Optimization Algorithms in Variational Monte Carlo for Molecular Excited States

by

Leon Otis

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Eric Neuscamman, Co-chair
Professor Jeffrey Neaton, Co-chair
Professor Steven Louie
Professor Martin Head-Gordon

Spring 2022

Optimization Algorithms in Variational Monte Carlo for Molecular Excited States

Copyright 2022
by
Leon Otis

Abstract

Optimization Algorithms in Variational Monte Carlo for Molecular Excited States

by

Leon Otis

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Eric Neuscamman, Co-chair

Professor Jeffrey Neaton, Co-chair

The accurate description of molecular excited states is an active frontier in the development of electronic structure theory. Compared to their ground state counterparts, electronic structure methods for excited states remain less well developed and many of the most widely used options struggle to provide accurate and predictive results across all varieties of excited states. This thesis presents a series of contributions to the development of variational Monte Carlo (VMC) as a broadly reliable tool for excited states, with a particular emphasis on wave function optimization algorithms.

We first present an analysis of current first and second order optimization techniques in VMC, demonstrating that each class of algorithm has strengths and weaknesses in terms of convergence and computational efficiency. The complementary nature of these traits motivates the design of a hybrid optimization algorithm, which can overcome the individual disadvantages of its constituent methods. After first developing this approach for ground state energy minimization, we extend it to the minimization of a variance-based objective function for the state-specific targeting of excited states. We demonstrate that employing VMC with the hybrid optimization algorithm can obtain highly accurate excitation energies across multiple classes of excited states, including single valence excitations, double excitations, and charge transfer. This level of reliability currently eludes all other electronic structure methods that can be applied beyond small molecules and enables VMC to serve as a widely applicable benchmark method for assessing which computationally cheaper methods are most trustworthy in various chemical contexts. We also continue our methodological work by presenting a systematic study of the issue of optimization stability in VMC and identify best practices for successful variance minimization that may aid the work of other researchers.

To my parents

Contents

Contents **ii**

List of Figures **iv**

List of Tables **ix**

1 Introduction **1**
1.1 The Schrodinger Equation and the Born-Oppenheimer Approximation 1
1.2 Overview of Electronic Structure Theory Methods 2
1.3 Electronic Excitations and Current Excited State Methods 5
1.4 Variational Monte Carlo 7
1.5 Stochastic Optimization Algorithms 10
1.6 Outline 12

2 Complementary First and Second Derivative Methods for Ansatz Optimization in Variational Monte Carlo **14**
2.1 Abstract 14
2.2 Introduction 14
2.3 Theory 17
2.4 Results 30
2.5 Conclusions and Outlook 47
2.6 Acknowledgements 48
2.7 Appendix A: Additional Energies 49
2.8 Appendix B: Details for N_2 Optimizations 50
2.9 Appendix C: Molecular Geometries 55
2.10 Appendix D: NCJF Gaussian Basis 55

3 A hybrid approach to excited-state-specific variational Monte Carlo and doubly excited states **59**
3.1 Abstract 59
3.2 Introduction 59
3.3 Theory 61

3.4 Results 68
3.5 Conclusion 83

4 Combining State-Specific Quantum Chemistry and Quantum Monte Carlo for Molecular Excited States 84
4.1 Abstract 84
4.2 Introduction 84
4.3 Theory 88
4.4 Results 93
4.5 Conclusions 105
4.6 Acknowledgments 106
4.7 Appendix 106

5 Optimization Stability in Variational Monte Carlo 112
5.1 Abstract 112
5.2 Introduction 112
5.3 Theory 114
5.4 Results 121
5.5 Conclusions 135
5.6 Acknowledgements 136
5.7 Appendix 136

6 Conclusions 137

Bibliography 139

List of Figures

1.1 Orbital spaces typically defined in CASSCF. The core, active, and virtual spaces are boxed in red, green, and blue respectively. 4

2.1 Flowchart depicting steps in the BLM algorithm to arrive at a parameter update. 21
2.2 Illustration of the difficulty faced by steepest descent in red on the lower left with its slow approach to the minimum. Accelerated descent in green on the upper right is able to progress more rapidly to the minimum with its memory of previous gradients. 24
2.3 Schematic depiction of a typical energy optimization using the hybrid method. The dashed box around a section of descent in green and BLM in red defines a macro-iteration of the method. 27
2.4 Schematic representation of gradient descent corrections in green to red BLM steps, which we have observed to reduce the uncertainty about the location of the final minimum. 27
2.5 Different methods' optimized energies relative to that of the LM for equilibrium N_2 when optimizing CI coefficients and the TJF. 32
2.6 Different methods' optimized energies relative to that of the LM for stretched N_2 when optimizing CI coefficients and the TJF. 33
2.7 Different methods' optimized energies relative to that of the traditional LM for equilibrium N_2 when all parameters are optimized simultaneously. See also Table 2.2. 34
2.8 Different methods' optimized energies relative to that of the traditional LM for stretched N_2 when all parameters are optimized simultaneously. See also Table 2.3. 35
2.9 The standard LM optimization for all parameter equilibrium N_2. 38
2.10 Values for the first off-diagonal F-matrix element (F_{01}), the first electron-nuclear TJF spline parameter (u_0), and the second orbital rotation variable (X_{02}) at each micro iteration of the "Hybrid 1" optimization for all parameter equilibrium N_2. 39
2.11 Standard Errors for the hybrid method and LM on all parameter equilibrium N_2 vs different optimizations' total sampling costs. 40
2.12 Standard Errors for the hybrid method and LM on all parameter stretched N_2 plotted against different optimizations' total sampling costs. 41
2.13 Converged energies in equilibrium N_2 before and after a final descent optimization. 42
2.14 Converged energies in stretched N_2 before and after a final descent optimization. 43

2.15 Equilibrium geometry of styrene. See Appendix C for structure coordinates. . . 43
2.16 Converged energies in equilibrium styrene before and after a final descent optimization. 44
2.17 Left panel: energies during stages 3 and 4 of the FNNF optimization. The descent energies are reported as the average over the last 50 RMSprop steps within each block of 100 RMSprop steps, whereas the BLM energies are the energy estimates on the random samples used for the BLM update steps. Right panel: change in the value of the F-matrix parameter that couples the two nitrogen atoms' counting regions over the first three macro iterations of stage 3, with each micro iteration corresponding to one RMSprop or BLM step. The nine BLM points on the right panel correspond to the first nine BLM points on the left panel. . . . 45

3.1 Examples of unstable and stable LM optimizations of Ω for CN5. We consider the singlet near -40.5 E_h, which is dominated by a HOMO to LUMO excitation, as well as the next triplet and singlet above it, which are dominated by a HOMO-2 to LUMO excitation. Initial wave function guesses were obtained from a diagonalization of the Hamiltonian in a 26 determinant space with pre-optimized one- and two-body Jastrow factors present. These Jastrows were optimized further along with the 26 determinant coefficients. For each state, the value of ω was set to the E-σ value from a one million sample evaluation using the initial wave function and held fixed throughout the optimization. The LM shifts were kept constant at 0.1 for the diagonal shift and 1 for the overlap-based shift throughout the optimizations with 50,000 samples per iteration in all cases. For the unstable cases, we found that using increased sampling effort alone did not ensure stability. A proposed parameter update was rejected if our correlated sampling assessment predicted it would raise the target function value, but all optimizations contain hundreds of accepted steps. Horizontal lines show the lowest 7 eigenenergies from our diagonalization, with solid lines for singlet states and dashed for triplets. . 69
3.2 Fractions of the CI expansion corresponding to the third singlet's dominant CSF over the course of LM optimizations. When starting from the energy eigenstate with only 1- and 2-body Jastrows are present, the weight of this CSF quickly collapses as the optimization drifts to a higher singlet. This optimization used 50,000 samples per iteration and starting from solely the dominant CSF with an additional 3-body Jastrow delayed, but did not prevent, the drift at this sampling effort. Upon increasing the sample size to 500,000 samples per iteration, the 3-body Jastrow wave function optimization became stable, as shown by the solid line. Note that the CI expansion fraction is simply the sum of the squares of the determinant coefficients within the given CSF divided by the sum of all the squared determinant coefficients. 71

3.3 Excitation energy of the 2 $^1\Sigma_g$ state in C_2 for LM and hybrid method on both MSJ and all parameter ansatzes. See also Table 3.1. 20 determinants were used for all ground state optimizations. The excited state optimizations used 23 determinants for both the LM and Hybrid MSJ cases, 71 for the LM All case, and 50 for Hybrid All case. The benchmark value is taken from a selected CI calculation using the CIPSI algorithm.[41] 73
3.4 Example optimization of target function Ω for all parameters using the hybrid method. Points for AD correspond to an average over 50 iterations while those for the Blocked LM are individual iterations. 74
3.5 Excitation energy of the 2 $^1A'$ state in nitroxyl for LM and hybrid method on MSJ ansatzes. See also Table 3.2. Both methods used 20 determinants the ground state and 40 for the excited state. Reference values for coupled cluster, CASPT2, and selected CI are taken from the work of Loos and coworkers.[41] They take the selected CI value as the theoretical best estimate. 76
3.6 Excitation energy of the 2 ^1A_g state in glyoxal for LM and hybrid method on MSJ ansatzes. See also Table 3.3. Both methods used 50 determinants the ground state and 20 for the excited state. Reference values for coupled cluster, CASPT2, and selected CI are taken from the work of Loos and coworkers.[41] Their theoretical best estimate is obtained by adding a 0.06 eV basis set correction to the selected CI result. 77
3.7 Excitation energy of the 3 $^1A'$ state in acrolein for LM and hybrid method on MSJ ansatzes. See also Table 3.4. Both methods used 20 determinants the ground state and 70 for the excited state. Reference values for coupled cluster, CASPT2, and selected CI are taken from the work of Loos and coworkers.[41] Their theoretical best estimate is obtained by adding a -.13 eV basis set correction to the selected CI result. 78
3.8 Excitation energy of 3 $^1A_1$ state in CPD for LM and hybrid method. See also Table 3.5. Both the LM and hybrid optimizations used 20 and 500 determinants for the ground and excited states, respectively. Literature values are taken from Schreiber et al.[190] for CASPT2 and CCSD LR and from Shen and Li[212] for MR-CISD+Q. 79
3.9 Excitation energy of the doubly excited state in ABN for LM and hybrid method. See also Table 3.7. Excitation energies for all other methods were obtained in Molpro. 20 determinants were used in all ground state optimizations. The excited state optimizations used 70, 100, 500, and 300 determinants for the LM SA, Hybrid SA, LM SS, and Hybrid SS cases respectively. These numbers of determinants were chosen to achieve an explicit variance match between the ground and excited states for each method. 82

4.1 Example of the linear regression used in variance matching. The points are energies and variances from individual blocked LM iterations in the optimization of the excited state of ketene. See Table 4.1 for how the variance matching affects the excitation energy prediction. 92
4.2 From left to right, excitation energies in thioformaldehyde, methanimine, and ketene for VMC and quantum chemistry approaches. See also Table 4.2. Reference values for coupled cluster, and selected CI are taken from the work of Loos and coworkers.[37] 94
4.3 Excitation energies in the carbon trimer and nitrosomethane in the left and right panels respectively. See also Table 4.3. For the carbon trimer, VMC used 100 determinants in the ground state and 500 for the excited state, while 500 and 1000 respectively were used for nitrosomethane. Reference values for coupled cluster, and selected CI are taken from the work of Loos and coworkers.[41] 97
4.4 Excitation energies in hexatriene and benzene in the left and right panels respectively. See also Table 4.4. In both molecules, the VMC used 500 determinants in ground state and 1000 for the excited state. Reference values for coupled cluster, CASPT2 and NEVPT2 are taken from the work of Loos and coworkers.[41] . . . 99
4.5 Excitation energy of the lowest CT state in ammonia-difluorine for VMC and quantum chemistry approaches. See also Table 4.5. The VMC used 300 determinants in the ground state and 500 for the excited state. 102
4.6 Excitation energies in uracil. The left panel depicts values for the 2 $^1A'$ state and the right panel for the 3 $^1A'$ state. See also Table 4.6. The VMC used 500 determinants the ground state and 1500 for each of the excited states. 104

5.1 Proposed LM updates to Jastrow parameter for varying total numbers of parameters. Each panel was produced with 200 proposed updates from independent LM runs on an excited state wave function in CN5. The same scale has been used in each panel, which leaves off some larger outliers in the 200 and 500 CI cases. From left to right, the total numbers of wave function parameters for the different cases are 150, 340, and 640. 100,000 samples were used to generate the LM matrices for each update. 117
5.2 Proposed LM updates to Jastrow parameter with diagonal shift of 100 for varying total numbers of parameters. Each panel shows 200 proposed updates from independent LM runs with an excited state wave function in CN5. From left to right, the total numbers of wave function parameters for the different cases are 150, 340, and 640. 100,000 samples were used to generate the LM matrices for each update. 118

5.3 Excitation energy in CN5 using the hybrid method to minimize the Ω target function. Reference values for SA-VMC,[91] CC3,[91] CASPT2,[88, 259] GW/BSE,[259] and sCI[11] are taken from the literature. For comparison, the basis sets used in the literature results were aug-cc-pVTZ for CC3 and GW/BSE, aug-cc-pVDZ for sCI, ANO-L-VTZP for CASPT2, and a double-ζ basis set minimally augmented with s and p diffuse functions on heavy atoms for SA-VMC. 123
5.4 Excited state energy for fixed shift LM, adaptive shift LM, and hybrid method. Starting point was unoptimized Jastrow and CI coefficients from SA-CASSCF. Green dashed line marks the end of the first stage where only Jastrow and CI parameters are optimized. 126
5.5 Objective function Ω for fixed shift LM, adaptive shift LM, and hybrid method. Starting point was unoptimized Jastrow and CI coefficients from SA-CASSCF.Green dashed line marks the end of the first stage where only Jastrow and CI parameters are optimized. 127
5.6 Excited state variance for fixed shift LM, adaptive shift LM, and hybrid method. Starting point was unoptimized Jastrow and CI coefficients from SA-CASSCF. Green dashed line marks the end of the first stage where only Jastrow and CI parameters are optimized. The separation between AD and blocked LM points is an artifact of autocorrelation. 128
5.7 Zoomed in view of excited state energy for adaptive shift LM, and hybrid method. Green dashed line marks the end of the first stage where only Jastrow and CI parameters are optimized. 130
5.8 Optimization of all parameters including orbitals using the hybrid method starting from the end of the LM fixed shift optimization of only the Jastrow and CI parameters. Green line is shown to the left to make clear these plots show only the stage with orbital optimization. 131
5.9 Optimization of all parameters including orbitals using the LM with fixed shifts while varying omega. Starting point was the optimized MSJ wave function from the preceding LM fixed shift optimization of only the Jastrow and CI parameters. Green line is shown to the left to make clear these plots show only the stage with orbital optimization. 132
5.10 Optimization of all parameters including orbitals using the hybrid method while varying omega. Starting point was the optimized MSJ wave function from the preceding hybrid method optimization of only the Jastrow and CI parameters. Green line is shown to the left to make clear these plots show only the stage with orbital optimization. The separation between AD and blocked LM variance points is an artifact of autocorrelation. 133
5.11 Varying values of omega during fixed shift LM and hybrid method optimizations of all parameters. 134

List of Tables

2.1 Rough memory cost scaling for the optimization methods we examine, with N the number of optimized parameters and N_b the number of blocks. 28
2.2 Energies,uncertainties, and sample numbers for optimization of all parameters in equilibrium N_2. 36
2.3 Energies,uncertainties, and sample numbers for optimization of all parameters in stretched N_2. 37
2.4 A summary of the VMC optimization stages in FNNF showing the number of determinants N_d included from HCI, which parameters are optimized, and the total number N_p of optimized parameters. Note that CI coefficients are optimized at every stage. Stages 2, 3, and 4 start from the parameter values from the previous stage, with newly added determinants' coefficients initialized to zero. We also report the number of iterations performed in each stage, which for stage 4 is simply the number of RMSprop steps. A hybrid iteration, on the other hand, consists of 100 RMSprop steps followed by three BLM steps. All RMSprop steps use 20,000 samples drawn from $|\Psi|^2$, while the BLM steps each use 1 million samples drawn from the $|\Phi|^2$ guiding function with ϵ set to 0.01. 42
2.5 Energies of the transition state of FNNF. 44
2.6 Precise Values for optimizing CI coefficients and traditional Jastrow factors in equilibrium N_2. 49
2.7 Precise Values for optimizing CI coefficients and traditional Jastrow factors in stretched N_2. 50
2.8 Step sizes for TJFCI equilibrium N_2 optimizations. 52
2.9 Step sizes for TJFCI stretched N_2 optimizations. 52
2.10 Step sizes for all parameter equilibrium N_2 optimizations. 53
2.11 Step sizes for all parameter stretched N_2 optimizations. 54
2.12 Structure of equilibrium styrene. Coordinates in Å. 55
2.13 Structure of FNNF transition state. Coordinates in Å. 55
2.14 Components of the matrix **A** for our systems. 56
2.15 Gaussian components for all parameter equilibrium N_2. 56
2.16 Gaussian basis functions for all parameter stretched N_2. 57
2.17 Gaussian basis functions for equilibrium styrene. 57
2.18 Gaussian basis functions for FNNF. 58

3.1 Excitation energies and uncertainties for the 2 $^1\Sigma_g$ state in C_2. Literature values for selected CI, CCSDT, CC3, and CASPT2 using an aug-cc-pVQZ basis set are included for comparison.[41] 73
3.2 Excitation energies and uncertainties for the 2 $^1A'$ state in nitroxyl. The included literature values for CCSDT, CC3, CASPT2, and selected CI all use an aug-cc-pVQZ basis set.[41] 76
3.3 Excitation energies and uncertainties for the 2 ^1A_g state in glyoxal. The included literature values use an aug-cc-pVDZ basis set for selected CI and aug-cc-pVQZ for CCSDT, CC3, and CASPT2.[41] 77
3.4 Excitation energies and uncertainties for the 3 $^1A'$ state in acrolein. The included literature values use a 6-31+G(d) basis set for selected CI, an aug-cc-pVTZ basis CC3, and an aug-cc-pVQZ basis for the three versions of CASPT2.[41] 78
3.5 Excitation Energies and uncertainties for 3 $^1A_1$ state in CPD. 80
3.6 Head to head comparison of LM and hybrid for CPD on the same excited state wave function at fixed ω. 80
3.7 Excitation Energies and uncertainties for ABN. 82

4.1 Energy and variance values in ketene before and after the variance matching procedure. As shown later in Table 4.2, the final excitation energy prediction is very accurate compared to the literature theoretical best estimate of 3.86 eV.[37] 92
4.2 Excitation energies for thioformaldehyde, methanimine, and ketene. The VMC uncertainties on the final digits are given in parentheses. The included literature values use an aug-cc-pVTZ basis set for sCI and coupled cluster, and the literature TBEs are obtained from applying a basis set correction using CC3 to the sCI result.[37] 95
4.3 Excitation energies for the carbon trimer and nitrosomethane. The VMC uncertainties on the final digits are given in parentheses. The included literature[41] values use an aug-cc-pVQZ basis set for the carbon trimer and the CC3 and CASPT2 values for nitrosomethane. The literature nitrosomethane CCSDT and sCI results used an aug-cc-pvTZ basis set. 98
4.4 Excitation energies for hexatriene and benzene. The VMC uncertainties on the final digits are given in parentheses. The included literature[41] values use an aug-cc-pVDZ basis set for CC3 and CCSDT in hexatriene and an aug-cc-pVQZ basis set for all varieties of CASPT2 and NEVPT2, except for PC-NEVPT2 in benzene, which used aug-cc-pVTZ. 100
4.5 Excitation energies for the CT state in ammonia-difluorine.The VMC uncertainty for the last digit is given in parentheses. 102
4.6 Excitation energies in eV for the 2 $^1A'$ and 3 $^1A'$ states in uracil. The VMC uncertainty for the last digit is given in parentheses. The literature results from Thiel[190] and Szalay[252] used a TZVP basis set, with the Thiel set using an aug-cc-pVQZ CC2 result as its reference. Details on the ANO type basis set used by Roos can be found in the original paper.[251] 105

4.7 Structure of thioformaldehyde. Coordinates in Å. 106
4.8 Structure of methanimine. Coordinates in Å. 107
4.9 Structure of ketene. Coordinates in Å. 107
4.10 Structure of carbon trimer. Coordinates in Å. 107
4.11 Structure of nitrosomethane. Coordinates in Å. 107
4.12 Structure of hexatriene. Coordinates in Å. 108
4.13 Structure of benzene. Coordinates in Å. 108
4.14 Structure of ammonia-fluorine. Coordinates in Å. 108
4.15 Structure of uracil. Coordinates in Å. 109
4.16 Macro-iterations used for GS and ES optimizations. Samples per iteration used for AD and blocked LM portions of the hybrid optimizations. Total samples for AD and LM with the costs for GS and ES combined. The nominal samples per iteration for blocked LM iterations is multiplied by 3 for computing the total sampling effort due to the need to run over the samples additional times to perform the automated parameter selection and the blocked LM algorithm. . . . 110
4.17 Uncertainties in the excitation energies due to stochastic averaging and variance matching extrapolation. Raw variance values for ground and excited states. Energy-variance slopes used in the extrapolations and uncertainties in the slopes due the uncertainty in the individual blocked LM optimization points. 111

5.1 Excitation energies for CN5 for state-specific VMC and literature results.[11, 88, 91, 259] Stochastic uncertainties on the last digit are given in parentheses. . . . 123
5.2 Structure of CN5. Coordinates in Å. 136

Acknowledgments

The road I have walked during my PhD has been a long one and I would not have reached the end without vast amounts of help and support from many sources. My first and deepest thanks must go to my advisor, Eric Neuscamman. Eric has been the greatest of mentors and he is an inspiring model of the type of scientist I hope to become one day. Unfailingly kind, patient, and generous with his advice and time, Eric has taught me more than I could have imagined when I first arrived in Berkeley, unaware of his existence, without any formal chemistry education since high school, and with only the vaguest awareness of electronic structure theory. The example of his deep thinking, probing questions, and eternal optimism that research problems can be solved much faster than they actually are has guided me to become a more careful and successful researcher. Beyond his guidance during the week-to-week challenges and failures of research, Eric has also been an invaluable source of career and life advice that has broadened my perspective on science and my time in grad school. I will always be deeply grateful to Eric for welcoming me into his group.

Next, I would like to thank Jeff Neaton for serving as my co-advisor in the Physics department and his useful advice on research and career development over the years. I also thank Steve Louie and Martin Head-Gordon for their service on my qualifying exam and thesis committees, and I greatly appreciate Holger Müller's willingness to serve on my qualifying exam on short notice after my original experimental member had a schedule conflict. Several administrative staff members: Anne Takizawa, Donna Sakima, Joelle Miles, and Laurie Mason have been enormously helpful in navigating bureaucratic requirements and in helping me get settled into life within both the Physics and Chemistry departments.

I also want to thank all Neuscamman group members, both those who came before me and those who arrived after, for making the group a wonderful environment for the past five years. Luning (Chris) Zhao, Beatrice Van Der Goetz, Sergio Pineda Flores, and Jacki Shea all kindly welcomed me into the group and helped me with learning the basics of doing calculations and wrapping my head around the alphabet soup of electronic structure methods. I also greatly benefited from conversations with Peter Walters and Lan Tran during their time as postdocs. The group has grown since those early years and I have enjoyed all the conversations and day-to-day collaborative work I have had in that time with Scott Garner, Connie Robinson, Becky Hanscam, Tarini Haridkar, Rachel Clune, Isaac Craig, Sonja Bumann, Harrison Tuckman, and Kat Muloma. Their hard work in research and dedication to making the group an evermore tightly-knit and supportive place have been inspiring and have made work with the group a source of stability and comfort while living through a devastating pandemic and worldwide upheaval. I look forward to seeing everything the group will accomplish in the years to come.

Finally, I wish to thank my parents for all their unceasing support and encouragement during grad school and all my life before. Their love has always been at the deepest core of my life and I would never have reached the end of this journey without it.

Chapter 1

Introduction

1.1 The Schrodinger Equation and the Born-Oppenheimer Approximation

Quantum mechanics provides a fundamental description of matter neatly encapsulated in the Schrodinger equation.

$$i\hbar\frac{\partial\left|\Psi(t)\right\rangle}{\partial t} = H\left|\Psi(t)\right\rangle \tag{1.1}$$

At the most general level, the Schrodinger equation relates the action of an operator called the Hamiltonian H upon a state $|\Psi\rangle$ to the evolution of that state $|\Psi\rangle$ in time. However, in the description of the atoms and molecules within all everyday matter, the Hamiltonian takes a particular form.

$$H = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_I \frac{1}{2M_I}\nabla_I^2 - \sum_i\sum_I \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \sum_{i>j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I>J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \tag{1.2}$$

The five terms of H describe the kinetic energies of the electrons and nuclei and the potential energies resulting from the electron-nuclear, electron-electron and nuclear-nuclear interactions. In equation 1.2 and throughout this thesis, we adopt atomic units where the electron charge and mass as well as $\hbar$ are set to 1. We index the electron and nuclear positions with i and I respectively, and M_I and Z_I are the nuclear masses and atomic numbers.

For solving the Schrodinger equation in this case, multiple simplifications are made in this thesis. First, we confine ourselves to finding solutions of the time-independent version of the Schrodinger equation $H\left|\Psi\right\rangle = E\left|\Psi\right\rangle$, seeking stationary states $|\Psi\rangle$ with energy eigenvalues E. Second, we employ the Born-Oppenheimer (BO) approximation,[1] which exploits the fact that nuclei have masses thousands of times greater than the electron mass and therefore, their positions can be regarded as essentially fixed compared to the motion of electrons. Within the BO approximation, we neglect the nuclear kinetic energy term and leave off

the now-constant nuclear-nuclear potential energy to arrive at the following form for the Hamiltonian.

$$H_e = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_i \sum_I \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{1.3}$$

The remaining three terms define the electronic Hamiltonian H_e and obtaining its eigenstates is still a formidable task due to the electron-electron repulsion term. In general, exact solutions cannot be obtained for non-trivial systems and the task of electronic structure theory is the development of efficient computational techniques for achieving approximate solutions.

1.2 Overview of Electronic Structure Theory Methods

Since the origins of electronic structure theory, a sprawling landscape of different methods has been created and this thesis will give only broad outlines of major approaches to contextualize the later chapters. Electronic structure theory can be divided at a high level between density functional based methods[2] and wave function based approaches.[1, 3] Density functional theory (DFT) relies on the formulation of the energy as a functional of the electron density and is the more widely used of the two paradigms due to its computational affordability. However, all functionals used within DFT are only approximate and the difficulty of improving the quality of the wide array of possible functionals[4, 5] serves as motivation for developing more predictive methods, particularly for cases where DFT has lower accuracy.

We focus more heavily on wave function methods, which generally possess the capacity for systematic improvability in their accuracy. The hierarchy of wave function techniques begins with Hartree-Fock theory. Hartree-Fock restricts the wave function ansatz to a single Slater determinant of one-electron orbitals and optimizes the orbital shapes to minimize the energy. Within this approximation of Hartree-Fock theory, each electron experiences only an averaged mean-field potential due to the other electrons, without dependence on their exact positions. This limitation leads to the definition of the difference between the true electronic energy and the Hartree-Fock energy as the correlation energy.

Electron correlation is frequently discussed in terms of strong and weak correlation. Strong correlation occurs in physical situations where the Hartree-Fock determinant is qualitatively inaccurate and other sets of multiple Slater determinants provide the dominant part of the wave function. Weak correlation can be viewed as smaller corrections to the qualitative picture provided by the major electron configurations such as the Hartree-Fock determinant. More sophisticated wave function methods aim to retrieve enough of the missing correlation energy to achieve quantitative accuracy, but maintaining accurate descriptions of both types of correlation remains a major challenge.

One intuitive approach to improving upon Hartree-Fock is to simply include more Slater determinants in what are known as Configuration Interaction (CI) methods. Additional Slater determinants can be obtained by considering configurations of electrons with excitations relative to the Hartree-Fock configuration.

$$|\Psi\rangle = |\Psi_{HF}\rangle + \sum_{i,a} C_i^a |\Psi_i^a\rangle + \sum_{i,j,a,b} C_{ij}^{ab} |\Psi_{ij}^{ab}\rangle + ... \tag{1.4}$$

In equation 1.4, we denote orbitals that are occupied in $|\Psi_{HF}\rangle$ with i and j and unoccupied orbitals with a and b. The additional determinants $|\Psi_i^a\rangle$ and $|\Psi_{ij}^{ab}\rangle$ are then respectively singly and doubly excited determinants relative to the Hartree-Fock determinant. The sums ensure all possible single and double excitations are included in the wave function, and in principle higher excitations can also be added. In full configuration interaction (FCI),[6–8] all possible levels of excitations in a given orbital set are included and the exact wave function in that finite orbital basis can be achieved. Unfortunately, FCI is prohibitively expensive beyond the smallest molecules and basis sets and more approximate forms of CI theory have been developed. A simple method of approximation is to truncate the addition of determinants at a particular level of excitation such as CI singles and doubles (CISD). An alternative is to develop criteria for identifying and including the most important determinants, which is the basis for a family of selected CI (sCI) methods.[9–13] An example of an sCI method is the heatbath CI algorithm,[10] which uses the selection criterion:

$$f(D_i) = \max_j(|H_{ij}c_j|) \tag{1.5}$$

where c_j is the coefficient of a determinant D_j already in the sCI wave function and H_{ij} is the Hamiltonian matrix element between the prospective determinant D_i and D_j. D_i will be added to the sCI wave function only if the largest $|H_{ij}c_j|$ after running over j is above some chosen threshold. The various sCI algorithms differ in the design of their criterion and other algorithmic details, but all fundamentally rely on this type of comparison for building up their wave function. Using a threshold of zero would lead to constructing the FCI wave function and while sCI algorithms are similarly exponentially scaling, their greater efficiency enables them to obtain essentially FCI quality results in somewhat larger systems.

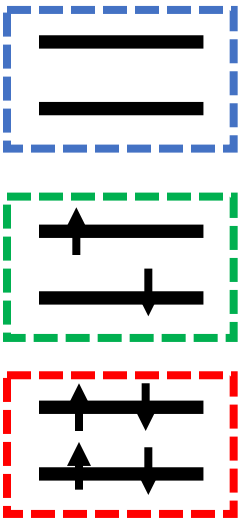


Figure 1.1: Orbital spaces typically defined in CASSCF. The core, active, and virtual spaces are boxed in red, green, and blue respectively.

A related idea is to perform FCI or a near-FCI quality level of sCI with a subset of orbitals and electrons called an active space. This is the basis of the complete active space self-consistent field (CASSCF) method. The low energy, doubly occupied core orbitals and the high energy, unoccupied virtual orbitals, depicted alongside the active space in Figure 1.1, are left out of the FCI calculation. Due to the cost of FCI, the active space is chosen to include only the most chemically important orbitals, which makes CASSCF non-black box and its results potentially sensitive to different active space choices by users of varying levels of chemical intuition and experience. When using FCI as the active space solver, the upper limit of CASSCF calculations is roughly 20 electrons in 20 orbitals.[14] Spaces of this size may be too small to hold all the important orbitals in some chemical systems. For example, molecules with transition metals, such as porphyrins, require two sets of d orbitals for an effective description,[15–18] which quickly grows the minimum size of a responsible active space choice. A major use of sCI in these types of larger molecules is to act as a solver for active spaces that are too expensive for a FCI solution.[12, 19] Another weakness of CASSCF is that while it succeeds at capturing all the electronic correlation associated with the active space, it lacks the ability to describe weak correlation effects involving orbitals outside the space. One approach for addressing this is to perform perturbation theory using the CASSCF wave function as a reference, which leads to the CASPT2 method.[20, 21]

Another prominent approach in electronic structure is coupled cluster (CC) theory.

$$|\Psi_{CC}\rangle = e^{\hat{T}} |\Psi_{HF}\rangle \tag{1.6}$$

The CC wave function is constructed from the Hartree-Fock determinant by applying an exponential of the cluster operator $\hat{T}$

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + ... = \sum_{ia} t_i^a \hat{c}_a^\dagger \hat{c}_i + \sum_{ijab} t_{ij}^{ab} \hat{c}_a^\dagger \hat{c}_b^\dagger \hat{c}_i \hat{c}_j + ... \tag{1.7}$$

The cluster operator can be broken up into terms that describe single, double, and higher levels of excitations from the occupied Hartree-Fock orbitals to the virtual orbitals. Including all levels of excitations will produce the full coupled cluster (FCC) wave function, which is equivalent to FCI. In practice, the cluster operator is truncated at some order such as $\hat{T} = \hat{T}_1 + \hat{T}_2$, which defines the CCSD method. While both CI and CC theory describe correlation through the inclusion of excited determinants, CC methods enjoy the advantage of being size consistent due to the wave function's exponential form. Satisfying this appealing property and CC's effective treatment of weak correlation have made its more accurate incarnations, particularly CCSD(T), popular benchmark methods for ground state quantum chemistry.[22]

1.3 Electronic Excitations and Current Excited State Methods

The theoretical machinery surveyed in the previous section was historically first developed for studying the ground states of physical systems, which correspond mathematically to the eigenstate of the electronic Hamiltonian with the lowest eigenvalue. However, the higher eigenstates, or electronically excited states, play a key role in a variety of biologically and technologically important phenomena. Two major examples are water splitting in photosynthesis[23] and the development of new solar cell materials.[24] The importance of excited states to these topics and these fields' role in developing renewable energy motivate the development of effective theoretical models to aid experimental work.

Molecular excited states appear in a variety of types including single valence excitations, double excitations, charge transfer (CT), Rydberg excitations and core excitations. Accurately describing all these classes of excited states is challenging task for any theoretical approach. Many of the methods introduced for the ground state have also been reformulated and applied to excited states. Excited state electronic structure methods are less fully developed compared to ground state theories, but an extensive array of approaches has already been produced and we discuss the most commonly utilized ones in this section.

On the density functional side of electronic structure, time-dependent density functional theory (TDDFT) builds upon ground state DFT and has been widely employed for excited state and absorption spectra studies.[25] However, its most common forms are known to been inaccurate for double excitations and charge transfer states,[26–29] and the large number of functional choices produces the same challenges for predictive power as in the ground state. TDDFT's difficulties for double excitations can be traced to the fact it is usually formulated as a linear response approach. The formalism of TDDFT requires computing the density-density response function $\chi(\mathbf{r}, \mathbf{r}', \omega)$, which is dependent on the frequency for a general

interacting system, but is taken to be frequency-independent in the widely used adiabatic approximation. Within the adiabatic approximation, only single excitations appear in the TDDFT spectrum, while double and higher excitations are completely absent.[26, 30, 31] For CT excitations, the simplest forms of TDDFT will severely underestimate excitation energies and fail to produce the asymptotic $\frac{1}{R}$ behavior as the distance between charge donor and acceptor molecules is increased. For larger donor-acceptor distances, the TDDFT excitation energy reduces to the difference of the Kohn-Sham donor and acceptor orbital energies, which do not include relaxation effects from the movement of charge.[28–30] While employing range-separated hybrid functionals with optimal tuning of the range parameter[32] improves the performance of TDDFT, CT excitations, particularly over intermediate distances, remain challenging. Taken together, the difficulties faced by TDDFT motivate the exploration of alternative methods that can both go beyond linear response and incorporate the effects of orbital relaxation.

As for wave function methods, FCI can of course provide exact answers for excited states, but the exponentially scaling cost again restricts its use to small molecules. In contrast, configuration interaction singles (CIS) is one of the computationally cheapest wave function techniques that can describe excited states, but it greatly overestimates CT excitation energies. While CIS correctly obtains the $\frac{1}{R}$ behavior that eludes standard TDDFT, it shares the same failure to account for orbital relaxations and predicts vertical excitation energies that are too high by an eV or more.[33] Seeking to improve upon CIS leads to methods with higher levels of correlation treatment, either other CI methods such as CIS(D),[34] or equation of motion coupled cluster (EOM-CC).[35]

EOM-CC theory is a popular wave function approach due to its effective treatment of weak correlation, which enables EOM-CCSD to provide a good combination of accuracy and affordablity in many cases.[36] Higher levels of cluster truncation can obtain very high accuracy on single valence excitations in small molecules.[37] As for more challenging classes of excited states, EOM-CCSD's inclusion of doubly excited configurations provides some ability to relax orbitals in the presence of a single charge transfer excitation though higher levels of EOM-CC may still further correct the excitation energy by several tenths of an eV.[38, 39] However, the problem of missing orbital relaxation will then recur for EOM-CCSD with double excitations and it may overestimate excitation energies by an eV or worse.[40] Including partial or even full triples can improve EOM-CC's accuracy for double excitations, but errors of half an eV may still occur for pure double excitations.[41]

The main alternative of active space approaches such as CASPT2 can obtain accurate excitation energies,[41, 42] most notably for the double excitations where coupled cluster struggles, but the same challenges related to active space choice and the limitations of perturbation theory seen in the ground state still apply. A new complication for excited states is the need to perform state-averaging where a common set of orbitals is used for multiple states. The resulting orbital shapes may be inaccurate in cases where the ground and excited state should have rather different relaxed orbitals such as in CT and core excitations.[43–45] After surveying the diversity of excited states and the individual limitations of the main current theoretical approaches, the need for more broadly reliable techniques becomes apparent.

Developing a more consistently accurate method across the different types of excited states would also aid in making comparisons between wave function methods to better understand the effects of each one's set of particular approximations.

One emerging approach for meeting this challenge is the development of state-specific electronic structure methods. The overarching philosophy is to craft techniques that are capable of providing a treatment of the correlation tailored to any particular state of interest and obtaining the orbital relaxation effects whose absence has hindered many past approaches. State-specific methods can still be organized in a hierarchical manner similar to ground state theories, beginning with an excited state mean field (ESMF) theory.[46]

$$|\Psi_{ESMF}\rangle = e^{\hat{X}} \sum_{ia} C_i^a \, |\Psi_i^a\rangle \tag{1.8}$$

The ESMF ansatz can be viewed as a CIS wave function that can acquire state-specific orbital relaxations through the minimization of a Lagrangian. The initial, albeit inefficient, implementation of ESMF used a Lagrangian of the form

$$L = \frac{\langle \Psi \,|\, (\omega - H)^2 \,|\, \Psi \rangle}{\langle \Psi \,|\, \Psi \rangle} + \vec{\mu} \cdot \nabla_p E \tag{1.9}$$

where ω is an input energy for targeting a particular state, $\nabla_p E$ is the parameter gradient of the energy, and $\vec{\mu}$ is a Lagrange multiplier. ESMF plays a role roughly analogous to Hartree-Fock's in the ground state and its effectiveness has more recently been enhanced through a self-consistent field formulation[47] and a generalization of the Lagrangian to include terms with the square of the energy gradient and other molecular properties.[48] More broadly, ESMF can serve as a platform for building state-specific versions of theories with more sophisticated treatments of correlation such as Møller-Plesset perturbation theory[49] and potentially in future coupled cluster theory.

Another notable state-specific method is a recent formulation of CASSCF.[45] This approach employs a root tracking procedure based on a combination of the variational principle $(\omega - E)^2$ and one-body reduced density matrices. The ability to optimize the CASSCF wave function for a particular excited state, which has very recently been improved to fully couple CI and orbital parameters,[50] sidesteps the limitations of state-averaging and provides an improved state-specific starting point for post-CASSCF methods that capture missing weak correlation. Part of this thesis investigates how to combine state-specific CASSCF with a more extensive correlation treatment to obtain a methodology that can fulfill our larger goal of consistently reliable descriptions of excited states.

1.4 Variational Monte Carlo

Quantum Monte Carlo (QMC) methods are a class of algorithms that can potentially fulfill this role. Many reasons for optimism can be found by surveying QMC's accomplishments in both ground and excited states. Historically, QMC approaches have earned a

reputation for obtaining highly accurate energies, beginning with early work on the electron gas[51] and continuing through numerous ground state applications to both molecules[52] and solids.[53] While the study of excited states remains less fully developed, several different QMC techniques have been applied in this setting, including Full Configuration Interaction QMC (FCIQMC),[54–56] Diffusion Monte Carlo (DMC),[57–63] and Auxiliary Field QMC (AFQMC).[64, 65] These approaches have obtained accurate excitation energies in many cases, but may face challenges due to high computational cost even by the standards of QMC or from situations where the excited state is not a symmetry ground state. Another difficulty for these projector QMC approaches is need to introduce approximations, such as the fixed node approximation in DMC,[53] to avoid the fermion sign problem where the projection obtains a lower energy bosonic solution. While the effects of these approximations can be reduced, for instance by obtaining more accurate trial wave functions for fixing the nodes in DMC, the remaining error can be difficult to characterize.

The work discussed in this thesis focuses instead on variational Monte Carlo (VMC), both in terms of its application to molecular excited states and necessary methodological improvements to enhance its success. VMC is based on a combination of the variational principle of quantum mechanics with Monte Carlo integration.[66] Working in position space, the energy for some wave function ansatz Ψ can be expressed as

$$E(\Psi) = \frac{\langle \Psi \mid H \mid \Psi \rangle}{\langle \Psi \mid \Psi \rangle} = \frac{\int d\mathbf{R} \Psi(\mathbf{R}) H \Psi(\mathbf{R})}{\int d\mathbf{R} \Psi(\mathbf{R})^2} = \frac{\int d\mathbf{R} \Psi(\mathbf{R})^2 E_L(\mathbf{R})}{\int d\mathbf{R} \Psi(\mathbf{R})^2} = \int d\mathbf{R} \rho(\mathbf{R}) E_L(\mathbf{R}) \quad (1.10)$$

The energy can then be estimated as a Monte Carlo average of the local energy values $E_L(\mathbf{R}) = \frac{H\Psi(\mathbf{R})}{\Psi(\mathbf{R})}$ determined from samples drawn from the distribution $\rho(\mathbf{R}) = \frac{\Psi(\mathbf{R})^2}{\int d\mathbf{R} \Psi(\mathbf{R})^2}$. The variational principle guarantees that for any approximate wave function, $E(\Psi)$ will be larger than the true ground state energy. In general, the functional form of Ψ will contain adjustable parameters such as the coefficients on Slater determinants that can be optimized to achieve a more accurate wave function and a lower energy.

The samples for Monte Carlo average can be obtained through the Metropolis algorithm, which generates a series of positions $\mathbf{R_i}$ where samples are taken. The steps of the algorithm are listed below:

1. Begin with a walker at a random position $\mathbf{R}$.
2. Generate a proposed position $\mathbf{R}'$ from a proposal distribution $T(\mathbf{R} \to \mathbf{R}')$.
3. Accept the proposed move with probability $A(\mathbf{R} \to \mathbf{R}') = \min(1, \frac{T(\mathbf{R}' \to \mathbf{R}) \rho(\mathbf{R}')}{T(\mathbf{R} \to \mathbf{R}') \rho(\mathbf{R})})$
4. Repeat from step 2 until all samples have been taken.

While we have described the Monte Carlo estimation of the energy and the Metropolis algorithm using the square of the wave function for the distribution $\rho(\mathbf{R})$, there are other

options. Generically, other probability distributions can be introduced by rewriting integrals as

$$I = \int f(\mathbf{R})\rho(\mathbf{R}) = \int \left(\frac{f(\mathbf{R})\rho(\mathbf{R})}{\tilde{\rho}(\mathbf{R})} \right) \tilde{\rho}(\mathbf{R}) \tag{1.11}$$

Samples can be drawn from the distribution $\tilde{\rho}(\mathbf{R})$ and use to evaluate the quantity of interest f and the ratio $\frac{\varrho}{\tilde{\rho}}$. This may be advantageous for estimation of quantities such as the energy variance, which itself has a diverging variance when $\rho(\mathbf{R})$ is used.[67, 68] A variety of modified guiding or importance sampling functions[69–71] that define a new distribution $\rho(\tilde{\mathbf{R}})$can be employed for more effective estimation, as will be seen later in this thesis.

Another nuance to stochastic estimation is the possibility of correlated sampling, which enables unbiased estimation of observables for a wave function Ψ' while using weighted samples from the distribution defined by a related reference wave function Ψ.[72] Correlated sampling can achieve smaller statistical errors on energy differences and other relative quantities than using independent sets of samples. This approach has been used for the calculation of forces and potential energy surfaces where Ψ and Ψ' are at different molecular geometries.[72, 73] For a fixed molecular geometry, correlated sampling may be used to compare values of the energy or other quantities for slightly different sets of wave function parameters.

VMC enjoys considerable flexibility in the forms of ansatzes that may be used,[53, 74–84] but in this thesis we shall mostly focus on the Multi-Slater Jastrow wave functions.

$$\Psi = \psi_J \sum_{i=0}^{N_D} c_i D_i \tag{1.12}$$

The sum of Slater determinants provide the essential antisymmetry of the electronic wave function and can be obtained from truncated CI calculations or active space methods such as CASSCF. The Jastrow factor ψ_J is an explicit function of electron positions and can capture correlation that would otherwise require the inclusion of a much larger number of Slater determinants. Jastrow factors are also able to enforce the correct cusp conditions in the wave function as electron-nuclei and electron-electron separations go to zero.[53] This particular wave function characteristic is challenging for CI and CC methods to achieve without the expense of using large basis sets and higher orders of truncation. Part of the motivation for the development of explicitly correlated R12/F12 methods[85] has been to better describe wave function cusps and obtain benefits analogous to the Jastrow's for CI and CC theory.

The ideas of VMC have been extended to excited states in both state-averaged[86–92] and state-specific[93–95] formulations, the latter of which will be a major focus of this thesis. A particular electronic state can be targeted through the use of the following functional.

$$\Omega(\Psi) = \frac{\langle \Psi \,|\, (\omega - H) \,|\, \Psi \rangle}{\langle \Psi \,|\, (\omega - H)^2 \,|\, \Psi \rangle} \tag{1.13}$$

For a chosen energy ω, minimization of Ω will target the state directly above ω allowing a wave function specifically tailored for that state to be obtained. The state-specificity enabled by this excited state variational principle allows VMC to avoid the pitfalls of state-averaging that can hamper CASSCF and CASPT2. In addition, multiconfigurational wave functions can be naturally employed in VMC, allowing for successful descriptions of doubly excited states. These wave functions can be systematically improved through the addition of more determinants and potentially higher body Jastrow factors[74, 77, 93, 96–103] or orbital optimization, which provides control over the quality of description for any state of interest. This control is essential for obtaining a balanced treatment of ground and excited state in order to achieve accurate excitation energies after cancellation of errors. These qualities of excited state-specific VMC position it to serve as a consistently reliable approach for studying different types of excited states and we explore this potential in this thesis.

At first glance, the VMC evaluation of the energy may seem relatively affordable with a formal scaling of $O(N^4)$ with system size. This scaling is quite favorable compared to methods such as coupled cluster where EOM-CCSD's cost goes as $O(N^6)$ and higher orders are even more expensive. VMC's scaling relies on efficient updates to the wave function as electrons are moved in the Metropolis algorithm. The Sherman-Morrison formula enables these updates in the case of a single Slater determinant[53, 104] and work with the table method[105, 106] has extended this efficiency to multi-determinant ansatzes. Unfortunately, the formal scaling neglects a substantial prefactor that leaves VMC in reality far slower than traditional quantum chemistry methods for small molecules. The natural parallelism of VMC with the division of sampling among compute nodes enables it to exploit the growth of highly parallel supercomputing resources, but its cost remains a challenge and motivates focusing VMC's applications on the cases that are most difficult for cheaper quantum chemistry techniques.

More broadly, method development in VMC remains an active endeavor that is crucial for realizing the full potential of this approach. Areas of work include improvements in wave function quality,[74–84, 107, 108] new sampling[67–71, 109] and regularization schemes[110, 111] to improve stochastic estimation of derivatives and other quantities, and designing new parameter optimization algorithms.[84, 112–119] This last topic of optimization is a common thread throughout all the work presented in this thesis.

1.5 Stochastic Optimization Algorithms

One particularly enduring challenge in VMC is performing the optimization of wave function parameters. The VMC literature contains a long trail of papers tracing the development of optimization methods, primarily for the ground state.[84, 112–123] The stochastic nature of VMC increases the difficulty of optimization compared to the analogous problem in deterministic quantum chemistry methods. As a result, VMC optimization algorithms are still the focus of highly active method development research, including the contributions in this thesis.

Optimization methods within VMC can be broadly divided between first-order and second-order approaches based on the level of derivative information they compute. Steepest descent and the Newton-Raphson method exemplify these two categories with the former using only a product of the current parameter gradient and a step size while the latter requires constructing and inverting the full Hessian. Equations 1.14 and 1.15 below define steepest descent and Newton-Raphson respectively for energy minimization.

$$p_i^{k+1} = p_i^k - \eta_i \frac{\partial E}{\partial p_i} \tag{1.14}$$

$$p_i^{k+1} = p_i^k - \tau_i \mathbf{h}^{-1}\mathbf{g} \tag{1.15}$$

Above, $\mathbf{h}$ is the full Hessian of partial derivatives $\frac{\partial^2 E}{\partial p_i \partial p_j}$ and $\mathbf{g}$ is the full parameter gradient. For the expense of their computation, Newton-Raphson is able to account for coupling between variational parameters, which is absent from steepest descent. One commonality between steepest descent and Newton-Raphson is the need for some step control in parameter updates, through η_i and τ_i respectively. The appropriate step size will vary with both the choice of algorithm and the type of parameter, but implementing some scheme for step control is generally essential for performing stable optimizations.

In between the extremes of steepest descent and Newton-Raphson are multiple algorithms that seek to either more effectively use first-order derivatives or more efficiently obtain second-order information to strike the right balance of optimization quality and computational affordability. Two prominent classes of methods within this spectrum are accelerated gradient descent techniques and the linear method. Accelerated descent encompasses a family of related algorithms[124] that have risen to prominence in the machine learning community and are based on retaining some memory of the objective function gradient on previous iterations in order to achieve faster convergence than steepest descent. The inclusion of gradients from earlier optimization iterations can be viewed as a way of accessing second-order curvature information at first-order cost. The linear method[113] is a more established optimization technique within VMC and has been regarded as the state-of-the-art in performance for at least a decade. It relies on a linear Taylor expansion of the wave function with respect to parameter changes and solving the generalized eigenvalue problem that results from minimizing the energy of the linear expansion. The matrices constructed by the linear method will be discussed in detail later, but can be viewed at a high level as containing some but not all of the second derivative information used in Newton-Raphson.

There are multiple considerations when selecting an optimization algorithm for use in VMC. The first is the ability to efficiently move toward the minimum of the objective function in parameter space within the available computational resources. While some algorithms such as gradient descent may have formal guarantees on convergence in the limit of infinite iterations, the minimum is generally not reached in practice for nontrivial ansatzes. An algorithm's ability to reach the minimum is also influenced by its degree of nonlinear bias due to stochastic estimation of derivatives or matrix elements. Methods with more nonlinear combinations of these stochastic estimates such as the linear method can be expected to converge

to a parameter space point offset from the true minimum.[125] We note that despite the misleading name, the linear method has significant nonlinearities, which can be recognized by observing that the high order characteristic polynomial corresponding to its eigenvalue problem grows only more nonlinear with more variational parameters. Stochastic estimation also has significant consequences for the stability of optimization algorithms. As the number of ansatz parameters increases, a fixed level of sampling effort will become insufficient to properly estimate wave function derivatives and poor parameter updates may occur, potentially leading to a complete failure of optimization. This issue can be addressed through careful algorithm design to prevent or at least reduce the incidence of poor optimization steps.

1.6 Outline

This dissertation makes several novel contributions to the use of VMC in electronic structure theory. We develop a new wave function optimization algorithm that overcomes the limitations of past first and second order approaches in VMC. We demonstrate that this new optimization method achieves superior minimization, stability, and computational efficiency compared to alternative algorithms. This optimizer is applied to both ground and excited states in a range of molecules and enables VMC to provide highly accurate excitation energy predictions across different types of excited states. The outline below details the work that will be presented in each chapter.

Chapter 2

Publication: Leon Otis and Eric Neuscamman. "Complementary First and Second Derivative Methods for Ansatz Optimization in Variational Monte Carlo." *Phys. Chem. Chem. Phys.* **21** 14491–14510 (2019).

Wave function optimization algorithms for ground-state VMC are surveyed. The performance of different flavors of accelerated descent is compared against the linear method for energy minimization in small molecule test cases. We introduce a new hybrid combination of accelerated descent and the blocked variant of the linear method. The hybrid method's superior statistical efficiency and ability to scale to larger numbers of parameters compared to the linear method are demonstrated.

Chapter 3

Publication: Leon Otis, Isaac M. Craig, and Eric Neuscamman. "A hybrid approach to excited-state-specific variational Monte Carlo and doubly excited states." *J. Chem. Phys.* **153** 234105 (2020).

The hybrid method is extended from energy-minimization to the minimization of an objective function for targeting excited states. We then apply this extension to the study

of doubly excited states and achieve good accuracy compared to benchmark results in small molecules. An application to a larger molecule, 4-aminobenzonitrile, which is too large for benchmark treatment by selected CI, and a separate discussion of optimization stability are included.

Chapter 4

Adapted from arXiv preprint: Leon Otis and Eric Neuscamman. "Combining State-Specific Quantum Chemistry and Quantum Monte Carlo for Molecular Excited States" arXiv:2111.07221

We conduct a wider benchmark survey of molecular excited states using VMC, demonstrating that VMC obtains consistently accurate excitation energies regardless of whether the excited state is single valence, doubly excited, or charge-transfer. On the methodological side, we use state-specific orbitals from a recent CASSCF approach and demonstrate we can avoid the expense of VMC orbital optimization while achieving high accuracy. The hybrid method is further refined with a scheme to reduce the number of parameters treated by the blocked linear method sections of the algorithm.

Chapter 5

We perform a systematic analysis of the optimization stability for variance-based minimization. With the test case of a cyanine dye molecule, we demonstrate that adaptive step control is crucial for stabilizing the linear method in challenging minimizations such as those requiring orbital optimization. We verify that the hybrid method remains stable in cases where the linear method fails and show that state-specific VMC obtains an excitation energy in the cyanine dye consistent with benchmark quantum chemistry and state-averaged VMC.

Chapter 6

We provide a summary of the work presented in earlier chapters, highlighting the methodological advances and application achievements of excited state-specific VMC. We close with some comments on future directions for excited-state VMC and the underlying optimization algorithms.

Chapter 2

Complementary First and Second Derivative Methods for Ansatz Optimization in Variational Monte Carlo

2.1 Abstract

We present a comparison between a number of recently introduced low-memory wave function optimization methods for variational Monte Carlo in which we find that first and second derivative methods possess strongly complementary relative advantages. While we find that low-memory variants of the linear method are vastly more efficient at bringing wave functions with disparate types of nonlinear parameters to the vicinity of the energy minimum, accelerated descent approaches are then able to locate the precise minimum with less bias and lower statistical uncertainty. By constructing a simple hybrid approach that combines these methodologies, we show that all of these advantages can be had at once when simultaneously optimizing large determinant expansions, molecular orbital shapes, traditional Jastrow correlation factors, and more nonlinear many-electron Jastrow factors.

