizing wave functions with any overlap with a reference state. This capability may be useful in some circumstances, particularly for strongly correlated systems and magnetic systems, in which energy eigenstates may not be easily representable by simple wave functions, but non-orthogonal basis states can be used to make relevant experimental predictions.[141] Such wave functions are also appropriate for usage in density matrix downfolding.[39]

The study on the benzene molecule revealed a few interesting physical insights. First, orbital optimization improves the nodal surface for excited states, particularly those with significant ionic character as compared to the ground state. Secondly, the classical association of the vertical excitation with the maximum intensity in spectroscopy leads to errors up to 0.4 eV in the excitation energies of benzene. A fully quantum treatment of the Franck-Condon principle, as shown here, brings the experimental estimates in closer alignment with coupled cluster and quantum Monte Carlo results.

Chapter 7

Tight binding model with lattice interactions for twisted bilayer graphene

7.1 Introduction

An exciting new task in condensed matter physics is understanding the microscopic mechanisms leading to a diverse set of electronic states in twisted bilayer graphene (TBLG). TBLG is a van der Waals structure that consists of two sheets of single layer graphene laid on top of each other and twisted with a relative twist angle θ. Below $\theta \sim 4^\circ$, the twist results in a large scale Moiré superlattice structure in TBLG. Within the Moiré regime, TBLG hosts putatively correlated superconducting and insulating phases [142, 143, 144, 145, 146] and anomalous Hall effects [147, 148] for certain "magic" values of θ, the first of which is near 1.05°. It is believed that these correlated phases emerge near the magic twist angles through the interplay of electronic and structural degrees of freedom.

Electronically, the unusual properties of TBLG are suggested to arise from band flattening at the Fermi level. The current understanding is that with sufficiently flat bands, even weak effective interactions can dominate the low-energy behavior, leading to the observed correlated states in TBLG near magic twist angles. The bulk of evidence for band flattening in TBLG is found in theoretical calculations [149, 150, 151, 152, 153, 154] where bands near the Fermi level are shown to flatten to a few meV bandwidth near magic twist angles. Experimentally, a recent combination of low-energy electron microscopy and angle-resolved photoemission spectroscopy measurements [155] also observe flat bands at charge neutrality with bandwidths of 30 ± 15 meV at $\theta = 1.3^\circ$.

Accompanying the changes in electronic structure near the magic twist angles are structural relaxations. In TBLG, structural relaxations enhance the size of low-energy AB regions and constrict those of AA regions and are paired with out-of-plane buckling, bringing together AB regions and pushing apart AA regions. This structural relaxation has been observed experimentally through scanning tunneling microscopy measurements on TBLG[156] and in theoretical calculations [149, 157, 158, 159, 160, 161, 162]. Perhaps most importantly, using simple tight binding models — such as the one of Moon and Koshino (MK)[163] — it has been proposed[149] that lattice relaxation is required to isolate the flat bands from the rest of the bands in the system.

Thus it appears that the structural and electronic degrees of freedom are tightly coupled in this system. However, it is unclear if more accurate treatments of the electronic and structural degrees of freedom would result in the same conclusions. While some density functional theory (DFT) calculations have been performed at the magic angle scale,[150, 158] these calculations are very computationally demanding and it is not feasible to perform many calculations to disentangle electronic and structural degrees of freedom.

In this chatper, I present a highly accurate local environment tight-binding (LETB) model for TBLG, fit to DFT calculations of 72 structural configurations of bilayer graphene. The LETB model is shown to reproduce DFT for structural configurations relevant to TBLG much more accurately than simple TB models such as the MK model. In the LETB model, both isolated flat bands and fragile topology[164, 165, 166, 167, 168, 169, 170] are observed both with and without lattice relaxation, in contrast to the simpler MK model, in which lattice relaxation is required for isolated flat bands with fragile topology. A Python package that generates LETB models for any configuration is made available.

7.2 Training data for twisted bilayer graphene model

7.2.1 Atomic configurations

Our goal is to develop a TB model for TBLG that correctly accounts for the variations in structure encountered at small twist angles. Our approach is to use deformed, untwisted, primitive cell bilayer graphene configurations. We use two deformation strategies in order to capture the variations in stacking pattern as well as in-plane and out-of-plane relaxations seen in small twist angle TBLG.

The first are in-plane and out-of-plane shifts, strains and shears,and the second are random atomic variations. Details of the deformation strategies follow.

We begin by introducing notation for the deformed atomic configurations. There are four atoms in our configurations with positions denoted by the vectors $\vec{a}_1, \vec{b}_1, \vec{a}_2, \vec{b}_2$, where a, b refer to the two distinct atoms in a given graphene layer, and $1, 2$ are layer indices. The real space lattice vectors for the primitive cell are denoted by $\vec{L}_1, \vec{L}_2$. These six vectors fully describe the atomic configurations.

The shifted, strained, and sheared configurations are best understood through the deformation equation

$$\begin{bmatrix} \vec{a}_1 \\ \vec{b}_1 \\ \vec{a}_2 \\ \vec{b}_2 \\ \vec{L}_1 \\ \vec{L}_1 \end{bmatrix} = \left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & a_0 & 0 \\ 0 & 0 & c_0 \\ 0 & a_0 & c_0 \\ \sqrt{3}a_0 & 0 & 0 \\ \frac{-\sqrt{3}}{2}a_0 & \frac{3}{2}a_0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & s & \Delta \\ 0 & s & \Delta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \cdot \begin{bmatrix} 1+\epsilon_{xx} & \epsilon_{xy} \\ \epsilon_{xy} & 1+\epsilon_{yy} \\ 0 & 0 \end{bmatrix} . \tag{7.1}$$

The first matrix on the right hand side denotes our reference configuration: the equilibrium AA configuration with $a_0 = 2.683$ Bohr, $c_0 = 6.646$ Bohr. The second matrix applies between layer shifts through the parameter s and inter-layer spacing variations through Δ. In our dataset, s takes three values - 0, a, $3a/2$ - which correspond to the AA, AB and SP bilayer configurations, and Δ takes three values - 0 Bohr, -0.149 Bohr, -0.126 Bohr - corresponding to the AA, AB, and SP layer spacings. The third matrix applies in-plane shears via ϵ_{xy} and in-plane strains in the x- and y-directions via $\epsilon_{xx}, \epsilon_{yy}$. All three parameters independently take three values, -0.01, 0, and 0.01, corresponding to 1% atomic position variations in-plane. The five parameters, $s, \Delta, \epsilon_{xx}, \epsilon_{xy}, \epsilon_{yy}$ can be varied together, leading to a large space of strained, sheared, and shifted configurations of bilayer graphene.

For the random configurations, the parameters $s, \Delta, \epsilon_{xx}, \epsilon_{xy}, \epsilon_{yy}$ are chosen at random from a uniform distribution between the corresponding ranges described in the previous paragraph. Once this new deformed configuration is constructed, we add an additional 6×3 random matrix R to the configuration to allow for arbitrary inter-atomic displacements not available in Eq 7.1. To ensure that the random displacements incurred by R are not too large, we require that the Frobenius norm $||R||_F = 0.01a_0$, corresponding to random variations that move the atomic configurations by a percent of the equilibrium atomic spacing.

We also employ twisted configurations in our dataset, but these are used only for model validation. Two twisted configurations are used, one at 9.4 degree twist, and the second at 4.4 degree twist. More details of the density functional calculations and validation procedure are presented in section III C.

7.2.2 Density functional theory (DFT) band structures

For each of the deformed geometries in our training data set, we computed the total SCF energies and band structure using Van der Waals DFT. We used the BEEF-VDW Van der Waals functional [171], a polarized triple-zeta all electron basis constructed for solid-state DFT calculations [172], and a 36 x 36 x 1 k-point grid. All DFT calculations were carried out using the PySCF package [173, 174].

To demonstrate the accuracy of the DFT calculations, we present a comparison of the DFT band structure to angle resolved photoemission spectroscopy (ARPES) measurements [6] in Fig 7.1 for the eqilibrium AB configuration. Near the Fermi level we find excellent agreement between DFT and ARPES, with DFT being within 0.1 eV of the experimentally measured excitations. The errors between DFT and ARPES increase away from K, but do not deviate more than 10% from the experimental measurements.

7.2.3 Wannierization of DFT band structure

From each DFT calculation we isolate the four π bands and Wannierize them to extract tight binding (TB) parameters. The disentanglement scheme of Souza, Marzari, and Vanderbilt [175] is used to isolate the four π bands from bands with C $1s$, $2s$, $2p_x$, $2p_y$ orbital character. The Maximally Localized Wannierization procedure [176, 177, 178] is then used to generate Maximally Localized Wannier Orbitals (MLWO), and the isolated π bands are projected onto the MLWOs to

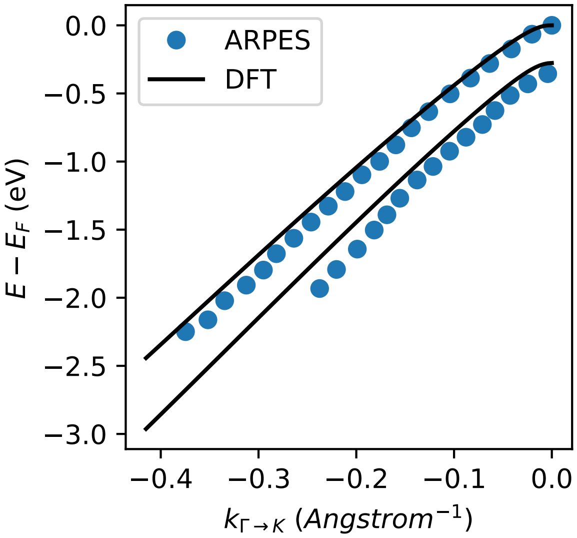


Figure 7.1: Comparison of DFT and ARPES band structures for AB bilayer graphene near the Fermi level. The band path goes in a line from the Γ to K point in the first Brillouin zone. ARPES measurements are digitized from Figure 3d of Joucken *et al.* [6]

get TB parameters. All Wannierization calculations are carried out using the Wannier90 [179] and pyWannier90 packages [174].

A comparison of the Wannierized band structure to DFT is shown in Fig 7.2. The figure shows the energy relative to the Fermi level for the DFT and Wannierized bands for the AB configuration. The Wannierized bands fall exactly on top the DFT bands, demonstrating the accuracy of both the disentanglement and MLWO schemes.

7.3 Local environment TB model (LETB)

We propose a local environment dependent TB parameterization which explicitly accounts for the nuclear configuration in the vicinity of atoms involved in hopping. The general form of the LETB is

$$H_{LETB} = \sum_{ij\sigma} t_{ij}^{LETB}(\vec{R}_i, \vec{R}_j, \{\vec{R}_{ij}\}) c_{i\sigma}^{\dagger} c_{j\sigma} + h.c. \tag{7.2}$$

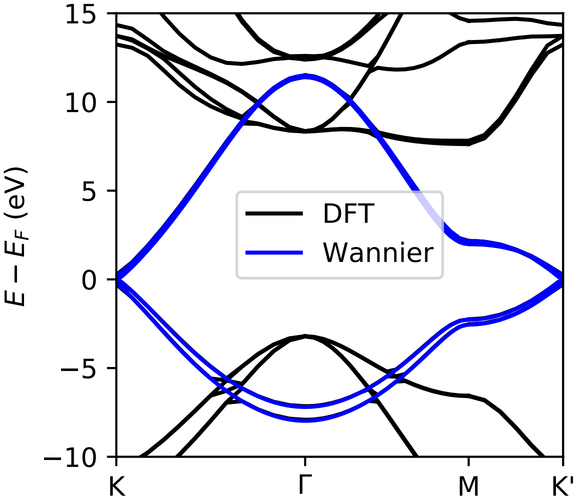


Figure 7.2: Comparison of DFT and Wannierized band structures using the MLWO scheme near the Fermi level for AB bilayer graphene. Only the π bands near the Fermi level are Wannierized.

