Absolutely Localized Huzinaga Projection Based Embedding for Efficient and Accurate Molecular Modeling

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Daniel Sterling Graham

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jason D. Goodpaster

May, 2022

Acknowledgements

My research advisor, Jason D. Goodpaster, was pivotal to my happiness and success while completing this work. His insight and knowledge were invaluable assets to my work and his support of my growth and development is abundant and clear. He has been a mentor, collaborator, and friend to me these years and I hope for many years to come.

My colleagues in the Goodpaster research group have sparked many important discussions both scientific and philosophical. Early in my graduate experience, Dhabih Chulhai and Xuelan Wen were vital sources of inspiration and joy. I cannot count the number of times Nancy Thao and Nick Erickson helped me navigate the administrative side of graduate school with kindness and haste. I would also like to thank my scientific collaborators, James J. Shepherd and Haylay R. Petras, for their diligence, wisdom, and patience.

Because my work was done with the help of many friends, I must thank Peter Clement, Ryan Daley, Stephen Dempsey, and Jacob Prat for transforming the Franklin house into a little Sun on the hill. I am grateful to my fellow band mates in Boy Crazy for teaching me the meaning of Rock. I have been blessed with support from my family throughout my life and their love was essential to this work. Finally, thank you Claire for your constant encouragement, dedication, and love.

We did it.

Dedication

To my family

Abstract

Within computational chemistry, Kohn-Sham density functional theory has become invaluable for low computational cost quantum mechanical calculations. However, computationally expensive post-Hartree-Fock wave function quantum methods are still required to accurately model many chemical systems. We are able to recreate post-Hartree-Fock wave function levels of computational accuracy while only incurring computational costs on par with density functional theory using our absolutely localized Huzinaga level-shift projection based wave function in density functional theory embedding method. Computationally accurate gas adsorption energies on transition metal clusters of metal organic frameworks, spin transition energies of iron models, full configuration interaction (FCI) level energies of small molecules on surfaces and analytical nuclear gradients of our embedding method have been developed and are reported here. We have demonstrated many valuable features of our embedding method such as systematic improvability, applicability to a breadth of chemical problems, and low computational cost for highly accurate calculations. Our Huzinaga embedding method pushes the boundaries of computational chemistry by enabling the calculation of accurate molecular energies for chemical systems previously beyond the scope of existing computational methods. We anticipate our method will be of interest to anyone studying large, complex systems that cannot be accurately modeled using density functional theory and are too large for traditional post-Hartree-Fock methods.

Contents

List of Tables

List of Figures

List of Abbreviations

CAS	Complete active space
CASSCF	Complete active space self-consistent field
CBS	Complete basis set
CC	Coupled cluster
CCSD(T)	Coupled cluster singles and doubles with perturbative triples
CI	Configuration interaction
DFT	Density functional theory
DMET	Density matrix emedding theory
FCI	Full configuration interaction
FCIQMC	Full configuration interaction quantum Monte Carlo
HF	Hartree-Fock
$i-$**FCIQMC**	Initiator adaptation of full configuration interaction quantum Monte Carlo
KS-DFT	Kohn-Sham density functional theory
MAE	Maximum absolute error
MCSCF	Multi-configurational self-consistent field theory
MOF	Metal organic framework
MP2	Second order Møller-Plesset perturbation theory
MRCI	Multireference configuration interaction
ONIOM	Our n-layered integrated molecular orbital and molecular mechanics
QMC	Quantum Monte Carlo
QM/MM	Quantum mechanics / molecular mechanics
QSoME	Quantum Solid state and Molecular Embedding
STE	Spin-transition energy
WF	Wave function
XC	Exchange-correlation

Chapter 1

Introduction

Computational modeling of chemical processes can be an invaluable tool for understanding complex reactions, screening large chemical databases, and guiding experiments to name a few key applications. In order to accurately model electronic effects, such as bond formation or electronic excitation, a quantum mechanical approach is usually required. Within the field of computational quantum mechanics, there are a wide variety of methods and implementations one could choose to model a chemical system. Density functional theory[4, 5] (DFT) is one of the most widely used quantum methods today, with computational scaling that allows modeling of systems with thousands of atoms while achieving relatively high accuracy through the use of an approximate exchange-correlation electron density functional. However there exist many important chemical systems, such as those with degenerate and near-degenerate states (e.g. stretched covalent bonds, transition metals), [6, 7, 8, 9, 10] and properties, such as spin-state energetics of transition metals,[11, 12, 13, 14, 15, 16, 17] that cannot be consistently and accurately described using current DFT methods. Furthermore, there is a myriad of approximate exchange-correlation density functionals to choose from when modeling a system, and thus determining a strategy to select the proper exchange-correlation functional can become a problem itself. Correlated wave function (WF) methods, such as coupled cluster (CC) and complete active space (CAS), do not rely on approximate exchange-correlation density functionals to model electronic effects, and so for some systems are more accurate than DFT. This improved accuracy comes at a high computational cost however, and many WF methods are limited to modeling a fraction of the number of

atoms one can model using DFT. The trade-off between accuracy and computational cost is one of the core challenges within the field of computational chemistry.

Quantum embedding methods take advantage of electronic localization to dramatically lower the computational cost of highly accurate calculations. Because many complex electronic interactions are limited to a relatively small region within the full chemical system, one can perform highly accurate calculations on a small subsystem while the remainder of the system is calculated using a less costly method. There are many quantum embedding methods in use today (for recent reviews see [18, 19, 20]), but most relevant to the work in this dissertation is DFT embedding[18, 21, 22, 23, 24, 25]. DFT embedding involves dividing a system into subsystems that are modeled using DFT. The interaction between subsystems is modeled by including a DFT-level embedding potential which can be calculated using non-additive kinetic energy functionals[26, 27, 28, 29, 30, 31, 32, 33, 34, 35] or through the use of a projection operator[36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 25]. DFT embedding also provides a framework for highly accurate WF-in-DFT embedding. For WF-in-DFT embedding one simulates each subsystem using DFT to generate an embedding potential for a small subsystem of interest. Then one performs a WF calculation on that small subsystem including the embedded external potential generated by the subsystem DFT calculation. Previous work in our group and others[47, 48, 49, 50, 41, 42, 46, 51] has demonstrated WF level accuracy for a substantial reduction in computational cost utilizing the WF-in-DFT embedding method. Our contributions to this rich field include the development and implementation of WF-in-DFT embedding utilizing a form of basis set truncation we call the absolutely localized basis set[47], and the use of the Huzinaga level-shift projection operator[52, 53] for WF-in-DFT embedding.[47] Our absolutely localized Huzinaga projection based WF-in-DFT embedding method continues to be an important development within this field, pushing the boundaries of accurate chemical simulation to larger and larger systems.

This dissertation follows the development of absolutely localized Huzinaga projection based WF-in-DFT embedding from early implementation and testing phases, to a robust computational chemistry method with a variety of useful features.

- Chapter 2 discusses the closed shell implementation of the method. Within this chapter we demonstrate important features of our embedding method such as

systematic improvability and high accuracy, provide practical guidance for use of the method based on experimental results, and calculate several energies of a complex metal organic framework (MOF) model.

- Chapter 3 demonstrates the open-shell implementation of the method. We again demonstrate accuracy and systematic improvability within the open shell implementation, discuss the nuances of unrestricted and restricted open shell embedding calculations, and demonstrate significantly reduced exchange-correlation functional dependence of our method for spin transition energies of transition metal complexes.

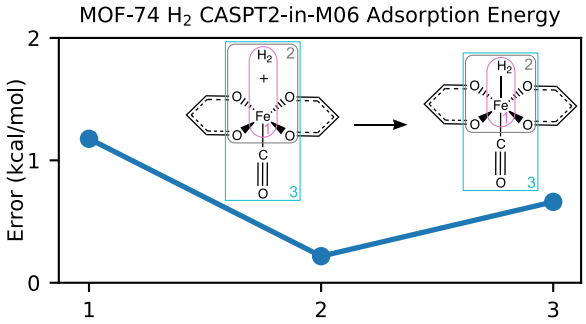
- Chapter 4 provides examples of extremely high accuracy $i-$FCIQMC -in-DFT embedding calculations. Through our collaborative research we demonstrate the power of the absolutely localized Huzinaga embedding method to reduce the computational cost and complexity for high level WF calculations.

- Chapter 5 presents the derivation of analytical nuclear gradients for our absolutely localized Huzinaga WF-in-DFT embedding method and provides evidence supporting correct derivation and implementation.

- Chapter 6 concludes the dissertaion with a discussion of the absolutely localized Huzinaga projection based WF-in-DFT method thus far, and provides recommendations for future work improving the method.

Chapter 2

Robust, accurate, and efficient: quantum embedding using the Huzinaga level-shift projection operator for complex systems



Reprinted with permission from the article by Daniel S. Graham, Xuelan Wen, Dhabih V. Chulhai and Jason D. Goodpaster, *J. Chem. Theory Comput.* 2020, 16, 4, 2284–2295.

2.1 Introduction

One of the fundamental challenges of quantum chemistry is balancing computational cost and accuracy. Kohn-Sham Density Functional Theory[4, 54] (KS-DFT) has been a computational chemistry mainstay as it balances cost and accuracy well for many chemical systems. However, KS-DFT relies on an approximate exchange-correlation functional which results in several well documented deficiencies including underestimation of chemical reaction barriers, and inaccurate description of degenerate and near-degenerate states, such as in transition metal systems and covalent bond dissociation.[6, 7, 8, 9, 10] These interactions are essential for accurately describing a variety of systems such as gas binding to metal organic frameworks (MOFs) and reaction energy barriers. Correlated wave function (WF) methods such as coupled cluster (CC)[55, 56] and complete active space (CAS)[57] have been shown to more accurately reproduce the aforementioned interactions.[58, 59] Additionally, most WF methods are systematically improvable: the accuracy of calculation may be improved through a well defined process (e.g. including additional excitations for CC calculations, or increasing the size of the CAS active space). Yet, for large systems the computational cost of most WF methods are several orders of magnitude larger than KS-DFT methods. Frequently, one is interested in chemical transformations that are localized to a small region of the overall system, such as bond formation or elimination, molecular adsorption, or bond rotation. Many embedding methods such as QM/MM[60], ONIOM[61], DMET[62, 18], embedded mean-field theory[63, 64, 65, 66, 67], Green's function embedding[68, 69, 18], partition DFT,[70, 71, 72] and DFT embedding[21, 18] among many others [73, 74, 75, 76, 77], take advantage of this intrinsic localization of chemical transformations to achieve substantially improved accuracy for a nominal additional computational cost. By dividing the total system into subsystems, important local interactions can be accurately modeled at significantly reduced computational cost. This approach is particularly advantageous when performing WF calculations, due to the steep computational scaling of WF methods.

DFT embedding provides a formally exact framework for subdividing a system where the interactions between subsystems are treated using DFT.[21, 78, 79, 80, 81, 34, 28]

The DFT subsystem interaction potential can then be used to easily embed a WF calculation within the DFT potential of the full system.[82, 38, 36, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93] Calculation of the subsystem interaction potential however, is the central challenge of DFT embedding methods. The interaction potential for DFT embedding differs from KS-DFT as subdividing a system introduces a non-additive kinetic energy component. This non-additive kinetic energy may be approximated[26, 5, 27, 29, 31, 21, 94, 95, 96, 30], numerically calculated[21, 28, 97], or eliminated all together through subsystem orbital orthogonalization[98, 47, 99, 43, 100, 101, 102]. The use of subsystem orbital orthogonalization methods for exact DFT embedding was studied by the Manby and Miller groups through the use of a constant shift μ-projection operator[36, 37, 38, 39, 40]. This μ-projection operator demonstrated impressive results and in a later paper, Kallay and co-workers suggested[44] the use of the Huzinaga[52, 53, 103] level-shift projection operator as an alternative to the μ-projection operator. Our group generalized the Huzinaga level-shift projection operator with a freeze-and-thaw localization scheme and demonstrated significant success using the absolutely localized basis on molecular ground[47] and excited[104] states, and ground state periodic systems.[48]

In order to make WF-in-DFT embedding feasible for large systems, the number of valence orbitals in the WF region must be managed. Including the basis functions of the full system in the embedded WF subsystem simply moves orbitals not occupied in the WF subsystem to the virtual space, which for CC calculations actually increases the computational cost upon embedding due to the higher scaling of CC methods with respect to virtual orbitals compared to occupied orbitals. There have been several methods proposed for reducing the number of basis functions in the WF subsystem while maintaining high accuracy such as basis set truncation[51, 105], "bottom up" basis set extension[46], dual basis set approximation[106], and concentric localization and truncation of virtual space[42]. Our absolutely localized basis scheme may be considered the most strict of the truncation methods mentioned, including only basis functions in the WF calculation which are centered on the atoms specified for the WF subsystem. While this constraint does increase the errors for absolute energies, we have found that absolutely localized basis truncation actually performs better for WF-in-DFT reaction energies than using the full system basis. [47] We argue this improved reaction energy is due to a systematic cancellation of error between products and reactants

enabled by the strict localization of the WF orbitals.

Expanding upon our previous study, the following results indicate that this method provides robust, systematically improvable results for a diverse test set, and thus has applicability to a wide variety of chemical systems. These systems include partitioning across two covalent bonds and a conjugated π network, and gas adsorption onto transition metals among others. Additionally, we provide recommendations for system partitioning to achieve the highest degree of chemical accuracy. Finally, we demonstrate the power of this method for multiconfigurational embedding by applying it to a MOF cluster system approaching the limit of traditional WF methods.

2.2 Theory

In subsystem DFT methods, the electron density matrix, γ, of a system is subdivided into two subsystems,

$$\gamma = \gamma^{\mathrm{A}} + \gamma^{\mathrm{B}} \tag{2.1}$$

where γ^{A} and γ^{B} are the electron density matrices of subsystem A and B respectively. In the absolutely localized basis, only basis functions centered on atoms in each subsystem are used to expand the Kohn-Sham orbitals,

$$\phi_i^{\mathrm{A}} = \sum_{\mu} C_{i\mu}^{\mathrm{A}} \chi_{\mu}^{\mathrm{A}}, \tag{2.2}$$

$$\phi_i^{\mathrm{B}} = \sum_{\mu} C_{i\mu}^{\mathrm{B}} \chi_{\mu}^{\mathrm{B}}, \tag{2.3}$$

where ϕ_i^{A}, ϕ_i^{B}, $C_{i\mu}^{\mathrm{A}}$, $C_{i\mu}^{\mathrm{B}}$, χ_{μ}^{A}, and χ_{μ}^{B} are the Kohn-Sham orbitals, orbital coefficients, and basis functions associated with subsystem A and B, respectively.

Thus in the absolutely localized basis, only basis functions centered on atoms in subsystem A are used to describe the electron density matrix of subsystem A. The absolutely localized basis is then also used when performing subsequent WF calculations. This choice of basis set expansion has previously been called the monomolecular basis[107]; however, as our subsystems are not monomers, we choose to refer to this as the absolutely localized basis. Level shift projection operators enforce orthogonality

via adding a projection operator to the subsystem Hamiltonian. Because of the general form, the embedding and projection potentials may be added to any subsystem Hamiltonian to calculate energy, either for DFT or WF theory methods. The Huzinaga projection operator is,

$$\mathbf{P}^{\mathrm{B}} = -\frac{1}{2}\left(\mathbf{F}^{\mathrm{AB}}\gamma^{\mathrm{B}}\mathbf{S}^{\mathrm{BA}} + \mathbf{S}^{\mathrm{AB}}\gamma^{\mathrm{B}}\mathbf{F}^{\mathrm{BA}}\right) \tag{2.4}$$

where,

$$\mathbf{F}^{\mathrm{AB}} = \langle\chi^{\mathrm{A}}|\hat{F}|\chi^{\mathrm{B}}\rangle \tag{2.5}$$

with $\hat{F}$ being the full system Fock operator and

$$\mathbf{S}^{\mathrm{AB}} = \langle\chi^{\mathrm{A}}|\chi^{\mathrm{B}}\rangle. \tag{2.6}$$

Thus, for subsystem A, orbitals within A that are not orthogonal to subsystem B are shifted to higher energies, and vice versa.

The projected Fock matrix of subsystem A embedded in subsystem B is,

$$\mathbf{f}^{\text{A-in-B}} = \mathbf{h}^{\text{A-in-B}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}] + \mathbf{J}[\gamma^{\mathrm{A}}] + \mathbf{v}_{\mathrm{xc}}[\gamma^{\mathrm{A}}] \tag{2.7}$$

where $\mathbf{J}$ is the electron Coulomb potential, $\mathbf{v}_{\mathrm{xc}}$ is the exchange-correlation (XC) potential, and the embedded core Hamiltonian is

$$\mathbf{h}^{\text{A-in-B}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}] = \mathbf{h} + \mathbf{J}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}] - \mathbf{J}[\gamma^{\mathrm{A}}] + \mathbf{v}_{\mathrm{xc}}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}] - \mathbf{v}_{\mathrm{xc}}[\gamma^{\mathrm{A}}] + \mathbf{P}^{\mathrm{B}}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}], \tag{2.8}$$

where $\mathbf{h}$ is the total one-electron Hamiltonian.

The form of our overall WF-in-DFT embedding energy,

$$E^{\text{Full}}_{\text{WF-in-DFT}} = E^{\text{Full}}_{\text{KS-DFT}} - E^{\mathrm{A}}_{\mathrm{DFT}} + E^{\mathrm{A}}_{\mathrm{WF}}, \tag{2.9}$$

uses an subtractive embedding framework like ONIOM[61], where $E^{\text{Full}}_{\text{WF-in-DFT}}$ is the total WF-in-DFT energy, $E^{\text{Full}}_{\text{KS-DFT}}$ is the canonical KS-DFT of the full system, $E^{\mathrm{A}}_{\mathrm{DFT}}$ is the DFT energy of subsystem A embedded in the full system,

$$E^{\mathrm{A}}_{\mathrm{DFT}} = \mathrm{Tr}\left(\gamma^{\mathrm{A}} \cdot \mathbf{h}^{\text{A-in-B}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}]\right) + \mathrm{J}[\gamma^{\mathrm{A}}] + E_{\mathrm{xc}}[\gamma^{\mathrm{A}}], \tag{2.10}$$

and $E^{\mathrm{A}}_{\mathrm{WF}}$ is the WF energy of subsystem A embedded in the DFT potential of the full system

$$E^{\mathrm{A}}_{\mathrm{WF}} = \langle \Psi^{\mathrm{A}} | \hat{H}^{\text{A-in-B}} | \Psi^{\mathrm{A}} \rangle. \tag{2.11}$$

Here $\hat{H}^{\text{A-in-B}}$ is

$$\hat{H}^{\text{A-in-B}} = \hat{h}^{\text{A-in-B}} + \hat{g}^{\mathrm{A}} \tag{2.12}$$

where $\hat{h}^{\text{A-in-B}}$ is equivalent to eq. 2.8 (now written in operator form) and $\hat{g}^{\mathrm{A}}$ is the two-electron operator for a given WF theory acting on the electrons of subsystem A. In practice, this involves simply performing the WF theory calculation by replacing the standard total one-electron Hamiltonian, $\mathbf{h}$, with the embedded one-electron Hamiltonian, $\mathbf{h}^{\text{A-in-B}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}]$. The final WF-in-DFT embedding energy form differs from several previous energy formulations[38, 51] by correcting using the fully relaxed, total KS-DFT energy of the system in a similar manner to Carter et. al.[34]

2.3 Computational Details

All organic molecule geometries were optimized using Gaussian 16[108] with the M06 functional[3] and aug-cc-pVTZ basis[109], and are reported in the Supporting Information (SI). The Fe-MOF-74 cluster geometry was calculated using Gaussian 16[108] following the procedure outlined by Lee and coworkers[110] and included in the SI. Single point DFT and CCSD(T) energy calculations were calculated using PySCF version 1.6[111] and CASPT2 calculations were done using Molpro 2019.2[112, 113, 114, 115, 116]. Embedded CASPT2 results were calculated using Molpro 2012.1[112, 114, 115, 116]. For all DFT calculations, PySCF grid level was set to 4. Full system WF energies used for comparison to the embedded results are included in Appendix A.

In order to generate a final WF-in-DFT embedding energy, our method utilizes a freeze and thaw[117] scheme and an ONIOM[61] energy formulation; this method is as follows: (1) We start with a full system KS-DFT calculation. (2) The resulting converged full system KS-DFT electron density matrix is used as the initial guess for subsystem electron density matrix. To generate the subsystem electron density matrix guess in the absolutely localized basis, the components of the full system electron density matrix consisting entirely of atomic orbitals centered on subsystem atoms are

extracted and normalized to create an initial subsystem electron density matrix. (3) The subsystem electron density is then relaxed following a freeze and thaw protocol, where one subsystem electron density is allowed to relax while the rest of the system electron density is frozen, then that subsystem is frozen and another subsystem is "thawed" and allowed to relax. During freeze and thaw relaxation, the full system Fock matrix, used in the projection operator, is recomputed each time all subsystems undergo a relaxation cycle. (4) Finally, once the subsystem electron densities are determined converged between freeze and thaw cycles, embedded WF calculations may be performed using DFT-in-DFT potentials. WF calculations use the DFT subsystem electron density matrix as a starting guess for an initial Hartree-Fock calculation followed by the subsequent WF calculation and utilize the embedded core Hamiltonian as defined in eq. 2.8. All embedding calculations were performed with our open source Quantum Solid state and Molecular Embedding (QSoME) code [118].

2.4 Results and Discussion

2.4.1 Systematic Improvability

A useful feature of many WF methods is their systematic improvability — there is a clearly defined procedure for improving results towards an accurate solution. For instance, one direction of systematic improvability is the description of the wave function; coupled cluster methods are improved by including additional excitations and multiconfigurational methods are improved by increasing the number of configurations. In contrast, KS-DFT does not benefit from systematic improvability, as there are a plethora of exchange-correlation functionals and no well-defined procedure for choosing a functional to improve accuracy. Therefore, within KS-DFT it is not apparent how to improve the accuracy of a particular calculation with a given exchange-correlation functional. Here, we show that unlike KS-DFT, WF-in-DFT provides systematic improvability. In order to determine the systematic improvability of the Huzinaga level-shift embedding method, embedding results were compared to full system WF energies while varying the size of the WF subsystem and the number of basis functions. Previously, it has been shown that other WF-in-DFT embedding methods exhibit this systematic improvability;[39] here, these calculations show that the accuracy of absolutely localized

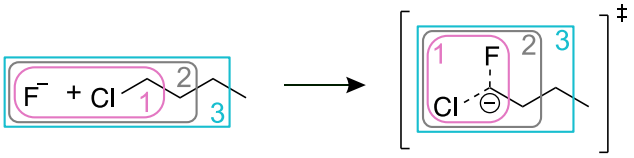


Figure 2.1: 1-chlorobutane S_N2 transition state reaction. Numbering indicates the size of WF subsystem, where the number corresponds to the number of carbon atom centers in the subsystem. The smallest WF subsystem includes only the carbon center most local to the S_N2 reaction. For other S_N2 reactions studied, WF subsystems similarly incrementally increase.

Huzinaga level-shift embedding method can be systematically improved by increasing the number of atoms in the WF subsystem.

Additionally, our WF-in-DFT methodologies retain all of the systematically improvability features of WF methods. As our method is generality applicable to any WF method, we know that we can always systematically improve the quality of the wave function. Furthermore, we have previously used the initiator adaptation of full configuration interaction quantum Monte Carlo ($i-$FCIQMC) in our embedding method;[119] therefore, our embedding scheme is compatible with other high-quality WF methodologies. In addition to the systematic improvability of the description of the wave function, the basis set can also be expanded to converge to the complete basis set limit. In practice, one typically does calculations at increasing basis set size and extrapolates to the complete basis set limit. It is not obvious that absolutely localized projection-based embedding will allow for the same extrapolation schemes. Therefore, we test basis set extrapolation and show that WF-in-DFT embedding is systematically improvable with respect to increasing the size of the basis.

The particular reactions we studied to determine systematic improvability were a series of S_N2 activation energies. The S_N2 activation energy was studied in a previous article from our group[47] and demonstrated the success of the embedding method when dividing a system across a single covalent bond. However, where the antecedent article only studied one system subdivision for the reaction, here the system is divided into subsystems of increasingly large WF subsystem sizes. These subsystem divisions are shown in Figure 2.1, where the size of the WF subsystem increases by including carbon centers along the alkane chain. The smallest WF subsystem includes only carbon 1, the

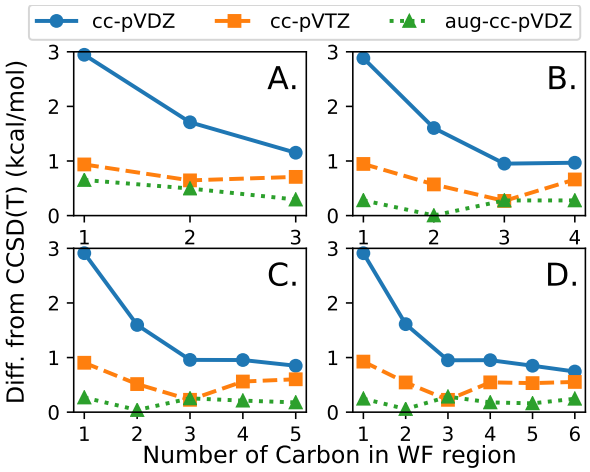


Figure 2.2: Absolute energy difference of WF-in-DFT embedding from full system CCSD(T) in the same basis set of S_N2 activation energy. Reactants are, (A) 1-chlorobutane, (B) 1-chloropentane, (C) 1-chlorohexane, and (D) 1-chloroheptane. Number of carbon in the WF subsystem corresponds to the WF subsystem subdivisions represented in figure 2.1. As previously mentioned, the full system CCSD(T) energies that were used for comparison to the embedded results are present in Appendix A.

halogen atoms, and the hydrogen bonded to carbon 1, as these are the atoms closest to the region of chemical change in the system. The WF subsystem is increased in size by including the carbon and its bonded hydrogen directly adjacent to the WF subsystem. Subsystems were charged to create closed shell fragments with WF subsystem given an additional -1 charge and DFT subsystem given an additional +1 charge. WF-in-DFT energies were calculated using CCSD(T) as the WF method and M06[3] as the DFT exchange-correlation functional. Four halogenated hydrocarbon systems energies were calculated using the incrementally increasing WF subsystem method described above. These embedding energies using the specified basis set were then compared to the CCSD(T) energy of the corresponding full system in the same basis set (Figure 2.2).

Using the cc-pVDZ basis set, WF-in-DFT energies systematically converged to the full system CCSD(T) energy with increasing size of the WF region to within 1 kcal/mol. For all alkane systems, energies consistently approach full system results until convergence of the energy with the inclusion of 3 carbon in the WF subsystem. Not only do energies converge with increasing WF subsystem size, but WF-in-DFT energies converge irrespective of total size of the system. Since convergence is not dependent on

the total size of the system, the size of the WF calculation may be limited to a much smaller subset of the entire system.

Additionally, from the data in Figure 2.2 it is apparent that the error does not always monotonically decay. For instance, in Figure 2.2D the cc-pVTZ error increases when increasing the WF subsystem size from 3 to 4 carbons. While the description of the total system for both the reactant and the transition state is improving with increasing the size of the WF region, the difference of the activation energy compared to canonical CCSD(T) calculations is not necessarily lower in error. This is due to difference in the cancellation of errors in the calculations. A similar effect is seen with the convergence of basis set size for activation energies in canonical CCSD(T) where the activation energy may not monotonically converge to a value, despite the fact that increasing basis set size always provides a better description of the wave function.[120] The discrepancy then is caused from the reactant and transition state converging differently to the complete basis set limit leading to slightly different cancellation of error in the transition barrier calculation. However, as evident in Figure 2.2 this differing cancellation of error only becomes apparent once the error is less than 1 kcal/mol. We discuss cancellation of error in greater detail in Section 2.4.2. Taken together, these results demonstrate that by increasing the number of atoms in the WF region the calculations converge to within 1 kcal/mol of the result obtained from the full WF calculation, regardless of the size of the basis.

We then tested how increasing the number of basis functions affected convergence. From the data in Figure 2.2 basis sets larger than cc-pVDZ converge with even fewer carbon atoms in the WF region. We also tested the accuracy of embedding for an extrapolation to the complete basis set (CBS) limit using cc-pVDZ and cc-pVTZ energies within the formula of Helgaker et al.[1, 2]. To do this, we replaced the WF correlation energy of the cc-pVTZ WF-in-DFT calculations with that from the CBS extrapolation of the WF region and compared these extrapolated WF-in-DFT results to the CSB extrapolated WF energy of the full systems. The extrapolated results, shown in Figure 2.3, lie very close to the WF-in-DFT cc-pVTZ results, which indicates that the extrapolation does not change the difference between WF-in-DFT and the full system calculation significantly. The similarity between cc-pVTZ and CBS energies for this system supports the use of basis set extrapolation methods within our embedding framework.