2.2 Introduction

The practical utility of widely used methods in electronic structure theory is in large part determined by the optimization algorithms they rely on. This basic theme has been repeated throughout the history of quantum chemistry, with methods as fundamental as Hartree-Fock theory becoming dramatically more useful with the development of superior solution methods such as the direct inversion of the iterative subspace. [126] Similar transformations have been seen in configuration interaction (CI) theory thanks to Davidson's method, [127] in the density matrix renormalization group (DMRG) approach thanks to (among other

innovations) the noise algorithm, [128] and in many other methods besides. As in the case of DMRG, it is usually not so simple as a single innovation in the numerical methods that transforms a theory from a promising proof of concept into a robust computational tool. Instead, such tools often arise as the result of a series of innovations, that, once combined, fit together in a way that makes them more than the sum of their parts.

In the context of quantum Monte Carlo (QMC), and more specifically in its variational (VMC) formulation, the introductions of stochastic reconfiguration [123] (SR) and the linear method [113] (LM) for trial function optimization marked large steps forward along the path to practical utility and reliability. However, recent research has revealed multiple options for bypassing these methods' memory bottlenecks, making clear that there is still a great deal of distance to cover in the maturation of VMC numerical methods. Some of these approaches [76, 79, 129] depend, like the original SR and LM formulations, on knowing at least some information about second derivatives, but by avoiding the construction of full Hessian-sized matrices they achieve dramatically lower memory footprints. Other even more recent approaches, most of which can be classified as accelerated descent (AD) methods, [116–118] avoid second derivative information entirely and are thus even more memory efficient, relying instead on a limited knowledge of the optimization's history of energy first derivatives or in one case just the signs of these derivatives. [84] In the present study, we explore the relative advantages of the (first derivative) AD and (second derivative) LM approaches and find that, when combined, they offer a highly complementary optimization strategy that appears to be both more robust and more efficient than either approach on its own.

The ability to optimize larger and more complicated wave function forms is becoming increasingly relevant due to rapid progress in other areas of VMC methodology. The introduction of the table method [105, 106] has increased the size of CI expansions that can be handled by more than an order of magnitude, and expansion lengths beyond 10,000 determinants are no longer unusual. A recent improvement to the table method [75, 76] now allows the molecular orbital basis to be optimized efficiently in the presence of these large expansions, while the resurgence of interest in selected CI methods [9–11, 130–132] has provided a convenient route to their construction. In addition to these CI-based advances, other wave function innovations have also led to growing demands on VMC optimization methods. Increasingly sophisticated correlation factors, such as those used in Hilbert space approaches [79, 80, 116–118, 133, 134] as well as a steady stream of developments in real space [96, 97, 99, 107, 108, 123] have also raised the demand for optimization approaches that can deal with large numbers of highly nonlinear parameters. Although less thoroughly explored, the treatment of correlation effects via back flow transformations also continues to receive attention and create new optimization challenges. [83, 84] Finally, in addition to these increases in ansatz sophistication, renewed interest in using excited state variational principles [93, 135] to expand QMC's excited state capabilities has led to its own collection of optimization difficulties. [63, 69, 70, 94, 95, 136, 137]

By supporting these various advances in QMC methodology, improved VMC optimization methods have the potential for large impacts in diverse areas of chemistry and solid state physics. Work on lattice models, for example, continues to push the boundaries on

how approximate wave functions are defined. [138, 139] In the area of molecular excited states, QMC methods offer promising new routes to high-accuracy treatments of both double excitations [94, 140] and charge transfer excitations, [70, 137] both of which continue to challenge conventional quantum chemistry methods. In QMC's traditional area of simulating real solids, applications of both VMC and projector Monte Carlo would benefit immediately from the ability to prepare more sophisticated trial wave functions. [53, 141, 142] Diffusion Monte Carlo (DMC) in particular would achieve higher accuracy using the better nodal surfaces determined by well-optimized ansatzes from VMC. Solid state simulations have provided demonstrations of QMC's ability to treat up to 1000 electrons[143, 144] and improved optimization of more variational parameters will support the continued study of larger systems. More generally, the ability of QMC to combine treatments of weak and strong electron correlation effects within a robust variational framework that operates near the basis set limit makes it a powerful general-purpose approach for difficult molecular and materials problems where high accuracy is necessary. By increasing the size and complexity of systems that fall into its purview, improvements in QMC wave function optimization methods therefore have the potential to move electronic structure simulation forward on a number of fronts.

The present study seeks to aid in this endeavor by focusing on the relative advantages of recently developed low-memory first and second derivative methods in VMC and in particular on how they can be used to complement each other. Unlike deterministic optimizations, in which second derivative methods are typically preferred so long as they are affordable, the situation is less straightforward when the objective function and its derivatives are statistically uncertain. One major concern is that, in practice, it can be more difficult to achieve low-uncertainty estimates of the second derivative terms that appear in the LM and its descendants. While this issue can be mitigated by the use of alternative approaches to importance sampling, these can increase uncertainty in the energy due to the loss of the zero-variance principle. Thus, as we will demonstrate, statistical precision tends to be higher when using AD methods, which is an advantage on top of their ability to converge to the minimum without the bias that arises from the LM's highly nonlinear matrix diagonalization. However, we will also see that in order to enjoy the advantages of a tighter and less biased final convergence, AD methods must first reach the vicinity of the minimum. For this task, we find that the LM and its low-memory variants outperform all of the first derivative methods that we tested, especially for optimizations in which the wave function contains different classes of parameters that vary greatly in their nonlinear character and how they couple to each other. Happily, we will see that a hybrid approach — in which AD and low-memory LM optimization steps are interwoven — excels both at reaching the vicinity of the minimum and producing unbiased final energies while simultaneously maintaining a high degree of statistical efficiency.

2.3 Theory

Variational Monte Carlo

VMC combines the variational principle of quantum mechanics with Monte Carlo evaluation of high dimensional integrals.[66] To study the ground state of a system, we pick a trial wave function Ψ of some particular form and seek to minimize its energy expectation value.

$$E(\Psi) = \frac{\langle \Psi \mid H \mid \Psi \rangle}{\langle \Psi \mid \Psi \rangle} \tag{2.1}$$

In the language of mathematical optimization, $E(\Psi)$ is an example of an objective function or cost function. For a typical system with N electrons, this expression contains integrals over $3N$ position space coordinates which for some wave functions can only be evaluated efficiently through Monte Carlo sampling rather than quadrature methods. We rewrite the energy as

$$E = \frac{\int d\mathbf{R}\Psi(\mathbf{R})H\Psi(\mathbf{R})}{\int d\mathbf{R}\Psi(\mathbf{R})^2} = \frac{\int d\mathbf{R}\Psi(\mathbf{R})^2 E_L(\mathbf{R})}{\int d\mathbf{R}\Psi(\mathbf{R})^2} = \int d\mathbf{R}\rho(\mathbf{R})E_L(\mathbf{R}) \tag{2.2}$$

where $E_L(\mathbf{R}) = \frac{H\Psi(\mathbf{R})}{\Psi(\mathbf{R})}$ is the local energy and $\rho(\mathbf{R}) = \frac{\Psi(\mathbf{R})^2}{\int d\mathbf{R}\Psi(\mathbf{R})^2}$ is the probability density. The zero-variance principle[145] makes $\rho(\mathbf{R})$ the most common choice of probability distribution for obtaining samples, but it is not the only option. For effective estimation of quantities beside the energy, such as the LM matrix elements, other importance sampling functions are often preferred. [67–69]

In our LM and blocked LM calculations in this study, we employ the importance sampling function (and the appropriately modified statistical estimate formulas [70])

$$|\Phi|^2 \equiv |\Psi|^2 + \frac{\epsilon}{M}\sum_I |D_I|^2 \tag{2.3}$$

in which the D_I are the M different S_z-conserving single excitations relative to the closed shell reference determinant. The logic behind this choice is that it puts some weight on configurations that are highly relevant for the orbital rotation parameters' wave function derivatives, as small orbital rotations can be approximated via the addition of singles. We find that this importance sampling function substantially reduces the uncertainty of the LM matrix elements corresponding to orbital rotations, which in turn helps reduce the update step uncertainty. For AD, we simply use traditional $|\Psi|^2$ importance sampling as in equation 2.2.

By the variational principle, we are guaranteed that E is an upper bound on the true ground state energy. Given some set of adjustable parameters in the functional form of Ψ, we expect that values of those parameters that yield a lower value of E to correspond to a wave function that is closer to the ground state. One could then imagine the abstract space produced by the possible values of all variational parameters. The set of optimal parameter

values that specify the wave function expression which minimizes E can be taken as a point in this space labeled by the vector $\mathbf{p}^*$. In general, the initial choice for parameters will not be at this energy minimum point, but at some other point $\mathbf{p_0}$. The problem of determining the best wave function in VMC calculations then relies on an optimization algorithm for finding $\mathbf{p}^*$ after starting from $\mathbf{p_0}$.

Within this framework, one of the most important considerations is that the optimization is inherently stochastic due to the introduction of noise through the Monte Carlo evaluation of the integral in equation 2.2. This forms a contrast with many other methods in electronic structure theory including Hartree-Fock, CI, and coupled cluster where various deterministic optimization schemes predominate.[1] Many of the algorithms commonly encountered in a deterministic quantum chemistry context such as steepest descent and the Newton-Raphson method, have been adapted for use in VMC.[121, 122, 146–148] However, there is now a need to be robust to the presence of noise. Historically, errors due to finite sampling led to numerical instabilities that prompted interest in minimizing variance[93, 149] instead of energy, but later optimization developments have sought to mitigate this issue and in this paper we only consider energy minimization. As we will now discuss in their respective sections, both the LM and gradient descent approaches possess features that enable them to operate stably in a stochastic setting.

The Linear Method

The LM[112, 113] begins with a first order Taylor expansion of the wave function. For a set of variational parameters given by vector $\mathbf{p}$, we have

$$\Psi(\mathbf{p}) = \Psi_0 + \sum_i \Delta p_i \Psi_i \tag{2.4}$$

where $\Psi_i = \frac{\partial \Psi(\mathbf{p})}{\partial p_i}$ and Ψ_0 is the wave function at the current parameter values. Note that $\Psi(\mathbf{p}) = \Psi(\mathbf{R}, \mathbf{p})$ depends on both the parameters $\mathbf{p}$ and electron positions $\mathbf{R}$, but we have suppressed the latter for convenience.

Finding the optimal changes to the parameters amounts to solving the generalized eigenvalue problem

$$\mathbf{H}\,\mathbf{c} = E\,\mathbf{S}\,\mathbf{c} \tag{2.5}$$

in the basis of the initial wave function Ψ_0 and its first order parameter derivatives $\{\Psi_1, \Psi_2, ...\}$. H and S are the Hamiltonian and overlap matrices in this basis with elements

$$H_{ij} = \langle \Psi_i \,|\, H \,|\, \Psi_j \rangle \tag{2.6}$$

$$S_{ij} = \langle \Psi_i \,|\, \Psi_j \rangle \tag{2.7}$$

The matrix diagonalization to solve this eigenproblem for eigenvector $\mathbf{c} = (1, \Delta\mathbf{p})$ then yields the updated parameter values $\mathbf{p_1} = \mathbf{p_0} + \Delta\mathbf{p}$. As the matrices $\mathbf{H}$ and $\mathbf{S}$ both contain a subset of the second derivative terms that would be present in a Newton-Raphson approach, [114]

the LM is most naturally categorized as a second-derivative method, and it certainly shares Newton-Raphson's difficulties with regards to dealing with matrices whose dimension grows as the number of variables.

For practical use with finite sampling, the LM must be stabilized to prevent unwisely large steps in parameter space. This is accomplished by adding shift values[113] to the matrix diagonal that effectively act as a trust radius scheme similar to those used with Newton-Raphson. In our implementation, the Hamiltonian is modified with two shift values meant to address distinct potential problems in the optimization.[150]

$$\mathbf{H} \rightarrow \mathbf{H} + c_I \mathbf{A} + c_S \mathbf{B} \tag{2.8}$$

The matrix elements of $\mathbf{A}$ are given by $A_{ij} = \delta_{ij}(1 - \delta_{i0})$ so that the shift c_I effectively gives an energy penalty to directions of change from the current wave function.[113] The second shift is intended to address problems that may arise if some wave function derivatives have norms that differ by orders of magnitude. In this situation, the single shift value c_I is insufficient to preserve a quick yet stable optimization. For a parameter with a large derivative norm, a sufficiently high value of c_I might prevent an excessively large change in its value. However, all other parameter directions with smaller derivative norms will be so heavily penalized by the large value of c_I that those parameters become effectively fixed. The purpose of the second $c_S\mathbf{B}$ term is to retain important flexibility in other parameter directions. We can write the matrix $\mathbf{B}$ as

$$\mathbf{B} = (\mathbf{Q^T})^{-1}\mathbf{T}\mathbf{Q}^{-1} \tag{2.9}$$

where

$$Q_{ij} = \delta_{ij} - \delta_{i0}(1 - \delta_{j0})S_{0j} \tag{2.10}$$

and

$$T_{ij} = (1 - \delta_{i0}\delta_{j0})[\mathbf{Q^T}\mathbf{S}\mathbf{Q}]_{ij} \tag{2.11}$$

The matrix $\mathbf{Q}$ provides a transformation to a basis where all update directions are orthogonal to the current wave function and the matrix $\mathbf{T}$ is the overlap matrix in this basis. The optimal choice of shift parameters c_I and c_S may depend on the particular optimization problem. In our implementation, an adaptive scheme adjusts the shifts on each iteration by comparing the energies calculated through correlated sampling on three different sets of shift values and choosing whichever shifts produced the lowest energy.

The LM has been successfully applied to a variety of systems to prepare good trial wave functions for DMC. [107, 108, 113–115, 151, 152] It has also been used in the variational optimization of a recent functional for targeting excited states.[70, 94, 95] However, it possesses a number of limitations, most notably a memory cost that scales with the square of the number of optimizable parameters due to the matrices it builds. Once these matrices are too large for storage in cache, the fact that each sample contributes to every matrix element dramatically slows the process of matrix construction. At present, routine use of the LM is limited to roughly 10,000 parameters, although exceptional calculations with up

to about 16,000 have been made.[105] Another shortcoming is the nonlinear bias of the LM. We are evaluating the elements of the Hamiltonian and overlap matrices stochastically and have a nonlinear relationship between them and our energy through the generally high order characteristic polynomial of the eigenvalue problem of equation 2.5. As a result, we in general expect the LM to converge to a point in parameter space slightly offset from the true minimum. This nonlinear bias has been studied for the LM in Hilbert space[125] and a similar issue arises in the context of Full Configuration Interaction QMC.[153] Both the memory constraint and the nonlinear bias of the LM become more severe for ansatzes with larger numbers of variational parameters, which spurs the search for potential alternatives. One approach suggested for memory reduction is to employ Krylov subspace methods for Eq. 2.5 to avoid building matrices, but it requires a drastically higher sampling effort due to the need for many matrix-vector multiplications and so we do not pursue the approach here.[79]

Blocked Linear Method

One recent approach to bypassing the memory bottleneck is known as the blocked linear method (BLM).[129] The first step of the algorithm is to divide the full set of parameters into N_b blocks. Next, a LM-style matrix diagonalization is carried out within each block and some number N_k of the resulting eigenvectors from the blocks are retained as good directions for constructing an approximation for the overall best update direction in the full parameter space. For a particular block of variables, the wave function expansion in the LM is given by

$$|\Psi_b\rangle = |\Psi_0\rangle + \sum_{i=1}^{M_b} c_i |\Psi^i\rangle \tag{2.12}$$

where $|\Psi^i\rangle$ is the wave function derivative with respect to the ith variable in the block, M_b is the number of variables in the block, and $|\Psi_0\rangle$ the current wave function as in the normal LM. We can perform the same matrix diagonalization done in the LM, only with parameters outside the block fixed. This yields a set of eigenvectors that we can use to construct another approximate expansion of the original wave function. We can construct a matrix $\mathbf{B}$ using the N_k eigenvectors with the lowest eigenvalues from each block and write a new expansion

$$|\tilde{\Psi}\rangle = \alpha |\Psi_0\rangle + \sum_{k=1}^{N_b} \sum_{j=1}^{N_k} A_{kj} \sum_{i=1}^{M_b} B_{ji}^{(b)} |\Psi^{i,b}\rangle \tag{2.13}$$

Having now pre-identified important directions within each block, the idea is that a subsequent LM-style diagonalization in the basis of these good directions (which yields the coefficients A_{kj}) should still provide a good update direction when re-expressed in the full parameter space.

In order to help retain most of the accuracy of the traditional LM, the first stage of the BLM computation includes N_o other good directions that are used to supply the current

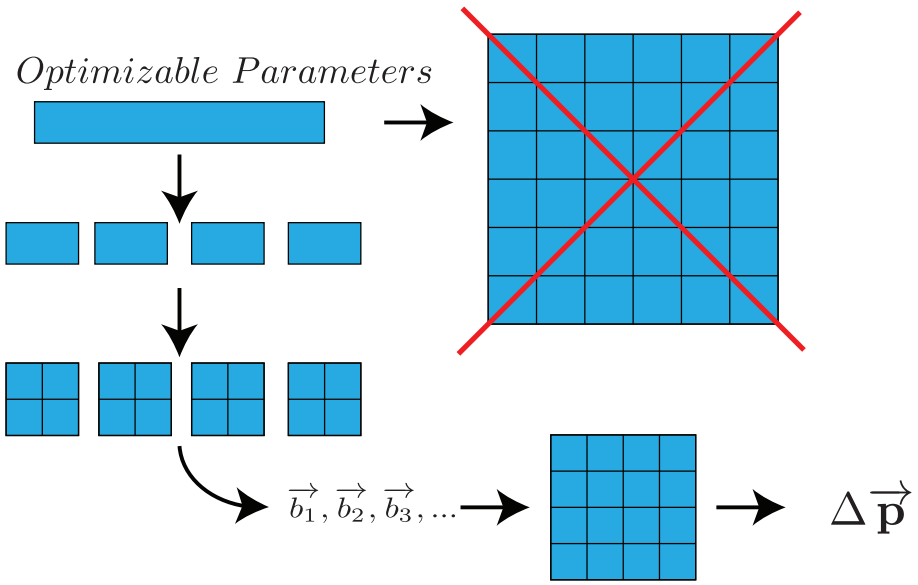


Figure 2.1: Flowchart depicting steps in the BLM algorithm to arrive at a parameter update.

block's diagonalization with information about how its variables are likely to couple to those in other blocks. In practice, important out-of-block directions are obtained by keeping a history of previous iterations' updates as the optimization progresses. We can rewrite the one block expansion introduced in equation 2.12 as

$$|\Psi_b\rangle = |\Psi_0\rangle + \sum_{i=1}^{M_b} c_i |\Psi^i\rangle + \sum_{j=1}^{N_o} \sum_{k=1,k\neq b}^{N_b} d_{jk} |\Theta_{jk}\rangle \tag{2.14}$$

where we take the

$$|\Theta_{jk}\rangle = \sum_{l=1}^{M_k} C_{jkl} |\Psi^{l,k}\rangle \tag{2.15}$$

as the linear combinations of wave function derivatives from other blocks that were identified as important based on previous iterations' updates. The additional term in the expansion allows us to account for couplings between variables in different blocks and enable the construction of a better space for the second diagonalization. We assemble the matrix $\mathbf{B}$ and $|\tilde{\Psi}\rangle$ and then seek to minimize $\frac{\langle \tilde{\Psi} | H | \tilde{\Psi} \rangle}{\langle \tilde{\Psi} | \tilde{\Psi} \rangle}$ with respect to variational parameters α and A_{kj} in our BLM wave function expansion in equation 2.13.

Figure 2.1 portrays the algorithmic steps described above. Some number of parameters too large to be handled by the standard LM is divided among different blocks whose diagonalizations produce the vectors $\vec{b_i}$ for the construction of the space of the final diagonalization that produces the parameter update. The BLM can be thought of as achieving

memory savings in the use of smaller matrices at the cost of having to run over the sample twice when the traditional LM must run over it just once. A more extensive description of the BLM and its precise memory usage can be found in its original paper.[129]

We divide parameters evenly among blocks, but one could implement the use of tailored blocks of varying sizes. It is advisable to choose the block size to be large enough to keep important parameters of the same type, such as all of those for a Jastrow factor, within the same block. This enables the expected strong coupling between them to be handled more accurately by the LM-style diagonalization within that block. While the BLM has been successfully applied up to about 25,000 parameters and found to closely reproduce the results of the standard LM, [129] it remains a relatively new method, and the present study will provide additional data on its efficacy.

Gradient Descent Methods

In the last few years, increasing attention[84, 116–118, 154] has been paid to optimization methods that use only first derivatives to optimize trial wave functions in VMC. One formulation for discussing these methods is to consider minimizing a Lagrangian of the form

$$\mathcal{L}(\Psi(\mathbf{p})) = \langle \Psi \mid H \mid \Psi \rangle - \mu(\langle \Psi \mid \Psi \rangle - 1) \tag{2.16}$$

where μ is a Lagrange multiplier and, in practice, a moving average of the local energy. There is no need to solve an eigenvalue problem as in the LM and the memory cost of these approaches scales linearly with the number of parameters. We also note that the stochastic evaluation of derivatives of this Lagrangian will lead to a smaller nonlinear bias compared to what is encountered in the LM. While there is some nonlinearity present in the product $\mu \langle \Psi \mid \Psi \rangle$, it is mild compared to the high order polynomials encountered in the solution of the LM eigenvalue problem and can be avoided entirely if desired through modest amounts of extra sampling. Minimization of this Lagrangian targets the ground state, but excited states can similarly be targeted with these optimization algorithms merely by using derivatives of one of the excited state functionals that have been developed.[93, 94, 135, 155]

The simplest method in this category is the steepest descent algorithm.

$$p_i^{k+1} = p_i^k - \eta_k \frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_i} \tag{2.17}$$

In this case, the value of each parameter on the $k+1$'th step is found simply by subtracting the statistically uncertain parameter derivative times a step size η_k. The step size can be taken as constant over all steps in the simplest case, but rigorous proofs on the convergence of stochastic gradient descent (SGD) rely on decaying step sizes satisfying $\sum_k \eta_k = \infty$ and $\sum_k \eta_k^2 < \infty$.[156]

It may be worth briefly commenting that the typical formulation of stochastic gradient descent as seen in the machine learning and mathematical optimization literature is slightly different from what we use here within VMC. In a common machine learning scenario,[156]

one has a training set of input data $\{x_1, x_2, ..., x_n\}$ and corresponding outputs $\{y_1, y_2, ..., y_n\}$ and wishes to minimize a loss function $Q(x, y; w)$ that measures the error produced by a model $f_w(x)$, which predicts $\widetilde{y}_i$ given x_i and is parameterized by variables w. For this setting, the SGD algorithm refers to evaluating the gradient of Q with a randomly chosen pair (x_j, y_j) from the given data set and then computing the parameter update according to $w_{k+1} = w_k - \eta_k \nabla_w Q(x_j, y_j)$. For our VMC optimization, we are dealing with a noisy gradient similar to what occurs in this machine learning problem, but the source of our noise is somewhat different and lies in our means of evaluating the underlying 3N dimensional integrals within our Lagrangian derivatives. Another important distinction is that in machine learning applications, complete convergence to the minimum is in fact undesirable because it will overfit the model to the training data and degrade its performance on new sets of test inputs. Much as SGD provides a computational speed up for machine learning problems, we are also able to operate gradient descent methods at a cheap per-iteration cost because we need only a modest number of samples to evaluate sufficiently precise Lagrangian derivatives compared to the Hamiltonian and overlap matrices in the LM. However, unlike the machine learning case, we do want to come as close as possible to the true minimum, and we will see that even reaching the vicinity of the minimum can be difficult for descent methods when typical VMC initial guesses are employed.

While steepest descent can be guaranteed to eventually reach the minimum of the Lagrangian even in a stochastic setting, its asymptotic convergence is very slow. For some intuition, one could imagine the landscape of the Lagrangian's values forming a very narrow valley near the true minimum. In this situation, steepest descent would produce parameter updates mostly back and forth along the sides of the valley with little improvement of parameter values in the direction directly toward the minimum. Due to the limitations of steepest descent, a number of other flavors of accelerated gradient descent (AD) have been developed that include a momentum term with information on previous values of the gradient. As illustrated in Figure 2.2, the general intuition is that this additional term provides some memory of the progression along narrow valleys that steepest descent lacks and thereby achieves swifter convergence. In addition, there are multiple schemes for adaptively varying the step sizes used in a manner that draws on the particular derivative values for each individual parameter as the optimization progresses. These methods have recently been applied successfully to Hilbert space QMC. In this study, we work in real space and investigate a combination of Nesterov momentum with RMSprop as presented by the Booth group [116, 154], a method using random step sizes from the Clark group[84], AMSGrad, recently used by the Sharma group[117, 118], as well as the ADAM optimizer[157].

We now lay out the precise expressions for each of these methods in turn. The RMSprop algorithm used by Booth and co-workers is given by the following recurrence relations.[116]

$$p_i^{k+1} = (1 - \gamma_k) q_i^{k+1} - \gamma_k q_i^k \tag{2.18}$$

$$q_i^{k+1} = p_i^k - \tau_k \frac{\partial \mathcal{L}(\mathbf{p})}{\partial p_i} \tag{2.19}$$

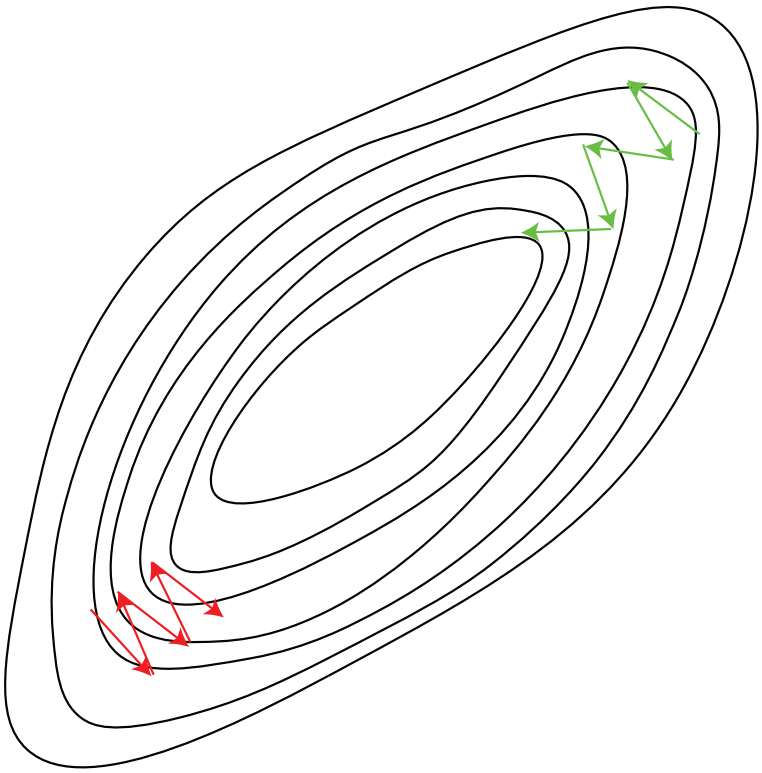


Figure 2.2: Illustration of the difficulty faced by steepest descent in red on the lower left with its slow approach to the minimum. Accelerated descent in green on the upper right is able to progress more rapidly to the minimum with its memory of previous gradients.

$$\lambda_0 = 0 \qquad \lambda_k = \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\lambda_{k-1}^2} \qquad \gamma_k = \frac{1 - \lambda_k}{\lambda_{k+1}} \tag{2.20}$$

$$\tau_k = \frac{\eta}{\sqrt{E[(\frac{\partial \mathcal{L}}{\partial p_i})^2]^{(k)}} + \epsilon} \tag{2.21}$$

$$E[(\partial \mathcal{L})^2]^{(k)} = \rho E\left[\left(\frac{\partial \mathcal{L}}{\partial p_i}\right)^2\right]^{(k-1)} + (1-\rho)\left(\frac{\partial \mathcal{L}}{\partial p_i}\right)^2 \tag{2.22}$$

Above, p_i^k denotes the value of the ith parameter on the kth step of the optimization, τ_k is a step size that is adaptively adjusted according to the RMSprop algorithm in equations 2.21 and 2.22. The running average of the square of parameter derivatives in the denominator of τ_k allows for the step size to decrease when the derivative is large, which should hedge against the possibility of taking excessively large steps. Conversely, a smaller denominator when the derivative is small allows for larger steps to be taken. The weighting in the running average is controlled by a factor ρ that can be thought of as the amount of memory retained of past

gradients for adjusting τ_k, and η again denotes the chosen initial step size. In order to avoid possible singularities when the gradient is very close to zero, a small positive number ϵ is included in the denominator of τ_k. Equation 2.18 shows the momentum effect in which the update for the parameter on the $k+1$ step depends on the update from the previous step as well as the current gradient. We also follow the Booth group in applying a damping factor to the momentum by replacing γ_k with $\gamma_k e^{-(\frac{1}{d})(k-1)}$. The quantity d effectively controls how quickly the momentum is turned off, which eventually turns the algorithm into SGD. The values of d,η, ρ, and ϵ may all be chosen by the user of the algorithm and are known as hyperparameters in the machine learning literature. In the results we present using this method, we have used $d = 100$, $\rho = .9$ and $\epsilon = 10^{-8}$. We have found adjusting these hyperparameters has relatively little influence on optimization performance compared to choices for step size η, but their influence could be explored more systematically.

The Clark group's algorithm takes a far simpler form

$$p_i^{k+1} = p_i^k - \alpha\eta \frac{\left|\frac{\partial \mathcal{L}}{\partial p_i^k}\right|}{\frac{\partial \mathcal{L}}{\partial p_i^k}} \tag{2.23}$$

and has been recently used with neural network wave functions in the context of the Hubbard model.[84] Here α is a random number in the interval $(0, 1)$ and η sets the overall scale of the random step size. The motivation for allowing the step size to be random is that it may help the optimization escape local minima that it encounters. Within VMC, this algorithm can be run with fewer samples per iteration even compared to other gradient descent based algorithms as only the sign of the derivative needs to be known, but it typically requires many more iterations to converge.

ADAM and AMSGrad are popular methods within the machine learning community[124, 157, 158] and have similar forms. ADAM is given by:

$$p_i^{k+1} = p_i^k - \eta \frac{m_i^k}{\sqrt{n_i^k}} \tag{2.24}$$

$$m_i^k = (1-\beta_1) m_i^{k-1} + \beta_1 \frac{\partial \mathcal{L}}{\partial p_i^k} \tag{2.25}$$

$$n_i^k = \beta_2\, n_i^{k-1} + (1-\beta_2) \left(\frac{\partial \mathcal{L}}{\partial p_i^k}\right)^2 \tag{2.26}$$

AMSGrad is a recent adaptive step size scheme developed in response to the limitations of ADAM [158] and has almost the same form except for a slightly different denominator.

$$n_i^k = \max\left(n_i^{k-1},\ \ (1-\beta_2)\, n_i^{k-1} + \beta_2 \left(\frac{\partial \mathcal{L}}{\partial p_i^k}\right)^2\right) \tag{2.27}$$

In our calculations, we have used $\beta_1 = 0.1$ and $\beta_2 = 0.01$ for both AMSGrad and ADAM in line with the choice made by the Sharma group.[117, 118] It may be worth noting that a

different convention appears in machine learning literature using $1-\beta_1$ and $1-\beta_2$ for what we and the Sharma group call β_1 and β_2. [124, 158]

Compared to the LM, these first derivative descent methods have some significant advantages. Their low memory usage and reduced nonlinear bias make them a natural fit for the large parameter sets that the LM struggles to handle. They are remarkably robust in the presence of noise and do not need special safeguards against statistical instabilities such as the LM's shifts. At a basic practical level, the descent methods are also far simpler to implement than the LM and especially its blocked variant. However, as we will see in our results, they often struggle to reach the vicinity of the minimum using a comparable sampling effort.

A Hybrid Optimization Method

In an attempt to retain the benefits of both the LM and the AD techniques, we have developed a hybrid optimization scheme that can be applied to large numbers of parameters. Our approach alternates between periods of optimization using AD and sections using the BLM. Among other advantages, this allows us to use gradient descent to identify the N_o previous important directions in parameter space that are used in the BLM via equation 2.14. The precise mixture of both methods can be flexibly altered, but a concrete example would be to first optimize for 100 iterations using RMSprop. By storing a vector of parameter value differences every 20 iterations, we would produce 5 vectors that can be used for equation 2.14 in some number (say three) steps of the BLM. After the execution of these BLM steps, the algorithm would return to another 100 iterations of descent and the process repeats until the minimum is reached. Figure 2.3 shows a generic depiction of how the ground state energy optimization may behave over the course of the hybrid method. There are extended sections of computationally cheap optimization using gradient descent interwoven with substantial energy improvement over a few BLM steps.

The use of AD and the BLM should naturally allow parameter sets beyond the traditional LM limit of about 10,000 variables to be addressed, a limit we will surpass in the present study in the diflurodiazene system. For now, Table 2.1 lays out how the memory cost of the methods we are considering scales with number of parameters N. Both the hybrid method and the BLM steps it contains have a memory scaling that is intermediate between that of the standard LM and the descent methods. The cost is given only approximately because while it is normally dominated by the cost of the N_b blocks in the BLM, there are additional contributions related to how many directions are retained from the first BLM diagonalization and how many old directions are used.[129]

One key motivation for including sections of AD, especially when the method is near convergence, is to counteract the noise we observe in LM updates. While the LM tends to converge in a relatively small number of steps, we find the individual energies still fluctuate from iteration to iteration by multiple mE_h, particularly when we are working with wave functions that possess many highly nonlinear parameters. Figure 2.4 shows a cartoon of this behavior near the minimum that prevents tight convergence. Unless the shifts are large

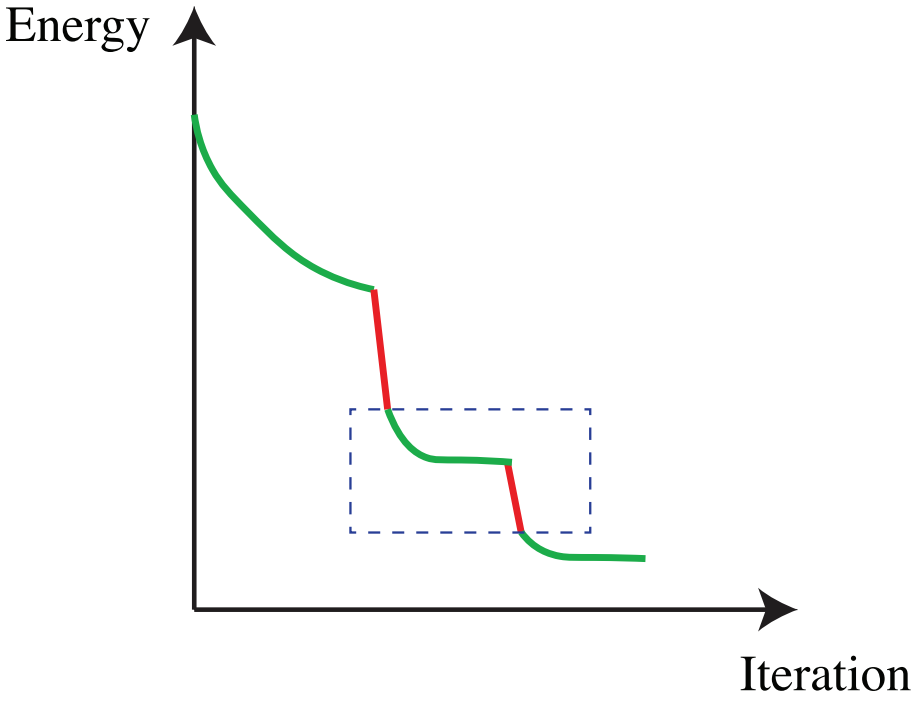


Figure 2.3: Schematic depiction of a typical energy optimization using the hybrid method. The dashed box around a section of descent in green and BLM in red defines a macro-iteration of the method.

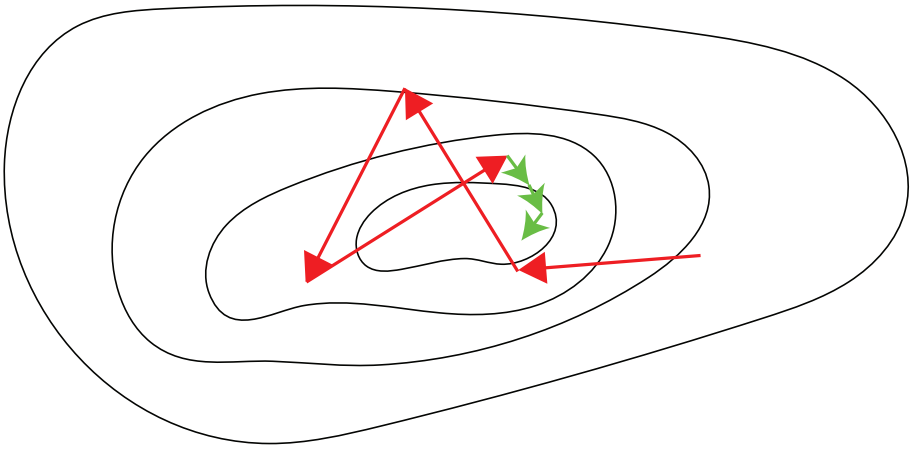


Figure 2.4: Schematic representation of gradient descent corrections in green to red BLM steps, which we have observed to reduce the uncertainty about the location of the final minimum.

Table 2.1: Rough memory cost scaling for the optimization methods we examine, with N the number of optimized parameters and N_b the number of blocks.

Method Type	Memory Cost
Standard Linear Method	$O(N^2)$
Blocked Linear Method	$\sim O\left(\frac{N^2}{N_b}\right)$
Hybrid Method	$\sim O\left(\frac{N^2}{N_b}\right)$
Descent Methods	$O(N)$

enough to constrain it to very small steps, the LM will tend to bounce around near the true minimum due to substantial (and biased) statistical uncertainties in its step direction. The resulting energy fluctuations lead to ambiguity in what to report as the definitive LM energy. One could take the absolute lowest energy reached on any iteration, but this is fairly unsatisfactory as it feels too dependent on a "lucky" step landing right on the minimum. Our practice has been to take an average over multiple steps at the end of the optimization when parameter values should be converged. However, this will generally include iterations with upward energy deviations due to the step uncertainties. The use of AD offers a way out of this dilemma because it can correct the errors in the LM steps by moving towards the minimum more smoothly and with less bias. As we shall demonstrate in our results, these considerations seem to give the hybrid method a statistical advantage over the LM by achieving lower error bars for the same computational cost. They are also the basis of our recommendation for finishing optimizations with a long section of pure AD, which we shall show tends to improve the energy and greatly diminish the final statistical uncertainty.

Wave Functions

An assessment of optimization methods' effectiveness requires consideration of the form of the wave function that they are applied to. Multi-Slater determinant wave functions have been a common choice of ansatz in QMC and are typically combined with Jastrow factors that help recover some electron correlation and describe particle cusps.[53] We specify our Multi-Slater Jastrow (MSJ) wave function with the following set of equations.

$$\Psi = \psi_{MS}\psi_J\psi_C \tag{2.28}$$

$$\psi_{MS} = \sum_{i=0}^{N_D} c_i D_i \tag{2.29}$$

$$\psi_J = \exp\left\{ \sum_i \sum_j \chi_k(|r_i - R_j|) + \sum_k \sum_{l>k} u_{kl}(|r_k - r_l|) \right\} \tag{2.30}$$

$$\psi_C = \exp\left(\sum_{IJ} F_{IJ} N_I N_J + \sum_K G_K N_K \right) \tag{2.31}$$

In equation 29 above, ψ_{MS} consists of N_D Slater determinants D_i with coefficients c_i. It can be generated by some other quantum chemistry calculation such as complete active space self-consistent field (CASSCF) or a selective CI method prior to the VMC optimization. In the one- and two-body Jastrow factor ψ_J, we have functions χ_k and u_{kl}, which are constructed from optimizable splines whose form is constrained so as to enforce any relevant electron-electron and electron-nuclear cusp conditions. [150]

While MSJ wave functions with these types of traditional Jastrow factors (TJFs) have been successfully used in many contexts, [53, 70, 76, 105, 113] more involved correlation factors can be considered. Typically, this involves the construction of many-body Jastrows factors, [74, 93, 98] which may involve various polynomials of interparticle distances [96–98] or an expansion in an atomic orbital basis [74, 99–103, 159] or a set of local counting functions. [107, 108] The latter case of many-body Jastrows, known as real space number-counting Jastrow factors (NCJF), is employed here as an example many-body Jastrow factor. In real space, Jastrow factors have historically been effective at encoding small changes to the wave function associated with weak correlation effects, [53] but work in Hilbert space and lattice model VMC reminds us that they can also be used to aid in the recovery of strong correlations. [140, 160, 161] One way to view NCJFs is as an attempt to develop a real space many-body Jastrow factor that can aid in recovering both strong and weak electron correlations. [108]

The form of our NCJFs in equation 2.31 has the same structure as previously proposed four-body Jastrow factors,[102] where N_I denotes the population of a region and the F_{IJ} and G_K are linear coefficients. The region populations are computed by summing the values of counting functions at each electron coordinate.

$$N_I = \sum_i C_I(\mathbf{r_i}) \tag{2.32}$$

In this work, we use a recently introduced [108] form for the counting functions consisting of normalized Gaussians.

$$C_I = \frac{g_I(\mathbf{r})}{\sum_j g_j(\mathbf{r})} \tag{2.33}$$

where

$$g_j(\mathbf{r}) = \exp\big((\mathbf{r} - \mu)^T \mathbf{A} (\mathbf{r} - \mu) + K\big) \tag{2.34}$$

describes a Gaussian about a center μ. By placing these normalized Gaussians at various centers, we can divide up space with a Voronoi tessellation. Schemes have been developed to generate partitions that either consist of regions centered on atoms or of finer grained

divisions of space that can capture correlation within an atomic shell. We make use of both types of partitioning methods for different wave functions in our study. For simplicity, we only consider optimization of the parameters F_{IJ} in the F-matrix of our NCJFs (the coefficients G_K can be eliminated with a basis transformation of the region populations N_I), [108] but in principle the parameters defining the Gaussians g_j could also be optimized. We provide details of the Gaussians used in our ansatzes in Appendix D.

We also consider the problem of optimizing the molecular orbital shapes alongside the other variational parameters. The ability to relax orbitals is important for successful study of many systems, particularly those involving excited state phenomena.[70] We make use of considerable theoretical and computational machinery based on the table method enhancements developed by Filippi and coworkers [75, 76] that enables efficient evaluation of orbital rotation derivatives in large MSJ wave functions. A rotation of molecular orbitals can be described with a unitary transformation with matrix $\mathbf{U}$ parameterized as the exponential of an antisymmetric matrix $\mathbf{X} = -\mathbf{X}^{\mathbf{T}}$

$$\mathbf{U} = \exp(\mathbf{X}) \tag{2.35}$$

Impressively, one can obtain all wave function derivatives with respect to the elements of $\mathbf{X}$ for a large multi-Slater determinant ansatz for a cost that is only slightly higher than that of the local energy evaluation. For the details of how this is accomplished, we refer the reader to the original publications. [75, 76] From the standpoint of parameter optimization, the main significance of the orbitals (and the NCJFs) lies in both their nonlinearity and their strong coupling to other optimizable parameters. In practice, we find that turning on the optimization of orbitals and NCJFs greatly enhances the difficulty of the optimization problem compared to MSJ optimizations in which only the CI coefficients and one- and two-body Jastrow parameters are varied.

2.4 Results

Multi-Slater Jastrow N_2

For a small initial test system, we consider the nitrogen dimer N_2 at the near-equilibrium and stretched bond lengths of 1.1 and 1.8 Å. The nitrogen dimer is a known example of a strongly correlated system and a common testing ground for quantum chemistry methods.[140, 161–166] The initial wave function ansatz consists of a modest number of Slater determinants (67 for the equilibrium geometry and 169 for the stretched, the result of a 0.01 cutoff limit on determinant coefficients) with traditional one-body and two-body Jastrow factors. The Jastrow splines provide 30 additional optimizable parameters via 10 point cubic b-splines with cutoff distances of 10 bohr for the electron-nuclear and same-spin and opposite-spin electron-electron components. The Slater determinant expansion is the result of a (10e,12o) CASSCF calculation in GAMESS[167] using BFD pseudopotentials and the corresponding VTZ basis set.[168] Due to the simplicity of the variable space in this case, we have employed

the $|\Psi|^2$ guiding function for all optimization methods, including the LM and BLM. See Appendix B for further computational details.

The first and simplest study we can make is to optimize our ansatzes with our multiple optimization techniques until convergence and compare final energies. In our results for N_2 and other systems, we focus on the absolute energies achieved by different optimization algorithms. Studies intending to provide chemical insight would of course require differences of computed energies and additional techniques may be needed to obtain balanced energies for accurate differences.[69, 70, 169] Better optimization algorithms are able to support such work by reducing other errors due to incomplete improvement of the wave function. Note that all of our VMC optimizations with different methods in this study were performed using our implementations within a development version of the QMCPACK software package.[150] As N_2 is a small enough system that the traditional LM can be easily employed, we take the approach of first obtaining a traditional LM optimization result and then using it as a reference against which to compare the performance of other methods. For the gradient descent methods, multiple optimizations were attempted with the initial step sizes tweaked from run to run based on a rough examination of how parameter values compared to the LM's results. We find that the chosen values for the step sizes and other hyperparameters in the gradient descent algorithms often leads to apparent convergence at different energies. It is therefore essential to make effective choices for these parameters, which in part seems to rely on one's experience with a given system.

Figures 2.5 and 2.6 show energy differences relative to the LM result when optimizing the equilibrium and stretched nitrogen dimer wave functions respectively. Tables providing the precise energies and statistical uncertainties as well as the step sizes used are shown in the appendices. First, we see the choice of step sizes can have a substantial influence on the quality of gradient descent results. In some cases, the same method can appear to converge to energies more than 20 mE_h apart when run with different initial step sizes. While many of the gradient descent optimizations clearly did not reach the minimum, the energy differences from the LM are only about 5 mE_h or less when looking at the runs that used what turned out to be the best choices for the hyperparameters. With further tweaking of the hyperparameters, we would guess that at least some of these descent methods could match the performance of the LM in this simple test case. Finally, we observe that the hybrid method performs about as well as the best descent optimizations, typically reaching energies that agree with the LM within error bars.

All parameter N_2

We now add a NCJF and enable orbital optimization in order to extend the comparison in a setting with a larger number and variety of nonlinear parameters. We will consider the relative merits of the optimization methods in much greater detail in this setting as it offers a clearer view of their differences. For the number-counting Jastrow factor, we generated a set of 16 counting regions with 8 octants per atom after dividing space in half with a plane bisecting the bond axis. The details are given in Appendix D, but we will

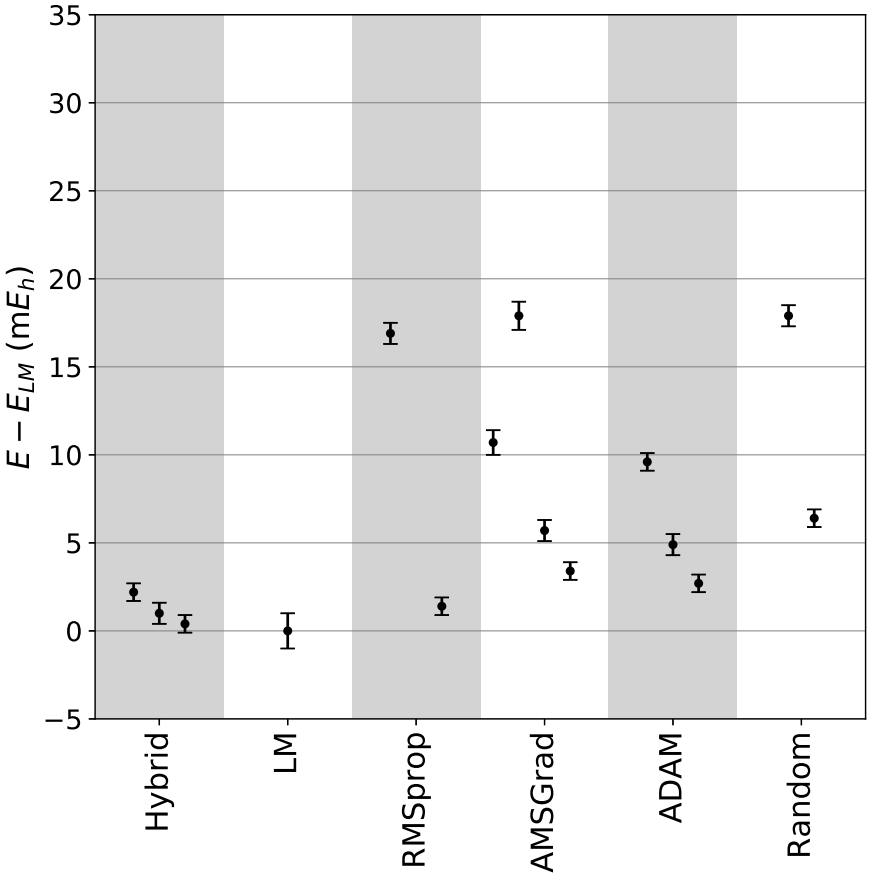


Figure 2.5: Different methods' optimized energies relative to that of the LM for equilibrium N_2 when optimizing CI coefficients and the TJF.

note here that this adds 135 F-matrix parameters to the optimization. Allowing for orbital optimization adds another 663 and 618 parameters for the equilibrium and stretched cases, respectively. Note that our implementation of orbital optimization in QMCPACK removes rotation parameters for orbitals that are not occupied in any determinant and also between orbitals occupied in all determinants, and so the precise number of rotation parameters is a function of the determinant expansion. With orbital optimization enabled, the choice of importance sampling function becomes an issue, and we now employ $|\Phi|^2$ for all LM and BLM steps with the ϵ weight set to 0.001.

Figures 2.7 and 2.8 show converged ground state energies relative to that of the LM. For this more difficult version of the nitrogen dimer, we find that the gradient descent methods are less effective. They now often yield energies that can be 10 mE_h or more above the LM's answer though we again find that choice of step size plays a significant role. The worst

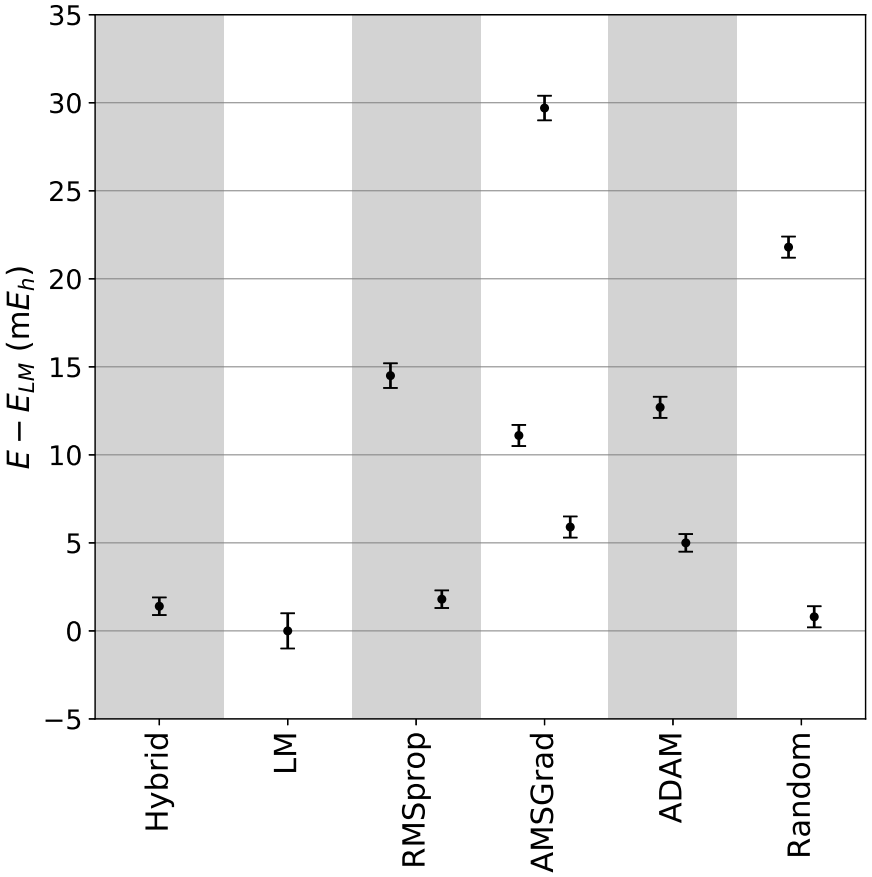


Figure 2.6: Different methods' optimized energies relative to that of the LM for stretched N_2 when optimizing CI coefficients and the TJF.

results for AMSGrad and ADAM were the result of choosing inappropriately large step sizes and simple reductions in the initial step size produced improvements in energy of tens of mE_h though the final result still remained well above the LM's. When we examined the optimizations over the course of their iterations, the gradient methods typically displayed some ability to quickly improve the wave function and energy initially, but they would then plateau and only very slowly improve the energy thereafter. Extrapolating from our results indicates that even if these gradient descent methods eventually converge to the minimum, they will only do so after thousands more iterations and at a computational cost well beyond that of the LM.