Here i, j are atomic indices for the pair of atoms with a hopping value of t_{ij}^{LETB}, σ is the spin index, $c_{i\sigma}^{\dagger}$ is the creation operator for a localized orbital on site i of p_z character with spin σ, $\vec{R}_i$, is the location of atom i, and $\{\vec{R}_{ij}\}$ is a set of nuclear positions within the local environment of atoms i, j. The definition of the local environment of two atoms i, j and functional dependence of t_{ij} on the local environment coordinates are made explicit in sections III A and III B.

The LETB can be contrasted with the MK model [163]:

$$\begin{aligned} H_{MK} &= \sum_{ij\sigma} t_{ij}^{MK}(\vec{R}_i - \vec{R}_j) c_{i\sigma}^{\dagger} c_{j\sigma} + h.c. \\ t_{ij}^{MK}(\vec{d}) &= V_{pp\pi}^{0} e^{(-\frac{|\vec{d}|-a_0}{\delta})} \Big[1 - (\frac{\vec{d}\cdot\hat{z}}{|\vec{d}|})^2\Big] \\ & \qquad V_{pp\sigma}^{0} e^{(-\frac{|\vec{d}|-d_0}{\delta})} (\frac{\vec{d}\cdot\hat{z}}{|\vec{d}|})^2 \end{aligned} \tag{7.3}$$

where the constants take values $V_{pp\pi}^{0} = -2.7$eV, $V_{pp\sigma}^{0} = 0.48$ eV, $a_0 = 2.683$ Bohr, $d_0 = 6.331$ Bohr, $\delta = 0.246 a_0$. Conceptually, this model bridges two exponentials: the $V_{pp\pi}$ exponential which corresponds to intra-layer hoppings and the $V_{pp\sigma}$ term corresponding to inter-layer hoppings. Unlike

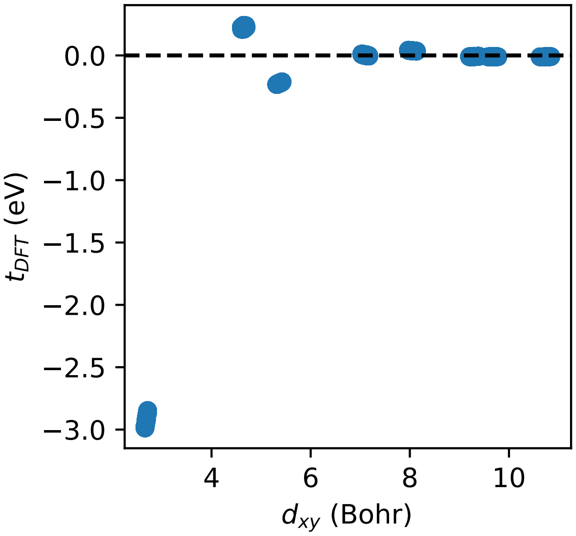


Figure 7.3: Plot of intralayer hoppings in our sampled configurations as a function of the in-plane distance d_{xy}. A clear separation of nearest neighbors is present, allowing for separate models to be fit for each term.

the LETB, the hoppings depend only on the pairwise displacement vector between hopping centers. Nonetheless, the MK paramaterization is a staple in the field of TBLG band structure calculations, and we use it as a point of comparison for model accuracy.

7.3.1 Intra-layer hopping

We begin by fitting the intra-layer hoppings, hoppings for atoms i, j within a single graphene sheet. In this work we only consider intra-layer hoppings up to third nearest neighbor, and fit separate models for the first, second, and third nearest neighbor hoppings. These choices were made after looking at the distance dependence of the intra-layer hoppings computed in DFT over the configurations in our training data set, as shown in Fig 7.3. Hoppings beyond the third nearest neighbor takes values on the order 0.01 eV, 10 times smaller than the third nearest neighbor, and are thus ignored. Further, a clear separation in distance is seen between the first, second, and third nearest neighbors, justifying our separate treatment of these three terms.

Each model is taken to be linear in a set of geometric descriptors which account for the variations

Intra-layer Hopping Models

1-NN	-9.68(4) + 2.52(1)a
2-NN	1.55(1) + 0.022(2)b - 0.66(1)h_1 - 0.20(1)h_2
3-NN	-1.23(1) + 0.04(1)c - 0.12(1)h + 0.23(1)l

Fang [157] Inter-layer Hopping Model

i	λ_i (meV)	ξ_i	x_i	κ_i
0	239(2)	2.12(2)		1.871(4)
3	-40(1)	3.8(4)	0.52(4)	
6	-5.9(7)	6.0(8)	1.52(1)	1.73(2)

Table 7.1: Table of cross validated fit parameters for the different terms in LETB. Cross validated errors are presented in parentheses.

in local nuclear environment. The linearity of the models follows from the fact that variations in hopping and distance seen in Fig 7.3 are within 1% of the equilibrium value for each hopping. The full set of descriptors in consideration for each model is presented in Fig 7.4. For the nearest neighbor hopping, the only descriptor is the pairwise distance between the two atoms. For the second and third nearest neighbor, we consider geometric descriptors which extend to a farther local environment and describe the triangular and hexagonal regions around the hopping centers, respectively. It should be noted that all descriptors here are distances computed after projection in the $x-y$ plane.

Descriptor selection for the second and third nearest neighbor models is carried out using the LASSO [180, 181] procedure. The LASSO method uses an altered cost function which adds an L_1 penalty to the standard least squares regression function. The L_1 penalty assigns zero coefficient to any descriptors which do not improve model quality. Therefore, the LASSO method is used as a systematic technique for descriptor selection when fitting linear models for the LETB.

After descriptor selection is carried out, the linear models are fit using cross validated (CV) ordinary least squares regression. The CV technique partitions the training data into n = 5, in this case, portions and sequentially fits on four portions and is tested on the remaining portion. This technique allows us to simultaneously regress the linear model, determine whether our models are overfit, and also compute statistical error bars for the fit parameters. A summary of the final regressed models with CV parameters and error bars is presented in Table 7.1.

The simplest is the nearest neighbor model fit which is linear in the distance a between the

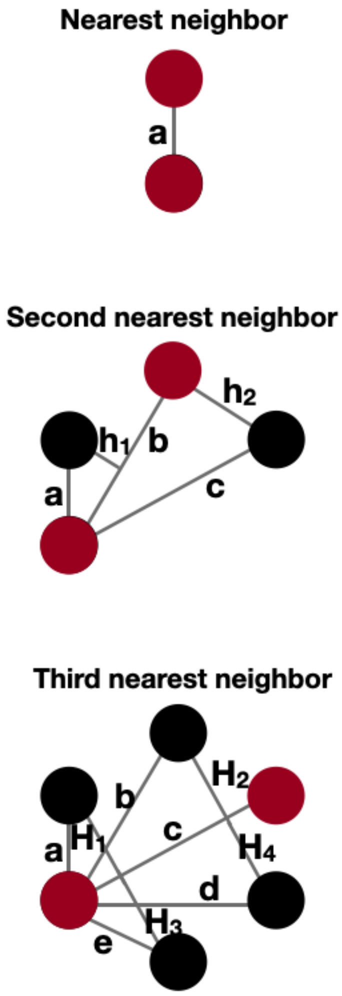


Figure 7.4: Schematic diagram listing all possible descriptors used to fit linear models for intralayer hopping. The red atoms denote the atomic pairs between - from left to right - nearest, next nearest, and third nearest neighbor atoms.

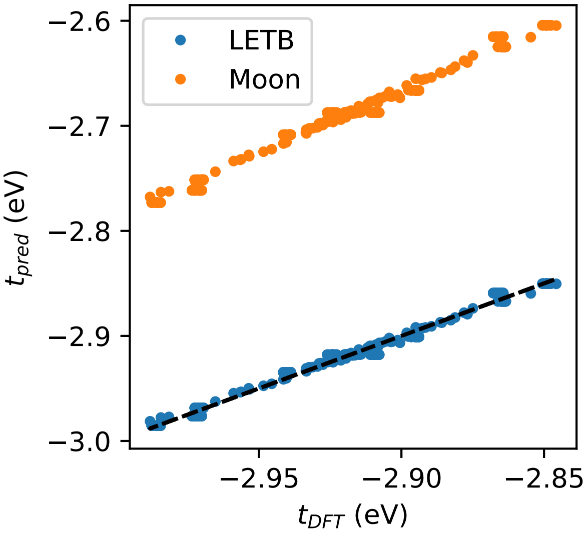


Figure 7.5: Plot of predicted versus computed DFT intralayer nearest neighbor hoppings. LETB refers to this work, and MK is using the MK parameterization. The black line indicates perfect agreement with the computed DFT values.

atoms. A comparison of the predicted LETB and MK hoppings to the computed DFT hoppings are shown in Fig 7.5. We find excellent agreement to the DFT data with our fit yielding an R^2 of 0.98. The MK parameterization, while following a similar trend to the LETB, consistently underestimates the magnitude of the hoppings by 10%.

The second nearest neighbor model is more involved, and is best described by $h_1,\ h_2$, with b yielding minor corrections. Again, a comparison of the model and DFT parameters can be seen in Fig 7.6. We find excellent agreement of the LETB with the DFT data yielding an R^2 of 0.92. The MK parameterization, however, completely fails to describe the variations seen in the DFT data. This is because the MK functional form does not account for the $h_1,\ h_2$ descriptors, which are the most important for the second nearest hopping. Instead, it relies on the descriptor b only, which is a very poor descriptor for the second nearest neighbor hopping.

The third nearest neighbor model in Fig 7.7 is the most complex, involving the size of the entire hexagonal environment around the hopping centers. The selected regressors for the LETB are $h = (H_1 + H_2 + H_3 + H_4)/4,\ l = (a + b + d + e)/4$ and c, with an excellent fit quality of $R^2 = 0.96$.

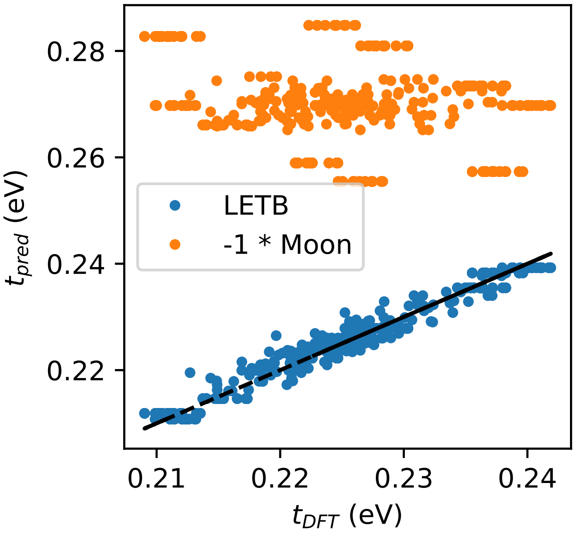


Figure 7.6: Plot of predicted versus computed DFT intralayer next nearest neighbor hoppings. LETB refers to this work, and MK is using the MK parameterization. Note that we present the MK hopping with a sign flip, as the model does not predict negative values. The black line indicates perfect agreement with the computed DFT values.

Similar to the first nearest neighbor, the MK parameterization consistently underestimates the hopping magnitude for the third nearest neighbor hopping. However, in this case the quantitative error is much larger, with a nearly 50% error between the MK and DFT values.