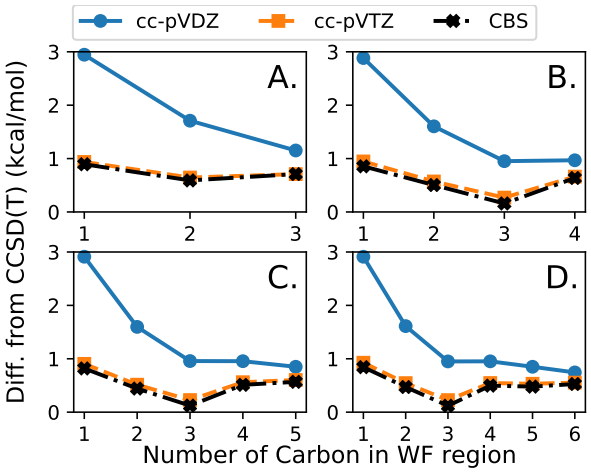


Figure 2.3: Absolute energy difference of S_N2 activation with complete basis extrapolation[1, 2]. Reactants are (A) 1-chlorobutane, (B) 1-chloropentane, (C) 1-chlorohexane, and (D) 1-chloroheptane.

WF-in-DFT energies are shown to converge to within 1 kcal/mol of the full system WF energy with both increasing size of WF subsystem, and increasing number of basis functions used to describe the system. This systematic improvability is analogous to the previously mentioned improvability of a WF calculation: there is a clear route to consistently improve the accuracy of the embedding calculation to within a 1 kcal/mol cutoff. The number of atoms in the WF subsystem may then be thought of as a convergence parameter where by increasing the number of atoms in the WF subsystem improves the accuracy of the embedding. These results demonstrate the systematic improvability of the Huzinaga embedding method.

2.4.2 Complex Subsystem Divisions

We have shown that the Huzinaga level-shift embedding method is systematically improvable and can closely recreate full system WF energies when subsystems are divided across a single covalent bond. However, in order to be broadly applicable to a variety of systems, the embedding method must be able to handle myriad interactions between subsystems. Embedding methods are often limited by how well the method can treat the interaction between subsystems. Here we demonstrate the robustness of the Huzinaga embedding method by dividing a system across two distinct covalent bonds and

subdividing across a delocalized double bond.

When subdividing a system, the resulting subsystems are defined by the atoms and electrons included in the subsystem. As such, when dividing a complex system into closed shell subsystems, there are a variety of possibilities for the electron distribution among subsystems. Practically, electrons are distributed among subsystems by specifying the number of electrons associated with each subsystem such that the total number of electrons add up to the total number of electrons in the full system. When subsystems are not covalently bound, it is reasonable to partition the subsystems such that they are charge neutral, for instance, two water molecules would have 10 electrons each. However, when partitioning across a covalent bond, the choice is less obvious. One possibility is to partition such that each fragment is closest to charge neutral. For instance, in Figure 2.4A in partition 1, one subsystem has a fluorine anion (10 electrons), chloride atom (17 electrons), carbon (6 electrons), and hydrogen (1 electron) for a total of 34 electrons and a charge of -1. Therefore, one option would be to include 34 electrons; however, this would lead to "dangling bonds" as the electrons in the bonds between subsystems are split between the WF and DFT subsystems. Thus, the other option would be to include 36 electrons and a charge of -3 such that the bonding electrons are included into the WF subsystem. We refer to this as "full bonds" partitioning. We reiterate for the results presented in Figure 2.3, to create closed shell subsystems we were forced to subdivide using "full bonds."

We tested these two options by expanding upon our previous S_N2 results, by studying the activation energy of an S_N2 reaction centered on a secondary carbon. When embedding this reaction in a WF subsystem, the system was subdivided following the scheme outlined in Figure 2.4A. The chosen subdivisions allow for direct comparisons to the previous primary S_N2 reaction results, while also demonstrating how partitioning across multiple bonds affect overall embedding energies. The WF method and DFT exchange-correlation functional were the same as the primary carbon S_N2 reaction embedding: CCSD(T) and M06[3], respectively. The basis used to generate the data in Figure 2.4B was aug-cc-pVDZ.

Figure 2.4 shows the results of various subsystem charging strategies. The "One covalent bond full bond" results (green dotted line) are those of the hexane S_N2 system in the previous section (Figure 2.3) with an aug-cc-pVDZ basis. The partitioning

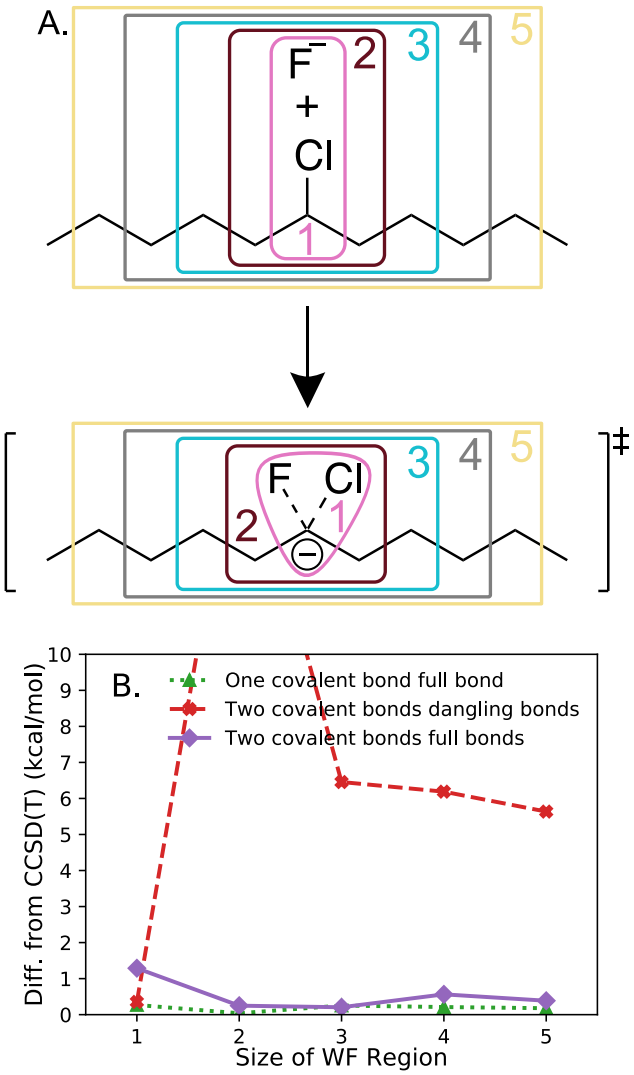


Figure 2.4: Subfigure (A) 6-chloroundecane reacting to form S_N2 transition state. Numbering indicates the size of the WF subsystem, where numbers identify the number of carbon away from the reaction in one direction along the chain. This numbering is to provide an analogous measurement to primary carbon S_N2 shown in Figure 2.1. Subfigure (B) S_N2 activation energy of primary and secondary carbons. Secondary carbon S_N2 reaction energies are shown where the subsystems are charged to include the electrons in the bond between subsystems within the WF region, and to not include those electrons in the WF subsystem. WF subsystem sizes for partitioning across a single bond are shown in Figure 2.1.

across two single covalent bonds results both follow the subdivision scheme outlined in Figure 2.4A, but the "Two covalent bonds dangling bonds" results (red dashes) use a neutral charge DFT subsystem and a -1 charge WF subsystem, while the "Two covalent bonds full bonds" results (purple line) show embedding when the WF subsystem has a -3 charge (1 electron added from each dangling bond) and the DFT subsystem has a +2 charge. From the data in Figure 2.4B, it is clear that some subsystem charging is necessary to maintain systematic improvability and significant similarity to full system WF results. We note, this strategy is similar to adding capping hydrogen atoms in a QM/MM calculation to prevent dangling bonds[121]. Once properly charged, subdividing a system across multiple covalent bonds gives similar results to subdividing a system across a single covalent bond, indicating that the accuracy of embedding energy is independent of number of bonds which connect the subsystems. After reviewing the electron density plots of the different subsystem charging strategies (Figure 2.5), we can see that including the electrons in the bond between subsystems in the WF subsystem localizes that bond within the WF subsystem.

The importance of including the electrons in the bond between the subsystems within the WF region may also be seen with the accuracy of the augmented basis sets. For most of the systems studied, augmented basis sets require the fewest atoms in the WF region to achieve energies below 1 kcal/mol of the full system WF calculation. We attribute this to the augmented basis functions accommodating the electrons in the bond between the subsystems and having additional flexibility to recreate these bond orbitals. From these results it would seem as though the more freedom the electron density has to recreate the KS-DFT electron density, the better the reaction energies. A natural conclusion would be to use the supersystem basis, rather than the absolutely localized basis for improved accuracy in WF-in-DFT reaction energies. In a previous paper, however, our group demonstrated higher accuracy for absolutely localized WF-in-DFT reaction energies than for supersystem basis WF-in-DFT reaction energies.[47] We attribute this improved accuracy to systematic error cancellation. When comparing the DFT-in-DFT energy of subdividing a system across multiple covalent bonds system to the fully relaxed KS-DFT energy, the role of error cancellation seems to be an important factor (Tables S1 and S2). For the undecane system with subsystem sizes 4 and 5, the absolute DFT-in-DFT energy of the dangling bonds partitioning is closer to the absolute

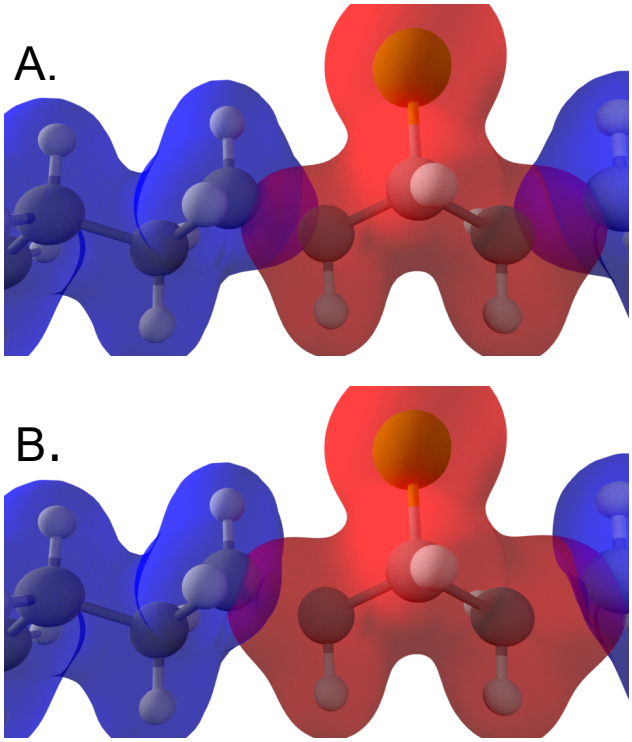


Figure 2.5: 6-chloroundecane system electron density plot with WF subsystem size 2. Subsystem electron densities for WF and DFT regions are shown in red and blue respectively. For neutral subsystems, subfigure (A), electrons comprising the bond are shared between subsystems. However, when charging the WF region with -2 and the DFT region with +2 the bond is almost entirely described by electrons in the WF region, subfigure (B).

KS-DFT than the full bond partitioning of the same sizes. However, for those system sizes the full bond partitioning DFT-in-DFT reaction energies are closer to the KS-DFT reaction energies than the dangling bond partitioning. While the DFT-in-DFT electron density in the aforementioned dangling bond partitioning is closer to the KS-DFT electron density, the systematic error in the electron density of the full bond partitioning enables cancellation of errors and better overall reaction energies.

This suggests that *error cancellation*, and not the ability to most accurately reproduce KS-DFT, dominates the accuracy of WF-in-DFT energy differences. In the extreme limit where the electrons treated at the WF level of theory are entirely different between products and reactants (for instance, core electrons in the product and valence electrons in the reactants), one would not expect accurate WF-in-DFT energy differences. Therefore, the opposite limit, where the electrons treated at the WF level of theory are the most similar between products and reactants is likely to produce the most accurate WF-in-DFT energy differences. This is precisely what absolute localization forces, the electrons are forced to localize on the atoms associated with the subsystem. Our results for the full bond partitioning are thus suggestive that the electrons in the bonding orbital between subsystems localize similarly between products and reactants, and this leads to better error cancellation and additional accuracy in WF-in-DFT energy differences. Thus we hypothesize there is an important balance between including enough of the system to encompass the region of interest, demonstrated by the convergence with size of subsystem, and maintaining good error cancellation, by consistently localizing the electron density in the WF region. Work in our group is ongoing to further elucidate the role of error cancellation for accurate WF-in-DFT reaction energies.

To further demonstrate the robustness of our embedding method, Huzinaga WF-in-DFT energies were calculated for a fluorine elimination reaction. This molecule, shown in Figure 2.6A, has a delocalized conjugated π system spanning the molecule. Reaction energies were calculated using CCSD(T) for WF subsystem, and M06[3] exchange-correlation functional for the DFT subsystem. The subsystems were charged to include the electrons in the bonds between subsystems entirely within the WF region which entailed a -2 charge when partitioning across a double bond and -1 charge when partitioning across a single bond. The results of embedding across a delocalized system (Figure 2.6), are largely similar to those of the single covalent bond partitioning: the

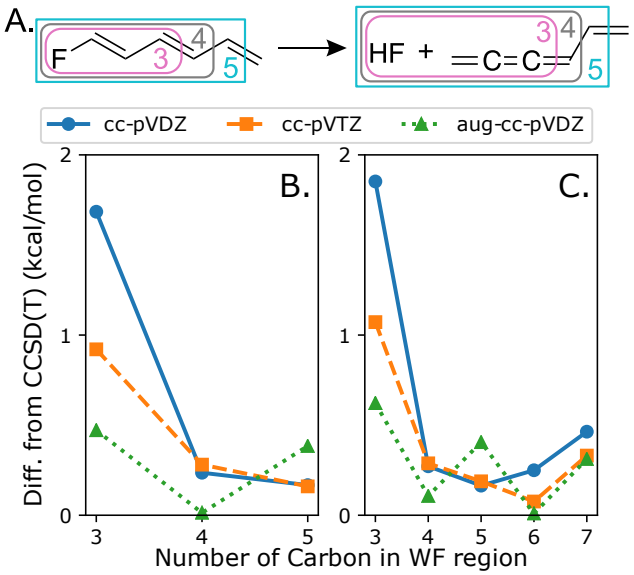


Figure 2.6: (A) Flourine elimination reaction. WF subsystem sizes are specified following a similar scheme to Figure 2.1. Here WF subsystem starts at 3, because 3 carbon undergo bonding changes during the reaction. Subfigures (B) and (C) show absolute energy difference of WF-in-DFT embedding from full system CCSD(T) for flourine elimination reaction. Reactants are (B) (1Z,3E)-1-fluorohexa-1,3,5-triene; and (C) (1Z,3E,5E)-1-fluoroocta-1,3,5,7-tetraene.

embedding method is systematically improvable with respect to the size of WF subsystem and number of basis functions, in addition to converging below 1 kcal/mol of the full system WF energy. Therefore, the Huzinaga WF-in-DFT Embedding scheme can handle delocalized π-bonding networks with the same accuracy as localized covalently bonded systems.

Complicated systems — for example, systems that need to divide across many bonds or across delocalized orbitals — have the potential to introduce additional errors in embedding methods. However, we have shown that the Huzinaga level-shift projection operator still performs as well for these more complicated systems as it does for simpler single covalent bond partitioned systems, further demonstrating the robustness of the method. We have shown that as long as the subsystems are divided such that the electrons in the bonds between subsystems are all included in the WF subsystem, then the results are accurate independent of the number of bonds or if the bond is a single or double bond. In all of these cases, the most important factor for obtaining accurate embedded WF energies is the size of the WF subsystem.

2.4.3 Multireference Embedding

Some of the most challenging systems for quantum chemistry are those with multireference character. Systems with degenerate orbitals or partially occupied states often require a multiconfigurational WF method to accurately describe. Multiconfigurational WF methods typically scale poorly with size of the system, oftentimes scaling exponentially with the number of electrons and basis functions. However, through absolutely localized Huzinaga WF-in-DFT embedding the WF subsystem includes only a fraction of the total system and therefore a fraction of the total electrons and basis functions. By localizing the multireference calculation to a subsystem, we demonstrate multireference energies for systems near the size limit of current non-embedded multireference methods with a greatly reduced computational cost. μ-projection based embedding has been successfully applied to large multireference systems[37, 122], here we demonstrate the accuracy and applicability of the Huzinaga projection operator.

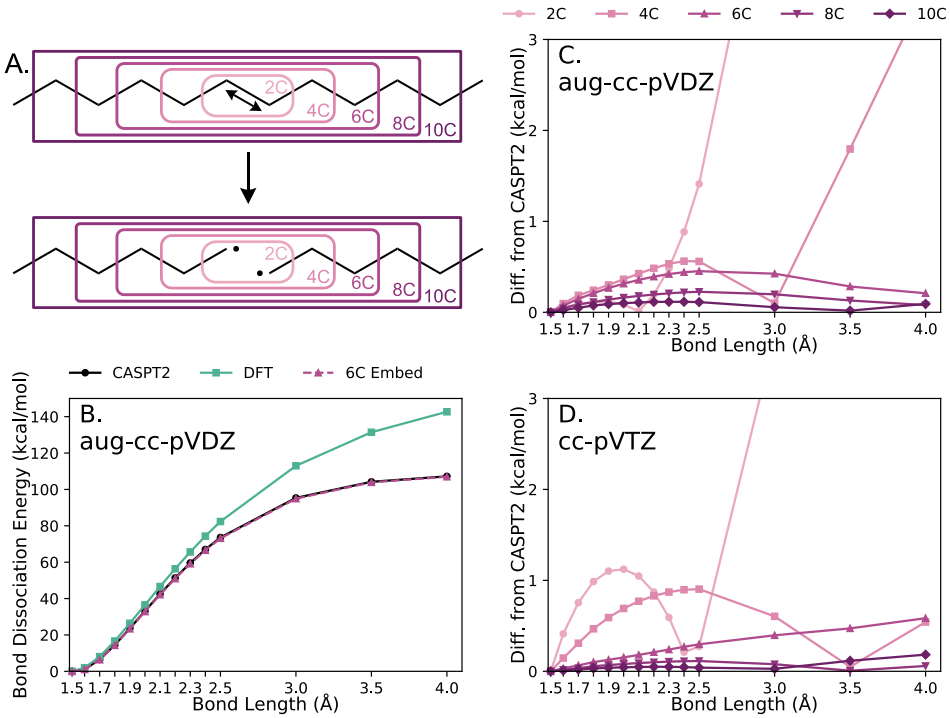


Figure 2.7: Absolute energy difference for the homolytic bond dissociation curve calculated using CASPT2 embedded in M06[3] DFT exchange-correlation functional. Subfigures show: (A) WF subsystem division; (B) Bond dissociation energy comparison between single reference method (DFT) and embedding (6C Embed) using aug-cc-pVDZ basis; (C) results of embedding with aug-cc-pVDZ; and (D) cc-pVTZ basis.

Homolytic Bond Cleavage

A relatively simple multireference reaction involves the homolytic bond cleavage of a carbon-carbon bond. As the bond elongates, the doubly occupied bonding orbital becomes two degenerate radical orbitals. Thus this system provides a good benchmark for determining how well the absolutely localized Huzinaga embedding method can embed a multireference WF method and similar systems have been used previously to benchmark embedded multireference methodologies.[123] Using the subsystem charging scheme from the previous section, we embedded a CASPT2 subsystem within a DFT potential utilizing the M06[3] functional (Figure 2.7). The results demonstrate similar desirable convergence behavior as all previous calculations: increasing the number of carbon in the WF region improves the energies until convergence. Additionally, relatively few carbon are necessary in the WF region in order to calculate full system WF level results. For those systems with fewer than 6 carbon in the WF region, results significantly differ from multireference energies only in regions of high multireference character.

Double Bond Rotation

Another potentially challenging multireference problem is the rotation of a system around a double bond, breaking the pi bond in the process to form a diradical (Figure 2.8A). We applied our Huzinaga CASPT2-in-M06 embedding procedure to this system with the charging scheme identified in the previous section. Our results (Figure 2.8) illustrate convergence with the size of the WF system, in addition to convergence to the full WF result with only 6 carbon in the WF region. Those systems with less than 6 carbon in the WF region only significantly deviate from full system results only at those regions of high multireference character, similar to the bond cleavage results. These results demonstrate the robustness of the method with respect to multireference WF calculations.

Fe-MOF-74 H_2 Adsorption

One of the most important contributions of this method is the ability to calculate WF energies of incredibly large, complex systems. Given that our method can calculate

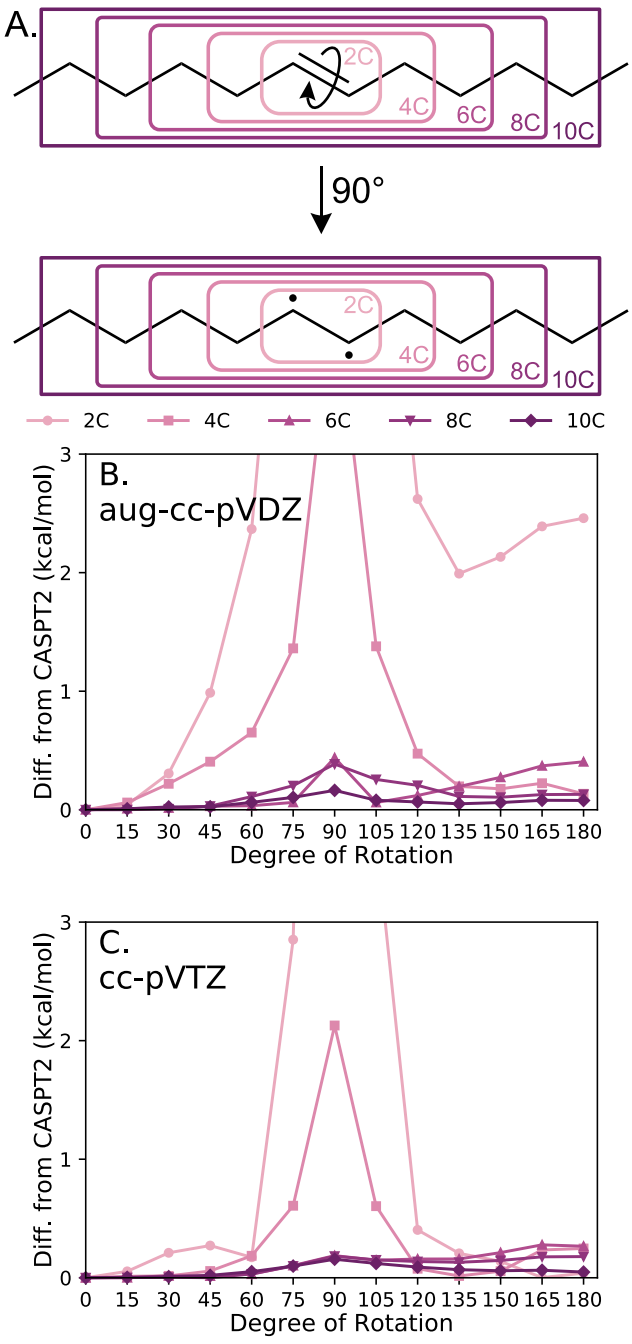


Figure 2.8: Absolute energy difference of rotation about a double bond calculated using CASPT2 embedded in M06[3] DFT exchange-correlation functional. At 90 rotation, pi bond is entirely broken and forms diradical. Sub-figures show: (A) WF subsystem division; (B) results of embedding with aug-cc-pVDZ; and (C) cc-pVTZ basis.

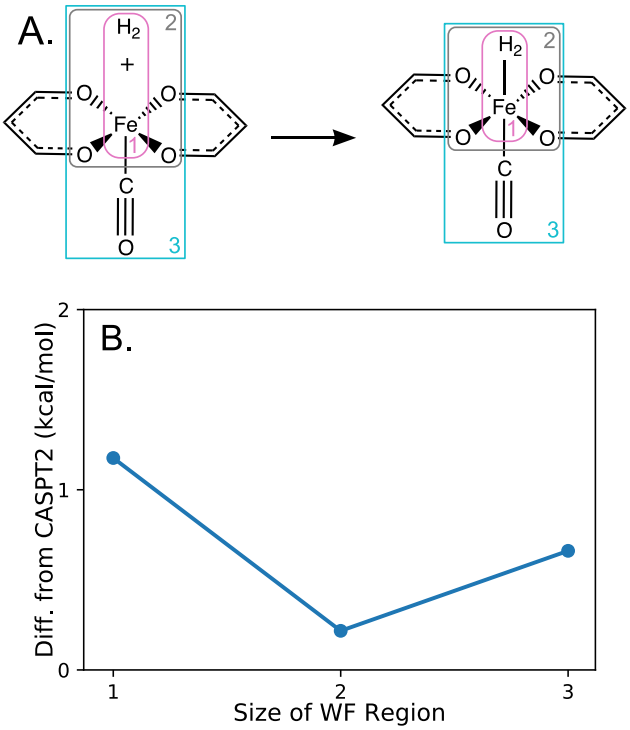


Figure 2.9: Subfigure (A) shows hydrogen gas adsorption on Fe-MOF-74. WF subsystems are specified with the smallest containing just the Fe and hydrogen adsorbant. Subfigure (B) shows absolute energy difference of hydrogen adsorption calculated using CASPT2 embedded in M06[3] DFT exchange-correlation functional and cc-pVDZ basis set.

multireference WF level energies and – with the correct subsystem charging – divide a system across complex interactions, we calculated the adsorption energy of hydrogen on a model cluster of Fe-MOF-74 for the singlet spin state. This small model has been used previously to represent the reaction center while remaining small enough to calculate full WF energies.[110] Initial CCSD(T) calculations on the model system indicated that a multireference WF method was necessary to adequately describe the adsorption interaction.

As with the previous multireference calculations, we embedded CASPT2 within DFT using the M06[3] functional. Since the hydrogen adsorption is localized to the Fe metal center, we subdivided embedding systems as illustrated in Figure 2.9 and limited the CAS active space to 6 electrons in 5 orbitals for the bare MOF model and 8 electrons

in 7 orbitals for the model with hydrogen bound. The active orbitals consist of the Fe 3d and bound hydrogen 1s orbitals. Figure 2.9A shows similar convergence seen with other systems studied in addition to very close agreement to full system CASPT2 results with only the metal center and hydrogen within the WF subsystem (subsystem 1). This embedding scheme also serves to reinforce our previously established rule of thumb for charging systems. For this MOF system, subsystem 1 has a +2 charge in the WF region, corresponding with the oxidation state of the Fe, and a -2 charge of the DFT region to maintain neutrality. Because there are no covalent bonds partitioned into subsystem 1, only ionic interactions, there is no need to move the electrons from a shared bond space. Subsystems 2 and 3 have a -6 charge of the WF region and a +6 charge of the DFT region since subsystem division occurs across four conjugated carbon-oxygen covalent bonds. Given our success when applying the method to this system, our embedding method is applicable to much larger, beyond current WF level calculations. We are actively applying our method to such systems.

2.5 Conclusions

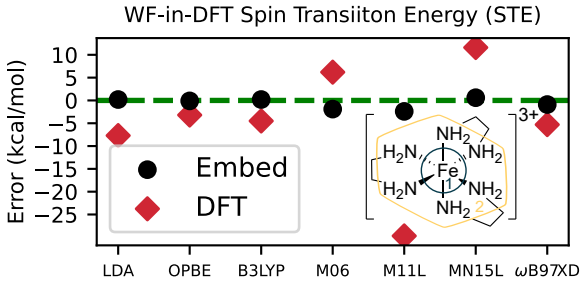
The absolutely localized Huzinaga level-shift projection operator method of DFT embedding is an efficient, robust, and systematically improvable embedding technique. Across a diverse set of test systems, the Huzinaga embedding scheme consistently approached the full system WF calculation energy; replicating the energy within 1 kcal/mol for most systems with a fraction of the full system basis functions. Additionally across all systems, as the size of the WF region increased, the embedded energy approached the full system WF energy, demonstrating systematic improvability. We also argue for the importance of balancing the size and flexibility of the WF subsystem with localized electron density leading to beneficial error cancellation. By performing the WF calculation using only the subsystem basis functions, far fewer computational resources were needed to calculate energies of complex systems. As a result, accurate single reference and multireference WF energies may be calculated on systems that previously were too large for all but DFT methods, as demonstrated with the Fe-MOF-74 cluster model.