A more careful comparison of the different methods can be made by referring to Tables 2.2 and 2.3, which list the precise converged energies and their error bars. We also report the total number of samples used in each optimization as a proxy for computational effort, noting

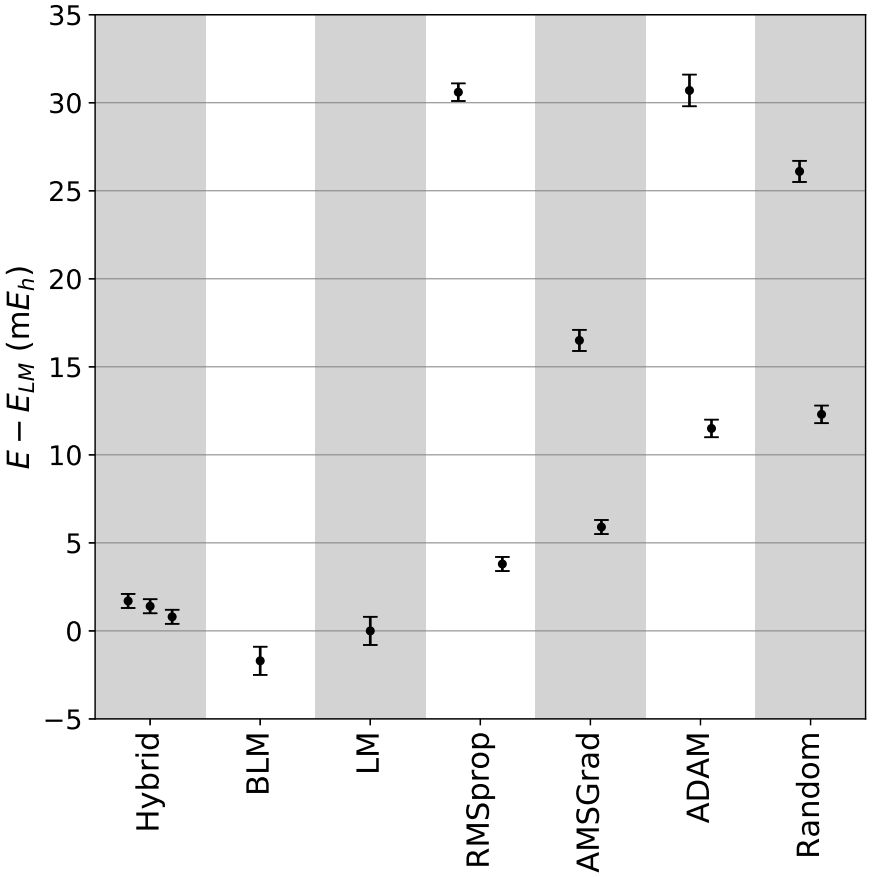


Figure 2.7: Different methods' optimized energies relative to that of the traditional LM for equilibrium N_2 when all parameters are optimized simultaneously. See also Table 2.2.

that for the BLM and the BLM portion of the hybrid method we double counted samples out of fairness as the BLM steps require running over their samples twice. In assessing cost, one must also consider the statistical uncertainty achieved, where we see that the LM and BLM are at a disadvantage. To help illustrate the update uncertainty contribution to this error, which we first discussed in the theoretical section above, we show the energy versus LM iteration for equilibrium N_2 in Figure 2.9. The fluctuations in energy from step to step, sometimes by as much as 2 mE_h, demonstrate the difficulty the LM faces from the uncertainty in its steps near the minimum. In this case, we see that the LM's final energy uncertainty is driven by the update step uncertainty rather than the uncertainty in evaluating the energy for a given set of parameter values at a particular iteration. We have observed similar behavior in the BLM and include the result of a BLM calculation in the tables. Note that the tabulated energies come from an average over the last ten optimization

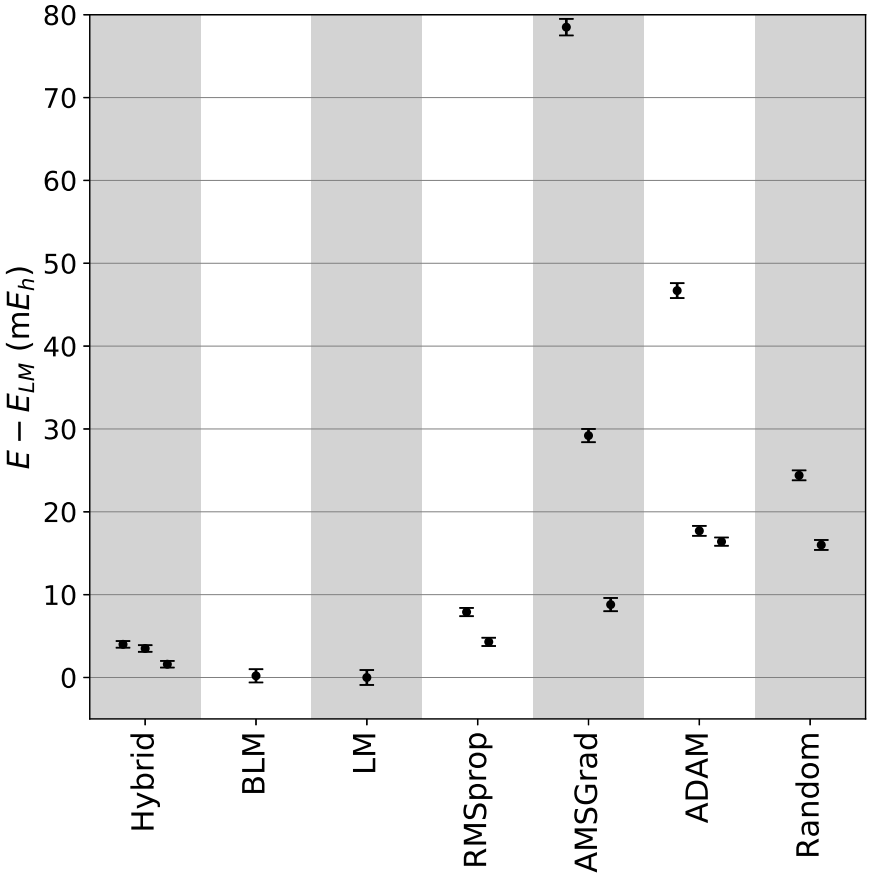


Figure 2.8: Different methods' optimized energies relative to that of the traditional LM for stretched N_2 when all parameters are optimized simultaneously. See also Table 2.3.

steps in the case of the standard LM and BLM and from an average over the last 50 descent steps in the case of the hybrid and pure descent methods.

From the tabulated data, we see that the hybrid optimization can achieve lower energies than the gradient descent methods using fewer samples, and that its results are typically within a few mE_h of the traditional LM. While the accelerated descent sections of the hybrid method provide some swift energy reductions early on when the wave function is still far from the minimum in parameter space, the BLM steps in the algorithm greatly accelerate the process of bringing the parameters near to the minimum, as can be seen in Figure 2.10. Looking at the electron-nuclear spline parameter and the orbital rotation parameter, we see typical cases in which rapid initial parameter movement during the early part of the first RMSprop stage transitions to much slower movement later in that stage, followed by very little movement at all in later RMSprop stages. Note that the latter can be explained

Table 2.2: Energies,uncertainties, and sample numbers for optimization of all parameters in equilibrium N_2.

Method	Energy (a.u.)	Uncertainty (a.u.)	Samples
Hybrid 1	-19.9263	0.0004	5,400,000
Hybrid 2	-19.9266	0.0004	10,000,000
Hybrid 3	-19.9272	0.0004	42,000,000
RMSprop 1	-19.8974	0.0005	20,000,000
RMSprop 2	-19.9242	0.0004	20,000,000
AMSGrad 1	-19.9115	0.0006	20,000,000
AMSGrad 2	-19.9221	0.0004	20,000,000
ADAM 1	-19.8973	0.0009	20,000,000
ADAM 2	-19.9165	0.0005	20,000,000
Random 1	-19.9019	0.0006	20,000,000
Random 2	-19.9157	0.0005	20,000,000
Linear Method	-19.9280	0.0008	40,000,000
Blocked Linear Method	-19.9297	0.0008	80,000,000
DF-BLM	-19.9293	0.0001	90,000,000
DF-Hybrid 1	-19.9290	0.0001	15,400,000
DF-Hybrid 2	-19.9290	0.0001	20,000,000
DF-Hybrid 3	-19.9293	0.0001	52,000,000

largely by the need to keep initial step sizes small in later stages to avoid significant upward deviations in the energy as the RMSprop method rebuilds its momentum history. In between these AD stages, the BLM updates move the parameter values in much larger steps, greatly accelerating convergence. This behavior makes the hybrid approach somewhat more black box as compared to the pure descent approaches, as the ability to get near the minimum with a modest sampling effort is much less dependent on the choice of the initial step sizes than for the AD methods. This conclusion is supported by the fact that the hybrid optimizations in Tables 2.2 and 2.3 used various initial step size settings (as discussed in Appendix B) and nonetheless produced lower energies than the pure descent methods in every case.

As discussed in our introduction of the hybrid method, another advantage is its ability to obtain a lower error bar at convergence than the LM for the same overall computational cost. This is a natural consequence of spending part of its sampling effort on gradient descent steps that correct for the BLM steps' uncertainty and bias (as illustrated earlier in Figure 2.4) and that hew closer to the zero-variance principle by importance sampling

Table 2.3: Energies,uncertainties, and sample numbers for optimization of all parameters in stretched N_2.

Method	Energy (a.u.)	Uncertainty (a.u.)	Samples
Hybrid 1	-19.6316	0.0004	5,400,000
Hybrid 2	-19.6321	0.0004	8,400,000
Hybrid 3	-19.6340	0.0004	49,200,000
RMSprop 1	-19.6277	0.0005	20,000,000
RMSprop 2	-19.6313	0.0005	20,000,000
AMSGrad 1	-19.5571	0.0010	20,000,000
AMSGrad 2	-19.6064	0.0008	20,000,000
AMSGrad 3	-19.6268	0.0008	20,000,000
ADAM 1	-19.5889	0.0009	20,000,000
ADAM 2	-19.6179	0.0006	20,000,000
ADAM 3	-19.6192	0.0005	20,000,000
Random 1	-19.6112	0.0006	20,000,000
Random 2	-19.6196	0.0006	20,000,000
Linear Method	-19.6356	0.0009	40,000,000
Blocked Linear Method	-19.6354	0.0008	80,000,000
DF-BLM	-19.6356	0.0001	90,000,000
DF-Hybrid 1	-19.6352	0.0001	15,400,000
DF-Hybrid 2	-19.6354	0.0001	18,400,000
DF-Hybrid 3	-19.6346	0.0001	59,200,000

with $|\Psi|^2$. To demonstrate this advantage explicitly, we ran additional sets of LM and hybrid optimizations adjusted to have essentially the same total number of samples. We then compare the standard error for the last ten LM steps and the last ten hybrid macro iterations in Figures 2.11 and 2.12, where we find that the hybrid has a substantially lower statistical uncertainty in every case. Assuming the usual $N^{-1/2}$ decay of uncertainty with sample size, the LM would require a factor of roughly four times more samples to reach the hybrid's uncertainty,

These statistical advantages in the final energy can be improved even further if we finish an optimization with a long section of pure descent. To demonstrate this, we have taken the final wave functions produced by the hybrid and BLM optimizations in Tables 2.2 and 2.3 and applied a further period of optimization using RMSprop with initial step sizes of 0.001 for all parameters. This "descent finishing" (DF) adds only a modest additional cost compared

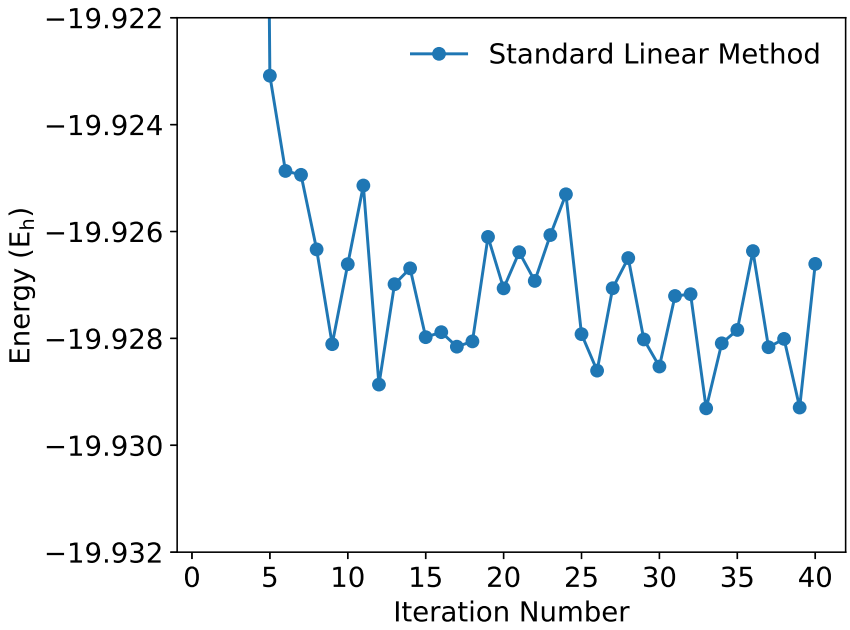


Figure 2.9: The standard LM optimization for all parameter equilibrium N_2.

to the preceding optimization and yields a large improvement in statistical uncertainty and, in many cases, an improvement in the final energy value as well. These advantages can be seen clearly in Figures 2.13 and 2.14, as well as in Tables 2.2 and 2.3, where we observe final error bars that are a factor of eight smaller than those of the LM. In terms of cost, this implies that the traditional LM would have required 64 times the original number of samples to achieve the DF-BLM or DF-hybrid precision. Put another way, we find that the DF-hybrid approach gives an equivalent or lower energy, with a much smaller error bar, at a substantially lower cost. Note that, in contrast, we find that this DF approach is not very effective when used in conjunction with the pure descent methods, where it essentially amounts to restarting the methods at the parameter values found after the first run of their optimization. While we do find that this restarting of the accumulation of momentum can improve the energy, the wave function parameters still do not reach their optimal values and the energy lowering vs total sampling cost is not competitive with the DF-hybrid.

Our study of the nitrogen dimer provides some clarity on the relative strengths of the LM and gradient descent, while also pointing the way to a more effective synthesis of the two. Gradient descent methods struggle in the presence of a variety of different highly nonlinear parameters, although they did perform better when we were only optimizing TJFs and CI coefficients. Among the descent methods, we found that the RMSprop approach came the closest to achieving the LM minimum energy. It is of course difficult to rule out

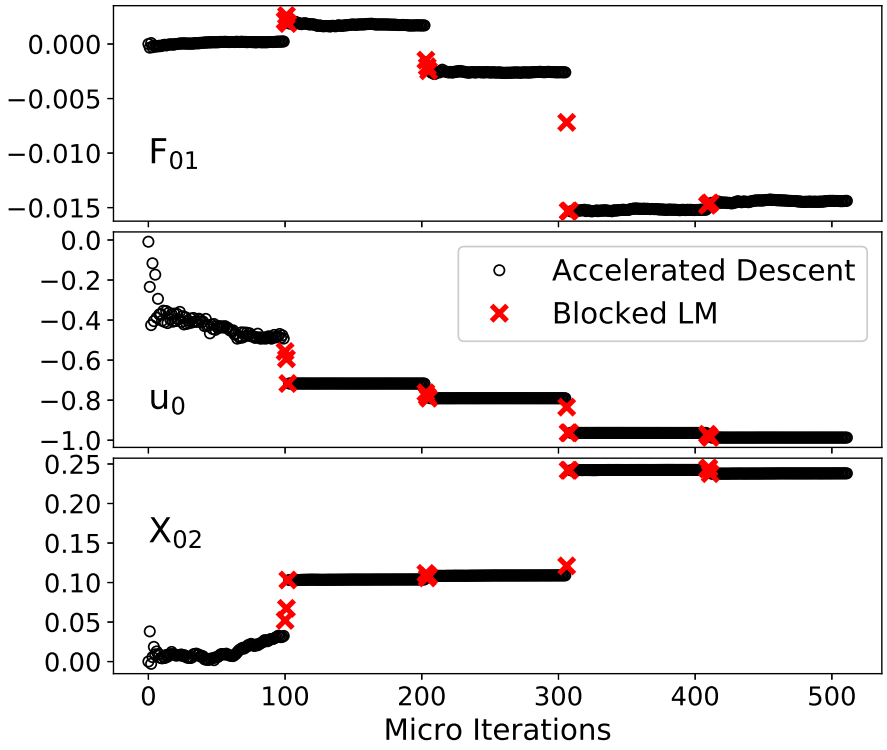


Figure 2.10: Values for the first off-diagonal F-matrix element (F_{01}), the first electron-nuclear TJF spline parameter (u_0), and the second orbital rotation variable (X_{02}) at each micro iteration of the "Hybrid 1" optimization for all parameter equilibrium N_2.

the possibility that this and other AD methods could reach the LM energy with additional sampling and more experimentation with the hyperparameters. However, it is far from obvious that this would be be cost-competitive, and the need to make careful and possibly system-specific choices for hyperparameters is somewhat antithetical to the general aspiration that an optimizer be as black-box as possible. For its part, the LM is more effective at moving parameters into the vicinity of the minimum, but tight convergence is then stymied by an unsatisfactory level of biased statistical uncertainty. As a side note, this behavior — in which the first derivative methods give better convergence once near the minimum but are at a relative disadvantage far from the minimum — is somewhat the reverse of what one would expect in deterministic optimization, where second derivative methods are at their strongest relative to first derivative methods during the final tight convergence in the vicinity of the minimum. Although things are reversed in the stochastic VMC case, we stress that the two classes of methodology are strongly complementary, as they compensate for each other's weaknesses. By using a low-memory version (BLM or hybrid) of the LM to get near to the minimum and then handing off to an accelerated descent method to achieve

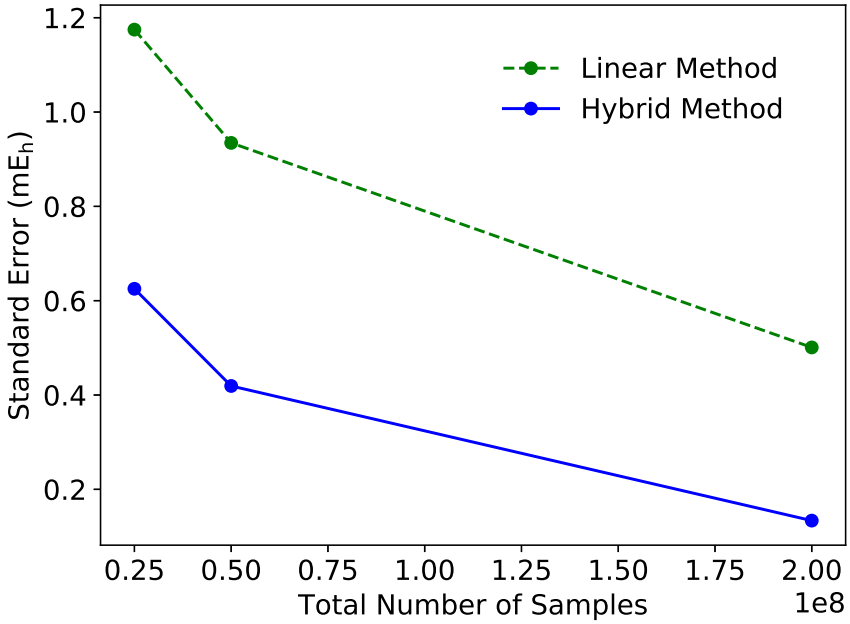


Figure 2.11: Standard Errors for the hybrid method and LM on all parameter equilibrium N_2 vs different optimizations' total sampling costs.

tight convergence, we find better overall performance than when working with either class of method on its own. These insights in hand, we will now apply this combined approach in a pair of larger and more challenging VMC optimization examples.

Styrene

We first turn to styrene at its equilibrium geometry (Figure 2.15) which offers an optimization with both more electrons and more variables, but in which the traditional LM is still quite achievable for comparison. As in N_2, we construct a multi-Slater wave function modified by both TJFs and a NCJF. To generate our Slater determinants, we have employed the heatbath selective CI (HCI) method as implemented in the Dice code by Sharma and coworkers. [10, 131] The orbital basis for the HCI calculation was produced via a (14e,14o) CASSCF calculation in Molpro [170] using a recently developed set of pseudopotentials and their corresponding double zeta basis.[171] In this CASSCF basis, HCI then correlated 32 electrons (out of a total of 40 electrons left over after applying pseudopotentials) in 64 orbitals. For our NCJF, we defined one counting region per atom, giving our F-matrix 135 optimizable parameters (see Appendix D for further NCJF details).

We optimized our wave function in a staged fashion using the standard LM, the BLM,

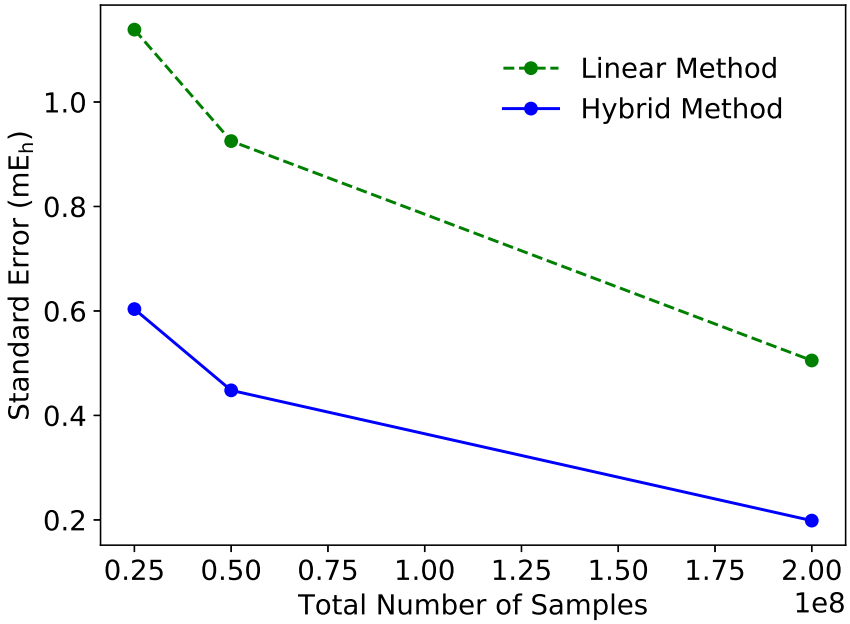


Figure 2.12: Standard Errors for the hybrid method and LM on all parameter stretched N_2 plotted against different optimizations' total sampling costs.

and the hybrid method. First, we conducted a partial optimization of the TJFs and the 100 most important CI coefficients. We then turned on the optimization of the orbitals and the NCJF's F-matrix, reaching a total of 4,570 parameters, most of them highly nonlinear. In the hybrid and BLM optimizations, the parameters were divided into 5 blocks and we used $N_k = 50$ and $N_o = 5$ for our numbers of kept directions and previous important directions, respectively. We used a value of 0.001 for ϵ in the $|\Phi|^2$ distribution for the LM and BLM sampling. These optimizations were then followed by 1,000 steps of RMSprop. As shown in Figure 2.16, we find that our hybrid method reaches a converged energy as low or better than that of the standard and blocked LM, and finishing our optimizations with descent provides a substantial improvement in the statistical uncertainty even in this more challenging case.

FNNF

We now turn our attention to a strongly correlated transition state of the the diflurodiazene (FNNF) *cis-trans* isomerization, where we test the hybrid optimization approach on a much larger determinant expansion. The FNNF isomerization can be thought of as a toy model molecule for larger systems such as photoswitches, which have potential uses in molecular machines [172, 173] and high-density memory storage. [174] In addition, FNNF itself is

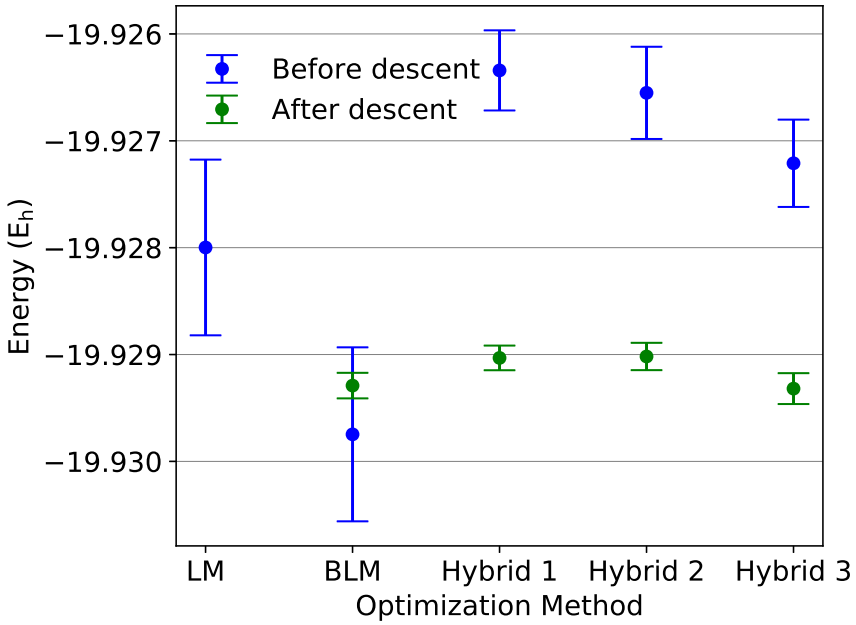


Figure 2.13: Converged energies in equilibrium N_2 before and after a final descent optimization.

Table 2.4: A summary of the VMC optimization stages in FNNF showing the number of determinants N_d included from HCI, which parameters are optimized, and the total number N_p of optimized parameters. Note that CI coefficients are optimized at every stage. Stages 2, 3, and 4 start from the parameter values from the previous stage, with newly added determinants' coefficients initialized to zero. We also report the number of iterations performed in each stage, which for stage 4 is simply the number of RMSprop steps. A hybrid iteration, on the other hand, consists of 100 RMSprop steps followed by three BLM steps. All RMSprop steps use 20,000 samples drawn from $|\Psi|^2$, while the BLM steps each use 1 million samples drawn from the $|\Phi|^2$ guiding function with ϵ set to 0.01.

Stage	Method	N_d	TJF	F-matrix	Orbitals	N_p	Iterations
1	Hybrid	10^2	✓			139	9
2	Hybrid	10^3	✓	✓		1048	4
3	Hybrid	10^4	✓	✓	✓	15,573	6
4	AD	10^4	✓	✓	✓	15,573	1,000

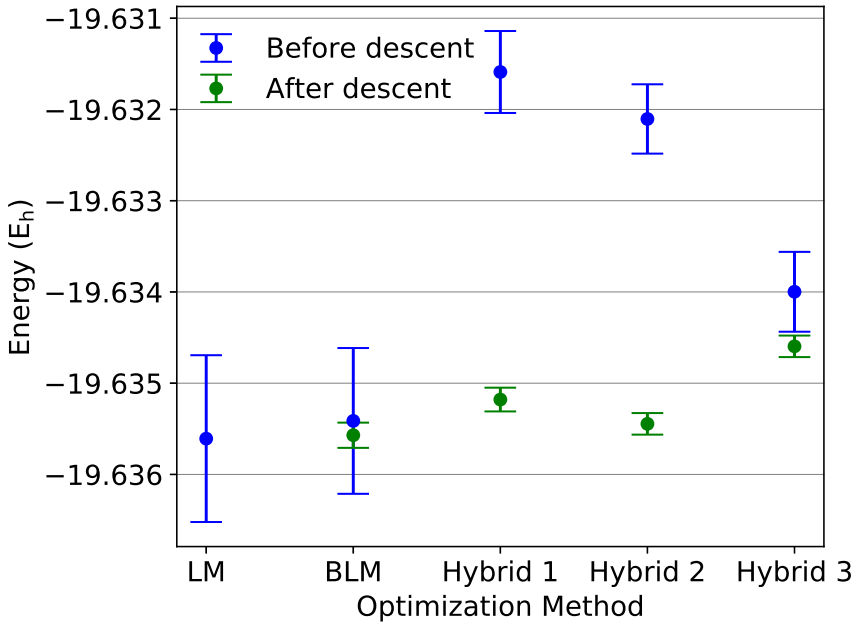


Figure 2.14: Converged energies in stretched N_2 before and after a final descent optimization.

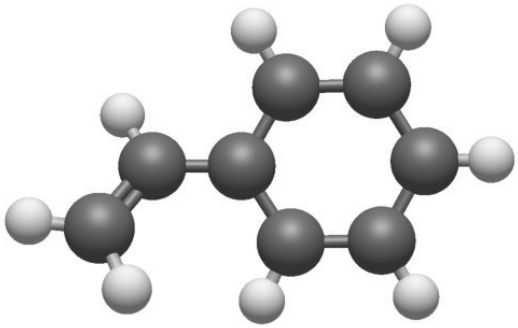


Figure 2.15: Equilibrium geometry of styrene. See Appendix C for structure coordinates.

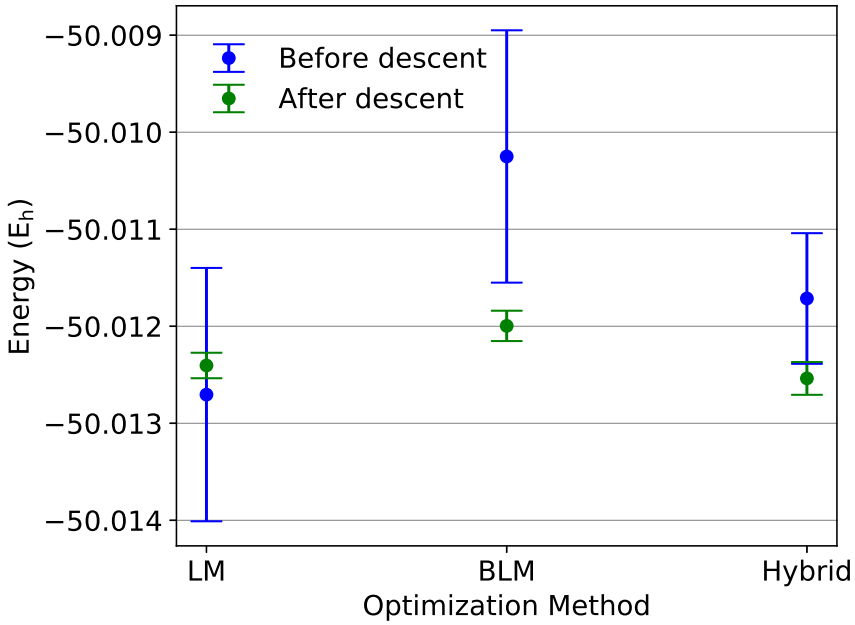


Figure 2.16: Converged energies in equilibrium styrene before and after a final descent optimization.

Table 2.5: Energies of the transition state of FNNF.

Method	Energy (a.u.)	Uncertainty (a.u.)
Hartree-Fock	-67.112730	
CASSCF	-67.359100	
VMC Stage 1	-68.1017	0.0011
VMC Stage 2	-68.1213	0.0009
VMC Stage 3	-68.1698	0.0006
VMC Stage 4	-68.1750	0.0002

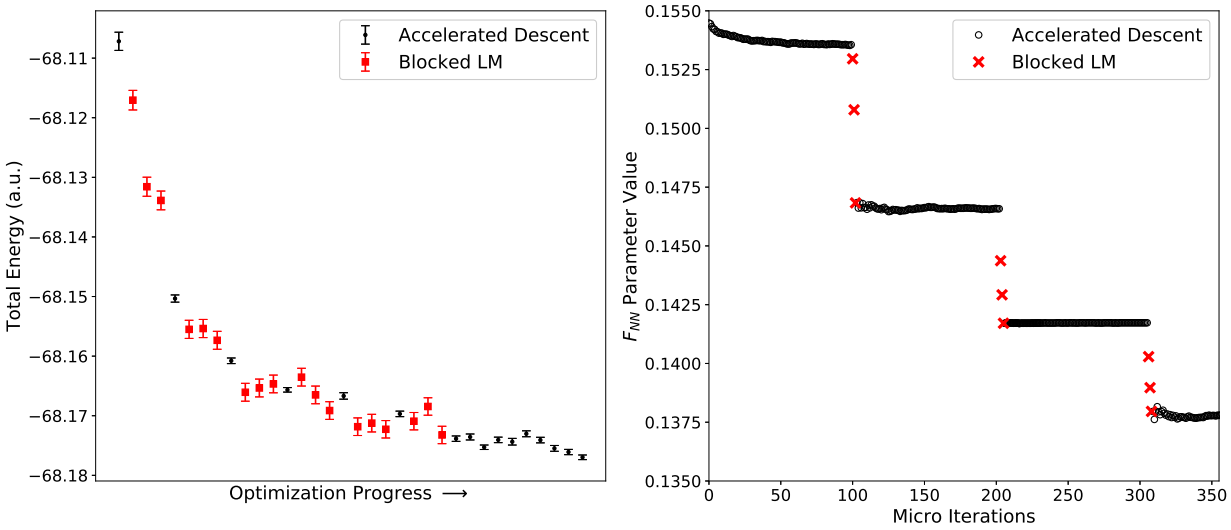


Figure 2.17: Left panel: energies during stages 3 and 4 of the FNNF optimization. The descent energies are reported as the average over the last 50 RMSprop steps within each block of 100 RMSprop steps, whereas the BLM energies are the energy estimates on the random samples used for the BLM update steps. Right panel: change in the value of the F-matrix parameter that couples the two nitrogen atoms' counting regions over the first three macro iterations of stage 3, with each micro iteration corresponding to one RMSprop or BLM step. The nine BLM points on the right panel correspond to the first nine BLM points on the left panel.

of interest as part of the synthesis of high energy polynitrogen compounds and has been the subject of multiple electronic structure studies.[175–177] Here we focus on its strongly correlated transition state, which is the direct analogue of the out-of-plane TS1 transition state in diazene. [178]

Our treatment of this transition state began by locating its geometry via an (8e,8o) CASSCF optimization in the cc-pVTZ basis using Molpro.[170] At this geometry (given in Appendix C) we then switch over to using BFD pseudopotentials and their corresponding triple zeta basis, [168] in which we use the Dice code [10, 131] to iterate an HCI calculation with 24 electrons distributed in the lowest 50 (8e,8o) CASSCF orbitals until its variational wave function has reached almost 2 million determinants. We then import the first 10,000 of these determinants into our VMC optimization and combine them with TJFs, atom-centered NCJFs, and orbital optimization, which produces an ansatz with over 15,000 variational parameters.

Our VMC optimization proceeds in stages as summarized in Table 2.4. This begins with TJFs and a 100-determinant ansatz from HCI, with later optimization stages adding more determinants and turning on the optimization of the NCJF and orbital rotation variables.

As in styrene, the strategy is to bring the parameters near to their optimal values with the help of the LM and then to perform a final unbiased relaxation via a long run of RMSprop AD. Due to the large number of variational parameters, we incorporate the LM via the hybrid scheme, with the BLM steps employing 2, 2, and 10 blocks during stages 1, 2, and 3, respectively. In stages 1 and 2, we used an initial RMSprop step size of 0.01 for TJFs and CI coefficients before setting it to 0.005 at the beginning of stage 3. For the F-matrix parameters, we began by setting the initial step size to 0.001, but after observing a significant rise and fall of the energy during the RMSprop section of the first hybrid macro iteration in stage 3, we reduced this to 0.0001 and also lowered the TJF and CI step size to 0.0005 for the last 4 macro iterations in that stage. For all steps in stage 4, we maintained the 0.0001 step size for the F-matrix parameters and lowered the initial step size for TJFs and CI coefficients to 0.0002. An initial step size of 0.0001 was used for orbital parameters throughout both stages 3 and 4. The BLM steps used the $|\Phi|^2$ guiding function with a value of 0.01 for ϵ.

The energies resulting from this staged optimization are shown in Table 2.5 and Figure 2.17. Unsurprisingly, stage 3 proved to be the most challenging and expensive stage, as it is where we hope to move all parameters near to their final values in a setting where the traditional LM would face severe memory bottlenecks. As seen in Figure 2.17, both the AD and BLM steps clearly work to lower the energy during the first two macro iterations of stage 3. In the last four macro iterations of stage 3, however, the energy decreases more slowly and it is less clear, at least when looking at the energetics, whether the BLM steps are still necessary. Instead, their importance is revealed by inspecting the movement of the F-matrix values within the NCJF, an example of which is shown in the right-hand panel of Figure 2.17. As in N_2, these parameters prove to be the most resistant to optimization via AD, and we clearly see that although AD does gradually move their values in the same direction as the BLM, the BLM steps dramatically accelerate their optimization. This effect is seen throughout all six macro iterations of stage 3, and so although the BLM energies are not obviously improving at the end of this stage, the inclusion of these steps is clearly still beneficial. Note that relaxing the NCJF after moving from a 1,000-determinant to a 10,000-determinant expansion is important, because the larger determinant expansion is better able to capture some of the correlation effects that the NCJF is encoding, and so we expect (and indeed see) that this diminishing of its role leads to smaller F-matrix values being optimal.

Although we have again found that it would be difficult for AD alone to provide a successful optimization of our ansatz, the statistical advantages of its incorporation are still quite clear. A close inspection of the sample sizes used in the optimization reveals that each of the AD and BLM points in the left panel of Figure 2.17 corresponds to averaging over 1 million random samples. Despite this equal sampling effort, the uncertainties for the AD energy estimates are about one third the size of those for the BLM, implying that a pure BLM approach would require an order of magnitude more sampling effort to produce similar results. To understand this statistical advantage, we need to remember two important differences between the AD and BLM steps. First, the nonlinearity of the LM and BLM eigenvalue problem leads to biases in the update steps that can both increase the step-to-step energy uncertainty and cause the method to optimize off-center from the true minimum.

Second, the use of an alternative guiding function for the BLM samples in order to mitigate this step uncertainty moves us away from the zero-variance regime enjoyed by traditional $|\Psi|^2$ sampling. If we were to instead employ traditional sampling, our energy estimates for a specific wave function would improve, but the BLM step uncertainty would increase sharply. As the AD methods do not suffer from these issues, they help us to further mitigate the BLM step uncertainty and to perform a final, high-precision relaxation during stage 4. In total, incorporating the AD steps in this case roughly doubles the number of samples required, but is well worthwhile given that it improves statistical efficiency by almost an order of magnitude.

While it is possible that the NCJF parameters are not quite converged in this particular optimization and that increasing iteration counts in stages 1 through 3 could further improve the energy, the lessons learned from investigating a large MSJ optimization for the FNNF transition state are already clear. While both the BLM and the AD methods can be used in this 10,000+ parameter regime, they bring highly complementary advantages to the optimization and so would appear to work better together than apart. In particular, the BLM helps optimize the parameters that change only very slowly during AD, whereas the statistical advantages of AD greatly increase precision at a given sample size and work to eliminate the statistical biases suffered by the BLM.

2.5 Conclusions and Outlook

We have found that a combination of first and second derivative optimization methods appears to work better than using either class of method on its own when minimizing the energies of wave functions in variational Monte Carlo. This is particularly true for wave functions with a wide variety of different types of nonlinear parameters, as for example when dealing simultaneously with traditional one- and two-body Jastrow factors, many-body Jastrow factors, and orbital relaxations. While the linear method and its low-memory variants show a superior ability to move these nonlinear parameters into the vicinity of their optimal values, accelerated descent methods prove much more capable of converging them tightly around the minimum. This situation stands as an interesting reverse of what is typically encountered in deterministic optimization, where second derivative methods are usually superior for tight final convergence and first derivative methods perform relatively better in the early stages of an optimization. The realities of working with statistically uncertain energies and energy derivatives turns this expectation on its head, both because of the need to stabilize the statistics of the linear method's second derivative elements through zero-variance-violating importance sampling schemes and due to the nonlinear biases that are induced when solving the linear method's eigenvalue problem. The linear method's ability to quickly move the parameters near the minimum, however, makes it appear that employing it as part of a hybrid approach is well worthwhile. Indeed, in our testing, hybridizing low-memory linear method variants with accelerated descent methods provides better energies with smaller statistical uncertainties at a lower computational cost when compared to the

stand-alone use of either the linear method or accelerated descent methods.

Looking forward, there are many questions still to be answered about the interplay between first and second derivative methods. For example, although the blocked linear method greatly reduces memory cost vs the traditional linear method, it is not clear that it can be applied effectively beyond 100,000 parameters in its current form. One thus wonders whether it is necessary to optimize all of the parameters during the linear method steps of the hybrid approach, or whether it may be possible to identify (perhaps during an ongoing optimization?) which parameters would benefit from linear method treatment and which would not. Were such a sorting possible, accelerated descent methods with their even lower memory footprint could be left to deal with most of the parameters, with only a relatively small subset treated by the linear method steps. Another important issue is making the hybrid approach as black box and user friendly as possible. Although we have tested it here with many different descent step size settings for the different parameter types, this has not been a systematic survey. More extensive testing may allow clear defaults to be settled upon so that users can reasonably expect a successful optimization without resorting to careful step size control. We look forward to investigating these exciting possibilities in future.

2.6 Acknowledgements

This work was supported by the Office of Science, Office of Basic Energy Sciences, the US Department of Energy, Contract No. DE-AC02-05CH11231. Calculations were performed using the Berkeley Research Computing Savio cluster and the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

2.7 Appendix A: Additional Energies

Table 2.6: Precise Values for optimizing CI coefficients and traditional Jastrow factors in equilibrium N_2.

Method	Energy (a.u.)	Uncertainty (a.u.)	Samples
Hybrid 1	-19.9083	0.0005	3,900,000
Hybrid 2	-19.9095	0.0006	5,400,000
Hybrid 3	-19.9101	0.0005	29,000,000
RMSprop 1	-19.8936	0.0006	20,000,000
RMSprop 2	-19.9091	0.0005	20,000,000
AMSGrad 1	-19.8998	0.0007	20,000,000
AMSGrad 2	-19.8926	0.0008	20,000,000
AMSGrad 3	-19.9048	0.0006	20,000,000
AMSGrad 4	-19.9071	0.0005	20,000,000
ADAM 1	-19.9009	0.0005	20,000,000
ADAM 2	-19.9056	0.0006	20,000,000
ADAM 3	-19.9078	0.0005	20,000,000
Random 1	-19.8926	0.0006	20,000,000
Random 2	-19.9041	0.0005	20,000,000
Linear Method	-19.9105	0.0010	40,000,000

Table 2.7: Precise Values for optimizing CI coefficients and traditional Jastrow factors in stretched N_2.

Method	Energy (a.u.)	Uncertainty (a.u.)	Samples
Hybrid 1	-19.6141	0.0005	27,000,000
RMSprop 1	-19.6010	0.0007	20,000,000
RMSprop 2	-19.6137	0.0005	20,000,000
AMSGrad 1	-19.6044	0.0006	20,000,000
AMSGrad 2	-19.5858	0.0007	20,000,000
AMSGrad 3	-19.6096	0.0006	20,000,000
ADAM 1	-19.6028	0.0006	20,000,000
ADAM 2	-19.6105	0.0005	20,000,000
Random 1	-19.5937	0.0006	20,000,000
Random 2	-19.6147	0.0006	20,000,000
Linear Method	-19.6155	0.0010	40,000,000

2.8 Appendix B: Details for N_2 Optimizations

The details for the optimizations behind the final energies in the main paper are presented below. In every case of N_2, all flavors of pure gradient descent based optimization were run for 2000 iterations at 10,000 samples per iteration except for the random step size method, which was run for 10,000 iterations at 2000 samples per iteration. The total sampling cost was then 20 million samples, half of the standard linear method's cost of 40 million over 40 steps. Given the descent methods' tendency to plateau after a few hundred iterations and then lower the energy by only a few mE_h afterward, we expect that running them longer to fully match or exceed the linear method's sampling effort would not yield a much better result in most cases. We found that it was often advantageous to allow for different types of parameters to be given different initial step sizes. Tables 2.8 through 2.11 list the step sizes for different descent optimizations in all cases of N_2. The values can be cross-referenced with the energy results in earlier tables to see which choices were most effective. Some amount of experimentation was necessary to build up intuition for what choices are effective, but we generally expect more nonlinear parameters such as those in the F-matrix and the orbitals to require smaller step sizes. We also found that RMSprop benefited from using larger step sizes compared to other descent algorithms. The energy may be significantly raised on early iterations, but tends to be quickly lowered and eventually brought to an improved result once enough gradient history has built up over more steps.

The details of the different hybrid method optimizations are slightly more involved and are discussed separately here. In all cases, the blocked linear method steps of the hybrid

optimization used 5 blocks with 5 directions from sections of RMSprop to provide coupling to variables outside a block and retained 30 directions from each block to construct the final space for determining the parameter update. These were also the settings given to the blocked linear method optimizations of N_2 that appear in the main text. All hybrid optimizations used the RMSprop method for their AD sections with the hyperparameters $d = 100$, $\rho = .9$ and $\epsilon = 10^{-8}$. The step sizes used in the AD sections varied over the course of the hybrid optimizations. We typically chose larger step sizes for the AD portion of the first macro-iteration in order to obtain more energy and parameter improvement at a low sampling cost before any use of the BLM and these are tabulated separately as "Hybrid-Initial". The smaller step sizes reported for the rest of hybrid optimization were used in the later macro-iterations to avoid rises in the energy that might occur before a sufficient gradient history was accumulated. We also list the step sizes in the long RMSprop optimization used to achieve the descent finalized energies. These were sometimes larger than those for the AD sections of the initial hybrid optimizations because the descent finalization was long enough for any early transient rises in the energy to recover.

We now specify how the hybrid method sampling costs reported in Tables 2.2 and 2.3 of the main paper were divided between AD and the BLM. In all parameter equilibrium N_2, Hybrid 1 consisted of 500 AD steps costing 3 million samples interwoven with 12 BLM steps that cost 2.4 million samples. Hybrid 2 consisted of the same sequence of steps, but had an increased sampling effort of 6 million samples on descent and 2.4 million on BLM. Hybrid 3 had a greatly increased sampling cost and consisted of 1400 AD steps for 11.2 million samples interwoven with 19 BLM steps costing 38 million. For all parameter stretched N_2, Hybrid 1 used the same sequence of steps and sampling cost breakdown as Hybrid 1 for the equilibrium case. Hybrid 2 consisted of 600 AD steps that cost 7 million samples and 15 BLM steps that cost 3 million. Hybrid 3 also had 600 AD steps, now using 12 million samples, and 15 BLM steps, now using 30 million samples. For all descent finalizations in N_2, we used 1000 steps of RMSprop at an additional cost of 10 million samples and took an average over the last 500 steps to obtain our reported energies and error bars.

Finally, we give the breakdown of the hybrid sampling costs in Tables 2.6 and 2.7 of Appendix A. For the equilibrium case, Hybrid 1 had 500 AD steps costing 1.5 million samples and 12 BLM steps costing 2.4 million samples. Hybrid 2 had the same combination of steps and BLM cost as Hybrid 1 while the AD steps used 3 million samples. Hybrid 3 used the same sequence of steps, but increased the AD and BLM sampling costs to 5 million and 24 million, respectively. In the stretched case, the hybrid optimization used 500 AD steps with 3 million samples and 12 BLM steps costing 24 million samples.

Table 2.8: Step sizes for TJFCI equilibrium N_2 optimizations.

Method	2 Body Jastrow	1 Body Jastrow	CI
RMSprop/Booth 1	0.01	0.01	0.005
RMSprop/Booth 2	0.05	0.05	0.01
AMSGrad 1	0.05	0.05	0.01
AMSGrad 2	0.05	0.05	0.05
AMSGrad 3	0.01	0.01	0.01
AMSGrad 4	0.001	0.001	0.001
ADAM 1	0.01	0.01	0.01
ADAM 2	0.005	0.005	0.005
ADAM 3	0.001	0.001	0.001
Random 1	0.01	0.01	0.01
Random 2	0.0005	0.0005	0.0005
Hybrid-Initial 1	0.1	0.1	0.01
Hybrid-Initial 2	0.1	0.1	0.01
Hybrid-Initial 3	0.1	0.1	0.01
Hybrid 1	0.001	0.001	0.001
Hybrid 2	0.005	0.005	0.005
Hybrid 3	0.005	0.005	0.005

Table 2.9: Step sizes for TJFCI stretched N_2 optimizations.

Method	2 Body Jastrow	1 Body Jastrow	CI
RMSprop/Booth 1	0.01	0.01	0.005
RMSprop/Booth 2	0.05	0.05	0.01
AMSGrad 1	0.05	0.05	0.01
AMSGrad 2	0.05	0.05	0.05
AMSGrad 3	0.01	0.01	0.01
ADAM 1	0.01	0.01	0.01
ADAM 2	0.005	0.005	0.005
Random 1	0.01	0.01	0.01
Random 2	0.001	0.001	0.001
Hybrid-Initial 1	0.1	0.1	0.1
Hybrid 1	0.005	0.005	0.005

Table 2.10: Step sizes for all parameter equilibrium N_2 optimizations.

Method	2 Body Jastrow	1 Body Jastrow	F-Matrix	CI	Orbitals
RMSprop/Booth 1	0.005	0.005	0.001	0.001	0.001
RMSprop/Booth 2	0.05	0.05	0.01	0.01	0.01
AMSGrad 1	0.05	0.05	0.005	0.01	0.001
AMSGrad 2	0.005	0.005	0.001	0.001	0.001
ADAM 1	0.05	0.05	0.005	0.01	0.001
ADAM 2	0.005	0.005	0.001	0.001	0.001
Random 1	0.001	0.001	0.001	0.001	0.001
Random 2	0.001	0.001	0.0005	0.001	0.0005
Hybrid-Initial 1	0.1	0.1	0.0001	0.01	0.01
Hybrid-Initial 2	0.1	0.1	0.0005	0.01	0.01
Hybrid-Initial 3	0.01	0.01	0.001	0.01	0.001
Hybrid 1	0.0001	0.0001	0.0001	0.0001	0.0001
Hybrid 2	0.001	0.001	0.0005	0.0005	0.0005
Hybrid 3	0.001	0.001	0.001	0.001	0.001
DF-Hybrid 1	0.001	0.001	0.001	0.001	0.001
DF-Hybrid 2	0.001	0.001	0.001	0.001	0.001
DF-Hybrid 3	0.001	0.001	0.001	0.001	0.001

Table 2.11: Step sizes for all parameter stretched N_2 optimizations.

Method	2 Body Jastrow	1 Body Jastrow	F-Matrix	CI	Orbitals
RMSprop/Booth 1	0.05	0.05	0.05	0.01	0.01
RMSprop/Booth 2	0.1	0.1	0.01	0.01	0.005
AMSGrad 1	0.05	0.05	0.05	0.01	0.01
AMSGrad 2	0.05	0.05	0.01	0.02	0.001
AMSGrad 3	0.005	0.005	0.001	0.002	0.001
ADAM 1	0.05	0.05	0.05	0.01	0.01
ADAM 2	0.05	0.05	0.01	0.02	0.001
ADAM 3	0.005	0.005	0.001	0.002	0.001
Random 1	0.001	0.001	0.001	0.001	0.001
Random 2	0.001	0.001	0.0005	0.001	0.0005
Hybrid-Initial 1	0.1	0.1	0.01	0.01	0.001
Hybrid-Initial 2	0.1	0.1	0.01	0.01	0.001
Hybrid-Initial 3	0.1	0.1	0.01	0.01	0.001
Hybrid 1	0.0001	0.0001	0.0001	0.0001	0.0001
Hybrid 2	0.001	0.001	0.0005	0.0005	0.0005
Hybrid 3	0.001	0.001	0.0005	0.0005	0.0005
DF-Hybrid 1	0.001	0.001	0.001	0.001	0.001
DF-Hybrid 2	0.001	0.001	0.001	0.001	0.001
DF-Hybrid 3	0.001	0.001	0.001	0.001	0.001

2.9 Appendix C: Molecular Geometries

Table 2.12: Structure of equilibrium styrene. Coordinates in Å.

C	1.392 95	0.000 00	0.000 00
C	2.160 42	−1.192 58	0.018 01
C	2.094 21	1.231 78	−0.019 14
C	3.565 85	−1.159 69	0.052 86
C	3.501 42	1.272 11	0.017 95
C	4.236 86	0.075 03	0.060 81
C	0.000 00	0.000 00	0.000 00
C	−0.795 15	−0.930 87	0.544 06
H	1.712 22	−2.111 61	−0.002 39
H	1.592 37	2.124 71	−0.047 53
H	4.098 18	−2.032 73	0.071 53
H	3.990 86	2.169 87	0.016 92
H	5.257 94	0.100 43	0.095 03
H	−0.463 24	0.771 12	−0.427 75
H	−0.434 31	−1.721 47	1.022 78
H	−1.782 40	−0.845 77	0.492 96

Table 2.13: Structure of FNNF transition state. Coordinates in Å.

N	0.499 39	−0.446 56	−0.593 77
N	0.570 66	0.412 24	0.5639
F	−0.390 84	0.145 63	−1.369 59
F	−0.398 07	−0.120 32	1.391 57

2.10 Appendix D: NCJF Gaussian Basis

The form for the three dimensional Gaussian basis functions of NCJFs in the main text can equivalently be written as $g_j(\mathbf{r}) = \exp\left(\mathbf{r}^T\mathbf{A}\mathbf{r} - 2\mathbf{B}^T\mathbf{r} + C\right)$ where $\mathbf{A}$ is a symmetric matrix defined by 6 parameters, $\mathbf{B}$ is a three-component vector, and C is a single dimensionless number. We used 16 basis functions for the all parameter cases of N_2, 16 in styrene, and 4 in FNNF. Their complete specifications are presented in Tables 2.14-2.18. The components $A_{xx}, A_{xy}, A_{xz}, A_{yy}, A_{yz}, A_{zz}$ with units of inverse square bohr are the same for each basis function within a particular system and are therefore listed separately in Table 2.14. Tables 2.15-2.18 contain components B_x, B_y, B_z, with units of inverse bohr and C for each system's basis functions.

Table 2.14: Components of the matrix **A** for our systems.

System	A_{xx}	A_{xy}	A_{xz}	A_{yy}	A_{yz}	A_{zz}
Equilibrium N_2	-6.9282	0.0	0.0	-6.9282	0.0	-6.9282
Stretched N_2	-6.9282	0.0	0.0	-6.9282	0.0	-6.9282
Styrene	-0.1	0.0	0.0	-0.1	0.0	-0.1
FNNF	-0.1	0.0	0.0	-0.1	0.0	-0.1

Table 2.15: Gaussian components for all parameter equilibrium N_2.

Basis Function	B_x	B_y	B_z	C
g_0	-0.8	-0.8	-0.8	-0.2771
g_1	0.8	-0.8	-0.8	-0.2771
g_2	-0.8	0.8	-0.8	-0.2771
g_3	0.8	0.8	-0.8	-0.2771
g_4	-0.8	-0.8	0.8	-0.2771
g_5	0.8	-0.8	0.8	-0.2771
g_6	-0.8	0.8	0.8	-0.2771
g_7	0.8	0.8	0.8	-0.2771
g_8	-4.4787	-0.8	-0.8	-11.2500
g_9	-2.8787	-0.8	-0.8	-4.5981
g_{10}	-4.4787	0.8	-0.8	-11.2500
g_{11}	-2.8787	0.8	-0.8	-4.5981
g_{12}	-4.4787	-0.8	0.8	-11.2500
g_{13}	-2.8787	-0.8	0.8	-4.5981
g_{14}	-4.4787	0.8	0.8	-11.2500
g_{15}	-2.8787	0.8	0.8	-4.5981

Table 2.16: Gaussian basis functions for all parameter stretched N_2.

Basis Function	B_x	B_y	B_z	C
g_0	-0.8	-0.8	-0.8	-0.2771
g_1	0.8	-0.8	-0.8	-0.2771
g_2	0.8	0.8	-0.8	-0.2771
g_3	0.8	0.8	-0.8	-0.2771
g_4	-0.8	-0.8	0.8	-0.2771
g_5	0.8	-0.8	0.8	-0.2771
g_6	-0.8	0.8	0.8	-0.2771
g_7	0.8	0.8	0.8	-0.2771
g_8	-5.8015	-0.8	-0.8	-22.7322
g_9	-4.2015	-0.8	-0.8	-11.8474
g_{10}	-5.8015	0.8	-0.8	-22.7322
g_{11}	-4.2015	0.8	-0.8	-11.8474
g_{12}	-5.8015	-0.8	0.8	-22.7322
g_{13}	-4.2015	-0.8	0.8	-11.8474
g_{14}	-5.8015	0.8	0.8	-22.7322
g_{15}	-4.2015	0.8	0.8	-11.8474

Table 2.17: Gaussian basis functions for equilibrium styrene.

Basis Function	B_x	B_y	B_z	C
g_0	-0.2632	0.0	0.0	-0.6929
g_1	-0.4083	0.2254	-0.003403	-2.1748
g_2	-0.3957	-0.2328	0.003617	-2.1081
g_3	-0.6738	0.2191	-0.009989	-5.0220
g_4	-0.6617	-0.2404	-0.003392	-4.9561
g_5	-0.8007	-0.01418	-0.01149	-6.4137
g_6	0.0	0.0	0.0	0.0
g_7	0.1503	0.1759	-0.1028	-0.6409
g_8	-0.3236	0.3990	0.0004516	-2.6392
g_9	-0.3009	-0.4015	0.008982	-2.5184
g_{10}	-0.7744	0.3841	-0.01352	-7.4750
g_{11}	-0.7542	-0.41005	-0.003197	-7.3691
g_{12}	-0.9936	-0.01898	-0.01796	-9.8794
g_{13}	0.08754	-0.1457	0.08083	-0.3543
g_{14}	0.08207	0.3253	-0.19328	-1.4992
g_{15}	0.3368	0.1598	-0.09316	-1.4767

Table 2.18: Gaussian basis functions for FNNF.