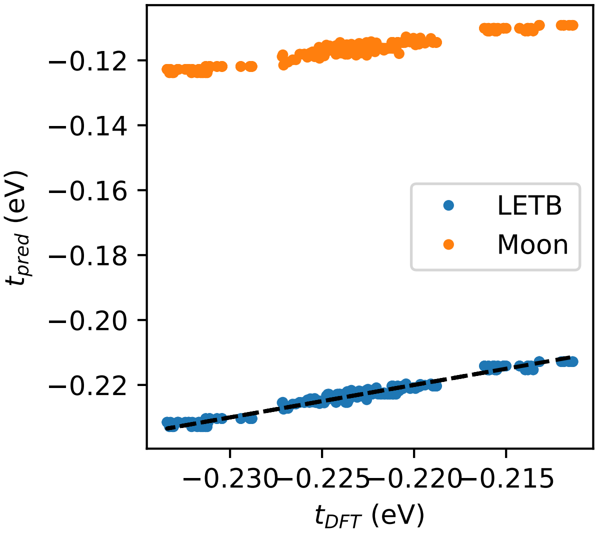


Figure 7.7: Plot of predicted versus computed DFT intralayer third nearest neighbor hoppings. LETB refers to this work, and MK is using the MK parameterization. The black line indicates perfect agreement with the computed DFT values.

7.3.2 Inter-layer hoppings

For the inter-layer hoppings, hopping for atoms i, j within separate layers, we use the parameterization proposed by Fang *et al.* [182]. The Fang model takes the following form

$$t_{ij}^{Fang}(\{\vec{R}_{ij}\}) = t^{Fang}(\vec{R}_i - \vec{R}_j, \theta_{u,ij}, \theta_{d,ij})$$

$$\begin{aligned} t^{Fang}(\vec{r}, \theta_u, \theta_d) = V_0(r_{xy}/a) + \\ V_3(r_{xy}/a)(\cos(3\theta_u) + \cos(3\theta_d)) + \\ V_6(r_{xy}/a)(\cos(6\theta_u) + \cos(6\theta_d)) + \cdot \end{aligned} \tag{7.4}$$

$$\begin{aligned} V_0(r) = \lambda_0 e^{-\xi_0 r^2} \cos(\kappa_0 r) \\ V_3(r) = \lambda_3 r^2 e^{-\xi_3 (r - x_3)^2} \\ V_6(r) = \lambda_6 e^{-\xi_6 (r - x_6)^2} \sin(\kappa_6 r) \end{aligned}$$

Here the angles $\theta_{u,ij}$, $\theta_{d,ij}$ account for the local environment effects between two atoms i, j. The first angle θ_u indicates the orientation of the nearest neighbor triangle of the *upper* sheet atom relative to the displacement vector $\vec{R}_i - \vec{R}_j$, and θ_d the orientation of the same triangle of the *lower* sheet atom. The constant $a = a_0/\sqrt{3} = 1.549$ Bohr, and all other constants in the expression are regression parameters.

While Fang *et al.* did fit the regression parameters in Eq 7.4, we do not use those parameters and instead fit the parameterization to our training data. We use a non-linear least squares algorithm to fit the parameters. As with the intra-layer parameters, we use a 5-fold CV to simultaneously fit parameters and assess the error bars for the parameters. The summary of the fit parameters and their CV error bars are presented in Table 7.1.

The results of the inter-layer fitting are shown in Figs 7.8, 7.9. Looking at the MK parameterization first, we find that the model performs well in two extreme regions - $d < 7$ Bohr and $d > 10$ Bohr - but fails to describe most of the variation in the region in between, near 8 Bohr. The Fang parameterization ameliorates this issue and is able to describe the variation in the DFT data consistently across all bond lengths. Our results indicate that local environmental effects are required to describe inter-layer hoppings in intermediate bond lengths near 8 Bohr.

7.3.3 Model validation

As a final check of model validity, in Figure 7.10 we compare the DFT band structures for TBLG at 9.4 and 4.4 degree twists against the band structures computed using LETB. We find excellent agreement between the DFT and LETB at all k points and at both twists for the four π bands near the Fermi surface. As such, the use of primitive cell bilayer configurations for training paired with a local TB approximation generalizes well to twisted bilayer.

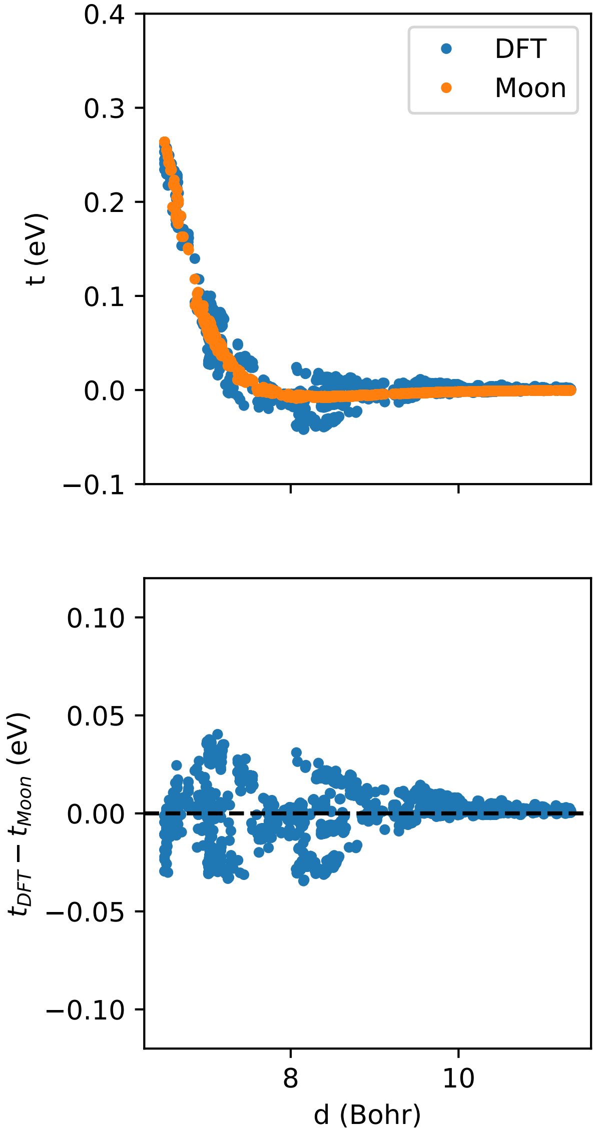


Figure 7.8: Plot of the predicted interlayer hoppings using the MK model, compared to DFT, as a function of atomic pairwise distance. The upper figure shows the hoppings, and the bottom the residual.

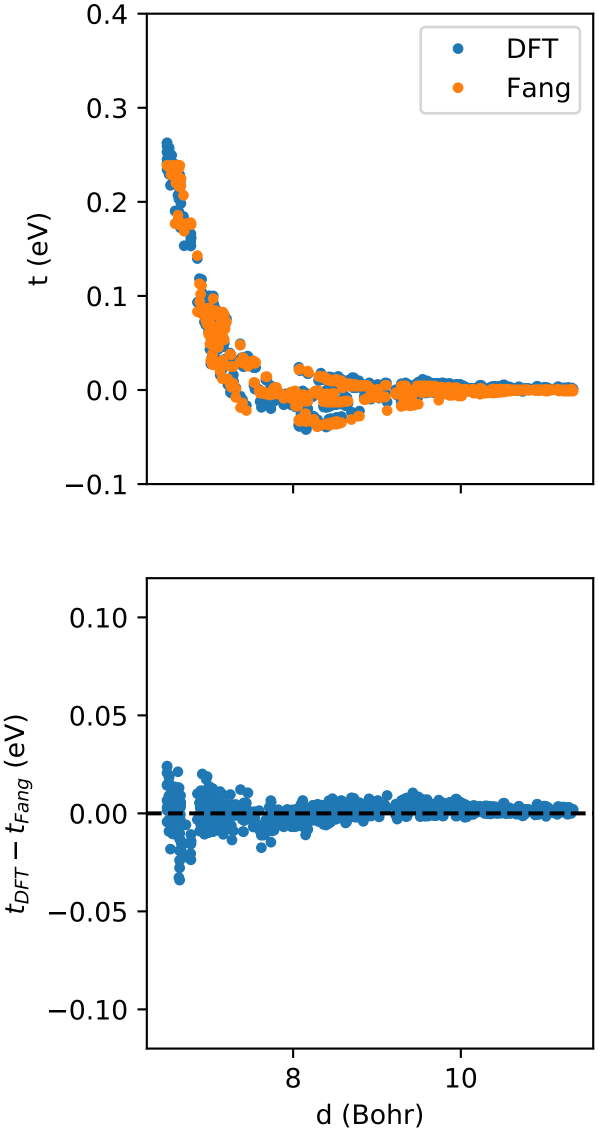


Figure 7.9: Plot of the predicted interlayer hoppings using the Fang model, compared to DFT, as a function of atomic pairwise distance. The upper figure shows the hoppings, and the bottom the residual.

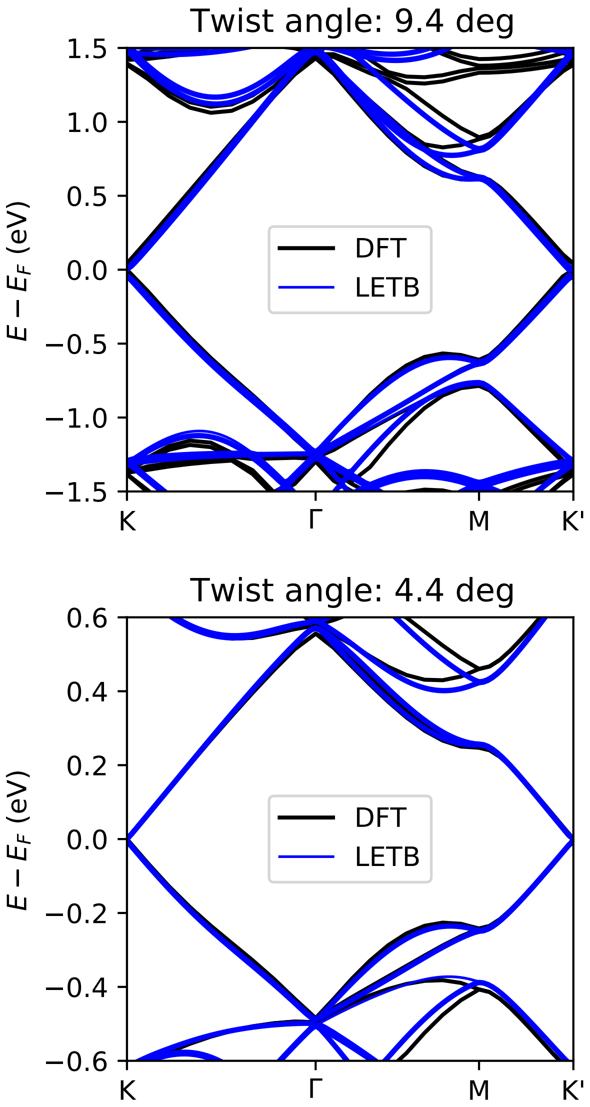


Figure 7.10: Comparison of LETB and MK band structure versus computed DFT band structures for 9.4 and 4.4 degree twist angles. Energy is relative to the Fermi level, and the standard $K \rightarrow \Gamma \rightarrow M \rightarrow K'$ path through the Moire supercell is shown.

7.4 Effects of accurate TB model on the electronic structure of TBLG

To study the effects of the LETB on our understanding of TBLG, we computed band structures for various small twist angles using LETB and the MK model with rigid and fully relaxed twisted geometries. We consider twist angles of 2, 1.47, 1.16, 1.08, 1.05 and 0.99 degrees. Rigid geometries consist of two sheets of unrelaxed graphene sheets twisted relative to each other with a commensurate twist angle. The fully relaxed geometries begin with the rigid geometries, and the atomic positions are determined using fully atomistic nonequillibrium molecular dynamics (MD) simulations. Comparison of computed band structures follow.

