Combining high- and low-level electronic structure theories for the efficient exploration of potential energy surfaces

Thesis by
Sebastian James Rice Lee

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy in Chemistry

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2021
Defended October 26, 2020

Sebastian James Rice Lee
ORCID: 0000-0001-7006-9378

ACKNOWLEDGEMENTS

I want to thank my advisor, Professor Thomas Miller, for his encouragement and inspiration, which has helped me become the scientist I am today. I want to thank my thesis committee, Professor Garnet Chan, Professor Mitchio Okumura, and Professor Ryan Hadt, for their guidance throughout my graduate student career.

I want to thank all members of the Miller Group who have made my time enjoyable at Caltech: Priscilla, Feizhi, Matt, Ioan, Philip, Leanne, Michiel, Jorge, Dan, Matthew, Xuecheng, Jeongmin, Sherry, Brooke, Ralph, Mike, Connie, Brett, Reid, Steve, Jiace, Saleh, Xinglong, Varun, Zhuoran, Roman, James, Marta, and Tamara. Special thanks to my friends Steven, Annet, Griffin, and Lindiwe for being there through the good and the bad.

I am grateful to the Resnick Sustainability Institute at Caltech for supporting me as a Resnick fellow. Special thanks to Neil Fromer and Heidi Rusina for all of their encouragement and support. I am also grateful to the Molecular Software Sciences Institute for supporting me as a MolSSI software fellow. I want to give a special thanks to my mentor Taylor Barnes for his guidance and support.

I am truly grateful to my family for all their support and patience. Thanks to my brothers Marcus and Trevor for our weekly video gaming sessions. And thank you to my parents for all your advice; it is not often that your parents can teach you the ins and outs of graduate school.

Finally, thank you Tamara for your love and support. You have made this experience all the better.

ABSTRACT

The efficient exploration and characterization of potential energy surfaces paves the way for the theoretical elucidation of complex chemical processes. A potential energy surface arises from the application of the Born–Oppenheimer approximation when solving the Schrödinger equation for a molecular system. The extraction of energies and nuclear gradients from the Schrödinger equation is typically cost-prohibitive, which has inspired a plethora of approximations. In this thesis, we present the development of embedding and machine learning methodologies that provide fast and accurate energies and nuclear gradients for different chemical classes by combining high- and low-level electronic structure theories. If a chemical change occurs in a spatially localized region, embedding strategies offer an effective approach for balancing accuracy and computational cost. We first consider embedded mean-field theory (EMFT), which seamlessly combines different mean-field theories for different subsystems to describe the whole molecular system. We analyze the errors in EMFT calculations that occur when subsystems employ different atomic-orbital basis sets. These errors can be alleviated by a Fock-matrix correction scheme or by following general basis set recommendations. Systems exhibiting a more complicated electronic structure require a systematically improvable level of theory for the subsystems, which can be realized by projection-based embedding. Projection-based embedding enables the description of a small part of a molecular system at the level of a correlated wavefunction method while the remainder of the system is described at the mean-field level. We go on to derive and numerically demonstrate the analytical nuclear gradients for projection-based embedding. If description of the entire system at the high level of theory is deemed necessary, molecular-orbital-based machine learning (MOB-ML) calculations offers a framework to predict accurate correlation energies at the cost of obtaining molecular orbitals. We go on to present the derivation, implementation, and numerical demonstration of MOB-ML analytical nuclear gradients. We demonstrate the developed methodologies by exploring potential energy surfaces of organic and transition-metal containing molecules.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1]S. J. R. Lee, T. Husch, F. Ding, and T. F. Miller III, "Analytical gradients for molecular-orbital-based machine learning", in preparation (2020).

[2]S. J. R. Lee, F. Ding, F. R. Manby, and T. F. Miller III, "Analytical gradients for projection-based wavefunction-in-DFT embedding", en, J. Chem. Phys. **151**, 064112 (2019) 10.1063/1.5109882.

[3]S. J. R. Lee, M. Welborn, F. R. Manby, and T. F. Miller III, "Projection-Based Wavefunction-in-DFT Embedding", Acc. Chem. Res. **52**, 1359–1368 (2019) 10.1021/acs.accounts.8b00672.

[4]S. J. R. Lee, K. Miyamoto, F. Ding, F. R. Manby, and T. F. Miller III, "Density-based errors in mixed-basis mean-field electronic structure, with implications for embedding and QM/MM methods", en, Chem. Phys. Lett. **683**, 375–382 (2017) 10.1016/j.cplett.2017.04.059.

In Ref. 1 S.J.R.L. derived the equations, wrote the computer code, carried out the numerical calculations, and contributed to the design of the research, the analysis of the data, and the preparation of the manuscript.

In Ref. 2 S.J.R.L. derived the equations, wrote the computer code, carried out the numerical calculations, designed the research, and contributed to the analysis of the data and the preparation of the manuscript.

In Ref. 3 S.J.R.L. had a lead role in the preparation of the manuscript.

In Ref. 4 S.J.R.L. carried out the numerical calculations and contributed to the design of the research, the analysis of the data, and the preparation of the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

NOMENCLATURE

a.u. atomic units.

AO. Atomic Orbital.

ASE. Atomic Simulation Environment.

BFGS. Broyden-Fletcher-Goldfarb-Shanno.

BSSE. Basis Set Superposition Error.

CCSD. Coupled Cluster Singles and Doubles.

CCSD(T). Coupled Cluster Singles and Doubles with perturbative Triples.

DF. Density Fitting.

DFT. Density Functional Theory.

EMFT. Embedded Mean-Field Theory.

FCDFT. Fock-Corrected Density Functional Theory.

FCEMFT. Fock-Corrected Embedded Mean-Field Theory.

GPR. Gaussian Process Regression.

HF. Hartree-Fock.

IBO. Intrinsic Bond Orbital.

IDPP. Image Dependent Pair Potential.

LMO. Localized Molecular Orbital.

MAE. Mean Absolute Error.

MD. Molecular Dynamics.

MINAO. MINimal Atomic Orbital basis.

MOB-ML. Molecular-Orbital-Based Machine Learning.

MP2. Second-order Moller-Plesset perturbation theory.

NAKP. Non-Additive Kinetic Potential.

NEB. Nudged Elastic Band.

OEP. Optimized Effective Potential.

ONIOM. Our own N-layered Integrated molecular Orbital and molecular Mechanics.

PNO-LCCSD. Pair Natural Orbital-Local Coupled Cluster Singles and Doubles.

PNO-LMP2. Pair Natural Orbital-Local Second-order Moller-Plesset perturbation theory.

QM/MM. Quantum Mechanics/Molecular Mechanics.

SCF. Self-Consistent Field.

WF. Wavefunction.

WF-in-DFT. Wavefunction-in-Density Functional Theory.

XC. Exchange Correlation.

Z-CPHF. Z-vector Coupled Perturbed Hartree–Fock.

Z-CPKS. Z-vector Coupled Perturbed Kohn-Sham.

Z-CPL. Z-vector Coupled Perturbed Localization.

Chapter 1

INTRODUCTION

Complex chemical systems present challenges to electronic structure theory, stemming from large system sizes, subtle interactions, and coupled dynamical timescales. New methods are needed to perform reliable, rigorous, and affordable electronic structure calculations for simulating the potential energy surfaces of such systems. These new methods need to have fast energy and gradient evaluations to efficiently explore and characterize potential energy surfaces, which is the foundation of the quantum chemical elucidation of complex reaction mechanisms via molecular dynamics simulations and minimum-energy and transition-state structure characterization.

For systems in which complicated chemical rearrangements (e.g. bond breaking and forming) occur in a local spatial region, an effective strategy for balancing accuracy and computational cost is to employ one of various multiscale embedding strategies [1–28]. Generally, embedding methodologies hinge on the condition that a system can be efficiently partitioned into a local subsystem that demands a high-level treatment and an environment that can be treated with a lower (and computationally less expensive) level of theory. By exploiting the intrinsic locality of molecular interactions, these approaches provide high accuracy for regions that demand it while avoiding the computational cost of a high-level calculation on the whole system.

One such embedding method is embedded mean-field theory (EMFT) [17, 23, 29], which flexibly embeds one high-level MF theory (e.g. hybrid-GGA) in another low-level MF theory (e.g. LDA) without needing to specify or fix the number of electrons in each subsystem. EMFT inherits the simple gradient theory of the parent mean-field theories and has been shown to provide hybrid-GGA quality results at a much reduced computational cost. In Chapter 2 we introduce the use of mixed-basis atomic orbital (AO) basis sets to further speed up the evaluation of EMFT energy and gradient calculations. Using mixed-basis sets in electronic structure calculations is a commonly employed strategy for reducing computational cost by describing different portions of the system with AO basis sets of different size [30–43]. Even for mean-field methods such as Hartree–Fock (HF), density functional theory (DFT) and

EMFT, the computational cost of these methods formally scales as $O(N^4)$, where N corresponds to the number of basis functions, so reducing the number of employed basis functions is a simple and tempting strategy to decrease computational cost. However, using mixed basis sets have shown in some cases to give rise to large errors in comparison to calculations performed with a single, uniform AO basis set [17, 41, 42]. In this chapter, we systematically analyze the sources of error that occur in mixed-based mean-field electronic structure calculations and highlight how the underlying errors can be largely eliminated by using Fock-matrix corrections or by avoiding the use of a minimal basis set in the low-level region.

Many chemical processes demand accurate, *ab initio* electronic structure theoretical descriptions from wavefunction theories, where mean-field methods such as DFT and EMFT prove to be inadequate. However, the routine calculation of accurate wave function energies is prohibited by their steep cost, e.g., coupled-cluster singles, doubles, and perturbative triples [CCSD(T)] scales as N^7 [44] and full configuration interaction scales as $N!$ [45]. In Chapter 3 we review projection-based embedding which provides a simple, robust, and accurate approach for describing a small part of a chemical system at the level of a correlated wavefunction method while the remainder of the system is described at the level of density functional theory. We go on to review a series of technical developments and applications illustrating the accuracy of projection-based embedding energy calculations. Next, in Chapter 4 we enable the full characterization of potential energy surfaces of chemical systems with projection-based embedding by deriving and numerically demonstrating its analytical nuclear gradients. We highlight the utility of the analytical gradients by calculating the minimum energy pathway for a hydride transfer in a cobalt-based molecular catalyst.

Even with the advances in embedding methodologies such as EMFT and projection-based embedding the modeling of many chemical processes still remains out of reach. EMFT lacks the accuracy required in many cases and projection-based embedding is still too computationally expensive to simulate large systems. In recent years, machine learning has opened up a new way of mitigating the cost of quantum chemical calculations [46–69]. One promising approach is molecular orbital based-machine learning (MOB-ML), which predicts the pair-wise sum of a wavefunction correlation energy at the cost of a mean-field calculation [70–72], giving us the best of both worlds in terms of accuracy and computational cost. In Chapter 5 we provide a brief overview of MOB-ML, and derive and numerically

demonstrate its analytical nuclear gradients enabling the full characterization of potential energy surfaces. We highlight that MOB-ML gradients yield high-quality optimized structures at a computational cost comparable to mean-field calculations (e.g., HF and DFT).

Tackling the Schrodinger equation has been the work of life times and has lead to many amazing developments allowing us to push the limits of quantum chemistry to further and further heights. The theoretical developments in this thesis further this pursuit by combining high- and low-level electronic structure theories for the efficient exploration of potential energy surfaces.

Chapter 2

DENSITY-BASED ERRORS IN MIXED-BASIS MEAN-FIELD ELECTRONIC STRUCTURE, WITH IMPLICATIONS FOR EMBEDDING AND QM/MM METHODS

Adapted from:

[1]S. J. R. Lee, K. Miyamoto, F. Ding, F. R. Manby, and T. F. Miller III, "Density-based errors in mixed-basis mean-field electronic structure, with implications for embedding and QM/MM methods", en, Chem. Phys. Lett. **683**, 375–382 (2017) 10.1016/j.cplett.2017.04.059.

In this chapter we consider mean-field electronic structure calculations with subsystems that employ different atomic-orbital basis sets. A major source of error arises in charge-manifestation reactions (including ionization, electron attachment, or deprotonation) due to electronic density artifacts at the subsystem interface. The underlying errors in the electronic density can be largely eliminated with Fock-matrix corrections or by avoiding the use of a minimal basis set in the low-level region. These corrections succeed by balancing the electronegativity of atoms at the subsystem interface, much as link-atoms in QM/MM calculations rely upon balancing the electronegativity of atoms in the truncated QM region.

2.1 Introduction

Mixed-basis electronic structure calculations have long been employed for obtaining diverse chemical properties, including nuclear magnetic resonance chemical shifts [30–38], reaction energies [39–42], solvation energies [42], and molecular electrostatic potentials [43]. While there have been many successful applications of mixed basis sets, they have been shown in some cases to give rise to large errors in comparison to calculations performed with a single, uniform AO basis set. In particular, DiLabio et al. [42] noted that properties such as proton affinities and electron affinities are particularly problematic when a minimal STO-3G basis set is used to describe portions of the system, whereas other processes such as bond dissociation energies are well behaved [41]. The underlying origin of these observations was not demonstrated, nor were strategies provided for anticipating and avoiding these sizable artifacts in mixed-basis electronic structure calculations.

It was also recently shown that embedded mean-field theory (EMFT), a DFT-based quantum embedding method for which the subsystem densities are not frozen, can also exhibit errors that are dominated by the use of different basis sets for the embedded subsystems [17]. A particularly striking illustration of the errors associated with mixed-basis-set calculations was found to be the deprotonation reaction of decanoic acid, akin to the previously observed mixed-basis errors in the calculation proton and electron affinities [42]. The development of rigorous and robust quantum embedding methods that allow for the unfrozen calculation of electronic densities thus demands an improved understanding of these errors and improved strategies for mitigating them, which is the focus of the current chapter.

In this chapter, we systematically analyze the sources of error caused in mixed-based mean-field electronic structure calculations. The quantum embedding framework of EMFT is employed to address this problem in the context of simple, illustrative chemical reactions. This chapter illustrates that artifacts in the calculated electronic density — akin to those that have been emphasized recently in both KS-DFT [73, 74] and quantum embedding methods [22] employing uniform AO basis sets — comprise a major source of error in the calculation of molecular properties with mixed-basis electronic structure methods. Furthermore, we demonstrate that Fock-matrix corrections [29] offer a promising strategy for mitigating these errors by addressing the root causes of the problem in the electronic density.

2.2 Methodology

Because mixed-basis KS-DFT calculations correspond to a special case of EMFT, and motivated by the fact that mixed AO basis sets are the dominant source of error in many EMFT calculations, we employ EMFT as a framework to analyze and mitigate this source of error. A brief review of EMFT is provided below, as well as the recently developed Fock-matrix correction method that will be employed as the low-level theory in some of the presented results.

Embedded Mean-Field Theory (EMFT)

EMFT employs a partitioning of the electronic density matrix at the level of the one-particle basis [17], such that

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}^{\mathrm{AA}} & \mathbf{D}^{\mathrm{AB}} \\ \mathbf{D}^{\mathrm{BA}} & \mathbf{D}^{\mathrm{BB}} \end{bmatrix}, \tag{2.1}$$

where $\mathbf{D}^{\mathrm{AA}}$ and $\mathbf{D}^{\mathrm{BB}}$ denotes the density matrix blocks for subsystems A and B, respectively, and $\mathbf{D}^{\mathrm{AB}}$ and $\mathbf{D}^{\mathrm{BA}}$ are the off-diagonal blocks. The total energy expression for EMFT is then

$$E^{\mathrm{EMFT}} = E^{\mathrm{low}}[\mathbf{D}] - E^{\mathrm{low}}[\mathbf{D}^{\mathrm{AA}}] + E^{\mathrm{high}}[\mathbf{D}^{\mathrm{AA}}], \tag{2.2}$$

where E^{low} and E^{high} refer to the energy functionals for the low-level and high-level theories, respectively. Unlike ONIOM [75, 76], the EMFT energy is obtained via minimization of the above energy expression with respect to the total density matrix, $\mathbf{D}$, yielding the usual self-consistent field (SCF) equation,

$$\mathbf{FC} = \mathbf{SC}\epsilon, \tag{2.3}$$

where $\mathbf{F}$ is the EMFT Fock matrix defined as the derivative of the EMFT energy with respect to the total density matrix; $\mathbf{C}$, $\mathbf{S}$, ϵ are the usual MO coefficients, overlap and MO eigenvalue matrices. The generality of the EMFT framework allows for the embedding of any mean-field theory into any other. This includes using relatively high-cost DFT methods (such as those employing hybrid, double-hybrid, or range-separated exchange-correlation functionals) for the high-level theory and low-cost DFT methods (such as GGA, LDA, or FCDFT) or tight-binding methods for the low-level theory. Since the high- and low-level theories in EMFT can employ different AO basis sets [17], it is clear that mixed-basis mean-field calculations can also be regarded as a special case of EMFT in which the same exchange correlation functional but a different basis set is employed for the two subsystems.

Fock-corrected DFT (FCDFT)

FCDFT is a parameterized version of KS-DFT in which calculations using an inexpensive XC functional form (such as LDA) and minimal basis (such as STO-3G) are corrected to reproduce KS-DFT results obtained with more accurate XC functionals and larger basis sets [29]. The parameterized corrections in FCDFT are included at the level of the Fock matrix, such that both the calculated energy and the density are improved; this feature contrasts with other parameterized DFT methods, such as the HF-3c approach developed by Grimme et al. [77], in which corrections are included to the energy following conventional SCF convergence of the minimal-basis density.

The energy functional for FCDFT is

$$E^{\mathrm{FCDFT}}[\tilde{\mathbf{D}}] = E^{\mathrm{DFT}}[\tilde{\mathbf{D}}] + \mathrm{tr}\tilde{\mathbf{D}}\tilde{\mathbf{L}} + U_{\mathrm{cor}}, \tag{2.4}$$

where $\tilde{\mathbf{D}}$ is the minimal-basis one-particle density matrix, E^{DFT} is the DFT energy evaluated with the inexpensive XC functional, $\tilde{\mathbf{L}}$ is the minimal-basis semi-empirical correction matrix, and U_{cor} is a sum of short-ranged pairwise interactions that is an explicit function of atom-atom distances, $\{\mathbf{R}_{IJ}\}$, and is independent of the electronic density matrix. The elements of $\tilde{\mathbf{L}}$ are given by

$$\tilde{L}_{\mu\nu} = \begin{cases} \delta_{\mu\nu}\epsilon_{\mu} & \mu, \nu \in I \\ F^{\mathrm{SK}}_{\mu\nu}(\mathbf{R}_{IJ}) & \mu \in I, \nu \in J \end{cases}, \tag{2.5}$$

where I and J index atoms, ϵ_{μ} are fitted, atom-specific diagonal energy shifts, and $F^{\mathrm{SK}}_{\mu\nu}(\mathbf{R}_{IJ})$ are fitted atom-atom interaction functions. Ref. 29 provides full details of FCDFT and shows that it can be straightforwardly employed as the low-level mean-field method in the EMFT framework.

In the current work, we employ the FCLDA implementation of FCDFT, for which E^{DFT} in Eq. 2.4 is evaluated using the LDA exchange-correlation functional. Only diagonal components in the correction matrix $\tilde{\mathbf{L}}$ are included, with the AO energy shifts, ϵ_{μ} parameterized to fit the valence orbital energies for a small training set of molecules, as described in the Computational Details Section. The term U_{cor} is neglected since it does not affect the electronic density and since it has little effect on the reaction energy due to the extremely small structural change in the environment subsystem for the reactions studied here.

2.3 Computational Details

All molecular geometries are identical to those reported by Fornace et al. [17], which are optimized at the B3LYP/6-311G** level of theory. Calculations are performed using atomic-orbital basis functions that are implemented in terms of spherical Gaussians, and unless otherwise specified, the restricted form of the SCF wavefunction is employed.

Standard KS-DFT calculations, including those with mixed basis sets, are performed using Molpro 2015.1 [78] without density fitting, employing the PBE exchange-correlation functional [79] with the cc-pVTZ [80], 6-31G [81], and STO-3G [82] basis sets. Throughout this work, KS-DFT mixed-basis calculations are denoted using the convention "PBE/large-basis:small-basis," where the large basis is used to describe subsystem A and the small basis is used to describe subsystem B. The exchange-correlation functional is evaluated using an adaptively generated quadrature grid that reproduces the energy of the Slater-Dirac functional to a specified threshold accuracy of 10^{-10} hartree. For the atomic electronegative analysis (Table 2.2), open-shell KS-DFT calculations are performed using Gaussian 09 with an unrestricted reference.

The current study employs the FCLDA/STO-3G implementation of FCDFT, for which the LDA exchange-correlation functional with the minimal STO-3G basis is used to evaluate E^{DFT} in Eq. 2.4. The FCDFT parameters, ϵ_μ, are obtained by minimizing the root-mean-square error of the FCLDA/STO-3G valence orbital energies with respect to KS-DFT at the PBE/cc-pVTZ level of theory. The needed parameters for atom-types H and C atoms are obtained by fitting over a training set composed of H_2, CH_4, C_2H_2, C_2H_4, C_2H_6, and benzene. All FCDFT parameters used in this study are provided in Table 2.1.

Table 2.1: Fock correction parameters for fitting FCLDA/STO-3G to PBE/cc-pVTZ

	Parameters (hartree)		
Atom	ϵ_{1s}	ϵ_{2s}	ϵ_{2p}
H	−0.059741		
C	−1.578564	0.077887	−0.087737

EMFT and FCEMFT calculations are performed using the developer version of Molpro [78, 83]. For the EMFT calculations, denoted as PBE-in-LDA/cc-pVTZ:STO-3G/DF:DF(s), the high-level theory corresponds to DFT with the PBE exchange-correlation functional, using the cc-pVTZ AO basis, and the Ahlrichs density-fitting

basis [84, 85]; the low-level theory corresponds to DFT with the LDA exchange-correlation functional, using the STO-3G AO basis, and the Ahlrichs density-fitting basis with only s-type functions. For the FCEMFT calculations, denoted PBE-in-FCLDA/cc-pVTZ:STO-3G/DF:DF(s), the same high-level theory is employed; the low-level theory corresponds to the previously described FCLDA/STO-3G implementation of FCDFT, using the Ahlrichs density fitting basis with only s-type functions. All EMFT and FCEMFT calculations employ the default threshold accuracy of 10^{-6} hartree for the quadrature grid.

Vacuum embedding calculations are performed with Gaussian 09 [86] using the PBE functional with the basis set cc-pVTZ. Vacuum embedding refers to terminating subsystem A with a link atom such that interactions with subsystem B are neglected. The terminal link atoms are positioned according to the default parameterization scheme in the Gaussian 09 implementation of the ONIOM method.

Figures that are plotted as a function of the number of carbon atoms in subsystem A assume that all non-carbon atoms attached to those carbon atoms are included in subsystem A; the remaining atoms are placed in subsystem B. Zero carbons in subsystem A corresponds to treating the full molecule with the smaller basis set.

2.4 Results and Discussion

Illustrating the mixed basis error

Fig. 2.1 compares two applications of mixed-basis KS-DFT. Panels (a)–(c) present results for the 1-chlorodecane/1-decanol substitution reaction, with subsystem A described using the cc-pVTZ basis set and with subsystem B described using either STO-3G or 6-31G. Panel (b) shows that as a function of the size of subsystem A, the error in the reaction energy relative to the full PBE/cc-pVTZ calculation rapidly converges for both mixed-basis methods. However, the similarity between the reaction-energy results does not extend to the calculated dipole moments: panel (c) shows that the PBE/cc-pVTZ:STO-3G gives substantially larger errors in the dipole moment for both the 1-chlorodecane (solid) and 1-decanol (dashed) than PBE/cc-pVTZ:6-31G, relative to the full PBE/cc-pVTZ results. The dipole-moment errors are quite small for the case of zero C-atoms in subsystem A, which corresponds to performing the full KS-DFT calculation with the smaller basis, indicating that the PBE/cc-pVTZ:STO-3G errors arise from the mismatch of basis sets rather than from the intrinsic inadequacy of the smaller basis set.

Panels (d)–(f) of Fig. 2.1 present the corresponding results for the deprotonation of

decanoic acid. For this case, the energy for the PBE/cc-pVTZ:STO-3G mixed-basis calculations converges very slowly as a function of the size of subsystem A. This convergence is greatly improved upon increasing the basis set size for subsystem B, as is seen from the PBE/cc-pVTZ:6-31G results. Interestingly, the reaction energy error is also avoided with the use of vacuum embedding, in which subsystem A is simply truncated with a hydrogen link atom. Panel (f) shows that PBE/cc-pVTZ:STO-3G likewise exhibits substantially larger errors in the decanoic acid (solid) and decanoate (dashed) dipole moments than PBE/cc-pVTZ:6-31G.