Basis Function	B_x	B_y	B_z	C
g_0	-0.09437	0.08439	0.1122	-0.2862
g_1	-0.1078	-0.07790	-0.1066	-0.2905
g_2	0.07386	-0.02752	0.2588	-0.7320
g_3	0.07522	0.02274	-0.2630	-0.7533

Chapter 3

A hybrid approach to excited-state-specific variational Monte Carlo and doubly excited states

3.1 Abstract

We extend our hybrid linear-method/accelerated-descent variational Monte Carlo optimization approach to excited states and investigate its efficacy in double excitations. In addition to showing a superior statistical efficiency when compared to the linear method, our tests on small molecules show good energetic agreement with benchmark methods. We also demonstrate the ability to treat double excitations in systems that are too large for a full treatment by selected configuration interaction methods via an application to 4-aminobenzonitrile. Finally, we investigate the stability of state-specific variance optimization against collapse to other states' variance minima and find that symmetry, ansatz quality, and sample size all have roles to play in achieving stability.

3.2 Introduction

Accurate predictions about doubly excited states remain a significant challenge for electronic structure methods. Although they are rare in the low-lying spectra of simple organic molecules, states with significant or near-total doubly excited character are not uncommon in aromatic systems, other π-conjugated settings, and transition metal compounds. Their difficulty can be understood by considering three factors. First, they tend to be strongly multi-reference in character, which frustrates traditional weakly-correlated quantum chemistry approaches. Second, predicting accurate excitation energies requires a method to also capture weak correlation effects, and capturing both strong and weak correlation in medium

to large systems remains an open challenge in electronic structure, despite the recent progress in smaller systems.[10, 11, 37, 41, 55, 131, 179–183] Third, they are excited states, which remain harder to treat than ground states. When faced with a doubly excited state, many excited state approaches are either not appropriate or difficult to afford. Time-dependent density functional theory (TDDFT), at least within the adiabatic approximation, is famously incapable of predicting double excitations,[27] while equation of motion coupled cluster with singles and doubles (EOM-CCSD) is much less accurate in these states than in single excitations. [40] Higher-level coupled cluster is more effective, [184] but the $O(N^8)$ and $O(N^{10})$ cost scalings of EOM-CCSDT and EOM-CCSDTQ make them difficult to use outside of small molecules. Similarly, selected configuration interaction (sCI) methods can establish benchmark results for double excitations in small molecules, [41] but their exponential scaling makes them difficult to extend to larger systems. For these reasons, more traditional multi-reference methods like CASPT2[20, 21] have long been favored when dealing with double excitations.[41, 185–196] These methods have their limitations as well — intruder states, smaller active space sizes than sCI, sometimes-sensitive state-averaging choices — and so the development of alternative high-accuracy approaches to double excitations, especially ones that rely on very different approximations that could cross-validate current methods' predictions, remains an important priority.

Quantum Monte Carlo (QMC) methods are, in principle, a promising alternative for doubly excited states and for difficult excited states more generally. Thanks to an ability to employ correlation factors and projector Monte Carlo methods to impart weak correlation effects on top of modest determinant expansions that capture the primary strong correlations, QMC approaches offer one route towards an integrated treatment of weak and strong correlation. Moreover, the ability of variational Monte Carlo (VMC) to work with excited states, either in a state-averaged [86, 87, 90, 91, 197] or a state-specific [93–95] manner, offers a clear route to extending these advantages to studies of doubly excited states. However, these advantages are only useful in practice if these sophisticated wave function forms can be optimized successfully for excited states. In ground states, recent years have seen significant improvements in the size and complexity of wave functions that can be treated by VMC wave function optimization, [79, 113, 116, 117, 119, 123, 129] from the development of the linear method (LM) to the adoption of accelerated descent (AD) methods and even the combination of the two. Excited state wave function optimization is less well developed, and variance-based state-specific approaches — which are useful when dealing with high-lying states [198] or in cases where large dipole changes raise concerns for state-averaging [70] — have been shown to face stability issues in some surprisingly simple settings. [91]

To help improve excited-state-specific wave function optimization in VMC, this study extends our hybrid LM/AD optimization approach [119] to a variance-based excited state objective function and tests both its stability and its efficacy for double excitations. As in the ground state, we find that AD provides useful support to the Blocked LM, and that their combination is more statistically efficient than the LM alone. In terms of stability against collapse to other states' variance minima, we find that the difficulty that AD faces in making large changes to the wave function [119] actually prevents it from causing stability

issues, but that the LM part of the algorithm faces the same challenges as have been seen [91] in Newton optimizations of the variance. We therefore focus our stability testing on the LM, and find that achieving stable optimizations is greatly aided by enforcing symmetries, improving the sophistication of the Jastrow factor, and using a sufficiently large sample size to avoid stochastically jumping out of a shallow minimum. In terms of efficacy for double excitations, we find the hybrid approach to be superior to the LM in all cases, usually in terms of statistical efficiency, but in one case also in terms of successfully finding the minimum at all. Comparisons against benchmark methods confirm VMC's accuracy, while a demonstration in a doubly excited state of 4-aminobenzonitrile shows that the approach can be expanded well beyond the reach of sCI methods.

3.3 Theory

Excited-State-Specific Variational Monte Carlo

While VMC has historically been used for the study of ground states, multiple functionals have been developed for the targeting and variational optimization of excited states. We focus on the recently developed functional[94],

$$\Omega(\Psi) = \frac{\langle \Psi \mid (\omega - H) \mid \Psi \rangle}{\langle \Psi \mid (\omega - H)^2 \mid \Psi \rangle} \tag{3.1}$$

which is minimized when Ψ is the eigenstate of lowest energy greater than ω. The formulation of VMC for this excited state functional proceeds analogously to the ground state case. Just as ground state VMC seeks to minimize the expectation value of the energy, written as

$$E(\Psi) = \frac{\langle \Psi \mid H \mid \Psi \rangle}{\langle \Psi \mid \Psi \rangle} = \frac{\int d\mathbf{R} \Psi(\mathbf{R}) H \Psi(\mathbf{R})}{\int d\mathbf{R} \Psi(\mathbf{R})^2} = \frac{\int d\mathbf{R} \Psi(\mathbf{R})^2 E_L(\mathbf{R})}{\int d\mathbf{R} \Psi(\mathbf{R})^2} = \int d\mathbf{R} \rho(\mathbf{R}) E_L(\mathbf{R}) \tag{3.2}$$

using the local energy $E_L(\mathbf{R}) = \frac{H\Psi(\mathbf{R})}{\Psi(\mathbf{R})}$ and probability density $\rho(\mathbf{R}) = \frac{\Psi(\mathbf{R})^2}{\int d\mathbf{R} \Psi(\mathbf{R})^2}$, we can write Ω as

$$\Omega(\Psi) = \frac{\int d\mathbf{R} \Psi(\mathbf{R})(\omega - H)\Psi(\mathbf{R})}{\int d\mathbf{R} \Psi(\mathbf{R})(\omega - H)^2 \Psi(\mathbf{R})} = \frac{\int d\mathbf{R} \rho(\mathbf{R})(\omega - E_L(\mathbf{R}))}{\int d\mathbf{R} \rho(\mathbf{R})(\omega - E_L(\mathbf{R}))^2}. \tag{3.3}$$

However, while the probability distribution ρ has a useful zero-variance property[145], it may be less useful than other importance sampling functions when estimating quantities such as the energy variance and matrix elements used in the optimization algorithms for minimizing Ω. [67–70] Within this work, we use the importance sampling function

$$|\Phi|^2 = |\Psi|^2 + c_1 \sum_i |\Psi^i|^2 + c_2 \sum_j |\Psi^j|^2 + c_3 \sum_k |\Psi^k|^2 \tag{3.4}$$

where c_1,c_2,c_3 are weights on sums of squares of wave function derivatives Ψ^i, Ψ^j, Ψ^k for Jastrow, CI, and orbital parameters, respectively. Effective choices for (c_1, c_2, c_3) may be system-dependent and require some experimentation in practice. In our results, we use $(0.0, 0.0001, 0.0)$ for our stability tests on a model cyanine dye, $(0.0004, 0.0002, 0.0)$ for the carbon dimer, $(0.0001, 0.0001, 0.0)$ for nitroxyl, glyoxal, and acrolein, and $(0.0001, 0.0, 0.0)$ for cyclopentadiene and 4-aminobenzonitrile. We have found these values enabled optimization to lower target function values than we could achieve using $|\Psi|^2$ and that using nonzero c_3 for the orbital parameter derivatives also resulted in poorer target function values. However, our exploration of these importance function parameters was not exhaustive and different values may also result in equally good target function values. The general intuition is to include wave function derivative terms, as importance sampling a linear combination of the current wave function probability density and that of its parameter derivatives allows us to obtain a better statistical estimate of the LM Hamiltonian. Whenever the parameters have large effects on the nodal surface, importance sampling $|\Psi|^2$ will lead to a large statistical uncertainty in this Hamiltonian. Importance sampling $|\Phi|^2$ allows us to better sample areas of parameter space where the current Ψ has negligible probability density, but linear updates to Ψ may not, obtaining a better Hamiltonian estimate and therefore better parameter updates in the process. Because this guiding function also has considerably fewer nodes, it can further help avoid the divergence in the variance of the variance[67, 68] encountered with $|\Psi|^2$ while also keeping the distribution close to $|\Psi|^2$.[69] We also employ clipping of the samples based on the value of the local energy[199, 200]. We compute a deviation of the form $\frac{1}{N}\sum|E_L(\mathbf{R}) - \bar{E}|$ where $\bar{E}$ is the mean local energy of all initial N samples. Samples with local energies more than 5 times the deviation from $\bar{E}$ are discarded and only the remainder are used in computations for the VMC optimization. We find that clipping can improve optimization performance, especially as more flexible ansatzes with larger numbers of parameters are considered, by reducing the statistical uncertainty in the matrix elements and derivatives used by the algorithms. Essentially, this guards against large parameter updates.

The use of Ω for excited states also requires careful handling of the input ω. For a generic fixed choice of ω, Ω will not be size-consistent and so ω must be updated to transform Ω into state-specific variance minimization, which is size-consistent.[95] There are multiple strategies for achieving this transformation to have both state-specificity and size-consistency, such as a linear interpolation between the initial fixed value of ω and the floating value of $E - \sigma$,[91, 95] or, as in this work, a series of fixed-ω optimizations with ω updated between each one until self-consistency between it and $E - \sigma$ is reached. The details of how ω is varied in a calculation is one potential source of instability in the optimization as the target function being minimized with respect to wave function parameters is now changing and there is the possibility of slipping outside the basin of convergence, particularly if ω is varied rapidly.

Linear Method

The LM[112–115] is based on a first order Taylor expansion of the wave function

$$\Psi(\mathbf{p}) = \Psi_0 + \sum_i \Delta p_i \Psi_i \tag{3.5}$$

using $\Psi_i = \frac{\partial \Psi(\mathbf{p})}{\partial p_i}$ for first order parameter derivatives of the wave function and Ψ_0 for the wave function at the current parameter values $\mathbf{p}$. Seeking to minimize the target function Ω with respect to $\mathbf{p}$ leads to the generalized eigenvalue problem

$$(\omega - \mathbf{H})\,\mathbf{c} = \lambda\,(\omega - \mathbf{H})^2\,\mathbf{c} \tag{3.6}$$

which can be solved to produce an update vector $\mathbf{c} = (1, \Delta\mathbf{p})$. The matrices are constructed in the basis of the initial wave function and its first order parameter derivatives, so we have matrix elements of the form

$$\langle \Psi_i \,|\, \omega - H \,|\, \Psi_j \rangle \tag{3.7}$$

and

$$\left\langle \Psi_i \,\middle|\, (\omega - H)^2 \,\middle|\, \Psi_j \right\rangle. \tag{3.8}$$

These matrix elements are evaluated within VMC using parameter derivatives of the wave function and the local energy, and we have found that employing the modified guiding function $|\Phi|^2$ can be crucial for obtaining accurate estimates and effective LM optimizations. In practice, the Hamiltonian matrix of the LM is modified with the addition of shift matrices[119, 150] which help prevent incautiously large parameter changes in the optimization and can noticeably influence the LM's stability and performance. Our implementation of the LM uses by default an adaptive scheme where three sets of shift values are used to produce candidate parameter updates and a correlated sampling procedure is used to determine which one, if any, should be taken to improve the target function value. This approach allows the value of the shifts to vary over the course of the optimization, with the aim of allowing the LM to safely take large steps in parameter space early on while automatically becoming more cautious when close to the minimum, where statistical uncertainty will eventually prevent steps from being able to resolve downhill.

Much recent work on VMC for excited states[69, 70, 87, 88, 94, 142, 198, 201] has relied on the LM for the task of wave function optimization. However, the LM has multiple limitations, particularly a memory cost that increases with the square of the number of optimizable parameters, and applications with more than 10,000 parameters are rare. The quadratic growth in the number of LM matrix elements exacerbates the nonlinear bias of the LM and eventually leads to the underdetermined regime where there are too few samples to effectively estimate the LM matrices and the step uncertainty is large. To help address these issues, a variant of the LM known as Blocked LM has been developed[129] that divides the parameter set into N_b blocks and performs a LM-style matrix diagonalization for each block. Some number N_k of the eigenvectors from each block are retained and combined together along with N_o other directions in parameter space for a second LM diagonalization to obtain

an update in the full parameter space. The lower dimension of the matrices constructed by the Blocked LM alleviates the issues faced by the standard LM at the cost of having to run over the samples twice instead of only once. Further details on the Blocked LM can be found in the original publication[129] and we describe our choices on the number of blocks and retained directions in the supplementary material. A recent study[119] of optimization methods for ground state optimization indicates that tighter and more statistically efficient convergence can be obtained through a hybrid combination of the Blocked LM and AD than when using the LM alone.

Hybrid Optimization

There are multiple flavors of AD optimization methods that require only first order parameter derivatives. These methods are widespread in the machine learning community and have been increasingly used in the context of VMC.[84, 116–119] They offer some appealing advantages compared to the LM, including a memory cost linear in parameter number, a lower per-iteration sampling cost to estimate derivatives, and a reduced nonlinear bias from the stochastic evaluation of those derivatives. However, comparisons[119] between AD methods and the LM indicate that the former may struggle to reach the minimum in parameter space at comparable computational effort, especially when the wave function contains many challenging nonlinear parameters.

An alternative approach that shows the potential to benefit from the strengths of both classes of methods is to take a hybrid combination that alternates between them.[119] Sections of gradient descent optimization can be used to identify important directions in parameter space, which can be provided as input to the Blocked LM to help it account for coupling between blocks of parameters. The inclusion of Blocked LM steps allows the hybrid method to more efficiently reach the vicinity of the minimum, while the sections of AD enable tighter convergence by correcting poor parameter updates by the Blocked LM due to step uncertainty. In addition, the use of AD and Blocked LM enables the optimization of larger parameter sets that are beyond the reach of the standard LM. In energy minimization we found it especially useful to follow the hybrid optimization with a final section of optimization using only descent, which has been found to more efficiently achieve lower final statistical uncertainties than optimization based solely on the LM.[119]

The hybrid scheme can be used with any of the variety of AD methods, but in this work, we use a combination of Nesterov momentum with the RMSprop algorithm[116] that was found to work well for energy minimization. It is specified by the following recurrence relations.

$$p_i^{k+1} = (1 - \gamma_k e^{-(\frac{1}{d})(k-1)}) q_i^{k+1} - \gamma_k e^{-(\frac{1}{d})(k-1)} q_i^k \tag{3.9}$$

$$q_i^{k+1} = p_i^k - \tau_k \frac{\partial \Omega(\mathbf{p})}{\partial p_i} \tag{3.10}$$

$$\lambda_0 = 0 \qquad \lambda_k = \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\lambda_{k-1}^2} \qquad \gamma_k = \frac{1 - \lambda_k}{\lambda_{k+1}} \tag{3.11}$$

$$\tau_k = \frac{\eta}{\sqrt{E[(\frac{\partial \Omega}{\partial p_i})^2]^{(k)}} + \epsilon} \tag{3.12}$$

$$E[(\partial\Omega)^2]^{(k)} = \rho E\left[\left(\frac{\partial \Omega}{\partial p_i}\right)^2\right]^{(k-1)} + (1-\rho)\left(\frac{\partial \Omega}{\partial p_i}\right)^2 \tag{3.13}$$

Equations 3.9 through 3.11 describe how a parameter p_i^k of the wave function on step k of the optimization is updated using knowledge of both the current and previous values of the target function derivatives. The step size τ_k is adaptively adjusted using a running average of the square of parameter derivatives according to equations 3.12 and 3.13, where ρ sets the relative weighting between the past average $E\left[\left(\frac{\partial \Omega}{\partial p_i}\right)^2\right]^{(k-1)}$ and the current value $\left(\frac{\partial \Omega}{\partial p_i}\right)^2$. The quantities d,η, ρ, and ϵ are hyperparameters whose values are chosen by the algorithm's user. For all our results, we have used $d = 100$, $\rho = 0.9$ and $\epsilon = 10^{-8}$. The value of η sets an overall scale for step size and can have a significant influence on optimization performance depending on the choices made for different types of parameters. Our choices for η are discussed in the supplementary material.

Optimization Stability

When working with an exact ansatz and an infinite sample size, any non-degenerate Hamiltonian eigenstate possesses its own variance minimum, [93] and an optimization with a guess sufficiently close to that minimum will be stable. In practice, sample sizes are finite, and the use of an approximate ansatz may lead some states' variance minima to be artificially shallow or to disappear entirely, and so there is a real possibility that state-specific variance minimization will be unstable. [91] If the minimum has indeed disappeared, then the only remedy is to improve the ansatz quality, as no amount of statistical precision will allow an optimization to find a minima that is not present. On the bright side, advances in VMC trial functions[74–77, 99, 102, 107, 108, 202] offer a strong toolkit for improving ansatz quality, albeit one whose use can increase optimization difficulty by increasing the number of variational parameters. If anything, this reality further motivates the development of optimizers, like the hybrid approach tested here, that are designed to handle large, challenging trial functions. In Section 3.4, we will explore the issue of optimization stability by starting with a very simple ansatz for which some states lack variance minima and then making improvements, either by enforcing symmetries or by enhancing the ansatz in order to make the variance minima re-emerge.

If the variance minimum is present but shallow, either due to an approximate ansatz or simply to a near-degeneracy in the spectrum, then the nature of the optimization update steps, and especially their statistical uncertainty, will determine whether an optimization is

stable. Strictly speaking, finite-sample-size VMC optimizations always contain at least some risk of selecting a step in the tail of the distribution of steps that, if taken, will move the optimization out of the basin of convergence for one variance minimum and into the basin for another. To help guard against this issue, a common practice in VMC, and one that we do use for our LM steps, is to verify on a new independent sample (usually using correlated sampling) that the step about to be taken does indeed lower the objective function. Even without statistical uncertainty, step size control is important in nonlinear optimization, an issue that is typically addressed by trust radius methods that guard against single steps making overly large parameter changes. [203] In the LM, one approach to this issue is a diagonal shift, [113] and our LM implementation employs both this shift and an overlap-based shift [150] for these purposes. In Section 3.4, we will see an example of an optimization that only becomes stable with a large enough sample size due to a shallow minimum created by a near-degeneracy.

Wave Functions

For demonstrating the hybrid method's effectiveness within our overall QMC methodology for excited states, we consider multiple types of parameters in our trial wave functions. Our ansatz is the Multi-Slater Jastrow wave function, which has the form

$$\Psi = \psi_{MS}\psi_J \tag{3.14}$$

$$\psi_{MS} = \sum_{i=0}^{N_D} c_i D_i \tag{3.15}$$

$$\psi_J = \exp\left\{ \sum_i \sum_j \chi_k(|r_i - R_j|) + \sum_k \sum_{l>k} u_{kl}(|r_k - r_l|) \right\} \tag{3.16}$$

where D_i are Slater determinants with coefficients c_i and the Jastrow factor ψ_J is constructed from splines that make up the functions χ_k and u_{kl} for the respective one- and two-body terms.[150] The MSJ ansatz is a common choice in QMC, but it can be augmented further to describe more correlation at the price of a more challenging optimization. Two means for doing so are to add a more complicated Jastrow factor and to optimize molecular orbital shapes.

A variety[74, 77, 93, 96–103] of many-body Jastrow factors have been considered for improving QMC ansatzes, but in this work we will limit ourselves to adding a three-body term and a number-counting factor. The three-body term is constructed from polynomials of interparticle distances and can be written in the following form.[77]

$$u(r_{\sigma I}, r_{\sigma' I}, r_{\sigma\sigma'}) = \sum_{l=0}^{M_{eI}} \sum_{m=0}^{M_{eI}} \sum_{n=0}^{M_{ee}} \gamma_{lmn}\, r_{\sigma I}^{l} r_{\sigma' I}^{m} r_{\sigma\sigma'}^{n} \,(r_{\sigma I} - \frac{r_c}{2})^3 (r_{\sigma' I} - \frac{r_c}{2})^3\, \Theta(r_{\sigma I} - \frac{r_c}{2}) \Theta(r_{\sigma' I} - \frac{r_c}{2}) \tag{3.17}$$

The maximum polynomial orders are set by M_{eI} and M_{ee} for the electron-ion and electron-electron distances respectively. The γ_{lmn} are the set of optimizable parameters in this polynomial and are subject to constraints for ensuring the Jastrow satisfies symmetry under exchange and cusp conditions, which can be found in the original publication.[77] Finally, the Theta functions require the three-body term become zero for electron-ion distances more than half a chosen cutoff distance r_c (10 bohr in our case).

Number-counting Jastrow factors are a recently developed[107, 108] ansatz component and can be thought of as a many-body Jastrow factor in real space that aims to recover both strong and weak correlation. They are based on a Voronoi partitioning of space, where the population of electrons in each region, N_I, is given by a sum of local counting functions C_I at each electron coordinate.

$$N_I = \sum_i C_I(\mathbf{r_i}) = \sum_i \frac{g_I(\mathbf{r})}{\sum_j g_j(\mathbf{r})} \tag{3.18}$$

where

$$g_j(\mathbf{r}) = \exp\big((\mathbf{r}-\boldsymbol{\mu})^T \mathbf{A} (\mathbf{r}-\boldsymbol{\mu}) + K\big) \tag{3.19}$$

are Gaussian basis functions about a center $\boldsymbol{\mu}$. With these populations N_I, we can construct the Jastrow factor

$$\psi_C = \exp\left(\sum_{IJ} F_{IJ} N_I N_J + \sum_K G_K N_K\right) \tag{3.20}$$

which can be tacked on to our overall expression for Ψ. The F_{IJ} and G_K are variational parameters though the latter is in practice eliminated with a basis transformation of the region populations and so we only optimize F_{IJ}.[108]

State-specific orbital optimization is useful for obtaining accurate VMC results on particular excited state phenomena including charge transfer[70] and core excitations[198] and can avoid some of the pitfalls of state-averaged approaches.[45] Recent work[75, 76] with the table method has enabled the efficient calculation of orbital parameter derivatives in MSJ wave functions even for large expansions and we refer the reader to the original publications for details. However, we do note that, despite these advances, orbital optimization remains the most challenging part of the optimization, likely due to its inherently high degree of nonlinearity and the fact that it alters the nodal surface. One potential alternative would be to obtain state-specific orbitals from another method[45] while optimizing only Slater coefficients and Jastrow factors, and we make a preliminary exploration of this idea in our results.

Variance Matching

While optimization of the parameters improves the absolute quality of the wave functions, the results are still approximate due to the limited ansatz and accurate determination of excitation energy differences requires cancellation of errors. This relies on a balanced treatment of both the ground and the excited state, which we attempt to obtain using variance

matching. As shown in previous work,[69, 70] this approach improves predicted excitation energies by optimizing ansatzes of different CI expansion lengths for the two states so that their variances are approximately equal. To facilitate interpolation, the variances for a series of excited state calculations at different expansion lengths can be fit to an analytic form such as the power law decay

$$\sigma^2(N) = c + \frac{d}{N^\alpha} \tag{3.21}$$

to determine parameters, c, d, and α. For a given expansion length for the ground state and a resulting variance σ_g^2, we can estimate the expansion length N^* that will yield a matching variance for the excited state and take the corresponding energy when computing our prediction for the excitation energy.[69] In practice, some additional varying of N by hand can be performed to find an explicit variance match. All our reported excitation energies were obtained using this explicit variance matching procedure.

3.4 Results

Computational Details

All our VMC calculations used an implementation of the described optimization algorithms within a development version of QMCPACK.[150, 204] A recently developed set of pseudopotentials[171] and associated basis sets were used for all molecules. For constructing our ansatzes, we have employed Molpro[170] and PySCF[205] for CASSCF calculations to generate Slater determinant expansions. In one case, we instead use CASSCF to provide orbitals for a selected CI calculation in Dice.[10, 131] CASPT2 calculations in Molpro[170] were used alongside other methods' literature values in benchmarking our VMC results. Specific active space and basis set choices are given in each system's section. Molecular geometries, absolute energies for our doubly excited state calculations, and additional optimization details can be found in the supplementary material.

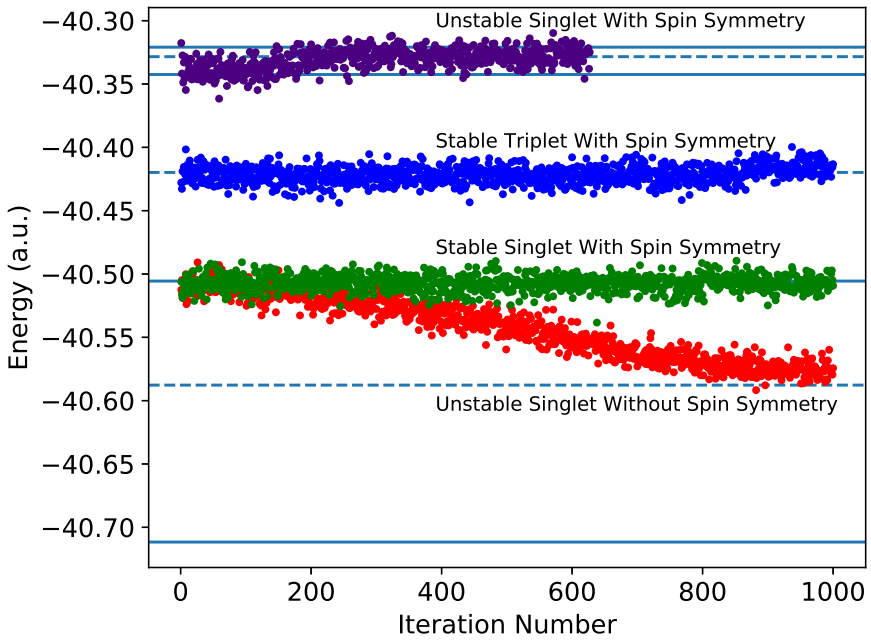


Figure 3.1: Examples of unstable and stable LM optimizations of Ω for CN5. We consider the singlet near -40.5 E_h, which is dominated by a HOMO to LUMO excitation, as well as the next triplet and singlet above it, which are dominated by a HOMO-2 to LUMO excitation. Initial wave function guesses were obtained from a diagonalization of the Hamiltonian in a 26 determinant space with pre-optimized one- and two-body Jastrow factors present. These Jastrows were optimized further along with the 26 determinant coefficients. For each state, the value of ω was set to the E-σ value from a one million sample evaluation using the initial wave function and held fixed throughout the optimization. The LM shifts were kept constant at 0.1 for the diagonal shift and 1 for the overlap-based shift throughout the optimizations with 50,000 samples per iteration in all cases. For the unstable cases, we found that using increased sampling effort alone did not ensure stability. A proposed parameter update was rejected if our correlated sampling assessment predicted it would raise the target function value, but all optimizations contain hundreds of accepted steps. Horizontal lines show the lowest 7 eigenenergies from our diagonalization, with solid lines for singlet states and dashed for triplets.

Stability in a Model Cyanine Dye

To explore the stability of Ω-based variance minimization, we have performed a series of LM tests on the model cyanine dye $C_3H_3(NH_2)_2^+$ (denoted hereafter as CN5) in which Filippi and coworkers discovered variance optimization instabilities for some choices of the trial function. [91] In particular, they showed that while a CAS(6e,5o) wave function was

stable, other ansatze from CAS(6e,10o) and CIPSI were not. They also showed that a small CSF expansion, constructed of HOMO-to-virtual excitations of B_1 symmetry, was especially prone to optimize to a different state than the one targeted by the initial guess due to the absence of variance minima in this simpler case. Here we perform similar tests to study the absence of variance minima, confirming that instabilities are present when the sample size is small and the trial function is simple, but also demonstrating that improvements in the trial function and sampling effort can overcome these instabilities. To start, we note that virtual orbitals outside the CASSCF active space are often not physical in their shapes, and so for the states we target, we make sure that the primary orbitals involved are within the (6e,5o) CASSCF active space. In this active space, we performed an equal-weight state-averaged CASSCF optimization of the lowest four B_1 singlet states using an aug-cc-pVDZ basis set in Molpro[170], after which we imported all determinants with weights above 0.05 in any state into our VMC ansatz. This resulted in a 26-determinant ansatz, or, in the cases where we enforced singlet spin symmetry in VMC, a 13-CSF ansatz. Note that this trial function differs from the one used previously, [91] which is intentional, as the previous approach examined some states whose dominant orbitals were virtual in the quantum chemistry calculations and so may not have been as well optimized as orbitals containing electrons. Here, we try to ensure that orbital quality is balanced between states by ensuring that the dominant orbitals of the states we test are within our CASSCF active space. Nonetheless, we can still find optimization instabilities when using this simple ansatz, which confirms that ansatz components beyond a small determinant expansion can be necessary to achieve stable variance optimization.

As seen in Figure 3.1, the stability of an optimization using our 26-determinant ansatz depends on which state is being targeted and in some cases on whether or not spin symmetry is enforced. First, consider the optimization that guesses the singlet near -40.5 E_h and then collapses to the triplet below it. This failure shows that, at least at this level of statistical resolution, the singlet in question lacks a local variance minimum in the variable space of the 26 different determinant coefficients. However, after enforcing spin symmetry by instead optimizing the 13 coefficients for the singlet CSFs, the variance minimum for this state reappears. Next, we note that when we optimize the 26 determinant coefficients starting with a guess for the triplet just below -40.4 E_h, the optimization is stable. Thus, in these cases, moving to a more sophisticated ansatz is not necessary so long as symmetries are enforced. The same cannot be said for the most difficult case we consider, in which a guess for the singlet just above -40.35 E_h is unstable and optimizes to a higher singlet even when spin symmetry is enforced. With no more symmetries to make use of, we must conclude that variance minimization is not stable for this state when using this simple ansatz.

In principle, variance minimization should become stable as the ansatz is improved and the sample size increased. The question, of course, is whether stability can be achieved in practice, especially given that overly aggressive embellishments of the ansatz may lead to an impossibly difficult optimization. First, let us improve the ansatz by adding the 3-body Jastrow factor introduced by Needs and coworkers[77], which we first optimize for the guessed CSF expansion with the CSF coefficients held fixed before finally optimizing

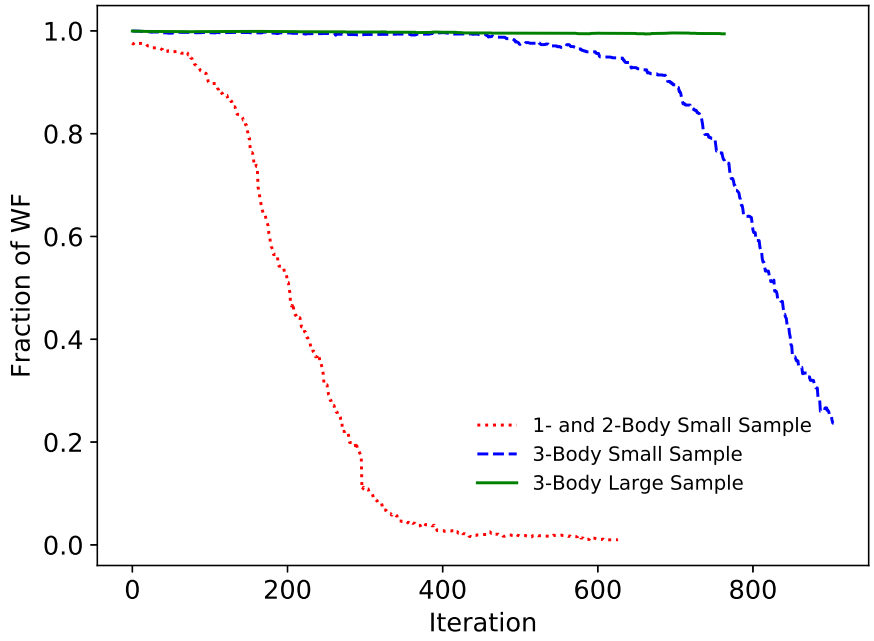


Figure 3.2: Fractions of the CI expansion corresponding to the third singlet's dominant CSF over the course of LM optimizations. When starting from the energy eigenstate with only 1- and 2-body Jastrows are present, the weight of this CSF quickly collapses as the optimization drifts to a higher singlet. This optimization used 50,000 samples per iteration and starting from solely the dominant CSF with an additional 3-body Jastrow delayed, but did not prevent, the drift at this sampling effort. Upon increasing the sample size to 500,000 samples per iteration, the 3-body Jastrow wave function optimization became stable, as shown by the solid line. Note that the CI expansion fraction is simply the sum of the squares of the determinant coefficients within the given CSF divided by the sum of all the squared determinant coefficients.

all variables together. As seen in Figure 3.2, which shows the key CSF weights during the final optimization in which all parameters are free to vary, this ansatz improvement does lead to a stable optimization, but only when a larger sample size is used, suggesting that the variance minimum is now present but shallow. Given how close in energy this state is to the singlet state above it (about 15 mE_h), a shallow minimum makes sense, which is a good reminder that optimization objective functions based on the energy and variance are at a disadvantage when states are nearly degenerate. For low-lying states, state-averaged energy minimization can often deal with this type of situation, [91] but near-degeneracies remain a challenge for higher-lying states or in situations where state-averaging introduces its own challenges. In future, it may therefore be worth considering ways to involve other

properties in the objective function, possibly via a VMC analogue of a generalized variational principle. [48] Even with this near-degeneracy, which we stress involves states significantly closer together in energy than any of those tested in the small-determinant-expansion case in the previous study, [91] we do see that a modest improvement in ansatz quality produces a stable Ω-based LM optimization.

While this finding is reassuring, we must stress that difficult optimization cases remain a serious challenge. For example, Filippi and coworkers found [91] that highly sophisticated wave functions derived from a (6,10) active space faced optimization instabilities even though wave functions derived from a smaller (6,5) active space did not, showing that in practice it is not always easy to predict when these instabilities will arise. This in mind, we have checked carefully for signs of optimization instability in the double excitations we now turn to, and while we did not observe any such issues in these states, finding a more complete resolution to this problem is clearly an important direction for future research.

Carbon Dimer

For our doubly excited state applications, we first consider the carbon dimer, a very heavily studied system for the testing and development of theoretical methods[41, 55, 206, 207]. We use the $2\ ^1\Sigma_g$ state, which is characterized by a HOMO to LUMO squared double excitation, as a simple starting test case for validating the hybrid method's results against the LM's and assessing the accuracy of VMC. In this case, selected CI methods are able to achieve millions of determinants and provide high quality benchmark data[41, 208] on the vertical excitation energy that we can use to assess our results with more compact wave functions. We consider only an equilibrium bond length of 1.248 Å.

To construct our ansatzes, we use a (8e,8o) CASSCF calculation in Molpro[170] with a carbon pseudopotential and the corresponding cc-pVTZ basis.[171] The resulting CI expansions are used in our variance matching procedure and we consider cases both for standard MSJ ansatzes using only one- and two-body Jastrows, and with an additional NCJF and orbital optimization. The NCJF was produced using a set of 16 counting regions composed of 8 octants for each carbon atom.[108]

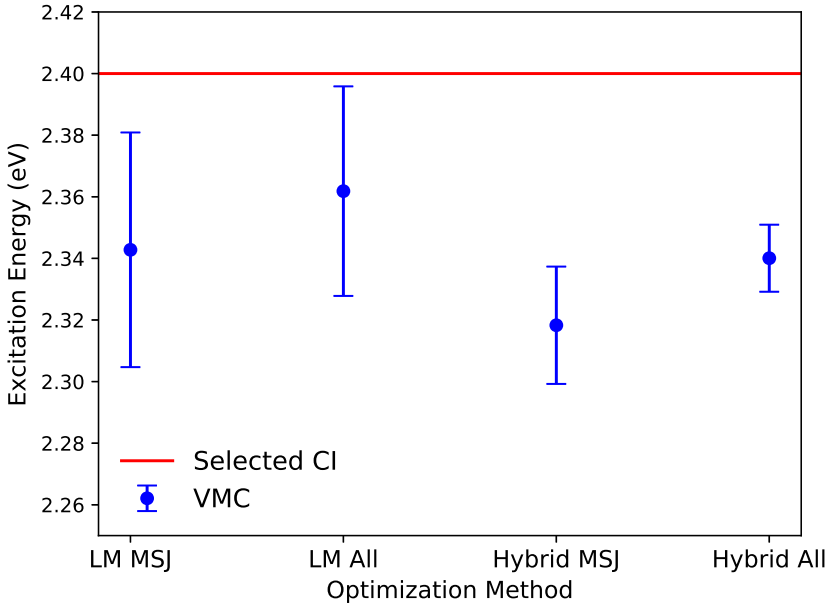


Figure 3.3: Excitation energy of the 2 $^1\Sigma_g$ state in C_2 for LM and hybrid method on both MSJ and all parameter ansatzes. See also Table 3.1. 20 determinants were used for all ground state optimizations. The excited state optimizations used 23 determinants for both the LM and Hybrid MSJ cases, 71 for the LM All case, and 50 for Hybrid All case. The benchmark value is taken from a selected CI calculation using the CIPSI algorithm.[41]

Table 3.1: Excitation energies and uncertainties for the 2 $^1\Sigma_g$ state in C_2. Literature values for selected CI, CCSDT, CC3, and CASPT2 using an aug-cc-pVQZ basis set are included for comparison.[41]

Method	Excitation Energy (eV)	Uncertainty (eV)	Total Samples
LM MSJ VMC	2.34	0.04	100,000,000
LM All VMC	2.36	0.03	100,000,000
Hybrid MSJ VMC	2.32	0.02	138,000,000
Hybrid All VMC	2.34	0.01	138,000,000
Selected CI	2.40		
CCSDT	2.87		
CC3	3.24		
CASPT2	2.50		

Figure 3.3 shows the predicted excitation energy achieved by the LM and the hybrid method on both the simpler MSJ ansatzes and on all parameter wave functions that include

the NCJF and orbital rotations. It is reassuring to find that the hybrid method's results agree with the LM's to within statistical uncertainty. In terms of accuracy, the VMC results are within about 0.05 eV of the selected CI value with the all parameter optimizations offering some improvement over the MSJ results. This has been achieved with very modest CI expansions, using less than 100 determinants in all cases, compared to the 5 million used to produce the benchmark energy.[41]

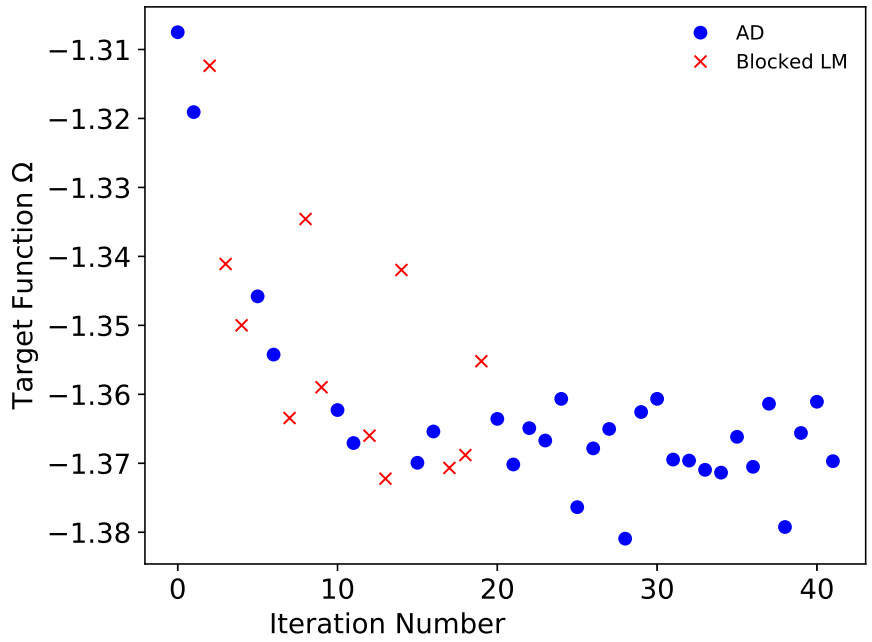


Figure 3.4: Example optimization of target function Ω for all parameters using the hybrid method. Points for AD correspond to an average over 50 iterations while those for the Blocked LM are individual iterations.

We also explore the potential benefits of the hybrid approach over a pure LM optimization. Figure 3.4 shows an example optimization by the hybrid method of all parameters including NCJF and orbitals. With these more challenging parameters, the Blocked LM is more prone to step uncertainty and upward fluctuations of the target function, which in this case we see the AD sections correct. The hybrid method also exhibits a statistical advantage over the LM based on the uncertainty and sampling data in Table 3.1. For slightly greater sampling effort, the hybrid method provides about a factor of 3 improvement in the uncertainty on the excitation energy, which would require a factor of 9 more samples to obtain solely with the LM. This advantage in computational efficiency persists in our results for more difficult systems.

Nitroxyl, Glyoxal, and Acrolein

One nuance to the study of doubly excited states is distinguishing between different types of such states. For instance, the state we have considered in the carbon dimer is multi-determinantal while some other molecules' states can be viewed as single reference, enabling higher order coupled cluster[41] or orbital optimized DFT[209] to obtain accurate excitation energies. The work of Loos and coworkers has categorized a set of doubly excited states using the percentage of singles amplitudes in CC3[41] and we briefly consider several of the same systems to compare our methodology against their benchmark results. Specifically, we consider the $2\ ^1\mathrm{A}'$ state in nitroxyl, the $2\ ^1\mathrm{A}_g$ state in glyoxal, and the $3\ ^1\mathrm{A}'$ state of acrolein. These systems exhibit some of the diversity of doubly excited states, with acrolein's state containing a high percentage of single excitations, while nitroxyl and glyoxal have almost none. For all systems, we use cc-pVTZ basis sets with pseudopotentials[171] and generate determinants for MSJ ansatzes from CASSCF calculations in Molpro.[170] Our active spaces are (12e,9o) for nitroxyl, (8e,6o) for glyoxal, and (10e,10o) for acrolein. Each CASSCF calculation was state-averaged over four singlet states.

The excitation energy predictions from the LM and the hybrid method are shown in Figures 3.5 through 3.7 with precise values given in Tables 3.2 through 3.4. We compare our results to the theoretical best estimates (TBE) from Loos and coworkers as well as their coupled cluster and CASPT2 values.[41] For this trio of systems, we find that we can obtain good accuracy with very modest wave functions and consistent results between the LM and hybrid method. For nitroxyl, we come within 0.03 eV of the TBE, while our excitation energies in glyoxal and acrolein are within about 0.2 eV. This level of accuracy outperforms the coupled cluster approaches as well as some versions of CASPT2 in the case of acrolein. As in the carbon dimer, these calculations use less than 100 determinants in all cases compared to the millions used in the benchmark selected CI calculations, which are restricted to smaller basis sets in glyoxal and acrolein. We note that our methodology allows for further systematic improvement through more sophisticated ansatzes and in order to consider its performance on larger systems beyond the reach of selected CI, we now turn to some molecules outside the benchmark set of Loos and coworkers.

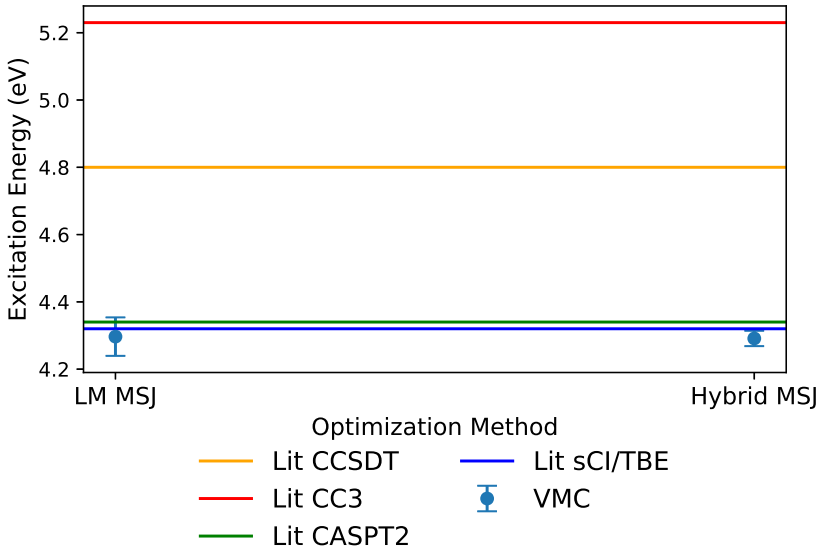


Figure 3.5: Excitation energy of the 2 ^{1}A$'$ state in nitroxyl for LM and hybrid method on MSJ ansatzes. See also Table 3.2. Both methods used 20 determinants the ground state and 40 for the excited state. Reference values for coupled cluster, CASPT2, and selected CI are taken from the work of Loos and coworkers.[41] They take the selected CI value as the theoretical best estimate.

Table 3.2: Excitation energies and uncertainties for the 2 ^{1}A$'$ state in nitroxyl. The included literature values for CCSDT, CC3, CASPT2, and selected CI all use an aug-cc-pVQZ basis set.[41]

Method	Excitation Energy (eV)	Uncertainty (eV)	Total Samples
LM MSJ VMC	4.30	0.06	100,000,000
Hybrid MSJ VMC	4.29	0.01	138,000,000
TBE/Selected CI	4.32		
CCSDT	4.8		
CC3	5.23		
CASPT2	4.34		

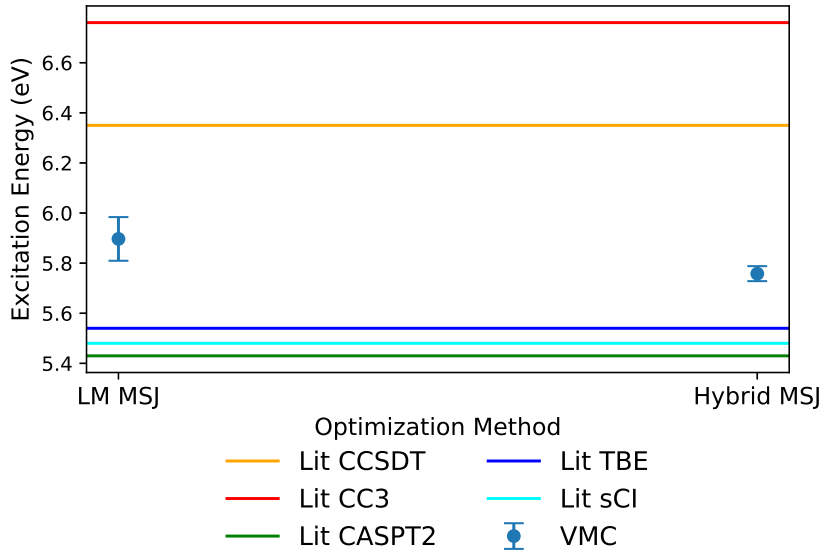


Figure 3.6: Excitation energy of the 2 ^1A_g state in glyoxal for LM and hybrid method on MSJ ansatzes. See also Table 3.3. Both methods used 50 determinants the ground state and 20 for the excited state. Reference values for coupled cluster, CASPT2, and selected CI are taken from the work of Loos and coworkers.[41] Their theoretical best estimate is obtained by adding a 0.06 eV basis set correction to the selected CI result.

Table 3.3: Excitation energies and uncertainties for the 2 ^1A_g state in glyoxal. The included literature values use an aug-cc-pVDZ basis set for selected CI and aug-cc-pVQZ for CCSDT, CC3, and CASPT2.[41]

Method	Excitation Energy (eV)	Uncertainty (eV)	Total Samples
LM MSJ VMC	5.90	0.09	140,000,000
Hybrid MSJ VMC	5.75	0.03	138,000,000
TBE	5.54		
Selected CI	5.48		
CCSDT	6.35		
CC3	6.76		
CASPT2	5.43		

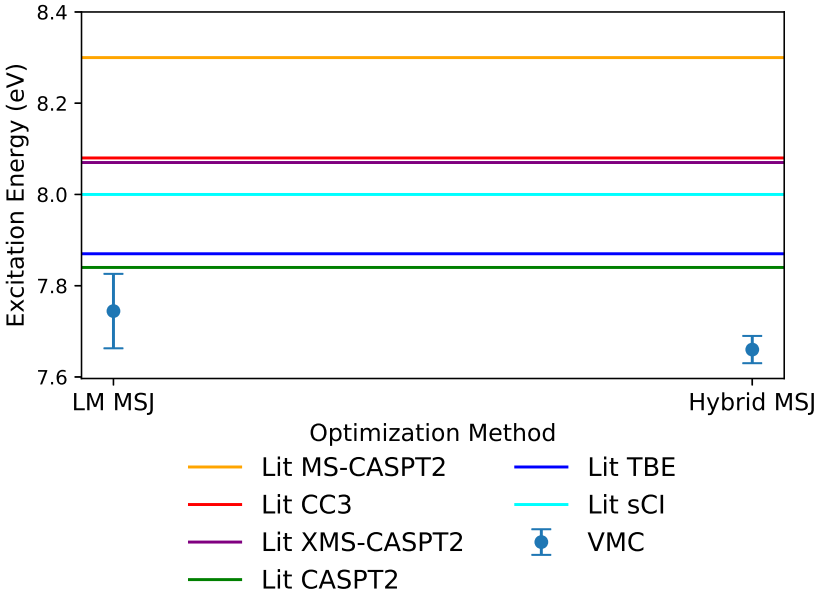


Figure 3.7: Excitation energy of the 3 ^{1}A$'$ state in acrolein for LM and hybrid method on MSJ ansatzes. See also Table 3.4. Both methods used 20 determinants the ground state and 70 for the excited state. Reference values for coupled cluster, CASPT2, and selected CI are taken from the work of Loos and coworkers.[41] Their theoretical best estimate is obtained by adding a -.13 eV basis set correction to the selected CI result.

Table 3.4: Excitation energies and uncertainties for the 3 ^{1}A$'$ state in acrolein. The included literature values use a 6-31+G(d) basis set for selected CI, an aug-cc-pVTZ basis CC3, and an aug-cc-pVQZ basis for the three versions of CASPT2.[41]

Method	Excitation Energy (eV)	Uncertainty (eV)	Total Samples
LM MSJ VMC	7.74	0.08	140,000,000
Hybrid MSJ VMC	7.66	0.03	138,000,000
TBE	7.87		
Selected CI	8.00		
CC3	8.08		
CASPT2	7.84		
MS-CASPT2	8.3		
XMS-CASPT2	7.84		

Cyclopentadiene

For a more challenging test of our methodology, we consider the doubly excited $3\ ^1A_1$ state of cyclopentadiene (CPD). This state has been repeatedly studied in theoretical benchmark investigations[190, 191, 210–213] and in some experimental investigations[214, 215]. As before, we construct multi-Slater wave functions and add traditional 1 and 2-body Jastrow factors. For this larger molecule, we use the heatbath selected CI (HCI) method[10, 131] in the Dice code to produce our CI expansions by correlating 26 electrons in the lowest 46 orbitals from a (6e,5o) CASSCF in Molpro with pseudopotentials and cc-pVTZ basis sets.[171]

In this system, we find that the LM fails to optimize the Ω functional as well as the hybrid method and leads to an inferior energy prediction. However, for relatively simple variance-matched multi-Slater Jastrow ansatzes of 20 and 500 determinants for the ground and excited states respectively, the hybrid method is able to achieve an excitation energy within about 0.1 to 0.2 eV of CASPT2 as seen in Figure 3.8 and Table 3.5.

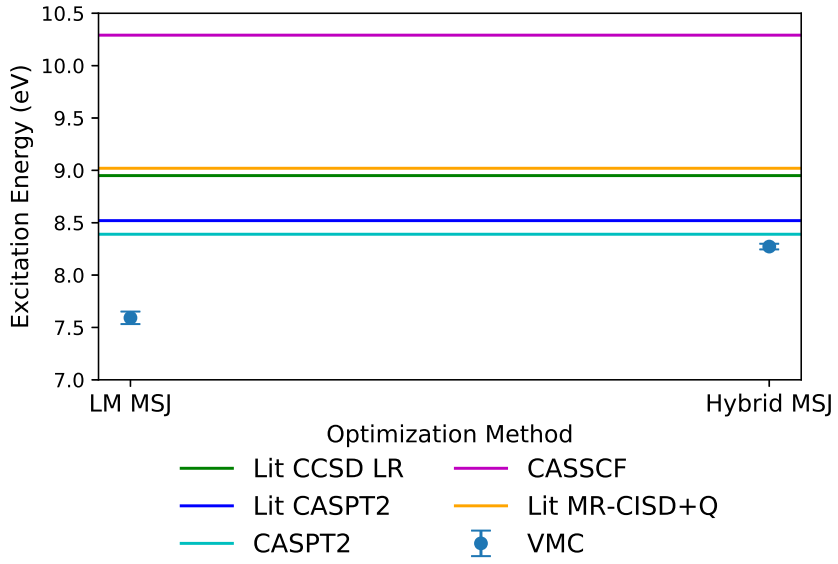


Figure 3.8: Excitation energy of $3\ ^1A_1$ state in CPD for LM and hybrid method. See also Table 3.5. Both the LM and hybrid optimizations used 20 and 500 determinants for the ground and excited states, respectively. Literature values are taken from Schreiber et al.[190] for CASPT2 and CCSD LR and from Shen and Li[212] for MR-CISD+Q.

Table 3.5: Excitation Energies and uncertainties for 3 $^1A_1$ state in CPD.

Method	Excitation Energy (eV)	Uncertainty (eV)	Total Samples
LM MSJ VMC	7.59	0.06	140,000,000
Hybrid MSJ VMC	8.27	0.03	210,000,000
CASSCF	10.29		
CASPT2	8.39		
Lit. CASPT2[190]	8.52		
Lit. CCSD LR[190]	8.95		
Lit. MR-CISD+Q[212]	9.02		

To elaborate on the failure of the LM in this case, we present data in Table 3.6 on a head to head comparison of the LM and the hybrid method on the same excited state wave function with a determinant expansion length of 50 and using the same fixed value of ω in the target function. We find that the hybrid method achieves a lower value of the target function, which corresponds to a significantly higher energy for the excited state. This results in its substantially better prediction for the excitation energy as both optimization methods give comparable results for the ground state calculation. We note that extensive experimentation with the technical details of the LM, including choices for the shifts, sampling effort, and guiding function failed to bring it into agreement with the hybrid method. When we started the LM with the hybrid method's *optimized* wave function, it remained at roughly those parameter values with the same target function value and energy as found by the hybrid method. This test indicates that the LM agrees that the hybrid method has found an optimal location in parameter space if it starts close enough, but is apparently unable to find it itself when starting from the unoptimized wave function.