7.4.1 Isolated flat bands

In Fig 7.11 we present the bandwidths at the Γ point of the flat bands for the LETB and MK model with rigid and relaxed geometries. Both models for both geometries yield flat bands with bandwidth below 600 meV for twist angles below 1.2 degrees. With both geometries, the MK parameterization achieves its first inflection point at 1.08 degree twist, whereas the LETB continues the downward trend through 0.99 degree twist. It is unclear where the first inflection point for the LETB should be, but it necessarily occurs below 0.99 degrees for both geometries. Defining the first magic twist angle as the first inflection point of the bandwidth with respect to twist, the LETB yields a first magic twist angle at least 10% smaller than the MK parameterization.

In Fig 7.12 we present the band gaps at the Γ point between the flat and dispersive bands for the LETB and MK model with rigid and relaxed geometries. Unlike the bandwidths, the band gaps exhibit qualitative differences between the different models and geometries. Beginning with the rigid geometry, we note that the LETB has a finite band gap for all twist angles. The MK parameterization, however, yields zero band gap for twists of 0.99, 1.05 and 1.08 degrees, coinciding with the first inflection point observed in Fig 7.11. With the relaxed MD geometries, both LETB and MK follow qualitatively similar trends, with band gaps nearly 10 times the rigid amount across all twist angles.

The computed band widths and gaps indicate that geometry relaxation is not required for energetically isolated flat bands in TBLG for small twist angles. Restricted to the MK parameterization, one may conclude that geometry relaxation is necessary to maintain isolated flat bands, consistent with the analysis of Nam and Koshino. Our results when using the more accurate LETB model - isolated flat bands below 1.08 degrees twist without geometric relaxation - instead demonstrate that geometric relaxation is not required to qualitatively describe flat band physics. Rather, geometric relaxation serves a quantitative role, enhancing the LETB band gap by nearly a factor of ten and reducing the LETB bandwidth by a factor of two.

7.4.2 Fragile topology

In Figure 7.13 we present orbital densities at twists of 2 and 1.05 degrees to demonstrate electron density localization in flat bands. Shown are the orbital densities of the lowest energy of the four flat bands at different high-symmetry k-points Γ, M, K with fully relaxed MD geometries using the LETB. We see a common localization pattern across the two twists - near the Γ point the density localizes in the AB coordinated region and away from the Γ point the density localizes in the AA coordinate region - with the magnitude of localization increasing as the twist decreases. This pattern of density localization has been observed in extended tight binding models with relaxed geometries [183].

The shared localization pattern of AB near Γ and AA elsewhere only deviates when using a rigid geometry with the MK parameterization as demonstrated in Fig 7.14. In this figure we show a twist angle of 1.05 degrees. The LETB results on the right column follow the localization pattern of AB near Γ and AA elsewhere, while the MK parameterization does not, yielding AA localization near Γ as well. The AA localization near Γ is also shared for twists of 0.99 and 1.08 degrees when using the MK parameterization and rigid geometries. It should be noted that the appearance of AA localization near Γ in the MK model with a rigid geometry coincides with the closing of the band gap observed in Fig 7.12.

The computed orbitals indicate that geometry relaxation is not required for fragile topology of the isolated flat bands. Here we use fragile topology characterized by the inability to create

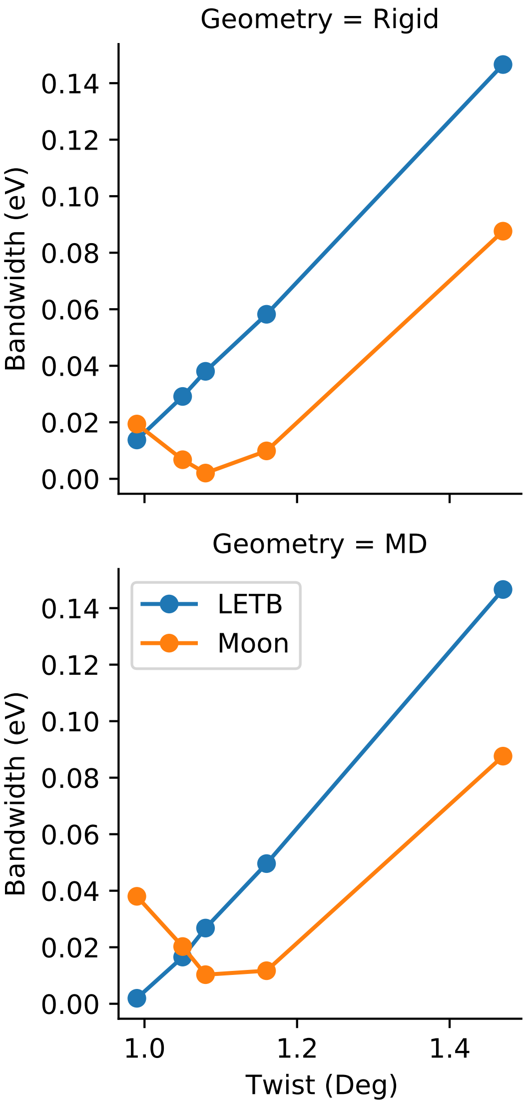


Figure 7.11: Computed flat band bandwidths for small twist angles near the magic twist angle using LETB and MK model parameterizations. The bandwidth is computed as the difference between the flat bands below and above the Fermi level at the Γ point. We also consider two different sets of lattice geometries: rigid twisted geometry and a fully relaxed, molecular dynamics geometry.

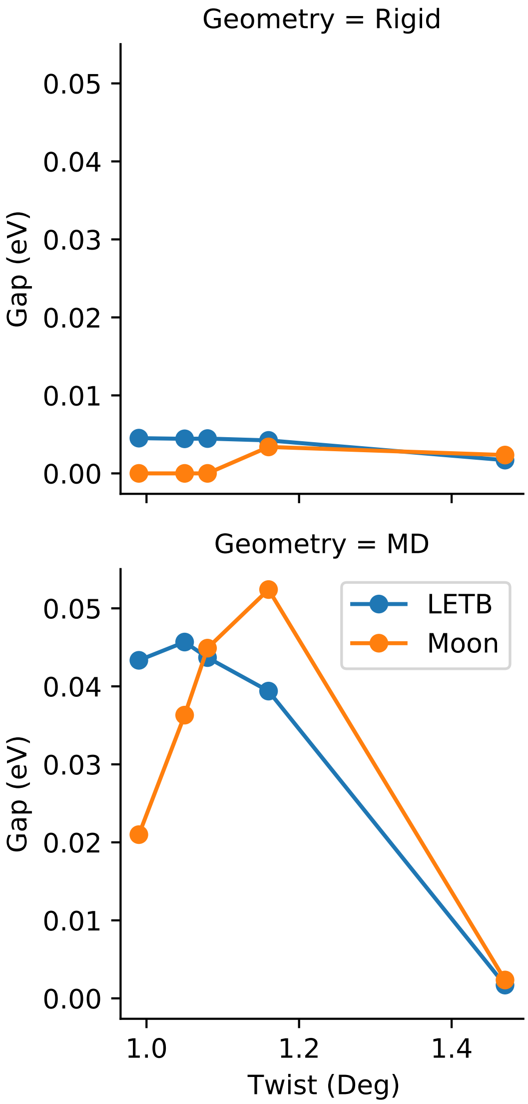


Figure 7.12: Computed band gap between flat and dispersive bands for small twist angles near the magic twist angle using LETB and MK model parameterizations. The bandwidth is computed as the difference between the flat bands above the Fermi level and the dispersive bands above the Fermi level at the Γ point. We also consider two different sets of lattice geometries: rigid twisted geometry and a fully relaxed, molecular dynamics geometry.

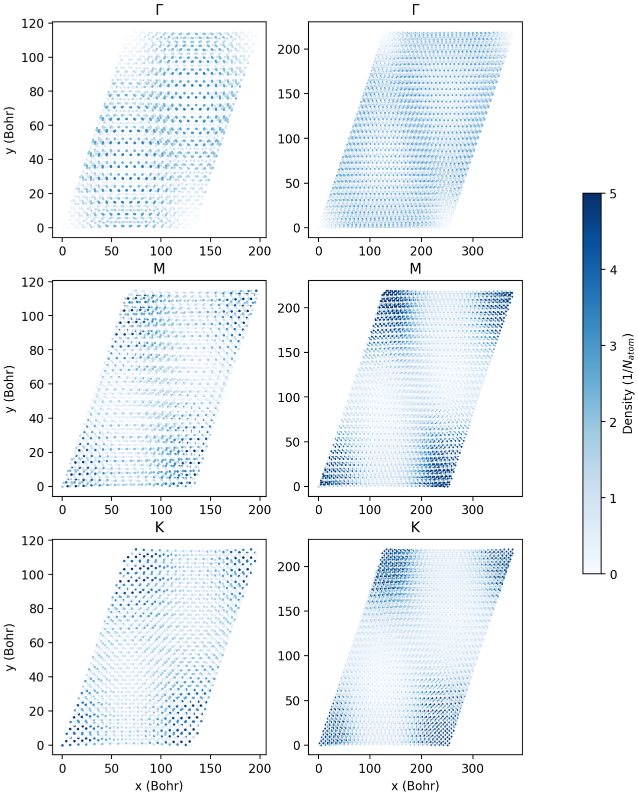


Figure 7.13: Orbital density for the fully relaxed molecular dynamics geometries and LETB hopping model of the lowest energy flat band at Γ, M, K points. Densities are shown for the 2 (left) and 1.05 (right) degree twists to demonstrate the orbital localization as the twist angle gets smaller.

symmetric, exponentially localized Wannier orbitals with only the flat bands in TBLG [166, 167]. Restricted to the MK parameterization, one may conclude that geometric relaxation is necessary to ensure a fragile topology of the flat bands, as rigid geometric twists can yield AA character orbitals across the first Brillouin zone and hence localized Wannier orbitals. Our results with the more accurate LETB paint a different picture, namely that AB character at Γ is ensured even with rigid geometries, preventing the localization of Wannier orbitals. As such, we have demonstrated that accurate TB model like LETB can maintain a fragile topology in TBLG without geometric relaxation.

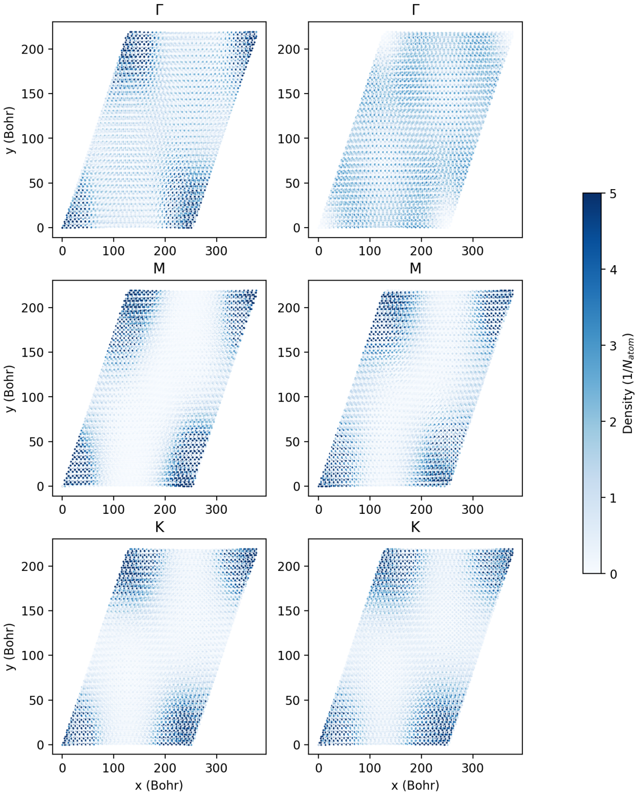


Figure 7.14: Orbital density for the rigid geometries of the lowest energy flat band at Γ, M, K points compared between the MK (left) and LETB (right) parameterizations. We can see clearly that the densities differ in orbital character near this twist angle between the two electronic models.