Fig. 2.1 yields two intriguing observations that will be explored in the following analysis. The first is that mixed-basis SCF calculations give rise to substantial errors in the electronic densities (as conveyed through the dipole moments) and in some cases also the reaction energies. The second is that even though the mixed-basis errors can largely be significantly reduced by going from a minimal basis set (such as STO-3G) to a marginally larger basis set (such as 6-31G), the complete neglect of basis functions and electronic density in subsystem B (as described using vacuum embedding and the use of a hydrogen link atom) can give better results than using a mixed-basis SCF description with a minimal basis set in subsystem B. Furthermore, we note that observations in Fig. 2.1 are completely consistent with those of DiLabio et al. [42], in that reaction energies involving charge-manifestation processes (in this case the deprotonation of decanoic acid) exhibit sizable mixed-basis errors, whereas reaction energies for processes that do not manifest charge (such as the 1-chlorodecane/1-decanol substitution reaction) are adequately described with mixed basis sets.

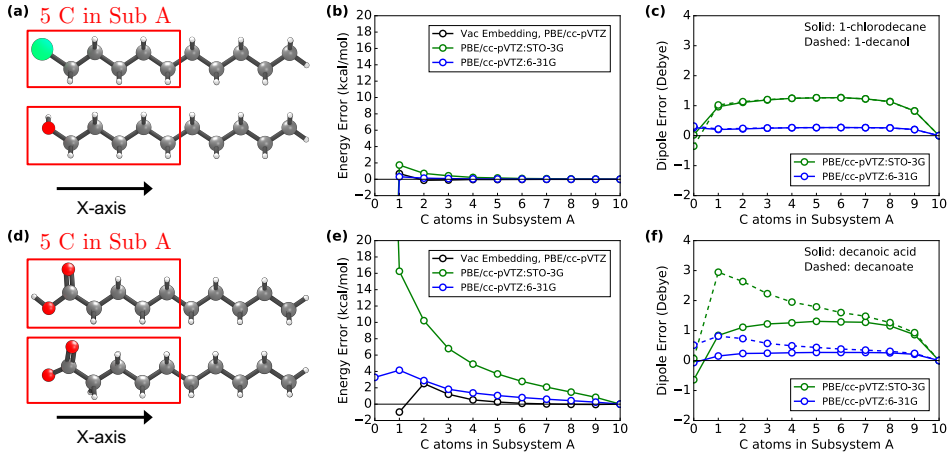

Figure 2.1: Geometries for (a) 1-chlorodecane and 1-decanol and (d) decanoic acid and decanoate. The molecular axis is specified by the x-axis. Error in the reaction energy as a function of the size of subsystem A for PBE/mixed-basis calculations of (b) the 1-chlorodecane/1-decanol substitution reaction and (e) the deprotonation of decanoic acid relative to the reaction energy calculated at the PBE/cc-pVTZ level of theory. Error in the dipole moment of (c) 1-chlorodecane and 1-decanol, and (f) decanoic acid and decanoate along the x-axis as a function of the size of subsystem A relative to the dipole moment calculated at the PBE/cc-pVTZ level of theory. The vacuum embedding calculation was performed with a hydrogen link atom. The reaction energy calculated at the PBE/cc-pVTZ level of theory is -71.8 kcal/mol for panel (b) and 356.1 kcal/mol for panel (e). The dipole moment along the x-axis calculated at the PBE/cc-pVTZ level of theory is -2.27 Debye for 1-chlorodecane, 1.21 Debye for 1-decanol, 0.343 Debye for decanoic acid and 22.8 Debye for decanoate. The solid lines and dashed lines in panel (c) represent 1-chlorodecane and 1-decanol respectively. The solid lines and dashed lines in panel (f) represent decanoic acid and decanoate respectively.

Mitigating the mixed-basis error using Fock-matrix corrections

Fig. 2.2 shows EMFT and FCEMFT results for the deprotonation of decanoic acid. For reference, the mixed-basis PBE/cc-pVTZ:STO-3G is included in replotted in panel (a) (green). The mixed-functional, mixed-basis EMFT results (blue) closely track the poor convergence of mixed-basis KS-DFT, as previously noted [17], indicating that the dominant source of error in these EMFT calculations is the use of different AO basis sets for the embedded subsystems. However, the mixed-functional, mixed-basis FCEMFT results (red) substantially improve this convergence in the calculated reaction energies for the deprotonation, indicating that Fock-matrix corrections significantly mitigate the errors in the reaction energy associated with the use of mixed-basis sets in the EMFT framework.

Similar trends are seen for the calculated dipole-moment errors in Fig. 2.2b. The mixed-functional, mixed-basis EMFT results again show results that are nearly identical to those obtained using mixed-basis KS-DFT (PBE/cc-pVTZ:STO-3G, Fig. 2.1f), whereas the errors in the dipole moment from FCEMFT are reduced.

Taken together, the results in Fig. 2.2 suggest that Fock-matrix corrections have the potential to significantly reduce both the reaction-energy and electronic-density errors arising in mixed-basis calculations SCF calculations in the EMFT framework, even with the use of a minimal STO-3G basis set in subsystem B.

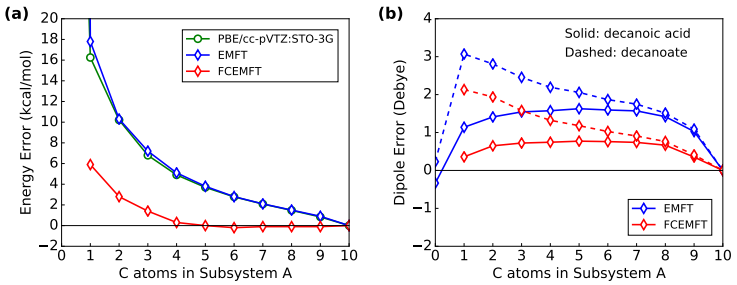

Figure 2.2: (a) Error in the reaction energy as a function of the size of subsystem A relative to the reaction energy calculated at the PBE/cc-pVTZ/DF level of theory. (b) Error in the dipole moment of decanoic acid and decanoate along the x-axis as a function of the size of subsystem A relative to the dipole moment calculated at the PBE/cc-pVTZ/DF level of theory. EMFT calculations correspond to PBE-in-LDA/cc-pVTZ:STO-3G/DF:DF(s) and FCEMFT calculations correspond to PBE-in-FCLDA/cc-pVTZ:STO-3G/DF:DF(s). The reaction energy calculated at the PBE/cc-pVTZ/DF level of theory is 356.2 kcal/mol. The dipole moment along the x-axis calculated at the PBE/cc-pVTZ/DF level of theory is 0.345 Debye for decanoic acid and 22.8 Debye for decanoate. The solid lines and dashed lines in panel (b) represent decanoic acid and decanoate respectively.

The underlying source of error: Density differences at the subsystem interface

The results in Figs. 2.1 and 2.2 suggest that electronic density errors introduced via mixed-basis KS-DFT and EMFT calculations, as manifested in the calculated dipole moments, give rise to reaction-energy errors for processes that involve large changes in Coulomb interactions, such as the deprotonation of decanoic acid. Fig. 2.2 further suggests that Fock-matrix corrections significantly reduce these errors in the calculated electronic densities, which in turn leads to improvements in the reaction energies. We now test this interpretation through a simple analysis of the electronic density differences associated with the mixed-basis calculations.

For decanoic acid and decanoate, Figs. 2.3a–2.3c present plots of the electronic density difference obtained from a mixed-basis KS-DFT calculation versus a full basis KS-DFT calculation performed at the PBE/cc-pVTZ level. All mixed-basis calculations in these plots correspond to the subsystem partitioning for which five carbon atoms are included in subsystem A (as illustrated in Fig. 2.1d). The electronic density distributions are reduced with respect to the y- and z-axes and are plotted as a function of the x-axis, along which the aliphatic chain is oriented; the position of each carbon atom along the x-axis is indicated by heavy black dots,

with vertical lines highlighting the fifth and sixth carbons that are separated by the subsystem interface. In comparing the density difference plots for decanoic acid and decanoate in Fig. 2.3c, the minor differences in geometry between the two structures is accounted for by offsetting the decanoate curve along the x-axis (by less than 0.1 Å) such that the positions of the fifth carbon for both structures coincide.

Considering first the density-difference plots for decanoic acid (Fig. 2.3a), it is clear that in the region of the acid proton (i.e., near carbon 1 at the left side of the plot), there is very little error introduced into the density through the use of mixed-basis calculations; the differences associated with both cc-pVTZ:STO-3G and cc-pVTZ:6-31G calculations essentially vanish for $x < -4$ bohr. However, in the interfacial region (carbons 4–6), the use of mixed basis-calculations with STO-3G is much worse than with 6-31G; indeed, cc-pVTZ:STO-3G exhibits the formation of a substantial dipole in the plotted density difference, with relative enhancement of the cc-pVTZ:STO-3G density in the interfacial region accompanied by relative depletion of this mixed-basis density further down the aliphatic chain.

Panel (b) illustrates that the same observations hold for decanoate, and panel (c) replots the cc-pVTZ:STO-3G density differences for both decanoic acid and decanoate to allow for direct comparison. The striking similarity for the mixed-basis density differences for the ion and neutral molecule indicate that the density error in the mixed-basis calculations is not caused by failure of the low-level system to account for polarization of the electronic density in the carbon-chain by the charge in the anion; instead, the observed mixed-basis density errors in panel (c) are essentially identical, irrespective of whether the system is charged or neutral. This is consistent with the previous observation that the cc-pVTZ:STO-3G mixed-basis calculations lead to substantial errors in the calculated dipole moments, regardless of whether the reaction involves manifestation of a charged species (Figs. 2.1c and 2.1f) and regardless of whether a large mixed-basis error in the reaction energy is correspondingly observed (Figs. 2.1b and 2.1e).

To more explicitly connect the mixed-basis density differences in Figs. 2.3a–2.3c to the calculated reaction energy errors for the acid deprotonation in Fig. 2.1e, we introduce a simple electrostatic model for the reaction energy error. This error is modeled as the electrostatic interaction of a single negative point charge on the carboxylate group of decanoate with the mixed-basis density difference,

$$E_{\mathrm{rxn}}^{(i)} = -\int_{x_{\mathrm{cut}}} \frac{q\,\Delta\rho^{(i)}(x)}{|x_q - x|}\,dx, \tag{2.6}$$

where $q = -1$ a.u. corresponds to the charge manifested on decanoate during the deprotonation reaction, $x_q = -10.2$ bohr is the approximate position of the manifested point charge (taken as the average x-coordinate of the two oxygen atoms in decanoate), $\Delta\rho^{(i)}(x)$ is the mixed-basis density difference for decanoate associated with a given number of carbon atoms i in subsystem A (which is plotted in Fig. 2.3b for the case of $\Delta\rho^{(5)}(x)$), and the overall negative sign arises from the negative charge of the electronic density. The integral in Eq. 2.6 is cutoff at $x_{\text{cut}} = -9.4$ bohr, to avoid the singularity in the integrand; the small magnitude of $\Delta\rho^{(i)}(x)$ in the region of the carboxylate group makes the results robust with respect to this cutoff position.

As a function of the number of carbon atoms included in subsystem A, Fig. 2.3d presents the results of this simple model with the reaction energy errors calculated from the mixed-basis KS-DFT calculations (replotted from Fig. 2.1e). The accuracy of the simple model is striking, quantitatively predicting the convergence properties of the full SCF calculations as a function of the subsystem size. This agreement demonstrates that it is the electrostatic interaction between the error in the mixed-basis density and the charge manifested in the deprotonation reaction that leads to the slow convergence of the mixed-basis reaction energy with respect to the subsystem size.

Fig. 2.4 presents the corresponding analysis for the EMFT and FCEMFT errors for the deprotonation reaction energy. The mixed-basis EMFT calculations yield large differences in the electronic density with respect to the full KS-DFT (PBE/cc-pVTZ) calculation, similar to the PBE/cc-pVTZ:STO-3G results in Figs. 2.3a and 2.3b. However, inclusion of the Fock-matrix corrections substantially reduces these density differences (Figs. 2.4a and 2.4b), even though a minimal basis set is still employed for describing subsystem B. Fig. 2.4c shows that, as seen previously for the mixed-basis KS-DFT calculations, the density difference obtained from EMFT for the reactant and product species is nearly identical. Finally, Fig. 2.4d demonstrates that the simple electrostatic model again accurately predicts the convergence of the reaction energy error as a function of subsystem size. As for the mixed-basis KS-DFT calculations, the dominant source of error in the reaction energy for EMFT is thus the formation of spurious density differences at the subsystem interface; FCEMFT largely corrects these errors in the reaction energy by reducing the errors in the underlying electronic density. These results emphasize the importance of developing methods, such as FCDFT, that provide corrections to both the energy and the density of a low-level theory; the fact that FCEMFT improves the convergence of

the reaction energy error with respect to system size is because it is also correcting the underlying electronic density distribution.

Finally, we note that the simple electrostatic model also explains the rapid convergence of the mixed-basis reaction energy errors for the 1-chlorodecane/1-decanol substitution reaction (Fig. 2.1b). Fig. 2.5 presents the PBE/cc-pVTZ:STO-3G mixed-basis density differences relative to the PBE/cc-pVTZ electronic density for the reactant and product, emphasizing that large errors in density are again observed, as is consistent with the large calculated errors in the dipole moments (Fig. 2.1c). However, since this reaction does not involve the manifestation of charge (i.e. $q = 0$ in Eq. 2.6), the simple model predicts that there will be no associated reaction energy error, as is consistent with the rapidly converging results with respect to subsystem size in Fig. 2.1b.

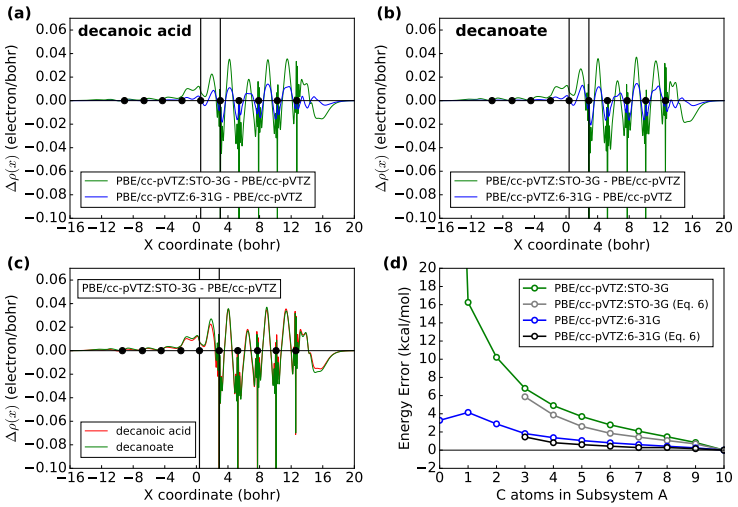

Figure 2.3: Density difference along the x-axis between the PBE/cc-pVTZ:STO-3G mixed-basis density with five carbons in subsystem A and the PBE/cc-pVTZ density for (a) decanoic acid and (b) decanoate. Heavy black dots correspond to the x coordinate of carbon atoms in the aliphatic chain, following the geometries in Fig. 2.1d. The vertical black lines highlight the boundary between subsystem A and subsystem B by identifying the fifth and sixth carbon in the chain. (c) Comparison of the PBE/cc-pVTZ:STO-3G mixed-basis densities with five carbons in subsystem A between decanoic acid and decanoate. The density differences are taken from panel (a) for decanoic acid and panel (b) for decanoate. (d) Reaction energy error as a function of the size of subsystem A for the electrostatic model in Eq. 2.6 and conventional PBE/mixed basis calculations for the deprotonation of decanoic acid.

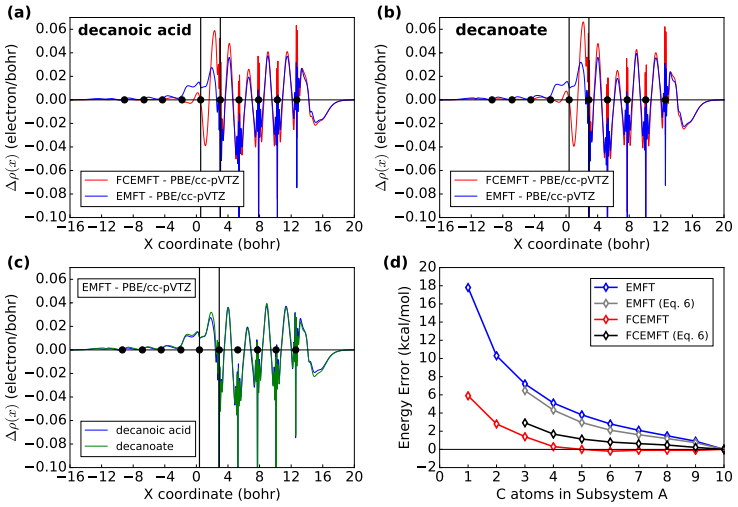

Figure 2.4: Density difference along the x-axis between the EMFT/FCEMFT density with five carbons in subsystem A and the PBE/cc-pVTZ density for (a) decanoic acid and (b) decanoate. EMFT calculations correspond to PBE-in-LDA/cc-pVTZ:STO-3G/DF:DF(s) and FCEMFT calculations correspond to PBE-in-FCLDA/cc-pVTZ:STO-3G/DF:DF(s). Heavy black dots correspond to the x coordinate of carbon atoms in the aliphatic chain, following the geometries in Fig. 2.1d. The vertical black lines highlight the boundary between subsystem A and subsystem B by identifying the fifth and sixth carbon in the chain. (c) Comparison of the EMFT densities with five carbons in subsystem A between decanoic acid and decanoate. The density differences are taken from panel (a) for decanoic acid and panel (b) for decanoate. (d) Reaction energy error as a function of size of subsystem A for the simple electrostatic model in Eq. 2.6, EMFT, and FCEMFT calculations for the deprotonation of decanoic acid.

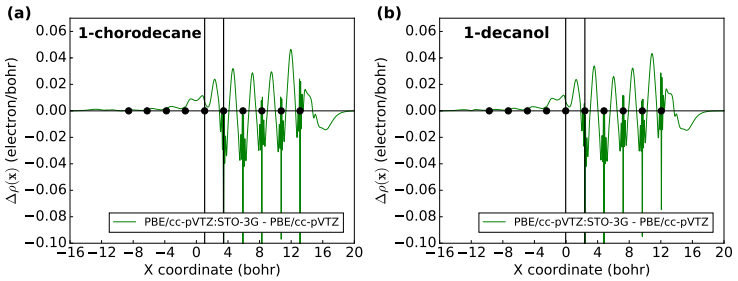

Figure 2.5: Density difference along the x-axis between the PBE/cc-pVTZ:STO-3G mixed-basis density with five carbons in subsystem A and the PBE/cc-pVTZ density for (a) 1-chlorodecane and (b) 1-decanol. Heavy black dots correspond to the x coordinate of carbon atoms in the aliphatic chain, following the geometries in Fig. 2.1a. The vertical black lines highlight the boundary between subsystem A and subsystem B by identifying the fifth and sixth carbon in the chain.

Implications for vacuum embedding and QM/MM link-atom selection

The analysis in the preceding section also provides insight into the surprisingly good convergence of the vacuum embedding reaction energies in Fig. 2.1e. Since vacuum embedding completely neglects electronic polarization effects in the environment, it might be expected to poorly describe a charge-manifestation reaction such as acid

deprotonation. However, the fact that mixed-basis errors arise from spurious density differences across the subsystem interface suggests that the choice of link atom in conventional QM/MM implementations implicitly corrects for this source of error.

Figure 2.6 illustrates this point using the decanoic acid deprotonation reaction energy error. For reference, the EMFT, mixed-basis KS-DFT, and vacuum embedding results with a hydrogen link atom are reproduced from Figs. 2.1e and 2.2a. Also plotted are vacuum embedding results in which fluorine is used for the link atom instead of hydrogen. Clearly, the use of fluorine as the link atom significantly degrades the vacuum embedding results, which become similar in magnitude, although opposite in sign, to the errors obtained in the mixed-basis KS-DFT and EMFT calculations without Fock-matrix corrections.

To mitigate the spurious density differences, and thus help to avoid the corresponding errors in charge-manifestation reactions, it is important to avoid introducing a significant discrepancy in the electronegativity at the interface, and this applies in both embedding and QM/MM calculations; indeed the strategy of tuning the properties of hydrogen link atoms to mitigate electronegativity mismatches has been employed in previous work [87–89]. Table 2.2 illustrates this point by reporting atomic electronegativities obtained at various levels of theory from the average of the atomic ionization potential and electron affinity. Comparison of the carbon and hydrogen electronegativities at the PBE/cc-pVTZ level shows that they are well matched, as is consistent with the rapid convergence of the vacuum embedding calculations with the choice of hydrogen for the link atom (Fig. 2.6). However, the fluorine atom at that level of theory is of course much more electronegative, leading to the poor convergence in the reaction energy seen in Fig. 2.6.

The electronegativity of the carbon atom at the PBE/STO-3G level also differs substantially from that of carbon at the PBE/cc-pVTZ level, explaining the poor convergence of the mixed-basis KS-DFT results in Fig. 2.6. Given that PBE/STO-3G carbon is much less electronegative than PBE/cc-pVTZ carbon, this result is also consistent with the fact that the mixed-basis KS-DFT errors converge from the opposite sign than the vacuum embedding results with the fluorine link atom in Fig. 2.6. Moreover, comparison of PBE/STO-3G carbon with LDA/STO-3G carbon in Table 2.2 reveals very similar electronegativities, which explains the nearly identical convergence behavior of EMFT (PBE-in-LDA/cc-pVTZ:STO-3G/DF:DF(s)) and mixed-basis KS-DFT in Fig. 2.6. Finally, it is clear that the inclusion of Fock-matrix corrections (FCLDA/STO-3G) brings the calculated electronegativity

of carbon into better agreement with PBE/cc-pVTZ, indicating that the mechanism by which the Fock-matrix corrections reduce the density difference and the reaction energy errors is by providing a better balance between the electronegativities of the atoms in the high- and low-level embedded subsystems.

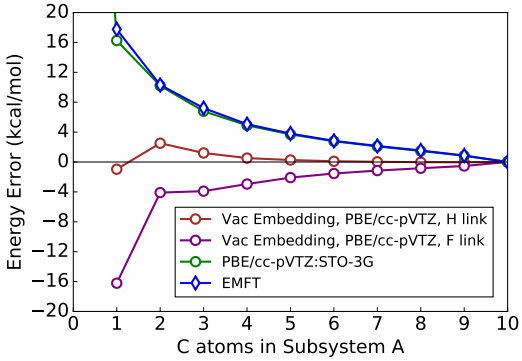

Figure 2.6: Error in the decanoic acid deprotonation reaction energy as a function of the size of subsystem A for mixed-basis, EMFT, and vacuum embedding calculations relative to the PBE/cc-pVTZ/DF level of theory. EMFT calculations correspond to PBE-in-LDA/cc-pVTZ:STO-3G/DF:DF(s).

Table 2.2: Atomic ionization potentials (IP), electron affinities (EA), and electronegativities at various levels of theory. All quantities reported in units of eV.

Atom	Method	IP	EA	Electronegativity
C	PBE/cc-pVTZ	11.53	0.95	6.24
	PBE/STO-3G	9.55	−5.93	1.81
	LDA/STO-3G	9.55	−5.84	1.85
	FCLDA/STO-3G	11.98	−3.48	4.25
F	PBE/cc-pVTZ	17.56	2.35	9.96
H	PBE/cc-pVTZ	13.60	−0.05	6.77

Negligible contributions from intramolecular BSSE

To complete our analysis, we consider the possible role of intramolecular basis-set superposition error (BSSE) in the slow convergence of the mixed-basis deprotonation reaction energy with respect to subsystem size. Intermolecular BSSE is known to be a serious effect for the evaluation of the binding energies, geometries, and vibrational modes for weakly bonded systems, such as the water dimer [90]. In addition, there are reports on the importance of intramolecular BSSE in terms of the energies and geometries for organic molecules [91, 92]. Since BSSE becomes

more severe when a small AO basis set is used, there is reason to expect that the mixed-basis calculations reported here with STO-3G in subsystem B will exhibit substantial artifacts due to BSSE. However, in determining whether BSSE gives rise to the slow convergence in the reaction energy for mixed-basis calculations, we focus not on the magnitude of the BSSE in any one molecular species, but instead on the degree to which BSSE effects cancel out between the reactant and the product.

In considering the effect of BSSE on mixed-basis calculations for the deprotonation of decanoic acid, we employ the counterpoise-correction method, [91, 92] considering the effect of ghost orbitals associated with the mixed basis set for the full molecule on an unrestricted KS-DFT calculation for the single carbon atom in subsystem B that is bonded to subsystem A, as this atom is expected to have the largest contribution to the intramolecular BSSE. For both the reactant and product of the deprotonation reaction, the lone carbon atom is described in the triplet state with STO-3G basis functions, the ghost orbitals associated with all other atoms in subsystem B are described using the STO-3G basis, and the ghost orbitals associated with all atoms in subsystem A are described using the cc-pVTZ basis. The PBE exchange-correlation functional and density fitting with the Ahlrichs density-fitting basis was used in all cases.

Fig. 2.7 reports the difference between the calculated BSSE for the reaction and product, as a function of the number of carbon atoms in subsystem A,

$$\Delta\text{BSSE} = \text{BSSE}^{\text{product}} - \text{BSSE}^{\text{reactant}}, \tag{2.7}$$

Although the individual BSSE to the reactant and produce species is quite large (i.e., as much as -37 kcal/mol, depending on subsystem size), this error almost exactly cancels between the reactant and product at each subsystem size, confirming that BSSE is not the source of the slow convergence of the reaction energy for the mixed-basis calculations of decanoic acid deprotonation.