Table 3.6: Head to head comparison of LM and hybrid for CPD on the same excited state wave function at fixed ω.

Method	Energy (a.u.)	Energy Uncertainty (a.u.)	Target Function $\Omega(\Psi)$ (a.u.)	Target Uncertainty (a.u.)
LM MSJ VMC	-31.541	0.0013	-0.747	0.0023
Hybrid MSJ VMC	-31.517	0.0007	-0.752	0.0004

4-Aminobenzonitrile

Our final system, 4-aminobenzonitrile (ABN), has been heavily studied as an example of intramolecular charge transfer(ICT) with many attempts to determine the geometry of the ICT state[216–219]. Here we test our ability to treat a doubly excited state at the ICT geometry. We selected this system for an initial exploration of possible benefits of using state-specific orbitals within VMC while forgoing orbital optimization. These orbitals were obtained from a recent state-specific CASSCF approach[45, 219] that employs a root tracker

based on combination of an excited state variational principle and density matrices. For a twisted geometry of ABN[219] (coordinates in supplementary material), we construct Multi-Slater wave functions from both state-averaged (over four states) and state-specific CASSCF calculations that use a (12e,11o) active space along with pseudpotentials and cc-pVDZ basis sets.[171] Both types of CASSCF calculations were performed in a development version of PySCF.[205] The excited state with double excitation character that we consider appears as the fourth CASSCF state in energy, directly above the ICT state.

Figure 3.9 shows the excitation energies obtained by the LM and hybrid method for the cases where we use state-averaged and state-specific CASSCF orbitals. In this instance, we find that the optimization methods agree with each other and that there is no clear difference between using state-averaged and state-specific orbitals within our VMC ansatzes. There is about a 0.4 eV difference between our VMC results and CASPT2, but in the absence of an experimental result or higher level benchmark, it is not obvious which is more accurate. The agreement between the state-averaged and state-specific VMC excitation energies may not be too surprising given that we also find little difference at the CASSCF level. While this is a null result for the usefulness of state-specific orbitals for this state in ABN, other cases may perform differently, including the ICT state. In terms of optimization, the across the board agreement in ABN offers further evidence that the hybrid method is at least as accurate as the LM, while continuing to provide better statistical efficiency.

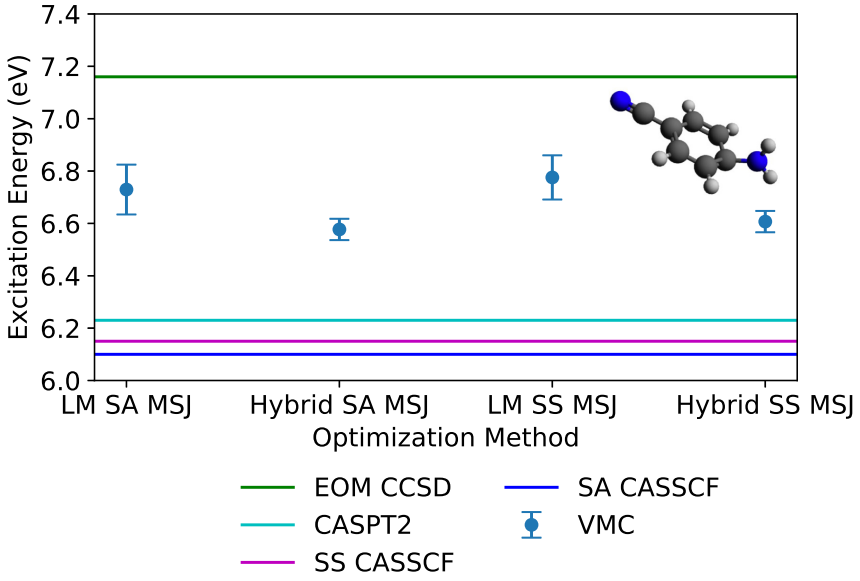


Figure 3.9: Excitation energy of the doubly excited state in ABN for LM and hybrid method. See also Table 3.7. Excitation energies for all other methods were obtained in Molpro. 20 determinants were used in all ground state optimizations. The excited state optimizations used 70, 100, 500, and 300 determinants for the LM SA, Hybrid SA, LM SS, and Hybrid SS cases respectively. These numbers of determinants were chosen to achieve an explicit variance match between the ground and excited states for each method.

Table 3.7: Excitation Energies and uncertainties for ABN.

Method	Excitation Energy (eV)	Uncertainty (eV)	Total Samples
LM SA MSJ VMC	6.73	0.10	140,000,000
Hybrid SA MSJ VMC	6.58	0.04	174,000,000
LM SS MSJ VMC	6.78	0.08	140,000,000
Hybrid SS MSJ VMC	6.61	0.04	192,000,000
SA CASSCF	6.10		
SS CASSCF	6.15		
SA CASPT2	6.23		
EOM-CCSD	7.16		

3.5 Conclusion

We have presented an extension of a hybrid LM/AD optimization approach to the case of excited-state-specific variance optimization and tested its efficacy for doubly excited states. As in energy minimization, we find the hybrid method to be more statistically efficient than the linear method, and, in one case, we were surprised to find it to be more effective at finding the variance minimum. Thanks to VMC's ability to combine a linear combination of determinants (for capturing strong correlation) with sophisticated correlation factors (for weak correlation) and its ability to explicitly balance wave function quality between different states, we find it to be highly accurate compared to theoretical benchmarks in our tests on doubly excited states. As it relies on far more modest determinant expansions than sCI methods, it can also be used to treat both strong and weak correlation in system sizes where capturing both through sCI is not currently possible. We have also performed some simple tests on the stability of state-specific variance minimization and found that using symmetry, increasing sample sizes, and improving the quality of the wave function approximation all play important roles in preventing collapse to other states.

Looking forward, there are multiple avenues for further improvement. Our optimization stability testing was limited to a relatively simple example, and while it does suggest steps that can be taken to alleviate stability concerns, it remains to be seen how effective these steps will be in general. The demonstration by Filippi and coworkers [91] that optimization instabilities can persist even when highly sophisticated CIPSI-based VMC trial functions are employed makes clear that there is more work to be done to resolve this issue. As variance-based approaches are particularly ill-suited to degenerate or near-degenerate states, it would be quite interesting to explore whether generalized variational principles that incorporate properties beyond the energy can be usefully adapted for VMC optimization. Another priority is improving user-accessibility, as the recent improvements in VMC optimization methodology have in many cases brought with them a significant increase in the methodological complexity. Finding ways to robustly automate choices for stability shifts, what balance to strike between linear method and accelerated descent steps, what hyperparameters to choose for the descent, and how to arrange variables into blocks within the blocked linear method would significantly simplify the practical application of these tools. The systematic study of effective importance sampling functions is another supporting step for excited state optimizations, where their use is more crucial than in the ground state. A third improvement would be to automatically stage the optimization of different parameters according to the statistical significance of their gradients, which would allow the noisiest and most difficult parameters to be handled at the end of the optimization without any direct human decision-making. Finally, although VMC optimization is becoming increasingly capable, it will likely be profitable to map out areas where difficult parameters like orbital shapes can be safely kept at their quantum chemistry values, whether from state-averaged or state-specific CASSCF. Given the high accuracy that VMC can offer for very challenging excited states, providing easy-to-use incarnations of the best available VMC optimization methods is a high priority.

Chapter 4

Combining State-Specific Quantum Chemistry and Quantum Monte Carlo for Molecular Excited States

4.1 Abstract

Through a combination of excited-state-specific quantum chemistry, selected configuration interaction, and variational Monte Carlo, we investigate an approach to predicting vertical excitation energies that achieves high accuracy across a broader range of states than either coupled cluster theory or multi-reference perturbation theory. The approach can be employed in settings beyond the current reach of selected configuration interaction or density matrix renormalization group and retains its high accuracy whether it is treating single excitations, double excitations, or charge transfer excitations. To address the different challenges that arise in these types of states, the method combines excited-state-specific orbital relaxations from complete active space self consistent field theory, an extended active space configuration selection, and improved approaches to optimization and variance-balancing within variational Monte Carlo. In addition to testing in smaller molecules where clear theoretical benchmarks exist, we use the method to offer new benchmark values in triple-zeta-or-better basis sets for a doubly excited state of benzene and two states, one of them partially-doubly-excited, in uracil.

4.2 Introduction

The wide variety of different excited states in molecules poses an ongoing challenge for traditional electronic structure methods. Many of the most widely applied and scalable methods may provide satisfyingly accurate predictions for one class of excited states, only to fail on others. When surveying the landscape of electronic structure techniques, this problem of inconsistent performance can be traced to the difficulty of simultaneously satisfying the

multiple requirements for accurate treatment of all classes of excited states. First, a method should have the means of correctly describing multiconfigurational character, which is more common among excited states compared to ground states. At the same time, for quantitatively accurate energies, the method must possess an ability to capture weak correlation effects. This effective treatment of strong and weak correlation should be maintained in a balanced fashion for both ground and excited state in order to obtain correct energy differences, and accounting for orbital relaxations may be necessary for cases where orbital shapes differ significantly between the two states. Any potential method with general-purpose applicability to excited states must also satisfy the above considerations in larger molecules and basis sets. After discussing the individual strengths of major quantum chemistry approaches, we present a methodology that satisfies these criteria and exploits wave function sparsity and state-specificity to deliver consistently accurate excitation energies across a wide range of molecules and types of excited states.

Time-dependent density functional theory (TDDFT) is by far the most commonly used approach for excited states and has played a leading role in the modeling of absorption spectra.[25] However, TDDFT has historically struggled, at least for its most common functional choices within the adiabatic approximation, to describe some classes of excited states accurately, with double excitations and charge transfer (CT) as two notorious examples.[26–29] While excited state DFT methods continue to benefit from extensive research[30, 209, 220–222] analyzing and improving approximate functionals, these categories of excited states remain challenging. One alternative approach is to employ Green's function techniques such as GW-BSE, which can obtain improved excitation energies over TDDFT,[223–225] but these methods in practice rely on an initial set of DFT eigenvalues and eigenstates and are sensitive to this choice of starting point. These challenges and the ever-growing array[4, 5] of possible functional choices motivate the development and use of methods with greater predictive power for difficult categories of excited states.

Among wave function approaches to excited states, coupled cluster theory is a natural initial option to consider, given its prowess for ground state properties and its effective treatment of weak correlation.[22, 226, 227] Excited state extensions of coupled cluster, particularly the equation-of-motion formulation,[35] have become widely used at different levels of cluster truncation. For single valence excitations, limiting the excitation level to singles and doubles with EOM-CCSD obtains balanced results with a systematic error of only a few tenths of an eV. [36] Higher orders can achieve even better accuracy. [37] However, for double excitations, EOM-CCSD produces substantially larger errors than for single excitations[40] and while higher orders of coupled cluster can obtain accurate results in some cases,[184] their errors can still be as large as half an eV or worse for pure double excitations.[41] Besides these inaccuracies, the high scalings of coupled cluster methods, $O(N^8)$ and $O(N^{10})$ for EOM-CCSDT and EOM-CCSDTQ respectively, hinders their application to larger molecules, even for favorable categories of excited states.

Coupled cluster's difficulties have helped motivate widespread use of multireference approaches such as CASPT2 to address doubly excited states.[41, 185–196] While CASPT2 faces its own limitations due to intruder states, sensitivity to active space and state-averaging

choices, and the limiting of dynamic correlation treatment to the PT2 level, it is still capable of obtaining accurate excitation energies, most notably for the pure double excitations where coupled cluster struggles the most.[41] However, for CT states, the situation is reversed, with the use of state-averaging[228] biasing the CASPT2 orbitals to the ground state and producing significant errors.[43–45] In contrast, recent efforts to build accurate benchmark sets for CT states, which are currently less common than for other classes of excited states, have relied upon coupled cluster.[38, 39, 229] This problem of inconsistent performance across the diverse array of possible excited states makes the development of more broadly reliable tools a key priority in electronic structure.

While a number of methods,[8, 230] including selected CI (sCI),[37, 41, 130, 132, 208] density matrix renormalization group (DMRG),[231, 232] and full configuration interaction quantum Monte Carlo (FCIQMC),[55, 56, 233–235] can achieve this reliability with systematic exploitation of wave function sparsity and/or compressibility, their high scaling limits their prospects for extensions beyond small systems. A recent study[230] of benzene is a useful demonstration of the state-of-art capabilities of this collection of high accuracy techniques. The determination of benzene's ground state energy in a double-ζ basis set is both an important example of the molecular sizes that these high-level approaches can now treat and a reminder of the challenges they face in larger systems and the larger basis sets needed for excited state accuracy. So far, the role of sCI, DMRG, and FCIQMC in larger molecules beyond benzene has been limited to mostly serving as (impressively large) active space solvers.[12, 19, 236–238] Thus, whether one looks to coupled cluster, multi-reference perturbation theory, or the modern array of sparsity and compressibility approaches, it remains challenging to deliver consistently accurate excitation energies in larger molecules, larger basis sets, and across widely differing classes of excitations.

In pursuit of this challenge, we adopt an approach that combines the advantages of excited-state-specific quantum chemistry, sCI, and variational Monte Carlo (VMC). The combination [62, 92, 169, 239–242] of sCI and QMC can be especially effective thanks to the determinant expansions' leveraging of wave function sparsity and the efficacy of Jastrow factors [74, 77, 93, 96–103, 107, 108] for addressing weak correlation effects that sCI is less efficient for. While excited state VMC approaches in this area that use state-averaging are highly effective in many systems, [86–92] we choose to employ state-specific formulations of VMC [93–95] in order to avoid the same challenges that state-averaged CASPT2 faces in charge transfer states. Further, although recent advances in VMC orbital optimization [75, 76] can be employed for state-specific optimization, [70, 198] our experience has been that orbital optimization significantly increases the difficulty and cost of the overall VMC ansatz optimization when compared to optimizing CI coefficients and Jastrow parameters alone. In the present study, we therefore hypothesize that orbitals taken from excited-state-specific complete active space self consistent field theory (SS-CASSCF) [45, 50, 219] will be sufficiently close to optimal that accurate excitation energy predictions can be made without further optimization within VMC. With these quantum-chemistry-derived excited state orbitals, we employ sCI in an extended active space to identify the electron configurations that are most helpful in treating the strong and the leading weak correlation effects. This

qualitatively correct but not quantitatively accurate wave function is then fed into excited-state-specific VMC, where the configuration interaction expansion is re-optimized alongside explicit (Jastrow) correlation factors. Finally, as these wave functions are still not exhaustively accurate in larger molecules and basis sets, we use VMC to balance the accuracy of the ground and excited state approximations through variance matching [69–71, 198] in an attempt to explicitly enforce error cancellation in the energy differences between ground and excited states.

This combination of quantum chemistry with VMC offers a number of key advantages. The explicitly multi-reference character is essential when treating difficult double excitations and in states with significant mixtures of singly and doubly excited character. By avoiding state averaging, charge transfer states are provided with fully appropriate post-excitation orbital relaxations and, in the results below, we find that stopping the orbital optimization after the SS-CASSCF stage is sufficient for high accuracy. For weak correlation effects, the use of Jastrow factors and a modest but sCI-chosen configuration interaction expansion appears to be sufficient for the error cancellation balance of variance matching to do its job and deliver 0.1 eV accuracies in all systems tested in which a reliable theoretical benchmark exists. This interplay between variance matching, Jastrow factors, and the limited CI expansion is crucial, as it allows us to push into larger molecules and basis sets than can be reached by approaches that rely on sCI alone and thus require much more exhaustive CI expansions. Finally, it is important to point out that the approximations employed in this approach are quite different than those used in EOM-CC, which allows us to validate coupled cluster results in some circumstances where it is not otherwise obvious (e.g. due to partial doubly excited character) how accurate coupled cluster should be. When these methods agree about a challenging excitation energy, such as in the $3^1A'$ state of uracil, the fact that they rely on very different assumptions and approximations increases our confidence in the accuracy of the prediction. We therefore hope that, in addition to its longstanding role in providing ground state reference data in both molecular[52] and solid-state[53] settings, quantum Monte Carlo can, in tandem with excited-state-specific quantum chemistry and coupled cluster theory, be increasingly helpful in this regard in challenging molecular excited states.

To make this combined approach as practical as possible, we have also explored technical improvements to the VMC optimization and variance matching algorithms. Methods for optimizing wave functions in VMC have improved steadily in recent years, [79, 113, 116, 117, 123, 129] with variants of the linear method (LM) and accelerated descent (AD) showing particular progress. Building on our recent work on a hybrid optimization approach that combines a blocked variant [129] of the LM with AD, [71, 119] we adopt a simple scheme for deciding which parameters are most likely to benefit from the LM optimization step and which can be left for AD to optimized. We also deploy a more practical variance matching approach that in most cases allows us to avoid building multiple CI expansions for interpolation purposes and thus simplifies calculations in most cases.

We investigate the effectiveness of the overall approach and these technical improvements by surveying a range of different types of excited states. By drawing upon recent benchmark

sets [37, 41] and our own quantum chemistry calculations, we are able to demonstrate the efficacy of our SS-CASSCF/sCI/VMC approach in single, double, and charge transfer excited states, reaching 0.1 eV accuracy in all cases where rigorous benchmarks are available. In some larger systems, where such benchmarks are not available, we are still able to establish close agreement between our approach and the highest-affordable levels of coupled cluster theory. In a challenging doubly-excited state of benzene where coupled cluster is simply not appropriate, we are able to offer some clarity amidst the range of predictions made by different variants of CASPT2. Taken together, the results indicate that the combination of state-specific quantum chemistry, sCI, and VMC can offer new clarity for excited states in a wide range of molecules that are either too difficult or too large for other high-accuracy electronic structure methods.

4.3 Theory

Excited State-Specific Variational Monte Carlo

The application of VMC to excited states is an active area of method development with multiple possible approaches, including the examples of both state-averaged energy minimization[86–92] and state-specific variance-based minimization shown below.[70, 71, 93–95, 198]

$$\Omega(\Psi) = \frac{\langle \Psi \,|\, (\omega - H) \,|\, \Psi \rangle}{\langle \Psi \,|\, (\omega - H)^2 \,|\, \Psi \rangle} \tag{4.1}$$

$$W(\Psi) = \left\langle \Psi \,\middle|\, (\omega - H)^2 \,\middle|\, \Psi \right\rangle \tag{4.2}$$

$$E^{SA} = \sum_I w_I \frac{\left\langle \Psi^I \,\middle|\, H \,\middle|\, \Psi^I \right\rangle}{\langle \Psi^I \,|\, \Psi^I \rangle} \tag{4.3}$$

We employ the state-specific objective function [94] Ω, which targets the lowest state above ω in energy when minimized. To achieve size-consistent results, the input ω must be updated to $E - \sigma$, the difference between the energy and the standard deviation, which transforms the optimization of Ω into state-specific variance minimization.[95] There are multiple approaches to varying ω, but in this work we simply conduct multiple fixed-ω calculations and change the value of ω between each one until self-consistency between it and $E - \sigma$ is reached. In practice, only a few such fixed-ω optimizations are needed for any one state.

In contrast to the deterministic methods that are predominantly used in electronic structure, QMC must obtain stochastic estimates of the energy and all related quantities, and practitioners have flexibility in choosing probability distributions for performing efficient Monte Carlo integration. While the distribution $\rho(\mathbf{R}) = \frac{\Psi(\mathbf{R})^2}{\int d\mathbf{R} \Psi(\mathbf{R})^2}$ is a common choice and

has a zero variance property[145] when Ψ is an exact eigenstate, it suffers from an infinite variance problem with approximate Ψ in the estimation of other key quantities, particularly the variance and wave function derivatives. To address this issue, regularization schemes,[110, 111] estimator modifications,[145, 243, 244] and other importance sampling functions[67–71, 109] have all been variously employed by the QMC community. We use the same importance sampling function

$$|\Phi|^2 = |\Psi|^2 + c_1 \sum_i |\Psi^i|^2 + c_2 \sum_j |\Psi^j|^2 + c_3 \sum_k |\Psi^k|^2 \tag{4.4}$$

that was previously used and more fully discussed in recent work.[71] The coefficients c_1,c_2,c_3 weight the sums of squares of wave function parameter derivatives, divided between Jastrow, CI, and orbital parameters respectively. We have found that setting $(c_1, c_2, c_3) = (0.0, 0.0001, 0.0)$ is a reasonably effective choice for the molecules we consider.

Hybrid Optimization

The stochastic nature of QMC increases the difficulty of its wave function ansatz optimization compared to the analogous problem encountered in deterministic methods. Optimization has been a long-running challenge in QMC method development for many years with a variety of algorithms being introduced. These include the linear method[112–115] and accelerated descent[84, 116–119] approaches. Neither of these classes of methods is fully satisfactory, with the LM possessing a tendency for poor parameter updates due to stochastic step uncertainty with large numbers of variables and AD suffering from slow convergence to optimal parameter values.

We have recently developed a hybrid combination of the blocked version of the LM[129] with AD, first for ground state energy minimization[119] and then for variance-based excited state optimization.[71] In this approach, alternating between sections of descent optimization and the blocked LM enables more efficient convergence to the minimum than using either method on its own. The blocked LM is able to move parameters to the vicinity of the minimum more swiftly than descent, but descent is able to correct for any poor parameter updates made by the blocked LM due to step uncertainty, which makes it useful for converging more tightly. Throughout all our optimizations, we alternate between 100 iterations of AD and 3 steps of blocked LM and refer to one set of both as a macro-iteration. A typical optimization consists of a small number of macro-iterations followed by 1100 iterations of AD with final energy, target function, and variance averages taken over the final 500 iterations. Details on the numbers of macro-iterations and sampling effort used can be found in Appendix B. These averages and the excited state optimizations are then used in our variance matching procedure, discussed below, to obtain a final excitation energy result. More extensive discussions of our hybrid method can be found in earlier papers[71, 119] and in this benchmarking study, we use the same combination of Nesterov momentum with RMSprop and the same hyperparameter choices as before.[71]

One new element of our optimization approach is to filter the parameters updated by the blocked LM according to the statistical significance of the target function's parameter derivatives, an approach has been successfully used in the past by the QMC community.[245] We use the criterion:

$$\left| \frac{< \frac{\partial \Omega}{\partial p} >}{\sigma_{\frac{\partial \Omega}{\partial p}}} \right| > 1 \tag{4.5}$$

The mean value of the parameter derivative must be at least as large as the standard deviation of that derivative estimate in order for the parameter to be included for the blocked LM update on that iteration. The number and identity of the parameters that are turned off by this filtration stage will vary from LM step to LM step in practice, but we have found that filtering out roughly a third of the parameter set on each blocked LM iteration is a typical occurrence in our optimizations.

We also note that the filtered parameter space is further reduced by the blocked LM algorithm,[129] which divides the parameter set into blocks for LM-style diagonalizations. A limited number of eigenvectors are retained from each block to form the space for a second diagonalization that determines the final parameter update. In our optimizations, we have found dividing the remaining parameters into 5 blocks and retaining 30 parameter directions per block for the second stage of the blocked LM to be an effective choice. The use of the initial filtration stage allows for an automated identification of a smaller variable space that the blocked LM can treat with reduced step uncertainty. The AD sections of the hybrid method continue to optimize all the variational parameters. This refinement of the hybrid method enhances the effectiveness of the blocked LM on moving the most important parameters while allowing AD to handle the remaining parameter adjustments.

Wave Functions

The wave function ansatzes we use in our VMC optimizations are constructed from a preceding set of quantum chemistry calculations. In particular, we obtain molecular orbitals from a recently developed SS-CASSCF, [45, 219] which uses a density matrix-based variational principle to track individual states during CASSCF's CI and orbital relaxation optimization steps. When successfully converged, SS-CASSCF can avoid root-flipping and obtain orbitals tailored specifically for the targeted state. We then employ sCI, specifically heatbath CI, [10, 131] to generate a set of the most important determinants for a state from a larger active space of our state-specific orbitals. We emphasize that converging the sCI is unnecessary and that in practice we often need only the first few hundred to few thousand most important determinants for effective initial wave functions within VMC. With the addition of one- and two-body Jastrow factors, we arrive at the form of the multi-Slater Jastrow (MSJ) ansatz that we use for all the molecules we consider in this work.

$$\Psi = \psi_{MS} \psi_J \tag{4.6}$$

$$\psi_{MS} = \sum_{i=0}^{N_D} c_i D_i \tag{4.7}$$

$$\psi_J = \exp\left\{ \sum_i \sum_j \chi_k(|r_i - R_j|) + \sum_k \sum_{l>k} u_{kl}(|r_k - r_l|) \right\} \tag{4.8}$$

We construct the Jastrow factor ψ_J with splines for the functions χ_k and u_{kl}.[150] As we will show in our results, only limited numbers of the determinants D_i are needed to obtain highly accurate excitation energies in many molecules.

In all our VMC results, only the Jastrow factors and the determinant coefficients c_i are optimized by our hybrid method, while the orbitals are left at their SS-CASSCF shapes. While recent developments[75, 76] with the table method have made orbital optimization at the VMC level more efficient, it remains more computationally demanding than the Jastrow and CI optimization, and significantly increases the difficulty of our optimization problem. Our decision to use SS-CASSCF orbital shapes and avoid any VMC orbital optimization enables us to obtain significant computational savings while retaining the benefits of state-specificity, particularly in cases with important orbital relaxation effects such as charge transfer. As our results show, leaving the orbitals at their SS-CASSCF shapes does not prevent us from obtaining accurate results, suggesting that the lion's share of orbital relaxation is present already in the active space picture.

Variance Matching

Another key part of obtaining a high level of accuracy with compact wave functions is ensuring a balanced treatment of both ground end excited state in order to benefit from cancellation of errors. Past work has shown that matching the variance between ground and excited state improves excitation energy predictions, but these studies[69–71, 198] employed various different protocols for obtaining a match and often required additional VMC optimizations at a variety of Slater determinant expansion lengths.

In this work, we introduce a simplified variance matching procedure that reduces the need for considering different numbers of determinants. Once we have a given ground state optimization with some number of determinants and a final variance, we perform an excited state optimization for some chosen CI expansion length. We plot the energies and variances obtained on the blocked LM iterations of this optimization, leaving aside those from the first macro-iteration when the wave function is still mostly unoptimized. A linear fit to these points provides us with a correction to the excited state energy once its variance has been extrapolated to match the ground state variance, and all the excitation energies we report were obtained from this protocol. Figure 4.1 and Table 4.1 provide an example of the procedure in the case of ketene.

The uncertainty in the excitation energy due to the extrapolation can be estimated from the effect of the stochastic uncertainties in the positions of the energy-variance points on the regression slope. The individual uncertainties in the points can define normal distributions

for generating sets of slightly shifted points with their own regression slopes and corresponding excitation energy results. We can then determine an uncertainty in our excitation energy prediction from the uncertainty in the slope used in the extrapolation. In the limits of either an explicit variance match between ground and excited state or a deterministic optimization from infinite sampling, this extrapolation uncertainty will go to zero. Table 4.17 in Appendix B provides values for this extrapolation uncertainty for all the systems we consider. All excitation energy uncertainties we report in the main text are the sum of the uncertainty from stochastic averaging and the extrapolation uncertainty.

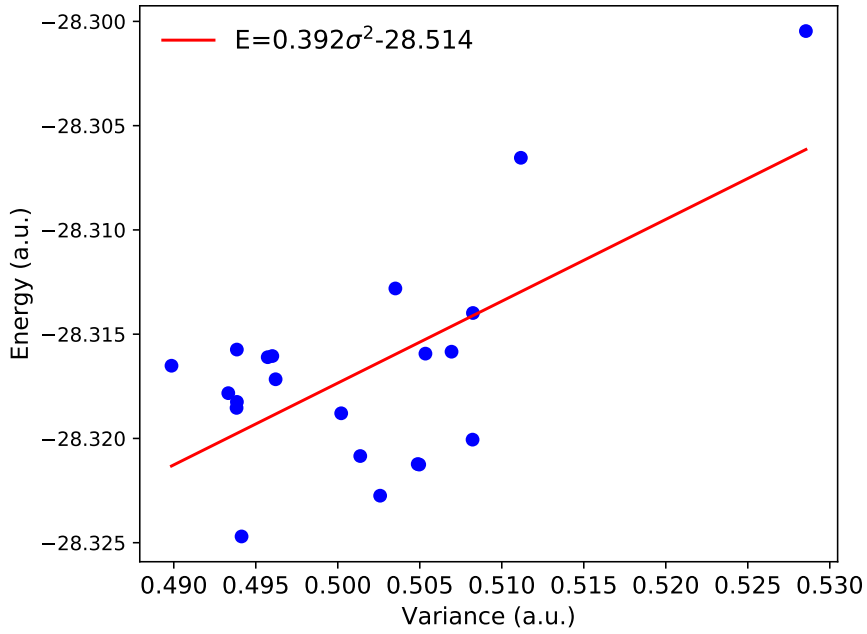


Figure 4.1: Example of the linear regression used in variance matching. The points are energies and variances from individual blocked LM iterations in the optimization of the excited state of ketene. See Table 4.1 for how the variance matching affects the excitation energy prediction.

Table 4.1: Energy and variance values in ketene before and after the variance matching procedure. As shown later in Table 4.2, the final excitation energy prediction is very accurate compared to the literature theoretical best estimate of 3.86 eV.[37]

	Energy (a.u.)	Variance (a.u.)	Excitation Energy (eV)
GS	−28.4813	0.456	
ES Before Match	−28.3226	0.494	4.32
ES After Match	−28.3375	0.456	3.91

This approach reduces the need to perform further optimizations with other numbers of determinants, a requirement when searching for an explicit variance match[71] or for performing an interpolation,[198] making the present methodology easier to apply. The respective numbers of determinants used for the ground and excited state optimizations are chosen based on experience, but can be systematically increased if necessary. In general, longer expansions will be needed to maintain the same level of accuracy in larger systems, and the results we have obtained after systematically increasing from smaller expansions may provide some guidance when considering other molecules of similar sizes.

4.4 Results

Computational Protocol and Technical Details

All our VMC calculations used a development version of QMCPACK.[150, 204] In all cases, we used the pseudopotentials and associated basis sets developed by Mitas and coworkers.[171] Molecular geometries were taken from literature benchmark sets[37, 38, 41, 190, 246] and coordinates are also listed in Appendix A. State-specific molecular orbitals were generated by an implementation of SS-CASSCF[45, 219] in a development version of PySCF.[205] We generated sCI expansions in expanded active spaces with Dice, [10, 131] iterating until at least a million determinants had been selected and then taking those with the largest coefficients into our VMC wave function. The particular active space and basis set choices for different molecules are given in the following sections for each set of excited states. For the charge transfer states we study, we have performed coupled cluster calculations in Molpro[170] and GAMESS[247] for comparisons against our VMC results.

Once our ansatzes were generated, the CI coefficients were optimized within VMC simultaneously with one- and two-body Jastrow factors. The Jastrow factor splines each consisted of 10 coefficients defining the function within a distance of 10 bohr. With two exceptions, the initial value of the Jastrow coefficients was set to zero in all cases and CI coefficients were begun at their values from sCI. In the case of nitrosomethane, the final 1000 determinant excited state optimization began from the result of a previous optimization of the Jastrow and 500 determinants, with the next 500 most important determinants from sCI added with initial coefficients of zero. Further technical details on optimization procedure and the uncertainties in our excitation energies can be found in Appendix B.

Valence Single Excitations: Thioformaldehyde, Methanimine, and Ketene

In testing our methodology, we begin by considering valence single excitations, specifically the $^1A_2(n \rightarrow \pi^*)$ excited state in thioformaldehyde, the $^1A''(n \rightarrow \pi^*)$ state in methanimine, and the $^1A_2(\pi \rightarrow \pi^*)$ state in ketene. For these three excited states, sCI and high level coupled cluster results[37] from Loos and coworkers are reliable references for assessing the

accuracy of our approach. We generated our SS-CASSCF orbitals using a (12e,10o) active space for thioformaldehyde, a (12e,12o) space for methanimine and a (8e,8o) space for ketene, with pseudopotentials and cc-pVTZ basis sets for all molecules.[171] The subsequent sCI calculations in Dice[10, 131] correlated 12 electrons in 17 orbitals for thioformaldehyde, 12 electrons in 18 orbitals for methanimine, and 16 electrons in 22 orbitals for ketene.

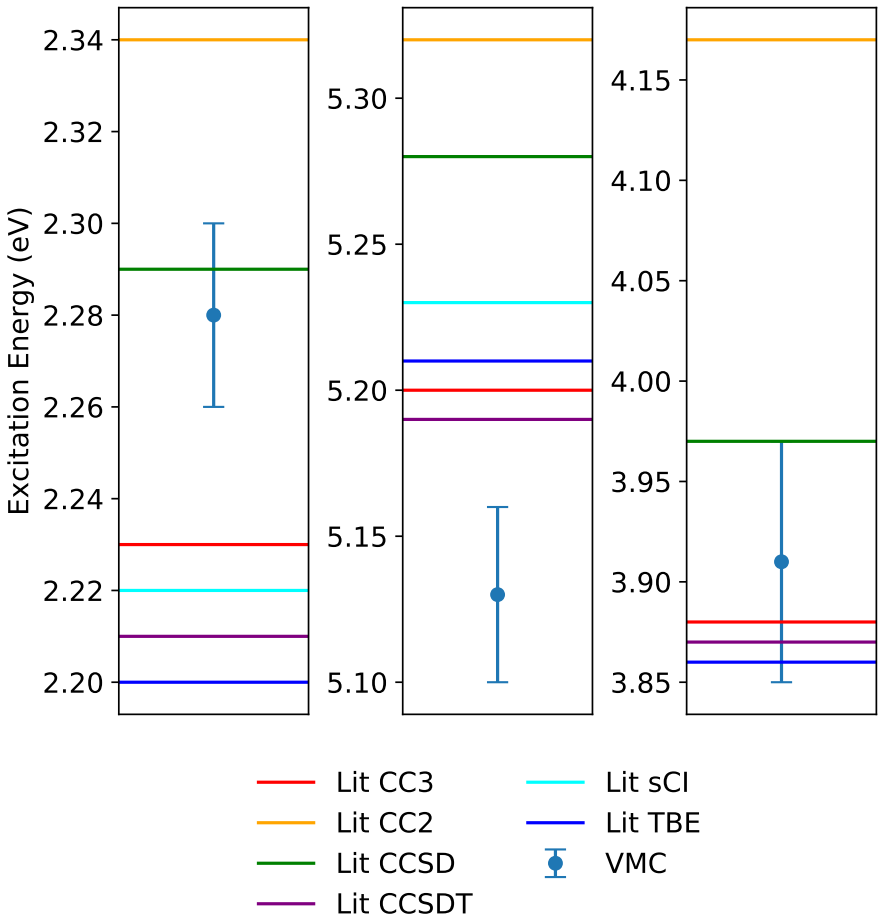


Figure 4.2: From left to right, excitation energies in thioformaldehyde, methanimine, and ketene for VMC and quantum chemistry approaches. See also Table 4.2. Reference values for coupled cluster, and selected CI are taken from the work of Loos and coworkers.[37]

Table 4.2: Excitation energies for thioformaldehyde, methanimine, and ketene. The VMC uncertainties on the final digits are given in parentheses. The included literature values use an aug-cc-pVTZ basis set for sCI and coupled cluster, and the literature TBEs are obtained from applying a basis set correction using CC3 to the sCI result.[37]

Method	Thioformaldehyde	Methanimine	Ketene
VMC	2.28(2)	5.13(3)	3.91(6)
CC2	2.34	5.32	4.17
CCSD	2.29	5.28	3.97
CC3	2.23	5.20	3.88
CCSDT	2.21	5.19	3.87
sCI	2.22	5.23	3.86
Lit TBE	2.20	5.21	3.86

Our results for this trio of small molecules, depicted visually in Figure 4.2 with precise values in Table 4.2, show that our VMC methodology is highly accurate. In all cases, our final excitation energies are within 0.1 eV of the theoretical best estimates determined by Loos and coworkers from basis set corrected sCI calculations. We also emphasize that our results require only relatively simple wave function ansatzes. We used 700 determinants for the ground state and 300 for the excited state in thioformaldehyde, 300 ground and 500 excited for methanimine, and 500 ground and 700 excited for ketene. Thioformaldehyde was somewhat unusual with the ground state having the higher variance for fewer determinants so a longer expansion was optimized for a better variance match with the excited state. The VMC wave functions used in these systems are significantly less complex than those obtainable by coupled cluster and the absolute energies from VMC are comparable to those of CCSD. The excitation energy differences from VMC are also most comparable to CCSD in terms of accuracy, as CC3 and CCSDT are within one or two hundredths of an eV of the sCI-based TBEs. While VMC does not obtain the exquisite accuracy of higher level coupled cluster in these cases, these results are a reassuring check that our approach agrees with very well established quantum chemistry treatments of this generally easier class of excited states, and we now turn our attention to more challenging categories.

Double Excitations: Carbon Trimer, Nitrosomethane, Hexatriene, and Benzene

We next address double excitations, which are a far more challenging class of excited states for coupled cluster. Past work[71] applying VMC to double excitations showed a high level of accuracy and we again draw from the wealth of benchmarking data[41] from Loos and coworkers, comparing our refined SS-CASSCF/VMC methodology to their quantum chemistry values in four molecules. We first consider two small systems, the carbon trimer and nitrosomethane, where highly reliable excitation energies can still be established by sCI, before studying hexatriene and benzene, which are considerably larger cases where less definitive

methods must be used. For our ansatz construction, SS-CASSCF orbitals were obtained from a (12e,12o) space for the carbon trimer, a (12e,9o) space for nitrosomethane, and (6e,6o) spaces for hexatriene and benzene. At the sCI stage, these spaces were expanded to (12e,18o), (18e,22o), (32e,46o), and (30e,46o) respectively. For all four molecules, we have used cc-pVTZ basis sets and pseudopotentials.[171]

Concentrating first on the two smaller systems, we see in Figure 4.3 and Table 4.3 that our VMC approach is again highly accurate and within 0.1 eV of the TBEs obtained from sCI. Both states can be classified as pure double excitations based on the low percentage of T_1 amplitudes in CC3[41] and we see that CC3 makes substantial errors of an eV or worse for this type of doubly excited state. Shouldering the expense of full triples with CCSDT only manages to eliminate about half of CC3's errors. In contrast, CASPT2's multireference nature makes it far more successful in this situation and it achieves comparable accuracy to our VMC results in both molecules.

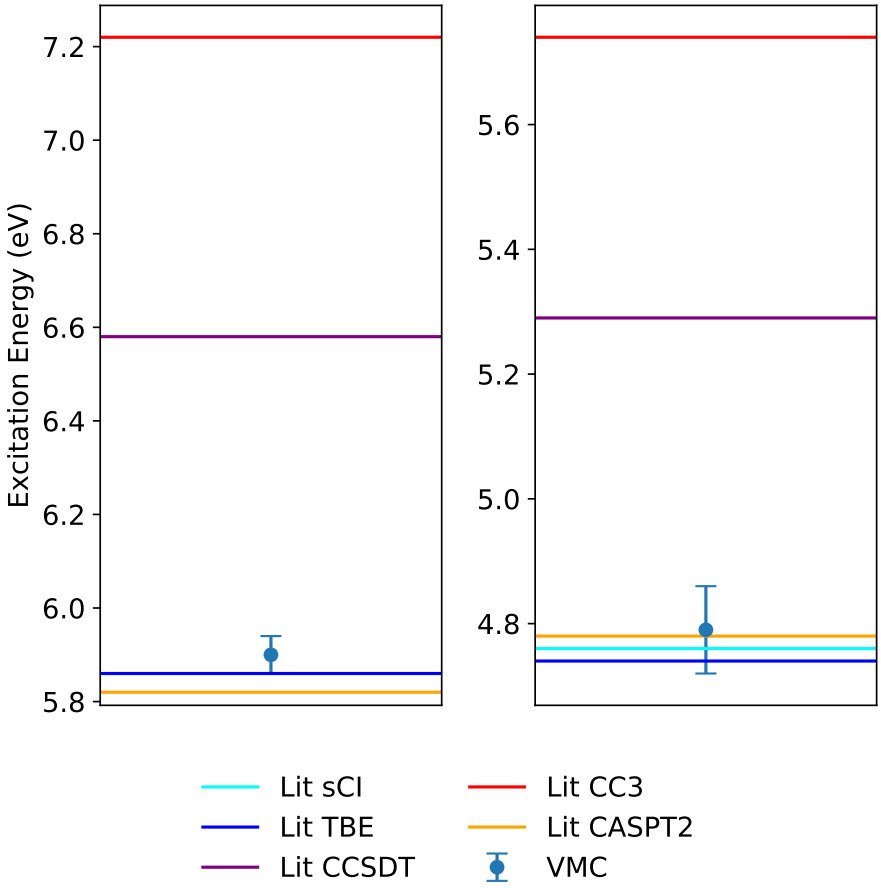


Figure 4.3: Excitation energies in the carbon trimer and nitrosomethane in the left and right panels respectively. See also Table 4.3. For the carbon trimer, VMC used 100 determinants in the ground state and 500 for the excited state, while 500 and 1000 respectively were used for nitrosomethane. Reference values for coupled cluster, and selected CI are taken from the work of Loos and coworkers.[41]

Table 4.3: Excitation energies for the carbon trimer and nitrosomethane. The VMC uncertainties on the final digits are given in parentheses. The included literature[41] values use an aug-cc-pVQZ basis set for the carbon trimer and the CC3 and CASPT2 values for nitrosomethane. The literature nitrosomethane CCSDT and sCI results used an aug-cc-pvTZ basis set.

Method	Carbon Trimer	Nitrosomethane
VMC	5.90(4)	4.79(7)
CC3	7.22	5.74
CCSDT	6.58	5.29
CASPT2	5.82	4.78
sCI	5.86	4.76
Lit TBE	5.86	4.74

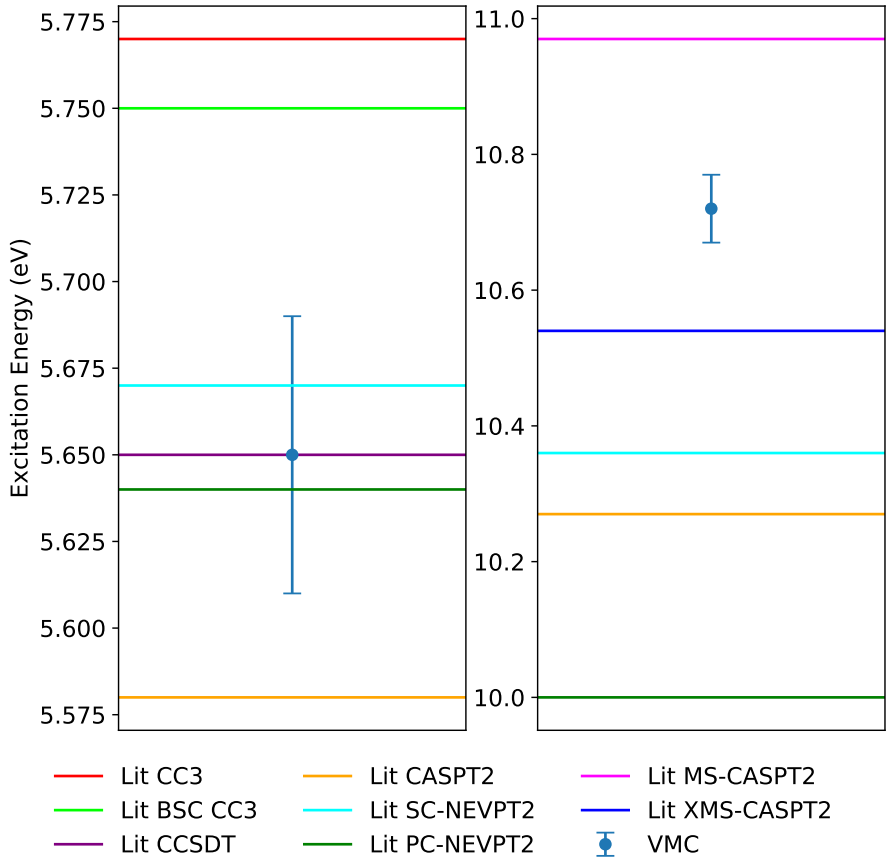


Figure 4.4: Excitation energies in hexatriene and benzene in the left and right panels respectively. See also Table 4.4. In both molecules, the VMC used 500 determinants in ground state and 1000 for the excited state. Reference values for coupled cluster, CASPT2 and NEVPT2 are taken from the work of Loos and coworkers.[41]

Table 4.4: Excitation energies for hexatriene and benzene. The VMC uncertainties on the final digits are given in parentheses. The included literature[41] values use an aug-cc-pVDZ basis set for CC3 and CCSDT in hexatriene and an aug-cc-pVQZ basis set for all varieties of CASPT2 and NEVPT2, except for PC-NEVPT2 in benzene, which used aug-cc-pVTZ.

Method	Hexatriene	Benzene
VMC	5.65(4)	10.72(5)
CC3	5.77	
BSC CC3	5.75	
CCSDT	5.65	
CASPT2	5.58	10.27
PC-NEVPT2	5.64	10.00
SC-NEVPT2	5.67	10.36
MS-CASPT2		10.97
XMS-CASPT2		10.54

For molecules at the scale of hexatriene and benzene, high accuracy excitation energies from sCI are now unaffordable and various flavors of coupled cluster and multireference perturbation theory are left as the main quantum chemistry tools. This leaves space for our VMC approach to provide reliable excitation energies that can help determine whichever other methods are most accurate in larger systems. For the hexatriene results shown in the left panel of Figure 4.4 and Table 4.4, we see that VMC is in closer agreement with coupled cluster, particularly CCSDT, instead of CASPT2. This aligns with the conclusions of Loos and coworkers that the 2 ^1A_g state has a high level of singles character, making coupled cluster with triples more reliable than it is for pure double excitations. In the case of the 2 $^1A_{1g}$ state in benzene, the literature[41] values consist only of different varieties of CASPT2 and NEVPT2, which show a significant spread of almost an entire eV in their excitation energy predictions without an obvious reason to favor any particular value. Our VMC result lands closest to the XMS-CASPT2 value, but is higher by about 0.18 eV and it may serve as a useful point of comparison for any other methods that are applied to this state in the future.

Charge Transfer: Ammonia-difluorine

With the accuracy of our method well established for valence single excitations and double excitations, we now show the same holds true for charge transfer states using the well-known ammonia-difluorine test system[38, 229, 248, 249] at a separation of 6 Å. Our wave function generation proceeds as before with a (2e,6o) active space for the SS-CASSCF and a (22e,40o) space for the sCI using pseudopotentials and an aug-cc-pVTZ basis set.[171] For comparison to our VMC, we performed δ-CR-EOMCC(2,3)D[250] calculations in GAMESS[247] and EOM-CCSD and CASPT2 calculations in Molpro.[170]

From Figure 4.5 and Table 4.5, we see that our VMC is in excellent agreement with the aug-cc-pVTZ δ-CR-EOMCC(2,3)D result. With only a cc-pVTZ basis set, the δ-CR-EOMCC(2,3)D excitation energy is 0.4 eV larger and EOM-CCSD has a substantial error of about 0.8 eV even with augmentation, showing the importance of both including some triples in the coupled cluster and using diffuse functions for quantitative accuracy. We also note that CASPT2 with a (2e,6o) active space, a 6 state state-average (necessary to even see the CT state), and IPEA shift of 0.25 has an error comparable to EOM-CCSD's, while increasing the active space to (6e,8o) with the same state averaging and shift improves the accuracy by only 0.14 eV. CASPT2's difficulties are likely due to the limitations of state-averaging for CT states, which we can avoid through our state-specific methodology. The high accuracy of our VMC in this CT example indicates that we can obtain reliable excitation energies regardless of the type of excited state, a versatility that neither coupled cluster nor CASPT2 can achieve.

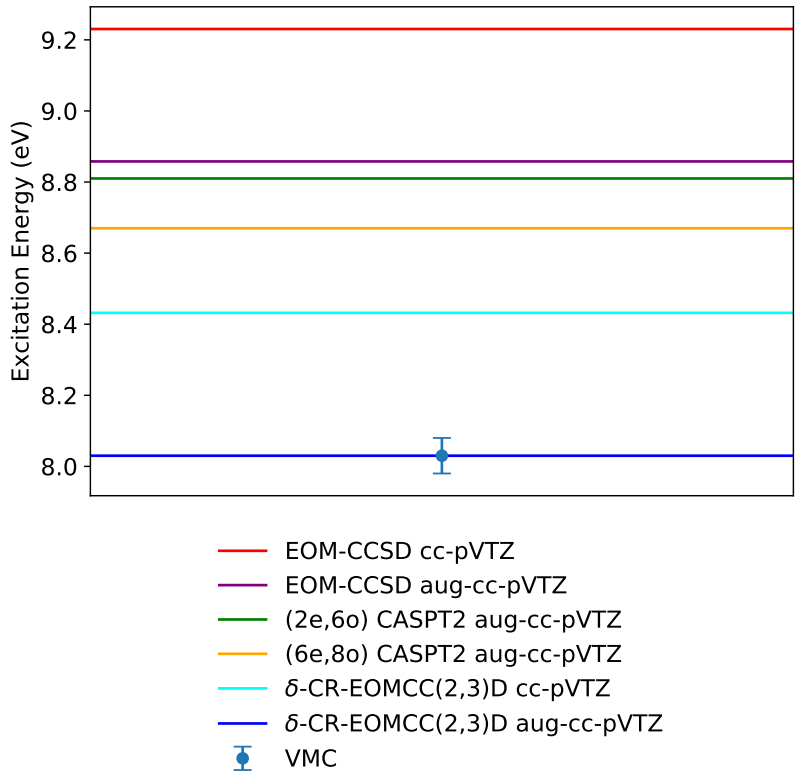


Figure 4.5: Excitation energy of the lowest CT state in ammonia-difluorine for VMC and quantum chemistry approaches. See also Table 4.5. The VMC used 300 determinants in the ground state and 500 for the excited state.

Table 4.5: Excitation energies for the CT state in ammonia-difluorine.The VMC uncertainty for the last digit is given in parentheses.

Method	Excitation Energy (eV)
VMC	8.03(5)
(2e,6o) CASPT2 aug-cc-pVTZ	8.81
(6e,8o) CASPT2 aug-cc-pVTZ	8.67
CCSD cc-pVTZ	9.23
CCSD aug-cc-pVTZ	8.86
δ-CR-EOMCC(2,3)D cc-pVTZ	8.43
δ-CR-EOMCC(2,3)D aug-cc-pVTZ	8.03

Larger System: Uracil

We have demonstrated our approach's high level of accuracy across multiple types of excited states and now apply it to a larger molecule. We first address the 2 $^1A'$ and 3 $^1A'$ states in uracil, which appear as part of the Thiel set[190] and other benchmarking studies.[251, 252] To generate our ansatzes, we used a (10e,8o) active space for the SS-CASSCF and a (42e,42o) space for the sCI with pseudopotentials and a cc-pVTZ basis set.[171]

These states have been previously studied with different levels of coupled cluster and CASPT2, giving a significant range of answers that spans 0.7 eV for the 2 $^1A'$ state and almost an entire eV for the 3 $^1A'$ state. The latter state is notable for having a significant, though non-dominant, amount of double excitation character, in the range of 15 to 20 percent based on CASSCF analysis, which might lead to hesitation over whether to trust coupled cluster or CASPT2. VMC is in a position to provide the necessary guidance, with our results ending up in good agreement with the CCSD(T) results of Szalay and coworkers[252] for both states and more generally supporting the higher excitation energy predictions of coupled cluster over the CASPT2 results of Roos[251] and Thiel.[190] We also note that while the numbers of determinants needed in VMC after systematic expansion are larger than those used for smaller molecules, they can remain modestly sized due to the error cancellation benefits of our variance matching procedure.

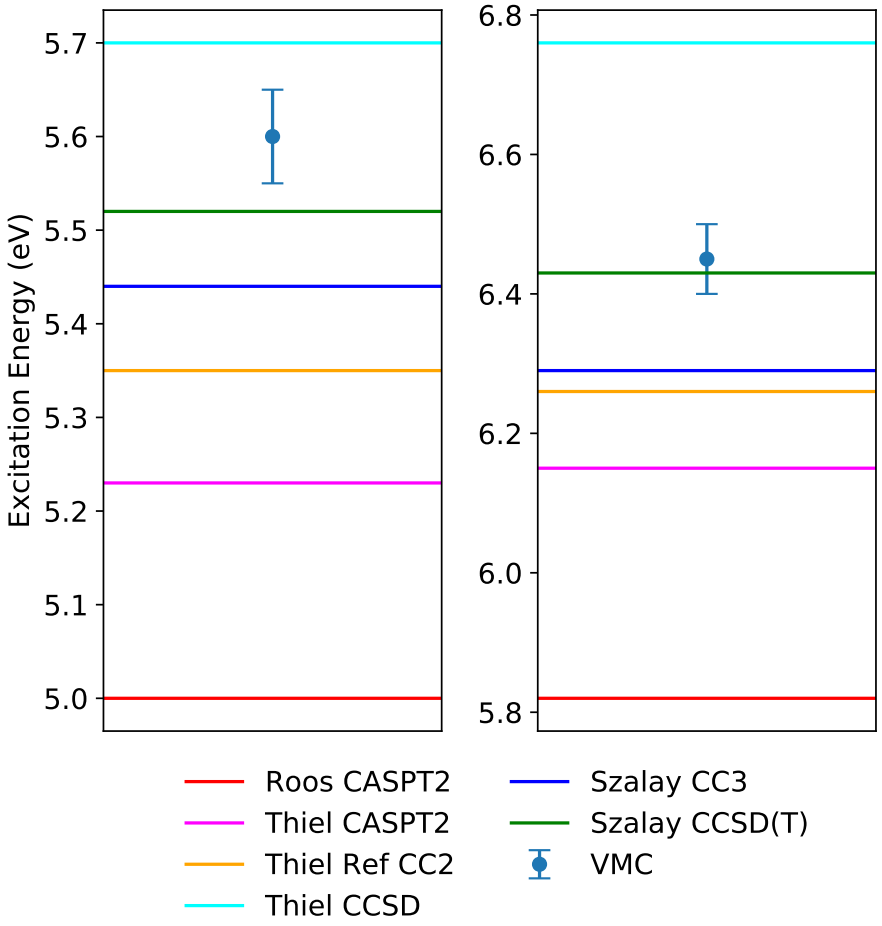


Figure 4.6: Excitation energies in uracil. The left panel depicts values for the 2 $^1A'$ state and the right panel for the 3 $^1A'$ state. See also Table 4.6. The VMC used 500 determinants the ground state and 1500 for each of the excited states.