7.5 Conclusion

We developed a local environment dependent tight binding (LETB) model using DFT and a simple Wannierization procedure, and demonstrated its superior accuracy to the state of the art MK parameterization. Our model achieves high accuracy as it explicitly includes the effect of the local environment around the hopping centers on the hopping value between them. This exceeds the ability of the MK parameterization which only takes into account the displacement vector between the hopping centers. As such, our model yields a significantly better fit to TB parameters computed in DFT and agrees with DFT computed TBLG band structures better than the MK parameterization.

Our model reveals important understanding of the twisted bilayer system, namely that isolated flat bands and fragile topology do not require geometric relaxation. We demonstrated that while the MK parameterization requires geometric relaxation to maintain energetic isolation of the flat and dispersive bands, the LETB yields flat isolated bands with both rigid and relaxed geometries. For the LETB, the geometry relaxation only quantitatively affects the band structure, increasing the magnitude of the band gap and decreasing band width. Further, we showed that the MK parameterization does not maintain the fragile topology of the flat bands with rigid geometries, whereas the LETB does. Combined, our results indicate that an accurate TB model can recover many effects that a less accurate model would attribute to geometric relaxation.

Chapter 8

Conclusion

Ab initio techniques are a powerful class of methods which can directly assess the relationships between microscopic particles and macroscopic phenomenon. In this work, I have focused on two different classes of *ab initio* techniques: first principles methods and effective models. First principles methods approximately solve for the eigenstates of 1.1, trading off accuracy for computational efficiency, and include methods like DFT, QMC, and SHCI. The effective model approach works by constructing a simpler effective Hamiltonian and Hilbert space that replace $\hat{H}_{ab}, \mathcal{H}$ and can be more easily solved. My thesis work has been focused on improving the state-of-the-art in first principles methods and effective models.

In my first project, presented in Chapter 4, I investigated a new trial wave function for use in QMC through my work on non-orthogonal determinant multi-Slater Jastrow wave functions [40]. This new wave function parameterization was motivated by the need for refined orbital relaxation in QMC ground state calculations for strongly correlated systems like transition metal materials. We found on calculations for the C_2 molecule that the non-orthogonal parameterization yielded significant decreases in energy compared to the state-of-the-art orthogonal parameterization. Recent QMC calculations on antiferromagnetic FeO set out to extend our simple demonstration by computing the benefit of orbital optimization for strongly correlated systems, with orbital optimization yielding the most accurate wave functions for VMC when compared to standard Slater-Jastrow or backflow parameterizations [184]. My work therefore stands as a contribution to the catalogue of QMC research demonstrating the importance of orbital optimization in accurate ground state calculations of strongly correlated systems.

In Chapter 5, I showed my development and testing of a regularized estimator for wave function parameter gradients [43]. This work was necessitated by an infinite variance problem in VMC orbital optimization, where the gradient of the total energy with respect to orbital parameters has a divergent statistical variance. My solution was a simple regularization procedure which provides an efficient method for stable optimization of ground state wave functions in QMC which can be quickly integrated into any QMC code. This project has had an immediate and concrete impact, with a recent project augmenting my regularization technique motivated by ”space warp” transformations [42] with great success. The space warp estimator yields qualitatively similar behavior to the regularized estimator I developed, and potentially has a lower variance. The development of the ”space warp” estimator demonstrates the flexibility and simplicity of the regularization technique — it is a method which can be easily improved, implemented, and tested.

My work on developing a penalized estimator for excited states in QMC, as shown in Chapter 6, stands as a significant improvement to the state of the art [44]. The new method addresses many of the shortcomings of current use QMC excited state techniques while maintaining computational efficiency and accuracy as shown by my computation of the excitation spectrum of benzene. A recent review article places the penalized estimator alongside the variance optimization of Neuscamman [185] and Filippi's state-averaged optimization [33] as the bleeding edge techniques for excited state calculation in QMC. These three methods form a suite of techniques that, together, have demonstrated the capabilities of QMC in accurately computing excited states from small molecules like benzene to solids like LiH.

And lastly, in Chapter 7 I illustrated my development of a tight binding model for TBLG with lattice interactions. This tight binding model went beyond the state-of-the-art as it was fit to *ab initio* data directly, using DFT band structures for various bilayer nuclear configurations as training data. The resultant tight binding model was shown to quantitatively describe the tight binding parameter variations with lattice configuration more accurately than the state-of-the art Moon and Koshino model [163]. Further, by computing band structures for TBLG from the fit tight binding model, I was able to demonstrate that lattice relaxation is neither a requirement for isolated band flattening nor the fragile topology of the isolated flat bands, in contrast to earlier calculations with innacurate tight binding models arguing the necessity for lattice relaxation.

The frontier of first principles methods and effective models is no doubt the accurate computation of properties of strongly correlated solids, and I believe my thesis work stands as a useful contribution to this goal. Accurate ground and excited state computation for strongly correlated materials will give us amazing insights into the emergent behaviors of exotic materials — unconventional superconductors, spin glasses, and topological materials, to name a few. Future researchers who want to accurately compute ground and excited state properties in QMC may use the wave functions, estimators, and excited state techniques I have developed, or adapt and improve upon them as seen fit. The development of accurate effective models with electron-electron, electron-lattice, and even electron-spin models will give us high level understanding of the effective atomic- or field-level descriptions which can be used to study exotic systems as well. Researchers who want to accurately model materials can take training data and a framework of modelling from my study on TBLG, adding in important pieces like effective electron-electron interactions. Whatever the future may hold, I hope that my work on accurate and efficient computation of ground and excited states from first principles, as well as models with electron-lattice interactions, will contribute to the progression of the field towards these exciting achievements.

References

[1] Adolfo Avella and Ferdinando Mancini. *Strongly Correlated Systems, Numerical Methods.* Springer-Verlag Berlin Heidelberg, 2013.

[2] Claudio Attaccalite and Sandro Sorella. Stable liquid hydrogen at high pressure by a novel ab initio molecular-dynamics calculation. *Phys. Rev. Lett.*, 100:114501, Mar 2008.

[3] Andrea Zen, Ye Luo, Sandro Sorella, and Leonardo Guidoni. Molecular properties by quantum monte carlo: An investigation on the role of the wave function ansatz and the basis set in the water molecule. *Journal of Chemical Theory and Computation*, 9(10):4332–4350, 2013.

[4] C. J. Umrigar. Observations on variational and projector monte carlo methods. *The Journal of Chemical Physics*, 143(16):164105, 2015.

[5] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, January 1989.

[6] Frédéric Joucken, Eberth A. Quezada-López, Jose Avila, Chaoyu Chen, John L. Davenport, Hechin Chen, Kenji Watanabe, Takashi Taniguchi, Maria Carmen Asensio, and Jairo Velasco. Nanospot angle-resolved photoemission study of bernal-stacked bilayer graphene on hexagonal boron nitride: Band structure and local variation of lattice alignment. *Phys. Rev. B*, 99:161406, Apr 2019.

[7] Richard M. Martin. Theoretical Background. In *Electronic Structure*, pages 52–71. Cambridge University Press, 2004.

[8] Kiel T. Williams, Yuan Yao, Jia Li, Li Chen, Hao Shi, Mario Motta, Chunyao Niu, Ushnish Ray, Sheng Guo, Robert J. Anderson, Junhao Li, Lan Nguyen Tran, Chia-Nan Yeh, Bastien Mussard, Sandeep Sharma, Fabien Bruneval, Mark van Schilfgaarde, George H. Booth, Garnet Kin-Lic Chan, Shiwei Zhang, Emanuel Gull, Dominika Zgid, Andrew Millis, Cyrus J. Umrigar, and Lucas K. Wagner. Direct comparison of many-body methods for realistic electronic hamiltonians. *Phys. Rev. X*, 10:011041, Feb 2020.

[9] Jolyon Aarons, Misbah Sarwar, David Thompsett, and Chris-Kriton Skylaris. Perspective: Methods for large-scale density functional calculations on metallic systems. *The Journal of Chemical Physics*, 145(22):220901, 2016.

[10] Chris-Kriton Skylaris, Peter D. Haynes, Arash A. Mostofi, and Mike C. Payne. Introducing onetep: Linear-scaling density functional simulations on parallel computers. *The Journal of Chemical Physics*, 122(8):084119, 2005.

[11] Lin Li, Ask H. Larsen, Nichols A. Romero, Vitali A. Morozov, Christian Glinsvad, Frank Abild-Pedersen, Jeff Greeley, Karsten W. Jacobsen, and Jens K. Nørskov. Investigation of catalytic finite-size-effects of platinum metal clusters. *The Journal of Physical Chemistry Letters*, 4(1):222–226, 2013. PMID: 26291235.

[12] D. Alfè, M. J. Gillan, and G. D. Price. Ab initio chemical potentials of solid and liquid solutions and the chemistry of the earth's core. *The Journal of Chemical Physics*, 116(16):7127–7136, 2002.

[13] Álvaro Ruiz-Serrano and Chris-Kriton Skylaris. A variational method for density functional theory calculations on metallic systems with thousands of atoms. *The Journal of Chemical Physics*, 139(5):054107, 2013.

[14] Lucas Wagner and Lubos Mitas. A quantum monte carlo study of electron correlation in transition metal oxygen molecules. *Chemical Physics Letters*, 370(3):412–417, 2003.

[15] Christopher J. Cramer and Donald G. Truhlar. Density functional theory for transition metals and transition metal chemistry. *Phys. Chem. Chem. Phys.*, 11:10757–10816, 2009.

[16] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum monte carlo simulations of solids. *Rev. Mod. Phys.*, 73:33–83, Jan 2001.

[17] Jind řich Kolorenč and Lubos Mitas. Quantum monte carlo calculations of structural properties of feo under pressure. *Phys. Rev. Lett.*, 101:185502, Oct 2008.

[18] Kateryna Foyevtsova, Jaron T. Krogel, Jeongnim Kim, P. R. C. Kent, Elbio Dagotto, and Fernando A. Reboredo. Ab initio quantum monte carlo calculations of spin superexchange in cuprates: The benchmarking case of $\mathrm{ca_2cuo_3}$. *Phys. Rev. X*, 4:031003, Jul 2014.

[19] Lucas K. Wagner and Peter Abbamonte. Effect of electron correlation on the electronic structure and spin-lattice coupling of high-T_c cuprates: Quantum monte carlo calculations. *Phys. Rev. B*, 90:125129, Sep 2014.

[20] Lucas K. Wagner. Ground state of doped cuprates from first-principles quantum monte carlo calculations. *Phys. Rev. B*, 92:161116, Oct 2015.

[21] Brian Busemeyer, Mario Dagrada, Sandro Sorella, Michele Casula, and Lucas K. Wagner. Competing collinear magnetic structures in superconducting fese by first-principles quantum monte carlo calculations. *Phys. Rev. B*, 94:035108, Jul 2016.

[22] Ángel Morales-García, Rosendo Valero, and Francesc Illas. An empirical, yet practical way to predict the band gap in solids by using density functional band structure calculations. *The Journal of Physical Chemistry C*, 121(34):18862–18866, 2017.

[23] Hai Xiao, Jamil Tahir-Kheli, and William A. Goddard. Accurate band gaps for semiconductors from density functional theory. *The Journal of Physical Chemistry Letters*, 2(3):212–217, 2011.

[24] Fabien Tran and Peter Blaha. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Phys. Rev. Lett.*, 102:226401, Jun 2009.

[25] Pragya Verma and Donald G. Truhlar. Hle16: A local kohn–sham gradient approximation with good performance for semiconductor band gaps and molecular excitation energies. *The Journal of Physical Chemistry Letters*, 8(2):380–387, 2017. PMID: 28033712.

[26] E. Pavarini, I. Dasgupta, T. Saha-Dasgupta, O. Jepsen, and O. K. Andersen. Band-structure trend in hole-doped cuprates and correlation with t_{cmax}. *Phys. Rev. Lett.*, 87:047003, Jul 2001.

[27] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, and A. I. Lichtenstein. Frequency-dependent local interactions and low-energy effective models from electronic structure calculations. *Phys. Rev. B*, 70:195104, Nov 2004.