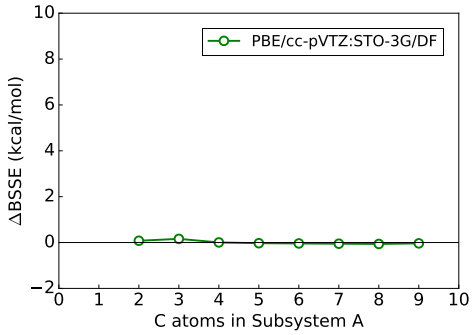

Figure 2.7: Differences in the BSSE corrections (Eq. 2.7) for the reactant and product of the deprotonation of decanoic acid, as a function of the size of subsystem A.

2.5 Conclusions

The use of mixed-based AO basis sets is a widely-used strategy for reducing the cost of electronic structure calculations, with relevance to KS-DFT, quantum embedding, and QM/MM calculations. The current chapter demonstrates that the use of mixed-basis approximations can give rise to substantial artifacts in the calculated electronic density, which can in turn give rise to large and slowly decaying errors in the calculation of reaction energies for processes that involve the manifestation of charge (including ionization, deprotonation, and electron attachment). In particular, for EMFT and other quantum embedding calculations for which the electronic density is not frozen, it is shown that artifacts in the calculated electronic density arise when the high- and low-level theories give an imbalanced description of atomic electronegativities. In the context of EMFT, it is shown that Fock-matrix corrections provide a straightforward means of addressing this imbalance, while preserving the efficiency of using a minimal basis to describe the low-level region. Finally, it is shown that similar issues arise in QM/MM embedding calculations, for which substantial artifacts in the reaction energy can arise if poor link-atom selection leads to an imbalance of its electronegativity with the atoms to which it is bonded.

Chapter 3

PROJECTION-BASED WAVEFUNCTION-IN-DFT EMBEDDING

Adapted from:

[1]S. J. R. Lee, M. Welborn, F. R. Manby, and T. F. Miller III, "Projection-Based Wavefunction-in-DFT Embedding", Acc. Chem. Res. **52**, 1359–1368 (2019) 10.1021/acs.accounts.8b00672.

This chapter reviews projection-based quantum embedding for electronic structure, which provides a formally exact method for density functional theory (DFT) embedding. The method also provides a rigorous and accurate approach for describing a small part of a chemical system at the level of a correlated wavefunction (WF) method while the remainder of the system is described at the level of DFT. A key advantage of projection-based embedding is that it can be formulated in terms of an extremely simple level-shift projection operator, which eliminates the need for any optimized effective potential calculation or kinetic energy functional approximation, while simultaneously ensuring that no extra programming is needed to perform WF-in-DFT embedding with an arbitrary WF method. The current work presents the theoretical underpinnings of projection-based embedding, describes use of the method for combining wavefunction and density-functional theories, and discusses technical refinements that have improved the applicability and robustness of the method.

Applications of projection-based WF-in-DFT embedding are also reviewed, with particular focus on recent work on transition-metal catalysis, enzyme reactivity, and battery electrolyte decomposition. In particular, we review the application of projection-based embedding for the prediction of electrochemical potentials and reaction pathways in a Co-centered hydrogen evolution catalyst. Projection-based WF-in-DFT calculations are shown to provide quantitative accuracy while greatly reducing the computational cost as compared to a reference coupled cluster calculation on the full system. Additionally, projection-based WF-in-DFT embedding is used to study the mechanism of citrate synthase; it is shown that projection-based WF-in-DFT largely eliminates the sensitivity of the potential energy landscape to the employed DFT exchange-correlation functional. Finally, we demonstrate the

use of projection-based WF-in-DFT to study electron transfer reactions associated with battery electrolyte decomposition. Projection-based WF-in-DFT embedding is used to calculate the oxidation potentials of neat ethylene carbonate (EC), neat dimethyl carbonate (DMC), and 1:1 mixtures of EC and DMC, to overcome qualitative inaccuracies in the electronic densities and ionization energies obtained from conventional DFT methods. By further embedding the WF-in-DFT description in a molecular mechanics point-charge environment, this work enables an explicit description of the solvent and ensemble averaging of the solvent configurations.

3.1 Introduction

The use of density functional theory (DFT) for the description of electronic structure has gained remarkable prevalence in recent years, due to its reasonable compromise between accuracy and computational cost [93]. However, the chemical sciences are permeated with systems for which the approximations of DFT fundamentally break down or for which the computational cost of DFT remains prohibitive for the molecular dynamics (MD) simulation of necessary length- and timescales. Although correlated wavefunction (WF) electronic structure methods, such as coupled-cluster theory, provide better accuracy than DFT for single point calculations on systems of modest size, they have been too expensive to allow for widespread use in terms of exploring conformational landscapes and reaction pathways.

To mitigate the trade-off between accuracy and computational cost, quantum embedding has emerged as a powerful strategy for modeling the electronic structure of complex systems. In embedding methods, a high-level quantum-mechanical description of a chemically active subsystem is embedded in a surrounding environment described using a more approximate theory. By exploiting the intrinsic locality of molecular interactions, this approach provides high accuracy for regions that demand it while avoiding the computational cost of a high-level calculation on the whole system. Notable examples of embedding include QM/MM [94, 95], ONIOM [76], fragmentation methods [11, 15, 96], density functional embedding [5–10, 13, 14, 97–105], and density matrix embedding [12, 17, 106, 107], although there are many manifestations of the idea.

Projection-based embedding [9] describes subsystem interactions at the level of DFT and allows for the partitioning of the subsystems across covalent and even conjugated bonds, and it enables the use of relatively small subsystem sizes for an embedded WF description. The current chapter aims to provide a practical review of the projection-based embedding method, including a description of its theory, implementation, applications, and limitations. Although we describe the methodological context for projection-based embedding, we also direct the reader to several reviews that provide a more complete description of alternative approaches [11, 76, 108–112].

3.2 Projection-Based Embedding

Quantum embedding methods developed within the framework of DFT offer a formally exact approach to electronic structure calculations in which complex chemical

problems are decomposed into the solution of individual smaller subsystems [103, 112]. Throughout this review, we shall use the term "exact" to denote that a DFT-in-DFT embedding calculation where both subsystems are treated using the same exchange-correlation (XC) functional yields the same result as a single Kohn-Sham (KS) DFT calculation performed over the full system. In principle, DFT embedding thus avoids the uncontrolled approximations (such as link atoms) that appear in widely used methods, such as QM/MM and ONIOM.

In practice, however, many DFT embedding studies employ substantial approximations in the description of subsystem interactions. The subsystem interaction potentials that emerge in the DFT embedding framework include non-additive kinetic potential (NAKP) terms that enforce Pauli exclusion between the electrons of the various subsystems [112]. Without knowledge of the exact functional for the non-interacting kinetic energy, this has typically required approximate NAKP treatments that break down in cases for which the subsystem densities significantly overlap (which include hydrogen-bonded or covalently bonded subsystems) [100, 112, 113], limiting applications to those involving weakly interacting subsystems. Although numerically exact DFT embedding methods have been developed that determine NAKP contributions via an optimized effective potential (OEP) inversion of the density [6, 8, 102, 104, 114–116], OEP inversion can be ill-conditioned and requires careful regularization protocols [10, 102, 117–120].

Projection-based embedding avoids these issues by providing a numerically exact DFT-in-DFT embedding framework that eliminates the NAKP contributions via the mutual orthogonalization of the subsystem molecular orbitals. Panel (a) of Fig. 3.1 outlines the general procedure of a projection-based embedding calculation. A KS-DFT calculation is first performed on the full system to self-consistently determine the KS orbitals and the corresponding Fock matrix, $\mathbf{F}$. The occupied KS orbitals are then localized, shown on the left side of panel (a) of Fig. 3.1. These localized molecular orbitals (LMOs) are partitioned into subsystems A (in red) and B (in blue), with corresponding atomic-orbital-basis density matrices, $\boldsymbol{\gamma}^{\mathrm{A}}$ and $\boldsymbol{\gamma}^{\mathrm{B}}$.

To determine the subsystem Fock matrix, $\mathbf{F}^{\mathrm{A}}$, that describes the electrons of subsystem A in the environment of the density matrix of subsystem B, we begin with

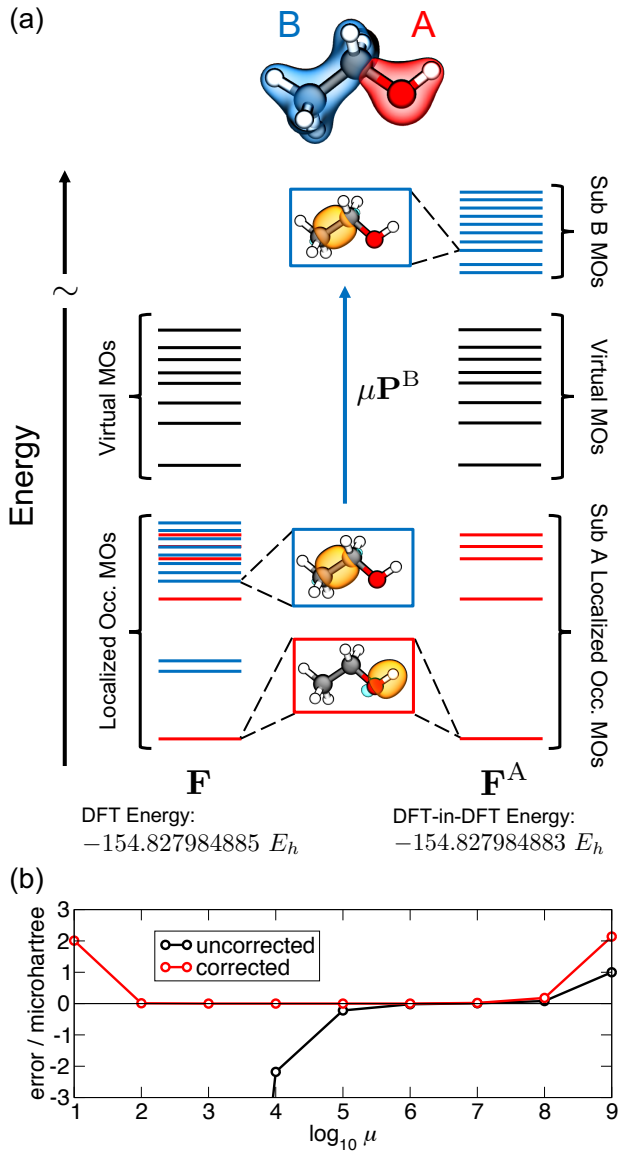

Figure 3.1: (a) Demonstration of projection-based embedding, using the example of embedding the 10 electrons of the –OH moiety of ethanol in the environment of the ethyl subsystem. **F** is the full system KS-DFT Fock matrix that is initially used to self-consistently determine the occupied KS MOs of the full system. The occupied MOs are then localized and grouped into subsystems A and B (red and blue respectively). $\mathbf{F}^{\mathrm{A}}$ is the embedded Fock matrix (Eq. 3.4) for the subsystem A electrons, which includes the projection operator, $\mu\mathbf{P}^{\mathrm{B}}$ (Eq. 3.5). Following projection, the subsystem A LMOs (red) are explicitly orthogonalized with respect to the subsystem B LMOs (blue), thus eliminating non-additive kinetic energy contributions. (b) Error in the uncorrected (Eq. 3.1) and corrected (Eq. 3.1 + Eq. 3.8) PBE-in-PBE/6-31G* energy expressions relative to full KS-DFT on ethanol using PBE/6-31G*, demonstrating that the perturbative correction (Eq. 3.8) yields essentially exact embedding energies over a wide range of μ values. Adapted with permission from Ref. 9. Copyright 2012 American Chemical Society.

the DFT-in-DFT energy expression for projection-based embedding,

$$\begin{aligned} E_{\text{DFT-in-DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}};\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right] &= E_{\text{DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}\right] \\ &+ E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}}+\boldsymbol{\gamma}^{\text{B}}\right] - E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}}\right] \\ &+ \text{tr}\left[(\tilde{\boldsymbol{\gamma}}^{\text{A}}-\boldsymbol{\gamma}^{\text{A}})\mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right]\right] \\ &+ \mu\text{tr}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}\mathbf{P}^{\text{B}}\right], \end{aligned} \tag{3.1}$$

where E_{DFT} denotes the KS-DFT energy evaluated using the bracketed density matrix, $\tilde{\boldsymbol{\gamma}}^{\text{A}}$ is the embedded subsystem A density matrix, and $\mathbf{P}^{\text{B}}$ is a projection operator that enforces the mutual orthogonalization of subsystem A and B LMOs. The embedding potential, $\mathbf{v}_{\text{emb}}$, describes all interactions between subsystems A and B,

$$\mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right] = \mathbf{g}\left[\boldsymbol{\gamma}^{\text{A}}+\boldsymbol{\gamma}^{\text{B}}\right] - \mathbf{g}\left[\boldsymbol{\gamma}^{\text{A}}\right]. \tag{3.2}$$

In general, $\mathbf{v}_{\text{emb}}$ would also include the difficult-to-evaluate NAKP contributions, but if the subsystem densities are constructed from disjoint subsets of orthogonal orbitals, these NAKP terms are exactly zero [9]. The matrix $\mathbf{g}$ is the density-matrix functional of two-electron terms, given by

$$\begin{aligned} \left(\mathbf{g}\left[\boldsymbol{\gamma}\right]\right)_{\kappa\nu} &= \sum_{\lambda\sigma}\gamma_{\lambda\sigma}\left[(\kappa\nu|\lambda\sigma) - \tfrac{1}{2}x(\kappa\lambda|\nu\sigma)\right] \\ &+ \left(\mathbf{v}_{\text{xc}}\left[\boldsymbol{\gamma}\right]\right)_{\kappa\nu}, \end{aligned} \tag{3.3}$$

where κ, ν, λ and σ label atomic orbital basis functions, $(\kappa\nu|\lambda\sigma)$ are two-electron repulsion integrals, x is the fraction of exact exchange, and $\mathbf{v}_{\text{xc}}$ is the exchange-correlation potential matrix.

The subsystem Fock matrix corresponding to variation of equation 3.1 with respect to $\tilde{\boldsymbol{\gamma}}^{\text{A}}$ is then

$$\begin{aligned} \mathbf{F}^{\text{A}} &= \frac{\partial}{\partial\tilde{\boldsymbol{\gamma}}^{\text{A}}}E_{\text{DFT-in-DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}};\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right] \\ &= \mathbf{h} + \mathbf{g}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}\right] + \mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right] + \mu\mathbf{P}^{\text{B}}, \end{aligned} \tag{3.4}$$

where $\mathbf{h}$ is the standard one-electron Hamiltonian. Self-consistent optimization of $\mathbf{F}^{\text{A}}$ with respect to $\tilde{\boldsymbol{\gamma}}^{\text{A}}$ recovers the original subsystem A density matrix, $\boldsymbol{\gamma}^{\text{A}}$ (Fig. 1, right column) for the case of DFT-in-DFT embedding when both subsystems are described using the same XC functional.

A practical way to enforce the orthogonality of the subsystem A orbitals to those in subsystem B is to introduce a level-shift operator of the form [9]

$$\mu\mathbf{P}^{\text{B}} = \mu\mathbf{S}\boldsymbol{\gamma}^{\text{B}}\mathbf{S}, \tag{3.5}$$

where $\mathbf{S}$ is the overlap matrix in the atomic orbital basis, and μ is a positive scalar number. The action of this operator is to level-shift the subsystem B LMOs to high energies so that they cannot hybridize with those of subsystem A (shown on the right side of panel (a) of Fig. 3.1) [121, 122]. In the $\mu \rightarrow \infty$ limit, Eq. 3.1 reduces to the KS-DFT energy for the full system, such that the projection-based approach is exact for DFT-in-DFT embedding.

Embedding methods that maintain orthogonality between subsystem orbitals have long been in use, including the Philips-Kleinman pseudopotential approach [121]. What had not been previously recognized is that these same strategies can be used to formulate a formally exact method for DFT embedding [9].

DFT-in-DFT Embedding

Before proceeding, it is worth expanding on several aspects of DFT-in-DFT embedding. First, it is clear from the preceding discussion that the projection-based approach allows for a description of DFT-in-DFT embedding with subsystems A and B evaluated using different XC functionals. Typically this involves using a more expensive (i.e. hybrid, meta-GGA, etc.) functional to describe subsystem A and a computationally cheaper (i.e. GGA or LDA) functional to describe subsystem B.

The procedure for this type of DFT-in-DFT embedding calculations begins with performing a low-level KS-DFT calculation on the full system, yielding $E_{\mathrm{DFT}}[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}]$. The resulting occupied MOs are localized and partitioned into subsystems A and B, which are used to form the matrices $\boldsymbol{\gamma}^{\mathrm{A}}$, $\boldsymbol{\gamma}^{\mathrm{B}}$, $\mathbf{P}^{\mathrm{B}}$, and $\mathbf{v}_{\mathrm{emb}}[\boldsymbol{\gamma}^{\mathrm{A}}, \boldsymbol{\gamma}^{\mathrm{B}}]$ and to evaluate $E_{\mathrm{DFT}}[\boldsymbol{\gamma}^{\mathrm{A}}]$; these quantities are unchanged during the self-consistent field (SCF) iterations for the embedded subsystem. The SCF iterations for the embedded subsystem are performed to optimize the subsystem density matrix, $\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}$. At each SCF iteration, $\mathbf{F}^{\mathrm{A}}$ is calculated (Eq. 3.4) and diagonalized; $\mathbf{g}[\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}]$ is the only term in Eq. 3.4 to be re-evaluated at each SCF iteration, and it is done so using the high-level XC functional. Finally, to obtain the total DFT-in-DFT energy, the converged subsystem density, $\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}$, is used to evaluate $E_{\mathrm{DFT}}[\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}]$ using the high-level XC functional, as well as the last two traces on the right hand side of Eq. 3.1.

In light of the quantities that must be iteratively re-evaluated during an embedding calculation that involves two different levels of theory, we provide a few additional comments regarding the derivation of the total DFT-in-DFT energy expression. The

starting point for Eq. 3.1 is the more transparent energy expression

$$E_{\text{DFT-in-DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}};\boldsymbol{\gamma}^{\text{B}}\right] = E_{\text{DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}+\boldsymbol{\gamma}^{\text{B}}\right] + \mu\text{tr}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}\mathbf{P}^{\text{B}}\right]. \tag{3.6}$$

Minimization of this total energy expression with respect to the density matrix for subsystem A would lead to an expression for the subsystem Fock matrix $\mathbf{F}^{\text{A}}$ that involves the costly re-evaluation of the embedding potential in terms of the high-level XC functional at each SCF iteration. To avoid this, a first-order expansion in $\tilde{\boldsymbol{\gamma}}^{\text{A}} - \boldsymbol{\gamma}^{\text{A}}$ is performed, yielding

$$\begin{aligned} E_{\text{DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}+\boldsymbol{\gamma}^{\text{B}}\right] - E_{\text{DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}\right] \approx \\ E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}}+\boldsymbol{\gamma}^{\text{B}}\right] - E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}}\right] \\ + \text{tr}\left[(\tilde{\boldsymbol{\gamma}}^{\text{A}}-\boldsymbol{\gamma}^{\text{A}})\mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right]\right]. \end{aligned} \tag{3.7}$$

Rearranging Eq. 3.7 and substituting it into Eq. 3.6 yields Eq. 3.1. Note that if both subsystems are described using the same XC functional, this perturbative approximation becomes exact in the limit of mutual orthogonalization of the subsystem orbitals, since $\tilde{\boldsymbol{\gamma}}^{\text{A}}$ approaches $\boldsymbol{\gamma}^{\text{A}}$. If different exchange-correlation functionals are employed for the two subsystems, then Eq. 3.7 can lead to density-driven errors associated with the fact that the density matrix obtained using the low-level theory is different from that obtained using the high-level theory [7, 10, 22, 74].

Finally, we discuss the convergence of the projection-based embedding description with respect to the level-shift parameter, μ. Although the projection operator in Eq. 3.4 only exactly enforces orthogonality between subsystem A and B orbitals in the limit of $\mu \to \infty$, finite values of μ in the range of 10^4 a.u. to 10^7 a.u. are consistently found to yield accurate results (panel (b) Fig. 3.1, black), regardless of chemical system [9]; a default choice of $\mu = 10^6$ a.u. has been found in almost all cases to yield microhartree-scale embedding errors. Nonetheless, if greater accuracy is needed, then a perturbative correction

$$\mu\text{tr}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}\mathbf{P}^{\text{B}}\right] \tag{3.8}$$

can be added to the DFT-in-DFT energy expression to account for the finiteness of μ [9], and this typically leads to sub-microhartree accuracy over a very large range of μ values (panel (b) Fig. 3.1, red). Errors associated with finite values of μ can also be avoided by enforcing the projection via explicit orthogonalization [123–127] of the subsystem orbitals, at some cost to the simplicity of the implementation.

Wavefunction-in-DFT Embedding

Beyond DFT-in-DFT embedding, the projection-based approach readily allows for wavefunction-in-DFT (WF-in-DFT) embedding, in which subsystem A is treated using a WF-level description and subsystem B is described at the DFT level [9]. Starting from the DFT-in-DFT energy expression in Eq. 3.1, the WF-in-DFT energy is simply obtained by substituting the DFT energy of subsystem A with the corresponding WF energy,

$$\begin{aligned} E_{\text{WF-in-DFT}}\left[\tilde{\Psi}^{\text{A}};\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right] &= E_{\text{WF}}\left[\tilde{\Psi}^{\text{A}}\right] \\ &+ \text{tr}\left[(\tilde{\boldsymbol{\gamma}}^{\text{A}}-\boldsymbol{\gamma}^{\text{A}})\mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right]\right] \\ &+ E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}}+\boldsymbol{\gamma}^{\text{B}}\right]-E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}}\right] \\ &+ \mu\text{tr}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}\mathbf{P}^{\text{B}}\right], \end{aligned} \tag{3.9}$$

where $\tilde{\Psi}^{\text{A}}$ is the WF for subsystem A, $\tilde{\boldsymbol{\gamma}}^{\text{A}}$ is the one-particle reduced density matrix corresponding to $\tilde{\Psi}^{\text{A}}$, and $E_{\text{WF}}[\tilde{\Psi}^{\text{A}}]$ is the WF energy of subsystem A [9, 14].

A projection-based WF-in-DFT embedding calculation proceeds as follows. A KS-DFT calculation is first performed over the full system. The resulting occupied MOs are localized and partitioned into two sets, corresponding to subsystems A and B. These sets are used to construct $\mathbf{h}^{\text{A-in-B}}$,

$$\mathbf{h}^{\text{A-in-B}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right] = \mathbf{h}+\mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right]+\mu\mathbf{P}^{\text{B}}, \tag{3.10}$$

which is an effective one-electron Hamiltonian containing the standard one-electron Hamiltonian, the embedding potential and the projection operator. Finally, a correlated WF calculation is performed on subsystem A wherein $\mathbf{h}^{\text{A-in-B}}$ replaces the standard one-electron Hamiltonian. The final WF-in-DFT energy is given by equation 3.9.

The WF calculation for subsystem A consists of two steps: first, a set of reference orbitals is generated and second, a correlated WF calculation is performed using those orbitals. The reference orbitals can be obtained either via Hartree-Fock (HF) or a multiconfigurational method. For the former case, the subsystem A post-HF calculation begins with HF-in-DFT embedding. The HF-in-DFT Fock matrix, $\mathbf{F}^{\text{A}}$, is derived by inserting a Slater determinant for the subsystem A WF into Eq. 3.9 and differentiating with respect to $\tilde{\boldsymbol{\gamma}}^{\text{A}}_{\text{HF}}$, giving

$$\begin{aligned} \mathbf{F}^{\text{A}} &= \frac{\partial}{\partial\tilde{\boldsymbol{\gamma}}^{\text{A}}_{\text{HF}}}E_{\text{HF-in-DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}_{\text{HF}};\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right] \\ &= \mathbf{h}^{\text{A-in-B}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right]+\mathbf{g}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}_{\text{HF}}\right], \end{aligned} \tag{3.11}$$

where $\mathbf{g}$ includes all of the usual HF two-electron terms, and $\mathbf{h}^{\text{A-in-B}}$ represents the effective one-electron Hamiltonian given by Eq. 3.10. Once the subsystem A HF MOs are optimized in the presence of the DFT embedding potential, $\mathbf{v}_{\text{emb}}$, they are used for the correlated subsystem A post-HF calculation. An analogous procedure holds for the case of multireference methods, wherein a multiconfigurational WF is substituted in place of the single Slater determinant in Eq. 3.9 [25, 128]. In this way, projection-based WF-in-DFT embedding can be readily performed with any existing WF method (or quantum impurity solver) simply by modifying the one-electron Hamiltonian in the WF method to include the projection-based embedding terms.