Table 4.6: Excitation energies in eV for the $2\ ^1A'$ and $3\ ^1A'$ states in uracil. The VMC uncertainty for the last digit is given in parentheses. The literature results from Thiel[190] and Szalay[252] used a TZVP basis set, with the Thiel set using an aug-cc-pVQZ CC2 result as its reference. Details on the ANO type basis set used by Roos can be found in the original paper.[251]

Method	$2\ ^1A'$	$3\ ^1A'$
VMC	5.60(5)	6.45(5)
Roos CASPT2	5.00	5.82
Thiel Ref CC2	5.35	6.26
Thiel CCSD	5.70	6.76
Thiel CASPT2	5.23	6.15
Szalay CC3	5.44	6.29
Szalay CCSD(T)	5.52	6.43

4.5 Conclusions

By constructing a method that combines the advantages of SS-CASSCF, sCI, and VMC, we find that it is possible to offer high accuracy excitation energy predictions across singly excited, doubly excited, and charge transfer states. Unlike CASPT2, this combined method avoids state averaging and is thus able to accommodate the post-excitation orbital relaxations that are so crucial in charge transfer. Unlike EOM-CC, the method is explicitly multi-configurational, making it much more reliable in doubly excited states. Unlike sCI on its own, the method is able to arrive at accurate excitation energies with much shorter determinant expansions thanks to the combined effect of Jastrow factors and variance matching. Together, these advantages allow the method to make reliable excitation energy predictions in a wider variety of excited states than either coupled cluster or multi-reference perturbation theory while also reaching into system sizes beyond the current limits of sCI without sacrificing basis set quality. In key systems like uracil, agreement between our method and the highest level coupled cluster that can be afforded allows us to make excitation energy predictions with a confidence that is bolstered by the significant differences in these methods' underlying approximations.

From a methodological perspective, our key finding is that SS-CASSCF orbitals are accurate enough that further optimization within VMC is not necessary for producing accurate excitation energies. The practical advantage of this finding is that, while more efficient than it used to be, orbital optimization within VMC remains more difficult than CI coefficient and Jastrow optimization, and by avoiding it we can therefore reduce computational cost. We also find that, in most cases, effective variance matching can be performed without the need for interpolations based on multiple wave function expansion lengths by instead using linear extrapolations based on optimization history. Finally, removing low-priority parameters from the blocked linear method steps while still optimizing them with accelerated descent

did not prevent high accuracies from being attained, which allows us to work with smaller sample sizes than would otherwise be necessary when estimating the elements of the linear method matrices.

Looking forward, we note that our VMC approach can be applied to additional excited state situations beyond those considered in this work. Core excitations are one area where VMC has already demonstrated high accuracy, [198] and work is currently underway on exploring whether the approach presented here will offer similar benefits in that context. Excited state absorption [253] is another setting that challenges both CASPT2 and coupled cluster and where the present approach could be usefully applied without any obvious additional difficulty. Finally, in solid-state systems, the discrete states of defects are another potential opportunity for applying our state-specific machinery. Methodologically, further improvements in user-accessibility remain an ongoing effort. Systematic determination and automation of effective choices in optimization details such as the mixture between accelerated descent and the blocked linear method, descent hyperparameters, and the choice of guiding function should streamline workflows in future. With such improvements, it should be possible to apply our SS-CASSCF/sCI/VMC methodology to an even wider range of challenging excited state questions.

4.6 Acknowledgments

This work was supported by the Office of Science, Office of Basic Energy Sciences, the U.S. Department of Energy, Contract No. DE-AC02-05CH11231. Computational work was performed with the Berkeley Research Computing Savio cluster, the LBNL Lawrencium cluster, and the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02- 05CH11231.

4.7 Appendix

Molecular Geometries

Table 4.7: Structure of thioformaldehyde. Coordinates in Å.

C	0.000 000	0.000 000	−1.104 273
S	0.000 000	0.000 000	0.514 631
H	0.000 000	0.918 957	−1.677 563
H	0.000 000	−0.918 957	−1.677 563

Table 4.8: Structure of methanimine. Coordinates in Å.

C	0.056 604	0.000 000	0.587 869
N	0.056 961	0.000 000	−0.686 225
H	−0.842 138	0.000 000	1.202 802
H	1.007 951	0.000 000	1.108 065
H	−0.899 369	0.000 000	−1.038 338

Table 4.9: Structure of ketene. Coordinates in Å.

C	0.000 000	0.000 000	−1.295 479
C	0.000 000	0.000 000	0.018 514
O	0.000 000	0.000 000	1.183 578
H	0.000 000	0.938 930	−1.818 814
H	0.000 000	−0.938 930	−1.818 814

Table 4.10: Structure of carbon trimer. Coordinates in Å.

C	0.000 000	0.000 000	0.000 000
C	0.000 000	0.000 000	1.298 313
C	0.000 000	0.000 000	−1.298 313

Table 4.11: Structure of nitrosomethane. Coordinates in Å.

C	−0.944 193	0.000 000	−0.567 405
N	−0.002 867	0.000 000	0.571 831
O	1.157 919	0.000 000	0.229 939
H	−0.409 287	0.000 000	−1.515 646
H	−1.574 151	0.882 677	−0.457 339
H	−1.574 151	−0.882 677	−0.457 339

Table 4.12: Structure of hexatriene. Coordinates in Å.

C	−0.000 131	0.672 998	0.000 000
C	0.000 131	−0.672 998	0.000 000
C	1.201 720	1.480 569	0.000 000
C	−1.201 720	−1.480 569	0.000 000
C	1.201 720	2.821 239	0.000 000
C	−1.201 720	−2.821 239	0.000 000
H	−0.949 240	1.200 063	0.000 000
H	0.949 240	−1.200 063	0.000 000
H	2.145 737	0.948 370	0.000 000
H	−2.145 737	−0.948 370	0.000 000
H	0.274 043	3.378 403	0.000 000
H	−0.274 043	−3.378 403	0.000 000
H	2.122 561	3.385 081	0.000 000
H	−2.122 561	−3.385 081	0.000 000

Table 4.13: Structure of benzene. Coordinates in Å.

C	0.000 000	1.392 503	0.000 000
C	−1.205 943	0.696 252	0.000 000
C	−1.205 943	−0.696 252	0.000 000
C	0.000 000	−1.392 503	0.000 000
C	1.205 943	−0.696 252	0.000 000
C	1.205 943	0.696 252	0.000 000
H	−2.141 717	1.236 521	0.000 000
H	−2.141 717	−1.236 521	0.000 000
H	0.000 000	−2.473 041	0.000 000
H	2.141 717	−1.236 521	0.000 000
H	2.141 717	1.236 521	0.000 000
H	0.000 000	2.473 041	0.000 000

Table 4.14: Structure of ammonia-fluorine. Coordinates in Å.

N	0.000 000	−0.234 913	−3.520 527
H	0.000 000	0.704 739	−3.904 943
H	0.813 763	−0.704 739	−3.904 943
H	−0.813 763	−0.704 739	−3.904 943
F	0.000 000	−0.234 913	2.479 473
F	0.000 000	−0.234 913	3.904 943

Table 4.15: Structure of uracil. Coordinates in Å.

H	−2.025 413	−1.517 742	0.000 000
H	−0.021 861	1.995 767	0.000 000
H	2.182 391	−1.602 586	0.000 000
H	−0.026 659	−2.791 719	0.000 000
C	−1.239 290	0.359 825	0.000 000
C	1.279 718	0.392 094	0.000 000
C	1.243 729	−1.064 577	0.000 000
C	0.055 755	−1.709 579	0.000 000
O	−2.308 803	0.954 763	0.000 000
O	2.287 387	1.092 936	0.000 000
N	−1.139 515	−1.026 364	0.000 000
N	0.000 000	0.978 951	0.000 000

Optimization Details

As discussed in the main text, we perform all our VMC optimizations with the hybrid method, alternating between sections of AD and blocked LM optimization. Each AD portion of a hybrid macro-iteration produces 5 vectors of parameter differences as input to the blocked LM algorithm as parameter coupling information.[119] As mentioned in section IIB, all blocked LM iterations divided a filtered parameter set into 5 blocks and retained 30 parameter directions for the second stage of the algorithm. In practice, shift values[150] c_i and c_s are added to blocked LM matrices to prevent unwisely large steps in parameter space. Our implementation of the LM uses an adaptive scheme[119] that compares target function values calculated through correlated sampling for three different sets of shift values and selects whichever update step and corresponding shifts give the lowest target function. For all optimizations, we set the initial values of c_i and c_s to 0.1 and 1.0 respectively.

The particular numbers of macro-iterations and samples for each molecule are listed in Table 4.16 and Table 4.17 gives details on the uncertainties resulting from the variance matching extrapolation procedure that we have used for all our excitation energy results. In Table 4.17, the energy-variance slope used for nitrosomethane is significantly smaller than the others due to the optimization starting from a preoptimzed wave function.

Table 4.16: Macro-iterations used for GS and ES optimizations. Samples per iteration used for AD and blocked LM portions of the hybrid optimizations. Total samples for AD and LM with the costs for GS and ES combined. The nominal samples per iteration for blocked LM iterations is multiplied by 3 for computing the total sampling effort due to the need to run over the samples additional times to perform the automated parameter selection and the blocked LM algorithm.

Molecule	Macro-iterations: (GS, ES)	Samples per iteration: (AD, LM)	Total Samples (10^6): (AD, LM)
Thioformaldehyde	(7,8)	$(3 \times 10^4, 1 \times 10^6)$	(111, 135)
Methanimine	(5,6)	$(3 \times 10^4, 1 \times 10^6)$	(99, 99)
Ketene	(6,8)	$(3 \times 10^4, 1 \times 10^6)$	(108, 126)
Carbon trimer	(5,5)	$(3 \times 10^4, 1 \times 10^6)$	(96, 99)
Nitrosomethane	(7,6)	$(3 \times 10^4, 1 \times 10^6)$	(105, 117)
Hexatriene	(8,8)	$(3 \times 10^4, 1 \times 10^6)$	(114, 144)
Benzene	(7,7)	$(3 \times 10^4, 1 \times 10^6)$	(108, 126)
Ammonia-difluorine	(4,6)	$(3 \times 10^4, 1 \times 10^6)$	(96, 90)
Uracil 2 $^1A'$	(7,8)	$(5 \times 10^4, 1 \times 10^6)$	(104, 135)
Uracil 3 $^1A'$	(7,8)	$(5 \times 10^4, 1 \times 10^6)$	(104, 135)

Table 4.17: Uncertainties in the excitation energies due to stochastic averaging and variance matching extrapolation. Raw variance values for ground and excited states. Energy-variance slopes used in the extrapolations and uncertainties in the slopes due the uncertainty in the individual blocked LM optimization points.

Molecule	Excitation Energy (eV)	Averaging Uncertainty (eV)	Extrapolation Uncertainty (eV)	GS Variance (a.u.)	ES Variance (a.u.)	Slope (a.u.)	Slope Uncertainty (a.u.)
Thioformaldehyde	2.28	0.02	0.001	0.178	0.177	0.4936	0.052
Methanimine	5.13	0.02	0.01	0.216	0.235	0.3727	0.026
Ketene	3.91	0.02	0.04	0.456	0.494	0.3924	0.039
Carbon trimer	5.90	0.02	0.02	0.209	0.195	0.5360	0.046
Nitrosomethane	4.79	0.03	0.04	0.527	0.563	0.0480	0.037
Hexatriene	5.65	0.03	0.01	0.605	0.591	0.5574	0.012
Benzene	10.72	0.03	0.02	0.570	0.600	0.4317	0.022
Ammonia-difluorine	8.03	0.03	0.02	1.066	1.102	0.3977	0.024
Uracil 2 $^1A'$	5.60	0.04	0.01	1.393	1.406	0.5165	0.027
Uracil 3 $^1A'$	6.45	0.04	0.01	1.393	1.405	0.4622	0.015

Chapter 5

Optimization Stability in Variational Monte Carlo

5.1 Abstract

We investigate the issue of optimization stability in variance-based state-specific variational Monte Carlo, discussing the roles of the objective function, the complexity of wave function ansatz, the amount of sampling effort, and the choice of minimization algorithm. Using a small cyanine dye molecule as a test case, we systematically perform minimizations using variants of the linear method as both a standalone algorithm and in a hybrid combination with accelerated descent. We demonstrate that adaptive step control is crucial for maintaining the linear method's stability when optimizing complicated wave functions and that the hybrid method enjoys both greater stability and minimization performance. As a verification of variance minimization's practical utility, we report an excitation energy in the cyanine dye that is in good agreement with both benchmark quantum chemistry values and results obtained from state-averaged energy-based variational Monte Carlo.

5.2 Introduction

The challenges of wave function optimization have been a long-running thread in the development of Quantum Monte Carlo (QMC) methods. Ground state variational Monte Carlo (VMC) alone has seen significant evolution both in the optimization algorithms themselves[84, 112–114, 116–123, 254] and the objective functions they are applied to.[53, 93, 113, 147, 149] Early work on VMC optimization implemented steepest descent[255, 256] and Newton-Raphson[147, 148] for relatively small numbers of parameters in atoms and small molecules. As optimization algorithms, steepest descent and Newton-Raphson mark out two extremes in terms of derivative information and computational cost as purely first-order and fully second-order methods respectively. Subsequent methods, including approximate Newton,[121, 122] the linear method (LM),[112–114] stochastic reconfiguration,[123, 254] and

accelerated descent (AD),[84, 116–118] can be broadly viewed as pursuing a balance between the advantages of second-order derivative information against the cost of obtaining it. This profusion of different techniques, spread among many research groups, has motivated recent efforts to systematically compare their performance and identify effective combinations of them.[119]

The need to achieve robust VMC optimization has also surfaced more recently in the context of excited state VMC. For this frontier in VMC, method development work has developed both energy-based state-averaged and variance-based state-specific objective functions for describing excited states. State-averaged energy minimization has been successful in many applications,[86–92] but a state-specific approach[70, 71, 94, 95, 198, 257] offers advantages in cases where the appropriate orbital shapes for the relevant states differ significantly. This limitation of state-averaging also appears in deterministic methods such as CASPT2[43–45] and is a barrier to successfully studying charge transfer states[38, 39, 229, 257] and core excitations.[198]

Methodological advancements have steadily occurred in recent years for both state-averaged and state-specific VMC. Many innovations can be readily adopted within both frameworks such as improvements in ansatz design with selected CI (sCI) expansions[92, 169, 258] and new optimization algorithms.[71, 119, 129, 257] In addition, multiple protocols[69, 92, 169, 198, 257] have been developed to ensure balanced treatments of different states, a prerequisite for obtaining accurate excitation energies. The progress on all these fronts has helped advance excited-state VMC from initial applications[60, 93, 94, 121] with small molecules and simple wave functions to much wider use on larger systems with more sophisticated ansatzes.[92, 257]

Given the hopes for developing VMC into a reliable tool for excited states, recent research[91] raising concerns about the stability of variance minimization in VMC merits close attention. In a small cyanine dye molecule, Filippi and coworkers found failures of variance minimization to converge to the correct excited state even for high quality wave function ansatzes. In some cases, a lowering of the variance was accompanied by a substantial rise in the energy as the optimization moved to some undesired state. This behavior was attributed to the shape of the variance's parameter space landscape, with the optimization finding "little or no barrier to escape from a local minimum or local plateau, eventually converging to a lower-variance state instead of the target state."[91] This hypothesis would imply that the same behavior might be observed in other excited state applications of variance minimization, posing a significant obstacle to successful excitation energy predictions. In the present study, we investigate instabilities in variance optimization in order to better understand where they come from and how they may be addressed.

The issue of stability in VMC optimization is multi-faceted, with technical considerations about the choice of objective function, the particular optimization algorithm at hand, the level of sampling effort, and the quality of wave function ansatz each playing a role. Deficiencies in any of these areas could potentially lead to failures of optimization, and no approach to optimization is likely to be successful in absolutely all circumstances. However, the relevant question for VMC practitioners is whether these instances of optimization fail-

ure pose an insurmountable barrier in practice or can be remedied with certain choices in methodological protocol.

At a broad conceptual level, the objective function being minimized and the particular wave function ansatz set the difficulty of a VMC optimization. Compared to the energy, the variance landscape in parameter space is more difficult to resolve at a given level of sampling effort and estimates of its derivatives may also be more uncertain. While increasing the amount of sampling enables systematic improvement of these estimates, the availability of computational resources inevitably places some practical upper limit. The shape of the landscape will also vary with the particular wave function form that is chosen. A minimum of zero variance is guaranteed for an exact eigenstate and some minimum can be expected to form when a wave function is systematically built up to this limit, as in a Slater determinant expansion. However, the depth and shape of the minimum are difficult to characterize and the ansatz parameters are unlikely to be set to the optimal values initially in practice. As a result, the particular optimization algorithms chosen to navigate the parameter landscape must be highly robust to noisy gradients and capable of handling increasing numbers of parameters. We shall see that different types of algorithms with varying levels of step control may be more or less vulnerable to destabilization from poor steps as the number of parameters and the noise of their updates grow.

To further elucidate the issue of optimization stability in VMC, we conduct systematic investigations of the LM in the case of the excited state of the cyanine dye molecule CN5. We focus on identifying circumstances where variance minimization with the LM fails and what can be done to resolve these failures. Beyond technical choices in operating the LM, we also consider the stability advantages offered by a hybrid combination of the LM's blocked variant and AD, particularly those rooted in adaptive step control. As we will see, multiple optimization choices play a role in optimization effectiveness, but the key determiner of stability appears to adaptive step control, analogous to using an adaptive trust radius. In particular, variance optimization instabilities do not appear to arise from pathologies in the objective function surface, since if this were the case, all optimization methods regardless of adaptive step control would eventually show instabilities as well. As our data below demonstrates, even cases previously reported as unstable become stable when using adaptive step size control.

5.3 Theory

Objective Functions in VMC

Analysis of optimization stability within VMC begins with consideration of the objective or target function being minimized. A variety of choices have been employed by VMC practitioners, for both ground and excited state studies, with the most common choices being the energy or the variance of energy for the system at hand.

$$E_V(\vec{\mathbf{p}}) = \frac{\int \Psi_T^2(\vec{\mathbf{p}}) E_L(\vec{\mathbf{p}}) d\vec{\mathbf{R}}}{\int \Psi_T^2(\vec{\mathbf{p}}) d\vec{\mathbf{R}}} \tag{5.1}$$

$$\sigma_E^2(\vec{\mathbf{p}}) = \frac{\int \Psi_T^2(\vec{\mathbf{p}}) (E_L(\vec{\mathbf{p}}) - E_V(\vec{\mathbf{p}}))^2 d\vec{\mathbf{R}}}{\int \Psi_T^2(\vec{\mathbf{p}}) d\vec{\mathbf{R}}} \tag{5.2}$$

In the above definitions, $E_L = \frac{H\Psi(\mathbf{R})}{\Psi(\mathbf{R})}$ is the local energy. Historically, numerical instabilities in early VMC energy minimizations motivated use of variance minimization,[53, 93, 149] but later algorithmic improvements, particularly the development of the linear method[113], have swung the pendulum back to energy minimization as the main approach for ground state studies. The choice of objective function is less settled for the study of excited states and both energy-based[86–92] and variance-based[70, 71, 93–95, 198] objective functions have been employed in this context.

$$\Omega(\Psi) = \frac{\langle \Psi \mid (\omega - H) \mid \Psi \rangle}{\langle \Psi \mid (\omega - H)^2 \mid \Psi \rangle} \tag{5.3}$$

$$W(\Psi) = \left\langle \Psi \,\middle|\, (\omega - H)^2 \,\middle|\, \Psi \right\rangle \tag{5.4}$$

$$E^{SA} = \sum_I w_I \frac{\left\langle \Psi^I \,\middle|\, H \,\middle|\, \Psi^I \right\rangle}{\langle \Psi^I \mid \Psi^I \rangle} \tag{5.5}$$

The functions Ω and W both use an input energy parameter ω that must be adjusted to transform them into the variance σ^2 in order to obtain size-consistent results.[95] For Ω, the value ω must eventually be set to the difference of the energy and standard deviation, $E - \sigma$, while for W, ω is set to the energy E. There are multiple strategies for varying omega, either allowing it to float based on the values of E and σ over the course of a single VMC optimization, or performing multiple VMC optimizations at fixed ω and changing the value after each one until self-consistency is achieved. From the standpoint of optimization stability, allowing the value of ω to vary is one potential source of instability as stochastic uncertainties in the value assigned to ω could possibly lead to the targeting of a different state and we investigate this possibility below. We also note that there are various options in the details of varying ω, such as allowing it to float only after a period of interpolation from its initial value[95] or controlling its value through a running average.

The effectiveness of stochastic estimation of different quantities is another factor in the stability implications of the choice of objective function. Sampling from the commonly chosen distribution $\rho(\mathbf{R}) = \frac{\Psi(\mathbf{R})^2}{\int d\mathbf{R} \Psi(\mathbf{R})^2}$ has a zero variance property[145] when Ψ is an exact eigenstate, but the use of approximate Ψ in practice leads to an infinite variance problem for the estimation of the variance itself as well as target function derivatives. While the infinite variance of the variance makes optimization of variance-based target functions more difficult, multiples approaches for mitigating the issue have been developed, including the use of other

importance sampling functions,[67–71, 109] modifications to estimators,[145, 243, 244], and regularization schemes.[110, 111] We employ an importance sampling function that we have successfully used in the minimization of the Ω objective function in other work.[71, 257]

$$|\Phi|^2 = |\Psi|^2 + c_1 \sum_i |\Psi^i|^2 + c_2 \sum_j |\Psi^j|^2 + c_3 \sum_k |\Psi^k|^2 \tag{5.6}$$

In $|\Phi|^2$, the coefficients c_1,c_2,c_3 weight the sums of squares of wave function parameter derivatives, for Jastrow, CI, and orbital parameters respectively and we set $(c_1, c_2, c_3) = (0.0, 0.0001, 0.0)$ throughout our results.

Parameter Optimization Algorithms

A variety of algorithms have been developed for performing parameter optimization in the context of VMC. Some of the most notable ones include the Newton method,[91, 115, 120–122] linear method,[112–115] stochastic reconfiguration,[123] and accelerated descent approaches.[84, 116–119] We focus our discussion on only the aspects of the LM and AD that are most relevant for optimization stability and further details on these methods can be found in the literature.[84, 112–119, 129]

The LM relies on a first order Taylor expansion of the trial wave function and in the case of minimizing Ω leads to a generalized eigenvalue problem

$$(\omega - \mathbf{H})\,\mathbf{c} = \lambda\,(\omega - \mathbf{H})^2\,\mathbf{c} \tag{5.7}$$

with matrix elements of the form

$$\langle \Psi_i \,|\, \omega - H \,|\, \Psi_j \rangle \tag{5.8}$$

and

$$\left\langle \Psi_i \,\middle|\, (\omega - H)^2 \,\middle|\, \Psi_j \right\rangle . \tag{5.9}$$

where Ψ_i and Ψ_j are wave function derivatives with respect to variational parameters. The LM requires stochastic estimates of these matrix elements to compute an update $\mathbf{c}$ to the wave function parameters. The size of the LM matrices will grow as the square of the number of parameters and for a fixed number of samples, the quality of the estimates of the growing number of matrix elements will decrease. This decline can potentially make the parameter updates from the LM diagonalization less reliable, risking failures of optimization. An example illustrating this issue is shown in Figure 5.1, which shows the spreads of proposed LM updates on the same Jastrow parameter for increasing total numbers of parameters. As more CI parameters are added, the spread of possible updates to this Jastrow parameter increases by orders of magnitude. The LM's eigenproblem effectively amplifies and couples the noise in the updates for individual parameters. Otherwise, the spread in the Jastrow update would remain constant regardless of how many updates are being sought for other

parameters. For larger numbers of parameters, naive acceptance of the poorer LM updates could lead to instabilities.

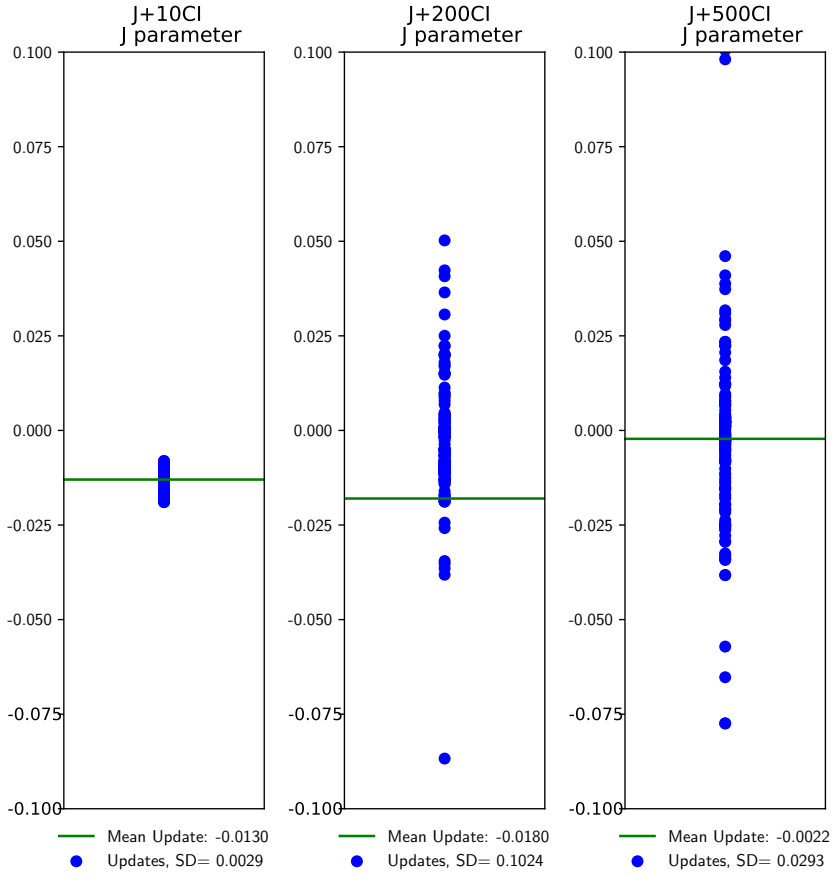


Figure 5.1: Proposed LM updates to Jastrow parameter for varying total numbers of parameters. Each panel was produced with 200 proposed updates from independent LM runs on an excited state wave function in CN5. The same scale has been used in each panel, which leaves off some larger outliers in the 200 and 500 CI cases. From left to right, the total numbers of wave function parameters for the different cases are 150, 340, and 640. 100,000 samples were used to generate the LM matrices for each update.

A number of technical modifications to the basic LM algorithm can reduce, though not eliminate, this problem. Shift values can be added to the LM Hamiltonian matrix to prevent overly large parameter changes, similar to the use of trust radius schemes in Newton-Raphson.[119, 150] Two different types of shifts, defined in equations 5.10 through 5.14, can

help address distinct problems in the LM optimization. The diagonal shift c_I helps stabilize the LM by effectively adding an energy penalty to parameter changes to the current wave function.[113] While this shift reduces the effective step size of the LM, it also changes the direction of the LM updates by rotating them to the steepest descent direction in the limit of an infinite shift that makes H diagonal. Figure 5.2 demonstrates the effect of a large diagonal shift in controlling the noise in parameter updates. The spread of proposed updates for the Jastrow parameter is very small and also remains constant across the different total numbers of parameters, which is the behavior expected in steepest descent.

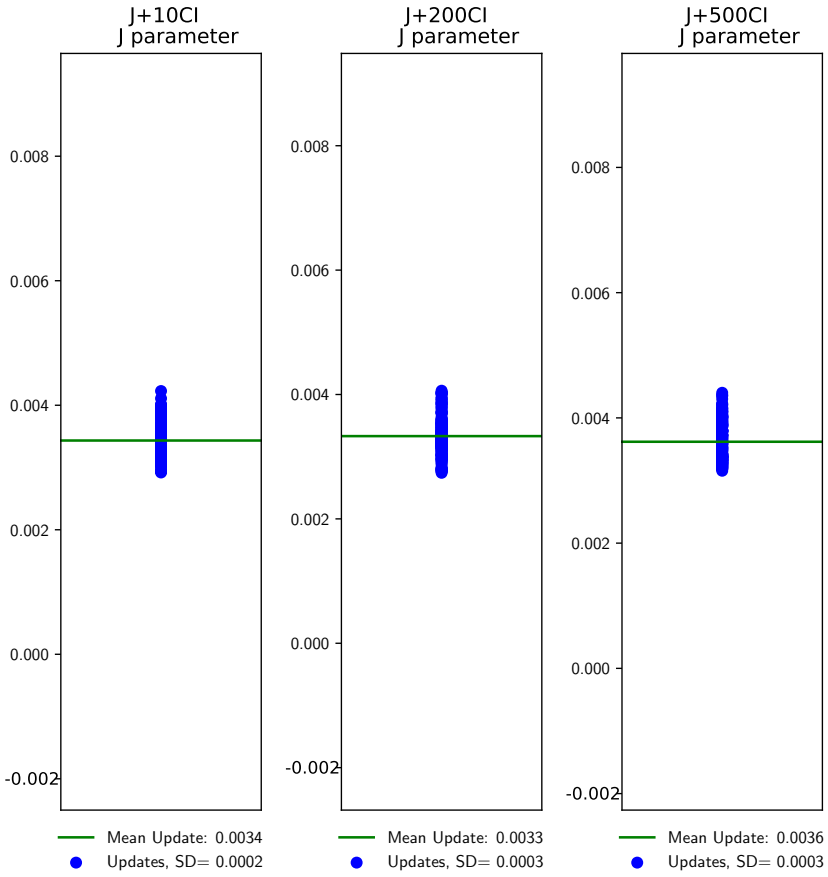


Figure 5.2: Proposed LM updates to Jastrow parameter with diagonal shift of 100 for varying total numbers of parameters. Each panel shows 200 proposed updates from independent LM runs with an excited state wave function in CN5. From left to right, the total numbers of wave function parameters for the different cases are 150, 340, and 640. 100,000 samples were used to generate the LM matrices for each update.

While the diagonal shift prevents excessive changes in parameters with large derivatives, assigning it a large enough value for this task may turn off the optimization of parameters with smaller, but still significant gradients. To avoid this scenario, the $c_S\mathbf{B}$ term helps retain flexibility in other parameter directions with smaller derivatives. The overlap shift c_S imposes an energy penalty based on the norm of parameter directions orthogonal to the current wave function so that steps in the directions of small derivatives are penalized less heavily and meaningful changes can still occur. In practice, the matrix $\mathbf{B}$ does not need to be constructed explicitly, as the LM generalized eigenvalue problem can be solved by Krylov iteration.[150]

$$\mathbf{H} \rightarrow \mathbf{H} + c_I\mathbf{A} + c_S\mathbf{B} \tag{5.10}$$

$$A_{ij} = \delta_{ij}(1 - \delta_{i0}) \tag{5.11}$$

$$\mathbf{B} = (\mathbf{Q^T})^{-1}\mathbf{T}\mathbf{Q}^{-1} \tag{5.12}$$

$$Q_{ij} = \delta_{ij} - \delta_{i0}(1 - \delta_{j0})S_{0j} \tag{5.13}$$

$$T_{ij} = (1 - \delta_{i0}\delta_{j0})[\mathbf{Q^T}\mathbf{S}\mathbf{Q}]_{ij} \tag{5.14}$$

In our own implementation of the linear method, the use of shifts is enhanced in combination with an adaptive scheme[113–115] where three sets of shift values are used to determine parameter updates, and a correlated sampling procedure either selects the candidate update expected to reduce the objective function the most, or rejects all the possible updates and raises the shifts on the next LM iteration for more cautious steps. Despite the ability to reject poor steps, we will show in our results that the LM can still fail to lower the objective function in some circumstances.

A more substantial modification to the LM is to divide the parameters into blocks and solve a set of smaller eigenvalue problems to produce the parameter update.[129] The blocked LM algorithm consists of two phases. First, an LM-style diagonalization is performed for each of the N_b blocks of parameters and the N_k lowest eigenvectors are retained from each block. In the second phase, the full set of N_bN_k eigenvectors plus N_o other parameter directions (either from previous blocked LM iterations or descent, in the case of the hybrid method) form the space for a final LM diagonalization that gives the overall update. This approach provides the blocked LM with a lower memory footprint compared to the standard LM, but reduces the flexibility of the directions it can explore in parameter space. By working with smaller matrices, the blocked LM may obtain more reliable estimates of its matrix elements for a fixed number of samples than the standard LM would and potentially achieve more reliable update steps. However, in the limit of more variational parameters with insufficient sampling effort, the blocked LM will share the same vulnerabilities as the standard LM.

AD methods are a class of first-order optimization algorithms that use only the gradient of the target function, retaining memory of gradients from previous iterations in order to converge more quickly than steepest descent. Many different flavors of AD have been widely used in the machine learning community and a number of them have been applied in the context of VMC.[84, 116–119] Below, we give the defining equations for a combination of RMSprop and Nesterov momentum as used in a number of VMC studies.[71, 116, 119, 257]

$$p_i^{k+1} = (1 - \gamma_k e^{-(\frac{1}{d})(k-1)}) q_i^{k+1} - \gamma_k e^{-(\frac{1}{d})(k-1)} q_i^k \tag{5.15}$$

$$q_i^{k+1} = p_i^k - \tau_k \frac{\partial \Omega(\mathbf{p})}{\partial p_i} \tag{5.16}$$

$$\lambda_0 = 0 \qquad \lambda_k = \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\lambda_{k-1}^2} \qquad \gamma_k = \frac{1 - \lambda_k}{\lambda_{k+1}} \tag{5.17}$$

$$\tau_k = \frac{\eta}{\sqrt{E[(\frac{\partial \Omega}{\partial p_i})^2]^{(k)}} + \epsilon} \tag{5.18}$$

$$E[(\partial\Omega)^2]^{(k)} = \rho E\left[\left(\frac{\partial \Omega}{\partial p_i}\right)^2\right]^{(k-1)} + (1 - \rho)\left(\frac{\partial \Omega}{\partial p_i}\right)^2 \tag{5.19}$$

The recurrence relations defined in equations 5.15 through 5.17 define the momentum in this algorithm. Obtaining the updated parameter p_i^{k+1} depends on the values of the target function gradient $\frac{\partial \Omega(\mathbf{p})}{\partial p_i}$ on previous iterations as well as current step, which accelerates the convergence compared to steepest descent. Equations 5.18 and 5.19 define the RMSprop adaptive modification of the step size τ_k, which is parameter-specific.

From a stability perspective, AD approaches offer a number of advantages compared to the LM. The relevant equations are far more linear than the LM's generalized eigenvalue problem, which reduces the dangers of step bias and uncertainty from nonlinear combinations of stochastic quantities. In practice, VMC optimization with AD can be conducted with a significantly lower per-iteration sampling effort than the LM without creating instabilities. While an individual AD step may be poor and raise the objective function, modest step sizes ensure that the change to the wave function is small, and typical AD optimizations use many hundreds or thousands of iterations to successfully minimize the objective function on average. However, these stabilizing features of AD also leave it slower to converge to the minimum than the LM.[119] This weakness has motivated the development of a hybrid approach that alternates between sections of AD and blocked LM optimization, in order to allow the blocked LM to more swiftly move parameters to their optimal values.[71, 119, 257]

Wave Functions

Another key aspect of VMC stability is the choice of wave function, which shapes the target function landscape and the difficulty of the optimization problem the various optimizer algorithms face. With the exact wave function and infinite sample size, variance minima will

exist for all the Hamiltonian eigenstates.[93] For approximate ansatzes and finite sampling effort, the minima may become shallow enough that some optimizers fail to target and converge to them or they may not exist at all. Therefore, as a practical matter, being able to systematically improve the wave function and the amount of sampling may make the difference between success and failure for a given optimization algorithm. We have provided a recent example of this applying the LM to simple wave functions in CN5, where adding a 3-body Jastrow factor and increasing sampling eliminated instabilities.[71]

In this work, we consider more complex wave functions in CN5, all of the Multi-Slater Jastrow form shown below. The Slater determinant expansion ψ_{MS} can be obtained from various active space calculations of different sizes, with larger spaces, potentially solved by sCI, producing longer and more accurate expansions.

$$\Psi = \psi_{MS}\psi_J \tag{5.20}$$

$$\psi_{MS} = \sum_{i=0}^{N_D} c_i D_i \tag{5.21}$$

$$\psi_J = \exp\left\{ \sum_i \sum_j \chi_k(|r_i - R_j|) + \sum_k \sum_{l>k} u_{kl}(|r_k - r_l|) \right\} \tag{5.22}$$

Throughout our results, we employ and optimize one- and two-body Jastrow factors, which are constructed with splines for the functions χ_k and u_{kl}.[150] For our stability analysis, we also optimize the orbitals of our Slater expansions, benefiting from recent methodological improvements with the table method.[75, 76] However, as we show in our results, orbital optimization can still be challenging for the LM, and we also consider VMC optimization with the orbitals left at shapes obtained from a recent state-specific CASSCF approach.[45, 50, 219] This approach of combining state-specific quantum chemistry with VMC has recently been shown to provide accurate excitation energies across a range of different types of excited states[257] and we provide another example of this for the excited state in CN5.

5.4 Results

Computational Details

We perform all VMC optimizations in a development version of QMCPACK.[150, 204] Our molecular geometry is identical to one recently used by Filippi and coworkers[91, 259] and we repeat the coordinates in the appendix for convenience. Basis set and pseudopotential choices are specified below for the particular wave function cases we consider. For generating our wave functions, we have performed CASSCF calculations in Molpro[170] and PySCF,[205] as well as sCI calculations in Dice.[10, 131] Our one- and two-body Jastrow factor splines each consist of 10 coefficients defining the function within a cutoff distance of 10 bohr.

Excitation Energies from Variance Minimization

We begin by simply checking the ability of variance-based excited state VMC to obtain accurate excitation energies with the hybrid method. We follow the same methodology that has recently been applied on a broader set of excited states.[257] We use a cc-pVTZ basis set with pseudopotentials[171] and obtained state-specific CASSCF orbitals in a (6e,5o) space. We then generate Slater determinants from heatbath CI in a (28e,40o) space of these orbitals. The orbitals are left at their SS-CASSCF shapes and we optimized only Jastrow parameters and CI coefficients in VMC with the hybrid method, using a recent parameter selection scheme[257] to optimize only parameters with significant derivatives within the blocked LM. For the blocked LM portions, we divided those parameters into 5 blocks and retained 30 parameter directions per block along with 5 directions from AD for the second phase of the blocked LM algorithm. No stability issues were observed in these optimizations. Our VMC wave functions use 500 determinants for the ground state and 1000 for the excited state. A recently introduced variance matching extrapolation procedure[257] is used to ensure a balanced treatment of the ground and excited states.

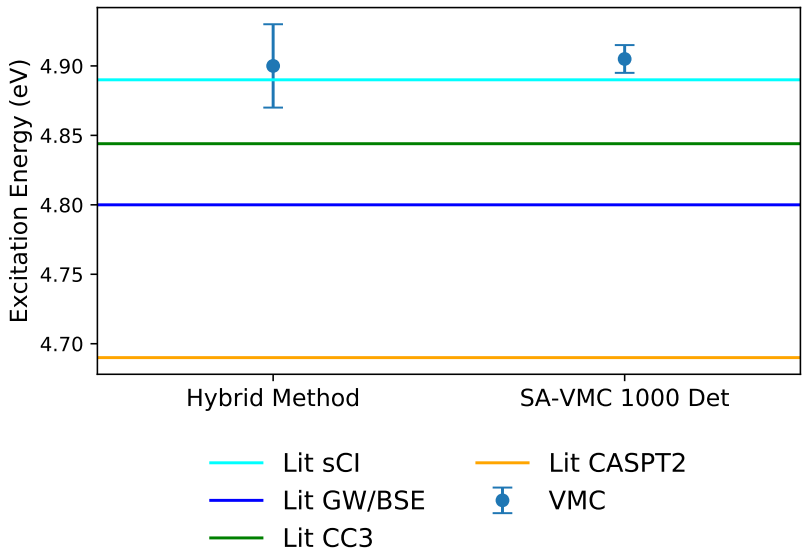


Figure 5.3: Excitation energy in CN5 using the hybrid method to minimize the Ω target function. Reference values for SA-VMC,[91] CC3,[91] CASPT2,[88, 259] GW/BSE,[259] and sCI[11] are taken from the literature. For comparison, the basis sets used in the literature results were aug-cc-pVTZ for CC3 and GW/BSE, aug-cc-pVDZ for sCI, ANO-L-VTZP for CASPT2, and a double-ζ basis set minimally augmented with s and p diffuse functions on heavy atoms for SA-VMC.

Table 5.1: Excitation energies for CN5 for state-specific VMC and literature results.[11, 88, 91, 259] Stochastic uncertainties on the last digit are given in parentheses.

Method	Excitation Energy (eV)
Hybrid Method VMC	4.90(3)
SA-VMC	4.91(1)
CASPT2	4.69
GW/BSE	4.80
CC3	4.84
sCI	4.89

As shown in Figure 5.3 and Table 5.1, we obtain a highly accurate excitation energy for this state in CN5 for our variance-based methodology. Our result is with 0.1 eV of literature results from sCI[11] and CC3[91] and statistically indistinguishable from a VMC

result from Filippi and coworkers based on minimizing the state-averaged energy.[91] Thus, as a practical matter, we have some reassurance that variance based minimizations with the hybrid method can achieve accurate excitation energies without stability concerns.

Stability Tests

To gain some more insight into cases where stability issues might arise, we consider a series of excited state optimizations with the LM and the hybrid method. Our focus is now mainly on whether these optimizations progress stably rather than whether they fully reach the minimum or the precise accuracy of the excitation energies that might be obtained from them. The differences in technical optimization choices between the stable and unstable outcomes we observe may assist the work of other VMC practitioners.

To construct an ansatz for this analysis, we consider a (6e,10o) active space in a cc-pVDZ basis set with BFD pseudopotentials.[168] This is the same basis, pseudopotential and active space for which stability issues were recently observed.[91] We use a 0.001 coefficient cutoff for a 4-state-average CASSCF to obtain a CI expansion of 1892 determinants for lowest excited state of CN5 and add one- and two-body Jastrow factors. The lowest excited state is a ground state within its symmetry and could have been described with a single state CASSCF, but here we allow determinants of other symmetries (which had coefficient above 0.001 for other states in the state-average but have zero coefficients for the lowest excited state) within our ansatz. The presence of these unnecessary determinants increases the difficulty of the optimization and the use of state-averaging may enhance the amount of orbital optimization being asked of VMC. While this ansatz is not as closely tailored for the excited state as quantum chemistry might allow, it remains a reasonable initial guess that still requires nontrivial optimization at the VMC level.

In order to assess stability issues that may appear gradually, we run our optimizations for a considerably larger number of iterations than we otherwise would for typical excitation energy predictions, aiming for many hundreds of LM iterations. We also provide the LM (and its blocked variant within hybrid method) with only 100,000 samples per iteration in order to make long optimizations more practical and somewhat favor the potential emergence of stability issues due to poorer stochastic estimates. For comparison, we note that for a molecule as large as CN5, our typical practices would provide a million samples per LM iteration, potentially more in cases with many ansatz parameters, and only optimize for somewhere in the range of 50 to 100 iterations. For the hybrid method, we use 30,000 samples per AD iteration, which is a typical amount otherwise, and use 17 macro-iterations, which is significantly longer than our usual range of 5 to 10.

We begin with a fixed value of $\omega = -41.475$ and minimize the objective function Ω in a staged approach, first optimizing the less difficult Jastrow and CI parameters and only turning on orbital optimization at the next stage. We choose to optimize all determinants from CASSCF alongside the initially unoptimized Jastrow for a total of 2030 variational parameters in our first stage of optimizations. We mention that in practical calculations, this stage could be performed even more cautiously, such as by optimizing the Jastrow with

only the 100 most important determinants and then adding the remainder in sets of a few hundred with zero initial coefficients, but optimizing them all at once from the start provides us with a more stringent stability test.

Three optimizer choices are considered: LM with fixed shifts, LM with adaptive shifts, and the hybrid method. These various options' approach to the issue of step size control is a key way of distinguishing them. In both varieties of the LM as well as the blocked LM steps within the hybrid method, a correlated sampling procedure is performed to allow the possibility of rejecting a proposed parameter update. However, the fixed shift version of the LM has no means of constraining its step size beyond what the chosen shift values($c_I = 0.1$ and $c_S = 1$) provide while the adaptive version can move from those initial values and alter both the size and direction of the LM updates. The hybrid method incorporates the benefits of the adaptive shift scheme in its use of the blocked LM and the step sizes employed during its AD sections dynamically adjust with the objective function gradients according to RMSprop. Throughout the hybrid method optimizations in this section, we divide parameters into 5 blocks and employ 30 parameter directions per block along with 5 directions from AD in the second phase of all blocked LM iterations.

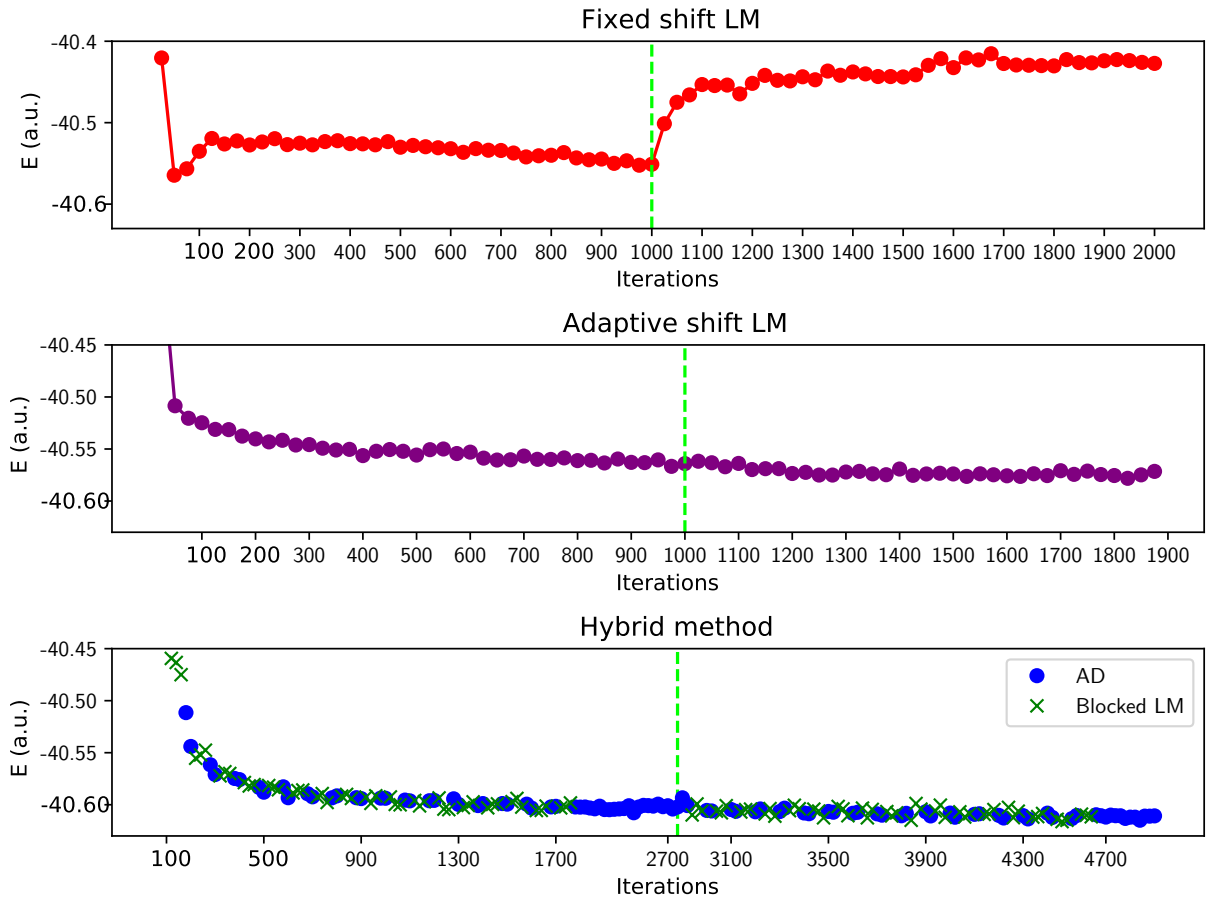


Figure 5.4: Excited state energy for fixed shift LM, adaptive shift LM, and hybrid method. Starting point was unoptimized Jastrow and CI coefficients from SA-CASSCF. Green dashed line marks the end of the first stage where only Jastrow and CI parameters are optimized.

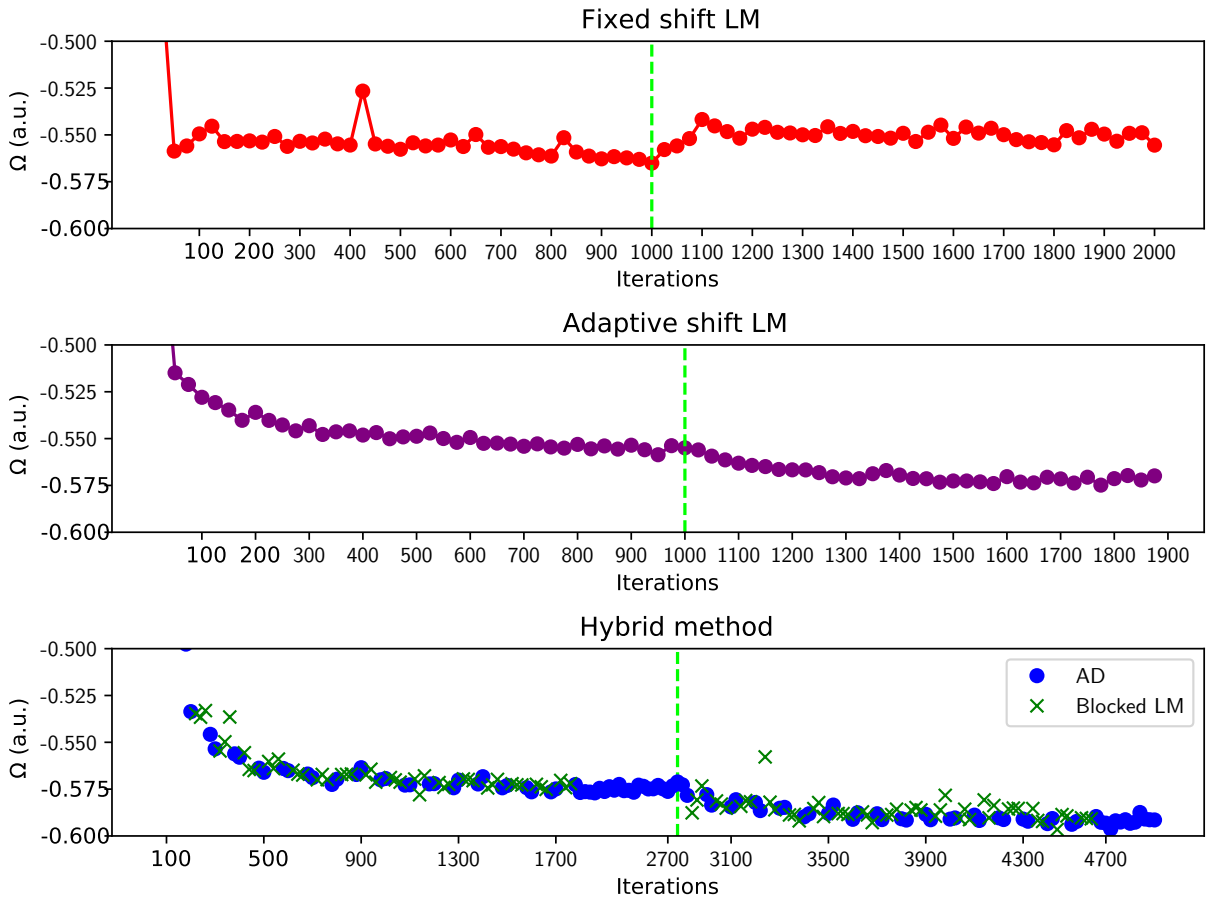


Figure 5.5: Objective function Ω for fixed shift LM, adaptive shift LM, and hybrid method. Starting point was unoptimized Jastrow and CI coefficients from SA-CASSCF.Green dashed line marks the end of the first stage where only Jastrow and CI parameters are optimized.

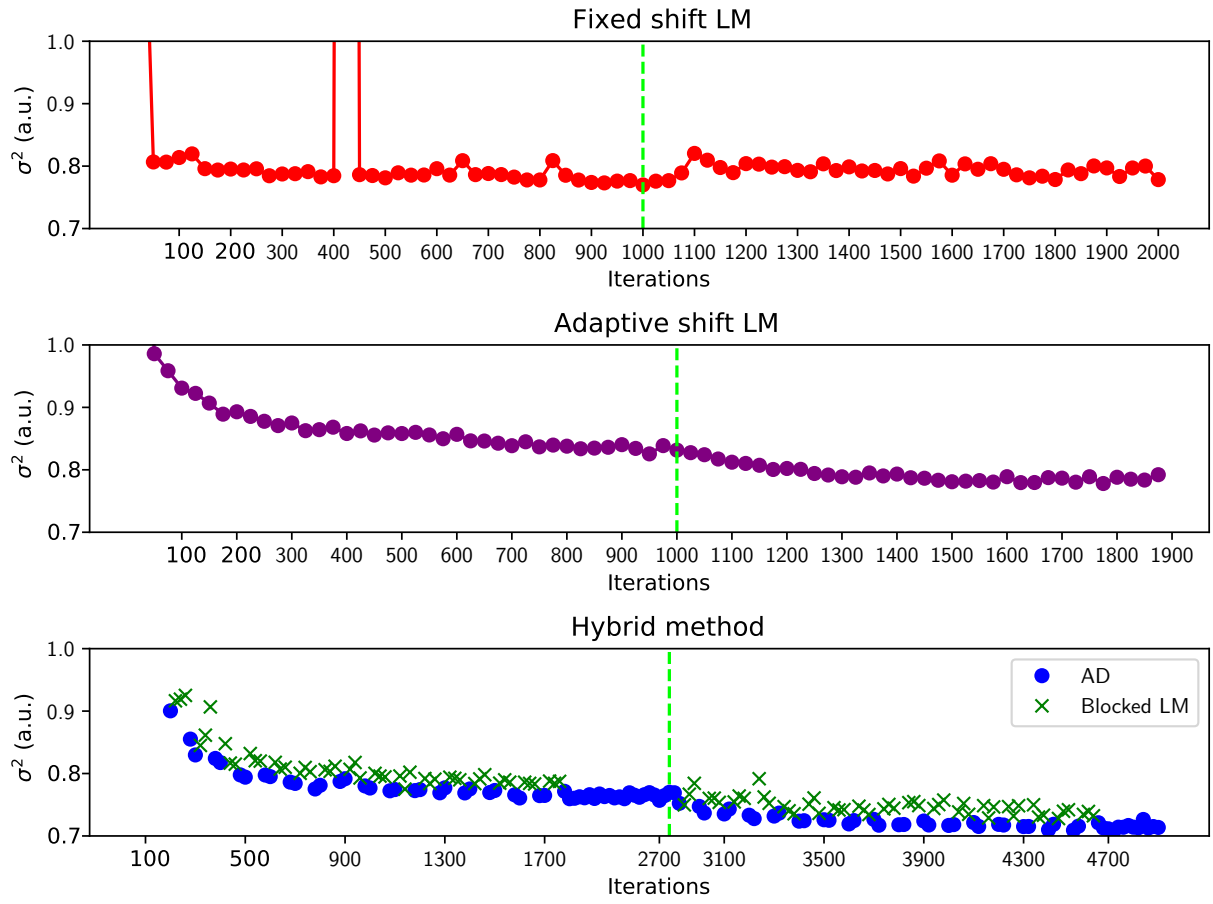


Figure 5.6: Excited state variance for fixed shift LM, adaptive shift LM, and hybrid method. Starting point was unoptimized Jastrow and CI coefficients from SA-CASSCF. Green dashed line marks the end of the first stage where only Jastrow and CI parameters are optimized. The separation between AD and blocked LM points is an artifact of autocorrelation.