[28] M.E. Casida and M. Huix-Rotllant. Progress in time-dependent density-functional theory. *Annual Review of Physical Chemistry*, 63(1):287–323, 2012. PMID: 22242728.

[29] Neepa T. Maitra. Perspective: Fundamental aspects of time-dependent density functional theory. *The Journal of Chemical Physics*, 144(22):220901, 2016.

[30] Lionel Lacombe and Neepa T. Maitra. Developing new and understanding old approximations in tddft. *Faraday Discuss.*, 224:382–401, 2020.

[31] Filip Vlahovic, Marko Peric, Maja Gruden-Pavlovic, and Matija Zlatar. Assessment of td-dft and lf-dft for study of d-d transitions in first row transition metal hexaaqua complexes. *The Journal of Chemical Physics*, 142(21):214111, 2015.

[32] 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. *Journal of the American Chemical Society*, 126(12):4007–4016, 2004. PMID: 15038755.

[33] Monika Dash, Jonas Feldt, Saverio Moroni, Anthony Scemama, and Claudia Filippi. Excited states with selected configuration interaction-quantum monte carlo: Chemically accurate excitation energies and geometries. *Journal of Chemical Theory and Computation*, 15(9):4896–4906, 2019. PMID: 31348645.

[34] Nick S. Blunt and Eric Neuscamman. Excited-state diffusion monte carlo calculations: A simple and efficient two-determinant ansatz. *Journal of Chemical Theory and Computation*, 15(1):178–189, 2019.

[35] Adam A. Holmes, C. J. Umrigar, and Sandeep Sharma. Excited states using semistochastic heat-bath configuration interaction. *The Journal of Chemical Physics*, 147(16):164111, 2017.

[36] Fabien Tran, Leila Kalantari, Boubacar Traoré, Xavier Rocquefelte, and Peter Blaha. Nonlocal van der waals functionals for solids: Choosing an appropriate one. *Phys. Rev. Materials*, 3:063602, Jun 2019.

[37] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Van der waals density functional for general geometries. *Phys. Rev. Lett.*, 92:246401, Jun 2004.

[38] Kyuho Lee, Éamonn D. Murray, Lingzhu Kong, Bengt I. Lundqvist, and David C. Langreth. Higher-accuracy van der waals density functional. *Phys. Rev. B*, 82:081101, Aug 2010.

[39] Huihuo Zheng, Hitesh J. Changlani, Kiel T. Williams, Brian Busemeyer, and Lucas K. Wagner. From real materials to model hamiltonians with density matrix downfolding. *Frontiers in Physics*, 6:43, 2018.

[40] Shivesh Pathak and Lucas K. Wagner. Non-orthogonal determinants in multi-slater-jastrow trial wave functions for fixed-node diffusion monte carlo. *The Journal of Chemical Physics*, 149(23):234104, 2018.

[41] Lucas K. Wagner and Lubos Mitas. Energetics and dipole moment of transition metal monoxides by quantum Monte Carlo. *Journal of Chemical Physics*, 126(3), 2007.

[42] Jesse van Rhijn, Claudia Filippi, Stefania De Palo, and Saverio Moroni. Energy derivatives in real-space diffusion monte carlo, 2021.

[43] Shivesh Pathak and Lucas K. Wagner. A light weight regularization for wave function parameter gradients in quantum monte carlo. *AIP Advances*, 10(8):085213, 2020.

[44] Shivesh Pathak, Brian Busemeyer, João N. B. Rodrigues, and Lucas K. Wagner. Excited states in variational monte carlo using a penalty method. *The Journal of Chemical Physics*, 154(3):034101, 2021.

[45] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.

[46] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[47] Krishnan Raghavachari and James B. Anderson. Electron correlation effects in molecules. *The Journal of Physical Chemistry*, 100(31):12960–12973, 1996.

[48] Filipp Furche and John P. Perdew. The performance of semilocal and hybrid density functionals in 3d transition-metal chemistry. *The Journal of Chemical Physics*, 124(4):044103, 2006.

[49] Nathan E. Schultz, Yan Zhao, and Donald G. Truhlar. Density functionals for inorganometallic and organometallic chemistry. *The Journal of Physical Chemistry A*, 109(49):11127–11143, 2005. PMID: 16331896.

[50] Elbek K. Kurbanov, Hannah R. Leverentz, Donald G. Truhlar, and Elizabeth A. Amin. Electrostatically embedded many-body expansion for neutral and charged metalloenzyme model systems. *Journal of Chemical Theory and Computation*, 8(1):1–5, 2012. PMID: 22639556.

[51] R. M. Lee, G. J. Conduit, N. Nemec, P. López Ríos, and N. D. Drummond. Strategies for improving the efficiency of quantum monte carlo calculations. *Phys. Rev. E*, 83:066706, Jun 2011.

[52] Julien Toulouse, Roland Assaraf, and Cyrus J. Umrigar. Chapter fifteen - introduction to the variational and diffusion monte carlo methods. In Philip E. Hoggan and Telhat Ozdogan, editors, *Electron Correlation in Molecules – ab initio Beyond Gaussian Quantum Chemistry*, volume 73 of *Advances in Quantum Chemistry*, pages 285–314. Academic Press, 2016.

[53] Andrea Zen, Ye Luo, Guglielmo Mazzola, Leonardo Guidoni, and Sandro Sorella. Ab initio molecular dynamics simulation of liquid water by quantum monte carlo. *The Journal of Chemical Physics*, 142(14):144111, 2015.

[54] Paul M. Zimmerman, Julien Toulouse, Zhiyong Zhang, Charles B. Musgrave, and C. J. Umrigar. Excited states of methylene from quantum monte carlo. *The Journal of Chemical Physics*, 131(12):124103, 2009.

[55] R. M. Grimes, B. L. Hammond, P. J. Reynolds, and W. A. Lester. Quantum monte carlo approach to electronically excited molecules. *The Journal of Chemical Physics*, 85(8):4749–4750, 1986.

[56] Miguel A. Morales, Jeremy McMinis, Bryan K. Clark, Jeongnim Kim, and Gustavo E. Scuseria. Multideterminant wave functions in quantum monte carlo. *Journal of Chemical Theory and Computation*, 8(7):2181–2188, 2012. PMID: 26588949.

[57] Claudio Genovese, Tomonori Shirakawa, Kousuke Nakano, and Sandro Sorella. General correlated geminal ansatz for electronic structure calculations: Exploiting pfaffians in place of determinants. *Journal of Chemical Theory and Computation*, 16(10):6114–6131, 2020. PMID: 32804497.

[58] Eric Neuscamman, C. J. Umrigar, and Garnet Kin-Lic Chan. Optimizing large parameter sets in variational quantum monte carlo. *Phys. Rev. B*, 85:045103, Jan 2012.

[59] Julien Toulouse and C. J. Umrigar. Full optimization of jastrow–slater wave functions with application to the first-row atoms and homonuclear diatomic molecules. *The Journal of Chemical Physics*, 128(17):174101, 2008.

[60] Alice Cuzzocrea, Anthony Scemama, Wim J. Briels, Saverio Moroni, and Claudia Filippi. Variational principles in quantum monte carlo: The troubled story of variance minimization. *Journal of Chemical Theory and Computation*, 16(7):4203–4212, 2020. PMID: 32419451.

[61] C. J. Umrigar. Observations on variational and projector Monte Carlo methods. *Journal of Chemical Physics*, 143(16):0–16, 2015.

[62] H.M. Taylor and S. Karlin. *An Introduction to Stochastic Modeling.* Academic Press, 1981.

[63] Andrea Zen, Sandro Sorella, Michael J. Gillan, Angelos Michaelides, and Dario Alfè. Boosting the accuracy and speed of quantum monte carlo: Size consistency and time step. *Phys. Rev. B*, 93:241118, Jun 2016.

[64] E. Y. Loh, J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar. Sign problem in the numerical simulation of many-electron systems. *Phys. Rev. B*, 41:9301–9307, May 1990.

[65] Norm M. Tubman, Jonathan L DuBois, Randolph Q. Hood, and Berni J. Alder. Prospects for release-node quantum monte carlo. *The Journal of Chemical Physics*, 135(18):184109, 2011.

[66] C. J. Umrigar, Julien Toulouse, Claudia Filippi, S. Sorella, and R. G. Hennig. Alleviation of the fermion-sign problem by optimization of many-body wave functions. *Phys. Rev. Lett.*, 98:110201, Mar 2007.

[67] Junhao Li, Matthew Otten, Adam A. Holmes, Sandeep Sharma, and C. J. Umrigar. Fast semistochastic heat-bath configuration interaction. *The Journal of Chemical Physics*, 149(21):214110, 2018.

[68] Arne Lüchow and James B. Anderson. Monte carlo methods in electronic structures for large systems. *Annual Review of Physical Chemistry*, 51(1):501–526, 2000. PMID: 11031291.

[69] W M C Foulkes, L Mitas, R J Needs, and G Rajagopal. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.*, 73(1):33–83, 2001.

[70] Julien Toulouse and C. J. Umrigar. Optimization of quantum Monte Carlo wave functions by energy minimization. *Journal of Chemical Physics*, 126(8):0–16, 2007.

[71] C. J. Umrigar, Julien Toulouse, Claudia Filippi, S. Sorella, and R. G. Hennig. Alleviation of the fermion-sign problem by optimization of many-body wave functions. *Physical Review Letters*, 98(11):1–4, 2007.

[72] Julien Toulouse and C. J. Umrigar. Full optimization of Jastrow-Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules. *Journal of Chemical Physics*, 128(17), 2008.

[73] Henrik Koch and Esper Dalgaard. Linear superposition of optimized non-orthogonal Slater determinants for singlet states. *Chemical Physics Letters*, 212(1-2):193–200, 1993.

[74] Jarrod R. McClean and Alan Aspuru-Guzik. Compact wavefunctions from compressed imaginary time evolution. *RSC Adv.*, 5(124):102277–102283, 2015.

[75] Hidekazu Goto, Masashi Kojo, Akira Sasaki, and Kikuji Hirose. Essentially exact ground-state calculations by superpositions of nonorthogonal slater determinants. *Nanoscale Research Letters*, 8(1):200, May 2013.

[76] William A. Goddard, Thom H. Dunning, William J. Hunt, and P. Jeffrey Hay. Generalized valence bond description of bonding in low-lying states of molecules. *Accounts of Chemical Research*, 6(11):368–376, 1973.

[77] Carlos A. Jimenez-Hoyos, R. Rodriguez-Guzman, and Gustavo E. Scuseria. Multi-component symmetry-projected approach for molecular ground state correlations. *The Journal of Chemical Physics*, 139(20):204102, 2013.

[78] Laimutis Bytautas, Carlos A. Jimenez-Hoyos, R. Rodriguez-Guzman, and Gustavo E. Scuseria. Potential energy curves for mo2: multi-component symmetry-projected hartree-fock and beyond. *Molecular Physics*, 112(14):1938–1946, 2014.

[79] Hao Shi, Carlos A. Jiménez-Hoyos, R. Rodríguez-Guzmán, Gustavo E. Scuseria, and Shiwei Zhang. Symmetry-projected wave functions in quantum monte carlo calculations. *Phys. Rev. B*, 89:125129, Mar 2014.

[80] Norikazu Tomita. Many-body wave functions approximated by the superposition of spin-projected nonorthogonal slater determinants in the resonating hartree-fock method. *Phys. Rev. B*, 69:045110, Jan 2004.

[81] Atsushi Ikawa, Shoji Yamamoto, and Hideo Fukutome. Orbital optimization in the resonating hartree-fock approximation and its application to the one dimensional hubbard model. *Journal of the Physical Society of Japan*, 62(5):1653–1668, 1993.