Included in the Supporting Information are all output files for the work presented here. This data set is also available at the Data Repository hosted at the University

of Minnesota.[124] These output files also contain the input files used to generate the output. Our QSoME code[118] is open-source and requires an interface to the open-source PySCF[111] program with an optional interface to Molpro[112, 113, 114, 115, 116]. The input is very simple: definition of the subsytems, the charge associated with the subsystem, followed by the standard requirements for DFT and WF calculations (such as basis set, WF method, and exchange-correlation functional). The combination of open-source code and simple input will allow researchers to apply this methodology to a wide range of applications and obtain high accuracy results at a significantly reduced computational cost.

Chapter 3

Huzinaga projection embedding for efficient and accurate energies of systems with localized spin-densities



Reprinted from the article by Daniel S. Graham, Xuelan Wen, Dhabih V. Chulhai and Jason D. Goodpaster, *J. Chem. Phys.* 156, 054112(2022) with the permission of AIP Publishing.

3.1 Introduction

The accurate simulation of systems with complex electronic character poses a significant challenge to modern computational methods. While Kohn-Sham density functional theory[54, 4] (KS-DFT) can provide accurate, computationally efficient results for many systems, current exchange-correlation functionals have significant limitations. It has been well documented that for reactions involving a transition metal[125, 126, 127, 10] and spin-state energetics of transition metal centers[11, 12, 13, 14, 15, 16, 17], the accuracy of KS-DFT is largely functional dependent with no obvious choice of which functional to use. Systems with degenerate and near-degenerate states, such as those with stretched covalent bonds and transition metals, are notoriously difficult to simulate accurately by modern KS-DFT methods.[6, 7, 8, 9, 10]

When studying systems with transition metals, correlated wave function (WF) methods such as coupled cluster (CC), and complete active space (CAS) are commonly considered more accurate. Unlike DFT, WF methods can be systematically improved (e.g. including additional excitations for CC calculations, or increasing the size of the CAS active space) to increase the accuracy of the calculation. However, WF calculations are often limited to a small model of the system of interest due to the high computational cost. This can be a successful strategy because of the localized nature of many reactions, but new errors may be introduced by dividing the full system into a smaller model and there is evidence that the process of choosing a model is not always a clear, systematically improvable process.[104]

Quantum embedding calculations seek to improve upon the small model simulations by including some influence of the full system on the final energy. Quantum embedding methods such as QM/MM[60], ONIOM[61], DMET[62, 18], embedded mean-field theory[63, 64, 65, 66, 67], Green's function embedding[68, 69, 18, 128], partition DFT,[70, 71, 72] and DFT embedding[21, 18, 22, 23, 24] among many others [73, 74, 75, 76, 77, 129] were designed to combine the benefits of high accuracy and systematic improvability from WF theory for a small subsystem, while including effects from the full system at a comparably negligible computational cost. Projection operator based DFT embedding has been developed by many groups with significant success.[36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 25] Recent work in our group has

focused on Huzinaga level-shift projection operator DFT embedding in the absolutely localized basis. We have demonstrated the efficiency, accuracy, and systematic improvability of this method for closed-shell molecular ground[47, 130] and excited states[104], and ground state properties of periodic systems[48]. We have found that the absolutely localized basis not only reduces the size of the valence orbital space thus decreasing the computational cost of the WF calculation, it also improves the total accuracy of the reaction energy due to favorable error cancellation.[130]

Given our success with the Huzinaga based DFT embedding for closed-shell systems, we have expanded the method to allow restricted and unrestricted ground state embedded energies of molecular systems. We demonstrate highly accurate embedding energies for radical reactions, gas adsorption onto transition metals, and transition metal spin-state energetics. We also discuss how transition metal energies using the Huzinaga embedding method are far less functional dependent than traditional KS-DFT calculations.

3.2 Theory

We have developed unrestricted and restricted open-shell Huzinaga level-shift projection based embedding. Both methods are extensions of the closed-shell method described previously. [130] In the sections that follow the full system is divided into two subsystems for simplicity, however the total system may be divided into any number of subsystems.

3.2.1 Unrestricted

For unrestricted embedding, the total system electron density matrix is divided into α and β spin components, $\gamma_\alpha, \gamma_\beta$ and those components are then subdivided into subsystem A, γ^{A}, and subsystem B, γ^{B} where,

$$\gamma_\alpha = \gamma_\alpha^{\mathrm{A}} + \gamma_\alpha^{\mathrm{B}}, \tag{3.1}$$

$$\gamma^{\mathrm{A}} = \gamma_\alpha^{\mathrm{A}} + \gamma_\beta^{\mathrm{A}}. \tag{3.2}$$

The definition of γ_β is analogous to equation 1 and γ^{B} is analogous to equation 2. These subsystem spin electron density matrices are defined in the absolutely localized basis.

The absolutely localized basis defines the Kohn-Sham orbitals in a subsystem using only those basis functions centered on atoms within that subsystem. In the absolutely localized basis, the α spin Kohn-Sham orbitals of subsystem A are defined as,

$$\phi_{i\alpha}^{\mathrm{A}} = \sum_{\mu} C_{i\mu\alpha}^{\mathrm{A}} \chi_{\mu}^{\mathrm{A}}, \tag{3.3}$$

where $\phi_{i\alpha}^{\mathrm{A}}$ are the α spin Kohn-Sham orbitals of subsystem A, $C_{i\mu\alpha}^{\mathrm{A}}$ are the spin orbital coefficients of subsystem A, and χ_{μ}^{A} are basis functions associated with subsystem A. This basis is also used for all embedded WF calculations in this paper.

In order to maintain orthogonality between subsystems, the unrestricted spin Huzinaga level-shift projection operator is added to the subsystem Hamiltonian. The unrestricted spin Huzinaga level-shift projection operator may be added to a subsystem Hamiltonian, allowing simple incorporation with any existing QM method utilizing an electron potential Hamiltonian. The form of the spin Huzinaga level-shift projection operator for subsystem A in the presence of subsystem B is,

$$\mathbf{P}_{\alpha}^{\mathrm{B}} = -\left(\mathbf{F}_{\alpha}^{\mathrm{AB}} \gamma_{\alpha}^{\mathrm{B}} \mathbf{S}^{\mathrm{BA}} + \mathbf{S}^{\mathrm{AB}} \gamma_{\alpha}^{\mathrm{B}} \mathbf{F}_{\alpha}^{\mathrm{BA}}\right) \tag{3.4}$$

where,

$$\mathbf{F}_{\alpha}^{\mathrm{AB}} = \langle \chi^{\mathrm{A}} | \hat{F}_{\alpha} | \chi^{\mathrm{B}} \rangle \tag{3.5}$$

with $\hat{F}_{\alpha}$ being the full system α spin Fock operators and,

$$\mathbf{S}^{\mathrm{AB}} = \langle \chi^{\mathrm{A}} | \chi^{\mathrm{B}} \rangle. \tag{3.6}$$

Including the unrestricted spin Huzinaga level-shift projection operator into the spin Fock matrix of subsystem A embedded into subsystem B results in the projected α spin Fock matrix,

$$\mathbf{f}_{\alpha}^{\text{A-in-B}} = \mathbf{h}_{\alpha}^{\text{A-in-B}}[\gamma_{\alpha}^{\mathrm{A}}, \gamma_{\beta}^{\mathrm{A}}, \gamma_{\alpha}^{\mathrm{B}}, \gamma_{\beta}^{\mathrm{B}}] + \mathbf{J}[\gamma^{\mathrm{A}}] + \mathbf{v}_{\mathrm{xc}}[\gamma_{\alpha}^{\mathrm{A}}, \gamma_{\beta}^{\mathrm{A}}] \tag{3.7}$$

where $\mathbf{J}$ is the electron Coulomb potential, $\mathbf{v}_{\text{xc}}$ is the exchange-correlation (XC) potential, and the embedded core Hamiltonian is

$$\begin{aligned}\mathbf{h}_{\alpha}^{\text{A-in-B}}[\gamma_{\alpha}^{\text{A}},\gamma_{\beta}^{\text{A}},\gamma_{\alpha}^{\text{B}},\gamma_{\beta}^{\text{B}}] = \mathbf{h} + \mathbf{J}[\gamma^{\text{A}}+\gamma^{\text{B}}] - \mathbf{J}[\gamma^{\text{A}}] + \mathbf{v}_{\text{xc}}[\gamma_{\alpha},\gamma_{\beta}] \\ -\mathbf{v}_{\text{xc}}[\gamma_{\alpha}^{\text{A}},\gamma_{\beta}^{\text{A}}] + \mathbf{P}_{\alpha}^{\text{B}}[\gamma_{\alpha}^{\text{A}},\gamma_{\beta}^{\text{A}},\gamma_{\alpha}^{\text{B}},\gamma_{\beta}^{\text{B}}],\end{aligned} \tag{3.8}$$

where $\mathbf{h}$ is the total one-electron Hamiltonian.

The embedded DFT energy of subsystem A, $E_{\text{DFT}}^{\text{A}}$, is calculated by,

$$\begin{aligned}E_{\text{DFT}}^{\text{A}} = \text{Tr}\left(\gamma_{\alpha}^{\text{A}}\cdot\mathbf{h}_{\alpha}^{\text{A-in-B}}[\gamma_{\alpha}^{\text{A}},\gamma_{\beta}^{\text{A}},\gamma_{\alpha}^{\text{B}},\gamma_{\beta}^{\text{B}}]\right) \\ +\text{Tr}\left(\gamma_{\beta}^{\text{A}}\cdot\mathbf{h}_{\beta}^{\text{A-in-B}}[\gamma_{\alpha}^{\text{A}},\gamma_{\beta}^{\text{A}},\gamma_{\alpha}^{\text{B}},\gamma_{\beta}^{\text{B}}]\right) \\ +\mathbf{J}[\gamma^{\text{A}}] + E_{\text{xc}}[\gamma_{\alpha}^{\text{A}},\gamma_{\beta}^{\text{A}}],\end{aligned} \tag{3.9}$$

and the WF energy of subsystem A embedded in the DFT potential of the full system, E_{WF}^{A}, is

$$E_{\text{WF}}^{\text{A}} = \langle\Psi_{\alpha}^{\text{A}}|\hat{H}_{\alpha}^{\text{A-in-B}}|\Psi_{\alpha}^{\text{A}}\rangle + \langle\Psi_{\beta}^{\text{A}}|\hat{H}_{\beta}^{\text{A-in-B}}|\Psi_{\beta}^{\text{A}}\rangle. \tag{3.10}$$

Here $\hat{H}_{\alpha}^{\text{A-in-B}}$ is

$$\hat{H}_{\alpha}^{\text{A-in-B}} = \hat{h}_{\alpha}^{\text{A-in-B}} + \hat{g}_{\alpha}^{\text{A}} \tag{3.11}$$

where $\hat{h}_{\alpha}^{\text{A-in-B}}$ is equivalent to equation 3.8 (now written in operator form) and $\hat{g}_{\alpha}^{\text{A}}$ is the α two-electron operator for a given WF theory acting on the electrons of subsystem A. Therefore the overall WF-in-DFT embedding energy is,

$$E_{\text{WF-in-DFT}}^{\text{Full}} = E_{\text{KS-DFT}}^{\text{Full}} - E_{\text{DFT}}^{\text{A}} + E_{\text{WF}}^{\text{A}}, \tag{3.12}$$

and uses an subtractive embedding framework like ONIOM[61], where $E_{\text{WF-in-DFT}}^{\text{Full}}$ is the total WF-in-DFT energy, $E_{\text{KS-DFT}}^{\text{Full}}$ is the canonical KS-DFT of the full system, $E_{\text{DFT}}^{\text{A}}$ is the DFT energy of subsystem A embedded in the full system.

3.2.2 Restricted Open-shell

For restricted open-shell embedding, the total system electron density matrix may be divided as in equations 3.1 and 3.2. The same absolutely localized basis is also employed,

however defined in the restricted sense as,

$$\phi_i^{\mathrm{A}} = \sum_{\mu} C_{i\mu}^{\mathrm{A}} \chi_{\mu}^{\mathrm{A}}, \tag{3.13}$$

where ϕ_i^{A} are the restricted open-shell Kohn-Sham orbitals of subsystem A, $C_{i\mu}^{\mathrm{A}}$ are the restricted open-shell orbital coefficients of subsystem A, and χ_{μ}^{A} are basis functions associated with subsystem A. The restricted open-shell Huzinaga level-shift projection operator is defined using the restricted open-shell Fock matrix, $\mathbf{F}_{\mathrm{RO}}$,

$$\mathbf{P}^{\mathrm{B}} = -\frac{1}{2}\left(\mathbf{F}_{\mathrm{RO}}^{\mathrm{AB}} \gamma^{\mathrm{B}} \mathbf{S}^{\mathrm{BA}} + \mathbf{S}^{\mathrm{AB}} \gamma^{\mathrm{B}} \mathbf{F}_{\mathrm{RO}}^{\mathrm{BA}}\right). \tag{3.14}$$

Here, the restricted open-shell Fock matrix is,

$$\mathbf{F}_{\mathrm{RO}} = \begin{array}{c} \\ \left(\begin{array}{ccc} \mathbf{F}_{cc}^{\mathrm{cs}} & \mathbf{F}_{co}^{\beta} & \mathbf{F}_{cv}^{\mathrm{cs}} \\ \mathbf{F}_{oc}^{\beta} & \mathbf{F}_{oo}^{\mathrm{cs}} & \mathbf{F}_{ov}^{\alpha} \\ \mathbf{F}_{vc}^{\mathrm{cs}} & \mathbf{F}_{vo}^{\alpha} & \mathbf{F}_{vv}^{\mathrm{cs}} \end{array}\right) \end{array} \begin{array}{c} \text{core } (c) \\ \text{open-shell } (o) \\ \text{virtual } (v) \end{array} \tag{3.15}$$

where $\mathbf{F}^{\alpha}$ and $\mathbf{F}^{\beta}$ are the α and β spin Fock matrices respectively, and $\mathbf{F}^{\mathrm{cs}}$ is $(\mathbf{F}^{\alpha}+\mathbf{F}^{\beta})/2$. Then we arrive at $\mathbf{F}_{\mathrm{RO}}^{\mathrm{AB}}$ as,

$$\mathbf{F}_{\mathrm{RO}}^{\mathrm{AB}} = \langle \chi^{\mathrm{A}} | \hat{F}_{\mathrm{RO}} | \chi^{\mathrm{B}} \rangle. \tag{3.16}$$

The restricted open-shell embedded core Hamiltonian is,

$$\begin{aligned} \mathbf{h}_{\mathrm{RO}}^{\text{A-in-B}}[\gamma_{\alpha}^{\mathrm{A}}, \gamma_{\beta}^{\mathrm{A}}, \gamma_{\alpha}^{\mathrm{B}}, \gamma_{\beta}^{\mathrm{B}}] = \mathbf{h} + \mathbf{J}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}] - \mathbf{J}[\gamma^{\mathrm{A}}] + \mathbf{v}_{\mathrm{xc,\ RO}}[\gamma_{\alpha}, \gamma_{\beta}] \\ - \mathbf{v}_{\mathrm{xc,\ RO}}[\gamma_{\alpha}^{\mathrm{A}}, \gamma_{\beta}^{\mathrm{A}}] + \mathbf{P}^{\mathrm{B}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}], \end{aligned} \tag{3.17}$$

where $\mathbf{v}_{\mathrm{xc,\ RO}}$ is the restricted open-shell exchange-correlation potential matrix defined as

$$\mathbf{v}_{\mathrm{xc,\ RO}}[\gamma_{\alpha}, \gamma_{\beta}] = \begin{array}{c} \\ \left(\begin{array}{ccc} \mathbf{v}_{\mathrm{xc}}^{\mathrm{cs}}[\gamma_{\alpha}, \gamma_{\beta}]_{cc} & \mathbf{v}_{\mathrm{xc}}[\gamma_{\beta}]_{co} & \mathbf{v}_{\mathrm{xc}}^{\mathrm{cs}}[\gamma_{\alpha}, \gamma_{\beta}]_{cv} \\ \mathbf{v}_{\mathrm{xc}}[\gamma_{\beta}]_{oc} & \mathbf{v}_{\mathrm{xc}}^{\mathrm{cs}}[\gamma_{\alpha}, \gamma_{\beta}]_{oo} & \mathbf{v}_{\mathrm{xc}}[\gamma_{\alpha}]_{ov} \\ \mathbf{v}_{\mathrm{xc}}^{\mathrm{cs}}[\gamma_{\alpha}, \gamma_{\beta}]_{vc} & \mathbf{v}_{\mathrm{xc}}[\gamma_{\alpha}]_{vo} & \mathbf{v}_{\mathrm{xc}}^{\mathrm{cs}}[\gamma_{\alpha}, \gamma_{\beta}]_{vv} \end{array}\right) \end{array} \begin{array}{c} \text{core } (c) \\ \text{open-shell } (o) \\ \text{virtual } (v) \end{array} \tag{3.18}$$

with

$$\mathbf{v}_{\mathrm{xc}}^{\mathrm{cs}}[\gamma_\alpha, \gamma_\beta] = \frac{\mathbf{v}_{\mathrm{xc}}[\gamma_\alpha] + \mathbf{v}_{\mathrm{xc}}[\gamma_\beta]}{2}. \tag{3.19}$$

The embedded DFT energy of subsystem A, $E_{\mathrm{DFT}}^{\mathrm{A}}$, is calculated by,

$$E_{\mathrm{DFT}}^{\mathrm{A}} = \mathrm{Tr}\left(\gamma^{\mathrm{A}} \cdot \mathbf{h}_{\mathrm{RO}}^{\text{A-in-B}}[\gamma_\alpha^{\mathrm{A}}, \gamma_\alpha^{\mathrm{B}}\gamma_\beta^{\mathrm{A}}, \gamma_\beta^{\mathrm{B}}]\right) + \mathrm{J}[\gamma^{\mathrm{A}}] + E_{\mathrm{xc}}[\gamma_\alpha^{\mathrm{A}}] + E_{\mathrm{xc}}[\gamma_\beta^{\mathrm{A}}], \tag{3.20}$$

and the WF energy of subsystem A embedded in the DFT potential of the full system, $E_{\mathrm{WF}}^{\mathrm{A}}$, is

$$E_{\mathrm{WF}}^{\mathrm{A}} = \langle \Psi^{\mathrm{A}} | \hat{H}_{\mathrm{RO}}^{\text{A-in-B}} | \Psi^{\mathrm{A}} \rangle. \tag{3.21}$$

Here $\hat{H}_{\mathrm{RO}}^{\text{A-in-B}}$ is

$$\hat{H}_{\mathrm{RO}}^{\text{A-in-B}} = \hat{h}_{\mathrm{RO}}^{\text{A-in-B}} + \hat{g}_{\mathrm{RO}}^{\mathrm{A}} \tag{3.22}$$

where $\hat{h}_{\mathrm{RO}}^{\text{A-in-B}}$ is equivalent to equation 3.17 (now written in operator form) and $\hat{g}_{\mathrm{RO}}^{\mathrm{A}}$ is the restricted open-shell two-electron operator for a given restricted open-shell WF theory acting on the electrons of subsystem A. The total WF-in-DFT embedding is the same as for unrestricted embedding, equation 3.12. We note that for both unrestricted and restricted open-shell embedding, if the full system basis is employed the exact KS-DFT energy and density would be obtained from the DFT-in-DFT embedding procedure as this formulation is formally exact.

3.3 Computational Details

The thiol-ene reaction molecular geometries were optimized using Gaussian 16[108] using the M06 functional[3] and aug-cc-pVDZ basis[109]. The Fe-MOF-74 cluster geometry was calculated using Gaussian 16[108] following the procedure outlined by Lee and coworkers.[110] Fe spin transition cluster geometries originate from a recent study by Radoń.[131] Furthermore, all geometries are reported in the Supporting Information (SI). Single point DFT and UCCSD(T) energy calculations were performed using PySCF version 1.7.4[111] and all ROCCSD(T) and CASPT2 calculations were done using Molpro 2019.2[112, 113, 114, 115, 116, 132, 133]. WF-in-DFT embedding energies were calculated using our Quantum Solid-state and Molecular Embedding code[134] (QSoME) via the same freeze-and-thaw scheme outlined in our previous paper. [130] All

relevant output files including full system energies are included in the SI. Interestingly, we have found that embedding two charged subsystems results in higher accuracy than embedding two open-shell subsystems. This result supports our previous findings and further details are provided in Appendix B.

3.4 Results and Discussion

With the addition of restricted open-shell and unrestricted embedding, a myriad of new chemical systems may be studied using our absolutely localized Huzinaga embedding method. Here we have chosen several open-shell systems as demonstrative examples of the broad applicability of our embedding method.

3.4.1 Radical Reactions

Systematic improvability, that is where energies improve through a well defined process, is an important feature of our embedding method. We have demonstrated previously that the method is systematically improvable for closed-shell systems[130], and here have chosen to study a simple radical thiol-ene reaction (Fig. 3.1) to demonstrate the same systematic improvability for our open-shell embedding method. The system is divided into a series of incrementally larger WF subsystems in order to understand how the energies changed as the size of the WF region grows. For this calculation, we chose CCSD(T) for the WF method and used the M06 DFT exchange-functional. We calculated the reaction energy using both restricted open-shell and unrestricted Huzinaga embedding and compared the result to the full system CCSD(T) reaction energy. As is evident from Fig. 3.2, our embedding method exhibits desirable systematic improvability with increasing WF subsystem size for restricted and unrestricted calculations. Furthermore, including only those atoms directly bonded to the reaction center is sufficient to recreate full system CCSD(T) results to within 1 kcal/mol, demonstrated by the high accuracy for WF subsystem size 2. A significant difference between CCSD(T) and embedding results for WF subsystem size 1 was expected based on our previous closed-shell embedding calculations [130]. Additionally, the difference between CCSD(T) and embedding results does not monotonically converge to exactly zero due to our use of the absolutely localized basis, however the difference does converge to well within accepted

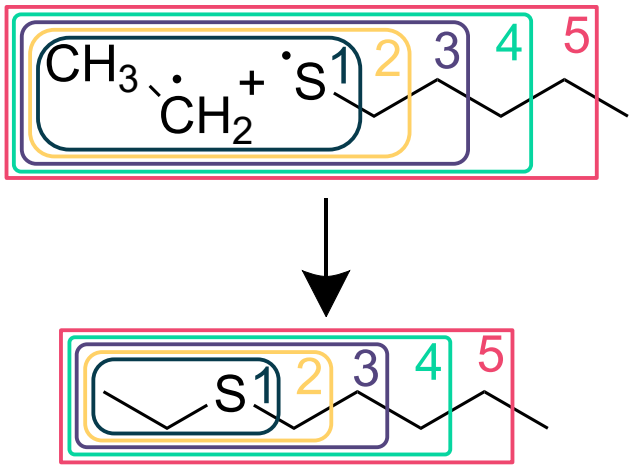


Figure 3.1: 1-(Ethylsulfyl)pentane formation by thiol-ene radical reaction. Numbering indicates the size of the WF subsystem. Implied hydrogen are included within the subsystem containing their associated bonded carbon.

chemical accuracy (1 kcal/mol). This result is expected and has been addressed in our previous paper about absolutely localized embedding[130].

3.4.2 Fe-MOF-74 Gas Adsorption

As was mentioned in the introduction, transition metal systems are notoriously difficult for DFT methods to calculate accurately. We have successfully calculated the energy of hydrogen gas adsorption on an closed shell singlet Fe-MOF-74 cluster model using our closed-shell method.[130] However, the ground state spin for this model is actually a quintet. Therefore, we now calculate the adsorption energy of the true ground electronic state. We divided the system following the scheme shown in Fig. 3.3 and embedded CASPT2 subsystems within M06 DFT. All five Fe 3d orbitals with seven electrons, and both adsorbate hydrogen 1s orbitals with two electrons were included as the active space for the CASPT2 calculation. Our results, shown in Fig. 3.4, once again demonstrate highly accurate WF results once all atoms adjacent to the reaction are included in the WF subsystem (subsystem size 3). Here for the first time we see an increase in error as the size of the WF subsystem is increased however, it is important to note that this error increase only occurs for restricted open-shell embedding. We attribute this exception to the smooth convergence of the embedding energy to WF results to the restricted

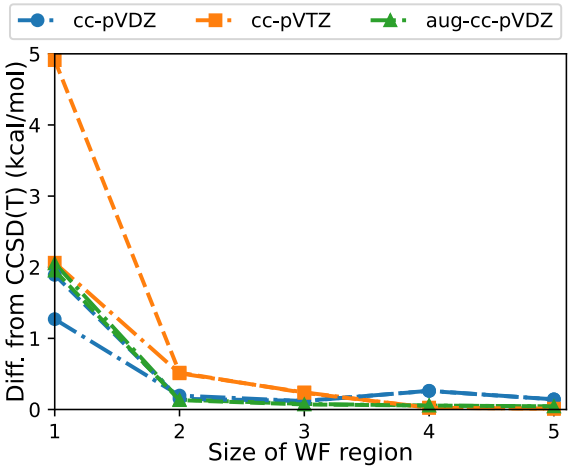


Figure 3.2: Absolute energy difference of CCSD(T) embedded in M06 for 1-(Ethylsulfyl)pentane formation by thiol-ene radical reaction compared to full system CCSD(T). Subsystem size divisions correspond with the division in Fig. 3.1. Dashed lines represent unrestricted embedding, dash dotted lines represent restricted open-shell embedding.

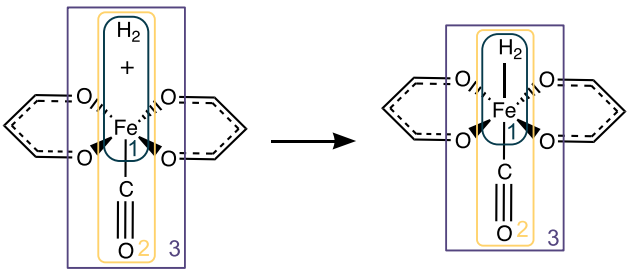


Figure 3.3: Hydrogen gas adsorption reaction on a Fe-MOF-74 cluster model. WF subsystem divisions are specified; the smallest subsystem includes only the Fe and hydrogen adsorbate.

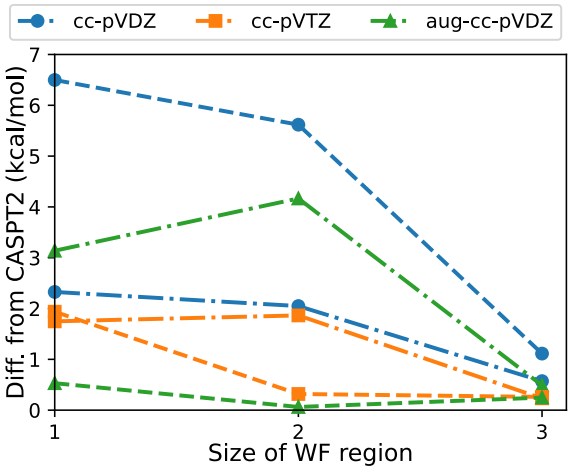


Figure 3.4: Absolute energy difference of CASPT2 embedded in M06 DFT hydrogen adsorption compared to full cluster CASPT2 results. Dashed lines represent unrestricted embedding, dash dotted lines represent restricted open-shell embedding.

open-shell orbital restriction in conjunction with the absolutely localized restriction. In our restricted open-shell embedding, the unpaired electron must be localized to the atoms in the WF subsystem while the DFT subsystem is described as a closed shell singlet because of our absolutely localized basis. This can result in increased errors due to forced localization of the spin density if the WF region is too small. We do note that there is a uniform convergence once the WF subsystem size is sufficiently large.

3.4.3 Spin Transition Energy

One particularly difficult property to accurately calculate for many systems is the spin-transition energy (STE), or vertical excitation energy. Spin-transition energies for systems with transition metals often have a significant dependency on the exchange-correlation functional.[135] The spin density is typically localized to the transition metal which makes spin-transition energy calculations an appealing target for embedding methods. To demonstrate the effectiveness of absolutely localized Huzinaga WF-in-DFT embedding, we calculated the STEs of two Fe cluster systems previously identified by Radon as having high functional dependence.[131] There are a variety of different WF methods we could chose when studying spin transition energies, such as CASPT2, CCSD(T) and multireference configuration interaction (MRCI) methods among many

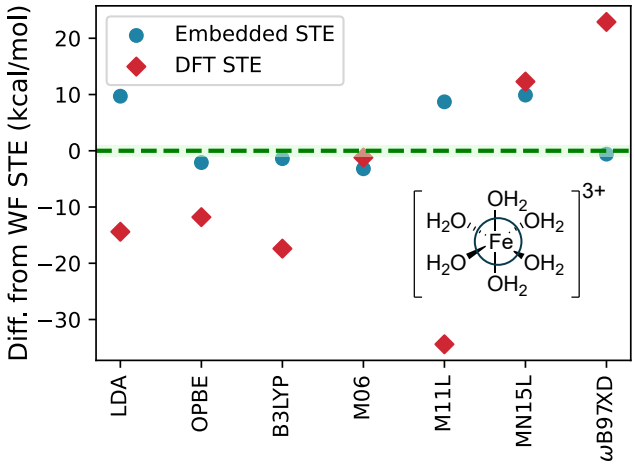


Figure 3.5: Spin-transition energy (STE) difference of embedding in a variety of DFT XC functionals compared to UCCSD(T). Green shaded region indicates 1 kcal/mol difference from full system WF results.

others. While there is debate about the accuracy of the particular WF methods for calculating the spin energetics of transition metal cluster systems[131, 136], we mean to show that our embedding method can recreate the spin transition energy of the chosen WF method on the full system. Therefore, we are less interested in which WF method is the most experimentally accurate, and instead here demonstrating that our embedding method can recreate the full system WF results. We compared UCCSD(T) calculations on the full system to embedding STEs using a variety of functionals.