Turning to the results, we consider the optimizations shown in Figures 5.4 through 5.6. Points on the LM plots represent the average of 25 LM iterations. On the hybrid method plots, dots are averages of 50 AD iterations and crosses are individual blocked LM iterations. The vertical scales have been chosen to make the varying stability behavior of the optimization algorithms more visible, which leaves off some points corresponding to early iterations. The vertical dashed line marks the end of the first stage of optimizing only Jastrow and CI parameters and the start of including orbital optimization. We note that while both stages of optimization are plotted together, the second phase was run as separate calculations. Initial values of the shifts were restored to $c_I = 0.1$ and $c_S = 1$ for the adaptive shift LM and the hybrid method, and the initial step sizes used in AD are changed. The initial hybrid method

optimization of CI and Jastrow parameters used step sizes of 0.01 and 0.1 respectively. Once orbitals were turned on, these were reduced to 0.005 and orbital parameters were given a step size of 0.0001.

Focusing on the left-hand Jastrow and CI side of the three figures, we see that all approaches successfully lower the target function with the energy and variance also decreasing in tandem. The fixed shift LM case in the upper panel of the three figures provides some reason for concern with a sustained rise in the target function and energy within the first 100 iterations, but the optimization later recovers. The adaptive shift LM and the hybrid method provide smoother optimization, but for this case with only the easier Jastrow and CI parameters, their refinements for step size control may be less necessary.

We then turn on orbital optimization and continue to optimize the other parameters starting from each method's Jastrow and CI result. We continue to keep ω fixed. With orbital optimization, our ansatz now has a total of 6163 variational parameters. We now see in the second half of Figures 5.4 through 5.6 that the behavior of the three methods more dramatically differs. The fixed shift LM shows a substantial rise in energy of roughly 80 milliHartree over the course of hundreds of iterations. This failure of optimization is clearly signaled by a rise in the objective function in the first hundred iterations of orbital optimization. Both the energy and objective function remain elevated during hundreds of the later iterations, indicating that the LM is unable to recover in this case. While the fixed shift LM in principle possesses the ability in the correlated sampling phase to detect and reject any step that would raise the objective function, this safeguard is clearly inadequate in this case. With poor estimates of the objective function at the current and proposed sets of ansatz parameters, the LM may mistakenly estimate that a step will lower the objective function only to discover otherwise once it is at the new point in parameter space and takes new samples. This danger is exacerbated if proposed parameter steps can remain incautiously large, which fixing the LM shifts permits. Our finding of optimization failure with limited step control aligns with other observations of variance minimization instabilities when using a modified Newton method with a fixed, non-adaptive step size parameter.[91]

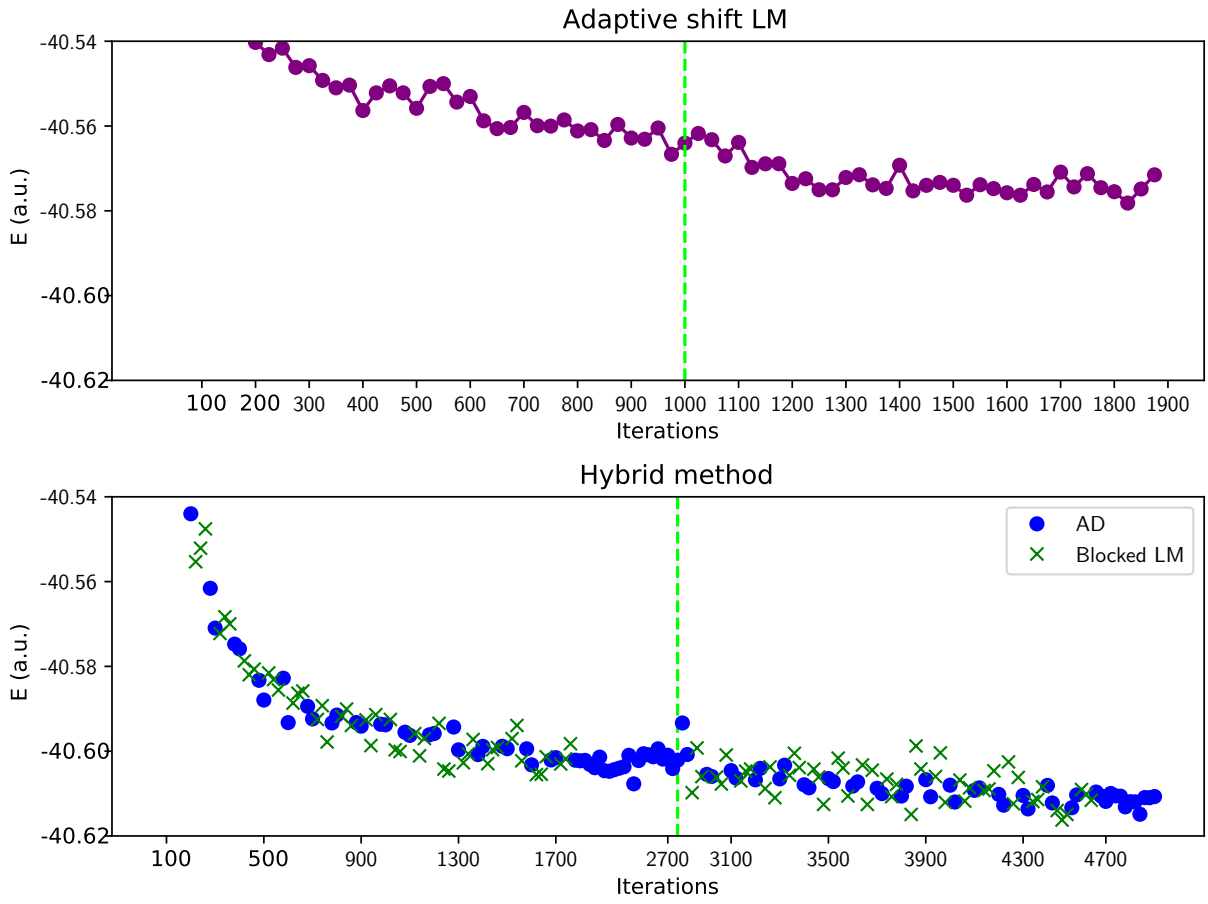


Figure 5.7: Zoomed in view of excited state energy for adaptive shift LM, and hybrid method. Green dashed line marks the end of the first stage where only Jastrow and CI parameters are optimized.

The change from fixed to adaptive shifts yields far better performance for the LM, with steady decreases in objective function, energy, and variance. We provide a zoomed in plot of the energy for the adaptive shift LM and hybrid method in Figure 5.7 to make clear that both approaches continue to lower the energy after enabling orbital optimization. The adaptive shift LM optimization makes only rather slow progress over many hundreds of iterations and after roughly 300 iterations with orbital optimization, the shifts are so large that only very minor changes are being made to the parameters in the remainder of the calculation. While this sluggish optimization would be undesirable in normal VMC applications, it clearly does not show any pathological behavior in terms of instability even though the adaptive shift LM is operating with only 100,000 samples per iteration. Although this low number of samples may hamper the LM's ability to determine effective parameter updates that would

minimize more quickly, any stability challenges it might pose have been overcome with enhanced step control in the algorithm's design. Essentially, the adaptive shifts prevent the optimizer from pushing past the point where statistical uncertainty precludes effective update steps. The hybrid method is similarly stable and able to achieve significantly lower values of Ω, energy, and variance. The blocked LM iterations show some scatter in the objective function and the variance, but do not derail the stability of the optimization. The separation between the blocked LM and AD variance points in Figure 5.6, also seen in Figure 5.10, is an autocorrelation artifact from the low number of samples used to evaluate the AD variances being averaged together in each point and diminishes if AD iterations are given more samples.

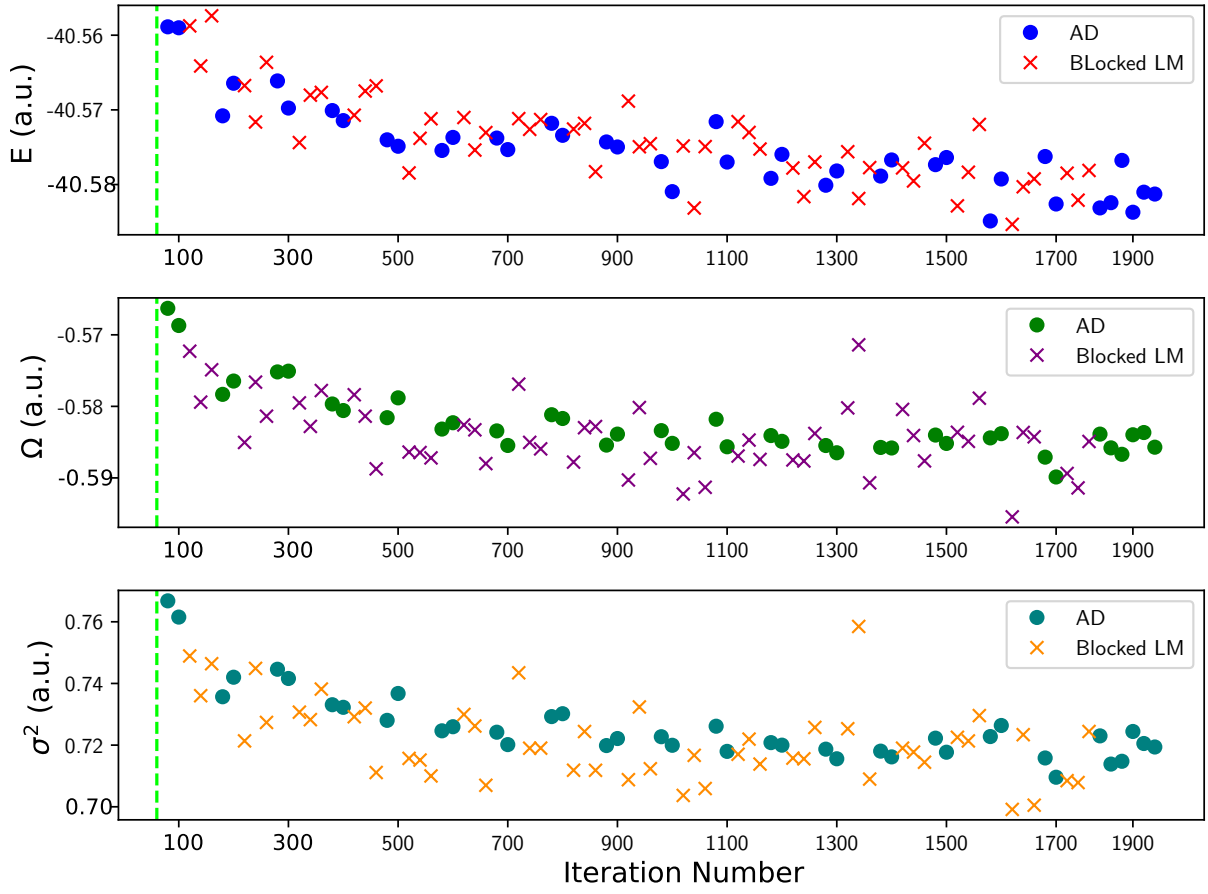


Figure 5.8: Optimization of all parameters including orbitals using the hybrid method starting from the end of the LM fixed shift optimization of only the Jastrow and CI parameters. Green line is shown to the left to make clear these plots show only the stage with orbital optimization.

We also consider the potential effects of the optimized MSJ starting point on the second stage of optimization. As depicted earlier in Figure 5.5, the hybrid method achieves a lower value of Ω than the two types of the LM during the first Jastrow and CI phase of optimization and therefore has a better initial wave function when orbital optimization is turned on. To check whether this difference has any stability consequences, we also performed a hybrid method optimization of all parameters beginning from the optimized MSJ ansatz obtained by the fixed shift LM. As shown in Figure 5.8, the hybrid method is still able to stably optimize from this inferior starting point. While the quality of the initial wave function may still influence how fully the hybrid method is able to minimize Ω, a somewhat poorer starting point is not sufficient to destabilize the hybrid algorithm.

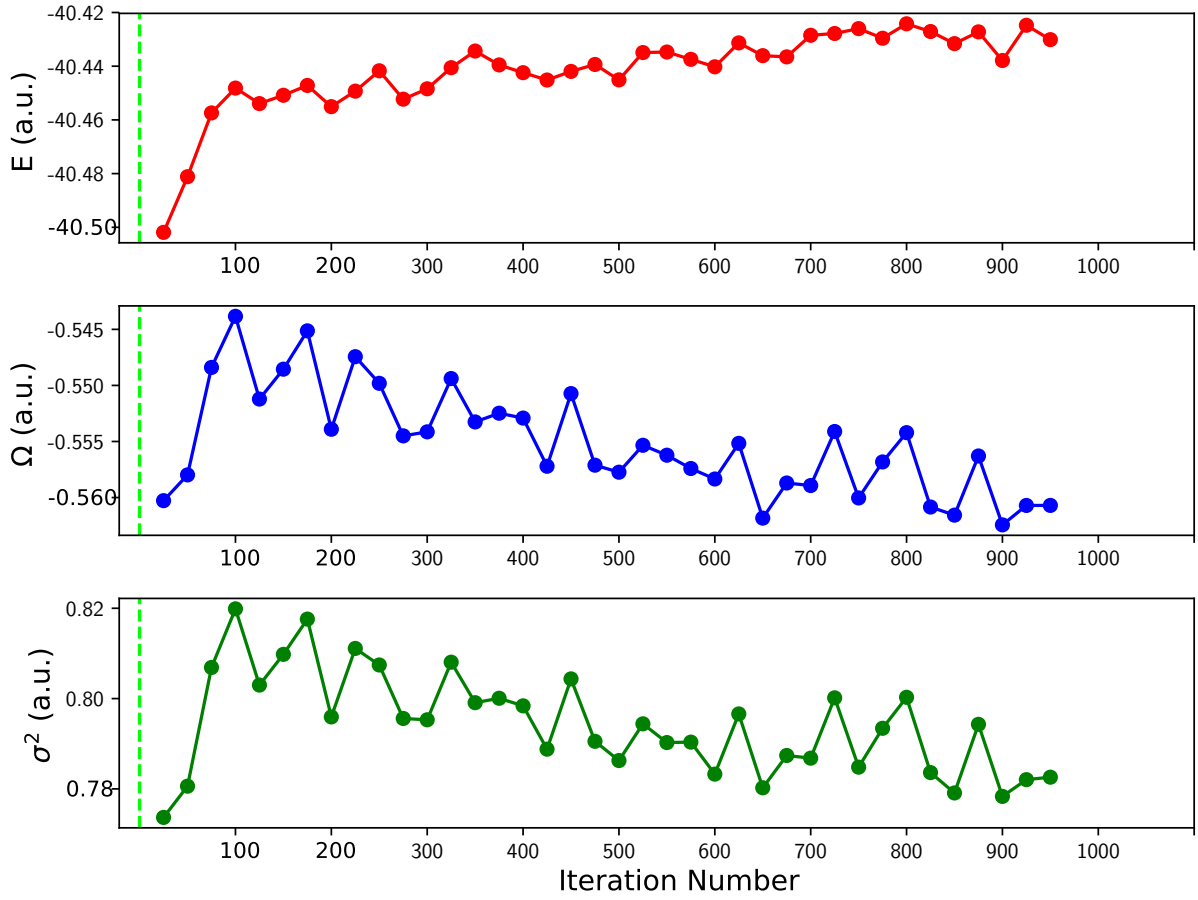


Figure 5.9: Optimization of all parameters including orbitals using the LM with fixed shifts while varying omega. Starting point was the optimized MSJ wave function from the preceding LM fixed shift optimization of only the Jastrow and CI parameters. Green line is shown to the left to make clear these plots show only the stage with orbital optimization.

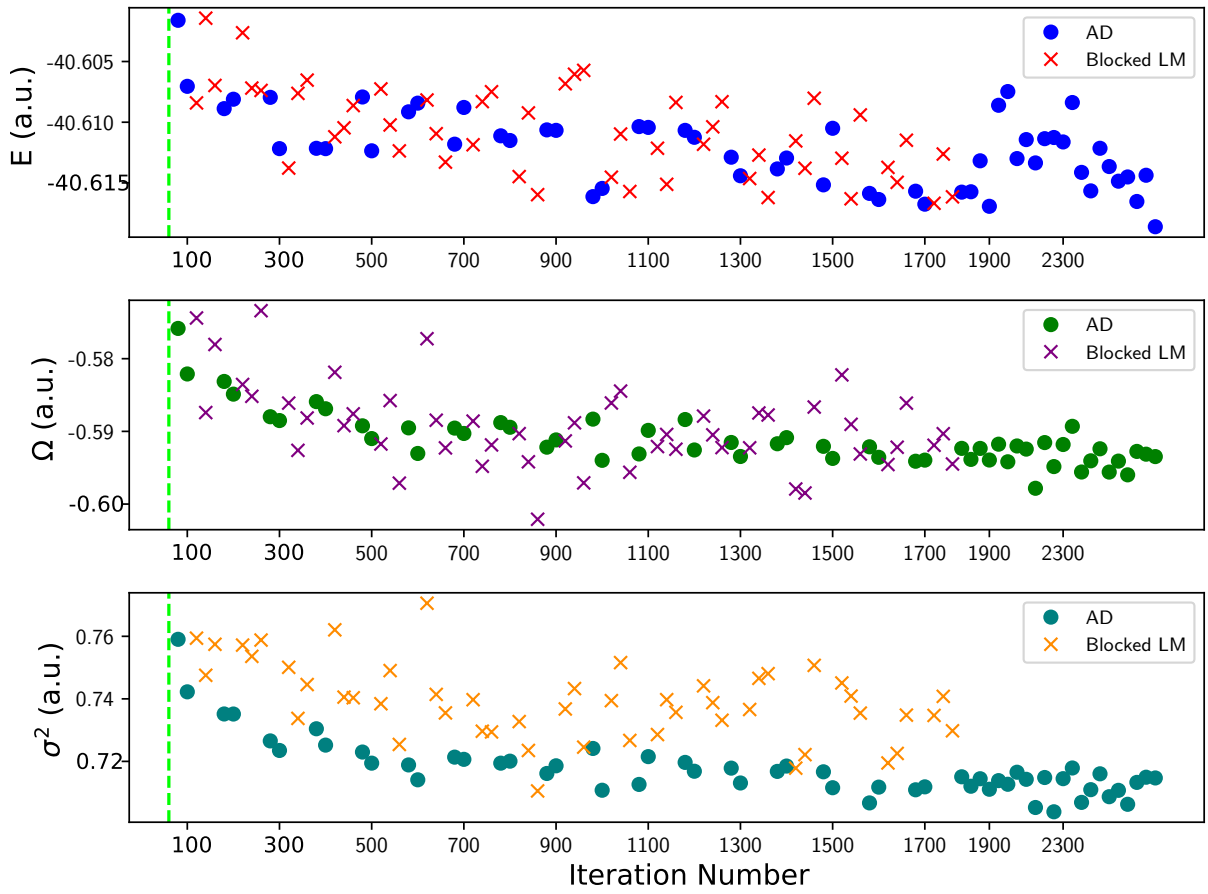


Figure 5.10: Optimization of all parameters including orbitals using the hybrid method while varying omega. Starting point was the optimized MSJ wave function from the preceding hybrid method optimization of only the Jastrow and CI parameters. Green line is shown to the left to make clear these plots show only the stage with orbital optimization. The separation between AD and blocked LM variance points is an artifact of autocorrelation.

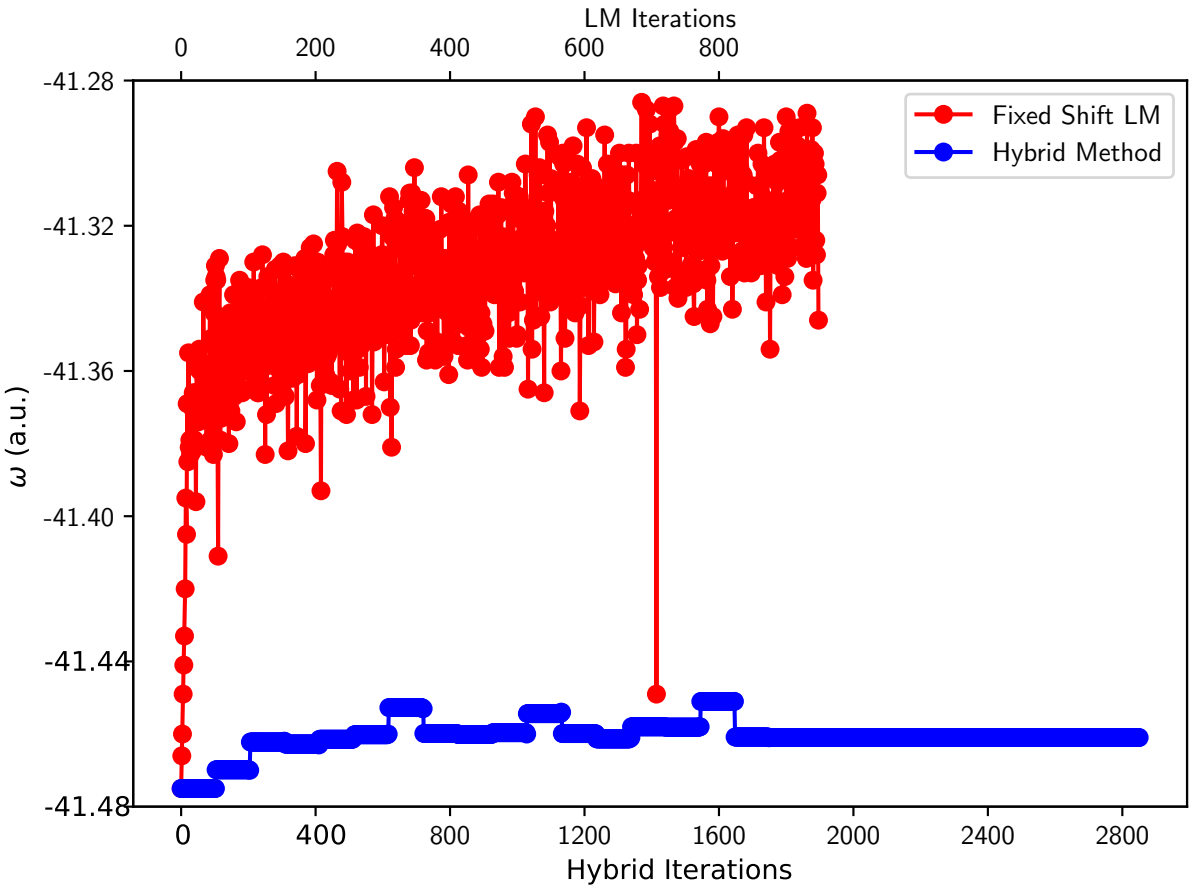


Figure 5.11: Varying values of omega during fixed shift LM and hybrid method optimizations of all parameters.

The optimizations shown so far have all used a fixed value of ω. While in practical calculations, a series of fixed ω calculations is always an option for moving ω to transform minimization of Ω into variance minimization, for completeness we also consider varying ω over the course of a single optimization. For the fixed shift LM and the hybrid method, we again turn on orbital optimization at their respective optimized Jastrow and CI starting points, now with schemes for varying ω. For the LM, during the first 10 iterations, we interpolate[95] ω between its initial value and $E - \sigma$, and afterward allow ω to float at the value of $E - \sigma$ set by the preceding iteration. For the hybrid method, we use the last 30 AD iterations in each macro-iteration to compute $E - \sigma$ for setting a new value of ω for the blocked LM steps and the AD section of the next macro-iteration. As seen in Figures 5.9 and 5.10, there is little difference in the methods' stability behavior from the fixed ω case. The fixed shift LM again shows a clear rise in the energy and variance while the hybrid

method still performs a stable minimization. As Figure 5.11 shows, the value of ω is dragged up significantly during the unstable LM optimization, while it only moves a little during the hybrid method. The hybrid method's stability suggests that varying ω and changing the objective function landscape are not the key factors in determining stability. Instead, we see again that adaptive step control makes the difference.

5.5 Conclusions

The sequences of optimizations we have performed point to several considerations when seeking to ensure stable and successful applications of variance-based state-specific VMC. Scenarios characterized by large numbers of difficult parameters, as in orbital optimization, pose a significant risk of optimization instability from poor parameter updates. This difficulty is exacerbated if the amount of sampling is insufficient for accurately determining the parameter gradients for those updates or the objective function values needed to assess and reject them. Our findings on stability failures of the LM add to other observations of challenges in variance minimization[91] and help clarify when such optimization difficulties are most likely to be encountered. The nature of the optimization failures we identify point to choices in optimizer design as a key factor in variance minimization instabilities rather than the shape of the variance surface itself. While the problem of optimization instability will remain a concern as researchers pursue studies of larger systems and more complex ansatzes, multiple tools are available for curbing its impact.

First, the ability to adaptively control step size and direction is a key element of robustness in challenging optimizations. Previously reported instabilities were seen in a setting that lacked adaptive step control, as were all the unstable runs in the present study. In contrast, optimizations in which we employ adaptive step control (either in the LM alone or in the hybrid method) proved stable. Second, algorithms can be designed to reduce the optimization burden on the portions most prone to instability. This idea is reflected in the hybrid method's structure, with a large portion of the optimization left to the more stable AD and the reduction of the size of the LM matrices by first including only parameters of significant gradients and then blocking. Finally, careful construction of wave functions to effectively describe correlation while remaining readily optimizable at the VMC level is advantageous. We have shown determining orbital shapes with state-specific quantum chemistry and optimizing only Jastrow and CI parameters is sufficient for an accurate excitation energy in CN5 and demonstrated the same idea for a broader set of molecules in recent work.[257] Circumventing VMC orbital optimization in this way avoids any stability challenges it might pose and greatly enhances computational efficiency. Taken together, these approaches provide reassurance that state-specific variance-based VMC can be performed in a robust and stable manner so long as adaptive step control is employed.

5.6 Acknowledgements

This work was supported by the Office of Science, Office of Basic Energy Sciences, the U.S. Department of Energy, Contract No. DE-AC02-05CH11231. Computational work was performed with the Berkeley Research Computing Savio cluster and the LBNL Lawrencium cluster.

5.7 Appendix

Table 5.2: Structure of CN5. Coordinates in Å.

C	0.000 000	0.000 000	0.340 620
C	0.000 000	1.188 120	−0.363 814
C	0.000 000	−1.188 120	−0.363 814
H	0.000 000	0.000 000	1.424 530
H	0.000 000	1.160 200	−1.449 150
H	0.000 000	−1.160 200	−1.449 150
N	0.000 000	2.389 120	0.164 413
N	0.000 000	−2.389 120	0.164 413
H	0.000 000	2.531 430	1.161 940
H	0.000 000	−2.531 430	1.161 940
H	0.000 000	3.211 380	−0.414 923
H	0.000 000	−3.211 380	−0.414 923

Chapter 6

Conclusions

In this dissertation, we have presented a series of four studies centered around the common theme of optimization algorithms in VMC, with an emphasis on applications to molecular excited states. We began with VMC method development in the ground state by performing a careful comparison of ansatz optimization algorithms. The two main classes of methods we considered, the LM and AD, have complementary strengths and weaknesses. While the LM is more effective than AD at swiftly moving challenging parameters to the vicinity of their optimal values, its vulnerability to step uncertainty hampers its ability to converge tightly. AD methods enjoy greater scalability to larger parameter numbers and robustness to noise, but remain relatively slow to converge when deployed on their own. This situation motivated the construction of the hybrid method, which was shown in its initial applications to be as accurate as the standard LM and more computationally efficient.

The hybrid method served as the main tool in the subsequent work we have presented here. We first extended the algorithm to optimization for excited states by minimizing an objective function based on the variance. Applying the hybrid method to doubly excited states in a set of small molecules allowed us to verify the accuracy of our approach against benchmark sCI excitation energies for this challenging category of excited states. The subsequent research in Chapter 4 applying state-specific VMC backed by the hybrid method to a wider array of excited states demonstrated that this methodology can reliably obtain accurate excitation energies for single valence, doubly excited, and charge transfer states. This versatility across multiple classes of excited states is not available for other quantum chemistry methods, particularly at the largest molecular sizes considered in this work. We have also continued methodological work for VMC by analyzing the issue of optimization stability in variance minimization and demonstrating the hybrid method's design allows it to remain stable in cases where the standard LM might fail.

Within the wider context of electronic structure, state-specific VMC can serve as a broadly reliable and systematically improvable tool for studying challenging excited states. Its ability to build upon a very different set of approximations compared to traditional quantum chemistry methods such as coupled cluster and CASPT2 enables VMC to provide guidance in cases where the others disagree. The development of the hybrid method

supports this role of VMC by ensuring robust optimizations in larger molecules where benchmarks from sCI and similarly high quality methods are unavailable and the reliability of the affordable quantum chemistry options may be unknown.

Looking to the future, there are multiple paths for both the application and development of excited-state VMC. The study of molecular excited states could first be pushed to larger systems of interest. One set of possibilities is charge transfer states in nucleobase dimers,[260–262] which are useful model systems for studying the photophysics of DNA,[263] a crucially important topic in the study of cancer. Further categories of excited states beyond those considered in this research such as core states[198] and Rydberg states are areas where VMC could provide similarly reliable results. There are also exciting possibilities in the solid-state, where VMC may be able to bring greater predictive power to the study of band gaps[142] and defect states. As VMC's capabilities continue to mature, it could potentially be used as a stage in high-throughput materials searches[264] after cheaper methods with less predictive power have been used.

On the methodological side, the optimization methods studied in this work may be combined with new objective functions such as a generalized variational principle[48] and with other ansatz forms including neural nets.[200, 265] More broadly, developing state-specific VMC to provide high quality results for observables besides excitation energies would further expand its usefulness. The hybrid method itself may be refined further as more applications clarify how to optimally set its many hyperparameters and design it to be run in a more black-box manner. Given the long history of optimization method development in QMC, the hybrid method may well be superseded one day by another approach, but it then will still have served as a useful standard for assessing new algorithms. The many opportunities for excited state VMC make the future of its role in electronic structure theory difficult to predict, but we hope that our contributions to optimization algorithms will play a useful part.

Bibliography

[1] Trygve Helgaker, Poul Jorgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory*. Chichester: John Wiley & Sons, 2000. ISBN: 9781118531471.

[2] W. Kohn, A. D. Becke, and R. G. Parr. "Density Functional Theory of Electronic Structure". In: *The Journal of Physical Chemistry* 100.31 (1996), pp. 12974–12980. DOI: 10.1021/jp9606691. eprint: https://doi.org/10.1021/jp9606691. URL: https://doi.org/10.1021/jp9606691.

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

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

[5] Adèle D. Laurent and Denis Jacquemin. "TD-DFT benchmarks: A review". In: *International Journal of Quantum Chemistry* 113.17 (2013), pp. 2019–2039. DOI: https://doi.org/10.1002/qua.24438. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.24438. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.24438.

[6] P.J. Knowles and N.C. Handy. "A new determinant-based full configuration interaction method". In: *Chemical Physics Letters* 111.4 (1984), pp. 315–321. ISSN: 0009-2614. DOI: https://doi.org/10.1016/0009-2614(84)85513-X. URL: https://www.sciencedirect.com/science/article/pii/000926148485513X.

[7] Peter J. Knowles and Nicholas C. Handy. "A determinant based full configuration interaction program". In: *Computer Physics Communications* 54.1 (1989), pp. 75–83. ISSN: 0010-4655. DOI: https://doi.org/10.1016/0010-4655(89)90033-7. URL: https://www.sciencedirect.com/science/article/pii/0010465589900337.

[8] Janus J. Eriksen. "The Shape of Full Configuration Interaction to Come". In: *The Journal of Physical Chemistry Letters* 12.1 (2021). PMID: 33356287, pp. 418–432. DOI: 10.1021/acs.jpclett.0c03225. eprint: https://doi.org/10.1021/acs.jpclett.0c03225. URL: https://doi.org/10.1021/acs.jpclett.0c03225.

[9] Norm M. Tubman et al. "A deterministic alternative to the full configuration interaction quantum Monte Carlo method". In: *J. Chem. Phys.* 145.4 (2016), p. 044112. DOI: 10.1063/1.4955109.

[10] Adam A. Holmes, Norm M. Tubman, and C. J. Umrigar. "Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling". In: *J. Chem. Theory Comput.* 12.8 (2016), pp. 3674–3680. ISSN: 15499626. DOI: 10.1021/acs.jctc.6b00407. eprint: 1606.07453.

[11] Yann Garniron et al. "Selected configuration interaction dressed by perturbation". In: *J. Chem. Phys.* 149.6 (2018), p. 064103.

[12] Daniel S. Levine et al. "CASSCF with Extremely Large Active Spaces Using the Adaptive Sampling Configuration Interaction Method". In: *Journal of Chemical Theory and Computation* 16.4 (2020). PMID: 32109055, pp. 2340–2354. DOI: 10.1021/acs.jctc.9b01255. eprint: https://doi.org/10.1021/acs.jctc.9b01255. URL: https://doi.org/10.1021/acs.jctc.9b01255.

[13] Norm M. Tubman et al. "Modern Approaches to Exact Diagonalization and Selected Configuration Interaction with the Adaptive Sampling CI Method". In: *Journal of Chemical Theory and Computation* 16.4 (2020). PMID: 32159951, pp. 2139–2159. DOI: 10.1021/acs.jctc.8b00536. eprint: https://doi.org/10.1021/acs.jctc.8b00536. URL: https://doi.org/10.1021/acs.jctc.8b00536.

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

[15] Kerstin Andersson and Björn O. Roos. "Excitation energies in the nickel atom studied with the complete active space SCF method and second-order perturbation theory". In: *Chemical Physics Letters* 191.6 (1992), pp. 507–514. ISSN: 0009-2614. DOI: https://doi.org/10.1016/0009-2614(92)85581-T. URL: https://www.sciencedirect.com/science/article/pii/000926149285581T.

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

[17] Christopher J. Stein, Vera von Burg, and Markus Reiher. "The Delicate Balance of Static and Dynamic Electron Correlation". In: *Journal of Chemical Theory and Computation* 12.8 (2016). PMID: 27409981, pp. 3764–3773. DOI: 10.1021/acs.jctc.6b00528. eprint: https://doi.org/10.1021/acs.jctc.6b00528. URL: https://doi.org/10.1021/acs.jctc.6b00528.

[18] Kristine Pierloot, Quan Manh Phung, and Alex Domingo. "Spin State Energetics in First-Row Transition Metal Complexes: Contribution of (3s3p) Correlation and Its Description by Second-Order Perturbation Theory". In: *Journal of Chemical Theory and Computation* 13.2 (2017). PMID: 28005368, pp. 537–553. DOI: 10.1021/acs.jctc.6b01005. eprint: https://doi.org/10.1021/acs.jctc.6b01005. URL: https://doi.org/10.1021/acs.jctc.6b01005.

[19] Sandeep Sharma et al. "Low-energy spectrum of iron–sulfur clusters directly from many-particle quantum mechanics". In: *Nature Chemistry* 6.10 (Oct. 2014), pp. 927–933. ISSN: 1755-4349. DOI: 10.1038/nchem.2041. URL: https://doi.org/10.1038/nchem.2041.

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

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

[22] Rodney J. Bartlett and Monika Musiał. "Coupled-cluster theory in quantum chemistry". In: *Rev. Mod. Phys.* 79 (1 Feb. 2007), pp. 291–352. DOI: 10.1103/RevModPhys.79.291. URL: https://link.aps.org/doi/10.1103/RevModPhys.79.291.

[23] Nicholas Cox et al. "Electronic structure of the oxygen-evolving complex in photosystem II prior to O-O bond formation". In: *Science* 345.6198 (2014), pp. 804–808. DOI: 10.1126/science.1254910. eprint: https://www.science.org/doi/pdf/10.1126/science.1254910. URL: https://www.science.org/doi/abs/10.1126/science.1254910.

[24] U. Bach et al. "Solid-state dye-sensitized mesoporous TiO2 solar cells with high photon-to-electron conversion efficiencies". In: *Nature* 395.6702 (Oct. 1998), pp. 583–585. ISSN: 1476-4687. DOI: 10.1038/26936. URL: https://doi.org/10.1038/26936.

[25] M.E. Casida and M. Huix-Rotllant. "Progress in Time-Dependent Density-Functional Theory". In: *Annual Review of Physical Chemistry* 63.1 (2012). PMID: 22242728, pp. 287–323. DOI: 10.1146/annurev-physchem-032511-143803. eprint: https://doi.org/10.1146/annurev-physchem-032511-143803. URL: https://doi.org/10.1146/annurev-physchem-032511-143803.

[26] Neepa T. Maitra et al. "Double excitations within time-dependent density functional theory linear response". In: *The Journal of Chemical Physics* 120.13 (2004), pp. 5932–5937. DOI: 10.1063/1.1651060. eprint: https://doi.org/10.1063/1.1651060. URL: https://doi.org/10.1063/1.1651060.

[27] Benjamin G. Levine et al. "Conical intersections and double excitations in time-dependent density functional theory". In: *Molecular Physics* 104.5-7 (2006), pp. 1039–1051. DOI: 10.1080/00268970500417762. eprint: https://doi.org/10.1080/00268970500417762. URL: https://doi.org/10.1080/00268970500417762.

[28] Andreas Dreuw, Jennifer L. Weisman, and Martin Head-Gordon. "Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange". In: *The Journal of Chemical Physics* 119.6 (2003), pp. 2943–2946. DOI: 10.1063/1.1590951. eprint: https://doi.org/10.1063/1.1590951. URL: https://doi.org/10.1063/1.1590951.

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

[30] Peter Elliott et al. "Perspectives on double-excitations in TDDFT". In: *Chemical Physics* 391.1 (2011). Open problems and new solutions in time dependent density functional theory, pp. 110–119. ISSN: 0301-0104. DOI: https://doi.org/10.1016/j.chemphys.2011.03.020. URL: https://www.sciencedirect.com/science/article/pii/S0301010411000966.

[31] David J. Tozer and Nicholas C. Handy. "On the determination of excitation energies using density functional theory". In: *Phys. Chem. Chem. Phys.* 2 (10 2000), pp. 2117–2121. DOI: 10.1039/A910321J. URL: http://dx.doi.org/10.1039/A910321J.

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

[33] Joseph E. Subotnik. "Communication: Configuration interaction singles has a large systematic bias against charge-transfer states". In: *J. Chem. Phys.* 135.7 (2011), p. 071104. ISSN: 00219606. DOI: 10.1063/1.3627152.

[34] Martin Head-Gordon et al. "A doubles correction to electronic excited states from configuration interaction in the space of single substitutions". In: *Chemical Physics Letters* 219.1 (1994), pp. 21–29. ISSN: 0009-2614. DOI: https://doi.org/10.1016/0009-2614(94)00070-0. URL: https://www.sciencedirect.com/science/article/pii/0009261494000700.

[35] John F. Stanton and Rodney J. Bartlett. "The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties". In: *The Journal of Chemical Physics* 98.9 (1993), pp. 7029–7039. DOI: 10.1063/1.464746. eprint: https://doi.org/10.1063/1.464746. URL: https://doi.org/10.1063/1.464746.

[36] Dániel Kánnár and Péter G. Szalay. "Benchmarking Coupled Cluster Methods on Valence Singlet Excited States". In: *Journal of Chemical Theory and Computation* 10.9 (2014). PMID: 26588520, pp. 3757–3765. DOI: 10.1021/ct500495n. eprint: https://doi.org/10.1021/ct500495n. URL: https://doi.org/10.1021/ct500495n.

[37] Pierre-François Loos et al. "A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks". In: *Journal of Chemical Theory and Computation* 14.8 (2018). PMID: 29966098, pp. 4360–4379. DOI: 10.1021/acs.jctc.8b00406. eprint: https://doi.org/10.1021/acs.jctc.8b00406. URL: https://doi.org/10.1021/acs.jctc.8b00406.

[38] Balázs Kozma et al. "A New Benchmark Set for Excitation Energy of Charge Transfer States: Systematic Investigation of Coupled Cluster Type Methods". In: *Journal of Chemical Theory and Computation* 16.7 (2020). PMID: 32502351, pp. 4213–4225. DOI: 10.1021/acs.jctc.0c00154. eprint: https://doi.org/10.1021/acs.jctc.0c00154. URL: https://doi.org/10.1021/acs.jctc.0c00154.

[39] Pierre-François Loos et al. "Reference Energies for Intramolecular Charge-Transfer Excitations". In: *Journal of Chemical Theory and Computation* 17.6 (2021). PMID: 33955742, pp. 3666–3686. DOI: 10.1021/acs.jctc.1c00226. eprint: https://doi.org/10.1021/acs.jctc.1c00226. URL: https://doi.org/10.1021/acs.jctc.1c00226.

[40] So Hirata, Marcel Nooijen, and Rodney J. Bartlett. "High-order determinantal equation-of-motion coupled-cluster calculations for electronic excited states". In: *Chemical Physics Letters* 326.3 (2000), pp. 255–262. ISSN: 0009-2614. DOI: https://doi.org/10.1016/S0009-2614(00)00772-7. URL: http://www.sciencedirect.com/science/article/pii/S0009261400007727.

[41] Pierre-François Loos et al. "Reference Energies for Double Excitations". In: *J. Chem. Theory Comput.* 15.3 (2019), pp. 1939–1956. DOI: 10.1021/acs.jctc.8b01205. eprint: https://doi.org/10.1021/acs.jctc.8b01205. URL: https://doi.org/10.1021/acs.jctc.8b01205.

[42] Rudraditya Sarkar et al. "Assessing the Performances of CASPT2 and NEVPT2 for Vertical Excitation Energies". In: *Journal of Chemical Theory and Computation* 0.0 (2022). PMID: 35333060, null. DOI: 10.1021/acs.jctc.1c01197. eprint: https://doi.org/10.1021/acs.jctc.1c01197. URL: https://doi.org/10.1021/acs.jctc.1c01197.

[43] Alex Domingo et al. "Metal-to-metal charge-transfer transitions: reliable excitation energies from ab initio calculations". In: *Theoretical Chemistry Accounts* 131.9 (Aug. 2012), p. 1264. ISSN: 1432-2234. DOI: 10.1007/s00214-012-1264-1. URL: https://doi.org/10.1007/s00214-012-1264-1.

[44] Benjamin Meyer et al. "Charge transfer processes: the role of optimized molecular orbitals". In: *Dalton Trans.* 43 (29 2014), pp. 11209–11215. DOI: 10.1039/C4DT00471J. URL: http://dx.doi.org/10.1039/C4DT00471J.

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

[46] Jacqueline A. R. Shea and Eric Neuscamman. "Communication: A mean field platform for excited state quantum chemistry". In: *The Journal of Chemical Physics* 149.8 (2018), p. 081101. DOI: 10.1063/1.5045056. eprint: https://doi.org/10.1063/1.5045056. URL: https://doi.org/10.1063/1.5045056.

[47] Tarini S. Hardikar and Eric Neuscamman. "A self-consistent field formulation of excited state mean field theory". In: *The Journal of Chemical Physics* 153.16 (2020), p. 164108. DOI: 10.1063/5.0019557. eprint: https://doi.org/10.1063/5.0019557. URL: https://doi.org/10.1063/5.0019557.

[48] Jacqueline A. R. Shea, Elise Gwin, and Eric Neuscamman. "A Generalized Variational Principle with Applications to Excited State Mean Field Theory". In: *J. Chem. Theory Comput.* 16.3 (2020). PMID: 32017562, pp. 1526–1540. DOI: 10.1021/acs.jctc.9b01105. eprint: https://doi.org/10.1021/acs.jctc.9b01105. URL: https://doi.org/10.1021/acs.jctc.9b01105.

[49] Rachel Clune, Jacqueline A. R. Shea, and Eric Neuscamman. "N5-Scaling Excited-State-Specific Perturbation Theory". In: *Journal of Chemical Theory and Computation* 16.10 (2020). PMID: 32816474, pp. 6132–6141. DOI: 10.1021/acs.jctc.0c00308. eprint: https://doi.org/10.1021/acs.jctc.0c00308. URL: https://doi.org/10.1021/acs.jctc.0c00308.

[50] Rebecca Hanscam and Eric Neuscamman. "Applying generalized variational principles to excited-state-specific complete active space self-consistent field theory". In: *arXiv* 2111.02590 (2021).

[51] D. M. Ceperley and B. J. Alder. "Ground State of the Electron Gas by a Stochastic Method". In: *Phys. Rev. Lett.* 45 (7 Aug. 1980), pp. 566–569. DOI: 10.1103/PhysRevLett.45.566. URL: https://link.aps.org/doi/10.1103/PhysRevLett.45.566.

[52] Brian M. Austin, Dmitry Yu Zubarev, and William A. Lester. "Quantum monte carlo and related approaches". In: *Chem. Rev.* 112.1 (2012), pp. 263–288. ISSN: 00092665. DOI: 10.1021/cr2001564.

[53] W. M. C. Foulkes et al. "Quantum Monte Carlo simulations of solids". In: *Rev. Mod. Phys.* 73.1 (2001), pp. 33–83. URL: http://journals.aps.org/rmp/abstract/10.1103/RevModPhys.73.33.

[54] Csaba Daday et al. "Full Configuration Interaction Excitations of Ethene and Butadiene: Resolution of an Ancient Question". In: *Journal of Chemical Theory and Computation* 8.11 (2012). PMID: 26605604, pp. 4441–4451. DOI: 10.1021/ct300486d. eprint: https://doi.org/10.1021/ct300486d. URL: https://doi.org/10.1021/ct300486d.

[55] N. S. Blunt et al. "An excited-state approach within full configuration interaction quantum Monte Carlo". In: *J. Chem. Phys.* 143.13 (2015), p. 134117. DOI: 10.1063/1.4932595.

[56] N. S. Blunt, George H. Booth, and Ali Alavi. "Density matrices in full configuration interaction quantum Monte Carlo: Excited states, transition dipole moments, and parallel distribution". In: *The Journal of Chemical Physics* 146.24 (2017), p. 244105. DOI: 10.1063/1.4986963. eprint: https://doi.org/10.1063/1.4986963. URL: https://doi.org/10.1063/1.4986963.

[57] A. J. Williamson et al. "Diffusion quantum Monte Carlo calculations of the excited states of silicon". In: *Phys. Rev. B* 57 (19 May 1998), pp. 12140–12144. DOI: 10.1103/PhysRevB.57.12140. URL: https://link.aps.org/doi/10.1103/PhysRevB.57.12140.

[58] Andrew J. Williamson et al. "Quantum Monte Carlo Calculations of Nanostructure Optical Gaps: Application to Silicon Quantum Dots". In: *Phys. Rev. Lett.* 89 (19 Oct. 2002), p. 196803. DOI: 10.1103/PhysRevLett.89.196803. URL: https://link.aps.org/doi/10.1103/PhysRevLett.89.196803.

[59] Alán Aspuru-Guzik et al. "Quantum Monte Carlo for electronic excitations of free-base porphyrin". In: *The Journal of Chemical Physics* 120.7 (2004), pp. 3049–3050. DOI: 10.1063/1.1646356. eprint: https://doi.org/10.1063/1.1646356. URL: https://doi.org/10.1063/1.1646356.

[60] Friedemann Schautz and Claudia Filippi. "Optimized Jastrow–Slater wave functions for ground and excited states: Application to the lowest states of ethene". In: *The Journal of Chemical Physics* 120.23 (2004), pp. 10931–10941. DOI: 10.1063/1.1752881. eprint: https://doi.org/10.1063/1.1752881. URL: https://doi.org/10.1063/1.1752881.

[61] Friedemann Schautz, Francesco Buda, and Claudia Filippi. "Excitations in photoactive molecules from quantum Monte Carlo". In: *The Journal of Chemical Physics* 121.12 (2004), pp. 5836–5844. DOI: 10.1063/1.1777212. eprint: https://doi.org/10.1063/1.1777212. URL: https://doi.org/10.1063/1.1777212.

[62] Anthony Scemama et al. "Excitation energies from diffusion Monte Carlo using selected configuration interaction nodes". In: *The Journal of Chemical Physics* 149.3 (2018), p. 034108. DOI: 10.1063/1.5041327. eprint: https://doi.org/10.1063/1.5041327. URL: https://doi.org/10.1063/1.5041327.

[63] Nick S Blunt and Eric Neuscamman. "Excited-state diffusion Monte Carlo calculations: a simple and efficient two-determinant ansatz". In: *J. Chem. Theory Comput.* 15.1 (2018), pp. 178–189.

[64] Wirawan Purwanto, Shiwei Zhang, and Henry Krakauer. "Excited state calculations using phaseless auxiliary-field quantum Monte Carlo: Potential energy curves of low-lying C2 singlet states". In: *The Journal of Chemical Physics* 130.9 (2009), p. 094107. DOI: 10.1063/1.3077920. eprint: https://doi.org/10.1063/1.3077920. URL: https://doi.org/10.1063/1.3077920.

[65] Fengjie Ma, Shiwei Zhang, and Henry Krakauer. "Excited state calculations in solids by auxiliary-field quantum Monte Carlo". In: *New Journal of Physics* 15.9 (Sept. 2013), p. 093017. DOI: 10.1088/1367-2630/15/9/093017. URL: https://doi.org/10.1088/1367-2630/15/9/093017.

[66] C. J. Umrigar. "Observations on variational and projector Monte Carlo methods". In: *J. Chem. Phys.* 143.16 (2015), p. 164105. ISSN: 00219606. DOI: 10.1063/1.4933112.

[67] J. R. Trail. "Heavy-tailed random error in quantum Monte Carlo". In: *Phys. Rev. E* 77 (1 Jan. 2008), p. 016703. DOI: 10.1103/PhysRevE.77.016703. URL: https://link.aps.org/doi/10.1103/PhysRevE.77.016703.

[68] J. R. Trail. "Alternative sampling for variational quantum Monte Carlo". In: *Phys. Rev. E* 77 (1 Jan. 2008), p. 016704. DOI: 10.1103/PhysRevE.77.016704. URL: https://link.aps.org/doi/10.1103/PhysRevE.77.016704.

[69] Paul J Robinson, Sergio D Pineda Flores, and Eric Neuscamman. "Excitation variance matching with limited configuration interaction expansions in variational Monte Carlo". In: *J. Chem. Phys.* 147.16 (2017), p. 164114.

[70] Sergio D. Pineda Flores and Eric Neuscamman. "Excited State Specific Multi-Slater Jastrow Wave Functions". In: *J. Phys. Chem. A* 123.8 (2019), pp. 1487–1497. DOI: 10.1021/acs.jpca.8b10671. arXiv: 1811.00583. URL: http://arxiv.org/abs/1811.00583.

[71] Leon Otis, Isabel Craig, and Eric Neuscamman. "A hybrid approach to excited-state-specific variational Monte Carlo and doubly excited states". In: *The Journal of Chemical Physics* 153.23 (2020), p. 234105. DOI: 10.1063/5.0024572. eprint: https://doi.org/10.1063/5.0024572. URL: https://doi.org/10.1063/5.0024572.

[72] C. J. Umrigar. "Two aspects of quantum monte carlo: Determination of accurate wavefunctions and determination of potential energy surfaces of molecules". In: *International Journal of Quantum Chemistry* 36.S23 (1989), pp. 217–230. DOI: https://doi.org/10.1002/qua.560360826. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560360826. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560360826.

[73] Claudia Filippi and C. J. Umrigar. "Correlated sampling in quantum Monte Carlo: A route to forces". In: *Phys. Rev. B* 61 (24 June 2000), R16291–R16294. DOI: 10.1103/PhysRevB.61.R16291. URL: https://link.aps.org/doi/10.1103/PhysRevB.61.R16291.

[74] Michele Casula and Sandro Sorella. "Geminal wave functions with Jastrow correlation: A first application to atoms". In: *J. Chem. Phys.* 119.13 (2003), pp. 6500–6511. ISSN: 00219606. DOI: 10.1063/1.1604379. arXiv: 0305169v1 [arXiv:cond-mat].

[75] Claudia Filippi, Roland Assaraf, and Saverio Moroni. "Simple formalism for efficient derivatives and multi-determinant expansions in quantum Monte Carlo". In: *J. Chem. Phys.* 144.19 (2016), p. 194105. ISSN: 00219606. DOI: 10.1063/1.4948778. arXiv: 1602.05490.

[76] Roland Assaraf, S. Moroni, and Claudia Filippi. "Optimizing the Energy with Quantum Monte Carlo: A Lower Numerical Scaling for Jastrow-Slater Expansions". In: *J. Chem. Theory Comput.* 13.11 (2017), pp. 5273–5281. ISSN: 15499626. DOI: 10.1021/acs.jctc.7b00648. arXiv: 1706.07588.

[77] N. D. Drummond, M. D. Towler, and R. J. Needs. "Jastrow correlation factor for atoms, molecules, and solids". In: *Phys. Rev. B* 70 (23 Dec. 2004), p. 235119. DOI: 10.1103/PhysRevB.70.235119. URL: https://link.aps.org/doi/10.1103/PhysRevB.70.235119.

[78] Eric Neuscamman et al. "Nonstochastic algorithms for Jastrow-Slater and correlator product state wave functions". In: *Phys. Rev. B* 84 (20 Nov. 2011), p. 205132. DOI: 10.1103/PhysRevB.84.205132. URL: https://link.aps.org/doi/10.1103/PhysRevB.84.205132.

[79] Eric Neuscamman, C. J. Umrigar, and Garnet Kin Lic Chan. "Optimizing large parameter sets in variational quantum Monte Carlo". In: *Phys. Rev. B* 85.4 (2012), p. 045103. ISSN: 10980121. DOI: 10.1103/PhysRevB.85.045103. arXiv: 1108.0900.

[80] Eric Neuscamman and Garnet Kin-Lic Chan. "Correlator product state study of molecular magnetism in the giant Keplerate Mo 72 Fe 30". In: *Phys. Rev. B* 86.6 (2012), p. 064402.

[81] Eric Neuscamman. "Size Consistency Error in the Antisymmetric Geminal Power Wave Function can be Completely Removed". In: *Phys. Rev. Lett.* 109 (20 Nov. 2012), p. 203001. DOI: 10.1103/PhysRevLett.109.203001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.109.203001.

[82] Yongkyung Kwon, D. M. Ceperley, and Richard M. Martin. "Effects of three-body and backflow correlations in the two-dimensional electron gas". In: *Phys. Rev. B* 48 (16 Oct. 1993), pp. 12037–12046. DOI: 10.1103/PhysRevB.48.12037. URL: https://link.aps.org/doi/10.1103/PhysRevB.48.12037.

[83] Michele Taddei et al. "Iterative backflow renormalization procedure for many-body ground-state wave functions of strongly interacting normal Fermi liquids". In: *Phys. Rev. B* 91.11 (2015), p. 115106.