[82] Lucas K. Wagner, Michal Bajdich, and Lubos Mitas. Qwalk: A quantum monte carlo program for electronic structure. *Journal of Computational Physics*, 228(9):3390 – 3404, 2009.

[83] M. Burkatzki, C. Filippi, and M. Dolg. Energy-consistent pseudopotentials for quantum Monte Carlo calculations. *Journal of Chemical Physics*, 126(23), 2007.

[84] M. Burkatzki, Claudia Filippi, and M. Dolg. Energy-consistent small-core pseudopotentials for 3d-transition metals adapted to quantum Monte Carlo calculations. *The Journal of Chemical Physics*, 129(16):164115, 2008.

[85] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin Lic Chan. PySCF: the Python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1), 2018.

[86] Claudia Filippi, Roland Assaraf, and Saverio Moroni. Simple formalism for efficient derivatives and multi-determinant expansions in quantum Monte Carlo. *Journal of Chemical Physics*, 144(19), 2016.

[87] Michele Casula. Beyond the locality approximation in the standard diffusion monte carlo method. *Phys. Rev. B*, 74:161102, Oct 2006.

[88] K. Kirby and B. Liu. The valence states of c2: A configuration interaction study. *The Journal of Chemical Physics*, 70(2):893–900, 1979.

[89] Sandro Sorella. Generalized lanczos algorithm for variational quantum monte carlo. *Phys. Rev. B*, 64:024512, Jun 2001.

[90] Michele Casula and Sandro Sorella. Geminal wave functions with jastrow correlation: A first application to atoms. *The Journal of Chemical Physics*, 119(13):6500–6511, 2003.

[91] C. J. Umrigar and Claudia Filippi. Energy and variance optimization of many-body wave functions. *Phys. Rev. Lett.*, 94:150201, Apr 2005.

[92] Roland Assaraf and Michel Caffarel. Zero-variance principle for monte carlo algorithms. *Phys. Rev. Lett.*, 83:4682–4685, Dec 1999.

[93] Roland Assaraf and Michel Caffarel. Computing forces with quantum monte carlo. *The Journal of Chemical Physics*, 113(10):4028–4034, 2000.

[94] Roland Assaraf and Michel Caffarel. Zero-variance zero-bias principle for observables in quantum monte carlo: Application to forces. *The Journal of Chemical Physics*, 119(20):10536–10552, 2003.

[95] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1989.

[96] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. Pyscf: the python-based simulations of chemistry framework. *WIREs Computational Molecular Science*, 8(1):e1340, 2018.

[97] PyQMC: Python library for real space quantum Monte Carlo. `https://github.com/WagnerGroup/pyqmc`.

[98] N. D. Drummond, J. R. Trail, and R. J. Needs. Trail-needs pseudopotentials in quantum monte carlo calculations with plane-wave/blip basis sets. *Phys. Rev. B*, 94:165170, Oct 2016.

[99] Junhao Li, Matthew Otten, Adam A. Holmes, Sandeep Sharma, and C. J. Umrigar. Fast semistochastic heat-bath configuration interaction. *J. Chem. Phys.*, 148:214110, 2018.

[100] Sandeep Sharma, Adam A. Holmes, Guillaume Jeanmairet, Ali Alavi, and C. J. Umrigar. Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory. *J. Chem. Theory Comput.*, 13:1595–1604, 2017.

[101] Adam A. Holmes, Norm M. Tubman, and C. J. Umrigar. Heat-bath configuration interaction: An efficient selected ci algorithm inspired by heat-bath sampling. *J. Chem. Theory Comput.*, 12:3674–3680, 2016.

[102] Michele Casula, Claudio Attaccalite, and Sandro Sorella. Correlated geminal wave function for molecules: An efficient resonating valence bond approach. *The Journal of Chemical Physics*, 121(15):7110–7126, 2004.

[103] J. R. Trail. Heavy-tailed random error in quantum monte carlo. *Phys. Rev. E*, 77:016703, Jan 2008.

[104] D. M. Ceperley and B. Bernu. The calculation of excited state properties with quantum Monte Carlo. *The Journal of Chemical Physics*, 89(10):6316–6328, November 1988. Publisher: American Institute of Physics.

[105] N. S. Blunt, Simon D. Smart, George H. Booth, and Ali Alavi. An excited-state approach within full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 143(13):134117, October 2015.

[106] Claudia Filippi, Maurizio Zaccheddu, and Francesco Buda. Absorption spectrum of the green fluorescent protein chromophore: A difficult case for ab initio methods? *Journal of Chemical Theory and Computation*, 5(8):2074–2087, 2009. PMID: 26613149.

[107] Monika Dash, Jonas Feldt, Saverio Moroni, Anthony Scemama, and Claudia Filippi. Excited states with selected configuration interaction-quantum monte carlo: Chemically accurate excitation energies and geometries. *Journal of Chemical Theory and Computation*, 15(9):4896–4906, 2019. PMID: 31348645.

[108] Robert Send, Omar Valsson, and Claudia Filippi. Electronic excitations of simple cyanine dyes: Reconciling density functional and wave function methods. *Journal of Chemical Theory and Computation*, 7(2):444–455, 2011. PMID: 26596164.

[109] Sergio D. Pineda Flores and Eric Neuscamman. Excited State Specific Multi-Slater Jastrow Wave Functions. *The Journal of Physical Chemistry A*, 123(8):1487–1497, February 2019.

[110] Luning Zhao and Eric Neuscamman. An efficient variational principle for the direct optimization of excited states. *Journal of Chemical Theory and Computation*, 12(8):3436–3440, 2016. PMID: 27379468.

[111] Jacqueline A. R. Shea and Eric Neuscamman. Size consistent excited states via algorithmic transformations between variational principles. *Journal of Chemical Theory and Computation*, 13(12):6078–6088, 2017. PMID: 29140699.

[112] Sergio D. Pineda Flores and Eric Neuscamman. Excited state specific multi-slater jastrow wave functions. *The Journal of Physical Chemistry A*, 123(8):1487–1497, 2019. PMID: 30702890.

[113] Nick S. Blunt and Eric Neuscamman. Excited-state diffusion monte carlo calculations: A simple and efficient two-determinant ansatz. *Journal of Chemical Theory and Computation*, 15(1):178–189, 2019.

[114] Alice Cuzzocrea, Anthony Scemama, Wim J. Briels, Saverio Moroni, and Claudia Filippi. Variational Principles in Quantum Monte Carlo: The Troubled Story of Variance Minimization. *Journal of Chemical Theory and Computation*, 16(7):4203–4212, July 2020.

[115] E.M. Stoudenmire and Steven R. White. Studying Two-Dimensional Systems with the Density Matrix Renormalization Group. *Annual Review of Condensed Matter Physics*, 3(1):111–128, March 2012.

[116] Michele Casula, Claudio Attaccalite, and Sandro Sorella. Correlated geminal wave function for molecules: An efficient resonating valence bond approach. *The Journal of Chemical Physics*, 121(15):7110–7126, October 2004.

[117] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum Monte Carlo simulations of solids. *Reviews of Modern Physics*, 73(1):33–83, January 2001.

[118] Lucas K. Wagner, Michal Bajdich, and Lubos Mitas. Qwalk: A quantum monte carlo program for electronic structure. *Journal of Computational Physics*, 228(9):3390 – 3404, 2009.

[119] J. Koster and S. Rahmann. Snakemake–a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522, October 2012.

[120] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. Pyscf: the python-based simulations of chemistry framework. *WIREs Computational Molecular Science*, 8(1):e1340, 2018.

[121] Björn O. Roos, Kerstin Andersson, and Markus P. Fülscher. Towards an accurate molecular orbital theory for excited states: the benzene molecule. *Chemical Physics Letters*, 192(1):5 – 13, 1992.

[122] Ove Christiansen, Henrik Koch, Asger Halkier, Poul Jo/rgensen, Trygve Helgaker, and Alfredo Sánchez de Merás. Large-scale calculations of excitation energies in coupled cluster theory: The singlet excited states of benzene. *The Journal of Chemical Physics*, 105(16):6921–6939, 1996.

[123] Marko Schreiber, Mario R. Silva-Junior, Stephan P. A. Sauer, and Walter Thiel. Benchmarks for electronically excited states: Caspt2, cc2, ccsd, and cc3. *The Journal of Chemical Physics*, 128(13):134110, 2008.

[124] M. Burkatzki, C. Filippi, and M. Dolg. Energy-consistent pseudopotentials for quantum monte carlo calculations. *The Journal of Chemical Physics*, 126(23):234105, 2007.

[125] Harrison Shull. Vibrational analysis of the 3400a triplet-singlet emission of benzene. *The Journal of Chemical Physics*, 17(3):295–303, 1949.

[126] J. P. Doering. Electronic energy levels of benzene below 7 ev. *The Journal of Chemical Physics*, 67(9):4065–4070, 1977.

[127] Ove Christiansen, John F. Stanton, and Jurgen Gauss. A coupled cluster study of the 1 1a1g andu 1 1b2u states of benzene. *The Journal of Chemical Physics*, 108(10):3987–4001, 1998.

[128] Robert J Bursill, Christopher Castleton, and William Barford. Optimal parametrisation of the pariser–parr–pople model for benzene and biphenyl. *Chemical Physics Letters*, 294(4):305 – 313, 1998.

[129] Carlo Adamo, Gustavo E. Scuseria, and Vincenzo Barone. Accurate excitation energies from time-dependent density functional theory: Assessing the pbe0 model. *The Journal of Chemical Physics*, 111(7):2889–2899, 1999.

[130] LAWRENCE D. ZIEGLER and BRUCE S. HUDSON. The vibronic spectroscopy of benzene: Old problems and new techniques. In EDWARD C. LIM, editor, *Excited States*, pages 41–140. Academic Press, 1982.

[131] B. Hajgató, D. Szieberth, P. Geerlings, F. De Proft, and M. S. Deleuze. A benchmark theoretical study of the electronic ground state and of the singlet-triplet split of benzene and linear acenes. *The Journal of Chemical Physics*, 131(22):224321, 2009.

[132] Koichi Ohno and Risako Takahashi. Excited-state vibrations of benzene and polycyclic aromatic hydrocarbons: simple force field models based on molecular orbital characteristics of hexagonal carbon networks. *Chemical Physics Letters*, 356(3):409 – 422, 2002.

[133] Pierre-François Loos, Martial Boggio-Pasqua, Anthony Scemama, Michel Caffarel, and Denis Jacquemin. Reference Energies for Double Excitations. *Journal of Chemical Theory and Computation*, 15(3):1939–1956, March 2019.

[134] Yongkyung Kwon, D. M. Ceperley, and Richard M. Martin. Effects of three-body and backflow correlations in the two-dimensional electron gas. *Physical Review B*, 48(16):12037, October 1993.

[135] P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs. Inhomogeneous backflow transformations in quantum Monte Carlo calculations. *Physical Review E*, 74(6):066701, December 2006.

[136] Luca F. Tocchio, Federico Becca, Alberto Parola, and Sandro Sorella. Role of backflow correlations for the nonmagnetic phase of the t–$t^{'}$ hubbard model. *Phys. Rev. B*, 78:041101, Jul 2008.

[137] Di Luo and Bryan K. Clark. Backflow transformations via neural networks for quantum many-body wave functions. *Phys. Rev. Lett.*, 122:226401, Jun 2019.

[138] David Pfau, James S. Spencer, Alexander G. D. G. Matthews, and W. M. C. Foulkes. *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks. *Physical Review Research*, 2(3):033429, September 2020.

[139] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, February 2017.

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

[141] Huihuo Zheng and Lucas K. Wagner. Computation of the correlated metal-insulator transition in vanadium dioxide from first principles. *Phys. Rev. Lett.*, 114:176401, Apr 2015.

[142] Yuan Cao, Valla Fatemi, Shiang Fang, Kenji Watanabe, Takashi Taniguchi, Efthimios Kaxiras, and Pablo Jarillo-Herrero. Unconventional superconductivity in magic-angle graphene superlattices. *Nature*, 556(7699):43–50, Apr 2018.