While projection-based embedding is exact for (same functional) DFT-in-DFT embedding, projection-based WF-in-DFT embedding is necessarily approximate. Some sources of error in WF-in-DFT embedding have been analyzed, including the approximate nature of the non-additive exchange-correlation energy [14] and density-driven errors in the underlying DFT calculation [22].

Technical Comments

Choice of Localization Method

Projection-based DFT-in-DFT embedding is formally exact with any disjoint partition of orthogonal orbitals. A multiscale embedding method, in which a spatially localized subsystem is treated at a different level of theory from its surroundings, can be constructed by selecting these subsets from a set of spatially localized orbitals.

Methods based on the Pipek-Mezey criterion [129], which maintain a chemically intuitive separation between orbitals (e.g. σ-type and π-type), have consistently been found to perform well in projection-based embedding [14, 18]. In particular, Knizia's intrinsic bond orbitals (IBOs) [130] have been found to be effective at producing compact LMOs that vary relatively smoothly with respect to nuclear coordinates [18]. We note the quality of the embedding depends on the degree to which the molecular orbitals can be localized; for this reason intrinsically delocalized systems, such as metals, remain a challenge [128].

Atomic Orbital Basis Set Truncation

As described thus far, projection-based WF-in-DFT embedding reduces the number of occupied LMOs that are correlated at the WF level, but leaves the virtual space untouched. Since the cost of WF methods scales steeply with the number of virtual

orbitals (e.g. $O(v^4)$ for CCSD), the spatial locality of subsystem A can be used to reduce the effective size of the virtual space for the embedded WF calculation. One strategy for doing so is to employ local correlation techniques for the WF method [20, 123]; local correlation methods include parameters that allow for the control of the length scale over which excitations are included [131]. A more general strategy for limiting the size of the virtual space in the WF calculation is to truncate the atomic orbital (AO) basis set employed in the WF-in-DFT calculation.

AO truncation for projection-based embedding have been devised to discard AOs on the basis of either distance from subsystem A [13] or magnitude of contribution to the Mulliken population of subsystem A [18]. The latter method has been found to be particularly simple and robust to employ in practice, determining whether to retain each AO via a single density threshold parameter: if the net Mulliken population – computed using the subsystem A density – of an AO is less than the specified threshold, it is removed from the basis set. In practice, we have found that a density threshold of 1×10^{-4} a.u. provides a good balance between speed and accuracy, but that system-specific sensitivity checks should always be performed.

The energy expression for a projection-based WF-in-DFT calculation with AO truncation is

$$\begin{aligned} &E^{\text{trun}}_{\text{WF-in-DFT}}\left[\tilde{\Psi}^{\text{A,trun}}; \bar{\boldsymbol{\gamma}}^{\text{A,trun}}; \boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right] = \\ &\quad E_{\text{WF}}\left[\tilde{\Psi}^{\text{A,trun}}\right] \\ &+ \text{tr}\left[(\tilde{\boldsymbol{\gamma}}^{\text{A,trun}} - \bar{\boldsymbol{\gamma}}^{\text{A,trun}})\mathbf{v}^{\text{trun}}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right]\right] \\ &- E^{\text{trun}}_{\text{DFT}}\left[\bar{\boldsymbol{\gamma}}^{\text{A,trun}}\right] + E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right] \\ &+ \mu\text{tr}\left[\tilde{\boldsymbol{\gamma}}^{\text{A,trun}}\mathbf{P}^{\text{B,trun}}\right], \end{aligned} \tag{3.12}$$

where terms superscripted by "trun" are represented in the truncated AO basis and those without are evaluated in the full basis. The matrices $\mathbf{v}^{\text{trun}}_{\text{emb}}$ and $\mathbf{P}^{\text{B,trun}}$ are first formed in the full AO basis (Eqs. 3.2 and 3.5) and then projected onto the truncated basis by removing the rows and columns that correspond to the truncated basis functions. The matrix $\bar{\boldsymbol{\gamma}}^{\text{A,trun}}$ is the subsystem A one-particle density optimized at the DFT level in the truncated basis. In Eq. 3.12, the leading order error due to AO truncation is corrected at the DFT-in-DFT level (see Eq. 11 in Ref. 18).

Atomic orbital truncation has been shown to greatly speedup up projection-based WF-in-DFT calculations at a small cost in accuracy in total and relative energies [18, 21, 24, 25, 132]. AO truncation can also be applied to DFT-in-DFT embedding to reduce the cost of the high-level DFT calculation.

Even-Handed Subsystem Partitioning

We now address the practical issue of how best to partition the LMOs between subsystems in applications to chemical reactions. It is convenient and chemically intuitive to associate a set of atoms with subsystem A, and then to automatically select LMOs corresponding to the subsystem A atoms. Typically, this is done by selecting all LMOs with a significant population on those atoms chosen to comprise subsystem A; populations are typically assigned using the atomic population scheme that corresponds to the localization method used to generate the LMOs (i.e., Mulliken populations for Pipek-Mezey localization [129] and intrinsic atomic orbital populations for IBO localization [130]).

This charge-selection strategy provides a good starting point for determining the subsystem A LMOs, but it becomes problematic when applied to processes for which charge-selected LMOs move into or out of subsystem A as a function of molecular geometry. When this occurs, substantial error is incurred and the projection-based embedded potential energy profile can become discontinuous. Such problems often arise in cases involving bond formation or breaking.

To address this problem, we have recently reported an "even-handed" LMO selection strategy [128] which forms a consensus set of subsystem A LMOs to be used at all geometries along a reaction coordinate. For every geometry, this set contains every LMO that is charge-selected at any geometry. The even-handed LMO selection procedure is automatic, uses information already available at the DFT level, and requires no user input beyond the set of atoms to be embedded (the same input as in the charge-selection method). Even-handed selection has been empirically demonstrated to result in smooth and quantitative energy profiles at the cost of only a few additional LMOs included in subsystem A [25, 128].

3.3 Selected Applications

Among the most important aspects of projection-based embedding is that it enables robust and efficient WF-in-DFT calculations in complex chemical systems. To date, projection-based embedding has been used in applications studies from the groups of the authors [16, 19–21, 24, 25, 132] and others [133–139], including applications to periodic systems [140, 141]. We now summarize applications of the method to transition-metal catalysis, enzyme catalysis, and electrochemistry.

Transition-Metal Catalysis

In a first example [20], WF-in-DFT embedding was employed to investigate a new class of cobalt-based catalysts for hydrogen evolution. A central challenge in the development of inorganic hydrogen-evolution catalysts is to avoid deleterious coupling of the energetics of metal-site reduction from the kinetics of metal-hydride formation.

Collaboration between theory and experiment identified a family of cobalt diimine-dioxime catalysts that shows promise for achieving this aim by introducing an intramolecular proton-shuttle via a pyridyl pendant group (Fig. 3.2B). For the intramolecular proton-shuttle reaction in this system, WF-in-DFT was found to converge with a high-level subsystem that included only the LMOs on the central transition-metal atoms and its first coordination sphere (Fig. 3.2C).

For this reaction, Fig. 3.2A demonstrates the degree to which (local CCSD(T))-in-DFT embedding (red) can remove the qualitative errors of DFT using the B3P86 functional (black), achieving quantitative agreement with local CCSD(T) performed over the full system. Furthermore, the excellent accuracy of the embedding calculation was achieved while reducing the computational cost of the full wavefunction calculation from 20 hours per energy evaluation down to just a single hour per energy evaluation [20]. As is illustrated in this application, transition metal complexes provide very fruitful application domain for projection-based embedding, given that they typically involve subtle electronic structure in the vicinity of the metal that demands a wavefunction theory description, while the surrounding ligand environment is typically both very expensive for wavefunction theories and adequately described using DFT. For these reasons, several other applications of projection-based WF-in-DFT embedding have also focused on transition-metal complexes [19, 25, 128, 135].

Enzyme Catalysis

Application of projection-based WF-in-DFT embedding to the reactivity of the citrate synthase enzyme illustrates the potential of the method to help elucidate biochemical mechanisms and predict pharmacological activity (Figure 3.3) [21, 24]. DFT is commonly used for the QM region of QM/MM calculations of enzyme systems because of its low cost and the ready availability of implementations. Prediction and understanding of activation barriers using DFT is hampered by the sometimes very large sensitivity on the choice of approximate exchange-correlation

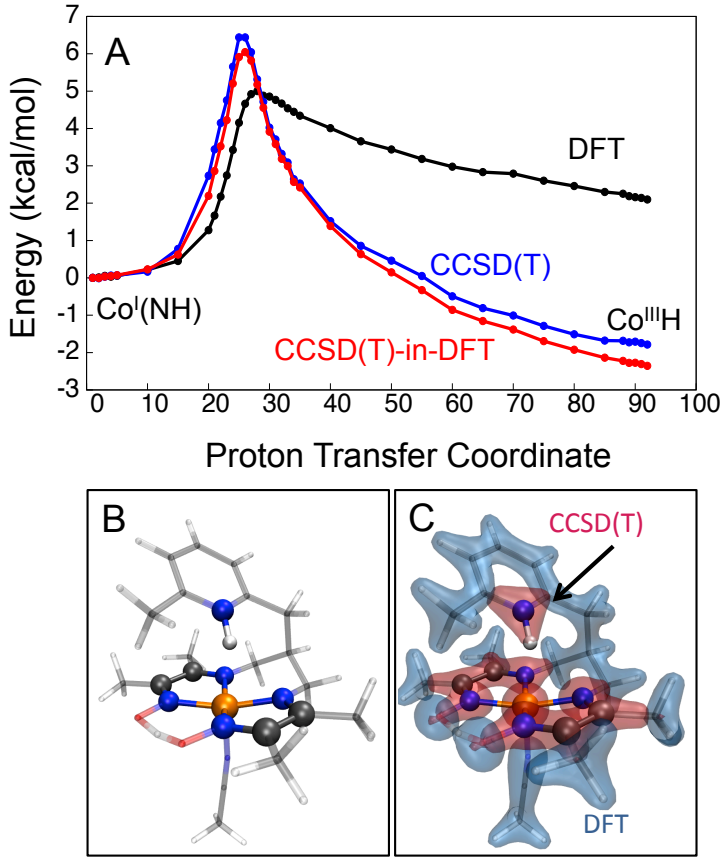

Figure 3.2: (A) Benchmark energy profiles for the rate-limiting intramolecular proton-transfer reaction in a new class of cobalt diimine-dioxime catalysts, obtained using CCSD(T) (blue), B3P86 (black), and CCSD(T)-in-B3P86 embedding (red). (B) Partitioning of the system in the CCSD(T)-in-B3P86 calculations into atoms that are treated using CCSD(T) (solid ball-stick) and B3P86 (stick). (C) The associated partitioning of the electronic density into subsystems that are treated using CCSD(T) (red) and B3P86 region (blue). Adapted with permission from Ref. 20. Copyright 2016 American Chemical Society.

functional. A combination of expertise, experience, and careful benchmarking can help identify a functional that should be reliable for a particular case, but we have found that WF-in-DFT embedding can almost completely eliminate this dependence with modest computational cost.

Specifically, the figure illustrates that whereas DFT with various XC functionals predict qualitatively different reaction energy profiles for the proton abstraction from acetyl-coenzyme A in the citrate synthase (panel A), CCSD(T)-in-DFT embedding provides nearly identical energy curves when the environment is described using DFT with the corresponding XC functionals. It is clear that in this case, the error in the DFT reaction profiles is associated with the local description of the chemical

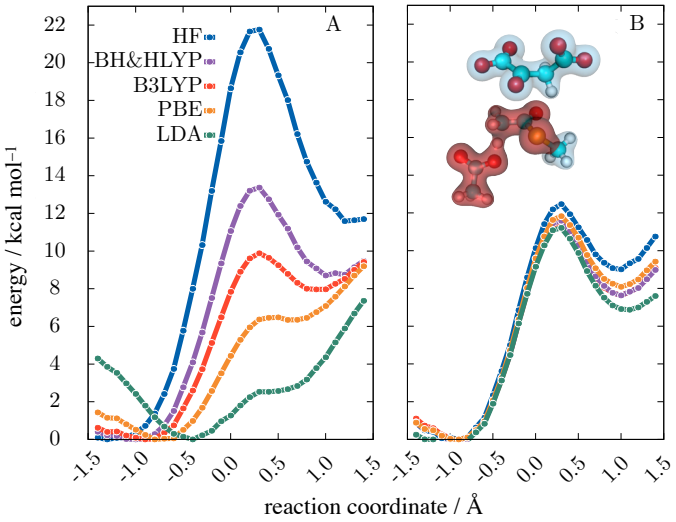

Figure 3.3: (A) Reaction profiles for proton abstraction from acetyl-coenzyme A in the citrate synthase enzyme, computed using QM/MM with standard DFT approaches (A) and projection-based CCSD(T)-in-DFT embedding (B). The dependence of the predicted activation barrier on choice of functional is almost completely eliminated through projection-based embedding. The QM region is shown as an inset, with the red-shaded density indicating the CCSD(T) region and the remaining environment described using DFT. Note that the CCSD(T)-in-B3LYP and CCSD(T)-in-BH&HLYP curves, red and purple, are indistinguishable. Adapted with permission from Ref. 21. Copyright 2016 American Chemical Society.

rearrangement [14], which is robustly corrected using the projection-based WF-in-DFT framework.

Battery Electrolytes

A central challenge in the refinement of lithium-ion batteries is to control cathode-induced oxidative decomposition of electrolyte solvents, such as ethylene carbonate (EC) and dimethyl carbonate (DMC). In recent work, projection-based embedding was used to study the oxidation potentials of neat EC, neat DMC, and 1:1 mixtures of EC and DMC, to overcome qualitative inaccuracies in the electronic densities and ionization energies obtained from conventional KS-DFT methods [16]. The embedding method was implemented as shown in Fig. 3.4, with a CCSD(T) description of the oxidized molecule, a DFT description of the surrounding molecules, and a molecular-mechanics (MM) description of more distant molecules. Configurations were sampled using classical MD trajectories on the MM force field, and approximately 2000 CCSD(T)-in-DFT-in-MM calculations were performed to obtained the thermal ensemble averages for the oxidation potentials.

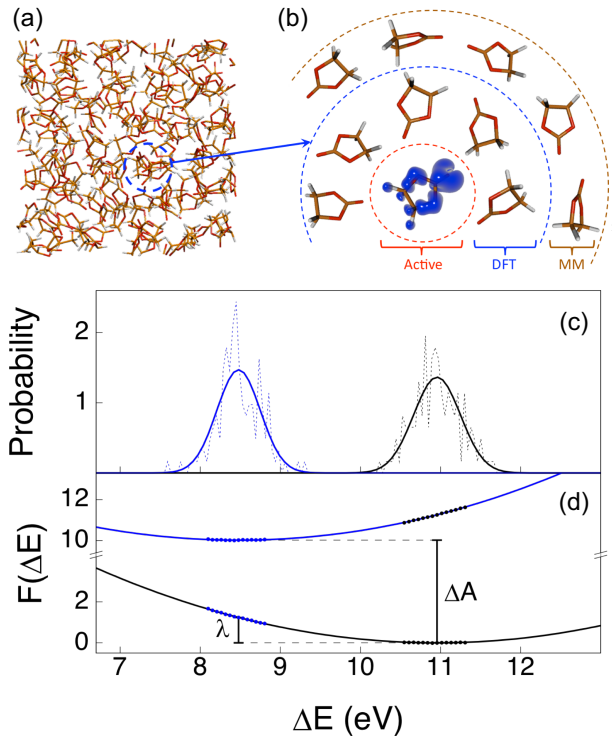

Figure 3.4: Summary of the embedding protocol in Ref. 16. (a) MD simulations are performed to generate the equilibrium ensemble of solvent configurations. (b) Illustration of the CCSD(T)-in-DFT-in-MM embedding protocol. (c) Equilibrium probability distributions, $P_M(\Delta E)$, of the vertical ionization energies, ΔE, of ethylene carbonyl (EC) molecules, calculated using CCSD(T)-in-B3LYP-in-MM embedding. "M" corresponds either to the reduced EC system (R, black) or the oxidized EC^+ system (O, blue). The distributions have similar standard deviations, demonstrating that the linear response approximation is accurate for this system. The best fit Gaussian distributions, $g_M(\Delta E)$, are indicated in solid lines. (d) Diabatic free energy profiles constructed from the equilibrium distributions shown in (c). The solid lines indicate the results from the Gaussian fits, while the points correspond to simulation data. Adapted with permission from Ref. 16. Copyright 2015 American Chemical Society.

It was shown that the ensemble-averaged distributions of vertical IEs are consistent with a linear response interpretation of the statistics of the solvent configurations (Fig. 3.4C,D), enabling determination of both the intrinsic oxidation potential of the solvents and the corresponding solvent reorganization energies. Interestingly, it was found that large contributions to the solvation properties of DMC originate from quadrupolar interactions, resulting in a much larger solvent reorganization energy than that predicted using simple dielectric continuum models. Demonstration that the solvation properties of EC and DMC are governed by fundamentally different intermolecular interactions provides useful insight into lithium-ion batteries electrolyte design, with relevance to electrolyte decomposition processes, solid-

electrolyte interphase formation, and the local solvation environment of lithium cations [16].

3.4 Outlook and Conclusions

Quantum embedding methods have long been recognized as a promising approach to achieving high-accuracy quantum chemical descriptions while preserving a tractable computational cost. As this chapter describes, projection-based WF-in-DFT embedding offers a simple and accurate strategy for reaching this goal that is of practical utility in many chemical applications areas. While many previous studies had recognized that subsystem embedding could be usefully employed via enforcement of subsystem orthogonalization [121, 122], the key advances of the projection-based embedding method [9] were to recognize that the strategy *(i)* could be used to formulate a formally exact method for DFT embedding, *(ii)* could be used to formulate a rigorous and accurate approach to WF-in-DFT embedding, and *(iii)* could be implemented via an extremely simple level-shift projection operator, such that no extra programming is needed to add a new WF method.

Continued technical advances, including atomic-orbital basis set truncation [18] and even-handed subsystem partitioning [128], have improved the efficiency and robustness of the method. The approach is implemented in the widely used Molpro software package [142], allowing straightforward application of projection-based WF-in-DFT embedding for quantum chemical studies [16, 19–21, 24, 25, 132–141].

Looking forward, we anticipate continued refinement of the projection-based embedding methodology, as well as increasingly widespread application in diverse areas of chemistry, biology, and materials science.

Chapter 4

ANALYTICAL GRADIENTS FOR PROJECTION-BASED WAVEFUNCTION-IN-DFT EMBEDDING

Adapted from:

[1]S. J. R. Lee, F. Ding, F. R. Manby, and T. F. Miller III, "Analytical gradients for projection-based wavefunction-in-DFT embedding", en, J. Chem. Phys. **151**, 064112 (2019) 10.1063/1.5109882.

In this chapter we present the derivation, implementation, and numerical demonstration of analytical nuclear gradients for projection-based wavefunction-in-density functional theory (WF-in-DFT) embedding. The gradients are formulated in the Lagrangian framework to enforce orthogonality, localization, and Brillouin constraints on the molecular orbitals. An important aspect of the gradient theory is that WF contributions to the total WF-in-DFT gradient can be simply evaluated using existing WF gradient implementations without modification. Another simplifying aspect is that Kohn-Sham (KS) DFT contributions to the projection-based embedding gradient do not require knowledge of the WF calculation beyond the relaxed WF density. Projection-based WF-in-DFT embedding gradients are thus easily generalized to any combination of WF and KS-DFT methods. We provide numerical demonstration of the method for several applications, including calculation of a minimum energy pathway for a hydride transfer in a cobalt-based molecular catalyst using the nudged-elastic-band method at the CCSD-in-DFT level of theory, which reveals large differences from the transition state geometry predicted using DFT.

4.1 Introduction

This chapter focuses on the development of analytical nuclear gradients for the projection-based embedding method [9, 27]. Projection-based embedding is a DFT-based embedding theory in which subsystem partitioning is performed in terms of localized Kohn-Sham (KS) molecular orbitals (LMOs). The method describes subsystem interactions at the level of KS and allows for the partitioning of the subsystems across covalent and even conjugated bonds, and it enables the use of relatively small subsystem sizes for an embedded WF description. A review of projection-based WF-in-DFT embedding is outlined in Chapter 3.

Projection-based embedding has proven to be a useful tool in a wide range of chemical contexts including transition-metal complexes [19, 20, 25, 128], protein active sites [21, 24], excited states [132, 135, 139] and condensed phase systems [16], among others [133, 134, 136–138, 140, 141]. The development of analytical nuclear gradients for projection-based embedding will expand its applicability to include geometry optimization, transition state searches, and potentially *ab initio* molecular dynamics. Analytical nuclear gradients already exist for a number of other embedding methodologies, including the incremental molecular fragmentation method [143], fragment molecular orbital method [144–146], QM/MM [94, 95, 108], ONIOM [75, 147–149], EMFT [17, 150], frozen density embedding [151–154], and subsystem DFT [155–157]. However, the projection-based approach provides a number of advantages for WF-in-DFT embedding calculations and leads to a distinct analytical gradient theory, which we derive and numerically demonstrate in several applications.

In section 4.2 we outline projection-based WF-in-DFT embedding and in section 4.2 we provide the derivation of its analytical nuclear gradients. Section 4.4 numerically validates the analytical nuclear gradient theory and its implementation in Molpro [78] via comparison with finite difference calculations, as well as presenting results for optimizing geometries in benchmark systems and the calculation of a minimum energy profile for an organometallic reaction using the nudged-elastic-band (NEB) method. We additionally provide the analytical nuclear gradient theory for WF-in-DFT embedding with atomic orbital (AO) truncation [18] in Appendices 4.7 and 4.7.

4.2 Projection-based Embedding Analytical Nuclear Gradients

Projection-based Embedding Energy Theory

Projection-based WF-in-DFT embedding relies on the partitioning the LMOs of a system into two subsystems. Subsystem A contains the LMOs that are treated using the WF method and subsystem B contains the remaining LMOs that are treated using KS. This WF-in-DFT procedure is accomplished by first performing a KS calculation on the full system to obtain a set of KS MOs. The occupied KS MOs are then localized and partitioned into subsystems A and B. Finally, subsystem A is treated using the WF method in the presence of the embedding potential created by the frozen LMOs of subsystem B. Note that the cost of the KS calculation on the full system is typically negligible in comparison to the subsystem WF calculation. This results in our working equation for projection-based WF-in-DFT embedding [9],

$$\begin{aligned} E_{\text{WF-in-DFT}}\left[\tilde{\Psi}^{\text{A}};\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right] = E_{\text{WF}}\left[\tilde{\Psi}^{\text{A}}\right] + \text{tr}\left[(\tilde{\mathbf{d}}^{\text{A}} - \boldsymbol{\gamma}^{\text{A}})\mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right]\right] \\ + E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right] - E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}}\right] + \mu\text{tr}\left[\tilde{\mathbf{d}}^{\text{A}}\mathbf{P}^{\text{B}}\right], \end{aligned} \tag{4.1}$$

where $\tilde{\Psi}^{\text{A}}$ and $E_{\text{WF}}[\tilde{\Psi}^{\text{A}}]$ are the WF and energy of subsystem A, $\tilde{\mathbf{d}}^{\text{A}}$ is the subsystem A one-particle reduced density matrix that corresponds to $\tilde{\Psi}^{\text{A}}$, E_{DFT} is the KS energy, and $\boldsymbol{\gamma}^{\text{A}}$ and $\boldsymbol{\gamma}^{\text{B}}$ are respectively the KS subsystem A and B one-particle densities that equal the full system KS density, $\boldsymbol{\gamma}$, when summed together. Throughout, we shall use a tilde to indicate quantities that have been calculated using the WF method. The embedding potential, $\mathbf{v}_{\text{emb}}$, is defined as

$$\mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right] = \mathbf{g}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right] - \mathbf{g}\left[\boldsymbol{\gamma}^{\text{A}}\right], \tag{4.2}$$

where $\mathbf{g}$ includes all KS two-electron terms,

$$(\mathbf{g}[\boldsymbol{\gamma}])_{\kappa\nu} = \sum_{\lambda\sigma}\gamma_{\lambda\sigma}\left((\kappa\nu|\lambda\sigma) - \frac{1}{2}x_f(\kappa\lambda|\nu\sigma)\right) + (\mathbf{v}_{\text{xc}}[\boldsymbol{\gamma}])_{\kappa\nu}, \tag{4.3}$$

and where κ, ν, λ and σ label atomic orbital basis functions, $(\kappa\nu|\lambda\sigma)$ are the two-electron repulsion integrals, x_f is the fraction of exact exchange and $\mathbf{v}_{\text{xc}}$ is the exchange-correlation (XC) potential matrix. The level-shift operator, $\mu\mathbf{P}^{\text{B}}$, is given by

$$\mu\mathbf{P}^{\text{B}} = \mu\mathbf{S}\boldsymbol{\gamma}^{\text{B}}\mathbf{S}, \tag{4.4}$$

where $\mathbf{S}$ is the overlap matrix. In the limit of $\mu \to \infty$, the LMOs that make up subsystems A and B are enforced to be exactly orthogonal, eliminating the

non-additive kinetic energy present in other embedding frameworks [99, 113]. In practice, finite values of μ in the range of 10^4 hartree to 10^7 hartree are found to provide accurate results regardless of chemical system [9]. If greater accuracy is needed, a perturbative correction outlined in Ref. 9 can be added to the WF-in-DFT energy expression to account for the finiteness of μ, but in practice, this correction is found to contribute negligibly to the total energy and is thus neglected here.