For the smaller, weak field ligand system shown in Fig. 3.5, the embedded WF region is limited to only include the Fe (larger divisions would be trivial with near all or all of the electrons in the WF region). There still exists some exchange-correlation functional dependency, but embedding reduces the deviation from more than 50 kcal/mol between functionals to less than 10 kcal/mol. Additionally, for all functionals except the M06 functional embedding improves upon the DFT results. In the case of the M06 functional, we believe the slight increase in error by 2 kcal/mol is due to the fact that the M06 functional closely recreates the WF results on its own and embedding introduces minor errors for small subsystem divisions as we have seen previously. Some embedding error for small WF subsystem sizes is expected however, and has been well documented by Bensberg and Neugebauer in a recent article[25]

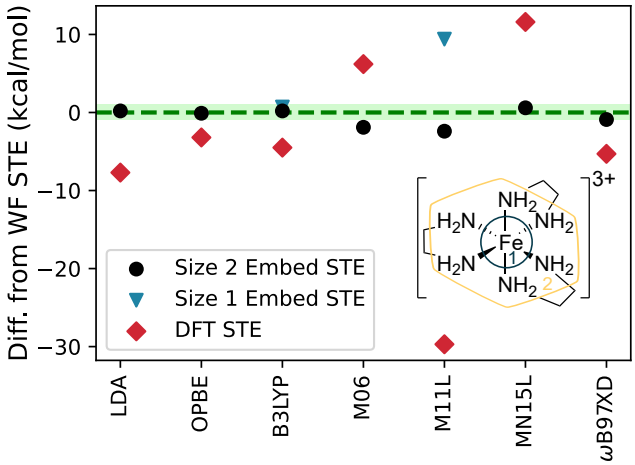


Figure 3.6: Spin-transition energy (STE) difference of embedding in a variety of DFT XC functionals compared to UCCSD(T). Green shaded region indicates 1 kcal/mol difference from full system WF results. The only subsystem size 1 embedding results presented are for B3LYP and M11L functionals as all other functionals failed to converge for subsystem size 1.

When looking at the strong field system shown in Fig. 3.6, it is clear that expanding the WF region does significantly improve the embedding results and in every case more closely recreates the UCCSD(T) results than the corresponding XC functional. This system proved difficult to converge for subsystem size 1, however following our general recommendation of including atoms adjacent to the region of chemical interest results in spin transition energies that are in close agreement with UCCSD(T) results for the full system, regardless of XC functional.

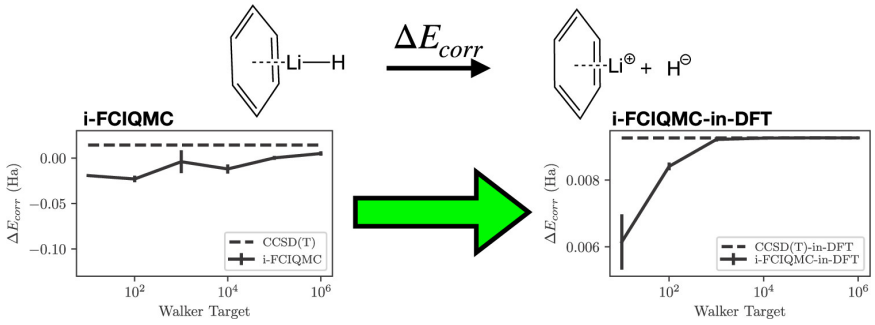
3.5 Conclusions

With this improved Huzinaga embedding method, we have demonstrated the effectiveness of the method for nearly every domain of Born-Oppenheimer localized chemical reaction: periodic or non-periodic, ground state and excited state, closed-shell or open-shell. Specifically here we have shown that our method can accurately recreate WF results for a variety of open-shell systems embedding a variety of QM methods. These WF results were achieved for a fraction of the computational cost of the full system WF calculation while retaining high accuracy. The spin transition energies are of particular

note as embedding can help break the functional dependency of calculations on transition metals. The absolutely localized Huzinaga WF-in-DFT method is an incredibly valuable tool for the chemical community and by publishing our code, user manual, and tutorial we seek to make the method as accessible as possible.

Chapter 4

Fully Quantum Embedding with Density Functional Theory for Full Configuration Interaction Quantum Monte Carlo



Reprinted with permission from the article by Hayley R. Petras, Daniel S. Graham, Sai Kumar Ramadugu, Jason D. Goodpaster, and James J. Shepherd *J. Chem. Theory Comput.* 2019, 15, 10, 5332–5342.

Author contributions: H.R.P and D.S.G. contributed equally to this paper.

4.1 Introduction

Catalysis often involves bond rearrangements at surfaces, a process featuring closely-separated energy minima, stretched bonds, and transition states. The electronic structure of these systems can become extremely complex; combined with energy differences that can be sub-millihartree, systematic study of catalytic bond rearrangements necessitates the development of new high-accuracy quantum chemistry methods. Although this is a subject of active and ongoing investigation, the high cost of wave function methods in particular prevents their widespread application.

One such method is full configuration interaction quantum Monte Carlo (FCIQMC) and its initiator adaptation ($i-$FCIQMC), which are both members of a family of particularly attractive high-accuracy electronic structure methods that seek to combine the exactness of full configuration interaction (FCI) with the increased speed achieved by quantum Monte Carlo (QMC).[137, 138] The first FCIQMC paper showed that the FCI ground-state wave function could be stochastically sampled due to the sparsity in the Hamiltonian; [137] it had previously been considered that there was no way to sample such a large vector as the exact FCI wavefunction. Since this pioneering work, many adaptions to FCIQMC and $i-$FCIQMC have been developed successfully for calculating correlation energies of a wide variety of benchmark systems. $i-$FCIQMC has already been used for a variety of applications on relatively small systems, including model systems (such as the Hubbard model[139, 140] and the uniform electron gas[141, 142]) and dimers (such as C_2[143] and Cr_2 [144]). It has also seen real applications that are more ambitious, such as iron porphyrins, which used a complete active space adaptation,[145] and fully periodic nickel oxide chains [146]. A significant amount of investigation has also been aimed at using the full scheme and initiator adaption of FCIQMC to stochastically sample reduced density matrices within the FCIQMC method.[147, 148, 149, 150, 151, 152] Altogether, FCIQMC and its adaptations seem well-poised for answering important questions about the electronic structures of complex chemical systems with high accuracy.

Unfortunately, like all of its high-accuracy cousin methods, $i-$FCIQMC is limited in its scope by its high cost: it can only treat relatively small system sizes (which here means number of electrons). Many further adaptations to $i-$FCIQMC have been

developed to allow for the application of $i-$FCIQMC to larger systems. These adaptations include a combination of complete active space self-consistent field (CASSCF) with $i-$FCIQMC [145], the semi-stochastic projector Monte Carlo method [153], model space QMC,[154, 155] heat-bath configuration interaction,[156] perturbation theory [157], stochastic multi-configurational self-consistent field theory (MCSCF) utilizing the FCIQMC methodology [150], use of a transcorrelated Hamiltonian with $i-$FCIQMC [158, 159], and combinations of the above methods, such as semistochastic heat-bath CI [160, 161, 162, 163].

These efforts are made all the more relevant because there are also varieties of FCIQMC which broaden its applicability. Density matrix QMC,[164] has been developed for temperature-dependent electronic structure. Additionally, several FCIQMC methods have been developed for use on excited states, such as changing the underlying propagator [165], or the Krylov-projected QMC method [166], utilizing a Löwdin partitioning technique[154], using a Gram–Schmidt procedure [167], and by restricting the population to the orthogonal complement of the low lying states [168]. The stochastic approach in the Slater determinant space has also been studied on the coupled cluster equations, called coupled cluster Monte Carlo. [169, 170, 171, 172] FCIQMC has also been adapted to treat the Clock Hamiltonian, to simulate the full time evolution of a quantum system. [173] A deterministic version of FCIQMC has been developed, [174] as well as a fast randomized iteration framework to essentially perform FCIQMC without walkers. [175] We also note that there are a number of methods which fall under the umbrella of selected configuration interaction (CI), where the CI is solved deterministically, which form a distinct and related family of methods.[176, 177, 178]

Quantum embedding methods were specifically developed to reduce the problem of scaling present in high level-methods such as $i-$FCIQMC. Embedding methods limit high-level calculations to a small subsystem that is embedded in the potential arising from the rest of the system, reducing the overall computational cost. When highly accurate embedding potentials are used, good accuracy can be achieved even when a subsystem is limited to a few atoms; therefore, embedding methodologies have been successfully applied to a wide variety of systems. [179, 180, 181, 182, 183, 184, 95, 96, 43, 64, 63, 65, 66, 67, 94, 74] Additionally, a large amount of work has been performed developing accurate embedding frameworks including quantum mechanics

/ molecular mechanics (QM/MM),[60] ONIOM,[61] density matrix embedding theory (DMET),[62] Green's function embedding,[68, 69] and density functional theory (DFT) embedding.[21, 78, 79, 80, 81, 34, 28] A recent review has considered the comparisons between DMET, Green's function embedding, and DFT embedding and we direct the interested reader to Ref.[18]. Many wave function methods such as density-matrix renormalization group (DMRG),[73] coupled cluster singles and doubles with perturbative triples (CCSD(T)),[36] second order Møller-Plesset perturbation theory (MP2),[43] and multireference configuration interaction (MRCI)[40] have been embedded as the high-level theory; this work presents the first use of $i-$FCIQMC embedding.

The quantum embedding for this work was done using projection-based embedding,[36] which is DFT embedding method. Projection-based embedding is one solution to the non-additive kinetic energy problem of DFT embedding.[39, 50] The initial projection operator applied to this problem was the μ projection operator developed by the Manby and Miller groups.[36] This projection operator allows two embedded DFT subsystems (DFT-in-DFT) to exactly recreate full-system Kohn-Sham DFT. However, when embedding a wavefunction (WF) subsystem within a DFT environment (WF-in-DFT), the number of orbitals in the WF subsystem is the same as the number of orbitals in the full system. Since WF methods scale poorly with number of orbitals, basis set truncation methods were developed to reduce computational cost.[105, 51] The more recent truncation method removes basis functions from a subsystem when the density of that subsystem is below a threshold, a manner that maintains a high degree of accuracy. By decoupling the WF calculation from the total size of the system, WF-level energies may be calculated for systems consisting of hundreds of atoms. The μ operator method has shown a high degree of accuracy for transition-metal and enzyme catalysis, and oxidation potentials of molecules in solution, among other systems of interest.[50, 38] Additionally, several groups have used the μ projection operator to embed multireference wavefunction methods for application to transition metal catalysts. [37, 122] These systems are inherently multireference; however, as the multireference character is localized to the metal center, μ embedding calculations were able to closely match experimental results.

Kállay and co-workers introduced the Huzinaga projection operator for DFT embedding;[44] however, that work truncated the orbitals by using local correlation

methods. We showed that the Huzinaga projection operator could be used for aggressive truncation of the orbital space, where the densities could be absolutely localized on the atomic basis functions centered on atoms within the subsystem.[47] This allows for high computational efficiency as the WF subsystem has a greatly reduced number of molecular orbitals. Huzinaga projection embedding has also been successfully extended to periodic systems,[48] allowing for cluster or periodic WF calculations embedded in a periodic DFT environment. Given that the absolutely localized basis used in Huzinaga projection-based embedding reduces the number of orbitals to only those centered on the atoms of interest, we here determine the effectiveness of $i-$FCIQMC on a absolutely localized subsystem within the embedding potential of the full system.

We are generally motivated to increase the range and scope of systems available for study with $i-$FCIQMC. With a view toward our long-term interests in the study of bond-breaking and bond rearrangement on surfaces relevant to heterogeneous catalysis, we here study bond dissociation for diatomic molecules containing ionic or covalent bonds (specifically, LiH and HF, respectively) physisorbed onto a benzene molecule using $i-$FCIQMC. This type of calculation (with $\sim$ 35 active electrons) is currently at the edge of applicability for $i-$FCIQMC; sometimes the system can be treated, and other times it cannot be treated. We show that embedding greatly alleviates the cost scaling of our model system. Specifically, data show that $i-$FCIQMC calculations performed on the full system (including both the diatomic molecule and the benzene molecule) fail to converge, whereas the system in which the benzene is represented by embedding converges with the same efficiency as an isolated molecular calculation. We analyze the type of convergence behaviors in $i-$FCIQMC and relate them to the differing electronic structures of the dissociation reactants and products. In addition, we explore the applicability of $i-$FCIQMC to a range of atomic separations of HF on benzene by calculating a dissociation curve using both $i-$FCIQMC and CCSD(T). We show that for HF on benzene, $i-$FCIQMC does not have the same failure CCSD(T) shows in regions of strong correlation.

4.2 Methods

4.2.1 $i-$FCIQMC

Full configuration interaction quantum Monte Carlo [137] and its initiator adaptation [138] attempt to solve for the ground-state wavefunction $|\Psi_0\rangle$ of the imaginary-time Schrödinger equation of a given Hamiltonian $\hat{H}$:

$$\frac{d|\Psi_0\rangle}{d\tau} = -\hat{H}|\Psi_0\rangle \tag{4.1}$$

where τ represents imaginary time. Beginning with a wavefunction that has non-zero overlap with the ground state, this equation can be solved in the long-imaginary-time limit to give the ground state wavefunction:

$$\lim_{\tau\to\infty} e^{-(\tau\hat{H}-S)}|D_0\rangle \propto |\Psi_0\rangle \tag{4.2}$$

where $|D_0\rangle$ is the reference Slater determinant, here taken to be the Hartree–Fock wavefunction. This relationship holds for any constant energy shift S. When long enough imaginary time τ has passed, S can be averaged, and the correlation energy ($E_{\text{corr}} = E_{\text{total}} - E_{\text{Hartree-Fock}}$) found.

The full configuration interaction wavefunction is typically written as a sum of Slater determinants, $|D_i\rangle$,

$$|\Psi_0\rangle = \sum_i c_i|D_i\rangle \tag{4.3}$$

As such, the imaginary time evolution operator acts in a determinant space.

Substituting Eq. 4.3 into Eq. 4.1 gives an expression which can be written as a finite difference

$$c_i^{m+1} - c_i^m = c_i^m\tau(-H_{ii} + S) - \sum_{j\neq i} c_j^m\tau H_{ij}. \tag{4.4}$$

Here, c_i^m is the coefficient of the i^{th} determinant at the m^{th} iteration of the simulation (after which $m\tau$ units of imaginary time have elapsed). The Hamiltonian is represented in the Slater determinant basis as:

$$H_{ij} = \langle D_i|\hat{H}|D_j\rangle. \tag{4.5}$$

In the original FCIQMC algorithm, the weight c_i takes integer values. [137] The walker population N_w is given by $N_w = \sum_i c_i$. When S is varied to keep the walker population constant, its average becomes an estimate of the total ground-state energy.

The population of particles evolves towards the ground state using the following three steps introduced by Booth et al[137]:

1. The particles with weight c_i are allowed to spawn from site i to a connected site j, where $H_{ij} \neq 0$ and $i \neq j$. The probability of spawning, $p(j|i)$ is uniform over the j which are connected by one or two electron excitations to i. The integer part of $\frac{H_{ij}\tau}{p(j|i)}$ (including its sign) is then added to the weight at j. The non-integer remainder r is added with probability $|r|$ as ± 1, where the sign comes from the sign of r.

2. Each particle with weight c_i changes its weight by $|S - H_{ii}|\tau$. As above, the integer part of $|S - H_{ii}|\tau$ is added to the weight at i. The non-integer remainder r treated as above.

3. Pairs of particles on the same site with opposite weight c_i annihilate each other and are removed from the simulation, leaving a population containing only a single sign on each site.

FCIQMC is not restricted to using only integer weights c_i. Real weights can be used; this adds a step to the above algorithm where the real weight is rounded off stochastically below a certain threshhold (here, 0.01), chosen to reduce stochastic error and raise efficiency [153].

The initiator adaption to FCIQMC, $i-$FCIQMC , separates the Slater determinant space into those with n_{add} (here, 3) or more walkers and those with fewer. If the origin of a spawning event (item 1. in the list above above) is not an "initiator" and the spawning is attempted onto a site without walkers, H_{ij} is zeroed. The result is a dynamically-modified Hamiltonian, which profoundly influences convergence of the simulation. A simulation is only converged in the limit when changing the walker population no longer changes the energy (i.e., $N_w \to \infty$). This is an important practical limitation that must be contended with when running an $i-$FCIQMCcalculation, and this ensures the wavefunction must be sampled with sufficient detail in order to attain statistical and

systematic convergence. As $N_w \to \infty$, the full configuration interaction (i.e. exact) limit is achieved; away from this limit, the calculation contains a small error termed the initiator error. This error typically converges as $\sim \exp(-\alpha N_w)$ and is challenging to extrapolate away. Reducing this error is crucial to the success of $i-$FCIQMC ; its prefactor/rate of decay is highly system dependent, and for larger systems can bottleneck the calculations.

4.2.2 Embedding

To perform $i-$FCIQMC-in-DFT embedding, the full system density is first split into two subsystems, subsystem A and subsystem B

$$\gamma^{\text{tot}} = \gamma^{\text{A}} + \gamma^{\text{B}} \tag{4.6}$$

where γ^{A} and γ^{B} are the densities matrices of subsystems A and B, respectively. We then obtain the DFT densities of the subsystems through a freeze-and-thaw algorithm.[47] This algorithm works by iteratively relaxing the density of subsystem A within the embedding potential and projection operator generated by the frozen density of subsystem B, and then freezing the subsystem A density and relaxing the subsystem B density within the embedding potential and projection operator generated by subsystem A until both subsystem densities have converged. The Fock matrix of subsystem A embedded in subsystem B can be written as

$$\mathbf{F}^{\text{A-in-B}} = \mathbf{h}^{\text{A-in-B}}[\gamma^{\text{A}}, \gamma^{\text{B}}] + \mathbf{g}[\gamma^{\text{A}}] \tag{4.7}$$

where $\mathbf{g}$ contains the Coulomb and exchange-correlation potential for DFT—and the embedded core Hamiltonian is

$$\mathbf{h}^{\text{A-in-B}}[\gamma^{\text{A}}, \gamma^{\text{B}}] = \mathbf{h} + \mathbf{g}[\gamma^{\text{A}} + \gamma^{\text{B}}] - \mathbf{g}[\gamma^{\text{A}}] + \mathbf{P}^{\text{B}} \tag{4.8}$$

where $\mathbf{h}$ is the one electron Hamiltonian, and thus contains the kinetic and nuclear potential operators for both subsystems, and $\mathbf{P}^{\text{B}}$ is the Huzinaga projection operator

for subsystem A, given by

$$\mathbf{P}^{\mathrm{B}} = -\frac{1}{2}\left(\mathbf{F}^{\mathrm{AB}}\gamma^{\mathrm{B}}\mathbf{S}^{\mathrm{BA}} + \mathbf{S}^{\mathrm{AB}}\gamma^{\mathrm{B}}\mathbf{F}^{\mathrm{BA}}\right), \tag{4.9}$$

where $\mathbf{F}^{\mathrm{AB}}$ and $\mathbf{S}^{\mathrm{AB}}$ are elements of the total Fock matrix and overlap matrix described over the basis functions of subsystems A and B. These equations are then analogously defined for the Fock matrix of B in A. Upon freeze-and-thaw convergence at the DFT level, the $\mathbf{h}^{\text{A-in-B}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}]$ is used as the one-electron Hamiltonian for the $i-$FCIQMC calculation; thus, embedding only influences the one-electron integrals for the $i-$FCIQMC calculation. The final embedding energy is then

$$E_{\mathrm{total}} = E^{\mathrm{total}}_{\mathrm{KS-DFT}} - E^{\mathrm{A}}_{\mathrm{DFT-in-DFT}} + E^{\mathrm{A}}_{\mathrm{iFCIQMC-in-DFT}}, \tag{4.10}$$

where $E^{\mathrm{total}}_{\mathrm{KS-DFT}}$ is the full-system Kohn-Sham (KS)-DFT energy, $E^{\mathrm{A}}_{\mathrm{DFT-in-DFT}}$ is the DFT energy of subsystem A embedded in the DFT potential of the rest of the system, and $E^{\mathrm{A}}_{\mathrm{iFCIQMC-in-DFT}}$ is the $i-$FCIQMC energy of subsystem A embedded in the DFT potential of the rest of the system.

4.2.3 Calculation details

The atomic coordinates of the systems under investigation were generated using the dispersion-corrected M06-D3 functional and the aug-cc-pVTZ basis set as implemented in Gaussian16. The geometries are presented in Appendix C. Six frozen orbitals were used for the C_6H_6-LiH canonical systems, and seven frozen orbitals were used for the C_6H_6-HF canonical systems. No frozen orbitals were used for the embedded integrals of either system, nor the isolated diatomics. In our implementation, QSoME was modified to output integrals for $i-$FCIQMC using PySCF, [111] which were then read into the HANDE software package.[185] For the dissociation curve of HF,MOLPRO [113] was also used taking advantage of an already-existing interface with QSoME. These integrals consisted of single-particle Hartree–Fock eigenvalues (ϵ_i) and electron repulsion integrals (v_{ijkl}).

The $i-$FCIQMC calculations were performed using the open-source code HANDE-QMC. For the C_6H_6-LiH system, an imaginary time step of 2×10^{-6} a.u. was used

with 200,000 reports and 20 Monte Carlo cycles between reports. For the C_6H_6−HF and C_6H_6−F^- systems, a smaller time step of 9×10^{-7} a.u. was used due to the additional electrons present, with 400,000 reports for the first three target populations and 600,000 reports for the largest three target populations. A larger time step of 0.002 a.u. was used for the isolated LiH, HF and the embedded systems, except for the 5 and 6 Å separations, which used a timestep of 0.0002. In order to converge the calculations with respect to the target population, a range of target populations between 10^1 and 10^6 was used.

Without the embedding algorithm, the LiH physisorbed on benzene system contains 34 electrons, requires 2.8×10^{41} determinants, and has a storage cost of 700 MB. After embedding is introduced, the subsystem treated with i−FCIQMC is reduced to 4 electrons and 2.9×10^4 determinants, with an integral storage cost of 440 KB.

4.3 Results and discussion

It is common for energy *differences* to yield better convergence (with respect to excitation rank, for example, in coupled cluster theory) than total energies themselves; this phenomenon, known as error cancellation, is a common benefit of running quantum-chemical calculations. In i−FCIQMC (in common with FCIQMC), a walker population of a given size (N_w) represents the wavefunction. The calculation is only exact if it is converged with respect to this walker number. An under-explored issue of i−FCIQMC calculations is that convergence is not faster for energy differences than for individual energies. The dissociation energies of LiH on benzene and HF on benzene represent two paradigmatic examples of how dissociation energies can be extremely challenging and costly to converge in i−FCIQMC due to a lack of error cancellation between reactants and products.

We hypothesize that adding benzene to straightforward LiH and HF dissociation energy calculations will cause i−FCIQMC to fail in a way that can be remedied by using embedding. To test our hypothesis, we calculate the energy changes associated

with four reactions:

$$\begin{aligned} \mathrm{LiH} &\longrightarrow \mathrm{Li^+ + H^-} \\ \mathrm{C_6H_6{-}LiH} &\longrightarrow \mathrm{C_6H_6{-}Li^+ + H^-} \\ \mathrm{HF} &\longrightarrow \mathrm{H^+ + F^-} \\ \mathrm{C_6H_6{-}HF} &\longrightarrow \mathrm{C_6H_6{-}F^- + H^+} \end{aligned} \tag{4.11}$$

Here, we are required to use the closed-shell ionic dissociation products by the embedding code; an open shell implementation is planned. We note that in $C_6H_6{-}HF$, the H atom is closest to the benzene ring and, following H^+ removal and geometry optimization, the F^- migrates into the plane of the ring. In particular, we reason that the dissociation energy of a LiH or HF molecule physisorbed to benzene will be significantly more difficult to calculate using $i{-}$FCIQMC due to non-monotonic energy convergence with system size N. In contrast with other methods, $i{-}$FCIQMC does not show error cancellation between systems that contain different numbers of electrons.

Figure 4.1 shows data we collected in support of our claim. This data is also presented in table form in Appendix C. Each of these plots is an initiator convergence plot, where the walker population is varied from 10^1 to 10^6, and the energy is computed using $i{-}$FCIQMC . We plot the $i{-}$FCIQMC energy differences between reactants and products for the LiH and HF dissociation reactions, and compare these differences to CCSD(T) dissociation energies. CCSD(T) can serve as a good benchmark for initiator convergence: initiator error can vary greatly over many orders of magnitude in energy, and CCSD(T) is generally thought to have systematic error only on the order of 1 millihartree.

Figure 4.1(a) shows that isolated LiH and HF dissociation energies rapidly converge as a function of walker number, showing complete convergence at 10^4 and 10^5 walkers, respectively. The $i{-}$FCIQMC and CCSD(T) results are in agreement with each other to within 1 millihartree for $N_w \geq 10^3$, and within 10 millihartree for the smaller target populations. The HF dissociation converges in an oscillatory manner, because HF is slightly slower to converge than F^-; in general, fine-grained oscillatory convergence has been shown in individual calculations. [141] The HF system contains more variability at lower walker numbers than the LiH system, as is expected due to the higher number

of electrons present in HF. As we expect, our results show that the isolated systems with small numbers of electrons converge with only modest convergence errors.

In contrast to the isolated molecules, convergence is difficult for the dissociation of molecules physisorbed on benzene. The convergence difficulties for these systems are shown in Fig. 4.1(b), where the oscillatory behavior observed in Fig. 4.1(a) is magnified; in the case of HF, we are not able to converge this calculation at all in order to obtain a reaction energy, as the energy difference between 10^5 and 10^6 walkers is approximately -0.0597 hartree. Physisorption onto benzene adds 30 electrons to the isolated molecules; thus, significantly harder convergence is unsurprising. Again, since $i-$FCIQMC does not show error cancellation between systems containing different numbers of electrons, C_6H_6-HF and $C_6H_6-F^-$ converge at different rates, which causes the energy difference between these two systems to be oscillatory. This is a key result of this manuscript that we explore later in further detail.

In Fig. 4.1(c), we present the results of the $i-$FCIQMC-in-DFT embedded systems. Since embedding decreases the number of electrons treated directly by $i-$FCIQMC, we are able to converge the $i-$FCIQMC energies of C_6H_6-LiH and C_6H_6-HF as easily as isolated LiH and HF. We see similar oscillatory behavior in the embedded calculations as we do for the isolated systems: Target populations 10^1 and 10^2 are still not very accurate. Fortunately, as we increase the target population, we see clear convergence. Comparing the three initiator curves across Fig. 4.1 reveals a similar convergence trend. This is a very encouraging result, as it shows the $i-$FCIQMC-in-DFT embedding gives convergent results while simultaneously reducing the cost of these calculations significantly.

As computational cost is proportional to walker number, the ability to converge a calculation at 10^3 walkers compared with leaving it unconverged at 10^6 walkers represents a cost savings of at least 1000x. Data we present in Appendix C additionally show a 1000x savings in memory.

We fully appreciate that there is an unquantified embedding error in these calculations. This causes a change in ordering of the C_6H_6-HF and C_6H_6-LiH dissociation energies between Fig. 4.1(b) and Fig. 4.1(c) at the CCSD(T) level. For completeness, we note that the difference between CCSD(T) embedded calculations and full-system calculations give us an estimate of the $i-$FCIQMC embedding error as 4.31 millihartree and 8.01 millihartree for LiH and HF, respectively. However, our previous studies have

shown that the embedding error can be further decreased by enlarging the wavefunction subsystem.[48] Although we are interested in quantifying the $i-$FCIQMC embedding error and using it to benchmark embedded CCSD(T), this analysis is beyond the scope of the proof-of-principle offered by this paper. We now analyze the sources of error and the way that embedding overcomes convergence difficulties in $i-$FCIQMC.