[143] Yuan Cao, Valla Fatemi, Ahmet Demir, Shiang Fang, Spencer L. Tomarken, Jason Y. Luo, Javier D. Sanchez-Yamagishi, Kenji Watanabe, Takashi Taniguchi, Efthimios Kaxiras, Ray C. Ashoori, and Pablo Jarillo-Herrero. Correlated insulator behaviour at half-filling in magic-angle graphene superlattices. *Nature*, 556(7699):80–84, Apr 2018.

[144] Y. Cao, J. Y. Luo, V. Fatemi, S. Fang, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero. Superlattice-induced insulating states and valley-protected orbits in twisted bilayer graphene. *Phys. Rev. Lett.*, 117:116804, Sep 2016.

[145] Yu Saito, Jingyuan Ge, Kenji Watanabe, Takashi Taniguchi, and Andrea F. Young. Independent superconductors and correlated insulators in twisted bilayer graphene. *Nature Physics*, 16(9):926–930, Sep 2020.

[146] Canxun Zhang, Tiancong Zhu, Salman Kahn, Shaowei Li, Birui Yang, Charlotte Herbig, Xuehao Wu, Hongyuan Li, Kenji Watanabe, Takashi Taniguchi, Stefano Cabrini, Alex Zettl, Michael P. Zaletel, Feng Wang, and Michael F. Crommie. Visualizing delocalized correlated electronic states in twisted double bilayer graphene. *Nature Communications*, 12(1):2516, May 2021.

[147] Jianpeng Liu and Xi Dai. Orbital magnetic states in moiré graphene systems. *Nature Reviews Physics*, 3(5):367–382, May 2021.

[148] Minhao He, Yuhao Li, Jiaqi Cai, Yang Liu, K. Watanabe, T. Taniguchi, Xiaodong Xu, and Matthew Yankowitz. Symmetry breaking in twisted double bilayer graphene. *Nature Physics*, 17(1):26–30, Jan 2021.

[149] Nguyen N. T. Nam and Mikito Koshino. Lattice relaxation and energy band modulation in twisted bilayer graphene. *Phys. Rev. B*, 96:075311, Aug 2017.

[150] G. Trambly de Laissardière, D. Mayou, and L. Magaud. Numerical studies of confined states in rotated bilayers of graphene. *Phys. Rev. B*, 86:125413, Sep 2012.

[151] E. Suárez Morell, J. D. Correa, P. Vargas, M. Pacheco, and Z. Barticevic. Flat bands in slightly twisted bilayer graphene: Tight-binding calculations. *Phys. Rev. B*, 82:121407, Sep 2010.

[152] Rafi Bistritzer and Allan H. MacDonald. Moiré bands in twisted double-layer graphene. *Proceedings of the National Academy of Sciences*, 108(30):12233–12237, 2011.

[153] Louk Rademaker, Dmitry A. Abanin, and Paula Mellado. Charge smoothening and band flattening due to hartree corrections in twisted bilayer graphene. *Phys. Rev. B*, 100:205114, Nov 2019.

[154] Grigory Tarnopolsky, Alex Jura Kruchkov, and Ashvin Vishwanath. Origin of magic angles in twisted bilayer graphene. *Phys. Rev. Lett.*, 122:106405, Mar 2019.

[155] Simone Lisi, Xiaobo Lu, Tjerk Benschop, Tobias A. de Jong, Petr Stepanov, Jose R. Duran, Florian Margot, Irène Cucchi, Edoardo Cappelli, Andrew Hunter, Anna Tamai, Viktor Kandyba, Alessio Giampietri, Alexei Barinov, Johannes Jobst, Vincent Stalman, Maarten Leeuwenhoek, Kenji Watanabe, Takashi Taniguchi, Louk Rademaker, Sense Jan van der Molen, Milan P. Allan, Dmitri K. Efetov, and Felix Baumberger. Observation of flat bands in twisted bilayer graphene. *Nature Physics*, 17(2):189–193, Feb 2021.

[156] Yuhang Jiang, Xinyuan Lai, Kenji Watanabe, Takashi Taniguchi, Kristjan Haule, Jinhai Mao, and Eva Y. Andrei. Charge order and broken rotational symmetry in magic-angle twisted bilayer graphene. *Nature*, 573(7772):91–95, Sep 2019.

[157] Shiang Fang, Stephen Carr, Ziyan Zhu, Daniel Massatt, and Efthimios Kaxiras. Angle-dependent *ab initio* low-energy hamiltonians for a relaxed twisted bilayer graphene heterostructure, 2019.

[158] Kazuyuki Uchida, Shinnosuke Furuya, Jun-Ichi Iwata, and Atsushi Oshiyama. Atomic corrugation and electron localization due to moiré patterns in twisted bilayer graphenes. *Phys. Rev. B*, 90:155451, Oct 2014.

[159] Shuyang Dai, Yang Xiang, and David J. Srolovitz. Twisted bilayer graphene: Moiré with a twist. *Nano Letters*, 16(9):5923–5927, 2016. PMID: 27533089.

[160] Sandeep K Jain, Vladimir Juričić, and Gerard T Barkema. Structure of twisted and buckled bilayer graphene. *2D Materials*, 4(1):015018, nov 2016.

[161] Giovanni Cantele, Dario Alfè, Felice Conte, Vittorio Cataudella, Domenico Ninno, and Procolo Lucignano. Structural relaxation and low-energy properties of twisted bilayer graphene. *Phys. Rev. Research*, 2:043127, Oct 2020.

[162] Stephen Carr, Shiang Fang, Ziyan Zhu, and Efthimios Kaxiras. Exact continuum model for low-energy electronic states of twisted bilayer graphene. *Phys. Rev. Research*, 1:013001, Aug 2019.

[163] Pilkyung Moon and Mikito Koshino. Energy spectrum and quantum hall effect in twisted bilayer graphene. *Phys. Rev. B*, 85:195458, May 2012.

[164] Hoi Chun Po, Haruki Watanabe, and Ashvin Vishwanath. Fragile topology and wannier obstructions. *Phys. Rev. Lett.*, 121:126402, Sep 2018.

[165] Junyeong Ahn, Dongwook Kim, Youngkuk Kim, and Bohm-Jung Yang. Band topology and linking structure of nodal line semimetals with Z_2 monopole charges. *Phys. Rev. Lett.*, 121:106403, Sep 2018.

[166] Hoi Chun Po, Liujun Zou, T. Senthil, and Ashvin Vishwanath. Faithful tight-binding models and fragile topology of magic-angle bilayer graphene. *Phys. Rev. B*, 99:195455, May 2019.

[167] Zhida Song, Zhijun Wang, Wujun Shi, Gang Li, Chen Fang, and B. Andrei Bernevig. All magic angles in twisted bilayer graphene are topological. *Phys. Rev. Lett.*, 123:036401, Jul 2019.

[168] Liujun Zou, Hoi Chun Po, Ashvin Vishwanath, and T. Senthil. Band structure of twisted bilayer graphene: Emergent symmetries, commensurate approximants, and wannier obstructions. *Phys. Rev. B*, 98:085435, Aug 2018.

[169] Junyeong Ahn, Sungjoon Park, and Bohm-Jung Yang. Failure of nielsen-ninomiya theorem and fragile topology in two-dimensional systems with space-time inversion symmetry: Application to twisted bilayer graphene at magic angle. *Phys. Rev. X*, 9:021013, Apr 2019.

[170] Fang Xie, Zhida Song, Biao Lian, and B. Andrei Bernevig. Topology-bounded superfluid weight in twisted bilayer graphene. *Phys. Rev. Lett.*, 124:167002, Apr 2020.

[171] Jess Wellendorff, Keld T. Lundgaard, Andreas Møgelhøj, Vivien Petzold, David D. Landis, Jens K. Nørskov, Thomas Bligaard, and Karsten W. Jacobsen. Density functionals for surface science: Exchange-correlation model development with bayesian error estimation. *Phys. Rev. B*, 85:235149, Jun 2012.

[172] T. Bredow M.F. Peintinger, D.V. Oliveira. Consistent gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations. *J. Comput. Chem.*, Mar 2013.

[173] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. Pyscf: the python-based simulations of chemistry framework. *WIREs Computational Molecular Science*, 8(1):e1340, 2018.

[174] Qiming Sun, Xing Zhang, Samragni Banerjee, Peng Bao, Marc Barbry, Nick S. Blunt, Nikolay A. Bogdanov, George H. Booth, Jia Chen, Zhi-Hao Cui, Janus J. Eriksen, Yang Gao, Sheng Guo, Jan Hermann, Matthew R. Hermes, Kevin Koh, Peter Koval, Susi Lehtola, Zhendong Li, Junzi Liu, Narbe Mardirossian, James D. McClain, Mario Motta, Bastien Mussard, Hung Q. Pham, Artem Pulkin, Wirawan Purwanto, Paul J. Robinson, Enrico Ronca, Elvira R. Sayfutyarova, Maximilian Scheurer, Henry F. Schurkus, James E. T. Smith, Chong Sun, Shi-Ning Sun, Shiv Upadhyay, Lucas K. Wagner, Xiao Wang, Alec White, James Daniel Whitfield, Mark J. Williamson, Sebastian Wouters, Jun Yang, Jason M. Yu, Tianyu Zhu, Timothy C. Berkelbach, Sandeep Sharma, Alexander Yu. Sokolov, and Garnet Kin-Lic Chan. Recent developments in the pyscf program package. *The Journal of Chemical Physics*, 153(2):024109, 2020.

[175] Ivo Souza, Nicola Marzari, and David Vanderbilt. Maximally localized wannier functions for entangled energy bands. *Phys. Rev. B*, 65:035109, Dec 2001.

[176] Nicola Marzari and David Vanderbilt. Maximally localized generalized wannier functions for composite energy bands. *Phys. Rev. B*, 56:12847–12865, Nov 1997.

[177] Nicola Marzari, Arash A. Mostofi, Jonathan R. Yates, Ivo Souza, and David Vanderbilt. Maximally localized wannier functions: Theory and applications. *Rev. Mod. Phys.*, 84:1419–1475, Oct 2012.

[178] Alberto Ambrosetti and Pier Luigi Silvestrelli. *Introduction to Maximally Localized Wannier Functions*, chapter 6, pages 327–368. John Wiley and Sons, Ltd, 2016.

[179] Arash A. Mostofi, Jonathan R. Yates, Giovanni Pizzi, Young-Su Lee, Ivo Souza, David Vanderbilt, and Nicola Marzari. An updated version of wannier90: A tool for obtaining maximally-localised wannier functions. *Computer Physics Communications*, 185(8):2309–2310, 2014.

[180] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[181] Fadil Santosa and William W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. and Stat. Comput.*, 1986.

[182] Shiang Fang and Efthimios Kaxiras. Electronic structure theory of weakly interacting bilayers. *Phys. Rev. B*, 93:235153, Jun 2016.

[183] Stephen Carr, Shiang Fang, Hoi Chun Po, Ashvin Vishwanath, and Efthimios Kaxiras. Derivation of wannier orbitals and minimal-basis tight-binding hamiltonians for twisted bilayer graphene: First-principles approach. *Phys. Rev. Research*, 1:033072, Nov 2019.

[184] Joshua P. Townsend, Sergio D. Pineda Flores, Raymond C. Clay, Thomas R. Mattsson, Eric Neuscamman, Luning Zhao, R. E. Cohen, and Luke Shulenburger. Starting-point-independent quantum monte carlo calculations of iron oxide. *Phys. Rev. B*, 102:155151, Oct 2020.

[185] Sergio D. Pineda Flores and Eric Neuscamman. Excited state specific multi-slater jastrow wave functions. *The Journal of Physical Chemistry A*, 123(8):1487–1497, 2019. PMID: 30702890.