Projection-based embedding can also be used for DFT-in-DFT embedding via a simplified version of Eq. 4.1. The working equation for projection-based DFT-in-DFT embedding is [9]

$$\begin{aligned} E_{\text{DFT-in-DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}};\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right] = E_{\text{DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}\right] + \text{tr}\left[\left(\tilde{\boldsymbol{\gamma}}^{\text{A}} - \boldsymbol{\gamma}^{\text{A}}\right)\mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}},\boldsymbol{\gamma}^{\text{B}}\right]\right] \\ + E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right] - E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}}\right] + \mu\text{tr}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}\mathbf{P}^{\text{B}}\right]. \end{aligned} \tag{4.5}$$

The only differences between WF-in-DFT and DFT-in-DFT embedding is that the first term on the RHS of Eq. 4.1 is replaced with the KS energy on subsystem A, $E_{\text{DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\text{A}}\right]$, and in the second and last terms $\tilde{\mathbf{d}}^{\text{A}}$ is reduced to the subsystem A KS density matrix, $\tilde{\boldsymbol{\gamma}}^{\text{A}}$.

Projection-based WF-in-DFT Embedding Gradient Theory

Since projection-based embedding is a non-variational theory, its analytical gradient is conveniently derived using a Lagrangian approach. We first construct a Lagrangian based on the projection-based WF-in-DFT energy. We then minimize the Lagrangian with respect to the variational parameters in the embedding energy, which include the subsystem A WF and the LMO coefficients. Then we show how to solve for each of the Lagrange multipliers and provide the working equation for the gradient of the total energy.

For consistency in notation, the MO coefficient matrix $\mathbf{C}$ refers to the entire set of KS MOs (occupied and virtual). The submatrix of $\mathbf{C}$ that refers to the (occupied) LMOs is denoted as $\mathbf{L}$ with column indices i, j, k, l. The submatrix of $\mathbf{C}$ that refers to the canonical virtual space is denoted as $\mathbf{C}_{\text{v}}$ with column indices a, b, c, d. The indices m, n, p, q are used to index generic molecular orbitals.

Total Energy Lagrangian

We now derive the total energy Lagrangian for projection-based WF-in-DFT embedding. Where appropriate we will provide WF method specific examples (e.g.

MP2) of general terms outlined in the equations. The WF-in-DFT Lagrangian is

$$\begin{aligned}\mathcal{L}\left[\mathbf{C}, \tilde{\Psi}^{\mathrm{A}}, \mathbf{\Lambda}, \mathbf{x}, \mathbf{z}^{\mathrm{loc}}, \mathbf{z}\right] &= E_{\text{WF-in-DFT}}\left[\tilde{\Psi}^{\mathrm{A}}; \boldsymbol{\gamma}^{\mathrm{A}}, \boldsymbol{\gamma}^{\mathrm{B}}\right] + \sum_{s} \Lambda_{s}^{\mathrm{WF,A}} c_{s} \\ &+ \sum_{pq} x_{pq}\left(\mathbf{C}^{\dagger}\mathbf{S}\mathbf{C} - \mathbf{1}\right)_{pq} + \sum_{i>j} z_{ij}^{\mathrm{loc}} r_{ij} + \sum_{ai} z_{ai}\left(\mathbf{F}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right]\right)_{ai}.\end{aligned} \tag{4.6}$$

The first term on the right hand side (RHS) of Eq. 4.6 is the projection-based WF-in-DFT embedding energy described by Eq. 4.1. The second term on the RHS of Eq. 4.6 contains any constraints, c_s, and the corresponding Lagrange multipliers, $\Lambda_s^{\mathrm{WF,A}}$, that arise from ensuring that the Lagrangian is variational with respect to parameters in the WF method. The third term on the RHS constrains the KS MOs, $\mathbf{C}$, to be orthonormal, which accounts for the basis set being atom centered; this term is commonly referred to as the Pulay force [158] and arises from the atomic orbital basis set being atom centered. The localization conditions, $r_{ij} = 0$, take into account how the KS MOs are localized before being selected for subsystems A and B. This is important because the LMOs will have a different dependence on nuclear perturbation than canonical MOs. In this work, we use Pipek-Mezey localization [129] to obtain LMOs. Generalization to other localization methods (e.g. Boys [159] and intrinsic bond orbitals [130]) is straightforward. The localization conditions for Pipek-Mezey are

$$r_{ij} = \sum_{C}\left(S_{ii}^{C} - S_{jj}^{C}\right)S_{ij}^{C} = 0 \quad \text{for all } i > j, \tag{4.7}$$

where C corresponds to an atom in the molecule. The matrices S^C are defined as

$$S_{kl}^{C} = \sum_{\alpha \in C}\sum_{\beta}\left(L_{\alpha,k} S_{\alpha\beta} L_{\beta,l} + L_{\alpha,l} S_{\alpha\beta} L_{\beta,k}\right), \tag{4.8}$$

where the summation over α is restricted to basis functions at atom C. The Brillouin conditions, $\left(\mathbf{F}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right]\right)_{ai} = 0$, reflect how the KS MOs are optimized before being used to construct subsystems A and B. The Brillouin conditions are only needed because subsystem B is frozen at the KS level of theory. However, due to the non-additivity of the XC potential, the Lagrange multipliers, $\mathbf{z}$, span the full virtual-occupied space.

The type and number of constraints applied to the WF method depend on the chosen method. For example, if the WF method is MP2 then the constraints are

$$\sum_{s} \Lambda_{s}^{\mathrm{MP2,A}} c_{s} = \sum_{pq} \tilde{x}_{pq}\left(\tilde{\mathbf{C}}^{\mathrm{A}\dagger}\mathbf{S}\tilde{\mathbf{C}}^{\mathrm{A}} - \mathbf{1}\right)_{pq} + \sum_{ai} \tilde{z}_{ai}\left(\mathbf{F}^{\mathrm{A}}\right)_{ai}\Big|_{i\in\mathrm{A}}, \tag{4.9}$$

where the first term on the RHS of Eq. 4.9 constrains the Hartree-Fock MOs, $\tilde{\mathbf{C}}^{\mathrm{A}}$, to be orthonormal, the condition $i \in \mathrm{A}$ restricts the sum to occupied MOs in subsystem A, and the second term on the RHS are the Brillouin conditions using the embedded Fock matrix, $\mathbf{F}^{\mathrm{A}}$. The embedded Fock matrix is defined as [9]

$$\mathbf{F}^{\mathrm{A}} = \mathbf{h} + \mathbf{g}\left[\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}\right] + \mathbf{v}_{\mathrm{emb}}\left[\boldsymbol{\gamma}^{\mathrm{A}}, \boldsymbol{\gamma}^{\mathrm{B}}\right] + \mu\mathbf{P}^{\mathrm{B}}, \tag{4.10}$$

where $\mathbf{h}$ is the standard one-electron Hamiltonian, $\mathbf{g}$ includes all of the usual HF two-electron terms and $\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}$ is the subsystem A HF one-particle density. These constraints also arise in the derivation of the MP2 analytical nuclear gradient [160].

For the projection-based WF-in-DFT energy to equal the Lagrangian, the Lagrangian must be minimized with respect to all of its parameters, including $\tilde{\Psi}^{\mathrm{A}}$, $\mathbf{C}$, and all of the Lagrange multipliers.

Minimizing the Lagrangian with respect to the variational parameters of the WF method

Upon minimizing the WF-in-DFT Lagrangian with respect to $\tilde{\Psi}^{\mathrm{A}}$, only terms associated with the first two terms on the RHS of Eq. 4.6 survive, all of which are familiar from the WF Lagrangian for the corresponding WF gradient theories.

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \tilde{\Psi}^{\mathrm{A}}} = \frac{\partial E_{\mathrm{WF}}\left[\tilde{\Psi}^{\mathrm{A}}\right]}{\partial \tilde{\Psi}^{\mathrm{A}}} + \mathrm{tr}\left[\frac{\partial \tilde{\mathbf{d}}^{\mathrm{A}}}{\partial \tilde{\Psi}^{\mathrm{A}}}\mathbf{v}_{\mathrm{emb}}\left[\boldsymbol{\gamma}^{\mathrm{A}}, \boldsymbol{\gamma}^{\mathrm{B}}\right]\right] \\ + \mu\mathrm{tr}\left[\frac{\partial \tilde{\mathbf{d}}^{\mathrm{A}}}{\partial \tilde{\Psi}^{\mathrm{A}}}\mathbf{P}^{\mathrm{B}}\right] + \frac{\partial}{\partial \tilde{\Psi}^{\mathrm{A}}}\sum_{s}\Lambda_{s}^{\mathrm{WF,A}}c_{s} = 0\end{aligned} \tag{4.11}$$

Since the embedding potential is independent of $\tilde{\Psi}^{\mathrm{A}}$, the Z-vector coupled perturbed Hartree–Fock (Z-CPHF) equations of any post-HF method are only impacted through the eigenvalues of the subsystem A HF WF. Therefore, the solutions for the WF Lagrange multipliers (e.g. $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{z}}$ for MP2 in Eq. 4.9) are obtained using the standard implementation of the WF gradient no matter what KS method is selected to describe subsystem B. However, if an alternative embedding potential is used that depends on the subsystem A WF, such as the Huzinaga constraint (i.e. Ref. 123), then the formulation of the WF gradient is changed; the Z-CPHF equations for a general WF method would need to be modified to include the contributions from the derivative of the embedding potential with respect to the subsystem A WF, $\tilde{\Psi}^{\mathrm{A}}$.

Minimizing the Lagrangian with respect to the MO coefficients

The remaining Lagrange multipliers, $\mathbf{z}^{\mathrm{loc}}$, $\mathbf{x}$, and $\mathbf{z}$ in Eq. 4.6, are determined by minimizing the WF-in-DFT Lagrangian with respect to the variational parameters

of the KS method, namely the MO coefficients, $\mathbf{C}$. Differentiation of the Lagrangian with respect to these parameters yields

$$\sum_{\mu} C_{\mu,p} \frac{\partial \mathcal{L}}{\partial C_{\mu,q}} = E_{pq} + \left(\mathbf{a}\left[\mathbf{z}^{\mathrm{loc}}\right]\right)_{pi} + (\mathbf{D}[\mathbf{z}])_{pq} + 2x_{pq} = 0, \tag{4.12}$$

where

$$E_{pq} = \sum_{\mu} C_{\mu,p} \left(\frac{\partial E_{\text{WF-in-DFT}}\left[\tilde{\Psi}^{\mathrm{A}}; \boldsymbol{\gamma}^{\mathrm{A}}, \boldsymbol{\gamma}^{\mathrm{B}}\right]}{\partial C_{\mu,q}} + \frac{\partial}{\partial C_{\mu,q}} \sum_{s} \Lambda_s^{\mathrm{WF,A}} c_s \right), \tag{4.13}$$

$$\begin{aligned} \left(\mathbf{a}\left[\mathbf{z}^{\mathrm{loc}}\right]\right)_{pi} &= \sum_{\mu} C_{\mu,p} \left(\sum_{k>l} z_{kl}^{\mathrm{loc}} \frac{\partial r_{kl}}{\partial C_{\mu,q}} \right) \\ &= \sum_{k>l} \mathcal{B}_{pi,kl} z_{kl}^{\mathrm{loc}}, \end{aligned} \tag{4.14}$$

$$\begin{aligned} (\mathbf{D}[\mathbf{z}])_{pq} &= \sum_{\mu} C_{\mu,p} \left(\sum_{ak} z_{ak} \frac{\partial \left(\mathbf{F}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right]\right)_{ak}}{\partial C_{\mu,q}} \right) \\ &= \sum_{ak} \mathcal{D}_{pq,ak} z_{ak} \\ &= \left(\mathbf{F}[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}]\mathbf{z}\right)_{pq}\Big|_{q\in\mathrm{occ}} + \left(\mathbf{F}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right]\mathbf{z}^{\dagger}\right)_{pq}\Big|_{q\in\mathrm{vir}} + 2(\mathbf{V}[\bar{\mathbf{z}}])_{pq}\Big|_{q\in\mathrm{occ}}, \end{aligned} \tag{4.15}$$

and

$$2x_{pq} = \sum_{\mu} C_{\mu,p} \left(\sum_{mn} x_{mn} \frac{\partial S_{mn}}{\partial C_{\mu,q}} \right). \tag{4.16}$$

The 4-dimensional tensors, $\mathcal{B}$ and $\mathcal{D}$, are expanded in Appendices 4.7 and 4.7, respectively, $\bar{\mathbf{z}}$ corresponds to $\mathbf{z} + \mathbf{z}^{\dagger}$, and $\mathbf{V}[\bar{\mathbf{z}}]$ includes all two-electron terms of the generalized Fock matrix and is shown explicitly in Appendix 4.7. Since the embedded Fock matrix, $\mathbf{F}^{\mathrm{A}}$, contains the embedding potential, $\mathbf{v}_{\mathrm{emb}}$, its derivative with respect to the MO coefficients, $\mathbf{C}$, is nonzero resulting in the WF relaxed density being needed to construct $\mathbf{E}$ in Eq. 4.13, which is explicitly shown in Appendix 4.7. Therefore, the subsystem A WF gradient only affects the embedding contributions to the gradient through the WF relaxed density.

We now show that solving for the Lagrange multipliers leads to familiar coupled perturbed equations. Combining the stationary conditions described by Eq. 4.12 with the auxiliary conditions $\boldsymbol{x} = \boldsymbol{x}^{\dagger}$ yields the linear Z-vector equations [160]

$$\left(1 - \mathcal{P}_{pq}\right)\left(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{a}\left[\mathbf{z}^{\mathrm{loc}}\right]\right)_{pq} = 0, \tag{4.17}$$

where $\mathcal{P}_{pq}$ permutes the indices p and q, which is used to solve for $\mathbf{z}$ and $\mathbf{z}^{\text{loc}}$. The matrix $\boldsymbol{x}$ is then obtained as

$$x_{pq} = -\frac{1}{4}\left(1 + \mathcal{P}_{pq}\right)\Big(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{a}\big[\mathbf{z}^{\text{loc}}\big]\Big)_{pq}. \tag{4.18}$$

The Lagrange multipliers $\mathbf{z}^{\text{loc}}$ pertain to the occupied-occupied MO space; considering only the occupied-occupied part of Eq. 4.17 yields

$$\left(1 - \mathcal{P}_{ij}\right)\Big(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{a}\big[\mathbf{z}^{\text{loc}}\big]\Big)_{ij} = 0. \tag{4.19}$$

Using the Brillouin conditions and the knowledge that $z_{ab} = z_{ij} = z_{ia} = 0$, Eq. 4.19 can be further simplified by showing that

$$\left(1 - \mathcal{P}_{ij}\right)(\mathbf{D}[\mathbf{z}])_{ij} = 0. \tag{4.20}$$

The solutions, $\mathbf{z}^{\text{loc}}$, are thus independent of $\mathbf{z}$, such that Eq. 4.19 reduces to

$$E_{ij} - E_{ji} + \sum_{k>l}\left(\mathcal{B}_{ij,kl} - \mathcal{B}_{ji,kl}\right)z^{\text{loc}}_{kl} = 0. \tag{4.21}$$

These are the Z-vector coupled perturbed localization (Z-CPL) equations, which are used to solve for $\mathbf{z}^{\text{loc}}$. Subsequently, $\mathbf{a}\big[\mathbf{z}^{\text{loc}}\big]$ can be computed according to Eq. 4.14.

The Lagrange multipliers $\mathbf{z}$ pertain to the virtual-occupied MO space; considering only the virtual-occupied part of Eq. 4.17 yields

$$\left(1 - \mathcal{P}_{ai}\right)\Big(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{a}\big[\mathbf{z}^{\text{loc}}\big]\Big)_{ai} = 0, \tag{4.22}$$

which further simplifies to

$$\Big(\mathbf{E} + \mathbf{a}\big[\mathbf{z}^{\text{loc}}\big] + \mathbf{F}\big[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\big]\mathbf{z} - \mathbf{z}\mathbf{F}\big[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\big] + 2\mathbf{V}[\bar{\mathbf{z}}]\Big)_{ai} = 0. \tag{4.23}$$

These are the Z-vector coupled perturbed Kohn-Sham (Z-CPKS) equations. Having solved the Z-CPL and Z-CPKS equations, the remaining Lagrangian multipliers associated with the orthogonality constraints, $\boldsymbol{x}$, can be obtained from Eq. 4.18.

Gradient of the Total Energy

Once the Lagrangian is minimized with respect to all variational parameters, the gradient of the total energy takes the form

$$\begin{aligned}\frac{\mathrm{d}E_{\text{WF-in-DFT}}}{\mathrm{d}q} = \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}q} &= \frac{\partial\mathcal{L}}{\partial q} + \frac{\partial\mathcal{L}}{\partial\tilde{\Psi}^{\text{A}}}\frac{\tilde{\Psi}^{\text{A}}}{\partial q} + \frac{\partial\mathcal{L}}{\partial\mathbf{C}}\frac{\partial\mathbf{C}}{\partial q} \\ &= \frac{\partial\mathcal{L}}{\partial q}.\end{aligned} \tag{4.24}$$

Since the Lagrangian is minimized with respect to the subsystem A WF and the KS LMO coefficients, calculation of the WF and KS LMO responses to nuclear perturbation, $\partial\tilde{\Psi}^{\mathrm{A}}/\partial q$ and $\partial\mathbf{C}/\partial q$ respectively, are avoided. This yields the following expression for the gradient,

$$\begin{aligned} E^{(q)}_{\text{WF-in-DFT}} &= E^{(q)}_{\mathrm{WF}}\big[\tilde{\Psi}^{\mathrm{A}}\big] + \sum_{\lambda\nu}\tilde{\mathbf{d}}^{\mathrm{A}}_{\lambda\nu}(\mathbf{v}^{(q)}_{\mathrm{emb}})_{\lambda\nu} \\ &+ \mu\sum_{\lambda\nu}\tilde{\mathbf{d}}^{\mathrm{A}}_{\lambda\nu}(\mathbf{P}^{\mathrm{B},(q)})_{\lambda\nu} + \sum_{s}\Lambda^{\mathrm{WF,A}}_{s}c^{(q)}_{s} \\ &- E^{(q)}_{\mathrm{DFT}}[\boldsymbol{\gamma}^{\mathrm{A}}] + E^{(q)}_{\mathrm{DFT}}[\boldsymbol{\gamma}^{\mathrm{A}}+\boldsymbol{\gamma}^{\mathrm{B}}] - \sum_{\lambda\nu}\boldsymbol{\gamma}^{\mathrm{A}}_{\lambda\nu}(\mathbf{v}^{(q)}_{\mathrm{emb}})_{\lambda\nu} \\ &+ \sum_{ij} z^{\mathrm{loc}}_{ij}r^{(q)}_{ij} + \sum_{ai} z_{ai}\Big(\mathbf{F}\big[\boldsymbol{\gamma}^{\mathrm{A}}+\boldsymbol{\gamma}^{\mathrm{B}}\big]\Big)^{(q)}_{ai} + \sum_{mn} x_{mn}S^{(q)}_{mn}, \end{aligned} \tag{4.25}$$

where the superscript (q) denotes the explicit derivative of the quantity with respect to a nuclear coordinate. Eq. 4.25 can be further simplified by folding $\sum_s \Lambda^{\mathrm{WF,A}}_s c^{(q)}_s$ into the first three terms on the RHS of Eq. 4.25, yielding

$$\begin{aligned} E^{(q)}_{\text{WF-in-DFT}} &= E^{q}_{\mathrm{WF}}\big[\tilde{\Psi}^{\mathrm{A}}\big] + \sum_{\lambda\nu}(\tilde{\mathbf{d}}^{\mathrm{A}}_{\mathrm{rel}})_{\lambda\nu}(\mathbf{v}^{(q)}_{\mathrm{emb}})_{\lambda\nu} + \mu\sum_{\lambda\nu}(\tilde{\mathbf{d}}^{\mathrm{A}}_{\mathrm{rel}})_{\lambda\nu}(\mathbf{P}^{\mathrm{B},(q)})_{\lambda\nu} \\ &- E^{(q)}_{\mathrm{DFT}}[\boldsymbol{\gamma}^{\mathrm{A}}] + E^{(q)}_{\mathrm{DFT}}[\boldsymbol{\gamma}^{\mathrm{A}}+\boldsymbol{\gamma}^{\mathrm{B}}] - \sum_{\lambda\nu}\boldsymbol{\gamma}^{\mathrm{A}}_{\lambda\nu}(\mathbf{v}^{(q)}_{\mathrm{emb}})_{\lambda\nu} \\ &+ \sum_{ij} z^{\mathrm{loc}}_{ij}r^{(q)}_{ij} + \sum_{ai} z_{ai}\Big(\mathbf{F}\big[\boldsymbol{\gamma}^{\mathrm{A}}+\boldsymbol{\gamma}^{\mathrm{B}}\big]\Big)^{(q)}_{ai} + \sum_{mn} x_{mn}S^{(q)}_{mn}. \end{aligned} \tag{4.26}$$

Here, $E^{q}_{\mathrm{WF}}[\tilde{\Psi}^{\mathrm{A}}]$ denotes the total derivative of the subsystem A WF energy with respect to nuclear coordinate, which can be directly calculated using existing WF gradient implementations, and $\tilde{\mathbf{d}}^{\mathrm{A}}_{\mathrm{rel}}$ is the WF-relaxed density for subsystem A. For example, the MP2-relaxed density is

$$\tilde{\mathbf{d}}^{\mathrm{A}}_{\mathrm{rel}} = \tilde{\mathbf{d}}^{\mathrm{A}} + \tilde{\mathbf{C}}^{\mathrm{A}}\tilde{\mathbf{z}}\tilde{\mathbf{C}}^{\mathrm{A},\dagger} = \tilde{\boldsymbol{\gamma}}^{\mathrm{A}} + \mathbf{d}^{(2)} + \tilde{\mathbf{C}}^{\mathrm{A}}\tilde{\mathbf{z}}\tilde{\mathbf{C}}^{\mathrm{A},\dagger}, \tag{4.27}$$

which contains the subsystem A Hartree-Fock density, $\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}$, the MP2 density matrix, $\mathbf{d}^{(2)}$, and the solutions of the subsystem A Brillouin conditions, $\tilde{\mathbf{C}}^{\mathrm{A}}\tilde{\mathbf{z}}\tilde{\mathbf{C}}^{\mathrm{A},\dagger}$. Eq. 4.26 can be expressed in terms of the WF gradient on subsystem A and the derivative AO integrals, yielding our final expression for the projection-based WF-in-DFT

analytical gradient,

$$\begin{aligned} E^{(q)}_{\text{WF-in-DFT}} &= E^{q}_{\text{WF}}\left[\tilde{\Psi}^{\text{A}}\right] \\ &+ \text{tr}\left[\mathbf{d}_{\text{a}}\mathbf{h}^{(q)}\right] + \text{tr}\left[\mathbf{X}\mathbf{S}^{(q)}\right] + \frac{1}{2}\sum_{\mu\nu\lambda\sigma} D_{\mu\nu\lambda\sigma}(\mu\nu|\lambda\sigma)^{(q)} \\ &+ E^{(q)}_{\text{xc}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right] - E^{(q)}_{\text{xc}}\left[\boldsymbol{\gamma}^{\text{A}}\right] \\ &+ \text{tr}\left[\left(\tilde{\mathbf{d}}^{\text{A}}_{\text{rel}} - \boldsymbol{\gamma}^{\text{A}}\right)\left(\mathbf{v}^{(q)}_{\text{xc}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right] - \mathbf{v}^{(q)}_{\text{xc}}\left[\boldsymbol{\gamma}^{\text{A}}\right]\right)\right]. \end{aligned} \tag{4.28}$$

The effective one-particle density $\mathbf{d}_{\text{a}}$ and effective two-particle density $\mathbf{D}$ are defined

$$\mathbf{d}_{\text{a}} = \boldsymbol{\gamma}^{\text{B}} + \mathbf{C}\mathbf{z}\mathbf{C}^{\dagger}, \tag{4.29}$$

and

$$\begin{aligned} D_{\mu\nu\lambda\sigma} &= \left(\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right)_{\mu\nu}(\mathbf{d}_{\text{b}})_{\lambda\sigma} - \gamma^{\text{A}}_{\mu\nu}(\mathbf{d}_{\text{c}})_{\lambda\sigma} \\ &\quad - \frac{1}{2}x_f\left(\left(\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right)_{\mu\lambda}(\mathbf{d}_{\text{b}})_{\nu\sigma} - \gamma^{\text{A}}_{\mu\lambda}(\mathbf{d}_{\text{c}})_{\nu\sigma}\right). \end{aligned} \tag{4.30}$$

The effective one-particle densities $\mathbf{d}_{\text{b}}$ and $\mathbf{d}_{\text{c}}$ are defined

$$\mathbf{d}_{\text{b}} = \boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}} + 2\mathbf{C}\mathbf{z}\mathbf{C}^{\dagger} + 2\tilde{\mathbf{d}}^{\text{A}}_{\text{rel}} - 2\boldsymbol{\gamma}^{\text{A}}, \tag{4.31}$$

and

$$\mathbf{d}_{\text{c}} = -\boldsymbol{\gamma}^{\text{A}} + 2\tilde{\mathbf{d}}^{\text{A}}_{\text{rel}}. \tag{4.32}$$

The matrix $\mathbf{X}$ is defined

$$\begin{aligned} \mathbf{X} &= \mathbf{C}\mathbf{x}\mathbf{C}^{\dagger} + \sum_{i>j}\frac{\partial r_{ij}}{\partial S_{\mu\nu}} z^{\text{loc}}_{ij} \\ &= \mathbf{X}^{\text{loc}} - \frac{1}{2}\mathbf{L}\left(\mathbf{E} + 2\mathbf{V}[\bar{\mathbf{z}}]\right)\mathbf{L}^{\dagger} - \frac{1}{2}\left(\mathbf{C}_{\text{v}}(\mathbf{z}\mathbf{F})\mathbf{L}^{\dagger} + \left(\mathbf{C}_{\text{v}}(\mathbf{z}\mathbf{F})\mathbf{L}^{\dagger}\right)^{\dagger}\right) \\ &\quad + \mu\left(\tilde{\mathbf{d}}^{\text{A}}_{\text{rel}}\mathbf{S}\boldsymbol{\gamma}^{\text{B}} + \boldsymbol{\gamma}^{\text{B}}\mathbf{S}\tilde{\mathbf{d}}^{\text{A}}_{\text{rel}}\right), \end{aligned} \tag{4.33}$$

where

$$\left(\mathbf{X}^{\text{loc}}\right)_{\mu\nu} = -\frac{1}{2}\left(\mathbf{L}\mathbf{a}\left[\mathbf{z}^{\text{loc}}\right]\mathbf{L}^{\dagger}\right)_{\mu\nu} + \sum_{i>j}\frac{\partial r_{ij}}{\partial S_{\mu\nu}} z^{\text{loc}}_{ij}. \tag{4.34}$$

The second term on the RHS of Eq. 4.34 is expanded in Appendix 4.7.