4.3.1 Analysis of different convergence behaviors in $i-$FCIQMC

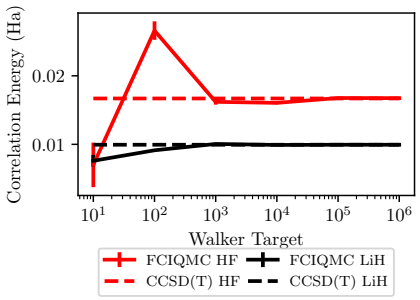
There are a number of analyses we can conduct in order to probe the extent of the non-convergent behavior described above in Fig. 4.1(b)—the case where all electrons in the benzene molecule are fully present in the $i-$FCIQMC calculation. In Fig. 4.2, the convergence of the reactants and products of dissociation for C_6H_6-LiH and C_6H_6-HF are shown. It can be seen from this figure that these calculations are not converged with respect to the number of walkers. This represents a particularly severe case where reactant and product energies actually cross over, which causes the energy differences to oscillate rather than converge smoothly, as observed in Fig. 4.1(b).

Both C_6H_6-LiH and C_6H_6-HF represent different types of challenges in convergence. In C_6H_6-HF, where reactants and products have the same number of electrons, each $i-$FCIQMC calculation appears to be smoothly converging as a function of walker number. Prior work has established the appearance of such smooth convergence as a stretched exponential in the walker population, $\exp(-N_w^{\alpha}), \alpha << 1.0$.[143] The decay parameters are highly system-dependent, and as such, two converging calculations could easily cross over one another. The general form of two converging calculations is:

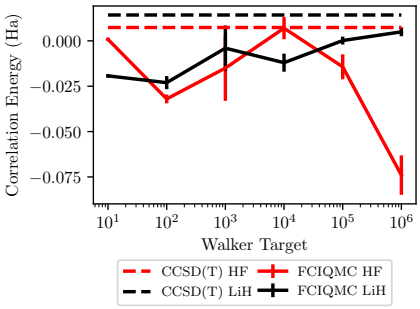
$$E_{\mathrm{corr,A-B}} = E_{corr,A} - E_{corr,B} + A_1 e^{-N_w^{\alpha 1}} - A_2 e^{-N_w^{\alpha 2}} \tag{4.12}$$

In the case of HF, the combined initiator error, $A_1 e^{-N_w^{\alpha 1}} - A_2 e^{-N_w^{\alpha 2}}$, obscures or is much larger than the term $E_{corr,A} - E_{corr,B}$. As a result, the reaction energy fails to converge, instead oscillating even at large walker numbers.

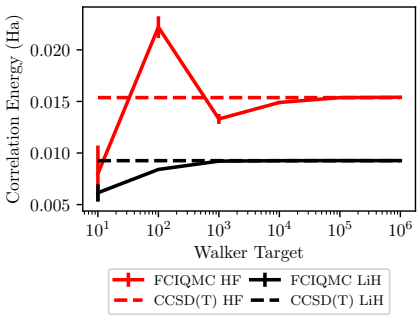
The underlying reason for the differences in convergence between C_6H_6-HF and $C_6H_6-F^-$ is not known. It seems likely that the form of the stretched exponential is itself related to excited state decays in imaginary time, although this has not been



(a)



(b)



(c)

Figure 4.1: Correlation energy contribution to the dissociation energies of cc-pVDZ LiH and HF for molecules that are (a) isolated (4 and 10 electrons respectively), (b) physisorbed to benzene (34 and 38 electrons respectively), (c) physisorbed to benzene and embedded (4 and 10 electrons treated explicitly with $i-$FCIQMC). The $i-$FCIQMC calculations, shown as solid lines, were performed with six target populations ranging from 10^1 to 10^6 on a logarithmic scale. Good agreement is achieved between $i-$FCIQMC and CCSD(T) for isolated and embedded systems.

established in the literature. Specifically, the overlap between the simulation wavefunction in imaginary time, $|\Psi(\tau)\rangle$, and the FCI excited states, $|\Psi_i\rangle$, is expected to decay exponentially in imaginary time:[143]

$$\langle\Psi(\tau)|\Psi_i\rangle = C_i \exp(-\tau(E_i - E_0)) \tag{4.13}$$

where E_i and E_0 are the excited state and ground state energy eigenvalues, respectively. In this picture, then, a simulation with insufficient walker population would have to get stuck somewhere between one state and another in a way that cannot be resolved by projecting out over more imaginary time steps, because there is not enough information in each timestep to afford resolution of the ground state.

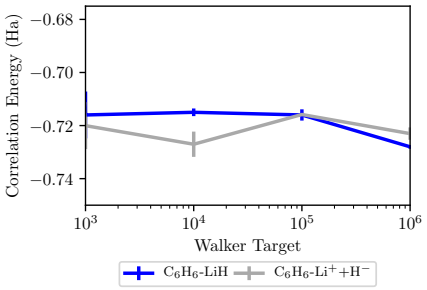
The case of C_6H_6−LiH is a little different, since C_6H_6−Li^+ exhibits oscillatory convergence already. This case of oscillatory fine structure has been seen before, such as in studies of the uniform electron gas. [141] This on its own hampers convergence, lending an oscillatory character to the reaction energy independent of whether these calculations are themselves converging to the correct energy.

4.3.2 Hartree–Fock Population

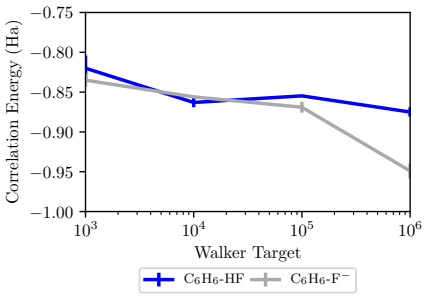
Another measure by which we can compare the isolated and embedded calculations is the number of walkers present on the Hartree–Fock determinant (shown in Fig. 4.3). This population is sometimes used as a means to determine convergence of an i−FCIQMC calculation, since, in the early phase of an i−FCIQMC calculation, it does not vary from its baseline of $[\mathcal{O}(1)]$ walker. The number of walkers on the Hartree–Fock determinant also confirms the different convergence behaviors of the full, isolated, and embedded systems: The embedded and isolated systems have Hartree–Fock populations that grow at the same rate, whereas the full system has a lot fewer of these kinds of walkers. In terms of the walker dynamics, the larger number of determinants in the full system depletes the signal present on the Hartree–Fock determinant and slows convergence.

4.3.3 Embedding and the sign problem in i−FCIQMC

The sign problem in FCIQMC has been related to the amount of spin frustration in the system: each Slater determinant in the system needs to find its sign over the course of



(a)



(b)

Figure 4.2: The initiator curves at walker numbers $N_w = 10^3$ through 10^6 for the products and reactants of the dissociation reactions of (a) LiH and (b) HF physisorbed on benzene.

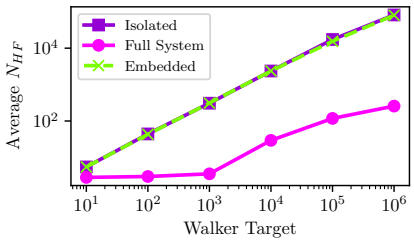
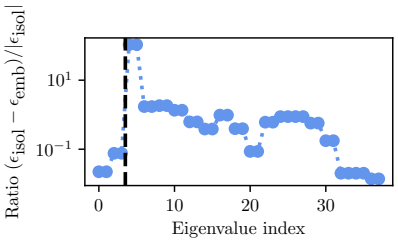
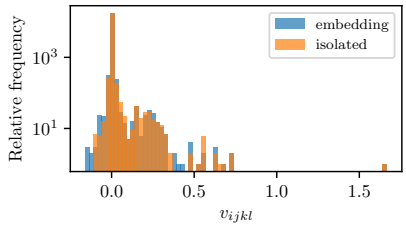


Figure 4.3: The population of walkers on the Hartree–Fock determinant in the $i-$FCIQMC calculation with respect to iteration for each target population of $N_w = 10^1$ to 10^6 for each of three LiH systems: isolated LiH, the full system C_6H_6-LiH and the embedded C_6H_6-LiH.



(a)



(b)

Figure 4.4: Changes in the LiH integral table for $i-$FCIQMC represented through (a) differences between eigenvalues ϵ_i for the embedding and isolated systems, where the black dashed line represents the division between occupied and virtual Hartree–Fock orbitals, and (b) electron repulsion integrals v_{ijkl} for both embedded and isolated systems.

a simulation.[140] Specifically, the eigenvalue of a matrix $H'_{ij} = \delta_{ij}H_{ij} - (1 - \delta_{ij})|H_{ij}|$, where δ_{ij} is the Kronecker delta, whose eigenstate has entirely non-negative components and contaminates solutions.

The signs in H come from the four-index integrals via the Slater–Condon rules, and so it is important to discuss whether there is a significant change in the integrals due to embedding. In Fig. 4.4, we show a comparison of two types of integrals that are passed between the embedding code and $i-$FCIQMC for isolated LiH compared with embedded C_6H_6-LiH. In this case, the LiH eigenvalues are generally lowered by between -0.01 hartree to -0.3 hartree by embedding. The specific ratio for each eigenvalue is plotted against its energy-ordered index in Fig. 4.4(a), showing that as the eigenvalue becomes higher in energy, it is also affected less by embedding. We show the effect on the electron repulsion integrals in Fig. 4.4(b), where the distribution of the ~ 1800 integrals is presented as a histogram. The molecular orbitals differ between the isolated case and the embedding case and this leads to the small changes in the electron repulsion integrals. From the plots above, we would expect that there is not an increase in the complexity of the sign problem, since most matrix elements remain unchanged.

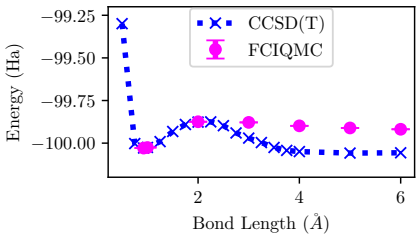
4.3.4 Application to bond stretching

Bond dissociation energy curves are frequently used to benchmark new developments in FCIQMC [143]. This is in part because CCSD(T) is known to fail due to the strong correlation which occurs as the bond is stretched, leading to certain determinants becoming closer in energy while being strongly coupled.[186] In order to highlight the potential benefits of FCIQMC to the study of catalysis we can therefore make comparison between FCIQMC and CCSD(T) for a bond dissociation curve.

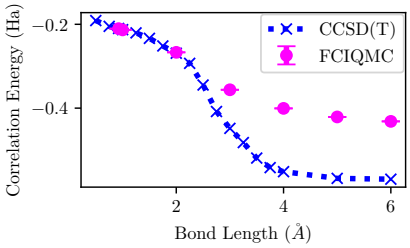
Here, we model the dissociation of H−F in C_6H_6-HF by increasing the H−F bond distance. This represents the following dissociation:

$$C_6H_6-HF \longrightarrow C_6H_6-H\cdots F \tag{4.14}$$

where the dissociation products, by contrast to Eq. (11), show dissociation by drawing the F^- away from the molecule with the rest of the geometry frozen.



(a)



(b)

Figure 4.5: Bond dissocation energy curves for cc-pVDZ hydrogen fluoride molecule embedded on benzene, showing $i-$FCIQMC has improved accuracy over CCSD(T) for (a) total energies and (b) correlation energies. These graphs show agreement between the two methods between the equilibrium separation and 2.00 Å, but the two methods diverge at longer separations. CCSD(T) calculations are shown as blue dashed lines and $i-$FCIQMC calculations are shown as fuchsia circles. The CCSD(T) calculations were performed on atomic separations from 0.50 Å to 4.00 Å in 0.25 Å increments, as well as the equilibrium separation of 0.92 Å and separations of 5 and 6 Å. $i-$FCIQMC calculations were added at the equilbrium geometry, 1.00, 2.00, 3.00, 4.00, 5.00 and 6.00 Å separation.

Figure 4.5 shows total and correlation energies calculated at different H−F separations. At above 2Å, the CCSD(T) energy decreases in a manner indicative of strong correlation. By contrast, FCIQMC energies appear to level off to an overall correlation contribution to the bond dissociation energy of approximately −0.221(1) Ha (between the equilibrium separation and 6Å). We note in passing that this is different from the previous correlation contribution to bond dissociation energy of 0.0154(2) Ha because these fragments are not the same (see Eq. (14) and Eq. (11)).

4.4 Conclusions

In summary, we here examined convergence difficulties present when using i−FCIQMC to calculate the electronic structure of large systems by exploring the bond dissociations of two prototypical molecules, LiH and HF, physisorbed to benzene. Since i−FCIQMC does not show error cancellation between systems with different numbers of electrons, the energy differences between reactants and products tended to oscillate. As a result, dissociation energies calculated from i−FCIQMC did not converge. To remedy the convergence issues that i−FCIQMC has with large systems, we embedded i−FCIQMC in DFT. We showed that this new embedded i−FCIQMC was better able to converge dissociation energies, giving results that agree with our CCSD(T) benchmarks. By way of an application, we have also shown the ability of i−FCIQMC -in-DFT to more accurately model dissociation curves at atomic separations greater than equilibrium than CCSD(T). This demonstrates the ability of the absolute localization approach for Huzinaga projection-based embedding to treat strongly correlated systems using high-level i−FCIQMC wavefunctions embedded in DFT.

Since embedded i−FCIQMC also reduces the number of electrons (and thus orbitals) in a calculation, embedded i−FCIQMC calculations run with substantially lower cost than full i−FCIQMC, alleviating the method's reduced-exponential cost scaling. Based on our results, we estimate the cost saving to be at least 1000x in compute time and 1000x in memory for the model systems studied here, whereas for larger systems, i−FCIQMC calculations can be computationally intractable while embedded i−FCIQMC calculations will remain feasible. There are applications for which CCSD(T) fails to give good answers, such as those involving strong correlation or bond breaking;

$i-$FCIQMC can treat these applications with high accuracy. As such, we believe that $i-$FCIQMC emdedded in DFT is a significant and realistic step forward for bringing $i-$FCIQMC towards the routine treatment of real applications, as DFT embedding both alleviates convergence concerns and dramatically reduces the cost of the method.

More broadly, the dissociation curve we calculated represents a situation where strong correlation (bond breaking) was treated by QMC and weak correlation (physisorption) was treated by embedding and DFT; this is likely the best-case scenario for our method. To extend this work, we would move towards real systems. A similar embedding approach as the one we take here has already shown promise for being applied to catalysis.[50] For example, quantum embedding was applied to a variety of Co-based catalysts to explore the coupling of the electronic structure of the transition metal to that of the ligand in the hydrogen evolution reaction,[93] and to explore the multireference character of these systems that presents challenges for DFT when calculating reaction barriers. [40] Also, Carter and coworkers have used a similar but distinct embedding method to study H_2 dissociation on Au and Al nanoparticles.[187, 188] In our example of HF dissociation, there was no barrier to dissociation, however in the general case a bond dissociation curve could be used to determine transition state energies and therefore kinetic barrier heights.

We believe that our work is particularly timely because there has been a call from prominent researchers studying oxygen reduction catalysts [189] to focus on the understanding and design of multi-functional active sites for next-generation catalysts and suggested embedding methods could help us get there. In examples such as bifunctional sites or in confinement, we may well expect strong correlation which is quantum mechanically coupled to the environment, and we expect FCIQMC-in-DFT embedding to find applications there.

There are limitations to the embedding approach which are relevant for high-accuracy modelling. The general case where this method will work is when the density from DFT is almost exact. This is especially true for the DFT region, whose density does not change in this method due to our using a frozen density approach. In practice, there are studies that have explored the severity of this approximation[39] and found that much can be gained by allowing the errors in the density outside the embedded (here, QMC) region to cancel. It is still very much an open question as to whether a

region including strong correlation could be left in the embedded region.

Since strong correlation is often investigated by way of model systems, we note the conditions to apply this approach to model systems as follows. The model system would need to have identifiable localized fragments (here, atoms) and orbitals that are associated with that local fragment which have a dot product rule. The formalism by which the subsystems are divided is exact within a Kohn–Sham formalism and partitions the subsystems to have integer numbers of electrons. In principle it would be possible to use non-integer subsystems which has been applied to 1-D hydrogen chain systems.[70, 190]

One other limitation of this work is that we have not analyzed the added error in the correlation energy introduced when undertaking embedding, since we believe it is outside of the scope of a proof-of-principle and deserves much more attention on its own. Since CCSD(T) can treat the full systems for the prototypical bond dissociations studied in this manuscript, we *could* have added a correction to our embedded $i-$FCIQMC arising from the CCSD(T) energy difference between the full and embedded systems; this may be a way forward for future work. It is also of note that the embedding error has been analyzed for CCSD-in-DFT in comparison to CCSD[47] and also for CCSD(T)-in-DFT in comparison with experiment;[48] we would expect comparable errors at this level of theory. We could also treat a system that is small enough to examine the full system with $i-$FCIQMC , resulting in our being able to benchmark the embedding error for the benefit of other practitioners.

Further work will be forthcoming where this embedding is further developed for excited states and EA/IP calculations; FCIQMC can also be interfaces with other types of calculations such as those with periodic boundary conditions for which a separate periodic code exists. Benchmarking $i-$FCIQMC in comparison to other high-accuracy methods (CCSD(T), DMRG, selected CI) to find the relative advantages and disadvantages of each method represents a very interesting open question. To facilitate this, integral files and output files can be found at https://doi.org/10.25820/data.001111.

In closing, we believe that this study highlights an important step forward for both $i-$FCIQMC and embedding. We believe that the work presented here brings the community one step closer to the routine application of high-accuracy electronic structure to study strongly-correlated systems of chemical and technological interest.

Chapter 5

Analytical Nuclear Gradients for Absolutely Localized Huzinaga WF-in-DFT Embedding

5.1 Introduction

Quantum embedding methods have been very successful in providing accurate energies for large, complex chemical systems at a relatively low cost. While highly accurate post-Hartree-Fock wave function (WF) methods scale prohibitively for large chemical systems, relatively low cost Kohn-Sham density functional theory[54, 4] (DFT) methods have well known accuracy limitations[125, 126, 127, 10, 11, 12, 13, 14, 15, 16, 17, 6, 7, 8, 9]. Quantum embedding methods are able to combine the high accuracy of WF methods with the favorable computational scaling of DFT. Some well studied quantum embedding methods such as our own n-layered integrated molecular orbital and molecular mechanics[61] (ONIOM), density matrix embedding theory[62, 18] (DMET), embedded mean-field theory[63, 64, 65, 66, 67], Green's function embedding[68, 69, 18, 128], partition DFT[70, 71, 72], and DFT embedding[21, 18, 22, 23, 24] among many others[73, 74, 75, 76, 77, 129], demonstrate highly accurate energies for significantly lower computational cost than similarly accurate WF methods. The success of many of these methods can be attributed to the fact that one can partition a chemical system

into a localized subsystem and perform a highly accurate calculation on that subsystem, while the remainder of the system is simulated using a lower cost method such as DFT.

We have previously demonstrated the high accuracy and low computational cost of our absolutely localized Huzinaga projection based WF-in-DFT embedding method.[47, 130, 191, 48, 104] We are able to recreate WF energies of bond formation, gas adsorption to transition metal centers, and spin-transition energies to give a few examples.[130, 191] Additionally, our method can be applied to periodic models[48], and can calculate properties for excited-state[104] and open-shell systems[191]. A key feature of our embedding method is the absolutely localized basis, which significantly reduces the computational cost of the WF subsystem while maintaining high accuracy. An important next step in the development of absolutely localized Huzinaga projection WF-in-DFT embedding is development of analytical nuclear gradients.

Analytical nuclear gradients for absolutely localized Huzinaga projection WF-in-DFT embedding will significantly expand the applicability of our method. Accurate molecular geometries are the foundation for a wide variety of chemical properties. With highly accurate analytical nuclear gradients, one can accurately optimize molecular geometries, perform transition state searches and simulate molecular dynamics. Analytical nuclear gradients have been successfully derived and implemented for a variety of embedding methods, including projection based WF-in-DFT embedding by Lee, Manby and Miller.[41] We believe analytical nnuclear gradients utilizing the absolutely localzed basis will be an important contribution to this field.

In the following sections we outline the absolutely localized Huzinaga WF-in-DFT energy equation, provide the derivation for the analytical nuclear gradients of our embedding method, and include a comparison of numerical and analytical nuclear gradients using our embedding method.

5.2 Theory

Absolutely localized Huzinaga projection WF-in-DFT involves dividing the full chemical system into subsystems. For simplicity, we will be discussing dividing the system into two subsystems, however the method could be applied to any number of subsystems. Our previously derived energy equation for absolutely localized Huzinaga embedding is

$$E^{\text{Full}}_{\text{WF-in-DFT}} = E^{\text{Full}}_{\text{DFT}} - E^{\text{I}}_{\text{DFT}} + E^{\text{I}}_{\text{WF}} \tag{5.1}$$

where $E^{\text{Full}}_{\text{WF-in-DFT}}$ is the embedding energy of the full system, $E^{\text{Full}}_{\text{DFT}}$ is the energy of the full system calculated at the DFT level of theory, $E^{\text{I}}_{\text{DFT}}$ is the energy of subsystem I calculated at the DFT level of theory, and E^{I}_{WF} is the energy of subsystem I calculated at the WF level of theory.[47, 130] The differentiation of $E^{\text{Full}}_{\text{WF-in-DFT}}$ with respect to atomic coordinate $\mathbf{R}$ is

$$\frac{\partial E^{\text{Full}}_{\text{WF-in-DFT}}}{\partial \mathbf{R}} = \frac{\partial E^{\text{Full}}_{\text{DFT}}}{\partial \mathbf{R}} - \frac{\partial E^{\text{I}}_{\text{DFT}}}{\partial \mathbf{R}} + \frac{\partial E^{\text{I}}_{\text{WF}}}{\partial \mathbf{R}}. \tag{5.2}$$

The first term on the right hand side (R.H.S.) can be calculated using previously available standard analytical nuclear gradients for DFT. Because of freeze-and-thaw absolute localization, the third term may also be solved by modifying standard analytical gradient solvers to substitute the gradient of the embedded core hamiltonian, $\mathbf{h}^{\text{I-in-II}}$, derived below, for the standard core hamiltonian. The second term on the R.H.S. must be derived for our absolutely localized Huzinaga embedding method. The embedded DFT energy of subsystem I is defined as,

$$E^{\text{I}}_{\text{DFT}} = \sum_{\mu\nu\in\text{I}} \gamma^{\text{I}}_{\mu\nu} h^{\text{I-in-II}}[\gamma^{\text{I}}, \gamma^{\text{II}}]_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma\in\text{I}} \gamma^{\text{I}}_{\mu\nu}\gamma^{\text{I}}_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - \frac{c_x}{2}(\mu\lambda|\nu\sigma)\right] + E^{\text{I}}_{\text{xc}} \tag{5.3}$$

where $\gamma^{I}_{\mu\nu}$ is the electron density matrix of subsystem I, c_x is a constant determining the amount of exact exchange, $(\mu\nu|\lambda\sigma)$ is the two-electron integral, and the exchange-correlation energy is a functional of electron density,$E^{\text{I}}_{\text{xc}}[\rho^{\text{I}}(\vec{r})]$ where

$$\rho^{\text{I}}(\vec{r}) = \sum_{\mu\nu\in\text{I}} \gamma^{\text{I}}_{\mu\nu} \phi^{\text{I}}_{\mu}(\vec{r})\phi^{\text{I}}_{\nu}(\vec{r}). \tag{5.4}$$

In the above equation, $\phi^{\text{I}}_{\mu}(\vec{r})$ are the atomic orbitals on fragment I. Finally the embedded core hamiltonian can be written,

$$\mathbf{h}^{\text{I-in-II}}[\gamma^{\text{I}}, \gamma^{\text{II}}] = \mathbf{h} + \mathbf{G}^{\text{I+II}}[\gamma^{\text{I}} + \gamma^{\text{II}}] - \mathbf{G}^{\text{I}}[\gamma^{\text{I}}] + \mathbf{P}^{\text{I-in-II}}[\gamma^{\text{I}} + \gamma^{\text{II}}]. \tag{5.5}$$

Here $\mathbf{h}$ is the full system one-electron Hamiltonian, the two electron potential $\mathbf{G}$ is

$$\mathbf{G}^{\mathrm{I}}[\gamma^{\mathrm{I}}] = \sum_{\lambda\sigma} \gamma^{\mathrm{I}}_{\lambda\sigma}[(\mu\nu|\lambda\sigma) - c_x(\mu\lambda|\nu\sigma)] + \mathbf{V}^{\mathrm{I}}_{\mathrm{xc}} \tag{5.6}$$

where

$$V^{\mathrm{I}}_{\mathrm{xc}}[\rho^{\mathrm{I}}(\vec{r})]_{\mu\nu} = \int \phi^{\mathrm{I}}_{\mu}(\vec{r}) \frac{\partial E^{\mathrm{I}}[\rho^{\mathrm{I}}(\vec{r})]}{\partial \rho^{\mathrm{I}}(\vec{r})} \phi^{\mathrm{I}}_{\nu}(\vec{r}) d\vec{r} \tag{5.7}$$

and

$$\begin{aligned} P^{\text{I-in-II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]_{\mu\nu} = -\frac{1}{2} \sum_{\lambda\sigma\in\mathrm{II}} (&h_{\mu\lambda}\gamma^{\mathrm{II}}_{\lambda\sigma} S_{\sigma\nu} + G^{\mathrm{I+II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]_{\mu\lambda}\gamma^{\mathrm{II}}_{\lambda\sigma} S_{\sigma\nu} \\ &+ S_{\nu\sigma}\gamma^{\mathrm{II}}_{\sigma\lambda} h_{\lambda\mu} + S_{\nu\sigma}\gamma^{\mathrm{II}}_{\sigma\lambda} G^{\mathrm{I+II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]_{\lambda\mu}). \end{aligned} \tag{5.8}$$

In the above equation, $S_{\sigma\nu}$ is the atomic orbital (AO) overlap between subsystems. Throughout this paper we will refer to AOs using greek letters ($\mu, \nu, \lambda, \sigma$,etc.), the entire molecular orbital (MO) space using p,q,r,s,..., the occupied MO space using i,j,k,l..., and the virtual MO space using a,b,c,d... for the sake of convenience.