[84] Di Luo and Bryan K. Clark. "Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions". In: *Phys. Rev. Lett.* 122 (22 June 2019), p. 226401. DOI: 10.1103/PhysRevLett.122.226401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.122.226401.

[85] Liguo Kong, Florian A. Bischoff, and Edward F. Valeev. "Explicitly Correlated R12/F12 Methods for Electronic Structure". In: *Chemical Reviews* 112.1 (2012). PMID: 22176553, pp. 75–107. DOI: 10.1021/cr200204r. eprint: https://doi.org/10.1021/cr200204r. URL: https://doi.org/10.1021/cr200204r.

[86] Felipe Cordova et al. "Troubleshooting time-dependent density-functional theory for photochemical applications: Oxirane". In: *J. Chem. Phys.* 127.16 (2007), p. 164111. DOI: 10.1063/1.2786997.

[87] Claudia Filippi, Maurizio Zaccheddu, and Francesco Buda. "Absorption Spectrum of the Green Fluorescent Protein Chromophore: A Difficult Case for ab Initio Methods?" In: *J. Chem. Theory Comput.* 5.8 (2009), pp. 2074–2087. DOI: 10.1021/ct900227j.

[88] Robert Send, Omar Valsson, and Claudia Filippi. "Electronic Excitations of Simple Cyanine Dyes: Reconciling Density Functional and Wave Function Methods". In: *J. Chem. Theory Comput.* 7.2 (2011). PMID: 26596164, pp. 444–455. DOI: 10.1021/ct1006295. eprint: https://doi.org/10.1021/ct1006295. URL: https://doi.org/10.1021/ct1006295.

[89] Riccardo Guareschi and Claudia Filippi. "Ground- and Excited-State Geometry Optimization of Small Organic Molecules with Quantum Monte Carlo". In: *Journal of Chemical Theory and Computation* 9.12 (2013). PMID: 26592286, pp. 5513–5525. DOI: 10.1021/ct400876y. eprint: https://doi.org/10.1021/ct400876y. URL: https://doi.org/10.1021/ct400876y.

[90] Riccardo Guareschi et al. "Introducing QMC/MMpol: Quantum Monte Carlo in polarizable force fields for excited states". In: *J. Chem. Theory Comput.* 12.4 (2016), pp. 1674–1683.

[91] Alice Cuzzocrea et al. "Variational Principles in Quantum Monte Carlo: The Troubled Story of Variance Minimization". In: *J. Chem. Theory Comput.* 16.7 (2020). PMID: 32419451, pp. 4203–4212. DOI: 10.1021/acs.jctc.0c00147. eprint: https://doi.org/10.1021/acs.jctc.0c00147. URL: https://doi.org/10.1021/acs.jctc.0c00147.

[92] Monika Dash et al. "Tailoring CIPSI Expansions for QMC Calculations of Electronic Excitations: The Case Study of Thiophene". In: *Journal of Chemical Theory and Computation* 17.6 (2021). PMID: 34029098, pp. 3426–3434. DOI: 10.1021/acs.jctc.1c00212. eprint: https://doi.org/10.1021/acs.jctc.1c00212. URL: https://doi.org/10.1021/acs.jctc.1c00212.

[93] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins. "Optimized Trial Wave Functions for QMC calculations". In: *Phys. Rev. Lett.* 60.17 (1988), pp. 1719–1722.

[94] Luning Zhao and Eric Neuscamman. "An Efficient Variational Principle for the Direct Optimization of Excited States". In: *J. Chem. Theory Comput.* 12 (2016), pp. 3436–3440. DOI: 10.1021/acs.jctc.6b00508.

[95] Jacqueline A R Shea and Eric Neuscamman. "Size Consistent Excited States via Algorithmic Transformations between Variational Principles". In: *J. Chem. Theory Comput.* 13.12 (2017), pp. 6078–6088. DOI: 10.1021/acs.jctc.7b00923.

[96] P. López Ríos et al. "Framework for constructing generic Jastrow correlation factors". In: *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 86.3 (2012), p. 036703. ISSN: 15393755. DOI: 10.1103/PhysRevE.86.036703.

[97] Arne Lüchow et al. "Generic expansion of the Jastrow correlation factor in polynomials satisfying symmetry and cusp conditions". In: *J. Chem. Phys.* 142.8 (2015), p. 084111. ISSN: 00219606. DOI: 10.1063/1.4909554.

[98] Chien-Jung Huang, C. J. Umrigar, and M. P. Nightingale. "Accuracy of electronic wave functions in quantum Monte Carlo: The effect of high-order correlations". In: *The Journal of Chemical Physics* 107.8 (1997), pp. 3007–3013. DOI: 10.1063/1.474658. eprint: https://doi.org/10.1063/1.474658. URL: https://doi.org/10.1063/1.474658.

[99] Michele Casula, Claudio Attaccalite, and Sandro Sorella. "Correlated geminal wave function for molecules: An efficient resonating valence bond approach". In: *J. Chem. Phys.* 121.15 (2004), pp. 7110–7126. ISSN: 00219606. DOI: 10.1063/1.1794632. arXiv: 0409644 [cond-mat].

[100] Fabio Sterpone et al. "Dissecting the hydrogen bond: A quantum monte carlo approach". In: *J. Chem. Theory Comput.* 4.9 (2008), pp. 1428–1434. ISSN: 15499618. DOI: 10.1021/ct800121e.

[101] Todd D. Beaudet et al. "Molecular hydrogen adsorbed on benzene: Insights from a quantum Monte Carlo study". In: *J. Chem. Phys.* 129.16 (2008), p. 164711. ISSN: 00219606. DOI: 10.1063/1.2987716.

[102] Mariapia Marchi et al. "Resonating valence bond wave function with molecular orbitals: Application to first-row molecules". In: *J. Chem. Phys.* 131.15 (2009), p. 154116. ISSN: 00219606. DOI: 10.1063/1.3249966.

[103] Andrea Zen et al. "Ab initio molecular dynamics simulation of liquid water by quantum Monte Carlo". In: *J. Chem. Phys.* 142.14 (2015), p. 144111. ISSN: 00219606. DOI: 10.1063/1.4917171.

[104] S. Fahy, X. W. Wang, and Steven G. Louie. "Variational quantum Monte Carlo nonlocal pseudopotential approach to solids: Formulation and application to diamond, graphite, and silicon". In: *Phys. Rev. B* 42 (6 Aug. 1990), pp. 3503–3522. DOI: 10.1103/PhysRevB.42.3503. URL: https://link.aps.org/doi/10.1103/PhysRevB.42.3503.

[105] Bryan K. Clark et al. "Computing the energy of a water molecule using multideterminants: A simple, efficient algorithm". In: *J. Chem. Phys.* 135.24 (2011), p. 244105. ISSN: 00219606. DOI: 10.1063/1.3665391. arXiv: 1106.2456.

[106] Miguel A. Morales et al. "Multideterminant wave functions in quantum Monte Carlo". In: *J. Chem. Theory Comput.* 8.7 (2012), pp. 2181–2188. ISSN: 15499618. DOI: 10.1021/ct3003404. arXiv: arXiv:1303.6676v1.

[107] Beatrice Van Der Goetz and Eric Neuscamman. "Suppressing Ionic Terms with Number-Counting Jastrow Factors in Real Space". In: *J. Chem. Theory Comput.* 13 (2017), pp. 2035–2042. DOI: 10.1021/acs.jctc.7b00158.

[108] Beatrice Van Der Goetz, Leon Otis, and Eric Neuscamman. "Clean and Convenient Tessellations for Number Counting Jastrow Factors". In: *J. Chem. Theory Comput.* 15.2 (2019), pp. 1102–1121. DOI: 10.1021/acs.jctc.8b01139. eprint: https://doi.org/10.1021/acs.jctc.8b01139. URL: https://doi.org/10.1021/acs.jctc.8b01139.

[109] Claudio Attaccalite and Sandro Sorella. "Stable Liquid Hydrogen at High Pressure by a Novel Ab Initio Molecular-Dynamics Calculation". In: *Phys. Rev. Lett.* 100 (11 Mar. 2008), p. 114501. DOI: 10.1103/PhysRevLett.100.114501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.100.114501.

[110] Shivesh Pathak and Lucas K. Wagner. "A light weight regularization for wave function parameter gradients in quantum Monte Carlo". In: *AIP Advances* 10.8 (2020), p. 085213. DOI: 10.1063/5.0004008. eprint: https://doi.org/10.1063/5.0004008. URL: https://doi.org/10.1063/5.0004008.

[111] Jesse van Rhijn et al. "Energy Derivatives in Real-Space Diffusion Monte Carlo". In: *Journal of Chemical Theory and Computation* 18.1 (2022). PMID: 34930005, pp. 118–123. DOI: 10.1021/acs.jctc.1c00496. eprint: https://doi.org/10.1021/acs.jctc.1c00496. URL: https://doi.org/10.1021/acs.jctc.1c00496.

[112] M. P. Nightingale and Vilen Melik-Alaverdian. "Optimization of ground- and excited-state wave functions and van der waals clusters". In: *Phys. Rev. Lett.* 87.4 (2001), p. 43401. ISSN: 10797114. DOI: 10.1103/PhysRevLett.87.043401.

[113] C. J. Umrigar et al. "Alleviation of the fermion-sign problem by optimization of many-body wave functions". In: *Phys. Rev. Lett.* 98.11 (2007), p. 110201. ISSN: 00319007. DOI: 10.1103/PhysRevLett.98.110201. arXiv: 0611094 [cond-mat].

[114] Julien Toulouse and C. J. Umrigar. "Optimization of quantum Monte Carlo wave functions by energy minimization". In: *J. Chem. Phys.* 126.8 (2007), p. 084102. ISSN: 00219606. DOI: 10.1063/1.2437215. arXiv: 0701039 [physics].

[115] Julien Toulouse and C. J. Umrigar. "Full optimization of Jastrow-Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules". In: *J. Chem. Phys.* 128.17 (2008), p. 174101. ISSN: 00219606. DOI: 10.1063/1.2908237. arXiv: arXiv:0803.2905v1.

[116] Lauretta R. Schwarz, A. Alavi, and George H. Booth. "Projector Quantum Monte Carlo Method for Nonlinear Wave Functions". In: *Phys. Rev. Lett.* 118.17 (2017), p. 176403. ISSN: 10797114. DOI: 10.1103/PhysRevLett.118.176403. arXiv: 1610.09326.

[117] Iliya Sabzevari and Sandeep Sharma. "Improved Speed and Scaling in Orbital Space Variational Monte". In: *J. Chem. Theory Comput.* 14 (2018), pp. 6276–6286. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.8b00780.

[118] Ankit Mahajan and Sandeep Sharma. "Symmetry-Projected Jastrow Mean-Field Wave Function in Variational Monte Carlo". In: *J. Phys. Chem. A* 123.17 (2019), pp. 3911–3921. DOI: 10.1021/acs.jpca.9b01583. eprint: https://doi.org/10.1021/acs.jpca.9b01583. URL: https://doi.org/10.1021/acs.jpca.9b01583.

[119] Leon Otis and Eric Neuscamman. "Complementary first and second derivative methods for ansatz optimization in variational Monte Carlo". In: *Phys. Chem. Chem. Phys.* 21 (27 2019), pp. 14491–14510. DOI: 10.1039/C9CP02269D. URL: http://dx.doi.org/10.1039/C9CP02269D.

[120] Mosé Casalegno, Massimo Mella, and Andrew M. Rappe. "Computing accurate forces in quantum Monte Carlo using Pulay's corrections and energy minimization". In: *The Journal of Chemical Physics* 118.16 (2003), pp. 7193–7201. DOI: 10.1063/1.1562605. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.1562605. URL: https://aip.scitation.org/doi/abs/10.1063/1.1562605.

[121] C. J. Umrigar and Claudia Filippi. "Energy and variance optimization of many-body wave functions". In: *Phys. Rev. Lett.* 94.15 (2005), p. 150201. ISSN: 00319007. DOI: 10.1103/PhysRevLett.94.150201. arXiv: 0412634 [cond-mat].

[122] Sandro Sorella. "Wave function optimization in the variational Monte Carlo method". In: *Phys. Rev. B* 71.24 (2005), p. 241103. ISSN: 10980121. DOI: 10.1103/PhysRevB.71.241103. arXiv: 0502553 [cond-mat].

[123] Sandro Sorella, Michele Casula, and Dario Rocca. "Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods". In: *J. Chem. Phys.* 127.1 (2007), p. 014105. ISSN: 00219606. DOI: 10.1063/1.2746035. arXiv: 0702349 [cond-mat].

[124] Sebastian Ruder. "An overview of gradient descent optimization algorithms". In: *arXiv.org* (2016), p. 1609.04747. ISSN: 0006341X. DOI: 10.1111/j.0006-341X.1999.00591.x.

[125] Luning Zhao and Eric Neuscamman. "Equation of Motion Theory for Excited States in Variational Monte Carlo and the Jastrow Antisymmetric Geminal Power in Hilbert Space". In: *J. Chem. Theory Comput.* 12.8 (2016), pp. 3719–3726. DOI: 10.1021/acs.jctc.6b00480.

[126] Peter Pulay. "Improved SCF convergence acceleration". In: *J. Comput. Chem.* 3.4 (1982), pp. 556–560.

[127] E R Davidson. "The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices". In: *J. Comput. Phys* 17 (1975), pp. 87–94.

[128] Steven R White. "Density matrix renormalization group algorithms with a single center site". In: *Phys. Rev. B* 72.18 (2005), p. 180403.

[129] Luning Zhao and Eric Neuscamman. "A Blocked Linear Method for Optimizing Large Parameter Sets in Variational Monte Carlo". In: *J. Chem. Theory Comput.* 13 (2017), pp. 2604–2611. DOI: 10.1021/acs.jctc.7b00119.

[130] Jeffrey B. Schriber and Francesco A. Evangelista. "Adaptive Configuration Interaction for Computing Challenging Electronic Excited States with Tunable Accuracy". In: *Journal of Chemical Theory and Computation* 13.11 (2017). PMID: 28892621, pp. 5354–5366. DOI: 10.1021/acs.jctc.7b00725. eprint: https://doi.org/10.1021/acs.jctc.7b00725. URL: https://doi.org/10.1021/acs.jctc.7b00725.

[131] Sandeep Sharma et al. "Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory". In: *J. Chem. Theory Comput.* 13.4 (2017), pp. 1595–1604. ISSN: 15499626. DOI: 10.1021/acs.jctc.6b01028. arXiv: 1610.06660v2.

[132] Alan D Chien et al. "Excited States of Methylene, Polyenes, and Ozone from Heat-Bath Configuration Interaction". In: *J. Phys. Chem. A* 122.10 (2018), pp. 2714–2722.

[133] Hitesh J Changlani et al. "Approximating strongly correlated wave functions with correlator product states". In: *Phys. Rev. B* 80.24 (2009), p. 245116.

[134] Fabio Mezzacapo et al. "Ground-state properties of quantum many-body systems: entangled-plaquette states and variational Monte Carlo". In: *New J. Phys.* 11.8 (2009), p. 083026.

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

[136] Eric Neuscamman. "Communication: Variation After Response in Quantum Monte Carlo". In: *J. Chem. Phys.* 145 (2016), p. 081103. DOI: 10.1063/1.4961686.

[137] Nick S Blunt and Eric Neuscamman. "Charge-transfer excited states: Seeking a balanced and efficient wave function ansatz in variational Monte Carlo". In: *J. Chem. Phys.* 147.19 (2017), p. 194101.

[138] Giuseppe Carleo and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks". In: *Science* 355.6325 (2017), pp. 602–606.

[139] Dmitrii Kochkov and Bryan K. Clark. "Variational optimization in the AI era: Computational Graph States and Supervised Wave-function Optimization". In: *arXiv.org* (2018), p. 1811.12423.

[140] Eric Neuscamman. "Improved optimization for the cluster Jastrow antisymmetric geminal power and tests on triple-bond dissociations". In: *J. Chem. Theory Comput.* 12.7 (2016), pp. 3149–3159. ISSN: 15499626. DOI: 10.1021/acs.jctc.6b00288. arXiv: 1603.06605.

[141] Shuai Zhang, Fionn D Malone, and Miguel A Morales. "Auxiliary-field quantum Monte Carlo calculations of the structural properties of nickel oxide". In: *J. Chem. Phys.* 149.16 (2018), p. 164102.

[142] Luning Zhao and Eric Neuscamman. "Variational Excitations in Real Solids: Optical Gaps and Insights into Many-Body Perturbation Theory". In: *Phys. Rev. Lett.* 123 (3 July 2019), p. 036402. DOI: 10.1103/PhysRevLett.123.036402. URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.036402.

[143] Arne Luchow and James B Anderson. "Monte Carlo Methods in Electronic Structures for Large Systems". In: *Annual Review of Physical Chemistry* 51 (2000), pp. 501–526.

[144] P R C Kent et al. "Finite-size errors in quantum many-body simulations of extended systems". In: *Phys. Rev. B* 59.3 (1999), pp. 1917–1929.

[145] Roland Assaraf and Michel Caffarel. "Zero-variance principle for monte carlo algorithms". In: *Physical Review Letters* 83.23 (1999), pp. 4682–4685. ISSN: 10797114. DOI: 10.1103/PhysRevLett.83.4682.

[146] A. Harju et al. "Stochastic gradient approximation: An efficient method to optimize many-body wave functions". In: *Phys. Rev. Lett.* 79.7 (1997), pp. 1173–1177. ISSN: 10797114. DOI: 10.1103/PhysRevLett.79.1173.

[147] Xi Lin, Hongkai Zhang, and Andrew M. Rappe. "Optimization of quantum Monte Carlo wave functions using analytical energy derivatives". In: *J. Chem. Phys.* 112.6 (2000), pp. 2650–2654. ISSN: 00219606. DOI: 10.1063/1.480839.

[148] Myung Won Lee, Massimo Mella, and Andrew M. Rappe. "Electronic quantum Monte Carlo calculations of atomic forces, vibrations, and anharmonicities". In: *The Journal of Chemical Physics* 122.24 (2005), p. 244103. DOI: 10.1063/1.1924690. eprint: https://doi.org/10.1063/1.1924690. URL: https://doi.org/10.1063/1.1924690.

[149] Paul R C Kent, R. J. Needs, and G Rajagopal. "Monte Carlo energy and variance-minimization techniques for optimizing many-body wave functions". In: *Phys. Rev. B* 59.19 (1999), p. 12344.

[150] Jeongnim Kim et al. "QMCPACK : An open source ab initio Quantum Monte Carlo package for the electronic structure of atoms, molecules, and solids". In: *J. Phys. Condens. Matter* 30.19 (2018), p. 195901. ISSN: 0953-8984. DOI: 10.1088/1361-648x/aab9c3.

[151] M. D. Brown et al. "Energies of the first row atoms from quantum Monte Carlo". In: *J. Chem. Phys.* 126.22 (2007), p. 224110. ISSN: 00219606. DOI: 10.1063/1.2743972. arXiv: 1011.4343.

[152] F. R. Petruzielo, Julien Toulouse, and C. J. Umrigar. "Approaching chemical accuracy with quantum Monte Carlo". In: *J. Chem. Phys.* 136.12 (2012), p. 124116. ISSN: 00219606. DOI: 10.1063/1.3697846. arXiv: 1202.0317.

[153] Nick S Blunt, Ali Alavi, and George H Booth. "Nonlinear biases , stochastically sampled effective Hamiltonians , and spectral functions in quantum Monte Carlo methods". In: *Phys. Rev. B* 98.8 (2018), p. 085118. DOI: 10.1103/PhysRevB.98.085118.

[154] Lauretta Rebecca Schwarz. "Projector Quantum Monte Carlo Methods for Linear and Non-linear Wavefunction Ansatzes". PhD thesis. University of Cambridge, 2017. DOI: https://doi.org/10.17863/CAM.13792. URL: https://aspace.repository.cam.ac.uk/handle/1810/267871.

[155] Hong Zhou Ye et al. "σ-SCF: A direct energy-targeting method to mean-field excited states". In: *J. Chem. Phys.* 147.21 (2017), p. 214104. ISSN: 00219606. DOI: 10.1063/1.5001262.

[156] Leon Bottou. "Stochastic Gradient Descent Tricks". In: *Neural Networks, Tricks of the Trade*. Neural Networks: Tricks of the Trade, 2nd. Lecture Notes in Computer Science (LNCS). Springer, Jan. 2012, pp. 430–445. URL: https://www.microsoft.com/en-us/research/publication/stochastic-gradient-tricks/.

[157] Diederik Kingma and Jimmy Ba. "ADAM: A Method for Stochastic Optimization". In: *arXiv.org* (2014), p. 1412.6980. ISSN: 13087711. DOI: 10.1063/1.4902458.

[158] Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. "On the Convergence of ADAM and Beyond". In: *International Conference on Learning Representations* (2018), pp. 1–23.

[159] Matteo Barborini, Sandro Sorella, and Leonardo Guidoni. "Structural optimization by quantum Monte Carlo: Investigating the low-lying excited states of ethylene". In: *J. Chem. Theory Comput.* 8.4 (2012), pp. 1260–1269. ISSN: 15499626. DOI: 10.1021/ct200724q.

[160] Martin C. Gutzwiller. "Correlation of Electrons in a Narrow s Band". In: *Phys. Rev.* 137 (6A Mar. 1965), A1726–A1735. DOI: 10.1103/PhysRev.137.A1726. URL: https://link.aps.org/doi/10.1103/PhysRev.137.A1726.

[161] Eric Neuscamman. "Communication: A Jastrow factor coupled cluster theory for weak and strong electron correlation". In: *J. Chem. Phys.* 139.18 (2013), p. 181101. ISSN: 00219606. DOI: 10.1063/1.4829536.

[162] Stephen R Langhoff and Ernest R Davidson. "Configuration Interaction Calculations". In: *Int. J. Quantum Chem.* VIII (1974), pp. 61–72.

[163] Elda Rossi et al. "A full-configuration benchmark for the N2 Molecule". In: *Chem. Phys. Lett.* 310 (1999), p. 530. ISSN: 00092614. DOI: 10.1016/S0009-2614(99)00791-5.

[164] Garnet Kin Lic Chan, Mihály Kállay, and Jürgen Gauss. "State-of-the-art density matrix renormalization group and coupled cluster theory studies of the nitrogen binding curve". In: *J. Chem. Phys.* 121.13 (2004), pp. 6110–6116. ISSN: 00219606. DOI: 10.1063/1.1783212.

[165] "Quantum Monte Carlo with Jastrow-valence-bond wave functions". In: *J. Chem. Phys.* 134.8 (2011), p. 084108. ISSN: 00219606. DOI: 10.1063/1.3555821.

[166] David A. Mazziotti. "Realization of Quantum Chemistry without Wave Functions through First-Order Semidefinite Programming". In: *Phys. Rev. Lett.* 93.21 (2004), p. 213001. ISSN: 0031-9007. DOI: 10.1103/physrevlett.93.213001.

[167] K. K. Baldridge et al. "General Atomic and Molecular Electronic Structure System". In: *J. Comput. Chem* 14.11 (1993), pp. 1347–1363.

[168] M Burkatzki, C Filippi, and M Dolg. "Energy-consistent pseudopotentials for quantum Monte Carlo calculations." In: *J. Chem. Phys.* 126.23 (2007), p. 234105. ISSN: 0021-9606. DOI: 10.1063/1.2741534. URL: http://www.ncbi.nlm.nih.gov/pubmed/17600402.

[169] Monika Dash et al. "Excited states with selected CI-QMC: chemically accurate excitation energies and geometries". In: *arXiv.org* (), pp. 1–12. arXiv: arXiv:1905.06737v1.

[170] H.-J. Werner et al. *MOLPRO, version 2019.1, a package of ab initio programs.* see http://www.molpro.net. Cardiff, UK, 2019.

[171] M. Chandler Bennett et al. "A new generation of effective core potentials for correlated calculations". In: *J. Chem. Phys.* 147.22 (2017), p. 224106. ISSN: 00219606. DOI: 10.1063/1.4995643.

[172] Maria Melanie Russew and Stefan Hecht. "Photoswitches: From molecules to materials". In: *Adv. Mater.* 22.31 (2010), pp. 3348–3360. ISSN: 09359648. DOI: 10.1002/adma.200904102.

[173] Kazushi Kinbara and Takuzo Aida. "Toward Intelligent Molecular Machines: Directed Motions of Biological and Artificial Molecules and Assemblies". In: *Chem. Rev.* 105.4 (2005), pp. 1377–1400. ISSN: 0009-2665. DOI: 10.1021/cr030071r.

[174] He Tian and Songjie Yang. "Recent progresses on diarylethene based photochromic switches". In: *Chem. Soc. Rev.* 33.2 (2004), pp. 85–97.

[175] Karl O. Christe et al. "The N2F+ Cation. an Unusual Ion Containing the Shortest Presently Known Nitrogen-Fluorine Bond". In: *Inorg. Chem.* 113.10 (1991), pp. 3795–3800. ISSN: 15205126. DOI: 10.1021/ja00010a023.

[176] Karl O. Christe et al. "Dinitrogen difluoride chemistry. Improved syntheses of cis- and trans-N2F2, Synthesis and characterization of N 2F+Sn2F9-, ordered crystal structure of N2F+Sb2F11-, High-level electronic structure calculations of cis-N 2F2". In: *Inorg. Chem.* 49.15 (2010), pp. 6823–6833. ISSN: 00201669. DOI: 10.1021/ic100471s.

[177] Timothy J. Lee et al. "Theoretical investigations of molecules composed only of fluorine, oxygen and nitrogen: determination of the equilibrium structures of FOOF, (NO)2 and FNNF and the transition state structure for FNNF cis-trans isomerization". In: *Theor. Chim. Acta* 75.2 (1989), pp. 81–98. DOI: 10.1007/BF00527711.

[178] Andrew M. Sand, Christine A. Schwerdtfeger, and David A. Mazziotti. "Strongly correlated barriers to rotation from parametric two-electron reduced-density-matrix methods in application to the isomerization of diazene". In: *J. Chem. Phys.* 136.3 (2012), p. 034112. ISSN: 00219606. DOI: 10.1063/1.3675683.

[179] George H. Booth, Alex J. W. Thom, and Ali Alavi. "Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space". In: *The Journal of Chemical Physics* 131.5 (2009), p. 054106. DOI: 10.1063/1.3193710. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.3193710. URL: https://aip.scitation.org/doi/abs/10.1063/1.3193710.

[180] Steven R. White and Richard L. Martin. "Ab initio quantum chemistry using the density matrix renormalization group". In: *The Journal of Chemical Physics* 110.9 (1999), pp. 4127–4130. DOI: 10.1063/1.478295. eprint: https://doi.org/10.1063/1.478295. URL: https://doi.org/10.1063/1.478295.

[181] Garnet Kin-Lic Chan and Martin Head-Gordon. "Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group". In: *The Journal of Chemical Physics* 116.11 (2002), pp. 4462–4476. DOI: 10.1063/1.1449459. eprint: https://doi.org/10.1063/1.1449459. URL: https://doi.org/10.1063/1.1449459.

[182] Yuki Kurashige and Takeshi Yanai. "High-performance ab initio density matrix renormalization group method: Applicability to large-scale multireference problems for metal compounds". In: *The Journal of Chemical Physics* 130.23 (2009), p. 234114. DOI: 10.1063/1.3152576. eprint: https://doi.org/10.1063/1.3152576. URL: https://doi.org/10.1063/1.3152576.

[183] Micah L. Abrams and C. David Sherrill. "Important configurations in configuration interaction and coupled-cluster wave functions". In: *Chemical Physics Letters* 412.1 (2005), pp. 121–124. ISSN: 0009-2614. DOI: https://doi.org/10.1016/j.cplett.2005.06.107. URL: http://www.sciencedirect.com/science/article/pii/S0009261405009826.

[184] Mark A. Watson and Garnet Kin-Lic Chan. "Excited States of Butadiene to Chemical Accuracy: Reconciling Theory and Experiment". In: *J. Chem. Theory Comput.* 8.11 (2012). PMID: 26605568, pp. 4013–4018. DOI: 10.1021/ct300591z. eprint: https://doi.org/10.1021/ct300591z. URL: https://doi.org/10.1021/ct300591z.

[185] Luis Serrano-Andrés et al. "Towards an accurate molecular orbital theory for excited states: Ethene, butadiene, and hexatriene". In: *The Journal of Chemical Physics* 98.4 (1993), pp. 3151–3162. DOI: 10.1063/1.465071. eprint: https://doi.org/10.1063/1.465071. URL: https://doi.org/10.1063/1.465071.

[186] Luis Serrano-Andrés et al. "A theoretical study of the electronic spectrum of thiophene". In: *Chemical Physics Letters* 211.1 (1993), pp. 125–134. ISSN: 0009-2614. DOI: https://doi.org/10.1016/0009-2614(93)80061-S. URL: http://www.sciencedirect.com/science/article/pii/000926149380061S.

[187] Kenichi Nakayama, Haruyuki Nakano, and Kimihiko Hirao. "Theoretical study of the $\pi \rightarrow \pi^*$ excited states of linear polyenes: The energy gap between 1 1Bu+ and 2 1Ag states and their character". In: *International Journal of Quantum Chemistry* 66.2 (1998), pp. 157–175. DOI: 10.1002/(SICI)1097-461X(1998)66:2<157::AID-QUA7>3.0.CO;2-U. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-461X%281998%2966%3A2%3C157%3A%3AAID-QUA7%3E3.0.CO%3B2-U. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-461X%281998%2966%3A2%3C157%3A%3AAID-QUA7%3E3.0.CO%3B2-U.

[188] Bojana Ostojić and Wolfgang Domcke. "Ab initio investigation of the potential energy surfaces involved in the photophysics of s-trans-1,3-butadiene". In: *Chemical Physics* 269.1 (2001), pp. 1–10. ISSN: 0301-0104. DOI: https://doi.org/10.1016/S0301-0104(01)00373-1. URL: http://www.sciencedirect.com/science/article/pii/S0301010401003731.

[189] Michal Dallos and Hans Lischka. "A systematic theoretical investigation of the lowest valence- and Rydberg-excited singlet states of trans-butadiene. The character of the 11Bu (V) state revisited". In: *Theoretical Chemistry Accounts* 112.1 (2004), pp. 16–26. ISSN: 1432-2234. DOI: 10.1007/s00214-003-0557-9. URL: https://doi.org/10.1007/s00214-003-0557-9.

[190] Marko Schreiber et al. "Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3". In: *J. Chem. Phys.* 128.13 (2008), p. 134110. DOI: 10.1063/1.2889385. eprint: https://doi.org/10.1063/1.2889385. URL: https://doi.org/10.1063/1.2889385.

[191] Mario R. Silva-Junior et al. "Benchmarks of electronically excited states: Basis set effects on CASPT2 results". In: *J. Chem. Phys.* 133.17 (2010), p. 174318. DOI: 10.1063/1.3499598. eprint: https://doi.org/10.1063/1.3499598. URL: https://doi.org/10.1063/1.3499598.

[192] Selin Duman et al. "Heavy Atom Free Singlet Oxygen Generation: Doubly Substituted Configurations Dominate S1 States of Bis-BODIPYs". In: *The Journal of Organic Chemistry* 77.10 (2012). PMID: 22530939, pp. 4516–4527. DOI: 10.1021/jo300051v. eprint: https://doi.org/10.1021/jo300051v. URL: https://doi.org/10.1021/jo300051v.

[193] Jin Wen et al. "An MS-CASPT2 Calculation of the Excited Electronic States of an Axial Difluoroborondipyrromethene (BODIPY) Dimer". In: *Journal of Chemical Theory and Computation* 14.8 (2018). PMID: 29874458, pp. 4291–4297. DOI: 10.1021/acs.jctc.8b00136. eprint: https://doi.org/10.1021/acs.jctc.8b00136. URL: https://doi.org/10.1021/acs.jctc.8b00136.

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

[195] Nadia Ben Amor, Adrien Soupart, and Marie Catherine Heitz. "Methodological CASPT2 study of the valence excited states of an iron-porphyrin complex". In: *Journal of Molecular Modeling* 23.2 (2017). ISSN: 09485023. DOI: 10.1007/s00894-017-3226-y.

[196] Pierre-François Loos et al. "A Mountaineering Strategy to Excited States: Highly Accurate Energies and Benchmarks for Medium Sized Molecules". In: *Journal of Chemical Theory and Computation* 16.3 (2020). PMID: 31986042, pp. 1711–1741. DOI: 10.1021/acs.jctc.9b01216. eprint: https://doi.org/10.1021/acs.jctc.9b01216. URL: https://doi.org/10.1021/acs.jctc.9b01216.

[197] Jonas Feldt and Claudia Filippi. "Excited-state calculations with quantum Monte Carlo". In: *arXiv preprint arXiv:2002.03622* (2020). arXiv: 2002.03622 [physics.chem-ph].

[198] Scott M. Garner and Eric Neuscamman. "A variational Monte Carlo approach for core excitations". In: *The Journal of Chemical Physics* 153.14 (2020), p. 144108. DOI: 10.1063/5.0020310. eprint: https://doi.org/10.1063/5.0020310. URL: https://doi.org/10.1063/5.0020310.

[199] C. J. Umrigar, M. P. Nightingale, and K. J. Runge. "A diffusion Monte Carlo algorithm with very small time-step errors". In: *J. Chem. Phys.* 99.4 (1993), pp. 2865–2890. DOI: 10.1063/1.465195. eprint: https://doi.org/10.1063/1.465195. URL: https://doi.org/10.1063/1.465195.

[200] David Pfau et al. "Ab initio solution of the many-electron Schrödinger equation with deep neural networks". In: *Phys. Rev. Research* 2 (3 Sept. 2020), p. 033429. DOI: 10.1103/PhysRevResearch.2.033429. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.2.033429.

[201] Paul M. Zimmerman et al. "Excited states of methylene from quantum Monte Carlo". In: *The Journal of Chemical Physics* 131.12 (2009), p. 124103. DOI: 10.1063/1.3220671. eprint: https://doi.org/10.1063/1.3220671. URL: https://doi.org/10.1063/1.3220671.

[202] Claudia Filippi and C. J. Umrigar. "Multiconfiguration wave functions for quantum Monte Carlo calculations of first-row diatomic molecules". In: *J. Chem. Phys.* 105.1 (1996), pp. 213–226. ISSN: 00219606. DOI: 10.1063/1.471865.

[203] D. C. Sorensen. "Newton's Method with a Model Trust Region Modification". In: *SIAM Journal on Numerical Analysis* 19.2 (1982), pp. 409–426. DOI: 10.1137/0719026. eprint: https://doi.org/10.1137/0719026. URL: https://doi.org/10.1137/0719026.

[204] P. R. C. Kent et al. "QMCPACK: Advances in the development, efficiency, and application of auxiliary field and real-space variational and diffusion quantum Monte Carlo". In: *J. Chem. Phys.* 152.17 (2020), p. 174105. DOI: 10.1063/5.0004860. eprint: https://doi.org/10.1063/5.0004860. URL: https://doi.org/10.1063/5.0004860.

[205] Qiming Sun et al. "PySCF: the Python-based simulations of chemistry framework". In: *WIREs Computational Molecular Science* 8.1 (2018), e1340. DOI: 10.1002/wcms.1340. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1340. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1340.

[206] Micah L. Abrams and C. David Sherrill. "Full configuration interaction potential energy curves for the X1Σg+, B1Δg, and B'1Σg+ states of C2: A challenge for approximate methods". In: *J. Chem. Phys.* 121.19 (2004), pp. 9211–9219. DOI: 10.1063/1.1804498.

[207] George H. Booth et al. "Breaking the carbon dimer: The challenges of multiple bond dissociation with full configuration interaction quantum Monte Carlo methods". In: *J. Chem. Phys.* 135.8 (2011), p. 084104. DOI: 10.1063/1.3624383. eprint: https://doi.org/10.1063/1.3624383. URL: https://doi.org/10.1063/1.3624383.

[208] Adam A. Holmes, C. J. Umrigar, and Sandeep Sharma. "Excited states using semistochastic heat-bath configuration interaction". In: *J. Chem. Phys.* 147.16 (2017), p. 164111. DOI: 10.1063/1.4998614. eprint: https://doi.org/10.1063/1.4998614. URL: https://doi.org/10.1063/1.4998614.

[209] Diptarka Hait and Martin Head-Gordon. "Excited State Orbital Optimization via Minimizing the Square of the Gradient: General Approach and Application to Singly and Doubly Excited States via Density Functional Theory". In: *Journal of Chemical Theory and Computation* 16.3 (2020). PMID: 32017554, pp. 1699–1710. DOI: 10.1021/acs.jctc.9b01127. eprint: https://doi.org/10.1021/acs.jctc.9b01127. URL: https://doi.org/10.1021/acs.jctc.9b01127.

[210] John D. Watts, Steven R. Gwaltney, and Rodney J. Bartlett. "Coupled-cluster calculations of the excitation energies of ethylene, butadiene, and cyclopentadiene". In: *J. Chem. Phys.* 105.16 (1996), pp. 6979–6988. DOI: 10.1063/1.471988. eprint: https://doi.org/10.1063/1.471988. URL: https://doi.org/10.1063/1.471988.

[211] Mario R. Silva-Junior et al. "Benchmarks for electronically excited states: Time-dependent density functional theory and density functional theory based multireference configuration interaction". In: *J. Chem. Phys.* 129.10 (2008), p. 104103. DOI: 10.1063/1.2973541. eprint: https://doi.org/10.1063/1.2973541. URL: https://doi.org/10.1063/1.2973541.

[212] Jun Shen and Shuhua Li. "Block correlated coupled cluster method with the complete active-space self-consistent-field reference function: Applications for low-lying electronic excited states". In: *J. Chem. Phys.* 131.17 (2009), p. 174101. DOI: 10.1063/1.3256297. eprint: https://doi.org/10.1063/1.3256297. URL: https://doi.org/10.1063/1.3256297.

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

[214] Robert P. Frueholz et al. "Electronic spectroscopy of 1,3-cyclopentadiene, 1,3-cyclohexadiene and 1,3-cycloheptadiene by electron impact". In: *J. Chem. Phys.* 70.4 (1979), pp. 2003–2013. DOI: 10.1063/1.437626. eprint: https://doi.org/10.1063/1.437626. URL: https://doi.org/10.1063/1.437626.

[215] R. McDiarmid, A. Sabljić, and J. P. Doering. "Valence transitions in 1,3-cyclopentadiene, 1,3-cyclohexadiene, and 1,3-cycloheptadiene". In: *J. Chem. Phys.* 83.5 (1985), pp. 2147–2152. DOI: 10.1063/1.449304. eprint: https://doi.org/10.1063/1.449304. URL: https://doi.org/10.1063/1.449304.

[216] Isabel Gómez et al. "Intramolecular Charge Transfer in 4-Aminobenzonitriles Does Not Necessarily Need the Twist". In: *J. Am. Chem. Soc.* 127.19 (2005). PMID: 15884954, pp. 7119–7129. DOI: 10.1021/ja042413w. eprint: https://doi.org/10.1021/ja042413w. URL: https://doi.org/10.1021/ja042413w.

[217] Mireia Segado, Isabel Gómez, and Mar Reguero. "Intramolecular charge transfer in aminobenzonitriles and tetrafluoro counterparts: fluorescence explained by competition between low-lying excited states and radiationless deactivation. Part I: A mechanistic overview of the parent system ABN". In: *Phys. Chem. Chem. Phys.* 18 (9 2016), pp. 6861–6874. DOI: 10.1039/C5CP04690D. URL: http://dx.doi.org/10.1039/C5CP04690D.

[218] Mireia Segado et al. "Intramolecular charge transfer in aminobenzonitriles and tetrafluoro counterparts: fluorescence explained by competition between low lying excited states and radiationless deactivation. Part II: influence of substitution on luminescence patterns". In: *Phys. Chem. Chem. Phys.* 18 (9 2016), pp. 6875–6884. DOI: 10.1039/C5CP04693A. URL: http://dx.doi.org/10.1039/C5CP04693A.

[219] Lan Nguyen Tran and Eric Neuscamman. "Improving excited-state potential energy surfaces via optimal orbital shapes". In: *J. Phys. Chem. A* 124 (2020), pp. 8273–8279. DOI: 10.1021/acs.jpca.0c07593.

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

[221] Diptarka Hait and Martin Head-Gordon. "Orbital Optimized Density Functional Theory for Electronic Excited States". In: *The Journal of Physical Chemistry Letters* 12.19 (2021). PMID: 33961437, pp. 4517–4529. DOI: 10.1021/acs.jpclett.1c00744. eprint: https://doi.org/10.1021/acs.jpclett.1c00744. URL: https://doi.org/10.1021/acs.jpclett.1c00744.

[222] John M. Herbert. "Density Functional Theory for Electronic Excited States". In: (2022). DOI: 10.48550/ARXIV.2204.10135. URL: https://arxiv.org/abs/2204.10135.

[223] Fabien Bruneval, Samia M. Hamed, and Jeffrey B. Neaton. "A systematic benchmark of the ab initio Bethe-Salpeter equation approach for low-lying optical excitations of small organic molecules". In: *The Journal of Chemical Physics* 142.24 (2015), p. 244101. DOI: 10.1063/1.4922489. eprint: https://doi.org/10.1063/1.4922489. URL: https://doi.org/10.1063/1.4922489.

[224] Denis Jacquemin, Ivan Duchemin, and Xavier Blase. "Benchmarking the Bethe–Salpeter Formalism on a Standard Organic Molecular Set". In: *Journal of Chemical Theory and Computation* 11.7 (2015). PMID: 26207104, pp. 3290–3304. DOI: 10.1021/acs.jctc.5b00304. eprint: https://doi.org/10.1021/acs.jctc.5b00304. URL: https://doi.org/10.1021/acs.jctc.5b00304.

[225] Tonatiuh Rangel et al. "An assessment of low-lying excitation energies and triplet instabilities of organic molecules with an ab initio Bethe-Salpeter equation approach and the Tamm-Dancoff approximation". In: *The Journal of Chemical Physics* 146.19 (2017), p. 194108. DOI: 10.1063/1.4983126. eprint: https://doi.org/10.1063/1.4983126. URL: https://doi.org/10.1063/1.4983126.

[226] Keld L. Bak et al. "Accuracy of atomization energies and reaction enthalpies in standard and extrapolated electronic wave function/basis set calculations". In: *The Journal of Chemical Physics* 112.21 (2000), pp. 9229–9242. DOI: 10.1063/1.481544. eprint: https://doi.org/10.1063/1.481544. URL: https://doi.org/10.1063/1.481544.

[227] Sonia Coriani et al. "The accuracy of ab initio molecular geometries for systems containing second-row atoms". In: *The Journal of Chemical Physics* 123.18 (2005), p. 184107. DOI: 10.1063/1.2104387. eprint: https://doi.org/10.1063/1.2104387. URL: https://doi.org/10.1063/1.2104387.

[228] Hans-Joachim Werner and Wilfried Meyer. "A quadratically convergent MCSCF method for the simultaneous optimization of several states". In: *The Journal of Chemical Physics* 74.10 (1981), pp. 5794–5801. DOI: 10.1063/1.440892. eprint: https://doi.org/10.1063/1.440892. URL: https://doi.org/10.1063/1.440892.

[229] Achintya Kumar Dutta et al. "Exploring the Accuracy of a Low Scaling Similarity Transformed Equation of Motion Method for Vertical Excitation Energies". In: *Journal of Chemical Theory and Computation* 14.1 (2018). PMID: 29206453, pp. 72–91. DOI: 10.1021/acs.jctc.7b00802. eprint: https://doi.org/10.1021/acs.jctc.7b00802. URL: https://doi.org/10.1021/acs.jctc.7b00802.

[230] Janus J. Eriksen et al. "The Ground State Electronic Energy of Benzene". In: *The Journal of Physical Chemistry Letters* 11.20 (2020). PMID: 33022176, pp. 8922–8929. DOI: 10.1021/acs.jpclett.0c02621. eprint: https://doi.org/10.1021/acs.jpclett.0c02621. URL: https://doi.org/10.1021/acs.jpclett.0c02621.

[231] Ö. Legeza, J. Roder, and B. A. Hess. "QC-DMRG study of the ionic-neutral curve crossing of LiF". In: *Molecular Physics* 101.13 (2003), pp. 2019–2028. DOI: 10.1080/0026897031000155625. eprint: https://www.tandfonline.com/doi/pdf/10.1080/0026897031000155625. URL: https://www.tandfonline.com/doi/abs/10.1080/0026897031000155625.

[232] Sandeep Sharma et al. "Spectroscopic accuracy directly from quantum chemistry: Application to ground and excited states of beryllium dimer". In: *The Journal of Chemical Physics* 140.10 (2014), p. 104112. DOI: 10.1063/1.4867383. eprint: https://doi.org/10.1063/1.4867383. URL: https://doi.org/10.1063/1.4867383.

[233] George H. Booth and Garnet Kin-Lic Chan. "Communication: Excited states, dynamic correlation functions and spectral properties from full configuration interaction quantum Monte Carlo". In: *The Journal of Chemical Physics* 137.19 (2012), p. 191102. DOI: 10.1063/1.4766327. eprint: https://doi.org/10.1063/1.4766327. URL: https://doi.org/10.1063/1.4766327.

[234] Alexander Humeniuk and Roland Mitrić. "Excited states from quantum Monte Carlo in the basis of Slater determinants". In: *The Journal of Chemical Physics* 141.19 (2014), p. 194104. DOI: 10.1063/1.4901020. eprint: https://doi.org/10.1063/1.4901020. URL: https://doi.org/10.1063/1.4901020.

[235] N. S. Blunt, Ali Alavi, and George H. Booth. "Krylov-Projected Quantum Monte Carlo Method". In: *Phys. Rev. Lett.* 115 (5 July 2015), p. 050603. DOI: 10.1103/PhysRevLett.115.050603. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.050603.

[236] Yuki Kurashige, Garnet Kin-Lic Chan, and Takeshi Yanai. "Entangled quantum electronic wavefunctions of the Mn4CaO5 cluster in photosystem II". In: *Nature Chemistry* 5.8 (Aug. 2013), pp. 660–666. ISSN: 1755-4349. DOI: 10.1038/nchem.1677. URL: https://doi.org/10.1038/nchem.1677.

[237] Fengyi Liu et al. "Multireference Ab Initio Density Matrix Renormalization Group (DMRG)-CASSCF and DMRG-CASPT2 Study on the Photochromic Ring Opening of Spiropyran". In: *Journal of Chemical Theory and Computation* 9.10 (2013). PMID: 26589164, pp. 4462–4469. DOI: 10.1021/ct400707k. eprint: https://doi.org/10.1021/ct400707k. URL: https://doi.org/10.1021/ct400707k.

[238] Werner Dobrautz et al. "Spin-Pure Stochastic-CASSCF via GUGA-FCIQMC Applied to Iron–Sulfur Clusters". In: *Journal of Chemical Theory and Computation* 17.9 (2021). PMID: 34469685, pp. 5684–5703. DOI: 10.1021/acs.jctc.1c00589. eprint: https://doi.org/10.1021/acs.jctc.1c00589. URL: https://doi.org/10.1021/acs.jctc.1c00589.

[239] Michel Caffarel et al. "Communication: Toward an improved control of the fixed-node error in quantum Monte Carlo: The case of the water molecule". In: *The Journal of Chemical Physics* 144.15 (2016), p. 151103. DOI: 10.1063/1.4947093. eprint: https://doi.org/10.1063/1.4947093. URL: https://doi.org/10.1063/1.4947093.

[240] A. Scemama et al. "Accurate nonrelativistic ground-state energies of 3d transition metal atoms". In: *The Journal of Chemical Physics* 141.24 (2014), p. 244110. DOI: 10.1063/1.4903985. eprint: https://doi.org/10.1063/1.4903985. URL: https://doi.org/10.1063/1.4903985.

[241] Anthony Scemama et al. "Influence of pseudopotentials on excitation energies from selected configuration interaction and diffusion Monte Carlo". In: *Results in Chemistry* 1 (2019), p. 100002. ISSN: 2211-7156. DOI: https://doi.org/10.1016/j.

rechem.2019.100002. URL: https://www.sciencedirect.com/science/article/pii/S2211715619300025.

[242] Anouar Benali et al. "Toward a systematic improvement of the fixed-node approximation in diffusion Monte Carlo for solids—A case study in diamond". In: *The Journal of Chemical Physics* 153.18 (2020), p. 184111. DOI: 10.1063/5.0021036. eprint: https://doi.org/10.1063/5.0021036. URL: https://doi.org/10.1063/5.0021036.

[243] Roland Assaraf and Michel Caffarel. "Computing forces with quantum Monte Carlo". In: *The Journal of Chemical Physics* 113.10 (2000), pp. 4028–4034. DOI: 10.1063/1.1286598. eprint: https://doi.org/10.1063/1.1286598. URL: https://doi.org/10.1063/1.1286598.

[244] Roland Assaraf and Michel Caffarel. "Zero-variance zero-bias principle for observables in quantum Monte Carlo: Application to forces". In: *The Journal of Chemical Physics* 119.20 (2003), pp. 10536–10552. DOI: 10.1063/1.1621615. eprint: https://doi.org/10.1063/1.1621615. URL: https://doi.org/10.1063/1.1621615.

[245] Kousuke Nakano et al. "TurboRVB: A many-body toolkit for ab initio electronic simulations by quantum Monte Carlo". In: *The Journal of Chemical Physics* 152.20 (2020), p. 204121. DOI: 10.1063/5.0005037. eprint: https://doi.org/10.1063/5.0005037. URL: https://doi.org/10.1063/5.0005037.

[246] Mickaël Véril et al. "QUESTDB: A database of highly accurate excitation energies for the electronic structure community". In: *WIREs Computational Molecular Science* 11.5 (2021), e1517. DOI: https://doi.org/10.1002/wcms.1517. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1517. URL: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1517.

[247] Giuseppe M. J. Barca et al. "Recent developments in the general atomic and molecular electronic structure system". In: *The Journal of Chemical Physics* 152.15 (2020), p. 154102. DOI: 10.1063/5.0005188. eprint: https://doi.org/10.1063/5.0005188. URL: https://doi.org/10.1063/5.0005188.

[248] Robert R. Lucchese and Henry F. Schaefer. "Charge-transfer complexes. Ammonia-molecular fluorine, ammonia-molecular chlorine, ammonia-chlorine fluoride, trimethylamine-molecular fluorine, trimethylamine-molecular chlorine, and trimethylamine-chlorine fluoride". In: *Journal of the American Chemical Society* 97.25 (1975), pp. 7205–7210. DOI: 10.1021/ja00858a001. eprint: https://doi.org/10.1021/ja00858a001. URL: https://doi.org/10.1021/ja00858a001.

[249] Yan Zhao and Donald G. Truhlar. "Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States". In: *The Journal of Physical Chemistry A* 110.49 (2006). PMID: 17149824, pp. 13126–13130. DOI: 10.1021/jp066479k. eprint: https://doi.org/10.1021/jp066479k. URL: https://doi.org/10.1021/jp066479k.

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

[251] Johan Lorentzon, Markus P. Fuelscher, and Bjoern O. Roos. "Theoretical Study of the Electronic Spectra of Uracil and Thymine". In: *Journal of the American Chemical Society* 117.36 (1995), pp. 9265–9273. DOI: 10.1021/ja00141a019. eprint: https://doi.org/10.1021/ja00141a019. URL: https://doi.org/10.1021/ja00141a019.

[252] Dániel Kánnár and Péter G. Szalay. "Benchmarking coupled cluster methods on singlet excited states of nucleobases". In: *Journal of Molecular Modeling* 20.11 (Nov. 2014), p. 2503. ISSN: 0948-5023. DOI: 10.1007/s00894-014-2503-2. URL: https://doi.org/10.1007/s00894-014-2503-2.

[253] Daniil A. Fedotov et al. "Excited-State Absorption of Uracil in the Gas Phase: Mapping the Main Decay Paths by Different Electronic Structure Methods". In: *Journal of Chemical Theory and Computation* 17.3 (2021). PMID: 33529532, pp. 1638–1652. DOI: 10.1021/acs.jctc.0c01150. eprint: https://doi.org/10.1021/acs.jctc.0c01150. URL: https://doi.org/10.1021/acs.jctc.0c01150.

[254] Sandro Sorella. "Generalized Lanczos algorithm for variational quantum Monte Carlo". In: *Phys. Rev. B* 64 (2 June 2001), p. 024512. DOI: 10.1103/PhysRevB.64.024512. URL: https://link.aps.org/doi/10.1103/PhysRevB.64.024512.

[255] Hongxin Huang and Zexing Cao. "A novel method for optimizing quantum Monte Carlo wave functions". In: *The Journal of Chemical Physics* 104.1 (1996), pp. 200–205. DOI: 10.1063/1.470889. eprint: https://doi.org/10.1063/1.470889. URL: https://doi.org/10.1063/1.470889.

[256] Hongxin Huang et al. "A novel quantum Monte Carlo strategy: Surplus function approach". In: *The Journal of Chemical Physics* 110.8 (1999), pp. 3703–3707. DOI: 10.1063/1.478260. eprint: https://doi.org/10.1063/1.478260. URL: https://doi.org/10.1063/1.478260.

[257] Leon Otis and Eric Neuscamman. "Combining State-Specific Quantum Chemistry and Quantum Monte Carlo for Molecular Excited States". In: *arXiv preprint arXiv:2111.07221* (2021).

[258] Monika Dash et al. "Perturbatively Selected Configuration-Interaction Wave Functions for Efficient Geometry Optimization in Quantum Monte Carlo". In: *Journal of Chemical Theory and Computation* 14.8 (2018). PMID: 29953810, pp. 4176–4182. DOI: 10.1021/acs.jctc.8b00393. eprint: https://doi.org/10.1021/acs.jctc.8b00393. URL: https://doi.org/10.1021/acs.jctc.8b00393.

[259] Paul Boulanger et al. "Fast and Accurate Electronic Excitations in Cyanines with the Many-Body Bethe–Salpeter Approach". In: *J. Chem. Theory Comput.* 10.3 (2014). PMID: 26580191, pp. 1212–1218. DOI: 10.1021/ct401101u. eprint: https://doi.org/10.1021/ct401101u. URL: https://doi.org/10.1021/ct401101u.

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

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

[262] Péter G. Szalay et al. "Benchmark Studies on the Building Blocks of DNA. 3. Watson–Crick and Stacked Base Pairs". In: *The Journal of Physical Chemistry A* 117.15 (2013). PMID: 23473108, pp. 3149–3157. DOI: 10.1021/jp3100975. eprint: https://doi.org/10.1021/jp3100975. URL: https://doi.org/10.1021/jp3100975.

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

[264] Kayahan Saritas et al. "Investigation of a Quantum Monte Carlo Protocol To Achieve High Accuracy and High-Throughput Materials Formation Energies". In: *J. Chem. Theory Comput.* 13.5 (2017). PMID: 28358499, pp. 1943–1951. DOI: 10.1021/acs.jctc.6b01179. eprint: https://doi.org/10.1021/acs.jctc.6b01179. URL: https://doi.org/10.1021/acs.jctc.6b01179.

[265] Jan Hermann, Zeno Schätzle, and Frank Noé. "Deep-neural-network solution of the electronic Schrödinger equation". In: *Nature Chemistry* 12.10 (Oct. 2020), pp. 891–897. ISSN: 1755-4349. DOI: 10.1038/s41557-020-0544-y. URL: https://doi.org/10.1038/s41557-020-0544-y.