The analytical nuclear gradient expression for projection-based DFT-in-DFT closely follows that for WF-in-DFT, with regard to evaluation of both the Lagrange multipliers (Eq. 4.12) and the final gradient (Eq. 4.28). To obtain the corresponding DFT-in-DFT expressions, $\tilde{\mathbf{d}}^{\text{A}}_{\text{rel}}$ becomes the subsystem A KS density

$$\tilde{\mathbf{d}}^{\text{A}}_{\text{rel}} = \tilde{\boldsymbol{\gamma}}^{\text{A}}, \tag{4.35}$$

which affects the evaluation of $\mathbf{E}$ in Eq. 4.12 (expanded in Eq. 4.39) and the evaluation of the final gradient expression, Eq. 4.28. Additionally, the first term on the RHS of the final gradient expression, Eq. 4.28, is replaced with the subsystem A KS gradient, $E^{q}_{\mathrm{DFT}}\left[\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}\right]$.

4.3 Computational Details

The implementation of projection-based WF-in-DFT embedding gradients is available in the 2019 general release of Molpro [78]. In all embedding calculations reported here, unless otherwise specified, the Pipek-Mezey localization method [129] is used with the core and occupied MOs localized together. The subsystem A region is chosen by including any LMOs with a net Mulliken population larger than 0.4 on the atoms associated with subsystem A, although more sophisticated partitioning algorithms have been introduced [128]. A level-shift parameter of $\mu = 10^6$ hartree is used for all embedding calculations. The perturbative correction to using a finite value of μ in Eq. 4.5 is less than 20 microhartrees for the applications presented here and thus not included (accomplished by specifying the option `HF_COR` = 0). Throughout this work, all embedding calculations are described using the nomenclature "(WF method)-in-DFT/basis," where the WF method describes subsystem A and the KS method describes subsystem B. For some embedding calculations a mixed-basis set is used and is denoted by "(WF method)-in-DFT/large-basis:small-basis," where the large basis is used to describe subsystem A and the small basis is used to describe subsystem B.

All SCF calculations employ a tighter threshold than default for MO convergence by specifying the option `ORBITAL` = 1×10^{-7} a.u. in Molpro. All KS calculations used in projection-based embedding are done without density fitting, employing the LDA [161, 162], PBE [79], PBE0 [163], and LDAX functionals with the def2-TZVPP, def2-SVP, def2-ASVP [164, 165], cc-pVDZ [80], and 6-31G [81] basis sets. Note that the def2-ASVP basis set used in Molpro is constructed by adding one set of even tempered diffuse functions to the def2-SVP basis set. The LDAX functional is constructed by including 50% exact exchange and reducing the weight of the DIRAC functional to 50% in the LDA functional. For the calculations in sections 4.4 and 4.4, the XC functional is evaluated on a fixed-pruned grid with index 7 (Ref. 166). For the optimized geometries shown in section 4.4, and the malondialdehyde calculations in section 4.4 the XC functional is evaluated on an adaptively generated quadrature grid that reproduces the energy of the Slater-Dirac functional to a specified threshold accuracy of $10^{-10}E_h$. All WF calculations are

performed with the frozen-core approximation, without density fitting, employing the MP2 [167], CCSD [168, 169], and CCSD(T) [170] correlation treatments with the def2-TZVPP, def2-SVP, cc-pVDZ and 6-31G basis sets. Even though the density fitting approximation is not used for the WF methods in this study, density fitted gradients are available for the aforementioned WF methods [171, 172]. The default values for integral screening were used in Molpro. For all Z-CPKS calculations an iterative subspace solver employing the Davidson algorithm [173, 174] is used with a convergence threshold of 1×10^{-6} a.u. For all Z-CPHF calculations needed for the subsystem A WF gradient an iterative solver with a convergence threshold of 1×10^{-7} a.u. is used. Grid weight derivatives are included for all gradient calculations involving the XC functional and potential.

For all geometry optimizations the number of LMOs in subsystem A is kept unchanged throughout the optimization. A natural way of enforcing this in future work is to employ even-handed partitioning [128], although this was not needed in the examples studied here; the default procedure based on net Mulliken population sufficed to keep subsystem A unchanged. All geometries are optimized using the translation-rotation-internal coordinate system devised by Wang and Song [175], which is available in the GeomeTRIC package [176]. Convergence parameters for the geometry optimizations follow the default parameters used by Molpro, namely that the maximum gradient value becomes less than 3×10^{-4} hartree/bohr and the energy change between adjacent steps becomes less than 1×10^{-6} hartree or the maximum component of the step displacement becomes less than 3×10^{-4} bohr. The maximum gradient value is evaluated in the Cartesian basis. All geometries are provided in the supporting information.

Nudged elastic band (NEB) [177] calculations are run using the implementation of the method in the atomic simulation environment (ASE) package [178]. All NEB calculations use Molpro forces which are provided through a Molpro calculator interface within the ASE package.

The intramolecular proton transfer of malondialdehyde is modeled with an NEB consisting of 15 images connected by springs with spring constants of 0.1 eV/Å^2. The CCSD/def2-ASVP, CCSD-in-LDA/def2-ASVP and LDA/def2-ASVP optimized NEBs used the image dependent pair potential (IDPP) method [179] as the initial guess for the band with reactant and product geometries previously optimized at the corresponding level of theory. All NEB calculations for malondialdehyde are converged with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update of the Hessian and by

enforcing that the maximum gradient value is less than 0.01 eV/Å^2.

The intramolecular proton transfer of the organometallic cobalt complex is modeled with an NEB consisted of images connected by springs with spring constants of 9 eV/Å^2. The PBE0/cc-pVDZ climbing image NEB [180], consisting of 26 images, used the IDPP method as the initial guess for the band with the reactant geometry previously optimized. The CCSD-in-PBE0/cc-pVDZ NEB consisting of 23 images, used its optimized reactant and the climbing image NEB converged at the PBE0/cc-pVDZ level of theory as its initial guess. Since the product is spatially far away from the reactant, an intermediate geometry between the transition state and the product is used as the endpoint of the NEB. This intermediate geometry is determined by initially converging a NEB with an extra image such that the maximum force dropped below 0.3 eV/Å^2. Then the second to last image is used as the new endpoint and a new NEB is converged. The PBE0/cc-pVDZ climbing image NEB is optimized using the FIRE [181] algorithm using a convergence criteria of 0.05 eV/Å^2 for the maximum gradient value. The projection-based WF-in-DFT embedding NEB is optimized at the CCSD-in-PBE0/cc-pVDZ level of theory using the FIRE algorithm with a convergence criteria of 0.25 eV/Å^2 for the maximum gradient value.

The CCSD-in-PBE0/cc-pVDZ calculations used for the NEB optimization are performed by specifying 51 occupied MOs to be in subsystem A using the `N_ORBITALS` option and by using AO truncation with a threshold of 1×10^{-3} a.u. An even-handed selection of AOs were used along the NEB by creating a union of the AOs that were selected for each image by the truncation procedure to ensure that the NEB traversed a smooth potential energy surface. The final, reported energies of the WF-in-DFT NEB are performed using the PNO-LCCSD/cc-pVDZ [182] and PNO-LCCSD-in-PBE0/cc-pVDZ levels of theory. Both the PNO-LCCSD and PNO-LCCSD-in-PBE0 calculations are performed with density fitting using the cc-pVTZ/JKFIT [183] (the def2-TZVPP/JKFIT [184] basis set was used for cobalt since the cc-pVTZ/JKFIT basis set was not available) and the cc-pVTZ/MP2FIT [185] density fitting basis sets. Tighter domain approximations were employed for all PNO-LCCSD calculations by specifying the `DOMOPT=TIGHT` option. Additionally, the Boughton-Pulay completeness criterion was used for the selection of the primary projected atomic orbitals domain by specifying the option `THRBP=1` and the Pipek-Mezey localization method was used. For the PNO-LCCSD-in-PBE0 calculations, AO truncation is not used, the core and valence DFT molecular orbitals are localized separately using the Pipek-Mezey localization method, and the subsystem A orbitals are selected using

the default procedure based on the Mulliken population threshold.

All calculations using AO truncation [18] ensure that at least one AO is kept per atom (specified by option `AO_PER_ATOM`) to make evaluation of the integral derivative contributions from the one electron Hamiltonian simpler within Molpro. This adds a negligible amount of AO functions than would have been selected using only the density threshold parameter [18] for the systems studied in this chapter. In all embedding geometry optimizations that employ AO truncation, the number of truncated AOs is fixed using the `STOREAO` option to ensure smoothness of the potential energy function. Upon convergence, the truncated AO list is reevaluated using the same density threshold parameter; if the number of kept AOs remains a subset of the original list of truncated AOs then the optimization is converged.

4.4 Results and Discussion

Comparison of Analytical and Numerical Gradients

The implementation of the projection-based WF-in-DFT analytical gradient is tested by comparison with the gradient evaluated by numerical finite difference for a distorted geometry of ethanol. The finite difference gradients are evaluated using a four-point central difference formula with a base step size of 0.01 bohr. The mean absolute error (MAE) between the analytical and finite difference gradients is reported for a range of embedding calculations in Table 4.1. These results show that the analytical nuclear gradient for projection-based WF-in-DFT embedding is essentially numerically exact with respect to the gradients calculated by finite difference. Comparison of the results obtained using HF over the full system versus using LDA over the full system illustrate that some of the finite difference error comes from the DFT exchange-correlation grid. Comparison of the HF-in-HF results with full HF and of the LDA-in-LDA results with full LDA illustrate the modest effect of using a large-but-finite value for the level-shift operator in projection based embedding. These results confirm the correct implementation of projection-based WF-in-DFT analytical nuclear gradients.

Optimized Geometries

Ethanol

As a proof of concept, CCSD-in-LDA/6-31G analytical nuclear gradients are employed to determine the ground state geometry of ethanol, which is shown in Fig. 4.1. For this simple case, the O-H moiety is treated by CCSD and the remainder of the

Table 4.1: Mean absolute error between the analytically and numerically determined embedding nuclear gradient for a distorted geometry of ethanol. The basis set 6-31G is used for all calculations. The distorted geometry of ethanol is provided in the supporting information.

Method	MAE (hartree/bohr)
HF	5.00×10^{-9}
HF-in-HF	4.61×10^{-8}
LDA	1.48×10^{-8}
LDA-in-LDA	7.23×10^{-8}
HF-in-LDA	5.24×10^{-8}
MP2-in-LDA	5.37×10^{-8}
CCSD-in-LDA	5.36×10^{-8}
CCSD(T)-in-LDA	5.26×10^{-8}
CCSD-in-LDA (AO)[a]	3.48×10^{-8}
CCSD(T)-in-LDA (AO)[a]	3.40×10^{-8}
CCSD(T)-in-PBE0	5.26×10^{-8}
CCSD(T)-in-PBE0 (AO)[a]	1.12×10^{-7}

[a]Calculations were performed with AO truncation with a density threshold of 1×10^{-1}a.u.

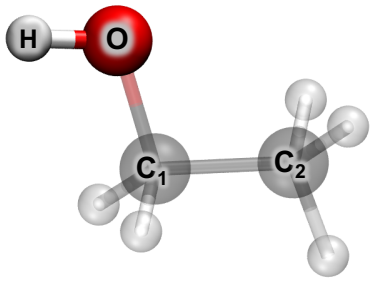

Figure 4.1: Optimized geometry of ethanol using projection-based CCSD-in-LDA/6-31G. The solid atoms (O and H) are in subsystem A and the transparent atoms are in subsystem B.

molecule is treated by LDA. Table 4.2 shows that the O-H bond length within subsystem A reproduces the CCSD predicted bond length of 0.979 Å and the remaining bonds within subsystem B reproduce the LDA predicted bond lengths. This indicates that the potential energy surface produced by projection-based embedding varies smoothly from CCSD-like interactions for subsystem A and LDA-like interactions for subsystem B. Interestingly, the C-O bond located at the boundary between subsystems A and B closely reproduces the LDA bond length and is not an interpolation between the CCSD and LDA bond lengths.

Table 4.2: Selected bond lengths and angles for ethanol (pictured in Fig.4.1) optimized at different levels of theory. Bond lengths are reported in units of Angstroms and angles are reported in units of degrees.

Method	r(O-H)	$\angle C_1OH$	r(C_1-C_2)	r(C_1-O)
LDA/6-31G	0.988	110.4	1.503	1.439
CCSD-in-LDA/6-31G	0.979	110.7	1.506	1.435
CCSD/6-31G	0.979	110.6	1.532	1.475

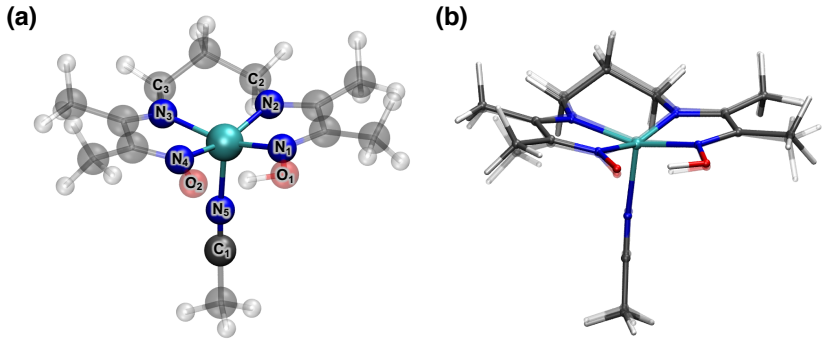

Figure 4.2: (a) The optimized geometry of the cobalt-based organometallic complex performed with projection-based CCSD-in-LDAX/def2-TZVPP:def2-SVP with AO truncation. The solid atoms (Co, N_1, N_2, N_3, N_4, N_5, and C_1) are included in subsystem A and the transparent atoms are included in subsystem B. (b) The LDAX/def2-TZVPP:def2-SVP optimized geometry (transparent) and the projection-based CCSD-in-LDAX/def2-TZVPP:def2-SVP with AO truncation optimized geometry (solid).

Cobalt-based Organometallic Complex

As a demonstration of embedding gradients with AO truncation, the geometry of the cobalt-based organometallic complex, shown in Fig. 4.2, is optimized. Fig. 4.2a shows the CCSD-in-LDAX/def2-TZVPP:def2-SVP optimized structure of the cobalt complex where the solid atoms are included in subsystem A and the transparent atoms are included in subsystem B. In Fig. 4.2b the optimized structures evaluated at the CCSD-in-LDAX/def2-TZVPP:def2-SVP (solid) and the LDAX/def2-TZVPP:def2-SVP (transparent) levels of theory are overlaid. While only modest differences are seen in the overall structure, Table 4.3 shows that the optimized bond lengths do change between the two levels of theory, both for the region within subsystem A and at the subsystem boundary. This indicates that the WF method is capable of relaxing the atoms in subsystem A even when they are strongly coordinated with subsystem B. It is also seen that the bond lengths across the boundary of subsystems A and B also differ from the LDAX geometry since the bonds in

Table 4.3: Selected bond lengths for the organometallic complex pictured in Fig.4.2 optimized at different levels of theory and their absolute difference (|Δ|). Bond lengths are reported in units of angstroms.

		LDAX	CCSD-in-LDAX	|Δ|
Sub A	r(Co-N_1)	1.836	1.846	0.010
	r(Co-N_2)	1.893	1.883	0.010
	r(Co-N_3)	1.932	1.951	0.019
	r(Co-N_4)	1.900	1.926	0.026
	r(Co-N_5)	1.978	2.026	0.048
Boundary	r(N_5-O_1)	1.317	1.355	0.038
	r(N_5-O_2)	1.262	1.301	0.039
	r(C_1-N_5)	1.131	1.150	0.019
	r(N_2-C_2)	1.430	1.458	0.028
	r(N_3-C_3)	1.428	1.458	0.030
Sub B	r(O_1-H)	1.025	1.030	0.005

question experience the effects of both the WF and KS methods. Finally, if a bond length associated with atoms in subsystem B is considered, such as the O_1-H bond, it is found to closely match the LDAX predicted bond length.

Malondialdehyde: Minimum Energy Reaction Pathway

The minimum energy reaction pathway for the proton transfer in malondialdehyde is determined using the NEB method. Fig. 4.3 shows that with minimal embedding (Fig. 4.3a) the CCSD-in-LDA/def2-ASVP reaction barrier, shown in Fig. 4.3b, is 4.85 kcal/mol which is within 1.5 kcal/mol of the CCSD/def2-ASVP reference reaction barrier of 6.12 kcal/mol. This is a vast improvement over the LDA/def2-ASVP result, which predicts an essentially barrierless reaction. In addition to correctly predicting the reaction barrier, Fig. 4.3c shows that the CCSD-in-LDA/def2-ASVP reaction pathway lies precisely on top of the CCSD/def2-ASVP pathway with only a small deviation in the basins. In contrast, Fig. 4.3d shows that the LDA/def2-ASVP reaction pathway and potential energy surface reveal errors in the location of the reactant and product basins, with the hydrogen-bond length vastly underestimated. This is consistent with the tendency of LDA to over stabilize hydrogen bonds.

Cobalt-based Organometallic Complex: Minimum Energy Reaction Pathway

The minimum energy reaction pathway for the intramolecular proton transfer in a cobalt diimine-dioxime catalyst (Fig. 4.4a) is now investigated. Previously, the

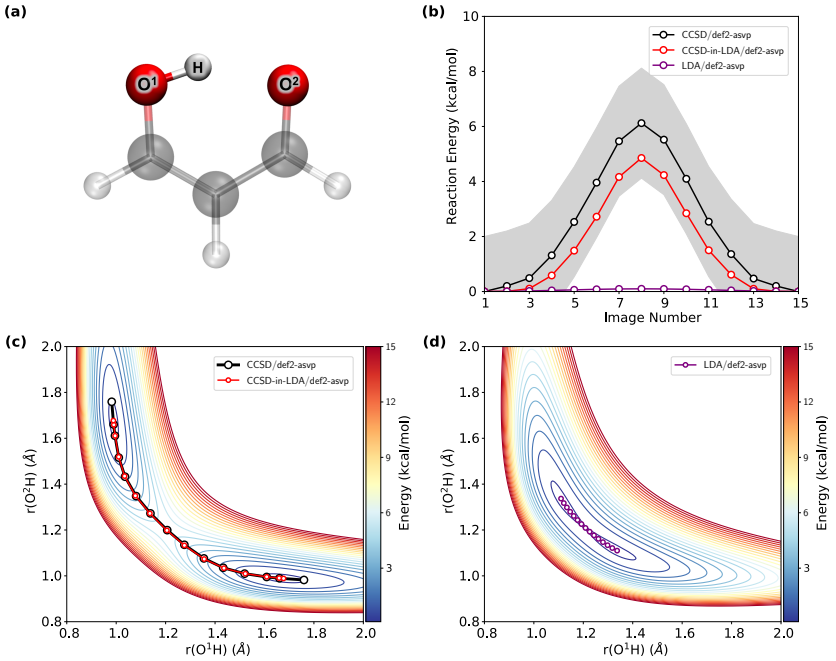

Figure 4.3: (a) The ground state geometry of malondialdehyde evaluated at the CCSD-in-LDA/def2-ASVP level of theory. The solid atoms are included in subsystem A and the transparent atoms are included in subsystem B. (b) The reaction barrier heights for the minimum energy reaction pathways for LDA/def2-ASVP, CCSD-in-LDA/def2-ASVP and CCSD/def2-ASVP. (c), (d) Also shown are the minimum energy reaction pathways of the proton transfer in malondialdehyde as a function of the distance of the proton from the oxygen atoms, O^1 on the x-axis, O^2 on the y-axis for the CCSD-in-LDA/def2-ASVP and CCSD/def2-ASVP levels of theory, (c), and for the LDA/def2-ASVP level of theory, (d).

reaction pathway for the transfer of the [-NH] to form a cobalt hydride had been investigated using geometries obtained using DFT [20]. Fig. 4.4b shows the energy profile for this reaction determined by various levels of theory. We observe that the reaction pathway determined by the NEB optimized at the PBE0/cc-pVDZ level of theory (purple curve) predicts a barrier height of 5.45 kcal/mol. However, when single-point PNO-LCCSD-in-PBE0/cc-pVDZ embedding energy calculations are run on the PBE0 optimized geometries (blue curve), the barrier height is lowered to 3.35 kcal/mol and the position of the transition state is shifted towards the reactant. The NEB optimized at the CCSD-in-PBE0/cc-pVDZ embedding level of theory (red curve) shows an even lower barrier height of 2.61 kcal/mol and predicts a substantially different transition state geometry (Fig. 4.4c) than the DFT result. The difference between the transition states predicted by the PBE0/cc-pVDZ and PNO-LCCSD-in-PBE0/cc-pVDZ levels of theory is clearly seen in Fig. 4.4d, which shows the projection of the NEB onto the two dimensions of the Co-H and N-H bonds

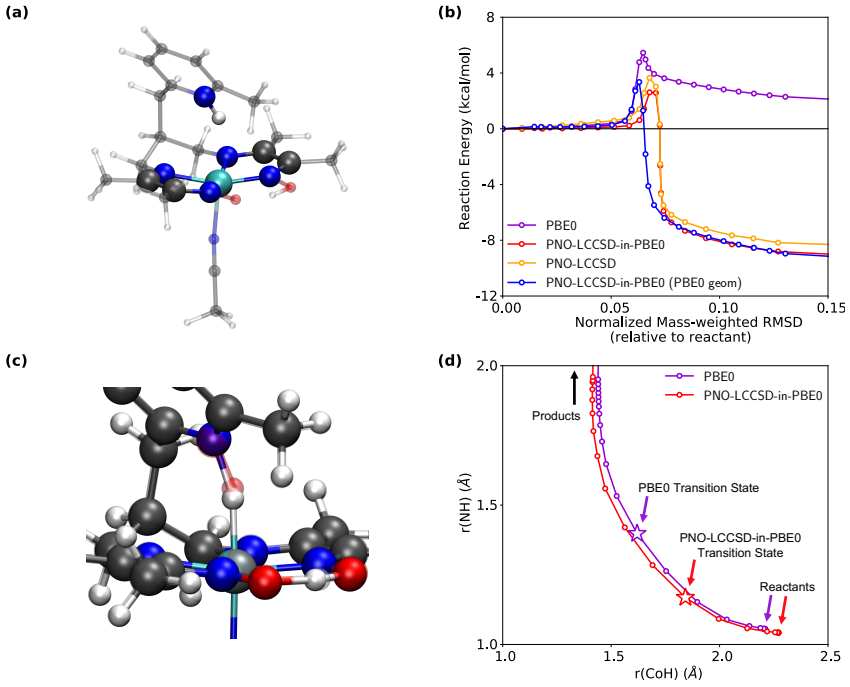

Figure 4.4: All calculations used the cc-pVDZ basis set. (a) The optimized geometry for a cobalt-based organometallic complex calculated at the CCSD-in-PBE0 level of theory with AO truncation. The solid atoms are included in subsystem A while the transparent atoms are in subsystem B. (b) The minimum energy reaction pathway for PBE0, the reaction pathway for PNO-LCCSD-in-PBE0 using the PBE0 geometries, the reaction pathway for PNO-LCCSD-in-PBE0 using CCSD-in-PBE0 geometries, and the reaction pathway for PNO-LCCSD using the CCSD-in-PBE0 geometries. The x-axis is a coordinate constructed by taking a normalized mass-weighted RMSD of all images along the pathway with respect to the respective reactant and product. In comparing the purple and blue curves versus the orange and red curves, note that the transition state position in these normalized coordinates is affected by changes in the geometries of the reactant and product. (c) A zoomed-in picture of the transition state geometries predicted by PBE0 (opaque atoms) and PNO-LCCSD-in-PBE0 (transparent atoms) levels of theory. The proton placement between the nitrogen and cobalt center at the PNO-LCCSD-in-PBE0 level of theory is highlighted in red. (d) The minimum energy reaction pathways of the proton transfer for PNO-LCCSD-in-PBE0 and PBE0 as a function of the distance of the proton from the cobalt atom on the x-axis, and the nitrogen atom on the y-axis. The placement of the transition states are highlighted for each level of theory.

and with the position of the transition state geometry indicated with stars. This result clearly shows the large degree to which commonly employed DFT transition state geometries can differ from the CCSD-quality result that is obtained using projection-based embedding.