The differentiation of $E^{\mathrm{I}}_{\mathrm{DFT}}$ with respect to atomic coordinate $\mathbf{R}$ leads to

$$\begin{aligned} \frac{\partial E^{\mathrm{I}}_{\mathrm{DFT}}}{\partial \mathbf{R}} = &\sum_{\mu\nu\in\mathrm{I}} \frac{\partial \gamma^{\mathrm{I}}_{\mu\nu}}{\partial \mathbf{R}} F^{\mathrm{I,emb}}_{\mu\nu} + \sum_{\mu\nu\in\mathrm{I}} \gamma^{\mathrm{I}}_{\mu\nu} \frac{\partial h^{\text{I-in-II}}[\gamma^{\mathrm{I}}, \gamma^{\mathrm{II}}]_{\mu\nu}}{\partial \mathbf{R}} \\ + \frac{1}{2} \sum_{\mu\nu\lambda\sigma\in\mathrm{I}} &\gamma^{\mathrm{I}}_{\mu\nu}\gamma^{\mathrm{I}}_{\lambda\sigma} \frac{\partial[(\mu\nu|\lambda\sigma) - \frac{c_x}{2}(\mu\lambda|\nu\sigma)]}{\partial \mathbf{R}} + \sum_{\mu\nu\in\mathrm{I}} \gamma^{\mathrm{I}}_{\mu\nu} \frac{\partial V^{\mathrm{I}}_{\mathrm{xc},\mu\nu}}{\partial \mathbf{R}}. \end{aligned} \tag{5.9}$$

Here $F^{\mathrm{I,emb}}_{\mu\nu}$ is the embedded Fock matrix of subsystem I,

$$F^{\mathrm{I,emb}}_{\mu\nu} = \mathbf{h}^{\text{I-in-II}}[\gamma^{\mathrm{I}}, \gamma^{\mathrm{II}}] + \mathbf{G}^{\mathrm{I}}[\gamma^{\mathrm{I}}], \tag{5.10}$$

and

$$\frac{\partial \mathbf{h}^{\text{I-in-II}}[\gamma^{\mathrm{I}}, \gamma^{\mathrm{II}}]}{\partial \mathbf{R}} = \frac{\partial \mathbf{h}}{\partial \mathbf{R}} + \frac{\partial \mathbf{G}^{\mathrm{I+II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]}{\partial \mathbf{R}} - \frac{\partial \mathbf{G}^{\mathrm{I}}[\gamma^{\mathrm{I}}]}{\partial \mathbf{R}} + \frac{\partial \mathbf{P}^{\text{I-in-II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]}{\partial \mathbf{R}} \tag{5.11}$$

where,

$$\frac{\mathbf{G}^{\mathrm{I}}[\gamma^{\mathrm{I}}]}{\partial \mathbf{R}} = \sum_{\lambda\sigma} \frac{\partial \gamma^{\mathrm{I}}_{\lambda\sigma}}{\partial \mathbf{R}}[(\mu\nu||\lambda\sigma) - c_x(\mu\lambda|\nu\sigma)] + \sum_{\lambda\sigma} \gamma^{\mathrm{I}}_{\lambda\sigma} \frac{\partial[(\mu\nu||\lambda\sigma) - c_x(\mu\lambda|\nu\sigma)]}{\partial \mathbf{R}} + \frac{\partial \mathbf{V}^{\mathrm{I}}_{\mathrm{xc}}}{\partial \mathbf{R}}. \tag{5.12}$$

The derivative of the projection operator may be written as,

$$\begin{aligned}
\frac{\partial P^{\text{I-in-II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]_{\mu\nu}}{\partial \mathbf{R}} = -\frac{1}{2} \sum_{\lambda\sigma \in \mathrm{II}} \Bigg(& \frac{\partial h_{\mu\lambda}}{\partial \mathbf{R}} \gamma^{\mathrm{II}}_{\lambda\sigma} S_{\sigma\nu} + h_{\mu\lambda} \frac{\partial \gamma^{\mathrm{II}}_{\lambda\sigma}}{\partial \mathbf{R}} S_{\sigma\nu} + h_{\mu\lambda} \gamma^{\mathrm{II}}_{\lambda\sigma} \frac{\partial S_{\sigma\nu}}{\partial \mathbf{R}} \\
& + \frac{\partial G^{\mathrm{I+II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]_{\mu\lambda}}{\partial \mathbf{R}} \gamma^{\mathrm{II}}_{\lambda\sigma} S_{\sigma\nu} + G^{\mathrm{I+II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]_{\mu\lambda} \frac{\partial \gamma^{\mathrm{II}}_{\lambda\sigma}}{\partial \mathbf{R}} S_{\sigma\nu} + G^{\mathrm{I+II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]_{\mu\lambda} \gamma^{\mathrm{II}}_{\lambda\sigma} \frac{\partial S_{\sigma\nu}}{\partial \mathbf{R}} \\
& + \frac{\partial S_{\nu\sigma}}{\partial \mathbf{R}} \gamma^{\mathrm{II}}_{\sigma\lambda} h_{\lambda\mu} + S_{\nu\sigma} \frac{\partial \gamma^{\mathrm{II}}_{\sigma\lambda}}{\partial \mathbf{R}} h_{\lambda\mu} + S_{\nu\sigma} \gamma^{\mathrm{II}}_{\sigma\lambda} \frac{\partial h_{\lambda\mu}}{\partial \mathbf{R}} \\
& + \frac{\partial S_{\nu\sigma}}{\partial \mathbf{R}} \gamma^{\mathrm{II}}_{\sigma\lambda} G^{\mathrm{I+II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]_{\lambda\mu} + S_{\nu\sigma} \frac{\partial \gamma^{\mathrm{II}}_{\sigma\lambda}}{\partial \mathbf{R}} G^{\mathrm{I+II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]_{\lambda\mu} + S_{\nu\sigma} \gamma^{\mathrm{II}}_{\sigma\lambda} \frac{\partial G^{\mathrm{I+II}}[\gamma^{\mathrm{I}} + \gamma^{\mathrm{II}}]_{\lambda\mu}}{\partial \mathbf{R}} \Bigg).
\end{aligned} \tag{5.13}$$

A further simplification can be made because the subsystem molecular orbitals are internally orthogonal[192] so,

$$\begin{aligned}
\frac{\partial E^{\mathrm{I}}_{\mathrm{DFT}}}{\partial \mathbf{R}} = & -2 \sum_{ij \in \mathrm{I}}^{\mathrm{occ}} S^{\mathrm{r,I}}_{ji} F^{\mathrm{I,emb}}_{ji} + \sum_{\mu\nu \in \mathrm{I}} \gamma^{\mathrm{I}}_{\mu\nu} \frac{\partial h^{\text{I-in-II}}[\gamma^{\mathrm{I}}, \gamma^{\mathrm{II}}]_{\mu\nu}}{\partial \mathbf{R}} \\
& + \frac{1}{2} \sum_{\mu\nu\lambda\sigma \in \mathrm{I}} \gamma^{\mathrm{I}}_{\mu\nu} \gamma^{\mathrm{I}}_{\lambda\sigma} \frac{\partial[(\mu\nu|\lambda\sigma) - \frac{c_x}{2}(\mu\lambda|\nu\sigma)]}{\partial \mathbf{R}} + \sum_{\mu\nu \in \mathrm{I}} \gamma^{\mathrm{I}}_{\mu\nu} \frac{\partial V^{\mathrm{I}}_{\mathrm{xc},\mu\nu}}{\partial \mathbf{R}}
\end{aligned} \tag{5.14}$$

where $F^{\mathrm{I,emb}}_{qp}$ is the embedded Fock matrix of subsystem I in MO form and $S^{\mathrm{r,I}}_{qp}$ is

$$S^{\mathrm{r,I}}_{qp} = \sum_{\mu\nu \in \mathrm{I}} C^{\mathrm{I}*}_{\mu q} \frac{\partial S^{\mathrm{I}}_{\mu\nu}}{\partial \mathbf{R}} C^{\mathrm{I}}_{\nu p}. \tag{5.15}$$

While most of these terms may be solved using standard gradient methods, the gradient of the subsystem electron density present in $\frac{\partial \mathbf{h}^{\text{I-in-II}}[\gamma^{\mathrm{I}}, \gamma^{\mathrm{II}}]}{\partial \mathbf{R}}$ must be derived and implemented for absolutely-localized Huzinaga projection embedding. This term contains the

gradient of subsystem density matrices, which may be written,

$$\frac{\partial \gamma^{\mathrm{I}}_{\mu\nu}}{\partial \mathbf{R}} = 2\sum_{i}^{\mathrm{occ}} \left(\frac{\partial C^{\mathrm{I}*}_{\mu i}}{\partial \mathbf{R}} C^{\mathrm{I}}_{\nu i} + C^{\mathrm{I}*}_{\mu i} \frac{\partial C^{\mathrm{I}}_{\nu i}}{\partial \mathbf{R}} \right). \tag{5.16}$$

The gradients of the MO coefficients may be written with a response term as,

$$\frac{\partial C^{\mathrm{I}}_{\mu i}}{\partial \mathbf{R}} = \sum_{m \in \mathrm{I}} U^{\mathrm{I}}_{mi} C^{\mathrm{I}}_{\mu m} \tag{5.17}$$

The response term U^{I}_{mi} must be determined by solving the coupled-perturbed equations.

5.2.1 Coupled-perturbed Solution

We can differentiate the embedded Fock matrix with respect to atomic coordinate $\mathbf{R}$ to get

$$\begin{aligned} \frac{\partial F^{\mathrm{I,emb}}_{qp}}{\partial \mathbf{R}} = &\sum_{r \in \mathrm{I}} (U^{\mathrm{I}}_{rq} F^{\mathrm{I,emb}}_{rp} + U^{\mathrm{I}}_{rp} F^{\mathrm{I,emb}}_{qr}) + F^{\mathrm{r,I,emb}}_{qp} \\ &+ \sum_{r \in \mathrm{I}} \sum_{i \in \mathrm{I}}^{\mathrm{occ}} U^{\mathrm{I}}_{ri} A^{\mathrm{'I,I}}_{qp,ri} + \sum_{r \in \mathrm{II}} \sum_{i \in \mathrm{II}}^{\mathrm{occ}} U^{\mathrm{II}}_{ri} A^{\mathrm{'I,II}}_{qp,ri}. \end{aligned} \tag{5.18}$$

Terms with superscript r contain gradient terms without U, so

$$F^{\mathrm{r,I,emb}}_{qp} = h^{\mathrm{r,I\text{-}in\text{-}II}}_{qp} + \sum_{\mu\nu \in \mathrm{I}} C^{\mathrm{I}*}_{\mu q} \sum_{\lambda\sigma \in \mathrm{I}} \gamma^{\mathrm{I}}_{\lambda\sigma} \frac{\partial [(\mu\nu|\lambda\sigma) - c_x(\mu\lambda|\nu\sigma)]}{\partial \mathbf{R}} C^{\mathrm{I}}_{\nu p} + V^{\mathrm{r,I}}_{\mathrm{xc},qp}, \tag{5.19}$$

and

$$\begin{aligned} V^{\mathrm{r,I}}_{\mathrm{xc},qp} = &\sum_{\mu\nu \in \mathrm{I}} C^{\mathrm{I}*}_{\mu q} \left(\int v^{\mathrm{I}}_{\mathrm{xc}}(\vec{r}) \frac{\partial \phi^{\mathrm{I}}_{\mu}(\vec{r}) \phi^{\mathrm{I}}_{\nu}(\vec{r})}{\partial \mathbf{R}} d\vec{r} \right. \\ &\left. + \sum_{\lambda\sigma \in \mathrm{I}} \gamma^{\mathrm{I}}_{\lambda\sigma} \int\int \phi^{\mathrm{I}}_{\mu}(\vec{r}) \phi^{\mathrm{I}}_{\nu}(\vec{r}) f^{\mathrm{I}}_{\mathrm{xc}}(\vec{r}, \vec{r}\,') \frac{\partial \phi^{\mathrm{I}}_{\lambda}(\vec{r}\,') \phi^{\mathrm{I}}_{\sigma}(\vec{r}\,')}{\partial \mathbf{R}} d\vec{r} d\vec{r}\,' \right) C^{\mathrm{I}}_{\nu p}. \end{aligned} \tag{5.20}$$

Here the exchange-correlation kernel, $f_{\rm xc}^{\rm I}(\vec{r}\,\vec{r}\,')$, results from the chain rule,

$$f_{\rm xc}^{\rm I}[\rho^{\rm I}](\vec{r},\vec{r}\,') = \frac{\partial v_{\rm xc}^{\rm I}[\rho^{\rm I}](\vec{r})}{\partial \rho^{\rm I}(\vec{r}\,')}. \tag{5.21}$$

Finally the derivative terms for the embedded core Hamiltonian that do not contain U may be written,

$$\begin{aligned} h_{qp}^{\rm r,I\text{-}in\text{-}II} = \sum_{\mu\nu\in {\rm I}} C_{\mu q}^{\rm I*} \Bigg(\frac{\partial h_{\mu\nu}}{\partial \mathbf{R}} + \sum_{\lambda\sigma} (\gamma^{\rm I}+\gamma^{\rm II})_{\lambda\sigma} \frac{\partial[(\mu\nu|\lambda\sigma) - c_x(\mu\lambda|\nu\sigma)]}{\partial \mathbf{R}} \\ - \sum_{\lambda\sigma\in {\rm I}} \gamma_{\lambda\sigma}^{\rm I} \frac{\partial[(\mu\nu|\lambda\sigma) - c_x(\mu\lambda|\nu\sigma)]}{\partial \mathbf{R}} \Bigg) C_{\nu p}^{\rm I} + V_{{\rm xc},qp}^{\rm r,I+II} - V_{{\rm xc},qp}^{\rm r,I} + P_{qp}^{\rm r,I\text{-}in\text{-}II}, \end{aligned} \tag{5.22}$$

and the projection terms are,

$$P_{qp}^{\rm r,I\text{-}in\text{-}II} = -\sum_{i\in {\rm II}}^{\rm occ} (F_{qi}^{\rm r} S_{ip} + F_{qi} S_{ip}^{\rm r} + S_{pi}^{\rm r} F_{iq} + S_{pi} F_{iq}^{\rm r}). \tag{5.23}$$

Here

$$S_{rp}^{\rm r} = \sum_{\mu\in {\rm II}\nu\in {\rm I}} C_{\mu r}^{\rm II*} \frac{\partial S_{\mu\nu}}{\partial \mathbf{R}} C_{\nu p}^{\rm I} \tag{5.24}$$

and

$$\mathbf{F}^{\rm r} = \mathbf{h}^{\rm r} + \mathbf{G}^{\rm r,\ I\text{-}in\text{-}II}[\gamma^{\rm I}+\gamma^{\rm II}] \tag{5.25}$$

where,

$$h_{qr}^{\rm r} = \sum_{\mu\in {\rm I}\nu\in {\rm II}} C_{\mu q}^{\rm I*} \frac{\partial h_{\mu\nu}}{\partial \mathbf{R}} C_{\nu r}^{\rm II}, \tag{5.26}$$

and

$$G^{\rm r,I\text{-}in\text{-}II}[\gamma^{\rm I}+\gamma^{\rm II}]_{qr} = \sum_{\mu\nu\in {\rm I}} C_{\mu q}^{\rm I*} \left(\sum_{\lambda\sigma} (\gamma^{\rm I}+\gamma^{\rm II})_{\lambda\sigma} \frac{\partial[(\mu\nu|\lambda\sigma) - c_x(\mu\lambda|\nu\sigma)]}{\partial \mathbf{R}} \right) C_{\nu r}^{\rm II} + V_{{\rm xc},qr}^{\rm r,I+II}. \tag{5.27}$$

The terms associated with U are,

$$\begin{aligned} A^{\text{'I,I}}_{qp,ri} &= 4(qp|ri) - c_x[(qr|pi) + (qi|pr)] + 4(qp|f_{xc}^{I+II}|ri) \\ -\frac{1}{2}\sum_{j\in \text{II}}^{\text{occ}} \big\{ &[4(qj|ri) - c_x[(qr|ji) + (qi|jr)] + 4(qj|f_{xc}^{I+II}|ri)]S_{jp} \\ &+S_{pj}[4(jq|ri) - c_x[(jr|qi) + (ji|qr)] + 4(jq|f_{xc}^{I+II}|ri)]\big\}, \end{aligned} \tag{5.28}$$

and

$$\begin{aligned} A^{\text{'I,II}}_{qp,ri} &= 4(qp|ri) - c_x[(qr|pi) + (qi|pr)] + 4(qp|f_{xc}^{I+II}|ri) \\ -\frac{1}{2}\sum_{j\in \text{II}}^{\text{occ}} \big\{ &[4(qj|ri) - c_x[(qr|ji) + (qi|jr)] + 4(qj|f_{xc}^{I+II}|ri)]S_{jp} \\ &+S_{pj}[4(jq|ri) - c_x[(jr|qi) + (ji|qr)] + 4(jq|f_{xc}^{I+II}|ri)]\big\} \\ &-(F_{qr}S_{ip} + F_{qi}S_{rp} + S_{pr}F_{iq} + S_{pi}F_{rq}). \end{aligned} \tag{5.29}$$

where $\mathbf{F}$ is the full system Fock matrix.

We can do an additional simplification after recognizing that for subsystems with orthogonal orbitals

$$U^{\text{I}}_{ji} + U^{\text{I}}_{ij} = -S^{\text{r,I}}_{ij}, \tag{5.30}$$

so we can rewrite equation 5.18,

$$\begin{aligned}
\frac{\partial F_{qp}^{\mathrm{I,emb}}}{\partial \mathbf{R}} &= (\epsilon_q^{\mathrm{I,emb}} - \epsilon_p^{\mathrm{I,emb}})U_{qp}^{\mathrm{I}} - \epsilon_p^{\mathrm{I,emb}} S_{qp}^{\mathrm{r,I}} + F_{qp}^{\mathrm{r,I,emb}} \\
&\quad - \sum_{ij\in \mathrm{I}}^{\mathrm{occ}} S_{ji}^{\mathrm{r,I}} \Bigg\{ 2(qp|ji) - c_x(qj|pi) + 2(qp|f_{xc}^{I+II}|ji) \\
&\qquad - \sum_{k\in \mathrm{II}}^{\mathrm{occ}} [2(qk|ji) - c_x(qj|ki) + 2(qk|f_{xc}^{I+II}|ji)] S_{kp} \\
&\quad - \sum_{k\in \mathrm{II}}^{\mathrm{occ}} S_{pk}[2(kq|ji) - c_x(kj|qi) + 2(kq|f_{xc}^{I+II}|ji)] \Bigg\} \\
&\qquad + \sum_{a\in \mathrm{I}}^{\mathrm{vir}} \sum_{i\in \mathrm{I}}^{\mathrm{occ}} U_{ai}^{\mathrm{I}} A'^{\mathrm{I,I}}_{qp,ai} \\
&\quad - \sum_{ij\in \mathrm{II}}^{\mathrm{occ}} S_{ji}^{\mathrm{r,II}} \Bigg\{ 2(qp|ji) - c_x(qj|pi) + 2(qp|f_{xc}^{I+II}|ji) \\
&\qquad - \sum_{k\in \mathrm{II}}^{\mathrm{occ}} [2(qk|ji) - c_x(qj|ki) + 2(qk|f_{xc}^{I+II}|ji)] S_{kp} \\
&\quad - \sum_{k\in \mathrm{II}}^{\mathrm{occ}} S_{pk}[2(kq|ji) - c_x(kj|qi) + 2(kq|f_{xc}^{I+II}|ji)] - (F_{qj}S_{ip} + S_{pj}F_{iq}) \Bigg\} \\
&\qquad + \sum_{a\in \mathrm{II}}^{\mathrm{vir}} \sum_{i\in \mathrm{II}}^{\mathrm{occ}} U_{ai}^{\mathrm{II}} A'^{\mathrm{I,II}}_{qp,ai}
\end{aligned} \tag{5.31}$$

where $\epsilon_q^{\mathrm{I,emb}}$ is the subsystem I orbital energy of MO q. Returning to the calculation of the gradient of the electron density, we can write it in terms of U as,

$$\begin{aligned}
\frac{\partial \gamma_{\mu\nu}^{\mathrm{I}}}{\partial \mathbf{R}} &= 2\sum_{i}^{\mathrm{occ}} \sum_{r} [U_{ri}^{\mathrm{I}} C_{\mu r}^{\mathrm{I}*} C_{\nu i}^{\mathrm{I}} + U_{ri}^{\mathrm{I}} C_{\mu i}^{\mathrm{I}*} C_{\nu r}^{\mathrm{I}}] \\
&= 2\sum_{ij}^{\mathrm{occ}} [U_{ji}^{\mathrm{I}} C_{\mu i}^{\mathrm{I}*} C_{\nu j}^{\mathrm{I}} + U_{ij}^{\mathrm{I}} C_{\mu i}^{\mathrm{I}*} C_{\nu j}^{\mathrm{I}}] + \sum_{i}^{\mathrm{occ}} \sum_{a}^{\mathrm{vir}} [U_{ai}^{\mathrm{I}} C_{\mu a}^{\mathrm{I}*} C_{\nu i}^{\mathrm{I}} + U_{ai}^{\mathrm{I}} C_{\mu i}^{\mathrm{I}*} C_{\nu a}^{\mathrm{I}}] \\
&= -2\sum_{ij}^{\mathrm{occ}} C_{\mu i}^{\mathrm{I}*} S_{ij}^{\mathrm{r,I}} C_{\nu j}^{\mathrm{I}} + 2\sum_{i}^{\mathrm{occ}} \sum_{a}^{\mathrm{vir}} [U_{ai}^{\mathrm{I}} C_{\mu a}^{\mathrm{I}*} C_{\nu i}^{\mathrm{I}} + U_{ai}^{\mathrm{I}} C_{\mu i}^{\mathrm{I}*} C_{\nu a}^{\mathrm{I}}]
\end{aligned} \tag{5.32}$$

From this equation it is clear that only need to solve for U_{ai}^{X} to determine the unknown gradient of the density matrix. Therefore we can formulate a matrix problem as,

$$\mathbf{AU} = \mathbf{B} \tag{5.33}$$

where

$$\begin{aligned} A_{ai,bj}^{\mathrm{I,I}} = \delta_{ab}\delta_{ij}(\epsilon_i^{\mathrm{I}} - \epsilon_a^{\mathrm{I}}) - \Big[& 4(ai|bj) - c_x[(ab|ij) + (aj|ib)] + 4(ai|f_{xc}^{I+II}|bj) \\ & -\frac{1}{2}\sum_{k\in\mathrm{II}}^{\mathrm{occ}} \big\{ [4(ak|bj) - c_x[(ab|kj) + (aj|kb)] + 4(ak|f_{xc}^{I+II}|bj)]S_{ki} \\ & + S_{ik}[4(ka|bj) - c_x[(kb|aj) + (kj|ab)] + 4(ka|f_{xc}^{I+II}|bj)] \big\} \Big], \end{aligned} \tag{5.34}$$

$$\begin{aligned} A_{ai,bj}^{\mathrm{I,II}} = - \Big[& 4(ai|bj) - c_x[(ab|ij) + (aj|ib)] + 4(ai|f_{xc}^{I+II}|bj) \\ & -\frac{1}{2}\sum_{k\in\mathrm{II}}^{\mathrm{occ}} \big\{ [4(ak|bj) - c_x[(ab|kj) + (aj|kb)] + 4(ak|f_{xc}^{I+II}|bj)]S_{ki} \\ & + S_{ik}[4(ka|bj) - c_x[(kb|aj) + (kj|ab)] + 4(ka|f_{xc}^{I+II}|bj)] \big\} \\ & - (F_{ab}S_{ji} + F_{aj}S_{bi} + S_{ib}F_{ja} + S_{ij}F_{ba}) \Big], \end{aligned} \tag{5.35}$$

and

$$
\begin{aligned}
B^{\mathrm{I}}_{0,ai} = F^{\mathrm{r,I,emb}}_{ai} - \epsilon^{\mathrm{I,emb}}_{i} S^{\mathrm{r},I}_{ai} \\
- \sum_{jk\in \mathrm{I}}^{\mathrm{occ}} S^{\mathrm{r,I}}_{kj} \Bigg\{ 2(ai|kj) - c_x(ak|ij) + 2(ai|f^{I+II}_{xc}|kj) \\
- \sum_{m\in \mathrm{II}}^{\mathrm{occ}} [2(am|kj) - c_x(ak|mj) + 2(am|f^{I+II}_{xc}|kj)] S_{mi} \\
- \sum_{m\in \mathrm{II}}^{\mathrm{occ}} S_{im} [2(ma|kj) - c_x(mk|aj) + 2(ma|f^{I+II}_{xc}|kj)] \Bigg\} \\
- \sum_{jk\in \mathrm{II}}^{\mathrm{occ}} S^{\mathrm{r,II}}_{kj} \Bigg\{ 2(ai|kj) - c_x(ak|ij) + 2(ai|f^{I+II}_{xc}|kj) \\
- \sum_{m\in \mathrm{II}}^{\mathrm{occ}} [2(am|kj) - c_x(ak|mj) + 2(am|f^{I+II}_{xc}|kj)] S_{mi} \\
- \sum_{m\in \mathrm{II}}^{\mathrm{occ}} S_{im} [2(ma|kj) - c_x(mk|aj) + 2(ma|f^{I+II}_{xc}|kj)] - (F_{ak} S_{ji} + S_{ik} F_{ja}) \Bigg\}
\end{aligned}
\tag{5.36}
$$

This matrix problem may be solved using any choice of matrix inversion program. Once $\mathbf{U}$ is known, the solution to the gradient of the embedded core Hamiltonian may be easily determined. Furthermore, the analytical nuclear gradients for WF-in-Hartree-Fock (WF-in-HF) theory are easily determined by setting the c_x parameter above to 1 and excluding any exchange-correlation terms.

5.3 Computational Details and Results

We implemented analytical nuclear gradients for absolutely localized Huzinaga projection WF-in-DFT embedding within our open source QSoME software, which uses packages from the computational chemistry package PySCF[111]. In order to validate our gradients, we compared our analytical nuclear gradients for a distorted ethanol geometry used in a previous study by Miller[41], to the numerical gradients calculated using a four-point central difference formula. For our embedding calculations, the WF region consisted of the OH comprising the hydroxyl. Apparent from Table 5.1, the differences between the analytical embedding gradient and numerical gradients are small and on

Method	MAE (hartree/bohrs)
HF	1.05433E-06
CCSD-in-HF	3.01538E-06
LDA	2.4955E-06
CCSD-in-LDA	1.25203E-06

Table 5.1: Maximum absolute error (MAE) of the difference between numerical and analytical nuclear gradients for distorted ethanol. All calculations done using 6-31g basis and a grid level of 5 within PySCF. In the case of Method HF or LDA, this indicates the entire system calculated using that method without embedding.

the same order as the analytical gradients already implemented in PySCF. These results indicate we have correctly implemented WF-in-DFT embedded analytical nuclear gradients and will be performing additional calculations in the near future to understand how accurate our method is for other chemical properties such as geometry optimizations and transition state searches.

Chapter 6

Conclusion and Future Work

The development of absolutely localized Huzinaga projection based WF-in-DFT embedding has seen significant progress in recent years as demonstrated by the work presented here. We have shown that the method is systematically improvable or put another way, the method provides a clear strategy for improving the accuracy of calculations. Furthermore, our Huzinaga WF-in-DFT embedding method can successfully incorporate a wide variety of WF and DFT methods within our embedding framework. This flexibility allows the method to be applicable when modeling many different kinds of chemical systems and properties. Also of importance, our method demonstrates WF method level accuracy for a fraction of the computational cost of the full system WF calculation. Given all of these beneficial features, the absolutely localized Huzinaga WF-in-DFT embedding method has significant potential to be a widely used method for accurately modeling complex chemical systems.

We suggest further development of the absolutely localized Huzinaga projection based WF-in-DFT embedding method should take place in several key areas. First, occasionally within this work we encountered a system that proved challenging for DFT-in-DFT freeze and thaw convergence. In order to achieve accurate results, the DFT-in-DFT calculations must be converged to a consistent energy minimum. Improving the freeze-and -thaw convergence through the use of dynamic damping, or a variation of a convergence acceleration algorithm could improve the scope of the method, in addition to reducing the computational cost of calculation convergence.

Second, work thus far has been focused on demonstrating the accuracy of the method

when compared to systems that we could calculate the full WF result as a point of comparison. Now that we have demonstrated the accuracy of our method, we should apply the method to much larger open problems within the chemistry community. Importantly, collaborating with experimental and computational chemists to understand a system which was previously beyond the scope of modern computational chemistry would be an exciting next step.

Finally, encouraging broad acceptance and utilization of our method is key to the community actually reaping the benefits of the improvements to accuracy and cost savings that our method provides. One of the best ways to invite other researchers to use our method is to make it accessible and easy to use. The method is already open source and documented on Github as QSoME, however implementing the method within other existing computational chemistry packages lowers the barrier for other researchers to apply our method to their specific problems. Implementing our method in other packages has the added benefit of reducing computational cost further by taking advantage of DFT and WF cost reduction features already implemented in other packages.

Absolutely localized Huzinaga WF-in-DFT embedding has been shown to be reliable, accurate and practical for many chemical problems. Our method provides a new computational tool for navigating the balance between accuracy and computational cost and we are excited to see what new insights the method will uncover in the future.

References

[1] Trygve Helgaker, Wim Klopper, Henrik Koch, and Jozef Noga. Basis-set convergence of correlated calculations on water. *The Journal of Chemical Physics*, 106(23):9639–9646, 6 1997.

[2] Asger Halkier, Trygve Helgaker, Poul Jørgensen, Wim Klopper, Henrik Koch, Jeppe Olsen, and Angela K. Wilson. Basis-set convergence in correlated calculations on Ne, N2, and H2O. *Chemical Physics Letters*, 286(3-4):243–252, 4 1998.

[3] Yan Zhao and Donald G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theoretical Chemistry Accounts*, 120(1-3):215–241, 5 2008.

[4] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140(4A):A1133–A1138, 11 1965.

[5] E. Fermi. A statistical method for the determination of some atomic properties and the application of this method to the theory of the periodic system of elements. *Zeitschrift fur Physik*, 48:73–79, 1928.

[6] Aron J. Cohen, P. Mori-Sanchez, and Weitao Yang. Insights into Current Limitations of Density Functional Theory. *Science*, 321(5890):792–794, 8 2008.

[7] R. O. Jones. Density functional theory: Its origins, rise to prominence, and future. *Reviews of Modern Physics*, 87(3):897–923, 8 2015.

[8] Haoyu S Yu, Shaohong L Li, and Donald G. Truhlar. Perspective: Kohn-Sham density functional theory descending a staircase. *The Journal of Chemical Physics*, 145(13):130901, 10 2016.

[9] Aron J Cohen, Paula Mori-Sánchez, and Weitao Yang. Challenges for Density Functional Theory. *Chemical Reviews*, 112(1):289–320, 1 2012.

[10] Klaus A Moltved and Kasper P Kepp. The Metal Hydride Problem of Computational Chemistry: Origins and Consequences. *The Journal of Physical Chemistry A*, 123(13):2888–2900, 4 2019.

[11] Abhik Ghosh and Peter R Taylor. High-level ab initio calculations on the energetics of low-lying spin states of biologically relevant transition metal complexes: a first progress report. *Current Opinion in Chemical Biology*, 7(1):113–124, 2 2003.