4.5 Conclusions

We present the derivation and numerical demonstration of analytical nuclear gradients for projection-based embedding both with and without AO truncation. A

key aspect of the gradient theory is that the WF contributions can be evaluated using existing WF gradient implementations without the need for modification or additional programming, thereby allowing projection-based WF-in-DFT embedding gradients to be easily generalized to any combination of WF and KS-DFT methods. It is demonstrated that projection-based embedding gradients produce accurate geometries for a variety of benchmark systems, including for bond-lengths that span the interface between subsystems. Furthermore, in applications to both malondialdehyde and a transition-metal catalyst, WF-in-DFT minimum energy pathways obtained via the NEB method reveal large errors in DFT-computed transition-state energies and geometries. Finally, we note that the Lagrangian framework presented here can be used to derive other analytical gradients of the projection-based WF-in-DFT energy with respect to quantities such as electric and magnetic fields.

4.6 Supporting Information

All geometries used in all tables and figures are provided in the supporting information of Ref. 28.

4.7 Appendix

Pipek-Mezey Localization

Equation 4.14 from the main text

$$\begin{aligned}\left(\mathbf{a}\left[\mathbf{z}^{\mathrm{loc}}\right]\right)_{pi} &= \sum_{\mu} C_{\mu,p}\left(\sum_{k>l} z_{kl}^{\mathrm{loc}} \frac{\partial r_{kl}}{\partial C_{\mu,q}}\right) \\ &= \sum_{k>l} \mathcal{B}_{pi,kl} z_{kl}^{\mathrm{loc}}\end{aligned} \tag{4.36}$$

corresponds to the derivative of the localization conditions, Eq. 4.7, with respect to $\mathbf{C}$, where

$$\mathcal{B}_{pi,kl} = \sum_{C}\left[\left(2S_{pk}^{C}\delta_{ki} - 2S_{pl}^{C}\delta_{li}\right)S_{kl}^{C} + \left(S_{kk}^{C} - S_{ll}^{C}\right)\left(S_{pl}^{C}\delta_{ki} + S_{pk}^{C}\delta_{li}\right)\right]. \tag{4.37}$$

Next, the overlap derivative contribution from the localization conditions from Eq. 4.34 is

$$\begin{aligned}\sum_{i>j} \frac{\partial r_{ij}}{\partial S_{\mu\nu}} z_{ij}^{\mathrm{loc}} = \sum_{i>j} z_{ij}^{\mathrm{loc}}\left(1 - \mathcal{P}_{ij}\right) \sum_{C}\Big[&2L_{\mu,i}L_{\nu,i}S_{ij}^{C} \\ &+ S_{ii}^{C}\left(L_{\mu,i}L_{\nu,j} + L_{\mu,j}L_{\nu,i}\right)\Big]\Big|_{\mu\in C},\end{aligned} \tag{4.38}$$

where $\mathcal{P}_{ij}$ permutes the indices i and j, and μ is restricted to atomic orbitals on atom C.

Orbital Derivatives of Projection-based WF-in-DFT Embedding Energy

This appendix provides additional details for the terms in Eqs. 4.13 and 4.15 of the main text. The derivative of the projection-based WF-in-DFT embedding energy and the WF constraints with respect to the MO coefficients shown in Eq. 4.13 is

$$\begin{aligned} E_{pq} &= \sum_{\mu} C_{\mu,p} \left(\frac{\partial E_{\text{WF-in-DFT}}\left[\tilde{\Psi}^{\text{A}}; \boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right]}{\partial C_{\mu,q}} + \frac{\partial}{\partial C_{\mu,q}} \sum_{s} \Lambda_{s}^{\text{WF,A}} c_{s} \right) \\ &= \sum_{\mu} C_{\mu,p} \left(\frac{\partial E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right]}{\partial C_{\mu,q}} - \frac{\partial E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}}\right]}{\partial C_{\mu,q}} + \text{tr}\left[\frac{\partial \left(\tilde{\mathbf{d}}_{\text{rel}}^{\text{A}} - \boldsymbol{\gamma}^{\text{A}}\right)}{\partial C_{\mu,q}} \mathbf{v}_{\text{emb}} \right] \right. \\ &\quad \left. + \text{tr}\left[\left(\tilde{\mathbf{d}}_{\text{rel}}^{\text{A}} - \boldsymbol{\gamma}^{\text{A}}\right) \frac{\partial \mathbf{v}_{\text{emb}}}{\partial C_{\mu,q}} \right] + \mu \text{tr}\left[\frac{\partial \tilde{\mathbf{d}}_{\text{rel}}^{\text{A}}}{\partial C_{\mu,q}} \mathbf{P}^{\text{B}} \right] + \mu \text{tr}\left[\tilde{\mathbf{d}}_{\text{rel}}^{\text{A}} \frac{\partial \mathbf{P}^{\text{B}}}{\partial C_{\mu,q}} \right] \right), \end{aligned} \tag{4.39}$$

where the partial derivative of the WF constraints causes the appearance of the WF relaxed density, $\tilde{\mathbf{d}}_{\text{rel}}^{\text{A}}$, in the last four terms on the RHS of Eq. 4.39. Equation 4.39 simplifies to

$$\begin{aligned} E_{pq} = 4\left(\mathbf{F}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right]\right)_{pq}\Big|_{q\in\text{occ}} - 4\left(\mathbf{F}\left[\boldsymbol{\gamma}^{\text{A}}\right]\right)_{pq}\Big|_{q\in\text{A}} - 4(\mathbf{v}_{\text{emb}})_{pq}\Big|_{q\in\text{A}} \\ + 4\left(\mathbf{M}\left[\tilde{\mathbf{d}}_{\text{rel}}^{\text{A}} - \boldsymbol{\gamma}^{\text{A}}\right]\right)_{pq}, \end{aligned} \tag{4.40}$$

where $q \in \text{occ}$ indicates that the index q is restricted to LMOs, $q \in \text{A}$ indicates that q is restricted to LMOs in subsystem A, and $\mathbf{F}$ is the KS Fock matrix evaluated with the bracketed density. The last term on the RHS of Eq. 4.40

$$(\mathbf{M}[\boldsymbol{\gamma}])_{pq} = \frac{1}{4} \sum_{\mu} C_{\mu,p} \left(\text{tr}\left[\boldsymbol{\gamma} \frac{\partial \mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right]}{\partial C_{\mu,q}} \right] + \mu \text{tr}\left[\boldsymbol{\gamma} \frac{\partial \mathbf{P}^{\text{B}}}{\partial C_{\mu,q}} \right] \right), \tag{4.41}$$

simplifies to

$$\begin{aligned} (\mathbf{M}[\boldsymbol{\gamma}])_{pq} = \sum_{\mu\nu} C_{\mu,p} \sum_{\lambda\sigma} \gamma_{\lambda\sigma} \left((\mu\nu|\lambda\sigma) - \tfrac{1}{2} x_f (\mu\lambda|\nu\sigma) \right) L_{\nu,q}\Big|_{q\in\text{B}} \\ + \mu\left(\mathbf{C}^{\dagger}\mathbf{P}[\boldsymbol{\gamma}]\mathbf{L}\right)_{pq}\Big|_{q\in\text{B}} + \left(\tilde{\mathbf{v}}_{\text{xc}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}, \boldsymbol{\gamma}\right]\right)_{pq}\Big|_{q\in\text{occ}} \\ + \left(\tilde{\mathbf{v}}_{\text{xc}}\left[\boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}\right]\right)_{pq}\Big|_{q\in\text{A}}, \end{aligned} \tag{4.42}$$

where

$$\mathbf{P}[\boldsymbol{\gamma}] = \mathbf{S}\boldsymbol{\gamma}\mathbf{S}. \tag{4.43}$$

In the current study, we employ both LDA and GGA exchange-correlation functionals; for the special case of LDA, the term $\tilde{\mathbf{v}}_{\text{xc}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}, \boldsymbol{\gamma}\right]$ assumes the form

$$\left(\tilde{\mathbf{v}}_{\text{xc}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}, \boldsymbol{\gamma}\right]\right)_{pq} = \sum_{mn} \left(pq|f_{\text{xc}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right]|mn \right) \boldsymbol{\gamma}_{mn}, \tag{4.44}$$

where f_{xc} is the XC kernel which is defined as the second derivative of the XC functional with respect to density.

The derivative of the Brillouin conditions in Eq. 4.15 can be expanded as follows.

$$\begin{aligned}
(\mathbf{D}[\mathbf{z}])_{pq} &= \sum_{ak} \mathcal{D}_{pq,ak} z_{ak} \\
&= \sum_{ak} z_{ak} \Bigg[\left(\mathbf{F}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right]\right)_{pk} \delta_{aq} + \left(\mathbf{F}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right]\right)_{ap} \delta_{kq} \\
&\qquad + 2 \sum_{l} \delta_{ql} \Big(2(ak|pl) - \tfrac{1}{2} x_f (ap|kl) - \tfrac{1}{2} x_f (al|kp) \Big) \\
&\qquad + \sum_{\mu\lambda\sigma} C_{\mu,p} C_{\lambda,a} \frac{\partial \left(\mathbf{v}_{\mathrm{xc}}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right]\right)_{\lambda\sigma}}{\partial C_{\mu,q}} L_{\sigma,k} \Bigg] \\
&= \left(\mathbf{F}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right]\mathbf{z}\right)_{pq}\Big|_{q\in\mathrm{occ}} + \left(\mathbf{F}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right]\mathbf{z}^{\dagger}\right)_{pq}\Big|_{q\in\mathrm{vir}} \\
&\quad + 2(\mathbf{V}[\bar{\mathbf{z}}])_{pq}\Big|_{q\in\mathrm{occ}},
\end{aligned} \tag{4.45}$$

where $\bar{\mathbf{z}} = \mathbf{z} + \mathbf{z}^{\dagger}$ and $\mathbf{V}[\bar{\mathbf{z}}]$ is defined as

$$\begin{aligned}
(\mathbf{V}[\bar{\mathbf{z}}])_{pq}\Big|_{q\in\mathrm{occ}} &= \sum_{mn} \bar{\mathbf{z}}_{mn} \Big((mn|pq) - \tfrac{1}{2} x_f (mp|nq) \Big)\Big|_{q\in\mathrm{occ}} \\
&\quad + \left(\tilde{\mathbf{v}}_{\mathrm{xc}}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}, \bar{\mathbf{z}}\right]\right)_{pq}\Big|_{q\in\mathrm{occ}}.
\end{aligned} \tag{4.46}$$

Atomic Orbital Truncation

Projection-based WF-in-DFT embedding reduces the cost of the WF calculation on subsystem A by reducing the number of LMOs that are correlated at the WF level, but thus far leaves the virtual space untouched. However, the scaling of most WF methods is dominated by the number of virtual MOs (e.g. $\mathcal{O}(v^4)$ for CCSD). One strategy has been to employ local correlation WF methods such as PNO-LMP2 [186] and PNO-LCCSD [187, 188] to describe subsystem A since these methods are able to leverage the reduced number of LMOs to significantly lower the number of occupied-virtual orbital pairs that need to be included, resulting in a cheap and accurate WF calculation. However, a real advantage of projection-based embedding hinges on being able to use any WF method to describe subsystem A. Therefore, having a more general approach to reduce the cost of the WF calculation on the subsystem A is desirable.

The AO truncation scheme devised by Bennie et al. [18] provides a simple way to significantly reduce the cost of the WF calculation by reducing the size of the

basis used to describe subsystem A. The AOs that are discarded are selected through a single density threshold parameter: if the net Mulliken population, computed using the subsystem A density, of an AO is less than the specified threshold, it is removed from the basis set. This scheme has shown to greatly speedup up WF-in-DFT calculations at a small cost in accuracy in total and relative energies [18]. Additionally, it has the nice feature that given a fixed subsystem A, the size of the truncated subsystem A basis scales asymptotically as the size of the environment grows. This basis set modification does not cause any complications in the evaluation of the subsystem A WF gradient so existing implementations can be used without any modifications. The energy expression for a projection-based WF-in-DFT calculation with AO truncation using the so-called type-in-type correction [18] is

$$\begin{aligned} E^{\text{trun}}_{\text{WF-in-DFT}}\left[\tilde{\Psi}^{\text{A,trun}}; \bar{\boldsymbol{\gamma}}^{\text{A,trun}}; \boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right] &= E_{\text{WF}}\left[\tilde{\Psi}^{\text{A,trun}}\right] \\ &- E^{\text{trun}}_{\text{DFT}}\left[\bar{\boldsymbol{\gamma}}^{\text{A,trun}}\right] + E_{\text{DFT}}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right] \\ &+ \text{tr}\left[\left(\tilde{\mathbf{d}}^{\text{A,trun}} - \bar{\boldsymbol{\gamma}}^{\text{A,trun}}\right)\mathbf{v}^{\text{trun}}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right]\right] \\ &+ \text{tr}\left[\left(\tilde{\mathbf{d}}^{\text{A,trun}} - \bar{\boldsymbol{\gamma}}^{\text{A,trun}}\right)\mathbf{P}^{\text{B,trun}}\right], \end{aligned} \tag{4.47}$$

where $\tilde{\Psi}^{\text{A,trun}}$ is the subsystem A WF in the truncated basis, $\bar{\boldsymbol{\gamma}}^{\text{A,trun}}$ is the KS subsystem A one-particle density in the truncated basis, $\boldsymbol{\gamma}^{\text{A}}$ and $\boldsymbol{\gamma}^{\text{B}}$ are the KS subsystem A and B one-particle densities in the full basis respectively, $\tilde{\mathbf{d}}^{\text{A,trun}}$ is the subsystem A one-particle reduced density matrix that corresponds to $\tilde{\Psi}^{\text{A,trun}}$, $\mathbf{v}^{\text{trun}}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right]$ is the embedding potential in the truncated basis which is evaluated by

$$\mathbf{v}^{\text{trun}}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right] = \mathbf{P}_{\text{t}}^{\dagger}\mathbf{v}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right]\mathbf{P}_{\text{t}}, \tag{4.48}$$

and $\mathbf{P}^{\text{B,trun}}$ is the projection operator in the truncated basis which is evaluated by

$$\mathbf{P}^{\text{B,trun}} = \mathbf{P}_{\text{t}}^{\dagger}\mathbf{P}^{\text{B}}\mathbf{P}_{\text{t}}. \tag{4.49}$$

Here, $\mathbf{P}_{\text{t}}$ is the rectangular matrix that maps the full basis to the truncated basis which is created by starting with identity matrix and deleting columns corresponding to thrown away AO functions. We note that the even though the notation for Eq. 4.47 is different from the one used in Ref. 18 the approach is identical.

Projection-based WF-in-DFT Gradient Theory with AO Truncation

Total Energy Lagrangian

We now derive the total energy Lagrangian for for projection-based WF-in-DFT embedding with AO truncation. The WF-in-DFT AO truncation Lagrangian is

$$\begin{aligned}\mathcal{L} =& E^{\text{trun}}_{\text{WF-in-DFT}}\left[\tilde{\Psi}^{\text{A,trun}}; \bar{\boldsymbol{\gamma}}^{\text{A,trun}}; \boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right] + \sum_{s} \Lambda^{\text{WF,A}}_{s} c_s \\ &- \sum_{ij\in\bar{\text{A}}} \bar{\epsilon}^{\text{A}}_{ij}\left[\bar{\mathbf{C}}^{\text{A}\dagger}\mathbf{S}\bar{\mathbf{C}}^{\text{A}} - \mathbf{1}\right]_{ij} + \sum_{pq} x_{pq}\left[\mathbf{C}^{\dagger}\mathbf{S}\mathbf{C} - \mathbf{1}\right]_{pq} + \sum_{i>j} z^{\text{loc}}_{ij} r_{ij} \\ &+ \sum_{ai} z_{ai}\left(\mathbf{F}\left[\boldsymbol{\gamma}^{\text{A}} + \boldsymbol{\gamma}^{\text{B}}\right]\right)_{ai},\end{aligned} \tag{4.50}$$

where the bar superscript refers to subsystem A quantities optimized by the KS functional in the truncated basis. The constraints that appear in Eq. 4.50 are all the same as those that appear in Eq. 4.6 from the main text, except for the third term on the RHS of Eq. 4.50. This term constrains the MOs, $\bar{\mathbf{C}}^{\text{A}}$, to be orthogonal.

Minimizing the Lagrangian with respect to the variational parameters of the WF method – $\tilde{\Psi}^{\text{A,trun}}$.

Minimizing the WF-in-DFT AO truncation Lagrangian with respect to $\tilde{\Psi}^{\text{A,trun}}$ simplifies to the minimization of the subsystem A WF energy and the WF constraints (as explained in section 4.2), which corresponds to the conventional WF Lagrangian used to derive WF gradient theories, albeit in the truncated basis.

Minimizing the Lagrangian with respect to the MO coefficients, $\bar{\mathbf{C}}^{\text{A}}$.

The minimization of the Lagrangian with respect to the optimized KS MO coefficients in the truncated basis, $\bar{\mathbf{C}}^{\text{A}}$, results in the SCF equations using the embedded Fock matrix.

$$\begin{aligned}\sum_{\mu} \bar{C}^{\text{A}}_{\mu,i}\Bigg(&\frac{\partial E^{\text{trun}}_{\text{DFT}}[\bar{\boldsymbol{\gamma}}^{\text{A,trun}}]}{\partial \bar{C}^{A}_{\mu,j}} + \frac{\partial \text{tr}\left[\bar{\boldsymbol{\gamma}}^{\text{A,trun}}\mathbf{v}^{\text{trun}}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right]\right]}{\partial \bar{C}^{A}_{\mu,j}} \\ &+ \mu\frac{\partial \text{tr}\left[\bar{\boldsymbol{\gamma}}^{\text{A,trun}}\mathbf{P}^{\text{B,trun}}\right]}{\partial \bar{C}^{A}_{\mu,j}}\Bigg) = \sum_{\mu} \bar{C}^{\text{A}}_{\mu,i} \sum_{kl\in\bar{\text{A}}} \bar{\epsilon}^{\text{A}}_{kl}\frac{\partial \bar{S}^{A}_{kl}}{\partial \bar{C}^{\text{A}}_{\mu,j}}\end{aligned} \tag{4.51}$$

$$\begin{aligned}\sum_{\mu\nu} \bar{C}^{A}_{\mu,i}\Bigg(&\left(\mathbf{F}\left[\bar{\boldsymbol{\gamma}}^{\text{A,trun}}\right]\right)_{\mu\nu} + \left(\mathbf{v}^{\text{trun}}_{\text{emb}}\left[\boldsymbol{\gamma}^{\text{A}}, \boldsymbol{\gamma}^{\text{B}}\right]\right)_{\mu\nu} \\ &+ \mu\left[\mathbf{P}^{\text{B,trun}}\right]_{\mu\nu}\Bigg)\bar{C}^{A}_{\nu,j} = \frac{1}{2}\bar{\epsilon}^{\text{A}}_{ij}\Big|_{ij\in\text{A}}\end{aligned} \tag{4.52}$$

$$\left(\mathbf{F}^{\mathrm{A}}\left[\bar{\boldsymbol{\gamma}}^{\mathrm{A,trun}}\right]\right)_{ij}\Big|_{ij\in\mathrm{A}} = \frac{1}{2}\bar{\epsilon}^{\mathrm{A}}_{ij}\Big|_{ij\in\mathrm{A}}. \tag{4.53}$$

Therefore, the Lagrange multipliers $\frac{1}{2}\bar{\epsilon}^{\mathrm{A}}$ are simply the MO eigen energies of the KS optimized subsystem A MOs in the truncated basis.

Minimize the Lagrangian with respect to the MO coefficients, C.

The minimization of the Lagrangian with respect to the KS MO coefficients in the full basis, $\mathbf{C}$, is

$$\sum_{\mu} C_{\mu,p}\frac{\partial \mathcal{L}}{\partial C_{\mu,q}} = E_{pq} + \left(\mathbf{a}\left[\mathbf{z}^{\mathrm{loc}}\right]\right)_{pi} + (\mathbf{D}[\mathbf{z}])_{pq} + 2x_{pq} = 0, \tag{4.54}$$

where only the matrix $\mathbf{E}$ differs from the ones outlined in Eqns. 4.13-4.16.

$$E_{pq} = 4\left(\mathbf{F}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right]\right)_{pq}\Big|_{q\in\mathrm{occ}} + 4\left(\mathbf{M}\left[\mathbf{P}_{\mathrm{t}}\left(\tilde{\mathbf{d}}^{\mathrm{A,trun}}_{\mathrm{rel}} - \bar{\boldsymbol{\gamma}}^{\mathrm{A,trun}}\right)\mathbf{P}_{\mathrm{t}}^{\dagger}\right]\right)_{pq}. \tag{4.55}$$

With the updated $\mathbf{E}$ matrix, the Lagrangian multipliers are solved in the same way as outlined for the WF-in-DFT Lagrangian multipliers.

Gradient of the Total Energy

Once the Lagrangian is minimized with respect to all variational parameters, the gradient of the energy with respect to nuclear coordinate, q, takes the form

$$\begin{aligned} E^{\mathrm{trun},(q)}_{\text{WF-in-DFT}} &= E^{\mathrm{trun},q}_{\mathrm{WF}}\left[\tilde{\Psi}^{\mathrm{A,trun}}\right] - E^{\mathrm{trun},q}_{\mathrm{DFT}}\left[\bar{\boldsymbol{\gamma}}^{\mathrm{A,trun}}\right] \\ &+ \mathrm{tr}\left[\mathbf{d}_{\mathrm{a}}\mathbf{h}^{(q)}\right] + \mathrm{tr}\left[\mathbf{X}\mathbf{S}^{(q)}\right] + \frac{1}{2}\sum_{\mu\nu\lambda\sigma} D_{\mu\nu\lambda\sigma}(\mu\nu|\lambda\sigma)^{(q)} \\ &+ E^{(q)}_{\mathrm{xc}}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right] + \mathrm{tr}\left[\mathbf{d}_{\mathrm{c}}\left(\mathbf{v}^{(q)}_{\mathrm{xc}}\left[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right] - \mathbf{v}^{(q)}_{\mathrm{xc}}\left[\boldsymbol{\gamma}^{\mathrm{A}}\right]\right)\right], \end{aligned} \tag{4.56}$$

where the first two terms on the RHS of Eq. 4.56, $E^{\mathrm{trun},q}_{\mathrm{WF}}$ and $E^{\mathrm{trun},q}_{\mathrm{DFT}}$, are the total derivative of the truncated subsystem A WF and KS energy respectively, minus the embedding contribution, $\mathbf{v}^{\mathrm{trun,(q)}}_{\mathrm{emb}}$. These two terms are calculated using existing gradient implementations, whereas the embedding contribution has been folded into the remaining terms. The effective one-particle densities $\mathbf{d}_{\mathrm{a}}$, $\mathbf{d}_{\mathrm{b}}$ and $\mathbf{d}_{\mathrm{c}}$ are

$$\mathbf{d}_{\mathrm{a}} = \boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}} + \mathbf{C}\mathbf{z}\mathbf{C}^{\dagger}, \tag{4.57}$$

$$\mathbf{d}_{\mathrm{b}} = \boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}} + 2\mathbf{C}\mathbf{z}\mathbf{C}^{\dagger} + 2\mathbf{d}_{\mathrm{c}}, \tag{4.58}$$

and

$$\mathbf{d}_{\mathrm{c}} = \mathbf{P}_{\mathrm{t}}\left(\tilde{\mathbf{d}}^{\mathrm{A,trun}}_{\mathrm{rel}} - \bar{\boldsymbol{\gamma}}^{\mathrm{A,trun}}\right)\mathbf{P}_{\mathrm{t}}^{\dagger}. \tag{4.59}$$

The effective two-particle density $\mathbf{D}$ is

$$\begin{aligned} D_{\mu\nu\lambda\sigma} = &\left(\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right)_{\mu\nu}(\mathbf{d}_{\mathrm{b}})_{\lambda\sigma} - 2\gamma^{\mathrm{A}}_{\mu\nu}(\mathbf{d}_{\mathrm{c}})_{\lambda\sigma} \\ &- \frac{1}{2}x_f\left(\left(\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}\right)_{\mu\lambda}(\mathbf{d}_{\mathrm{b}})_{\nu\sigma} - 2\gamma^{\mathrm{A}}_{\mu\lambda}(\mathbf{d}_{\mathrm{c}})_{\nu\sigma}\right). \end{aligned} \tag{4.60}$$

The matrix $\mathbf{X}$ is

$$\begin{aligned} \mathbf{X} = &\mathbf{X}^{\mathrm{loc}} - \frac{1}{2}\mathbf{L}\big(\mathbf{E} + 2\mathbf{V}[\bar{\mathbf{z}}]\big)\mathbf{L}^{\dagger} - \frac{1}{2}\left(\mathbf{C}_{\mathrm{v}}(\mathbf{zF})\mathbf{L}^{\dagger} + \left(\mathbf{C}_{\mathrm{v}}(\mathbf{zF})\mathbf{L}^{\dagger}\right)^{\dagger}\right) \\ &+ \mu\left(\mathbf{d}_{\mathrm{c}}\mathbf{S}\boldsymbol{\gamma}^{\mathrm{B}} + \boldsymbol{\gamma}^{\mathrm{B}}\mathbf{S}\mathbf{d}_{\mathrm{c}}\right), \end{aligned} \tag{4.61}$$

where

$$\left(\mathbf{X}^{\mathrm{loc}}\right)_{\mu\nu} = -\frac{1}{2}\left(\mathbf{La}\left[\mathbf{z}^{\mathrm{loc}}\right]\mathbf{L}^{\dagger}\right)_{\mu\nu} + \sum_{i>j}\frac{\partial r_{ij}}{\partial S_{\mu\nu}}z^{\mathrm{loc}}_{ij}. \tag{4.62}$$

Chapter 5

ANALYTICAL GRADIENTS FOR MOLECULAR-ORBITAL-BASED MACHINE LEARNING

Adapted from:

[1]S. J. R. Lee, T. Husch, F. Ding, and T. F. Miller III, "Analytical gradients for molecular-orbital-based machine learning", in preparation (2020).