[12] Jeremy N. Harvey. DFT Computation of Relative Spin-State Energetics of Transition Metal Compounds. In *Structure and Bonding*, volume 112, pages 151–184. Springer, Berlin, Heidelberg, 2004.

[13] Abhik Ghosh. Transition metal spin state energetics and noninnocent systems: challenges for DFT in the bioinorganic arena. *JBIC Journal of Biological Inorganic Chemistry*, 11(6):712–724, 9 2006.

[14] Thomas F. Hughes and Richard A. Friesner. Correcting Systematic Errors in DFT Spin-Splitting Energetics for Transition Metal Complexes. *Journal of Chemical Theory and Computation*, 7(1):19–32, 1 2011.

[15] Latévi Max Lawson Daku, Francesco Aquilante, Timothy W. Robinson, and Andreas Hauser. Accurate Spin-State Energetics of Transition Metal Complexes. 1. CCSD(T), CASPT2, and DFT Study of [M(NCH) 6] 2+ (M = Fe, Co). *Journal of Chemical Theory and Computation*, 8(11):4216–4231, 11 2012.

[16] Mariusz Radoń. Revisiting the role of exact exchange in DFT spin-state energetics of transition metal complexes. *Phys. Chem. Chem. Phys.*, 16(28):14479–14488, 7 2014.

[17] Mariusz Radoń. Spin-State Energetics of Heme-Related Models from DFT and Coupled Cluster Calculations. *Journal of Chemical Theory and Computation*, 10(6):2306–2321, 6 2014.

[18] Qiming Sun and Garnet Kin-Lic Chan. Quantum Embedding Theories. *Accounts of Chemical Research*, 49(12):2705–2712, 12 2016.

[19] Adam Wasserman and Michele Pavanello. Quantum embedding electronic structure methods. *International Journal of Quantum Chemistry*, 120(21), 11 2020.

[20] Leighton O. Jones, Martín A. Mosquera, George C. Schatz, and Mark A. Ratner. Embedding Methods for Quantum Chemistry: Applications from Materials to Life Sciences. *Journal of the American Chemical Society*, 142(7):3281–3295, 2 2020.

[21] Christoph R. Jacob and Johannes Neugebauer. Subsystem density-functional theory. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(4):325–362, 7 2014.

[22] Bryan T. G. Lau, Gerald Knizia, and Timothy C. Berkelbach. Regional Embedding Enables High-Level Quantum Chemistry for Surface Science. *The Journal of Physical Chemistry Letters*, 12(3):1104–1109, 1 2021.

[23] Wenfei Li, Ming Chen, Eran Rabani, Roi Baer, and Daniel Neuhauser. Stochastic embedding DFT: Theory and application to p -nitroaniline in water. *The Journal of Chemical Physics*, 151(17):174115, 11 2019.

[24] Peter V. Sushko, Chen Huang, Niranjan Govind, and Karol Kowalski. Chapter 4. Embedding Methods in Materials Discovery. In *Computational Materials Discovery*, pages 87–116. Royal Society of Chemistry, 10 2018.

[25] Moritz Bensberg and Johannes Neugebauer. Density functional theory based embedding approaches for transition-metal complexes †. *Phys. Chem. Chem. Phys*, 22(45):26093, 2020.

[26] L H Thomas. The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23(05):542, 1 1927.

[27] C. F v Weizsäcker. Zur Theorie der Kernmassen. *Zeitschrift für Physik*, 96(7-8):431–458, 7 1935.

[28] Jason D. Goodpaster, Nandini Ananth, Frederick R. Manby, and Thomas F. Miller. Exact nonadditive kinetic potentials for embedded density functional theory. *The Journal of Chemical Physics*, 133(8):084103, 8 2010.

[29] Samuel Fux, Karin Kiewisch, Christoph R. Jacob, Johannes Neugebauer, and Markus Reiher. Analysis of electron density distributions from subsystem density functional theory applied to coordination bonds. *Chemical Physics Letters*, 461(4-6):353–359, 8 2008.

[30] Andreas W. Götz, S. Maya Beyhan, and Lucas Visscher. Performance of Kinetic Energy Functionals for Interaction Energies in a Subsystem Formulation of Density Functional Theory. *Journal of Chemical Theory and Computation*, 5(12):3161–3174, 12 2009.

[31] S. Maya Beyhan, Andreas W. Götz, Christoph R. Jacob, and Lucas Visscher. The weak covalent bond in NgAuF (Ng=Ar, Kr, Xe): A challenge for subsystem density functional theory. *The Journal of Chemical Physics*, 132(4):044114, 1 2010.

[32] Johannes Neugebauer, Manuel J. Louwerse, Evert Jan Baerends, and Tomasz A. Wesolowski. The merits of the frozen-density embedding scheme to model solvatochromic shifts. *The Journal of Chemical Physics*, 122(9):094115, 3 2005.

[33] M. Levy. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proceedings of the National Academy of Sciences of the United States of America*, 76(12):6062–5, 12 1979.

[34] Chen Huang, Michele Pavone, and Emily A. Carter. Quantum mechanical embedding theory based on a unique embedding potential. *The Journal of Chemical Physics*, 134(15):154110, 4 2011.

[35] Rollin A King and Nicholas C Handy. Kinetic energy functionals from the Kohn–Sham potential. *Physical Chemistry Chemical Physics*, 2(22):5049–5056, 2000.

[36] Frederick R. Manby, Martina Stella, Jason D. Goodpaster, and Thomas F. Miller. A Simple, Exact Density-Functional-Theory Embedding Scheme. *Journal of Chemical Theory and Computation*, 8(8):2564–2568, 8 2012.

[37] Alon Chapovetsky, Matthew Welborn, John M. Luna, Ralf Haiges, Thomas F. Miller, and Smaranda C. Marinescu. Pendant Hydrogen-Bond Donors in Cobalt Catalysts Independently Enhance CO 2 Reduction. *ACS Central Science*, 4(3):397–404, 3 2018.

[38] Jason D. Goodpaster, Taylor A. Barnes, Frederick R. Manby, and Thomas F. Miller. Density functional theory embedding for correlated wavefunctions: Improved methods for open-shell systems and transition metal complexes. *The Journal of Chemical Physics*, 137(22):224113, 12 2012.

[39] Jason D. Goodpaster, Taylor A. Barnes, Frederick R. Manby, and Thomas F. Miller. Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *The Journal of Chemical Physics*, 140(18):18A507, 5 2014.

[40] Matthew Welborn, Frederick R. Manby, and Thomas F. Miller. Even-handed subsystem selection in projection-based embedding. *The Journal of Chemical Physics*, 149(14):144101, 10 2018.

[41] Sebastian J. R. Lee, Feizhi Ding, Frederick R. Manby, and Thomas F. Miller. Analytical gradients for projection-based wavefunction-in-DFT embedding. *The Journal of Chemical Physics*, 151(6):064112, 8 2019.

[42] Daniel Claudino and Nicholas J Mayhall. Simple and Efficient Truncation of Virtual Spaces in Embedded Wave Functions via Concentric Localization. *Journal of Chemical Theory and Computation*, 15(11):6085–6096, 11 2019.

[43] Daniel Claudino and Nicholas J. Mayhall. Automatic Partition of Orbital Spaces Based on Singular Value Decomposition in the Context of Embedding Theories. *Journal of Chemical Theory and Computation*, 15(2):1053–1064, 2 2019.

[44] Bence Hégely, Péter R. Nagy, György G. Ferenczy, and Mihály Kállay. Exact density functional and wave function embedding schemes based on orbital localization. *The Journal of Chemical Physics*, 145(6):064107, 8 2016.

[45] Moritz Bensberg and Johannes Neugebauer. Orbital alignment for accurate projection-based embedding calculations along reaction paths. *Journal of Chemical Theory and Computation*, 16(6):3607–3619, 6 2020.

[46] Moritz Bensberg and Johannes Neugebauer. Automatic basis-set adaptation in projection-based embedding. *The Journal of Chemical Physics*, 150(18):184104, 5 2019.

[47] Dhabih V. Chulhai and Jason D. Goodpaster. Improved Accuracy and Efficiency in Quantum Embedding through Absolute Localization. *Journal of Chemical Theory and Computation*, 13(4):1503–1508, 4 2017.

[48] Dhabih V. Chulhai and Jason D. Goodpaster. Projection-Based Correlated Wave Function in Density Functional Theory Embedding for Periodic Systems. *Journal of Chemical Theory and Computation*, 14(4):1928–1942, 4 2018.

[49] Thomas Dresselhaus and Johannes Neugebauer. Part and whole in wavefunction/DFT embedding. *Theoretical Chemistry Accounts*, 134(8):97, 8 2015.

[50] Sebastian J.R. R. Lee, Matthew Welborn, Frederick R. Manby, and Thomas F. Miller. Projection-Based Wavefunction-in-DFT Embedding. *Accounts of Chemical Research*, 52(5):acs.accounts.8b00672, 4 2019.

[51] Simon J. Bennie, Martina Stella, Thomas F. Miller, and Frederick R. Manby. Accelerating wavefunction in density-functional-theory embedding by truncating the active basis set. *The Journal of Chemical Physics*, 143(2):024105, 7 2015.

[52] S Huzinaga and A A Cantu. Theory of Separability of Many-Electron Systems. *The Journal of Chemical Physics*, 55(12):5543–5549, 12 1971.

[53] E Francisco, A. Martín Pendás, and W H Adams. Generalized Huzinaga building-block equations for nonorthogonal electronic groups: Relation to the Adams–Gilbert theory. *The Journal of Chemical Physics*, 97(9):6504–6508, 11 1992.

[54] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871, 11 1964.

[55] Jiří Čížek. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *The Journal of Chemical Physics*, 45(11):4256–4266, 12 1966.

[56] J. Čížek and J. Paldus. Correlation problems in atomic and molecular systems III. Rederivation of the coupled-pair many-electron theory using the traditional quantum chemical methodst. *International Journal of Quantum Chemistry*, 5(4):359–379, 7 1971.

[57] Hans-Joachim Werner and Wilfried Meyer. A quadratically convergent multiconfiguration–self-consistent field method with simultaneous optimization of orbitals and CI coefficients. *The Journal of Chemical Physics*, 73(5):2342–2356, 9 1980.

[58] Marcel Swart, Miquel Solà, and F. Matthias Bickelhaupt. Energy landscapes of nucleophilic substitution reactions: A comparison of density functional theory and coupled cluster methods. *Journal of Computational Chemistry*, 28(9):1551–1560, 7 2007.

[59] Miroslav Rubeš, Lukáš Grajciar, Ota Bludský, Andrew D. Wiersum, Philip L. Llewellyn, and Petr Nachtigall. Combined Theoretical and Experimental Investigation of CO Adsorption on Coordinatively Unsaturated Sites in CuBTC MOF. *ChemPhysChem*, 13(2):488–495, 2 2012.

[60] Arieh Warshel and Michael Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227–249, 5 1976.

[61] Mats Svensson, Stéphane Humbel, Robert D. J. Froese, Toshiaki Matsubara, Stefan Sieber, and Keiji Morokuma. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(t -Bu) 3) 2 + H 2 Oxidative Addition. *The Journal of Physical Chemistry*, 100(50):19357–19363, 1 1996.

[62] Gerald Knizia and Garnet Kin-Lic Lic Chan. Density Matrix Embedding: A Strong-Coupling Quantum Embedding Theory. *Journal of Chemical Theory and Computation*, 9(3):1428–1432, 3 2013.

[63] Feizhi Ding, Frederick R. Manby, and Thomas F. Miller. Embedded Mean-Field Theory with Block-Orthogonalized Partitioning. *Journal of Chemical Theory and Computation*, 13(4):1605–1615, 4 2017.

[64] Feizhi Ding, Takashi Tsuchiya, Frederick R. Manby, and Thomas F. Miller. Linear-Response Time-Dependent Embedded Mean-Field Theory. *Journal of Chemical Theory and Computation*, 13(9):4216–4227, 9 2017.

[65] Kaito Miyamoto, Thomas F. Miller, and Frederick R. Manby. Fock-Matrix Corrections in Density Functional Theory and Use in Embedded Mean-Field Theory. *Journal of Chemical Theory and Computation*, 12(12):5811–5822, 12 2016.

[66] Mark E. Fornace, Joonho Lee, Kaito Miyamoto, Frederick R. Manby, and Thomas F. Miller. Embedded Mean-Field Theory. *Journal of Chemical Theory and Computation*, 11(2):568–580, 2 2015.

[67] Mark E. Fornace, Joonho Lee, Kaito Miyamoto, Frederick R. Manby, and Thomas F. Miller. Correction to Embedded Mean-Field Theory. *Journal of Chemical Theory and Computation*, 11(8):3968–3968, 8 2015.

[68] Giovanni Onida, Lucia Reining, and Angel Rubio. Electronic excitations: density-functional versus many-body Green's-function approaches. *Reviews of Modern Physics*, 74(2):601–659, 6 2002.

[69] Wael Chibani, Xinguo Ren, Matthias Scheffler, and Patrick Rinke. Self-consistent Green's function embedding for advanced electronic structure methods based on a dynamical mean-field concept. *Physical Review B*, 93(16):165106, 4 2016.

[70] Peter Elliott, Kieron Burke, Morrel H. Cohen, and Adam Wasserman. Partition density-functional theory. *Physical Review A*, 82(2):024501, 8 2010.

[71] Jonathan Nafziger and Adam Wasserman. Density-Based Partitioning Methods for Ground-State Molecular Calculations. *The Journal of Physical Chemistry A*, 118(36):7623–7639, 9 2014.

[72] Sara Gómez, Jonathan Nafziger, Albeiro Restrepo, and Adam Wasserman. Partition-DFT on the water dimer. *The Journal of Chemical Physics*, 146(7):074106, 2 2017.

[73] Erik D. Hedegård and Markus Reiher. Polarizable Embedding Density Matrix Renormalization Group. *Journal of Chemical Theory and Computation*, 12(9):4242–4253, 9 2016.

[74] Adrian H. Mühlbach and Markus Reiher. Quantum system partitioning at the single-particle level. *The Journal of Chemical Physics*, 149(18):184104, 11 2018.

[75] Dalibor Hršak, Jógvan Magnus Haugaard Olsen, and Jacob Kongsted. Polarizable Density Embedding Coupled Cluster Method. *Journal of Chemical Theory and Computation*, 14(3):1351–1360, 3 2018.

[76] Tanner Culpitt, Kurt R Brorsen, Michael V Pak, and Sharon Hammes-Schiffer. Multicomponent density functional theory embedding formulation. *The Journal of Chemical Physics*, 145(4):044106, 7 2016.

[77] Kaushik D Nanda and Anna I Krylov. The effect of polarizable environment on two-photon absorption cross sections characterized by the equation-of-motion coupled-cluster singles and doubles method combined with the effective fragment potential approach. *The Journal of Chemical Physics*, 149(16):164109, 10 2018.

[78] Tomasz A. Wesolowski and Yan Alexander Wang. *Recent Progress in Orbital-free Density Functional Theory*, volume 6 of *Recent Advances in Computational Chemistry*. WORLD SCIENTIFIC, 5 2013.

[79] Tomasz A. Wesołowski. One-Electron Equations for Embedded Electron Density: Challenge for Theory and Practical Payoffs in Multi-Level Modelling of Complex

Polyatomic Systems. In *Computational Chemistry: Reviews of Current Trends*, pages 1–82. World Scientific, 6 2006.

[80] Johannes Neugebauer. Chromophore-specific theoretical spectroscopy: From subsystem density functional theory to mode-specific vibrational spectroscopy. *Physics Reports*, 489(1-3):1–87, 4 2010.

[81] Weitao Yang. Direct calculation of electron density in density-functional theory. *Physical Review Letters*, 66(11):1438–1441, 3 1991.

[82] Jin Cheng, Florian Libisch, Kuang Yu, Mohan Chen, Johannes M. Dieterich, and Emily A. Carter. Potential Functional Embedding Theory at the Correlated Wave Function Level. 1. Mixed Basis Set Embedding. *Journal of Chemical Theory and Computation*, 13(3):1067–1080, 3 2017.

[83] Florian Libisch, Chen Huang, and Emily A. Carter. Embedded Correlated Wavefunction Schemes: Theory and Applications. *Accounts of Chemical Research*, 47(9):2768–2775, 9 2014.

[84] N. Govind, Y.A. A Wang, A.J.R. J R da Silva, and E.A. A Carter. Accurate ab initio energetics of extended systems via explicit correlation embedded in a density functional environment. *Chemical Physics Letters*, 295(1-2):129–134, 10 1998.

[85] Niranjan Govind, Yan Alexander Wang, and Emily A. Carter. Electronic-structure calculations by first-principles density-based embedding of explicitly correlated systems. *The Journal of Chemical Physics*, 110(16):7677–7688, 4 1999.

[86] Sebastian Höfener and Lucas Visscher. Calculation of electronic excitations using wave-function in wave-function frozen-density embedding. *The Journal of Chemical Physics*, 137(20):204120, 11 2012.

[87] André Severo Pereira Gomes, Christoph R. Jacob, and Lucas Visscher. Calculation of local excitations in large systems by embedding wave-function theory in density-functional theory. *Physical Chemistry Chemical Physics*, 10(35):5353, 2008.

[88] Yuriy G Khait and Mark R Hoffmann. Embedding theory for excited states. *The Journal of Chemical Physics*, 133(4):044107, 7 2010.

[89] Thorsten Klüner, Niranjan Govind, Yan Wang, and Emily A. Carter. Prediction of Electronic Excited States of Adsorbates on Metal Surfaces from First Principles. *Physical Review Letters*, 86(26):5954–5957, 6 2001.

[90] T. Klüner, N. Govind, Y. A. Wang, and E. A. Carter. Klüner et al. Reply:. *Physical Review Letters*, 88(20):209702, 5 2002.

[91] André Severo Pereira Gomes and Christoph R. Jacob. Quantum-chemical embedding methods for treating local electronic excitations in complex chemical systems. *Annual Reports Section "C" (Physical Chemistry)*, 108(0):222, 5 2012.

[92] Patrick Huang and Emily A. Carter. Advances in Correlated Electronic Structure Methods for Solids, Surfaces, and Nanostructures. *Annual Review of Physical Chemistry*, 59(1):261–290, 5 2008.

[93] Pengfei Huo, Christopher Uyeda, Jason D. Goodpaster, Jonas C. Peters, and Thomas F. Miller. Breaking the Correlation between Energy Costs and Kinetic Barriers in Hydrogen Evolution via a Cobalt Pyridine-Diimine-Dioxime Catalyst. *ACS Catalysis*, 6(9):6114–6123, 9 2016.

[94] Pablo Ramos, Markos Papadakis, and Michele Pavanello. Performance of Frozen Density Embedding for Modeling Hole Transfer Reactions. *The Journal of Physical Chemistry B*, 119(24):7541–7557, 6 2015.

[95] Michele Pavanello and Johannes Neugebauer. Modelling charge transfer reactions with the frozen density embedding formalism. *The Journal of Chemical Physics*, 135(23):234103, 12 2011.

[96] Michele Pavanello. On the subsystem formulation of linear-response time-dependent DFT. *The Journal of Chemical Physics*, 138(20):204118, 5 2013.

[97] O Roncero, M. P. de Lara-Castells, P Villarreal, F Flores, J Ortega, M Paniagua, and A Aguado. An inversion technique for the calculation of embedding potentials. *The Journal of Chemical Physics*, 129(18):184104, 11 2008.

[98] Mark S. Gordon, Dmitri G. Fedorov, Spencer R. Pruitt, and Lyudmila V. Slipchenko. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chemical Reviews*, 112(1):632–672, 1 2012.

[99] Dhabih V Chulhai and Lasse Jensen. Frozen Density Embedding with External Orthogonality in Delocalized Covalent Systems. *Journal of Chemical Theory and Computation*, 11(7):3080–3088, 7 2015.

[100] Yuriy G. Khait and Mark R. Hoffmann. On the Orthogonality of Orbitals in Subsystem Kohn–Sham Density Functional Theory. In *Annual Reports in Computational Chemistry*, volume 8, pages 53–70. Elsevier Ltd, 2012.

[101] Patrick K. Tamukong, Yuriy G. Khait, and Mark R. Hoffmann. Density Differences in Embedding Theory with External Orbital Orthogonality. *The Journal of Physical Chemistry A*, 118(39):9182–9200, 10 2014.

[102] Tanner Culpitt, Kurt R Brorsen, and Sharon Hammes-Schiffer. Communication: Density functional theory embedding with the orthogonality constrained basis set expansion procedure. *The Journal of Chemical Physics*, 146(21):211101, 6 2017.

[103] Jógvan Magnus Haugaard Olsen, Casper Steinmann, Kenneth Ruud, and Jacob Kongsted. Polarizable Density Embedding: A New QM/QM/MM-Based Computational Strategy. *The Journal of Physical Chemistry A*, 119(21):5344–5355, 5 2015.

[104] Xuelan Wen, Daniel S. Graham, Dhabih V. Chulhai, and Jason D. Goodpaster. Absolutely Localized Projection-Based Embedding for Excited States. *Journal of Chemical Theory and Computation*, 16(1):385–398, 1 2020.

[105] Taylor A. Barnes, Jason D. Goodpaster, Frederick R. Manby, and Thomas F. Miller. Accurate basis set truncation for wavefunction embedding. *The Journal of Chemical Physics*, 139(2):024103, 7 2013.

[106] Bence Hégely, Péter R Nagy, and Mihály Kállay. Dual Basis Set Approach for Density Functional and Wave Function Embedding Schemes. *Journal of Chemical Theory and Computation*, 14(9):4600–4615, 9 2018.

[107] Marcin Dułak and Tomasz A. Wesołowski. Interaction energies in non-covalently bound intermolecular complexes derived using the subsystem formulation of density functional theory. *Journal of Molecular Modeling*, 13(6-7):631–642, 6 2007.

[108] M J Frisch, G W Trucks, H B Schlegel, G E Scuseria, M A Robb, J R Cheeseman, G Scalmani, V Barone, G A Petersson, H Nakatsuji, X Li, M Caricato, A V Marenich, J Bloino, B G Janesko, R Gomperts, B Mennucci, H P Hratchian, J V Ortiz, A F Izmaylov, J L Sonnenberg, D Williams-Young, F Ding, F Lipparini, F Egidi, J Goings, B Peng, A Petrone, T Henderson, D Ranasinghe, V G Zakrzewski, J Gao, N Rega, G Zheng, W Liang, M Hada, M Ehara, K Toyota, R Fukuda, J Hasegawa, M Ishida, T Nakajima, Y Honda, O Kitao, H Nakai, T Vreven, K Throssell, J A Montgomery Jr., J E Peralta, F Ogliaro, M J Bearpark, J J Heyd, E N Brothers, K N Kudin, V N Staroverov, T A Keith, R Kobayashi, J Normand, K Raghavachari, A P Rendell, J C Burant, S S Iyengar, J Tomasi, M Cossi, J M Millam, M Klene, C Adamo, R Cammi, J W Ochterski, R L Martin, K Morokuma, O Farkas, J B Foresman, and D J Fox. Gaussian16 Revision A.03, 2016.

[109] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1 1989.

[110] Kyuho Lee, William C Isley, Allison L Dzubak, Pragya Verma, Samuel J Stoneburner, Li-Chiang Lin, Joshua D Howe, Eric D Bloch, Douglas A Reed, Matthew R Hudson, Craig M Brown, Jeffrey R Long, Jeffrey B Neaton, Berend Smit, Christopher J Cramer, Donald G Truhlar, and Laura Gagliardi. Design of a Metal–Organic Framework with Enhanced Back Bonding for Separation of N 2 and CH 4. *Journal of the American Chemical Society*, 136(2):698–704, 1 2014.

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

[112] Hans-Joachim Werner, Peter J. Knowles, Gerald Knizia, Frederick R. Manby, and Martin Schütz. Molpro: a general-purpose quantum chemistry program package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(2):242–253, 3 2012.

[113] H.-J. Werner, P J Knowles, G Knizia, F R Manby, M Schütz, P Celani, W Györffy, D Kats, T Korona, R Lindh, A Mitrushenkov, G Rauhut, K R Shamasundar, T B Adler, R D Amos, S J Bennie, A Bernhardsson, A Berning, D L Cooper, M J O Deegan, A J Dobbyn, F Eckert, E Goll, C Hampel, A Hesselmann, G Hetzer, T Hrenar, G Jansen, C Köppl, S J R Lee, Y Liu, A W Lloyd, Q Ma, R A Mata, A J May, S J McNicholas, W Meyer, T F Miller III, M E Mura, A Nicklass, D P O'Neill, P Palmieri, D Peng, K Pflüger, R Pitzer, M Reiher, T Shiozaki, H Stoll, A J Stone, R Tarroni, T Thorsteinsson, M Wang, and M Welborn. MOLPRO, version 2019.2, a package of ab initio programs, 2019.

[114] Werner Győrffy, Toru Shiozaki, Gerald Knizia, and Hans-Joachim Werner. Analytical energy gradients for second-order multireference perturbation theory using density fitting. *The Journal of Chemical Physics*, 138(10):104104, 3 2013.

[115] Toru Shiozaki, Werner Győrffy, Paolo Celani, and Hans-Joachim Werner. Communication: Extended multi-state complete active space second-order perturbation theory: Energy and nuclear gradients. *The Journal of Chemical Physics*, 135(8):081106, 8 2011.

[116] Paolo Celani and Hans-Joachim Werner. Multireference perturbation theory for large restricted and selected active space reference wave functions. *The Journal of Chemical Physics*, 112(13):5546–5557, 4 2000.

[117] Tomasz Adam Wesolowski and Jacques Weber. Kohn-Sham equations with constrained electron density: an iterative evaluation of the ground-state electron density of interacting molecules. *Chemical Physics Letters*, 248(1-2):71–76, 1 1996.

[118] Jason D Goodpaster, Daniel S Graham, and Dhabih V Chulhai. Goodpaster/QSoME: Open-shell Embedding, 7 2019.

[119] Hayley R. Petras, Daniel S. Graham, Sai Kumar Ramadugu, Jason D. Goodpaster, and James J. Shepherd. Fully Quantum Embedding with Density Functional Theory for Full Configuration Interaction Quantum Monte Carlo. *Journal of Chemical Theory and Computation*, 15(10):5332–5342, 10 2019.

[120] Jason M. Gonzales, Wesley D. Allen, and Henry F. Schaefer. Model Identity S N 2 Reactions CH 3 X + X - (X = F, Cl, CN, OH, SH, NH 2 , PH 2): Marcus Theory Analyzed. *The Journal of Physical Chemistry A*, 109(46):10613–10628, 11 2005.

[121] U. Chandra Singh and Peter A. Kollman. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl + Cl- exchange reaction and gas phase protonation of polyethers. *Journal of Computational Chemistry*, 7(6):718–730, 12 1986.

[122] Ana P. de Lima Batista, Antonio G. S. de Oliveira-Filho, and Sérgio E. Galembeck. Photophysical properties and the NO photorelease mechanism of a ruthenium nitrosyl model complex investigated using the CASSCF-in-DFT embedding approach. *Physical Chemistry Chemical Physics*, 19(21):13860–13867, 5 2017.

[123] David J. Coughtrie, Robin Giereth, Daniel Kats, Hans-Joachim Werner, and Andreas Köhn. Embedded Multireference Coupled Cluster Theory. *Journal of Chemical Theory and Computation*, 14(2):693–709, 2 2018.

[124] Daniel S. Graham, Wen Xuelan, Dhabih V. Chulhai, and Jason D. Goodpaster. 2019GSEmbeddingPaperData, 2019.

[125] Per E. M. Siegbahn. The performance of hybrid DFT for mechanisms involving transition metal complexes in enzymes. *JBIC Journal of Biological Inorganic Chemistry*, 11(6):695–701, 9 2006.

[126] Miriam M. Quintal, Amir Karton, Mark A. Iron, A. Daniel Boese, and Jan M. L. Martin. Benchmark Study of DFT Functionals for Late-Transition-Metal Reactions †. *The Journal of Physical Chemistry A*, 110(2):709–716, 1 2006.

[127] Yury Minenkov, Giovanni Occhipinti, and Vidar R. Jensen. Metal-Phosphine Bond Strengths of the Transition Metals: A Challenge for DFT †. *The Journal of Physical Chemistry A*, 113(43):11833–11844, 10 2009.

[128] Alexander A. Rusakov, Sergei Iskakov, Lan Nguyen Tran, and Dominika Zgid. Self-Energy Embedding Theory (SEET) for Periodic Systems. *Journal of Chemical Theory and Computation*, 15(1):229–240, 1 2019.

[129] He Ma, Nan Sheng, Marco Govoni, and Giulia Galli. Quantum Embedding Theory for Strongly Correlated States in Materials. *Journal of Chemical Theory and Computation*, 17(4):2116–2125, 4 2021.

[130] Daniel S. Graham, Xuelan Wen, Dhabih V. Chulhai, and Jason D. Goodpaster. Robust, Accurate, and Efficient: Quantum Embedding Using the Huzinaga Level-Shift Projection Operator for Complex Systems. *Journal of Chemical Theory and Computation*, 16(4):2284–2295, 4 2020.

[131] Mariusz Radon´the Radon´. Benchmarking quantum chemistry methods for spin-state energetics of iron complexes against quantitative experimental data †. *Phys. Chem. Chem. Phys*, 4854:4854, 2019.

[132] Peter J. Knowles, Claudia Hampel, and Hans-Joachim Werner. Coupled cluster theory for high spin, open shell reference wave functions. *The Journal of Chemical Physics*, 99(7):5219–5227, 10 1993.

[133] Peter J. Knowles, Claudia Hampel, and Hans-Joachim Werner. Erratum: "Coupled cluster theory for high spin, open shell reference wave functions" [J. Chem. Phys. 99 , 5219 (1993)]. *The Journal of Chemical Physics*, 112(6):3106–3107, 2 2000.

[134] Jason D. Goodpaster, Daniel S. Graham, Dhabih V. Chulhai, and Wen Xuelan. Goodpaster/QSoME: Open-shell Embedding, 2021.

[135] Jordi Cirera, Mireia Via-Nadal, and Eliseo Ruiz. Benchmarking Density Functional Methods for Calculation of State Energies of First Row Spin-Crossover Molecules. *Inorganic Chemistry*, 57(22):14097–14105, 11 2018.

[136] Dayou Zhang and Donald G. Truhlar. Spin Splitting Energy of Transition Metals: A New, More Affordable Wave Function Benchmark Method and Its Use to Test Density Functional Theory. *Journal of Chemical Theory and Computation*, 16(7):4416–4428, 7 2020.

[137] 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. *The Journal of Chemical Physics*, 131(5):054106, 8 2009.

[138] Deidre Cleland. *The initiator Full Configuration Interaction Quantum Monte Carlo method : Development and applications to molecular systems*. PhD thesis, University of Cambridge, 2012.

[139] Lauretta R. Schwarz, George H. Booth, and Ali Alavi. Insights into the structure of many-electron wave functions of Mott-insulating antiferromagnets: The three-band Hubbard model in full configuration interaction quantum Monte Carlo. *Physical Review B*, 91(4):045139, 1 2015.

[140] J. S. Spencer, N. S. Blunt, and W. M.C. Foulkes. The sign problem and population dynamics in the full configuration interaction quantum Monte Carlo method. *The Journal of Chemical Physics*, 136(5):054110, 2 2012.

[141] James J. Shepherd, George H. Booth, and Ali Alavi. Investigation of the full configuration interaction quantum Monte Carlo method using homogeneous electron gas models. *The Journal of Chemical Physics*, 136(24):244101, 6 2012.

[142] James J Shepherd, George Booth, Andreas Grüneis, and Ali Alavi. Full configuration interaction perspective on the homogeneous electron gas. *Physical Review B*, 85(8):081103, 2 2012.

[143] George H. Booth, Deidre Cleland, Alex J. W. Thom, and Ali Alavi. Breaking the carbon dimer: The challenges of multiple bond dissociation with full configuration interaction quantum Monte Carlo methods. *The Journal of Chemical Physics*, 135(8):084104, 8 2011.

[144] George H. Booth, Simon D. Smart, and Ali Alavi. Linear-scaling and parallelisable algorithms for stochastic quantum chemistry. *Molecular Physics*, 112(14):1855–1869, 7 2014.

[145] Giovanni Li Manni, Simon D. Smart, and Ali Alavi. Combining the Complete Active Space Self-Consistent Field Method and the Full Configuration Interaction Quantum Monte Carlo within a Super-CI Framework, with Application to Challenging Metal-Porphyrins. *Journal of Chemical Theory and Computation*, 12(3):1245–1258, 3 2016.

[146] George H. Booth, Andreas Grüneis, Georg Kresse, and Ali Alavi. Towards an exact description of electronic wavefunctions in real solids. *Nature*, 493(7432):365–370, 12 2012.

[147] Catherine Overy, George H. Booth, N. S. Blunt, James J. Shepherd, Deidre Cleland, and Ali Alavi. Unbiased reduced density matrices and electronic properties from full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 141(24):244117, 12 2014.

[148] George H. Booth, Deidre Cleland, Ali Alavi, and David P. Tew. An explicitly correlated approach to basis set incompleteness in full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 137(16):164112, 10 2012.

[149] Robert E. Thomas, Daniel Opalka, Catherine Overy, Peter J. Knowles, Ali Alavi, and George H. Booth. Analytic nuclear forces and molecular properties from full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 143(5):054108, 8 2015.

[150] Robert E. Thomas, Qiming Sun, Ali Alavi, and George H. Booth. Stochastic Multiconfigurational Self-Consistent Field Theory. *Journal of Chemical Theory and Computation*, 11(11):5316–5325, 11 2015.

[151] 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. *The Journal of Chemical Physics*, 146(24):244105, 6 2017.

[152] N. S. Blunt, Simon D. Smart, J. A. F. Kersten, J. S. Spencer, George H. Booth, and Ali Alavi. Semi-stochastic full configuration interaction quantum Monte Carlo: Developments and application. *The Journal of Chemical Physics*, 142(18):184107, 5 2015.

[153] F. R. Petruzielo, A. A. Holmes, Hitesh J. Changlani, M. P. Nightingale, and C. J. Umrigar. Semistochastic Projector Monte Carlo Method. *Physical Review Letters*, 109(23):230201, 12 2012.

[154] Seiichiro Ten-no. Stochastic determination of effective Hamiltonian for the full configuration interaction solution of quasi-degenerate electronic states. *The Journal of Chemical Physics*, 138(16):164126, 4 2013.

[155] Yuhki Ohtsuka and Seiichiro Ten-no. A study of potential energy curves from the model space quantum Monte Carlo method. *The Journal of Chemical Physics*, 143(21):214107, 12 2015.

[156] 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. *Journal of Chemical Theory and Computation*, 12(8):3674–3680, 8 2016.

[157] Nick S. Blunt. Communication: An efficient and accurate perturbative correction to initiator full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 148(22):221101, 6 2018.

[158] Sandeep Sharma, Takeshi Yanai, George H. Booth, C. J. Umrigar, and Garnet Kin-Lic Chan. Spectroscopic accuracy directly from quantum chemistry: Application to ground and excited states of beryllium dimer. *The Journal of Chemical Physics*, 140(10):104112, 3 2014.

[159] Hongjun Luo and Ali Alavi. Combining the Transcorrelated Method with Full Configuration Interaction Quantum Monte Carlo: Application to the Homogeneous Electron Gas. *Journal of Chemical Theory and Computation*, 14(3):1403–1411, 3 2018.

[160] Sandeep Sharma, Adam A. Holmes, Guillaume Jeanmairet, Ali Alavi, and C. J. Umrigar. Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory. *Journal of Chemical Theory and Computation*, 13(4):1595–1604, 4 2017.

[161] Adam A. Holmes, C. J. Umrigar, and Sandeep Sharma. Excited states using semistochastic heat-bath configuration interaction. *The Journal of Chemical Physics*, 147(16):164111, 10 2017.

[162] Junhao Li, Matthew Otten, Adam A. Holmes, Sandeep Sharma, and C. J. Umrigar. Fast semistochastic heat-bath configuration interaction. *The Journal of Chemical Physics*, 149(21):214110, 12 2018.

[163] Alan D. Chien, Adam A. Holmes, Matthew Otten, C. J. Umrigar, Sandeep Sharma, and Paul M. Zimmerman. Excited States of Methylene, Polyenes, and Ozone from Heat-Bath Configuration Interaction. *The Journal of Physical Chemistry A*, 122(10):2714–2722, 3 2018.

[164] N. S. Blunt, T. W. Rogers, J. S. Spencer, and W. M. C. Foulkes. Density-matrix quantum Monte Carlo method. *Physical Review B*, 89(24):245124, 6 2014.

[165] George H Booth and Garnet Kin-Lic Chan. Communication: Excited states, dynamic correlation functions and spectral properties from full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 137(19):191102, 11 2012.

[166] N. S. Blunt, Ali Alavi, and George H. Booth. Krylov-Projected Quantum Monte Carlo Method. *Physical Review Letters*, 115(5):050603, 7 2015.

[167] N. S. Blunt, Simon D. Smart, George H. Booth, and Ali Alavi. An excited-state approach within full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics*, 143(13):134117, 10 2015.

[168] Alexander Humeniuk and Roland Mitrić. Excited states from quantum Monte Carlo in the basis of Slater determinants. *The Journal of Chemical Physics*, 141(19):194104, 11 2014.

[169] Alex J. W. Thom. Stochastic Coupled Cluster Theory. *Physical Review Letters*, 105(26):263004, 12 2010.

[170] R. S. T. Franklin, J. S. Spencer, A. Zoccante, and A. J. W. Thom. Linked coupled cluster Monte Carlo. *The Journal of Chemical Physics*, 144(4):044111, 1 2016.

[171] J. Emiliano Deustua, Jun Shen, and Piotr Piecuch. Converging High-Level Coupled-Cluster Energetics by Monte Carlo Sampling and Moment Expansions. *Physical Review Letters*, 119(22):223003, 11 2017.

[172] Charles J. C. Scott and Alex J. W. Thom. Stochastic coupled cluster theory: Efficient sampling of the coupled cluster expansion. *The Journal of Chemical Physics*, 147(12):124105, 9 2017.

[173] Jarrod R. McClean and Alán Aspuru-Guzik. Clock quantum Monte Carlo technique: An imaginary-time method for real-time quantum dynamics. *Physical Review A*, 91(1):012311, 1 2015.

[174] Norm M. Tubman, Joonho Lee, Tyler Y. Takeshita, Martin Head-Gordon, and K. Birgitta Whaley. A deterministic alternative to the full configuration interaction quantum Monte Carlo method. *The Journal of Chemical Physics*, 145(4):044112, 7 2016.

[175] Samuel M. Greene, Robert J. Webber, Jonathan Weare, and Timothy C. Berkelbach. Beyond Walkers in Stochastic Quantum Chemistry: Reducing Error Using Fast Randomized Iteration. *Journal of Chemical Theory and Computation*, 15(9):4834–4850, 9 2019.

[176] Anthony Scemama, Anouar Benali, Denis Jacquemin, Michel Caffarel, and Pierre-François Loos. Excitation energies from diffusion Monte Carlo using selected configuration interaction nodes. *The Journal of Chemical Physics*, 149(3):034108, 7 2018.

[177] Yann Garniron, Anthony Scemama, Emmanuel Giner, Michel Caffarel, and Pierre-François Loos. Selected configuration interaction dressed by perturbation. *The Journal of Chemical Physics*, 149(6):064103, 8 2018.

[178] Monika Dash, Saverio Moroni, Anthony Scemama, and Claudia Filippi. Perturbatively Selected Configuration-Interaction Wave Functions for Efficient Geometry Optimization in Quantum Monte Carlo. *Journal of Chemical Theory and Computation*, 14(8):4176–4182, 8 2018.

[179] Samuel A. French, Alexey A. Sokol, Stefan T. Bromley, C. Richard A. Catlow, Stephen C. Rogers, Frank King, and Paul Sherwood. From CO2 to Methanol by Hybrid QM/MM Embedding. *Angewandte Chemie International Edition*, 40(23):4437, 12 2001.

[180] S. A. French, A. A. Sokol, S. T. Bromley, C. R. A. Catlow, S. C. Rogers, and P. Sherwood. Assignment of the complex vibrational spectra of the hydrogenated ZnO polar surfaces using QM/MM embedding. *The Journal of Chemical Physics*, 118(1):317–320, 1 2003.

[181] Lung Wa Chung, W. M.C. C. Sameera, Romain Ramozzi, Alister J. Page, Miho Hatanaka, Galina P. Petrova, Travis V. Harris, Xin Li, Zhuofeng Ke, Fengyi Liu, Hai-Bei Bei Li, Lina Ding, and Keiji Morokuma. The ONIOM Method and Its Applications. *Chemical Reviews*, 115(12):5678–5796, 6 2015.

[182] Thom Vreven and Keiji Morokuma. Investigation of the S 0 ?S 1 excitation in bacteriorhodopsin with the ONIOM(MO:MM) hybrid method. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 109(3):125–132, 4 2003.

[183] Yogesh V. Joshi and Kendall T. Thomson. Embedded cluster (QM/MM) investigation of C6 diene cyclization in HZSM-5. *Journal of Catalysis*, 230(2):440–463, 3 2005.

[184] Alexey A. Sokol, Stefan T. Bromley, Samuel A. French, C. Richard A. Catlow, and Paul Sherwood. Hybrid QM/MM embedding approach for the treatment of localized surface states in ionic materials. *International Journal of Quantum Chemistry*, 99(5):695–712, 1 2004.

[185] James S. Spencer, Nick S. Blunt, Seonghoon Choi, Jiří Etrych, Maria-Andreea Filip, W M C Foulkes, Ruth S T Franklin, Will J. Handley, Fionn D. Malone,

Verena A. Neufeld, Roberto Di Remigio, Thomas W. Rogers, Charles J C Scott, James J. Shepherd, William A. Vigor, Joseph Weston, RuQing Xu, and Alex J W Thom. The HANDE-QMC Project: Open-Source Stochastic Quantum Chemistry from the Ground State Up. *Journal of Chemical Theory and Computation*, 15(3):1728–1742, 3 2019.

[186] Rodney J. Bartlett and Monika Musiał. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, 79(1):291–352, 2 2007.

[187] Shaunak Mukherjee, Florian Libisch, Nicolas Large, Oara Neumann, Lisa V. Brown, Jin Cheng, J. Britt Lassiter, Emily A. Carter, Peter Nordlander, and Naomi J. Halas. Hot Electrons Do the Impossible: Plasmon-Induced Dissociation of H 2 on Au. *Nano Letters*, 13(1):240–247, 1 2013.

[188] Linan Zhou, Chao Zhang, Michael J. McClain, Alejandro Manjavacas, Caroline M. Krauter, Shu Tian, Felix Berg, Henry O. Everitt, Emily A. Carter, Peter Nordlander, and Naomi J. Halas. Aluminum Nanocrystals as a Plasmonic Photocatalyst for Hydrogen Dissociation. *Nano Letters*, 16(2):1478–1484, 2 2016.

[189] Ambarish Kulkarni, Samira Siahrostami, Anjli Patel, and Jens K. Nørskov. Understanding Catalytic Activity Trends in the Oxygen Reduction Reaction. *Chemical Reviews*, 118(5):2302–2312, 3 2018.

[190] Rougang Tang, Jonathan Nafziger, and Adam Wasserman. Fragment occupations in partition density functional theory. *Physical Chemistry Chemical Physics*, 14(21):7780, 5 2012.

[191] Daniel S. Graham, Xuelan Wen, Dhabih V. Chulhai, and Jason D. Goodpaster. Huzinaga projection embedding for efficient and accurate energies of systems with localized spin-densities. *The Journal of Chemical Physics*, 156(5):054112, 2 2022.

[192] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley. Derivative studies in hartree-fock and møller-plesset theories. *International Journal of Quantum Chemistry*, 16(S13):225–241, 6 1979.

Appendix A

Appendix of: Robust, Accurate, and Efficient: Quantum Embedding Using the Huzinaga Level-Shift Projection Operator for Complex Systems

A.1 WF Reaction Energies

Here we present the full system WF energies for the systems studied in the paper. All Figures in the main text show the difference between the embedding energies, and the full system WF energies in the below tables.

The CCSD(T) activation energy of the secondary carbon S_N2 system shown in Figure 2.4 is -18.10 kcal/mol.

CCSD(T) Results of S_N2 Activation Energy (kcal/mol)			
Length of Carbon Chain	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ
4	-34.32	-26.04	-14.22
5	-37.26	-27.86	-15.96
6	-37.36	-27.99	-16.10
7	-37.47	-28.09	-16.18

Table A.1: CCSD(T) activation energies of the S_N2 system corresponding to Figures 2.1, 2.2, and 2.3.

CCSD(T) Results of Fluorine Elimination Reaction Energy (kcal/mol)			
Length of Carbon Chain	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ
6	39.88	36.45	36.42
8	39.79	36.27	36.28

Table A.2: CCSD(T) reaction energies of the fluorine elimination reaction corresponding to Figure 2.6.

CASPT2 Bond Dissociation Energy (kcal/mol)		
Bond Distance (Å)	cc-pVTZ	aug-cc-pVDZ
1.5	0	0
1.6	1.63	1.08
1.7	7.49	6.52
1.8	15.74	14.48
1.9	25.12	23.64
2.0	34.81	33.15
2.1	44.31	42.49
2.2	53.30	51.36
2.3	61.59	59.56
2.4	69.07	66.98
2.5	75.70	73.59
3.0	97.12	95.34
3.5	105.27	104.18
4.0	107.83	107.19

Table A.3: CASPT2 bond dissociation energies for the system shown in Figure 2.7. All energies are relative to equilibrium bond distance (1.5 Å), which is why the first row is all 0.

CASPT2 Rotation Energy (kcal/mol)		
Degree of Rotation	cc-pVTZ	aug-cc-pVDZ
0	0	0
15	2.78	2.79
30	11.06	11.09
45	24.56	24.57
60	42.47	42.45
75	61.87	61.92
90	71.98	72.24
105	62.16	62.21
120	42.62	42.49
135	24.61	24.42
150	11.45	11.30
165	4.02	3.96
180	1.76	1.76

Table A.4: CASPT2 rotation energies for the system shown in Figure 2.8. All energies are relative to 0 degree rotation, which is why the first row is all 0.

The CASPT2 energy of the Fe-MOF-74 system shown in Figure 2.9 is -3.07 kcal/mol. The energy is calculated as

$$\Delta E_{CASPT2} = E^{\text{MOF with H}_2\text{bound}}_{\text{CASPT2}} - (E^{\text{MOF}}_{\text{CASPT2}} + E^{\text{H}_2}_{\text{CASPT2}}). \tag{A.1}$$

A.2 DFT-in-DFT Difference

In order to analyze the effect of error cancellation, we compared the DFT-in-DFT energy of the undecane system in the aug-cc-pVDZ basis to the full KS-DFT energy of the same system for both the reactant, 1-chloroundecane and transition state, 1-chlorofluoroundecane. In tables A.5 and A.6 we compare the absolute DFT-in-DFT energy to the absolute KS-DFT energy,

$$\Delta E^{\text{reactant}} = E^{\text{reactant}}_{\text{KS-DFT}} - E^{\text{reactant}}_{\text{DFT-in-DFT}} \tag{A.2}$$

Charged Subsystems			
Subsystem Size	Reactant DFT Diff.	Trans. DFT Diff	Rxn DFT Diff
1	-199.2293612	-223.8421388	-24.61277762
2	-186.0572263	-210.0586433	-24.00141692
3	-190.2829682	-210.0476746	-19.76470635
4	-196.8392577	-212.8787538	-16.03949612
5	-207.4704191	-219.9877578	-12.51733862

Table A.5: While absolute DFT-in-DFT energies do not converge to the KS-DFT absolute energy, the reaction energies converge with increasing subsystem size. All energies reported in kcal/mol.

Neutral Subsystems			
Subsystem Size	Reactant DFT Diff.	Trans. DFT Diff	Rxn DFT Diff
1	-219.2802954	-232.832758	-13.55246262
2	-186.3463784	-218.2616119	-31.91523357
3	-189.6231771	-211.6679109	-22.04473387
4	-194.7568487	-212.7193562	-17.96250752
5	-199.5069421	-215.4697389	-15.9627968

Table A.6: Absolute DFT-in-DFT energy differences are similar to the charged subsystems, however the reaction energy differences are larger for most systems. All energies reported in kcal/mol.

in addition to the reaction DFT-in-DFT energy compared with the KS-DFT reaction energy,

$$\Delta E^{\text{rxn}} = (E^{\text{transition state}}_{\text{KS-DFT}} - E^{\text{reactant}}_{\text{KS-DFT}}) - (E^{\text{transition state}}_{\text{DFT-in-DFT}} - E^{\text{reactant}}_{\text{DFT-in-DFT}}). \tag{A.3}$$

Important to note is that the absolute energy differences are lower for the neutral subsystems at subsystem size 4 and 5, but the reaction energy differences are lower for the charged subsystems, indicating some level of error cancellation.

Appendix B

Appendix of: Huzinaga Projection Embedding for Efficient and Accurate Energies of Systems with Localized Spin-densities

B.1 Unrestricted S_N2 Subsystem Charging Analysis

Here we present the results of our S_N2 study utilizing different subsystem electron distribution schemes. The two schemes studied here are charging subsystems to maintain closed-shell subsystems as discussed in our previous paper[130], and specifying one subsystem as a doublet with an unpaired alpha electron and the other subsystem as a doublet with an unpaired beta electron. These two schemes will be referred to as charged subsystems and spin subsystems respectively.

As is clear by the lack of convergence to the CCSD(T) result for the spin subsystem scheme, accounting for cut bonds by embedding two doublets is less favorable than our previously identified charging embedding scheme.

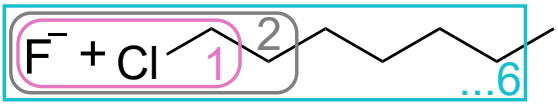


Figure B.1: Subsystem division diagram for reactant of 1-chloroheptane S_N2 reaction. Subsystem divisions are numbered according to how many carbon are included in the WF region.

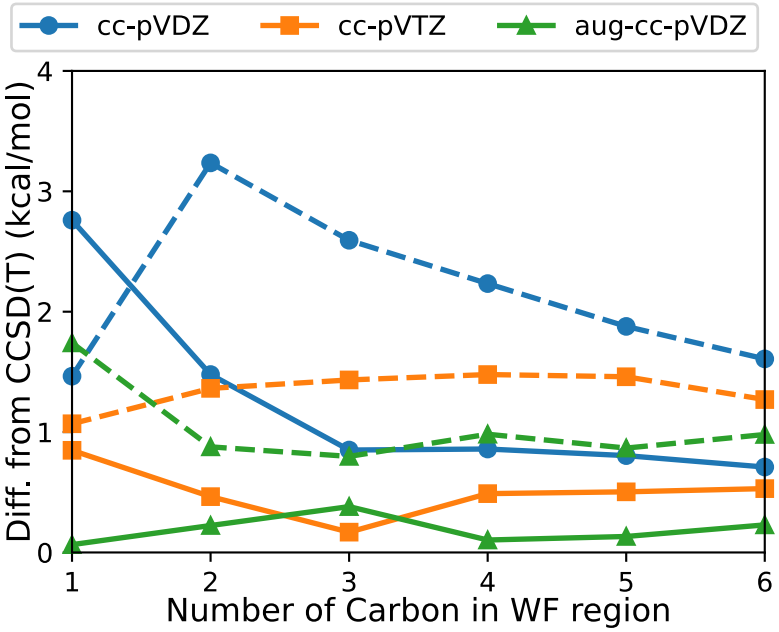


Figure B.2: S_N2 Activation energy of 1-chloroheptane reaction following the subsystem division scheme in Figure B.1. Charged subsystem scheme previously reported by our group are points connected by solid line, while the results of spin subsystem embedding are shown by points connected using the dashed lines.

Appendix C

Appendix of: Fully quantum embedding with density functional theory for full configuration interaction quantum Monte Carlo

Table C.1: Sizes of space are dramatically reduced by embedding.

	Before embedding				After embedding			
	N	M	N_{dets}	Storage	N	M	N_{dets}	storage
Li^+ Benzene	32	122	10^{39}	1100 MB	2	14	10^2	228 KB
LiH Benzene	34	127	10^{41}	1300 MB	4	19	10^4	735 KB
F^- Benzene	38	121	10^{43}	2200 MB	10	14	10^6	228 KB
HF Benzene	38	126	10^{44}	2600 MB	10	19	10^8	735 KB

Table C.2: Energy data for canonical calculations shown in the manuscript. Energies are in Hartree. FCIQMC energies have errors in the parentheses in the final digit.

Species	CCSD(T) total energy	CCSD(T) correlation energy	$i-$FCIQMC correlation energy (10^6 walkers)
LiH Benzene	-239.6194835481280	-0.85600348827805	-0.7280(6)
Li^+ Benzene	-238.8755497917899	-0.82071622974315	-0.702(2)
HF Benzene	-331.8135808511175	-1.029661027898815	-0.875(5)
F^- Benzene	-331.1870787004810	-1.022303198219452	-0.949(9)
LiH	-8.0147569	-0.031080149	-0.031081(1)
Li^+	-7.2362237	-0.000105103	-0.000108(8)
HF	-100.23007	-0.211019285	-0.21152(1)
F^-	-99.56031	-0.194326068	-0.194779(8)
H^-	-0.4698568	-0.021033058	-0.0210328(2)

Table C.3: Energy data for embedded calculations shown in the manuscript. Energies are in Hartree. FCIQMC energies have errors in the parentheses in the final digit.

Species	Total Embedding Energy	CCSD(T) correlation energy	$i-$FCIQMC correlation energy (10^6 walkers)
Embedded LiH Benzene	-240.09956	-0.030348178	-0.030352(3)
Embedded Li^+ Benzene	-239.3551679	-5.84527E-05	-0.000059(4)
Embedded HF Benzene	-332.29869	-0.209871759	-0.21036(1)
Embedded F^- Benzene	-331.6769	-0.194501559	-0.194954(9)
HF Benzene 0.9248454 Å	-332.2986941	-0.209871759	-0.21036(1)
HF Benzene 1 Å	-332.2932769	-0.212779343	-0.21334(1)
HF Benzene 2 Å	-332.1192635	-0.26900167	-0.26687(4)
HF Benzene 3 Å	-332.1886709	-0.448253972	-0.35621(3)
HF Benzene 4 Å	-332.2467766	-0.551591512	-0.4006(1)

Table C.4: Equilibrium geometries, xyz format (Angstroms)

14

FHBenzene

C	-0.01655	-0.81006	-1.14622
C	0.6426	-1.37963	-0.06713
C	1.22868	-0.56973	0.89043
C	1.15647	0.8089	0.7732
C	0.49836	1.38024	-0.30253
C	-0.08851	0.57101	-1.26435
H	-0.47577	-1.44158	-1.89571
H	0.69394	-2.45623	0.0278
H	1.73807	-1.01423	1.73546
H	1.60966	1.43998	1.52639
H	0.43719	2.4568	-0.39193
H	-0.6022	1.01591	-2.1069
F	-2.46678	0.00079	0.83583
H	-1.72621	-0.01215	0.282

13

FBenzene

C	2.13783	-0.00067	0.00023
C	1.43737	1.19593	0.00005
C	0.05056	1.18951	-0.00019
C	-0.67026	0.00073	-0.00037
C	0.04942	-1.18887	-0.00025
C	1.43625	-1.19659	0.00011
H	3.22227	-0.00116	0.0005
H	1.97827	2.13699	0.00011
H	-0.49077	2.12957	-0.00024
H	-1.7952	0.00147	-0.00044
H	-0.49282	-2.1284	-0.0004

H	1.97622	-2.13819	0.00027
F	-3.44944	-0.00005	0.0003

14

LiHBenzene

C	1.29221	-0.5086	-0.22611
C	1.08617	0.86415	-0.22792
C	-0.20571	1.37202	-0.22858
C	-1.29207	0.50774	-0.22762
C	-1.08606	-0.86497	-0.22637
C	0.20591	-1.37287	-0.22566
H	2.29857	-0.90431	-0.21402
H	1.93183	1.53848	-0.21699
H	-0.36627	2.44166	-0.21821
H	-2.29837	0.90362	-0.2167
H	-1.93169	-1.53933	-0.21427
H	0.36638	-2.44251	-0.21308
Li	-0.00084	0.00226	1.95974
H	-0.00066	0.01074	3.58758

13

LiBenzene

C	-1.24317	0.62708	-0.12684
C	-1.16464	-0.76316	-0.12603
C	0.0787	-1.39001	-0.12624
C	1.24337	-0.62718	-0.12623
C	1.16499	0.76306	-0.12605
C	-0.07839	1.38986	-0.12637
H	-2.20885	1.1148	-0.1352
H	-2.0697	-1.35594	-0.13443
H	0.13937	-2.47028	-0.13466
H	2.20911	-1.11482	-0.13304

H	2.07007	1.35579	-0.13332
H	-0.13911	2.47013	-0.13478
Li	-0.00201	0.00082	1.784

2
FH Isolated Geometry

F	-2.46678	0.00079	0.83583
H	-1.72621	-0.01215	0.282

2
LiH Isolated

Li	-0.00084	0.00226	1.95974
H	-0.00066	0.01074	3.58758