Molecular-orbital-based machine learning (MOB-ML) enables the prediction of accurate correlation energies at the cost of obtaining molecular orbitals. Here, we present the derivation, implementation, and numerical demonstration of MOB-ML analytical nuclear gradients which are formulated in a general Lagrangian framework to enforce orthogonality, localization, and Brillouin constraints on the molecular orbitals. The MOB-ML gradient framework is general with respect to the regression technique (e.g., Gaussian process regression or neural networks) and the MOB feature design. We show that MOB-ML gradients are highly accurate compared to other ML methods on the ISO17 data set while only being trained on energies for hundreds of molecules compared to energies and gradients for hundreds of thousands of molecules for the other ML methods. The MOB-ML gradients are shown to yield high-quality optimized structures, at a computational cost for the gradient evaluation that is comparable to Hartree-Fock theory or hybrid DFT.

5.1 Introduction

This chapter focuses on the development of analytical nuclear gradients for the molecular-orbital-based machine learning (MOB-ML) method [70–72]. MOB-ML relies on information of local molecular orbitals to predict the pair-wise sum of a post-Hartree–Fock correlation energy at drastically reduced cost. However, taking the derivative of the MOB-ML energy is similarly as challenging as taking the derivative for non-canonical wave function-based correlation methods due to the complexity arising from the invoked local approximations and the non-variational nature inherent to MOB-ML. There exists only a handful of local wave function-based correlation methods for which this effort has been performed [160, 189]. In this work, we establish a general Lagrangian framework to obtain the analytical nuclear gradients of a MOB-ML energy. The framework enforces orthogonality, localization, and Brillouin constraints on the molecular orbitals (section 5.2). A noteworthy aspect of this framework is that it allows taking the gradient of the MOB-ML energy irrespective of which energy method MOB-ML was trained to predict. Consequently, we are able to take the gradient of a MOB-ML method trained to predict an accurate wave function theory for which gradients have not been derived (yet) without any modification of the framework. Furthermore, the computational cost of evaluating a MOB-ML gradient is comparable to the one of evaluating a Hartree–Fock (HF) gradient, or equivalently, a hybrid density functional theory (DFT) gradient, making it orders of magnitude faster than evaluating the gradients of *ab initio* wave function theories. We will numerically validate the gradient theory by comparison to energy finite difference in section 5.4. After assuring the correctness of our gradient theory, we compare the MOB-ML gradients to the ones obtained by the reference theories MOB-ML was trained on. Specifically, we will show that we obtain accurate and systematically improvable gradients for potential energy surfaces of single molecules and for diverse sets of molecules (section 5.4). The comparison to other standard ML methods on the example of the ISO17 data set highlights the data efficiency and high transferability of MOB-ML regarding nuclear gradients.

5.2 MOB-ML Analytical Nuclear Gradients

MOB-ML Energy Theory

MOB-ML relies on molecular orbital information from a HF calculation to predict a wave function correlation energy. The working equation for the MOB-ML energy

is [70–72]

$$E_{\text{MOB-ML}}[\mathbf{f}] = E_{\text{corr}}[\mathbf{f}] + E_{\text{HF}}, \tag{5.1}$$

where E_{HF} is the HF energy, and $E_{\text{corr}}[\mathbf{f}]$ is the machine-learned correlation energy,

$$E_{\text{corr}}[\mathbf{f}] = \sum_i \epsilon_{ii}[\mathbf{f}_i] + 2\sum_{i>j} \epsilon_{ij}\left[\mathbf{f}_{ij}\right]. \tag{5.2}$$

The matrix of feature vectors, $\mathbf{f}$, is divided into two sub-classes. The first sub-class is made up by the diagonal components of $\mathbf{f}$, $\mathbf{f}_i$, which represent the valence-occupied orbital i. The second sub-class is made up by the off-diagonal components of $\mathbf{f}$, $\mathbf{f}_{ij}$, which represent the interaction between the valence-occupied orbitals i and j. Both diagonal and off-diagonal feature vectors are composed of elements from the HF Fock matrix in the MO basis, $\mathbf{F}$, and the MO repulsion integrals, $\boldsymbol{K}$, where

$$[\boldsymbol{K}^{pq}]_{mn} = (pq|mn) = \sum_{\mu\nu\kappa\sigma} C_{\mu p} C_{\nu q} C_{\kappa m} C_{\sigma n} (\mu\nu|\kappa\sigma). \tag{5.3}$$

Here, $(\mu\nu|\kappa\sigma)$ are the four-center atomic orbital integrals with μ, ν, κ, and σ representing atomic orbital indices. We restrict the MO indices of $\mathbf{F}$ and $\boldsymbol{K}$ to the valence-occupied and valence-virtual MOs and we only include 2-center Coulomb- and exchange-type MO integrals, $[\boldsymbol{K}^{pp}]_{qq}$ and $[\boldsymbol{K}^{pq}]_{pq}$ respectively. The feature vectors are further processed following the procedure outlined in Ref. 190.

MOB-ML Gradient Theory

Lagrangian framework

MOB-ML is a non-variational theory which allows us to derive its analytical nuclear gradient within a Lagrangian framework,

$$\begin{aligned} \frac{\mathrm{d}E_{\text{MOB-ML}}}{\mathrm{d}q} = \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}q} &= \frac{\partial \mathcal{L}}{\partial q} + \frac{\partial \mathcal{L}}{\partial \mathbf{C}} \frac{\partial \mathbf{C}}{\partial q} \\ &= \frac{\partial \mathcal{L}}{\partial q}, \end{aligned} \tag{5.4}$$

where q refers to nuclear coordinate. The calculation of the nuclear response of the HF MOs, $\partial\mathbf{C}/\partial q$, is avoided because the Lagrangian $\mathcal{L}$ is minimized with respect to all of its variational parameters which are the MO coefficients, $\mathbf{C}$. The MOB-ML energy Lagrangian is

$$\begin{aligned} \mathcal{L}\left[\mathbf{C}, \mathbf{x}, \mathbf{z}, \mathbf{z}^{\text{core}}, \mathbf{z}^{\text{val-occ}}, \mathbf{z}^{\text{val-vir}}, \boldsymbol{\lambda}\right] &= E_{\text{MOB-ML}}[\mathbf{f}] + \sum_{pq} x_{pq}\Big(\mathbf{C}^\dagger \mathbf{S}\mathbf{C} - \mathbf{I}\Big)_{pq} \\ &+ \sum_{ai} z_{ai} F_{ai}\Big|_{i\in\text{occ}, a\in\text{vir}} + \sum_{ri} z_{ri}^{\text{core}} F_{ri} + \sum_{i>j} z_{ij}^{\text{val-occ}} r_{ij} + \sum_{ab} z_{ab}^{\text{val-vir}} r_{ab} \\ &+ \sum_{wa} \lambda_{wa} P_{wa}, \end{aligned} \tag{5.5}$$

where $\mathbf{x}$, $\mathbf{z}$, $\mathbf{z}^{\text{core}}$, $\mathbf{z}^{\text{val-occ}}$, $\mathbf{z}^{\text{val-vir}}$, and $\boldsymbol{\lambda}$ are the Lagrange multipliers. We refer to the core MOs with column indices r, s, to the valence-occupied localized MOs (LMOs) with column indices i, j, k, l, to the valence-virtual LMOs with column indices a, b, and to the non-valence-virtual MOs with column indices w, x. The indices m, n, p, q are used to index generic molecular orbitals. The first term on the right hand side (RHS) of Eq. 5.5 is the MOB-ML energy described by Eq. 5.1. The second term on the RHS constrains the HF MOs, $\mathbf{C}$, to be orthonormal, which is commonly referred to as the Pulay force [158]. The third term on the RHS is known as the Brillouin conditions, which account for the dependence of the correlation energy on the HF optimized molecular orbitals. The frozen-core conditions, $F_{ri} = 0$, account for neglecting the correlation energy contributions from the core orbitals. The localization conditions, $r_{ij} = 0$ and $r_{ab} = 0$, account for how the valence-occupied and valence-virtual MOs are localized respectively. In this work, we employ Foster-Boys localization [159] and intrinsic bond orbitals (IBO) localization [130], but it is straightforward to generalize to other localization methods. The valence virtual conditions, $P_{wa} = 0$, reflect how the valence virtual MOs are obtained through a unitary transformation of the virtual MOs. This unitary transformation corresponds to the column space of a projection matrix formed by projecting the virtual MOs onto the IAOs. The complementary null space of this projection matrix corresponds to the non valence-virtual orbitals. This projection matrix is defined as

$$\mathbf{P} = \mathbf{C}_{\text{vir}}^{\text{IAO},\dagger}\mathbf{C}_{\text{vir}}^{\text{IAO}}, \tag{5.6}$$

where

$$\mathbf{C}_{\text{vir}}^{\text{IAO}} = \mathbf{X}_{\text{occ}}^{\text{IAO},\dagger}\mathbf{S}_1\mathbf{C}_{\text{vir}}, \tag{5.7}$$

and where $\mathbf{C}_{\text{vir}}$ is the virtual MO coefficient matrix. The matrix $\mathbf{X}^{\text{IAO}}$ transforms between the original AO and IAO basis sets and is expanded in Appendix 5.7. All together, this yields the following analytical nuclear gradient,

$$\begin{aligned}\frac{\mathrm{d}E_{\text{MOB-ML}}}{\mathrm{d}q} = {} & E_{\text{ML}}^{(q)} + E_{\text{HF}}^{(q)} + \sum_{pq} x_{pq}\left(\mathbf{C}^{\dagger}\mathbf{S}^{(q)}\mathbf{C}\right)_{pq} + \sum_{ai} z_{ai}F_{ai}^{(q)}\Big|_{a\in\text{vir},i\in\text{occ}} \\ & + \sum_{ri} z_{ri}^{\text{core}}F_{ri}^{(q)} + \sum_{i>j} z_{ij}^{\text{val-occ}}r_{ij}^{(q)} + \sum_{ab} z_{ab}^{\text{val-vir}}r_{ab}^{(q)} \\ & + \sum_{wa} \lambda_{wa}P_{wa}^{(q)},\end{aligned} \tag{5.8}$$

where the superscript (q) denotes the explicit derivative of the quantity with respect to a nuclear coordinate. Eq. 5.8 is the general MOB-ML analytical nuclear gradient and we will now outline how to determine the Lagrange multipliers for our particular use case to arrive at a final working equation.

Minimizing the Lagrangian with respect to MO coefficients

All of the Lagrange multipliers ($\mathbf{x}$, $\mathbf{z}$, $\mathbf{z}^{\text{core}}$, $\mathbf{z}^{\text{val-occ}}$, $\mathbf{z}^{\text{val-vir}}$ and $\boldsymbol{\lambda}$) are determined by minimizing the MOB-ML Lagrangian with respect to its variational parameters, which are the MO coefficients, $\mathbf{C}$. Differentiating the Lagrangian with respect to these parameters yields

$$\begin{aligned}\sum_{\mu} C_{\mu p}\frac{\partial \mathcal{L}}{\partial C_{\mu q}} &= E_{pq} + 2x_{pq} + (\mathbf{D}[\mathbf{z}])_{pq} + (\mathbf{D}[\mathbf{z}^{\text{core}}])_{pq} + \left(\mathbf{a}\left[\mathbf{z}^{\text{val-occ}}\right]\right)_{pq} \\ &+ \left(\mathbf{a}\left[\mathbf{z}^{\text{val-vir}}\right]\right)_{pq} + (\mathbf{D}[\boldsymbol{\lambda}])_{pq} = 0,\end{aligned} \tag{5.9}$$

where

$$\begin{aligned}E_{pq} &= \sum_{\mu} C_{\mu p}\frac{\partial (E_{\text{corr}}[\mathbf{f}] + E_{\text{HF}})}{\partial C_{\mu q}} \\ &= 4F_{pq}\Big|_{q\in\text{occ}} + \left(\mathbf{F}\bar{\mathbf{D}}^{\text{F}}\right)_{pq}\Big|_{q\in\text{loc}} + 2\left(\mathbf{g}\left[\mathbf{C}\bar{\mathbf{D}}^{\text{F}}\mathbf{C}^{\dagger}\right]\right)_{pq}\Big|_{q\in\text{occ}} \\ &+ 2\sum_{m}[\boldsymbol{K}^{pq}]_{mm}\left(D^{\text{J}}_{qm} + D^{\text{J}}_{mq}\right)\Big|_{mq\in\text{loc}} \\ &+ 2\sum_{m}[\boldsymbol{K}^{pm}]_{qm}\left(D^{\text{K}}_{qm} + D^{\text{K}}_{mq}\right)\Big|_{mq\in\text{loc}} \\ &+ \sum_{nm} R^{n}_{pm}\left(D^{\text{R},n}_{qm} + D^{\text{R},n}_{mq}\right)\Big|_{mq\in\text{val-occ}},\end{aligned} \tag{5.10}$$

$$\begin{aligned}(\mathbf{D}[\mathbf{z}])_{pq} &= \sum_{\mu} C_{\mu p}\left(\sum_{ai} z_{ai}\frac{\partial F_{ai}}{\partial C_{\mu q}}\right)\Big|_{i\in\text{occ},a\in\text{vir}} \\ &= (\mathbf{F}\mathbf{z})_{pq}\Big|_{q\in\text{occ}} + \left(\mathbf{F}\mathbf{z}^{\dagger}\right)_{pq}\Big|_{q\in\text{vir}} + 2(\mathbf{g}[\bar{\mathbf{z}}])_{pq}\Big|_{q\in\text{occ}},\end{aligned} \tag{5.11}$$

$$\begin{aligned}(\mathbf{D}[\mathbf{z}^{\text{core}}])_{pq} &= \sum_{\mu} C_{\mu p}\left(\sum_{rk} z^{\text{core}}_{rk}\frac{\partial F_{rk}}{\partial C_{\mu q}}\right) \\ &= (\mathbf{F}\mathbf{z}^{\text{core}})_{pq}\Big|_{q\in\text{val-occ}} + \left(\mathbf{F}\mathbf{z}^{\text{core},\dagger}\right)_{pq}\Big|_{q\in\text{core}} + 2(\mathbf{g}[\bar{\mathbf{z}}^{\text{core}}])_{pq}\Big|_{q\in\text{occ}},\end{aligned} \tag{5.12}$$

$$\left(\mathbf{a}\left[\mathbf{z}^{\text{val-occ}}\right]\right)_{pq} = \sum_{\mu} C_{\mu p}\left(\sum_{i>j} z^{\text{loc}}_{ij}\frac{\partial r_{ij}}{\partial C_{\mu q}}\right), \tag{5.13}$$

$$\left(\mathbf{a}\left[\mathbf{z}^{\text{val-vir}}\right]\right)_{pq} = \sum_{\mu} C_{\mu p}\left(\sum_{a>b} z^{\text{vir}}_{ab}\frac{\partial r_{ab}}{\partial C_{\mu q}}\right), \tag{5.14}$$

and

$$(\mathbf{D}[\boldsymbol{\lambda}])_{pq} = \sum_{\mu} C_{\mu p}\left(\sum_{wa}\lambda_{wa}\frac{\partial P_{wa}}{\partial C_{\mu q}}\right) = (\mathbf{P}\boldsymbol{\lambda})_{pq}\Big|_{q\in\text{non-val-vir}}. \tag{5.15}$$

Eqs. 5.13 and 5.14 are expanded in Appendices 5.7 and 5.7, respectively, $\mathbf{F}$ is the HF Fock matrix, $\mathbf{g}$ includes all of the usual HF two-electron terms, $\mathbf{R}^n$ is expanded in Appendix 5.7, the condition $q \in \text{loc}$ restricts the sum to valence-occupied and valence-virtual MOs, $\bar{\mathbf{z}} = \mathbf{z} + \mathbf{z}^\dagger$, $\bar{\mathbf{z}}^{\text{core}} = \mathbf{z}^{\text{core}} + \mathbf{z}^{\text{core},\dagger}$, and $\bar{\mathbf{D}}^{\text{F}} = \mathbf{D}^{\text{F}} + \mathbf{D}^{\text{F},\dagger}$. The matrices $\mathbf{D}^{\text{F}}$, $\mathbf{D}^{\text{J}}$, and $\mathbf{D}^{\text{K}}$ are calculated by

$$D^{\text{M}}_{pq} = \sum_i \frac{\partial \epsilon_{ii}[\mathbf{f}_i]}{\partial \mathbf{f}_i} \frac{\partial \mathbf{f}_i}{\partial M_{pq}}\Big|_{pq\in\text{loc}} + 2\sum_{i>j} \frac{\partial \epsilon_{ij}\left[\mathbf{f}_{ij}\right]}{\partial \mathbf{f}_{ij}} \frac{\partial \mathbf{f}_{ij}}{\partial M_{pq}}\Big|_{pq\in\text{loc}}, \tag{5.16}$$

where M_{pq} refers to F_{pq}, $[\boldsymbol{K}^{pp}]_{qq}$ and $[\boldsymbol{K}^{pq}]_{pq}$, respectively. The matrix $\mathbf{D}^{\text{R,n}}$ is

$$D^{\text{R},n}_{pq} = 2\sum_{i>j} \frac{\partial \epsilon_{ij}\left[\mathbf{f}_{ij}\right]}{\partial \mathbf{f}_{ij}} \frac{\partial \mathbf{f}_{ij}}{\partial R^n_{pq}}\Big|_{pq\in\text{val-occ}}. \tag{5.17}$$

The partial derivatives $\frac{\partial \epsilon_{ii}[\mathbf{f}_i]}{\partial \mathbf{f}_i}$ and $\frac{\partial \epsilon_{ij}\left[\mathbf{f}_{ij}\right]}{\partial \mathbf{f}_{ij}}$ on the RHS of Eqns. 5.16 and 5.17 are the derivatives of the machine learning prediction with respect to the feature vectors.

We want to highlight that any machine learning method (e.g. Gaussian process, regression clustering, neural net, etc.) that has this feature vector derivative of its prediction can be readily used in this gradient framework without any modifications to the framework itself. Furthermore, we want to emphasize that the following analytical nuclear gradient derivation generalizes to any type of feature vector design and construction as long as the feature elements are obtained from $\mathbf{F}$ and $\boldsymbol{K}$. The partial derivatives $\frac{\partial \mathbf{f}_i}{\partial M_{pq}}$, $\frac{\partial \mathbf{f}_{ij}}{\partial M_{pq}}$ and $\frac{\partial \mathbf{f}_{ij}}{\partial R^n_{pq}}$ are expanded in Appendix 5.7.

We now proceed to solve for each of the Lagrange multipliers. First, combining the stationary conditions described by Eq. 5.9 with the auxiliary conditions $\boldsymbol{x} = \boldsymbol{x}^\dagger$ yields the linear Z-vector equations

$$\left(1 - \mathcal{P}_{pq}\right)(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{D}[\mathbf{z}^{\text{core}}] + \mathbf{a}\left[\mathbf{z}^{\text{val-occ}}\right] + \mathbf{a}\left[\mathbf{z}^{\text{val-vir}}\right] + \mathbf{D}[\boldsymbol{\lambda}])_{pq} = 0, \tag{5.18}$$

where $\mathcal{P}_{pq}$ permutes the indices p and q, which is used to solve for $\mathbf{z}$, $\mathbf{z}^{\text{core}}$, $\mathbf{z}^{\text{val-occ}}$, $\mathbf{z}^{\text{val-vir}}$ and $\boldsymbol{\lambda}$. The matrix $\boldsymbol{x}$ is then obtained as

$$\begin{aligned} x_{pq} = -\tfrac{1}{4}\left(1 + \mathcal{P}_{pq}\right)\big(&\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{D}[\mathbf{z}^{\text{core}}] + \mathbf{a}\left[\mathbf{z}^{\text{val-occ}}\right] \\ &+ \mathbf{a}\left[\mathbf{z}^{\text{val-vir}}\right] + \mathbf{D}[\boldsymbol{\lambda}]\big)_{pq}. \end{aligned} \tag{5.19}$$

We show that the Lagrange multipliers $\mathbf{z}^{\text{val-occ}}$ are solved by considering the (valence-occupied)-(valence-occupied) part of Eq. 5.18, yielding

$$\left(1 - \mathcal{P}_{ij}\right)(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{D}[\mathbf{z}^{\text{core}}] + \mathbf{a}\left[\mathbf{z}^{\text{val-occ}}\right] + \mathbf{a}\left[\mathbf{z}^{\text{val-vir}}\right] + \mathbf{D}[\boldsymbol{\lambda}])_{ij} = 0. \tag{5.20}$$

Eq. 5.20 can be further simplified by showing that

$$\begin{aligned}
\big(1 - \mathcal{P}_{ij}\big)(\mathbf{D}[\mathbf{z}])_{ij} &= 0, \\
\big(1 - \mathcal{P}_{ij}\big)(\mathbf{D}[\mathbf{z}^{\text{core}}])_{ij} &= 0, \\
\big(1 - \mathcal{P}_{ij}\big)\Big(\mathbf{a}\big[\mathbf{z}^{\text{val-vir}}\big]\Big)_{ij} &= 0, \\
\big(1 - \mathcal{P}_{ij}\big)(\mathbf{D}[\boldsymbol{\lambda}])_{ij} &= 0.
\end{aligned} \tag{5.21}$$

As a result, $\mathbf{z}^{\text{val-occ}}$ is independent of all other Lagrange multipliers, which simplifies Eq. 5.20 to

$$E_{ij} - E_{ji} + \sum_{k>l} \big(\mathcal{B}_{ijkl} - \mathcal{B}_{jikl}\big) z_{kl}^{\text{val-occ}} = 0, \tag{5.22}$$

where the 4-dimensional tensor $\mathcal{B}$ is expanded in Appendix 5.7. The set of linear system of equations defined by Eq. 5.22 are the Z-vector coupled perturbed localization (Z-CPL) equations which are used to solve for $\mathbf{z}^{\text{val-occ}}$. Subsequently, Eq. 5.13 can be used to compute $\mathbf{a}\big[\mathbf{z}^{\text{val-occ}}\big]$.

Next we show that the Lagrange multipliers $\mathbf{z}^{\text{core}}$ are solved by considering the core-(valence-occupied) part of Eq. 5.18, yielding

$$(1 - \mathcal{P}_{ri})(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{D}[\mathbf{z}^{\text{core}}] + \mathbf{a}\big[\mathbf{z}^{\text{val-occ}}\big] + \mathbf{a}\big[\mathbf{z}^{\text{val-vir}}\big] + \mathbf{D}[\boldsymbol{\lambda}])_{ri} = 0, \tag{5.23}$$

which further simplifies to

$$E_{ri} - E_{ir} + \Big(\mathbf{a}\big[\mathbf{z}^{\text{val-occ}}\big]\Big)_{ri} + (\mathbf{F}\mathbf{z}^{\text{core}} - \mathbf{z}^{\text{core}}\mathbf{F})_{ri} = 0. \tag{5.24}$$

These are the Z-vector equations used to solve for $\mathbf{z}^{\text{core}}$. Subsequently, Eq. 5.12 can be used to calculate $\mathbf{D}[\mathbf{z}^{\text{core}}]$.

Next we show that the Lagrange multipliers $\mathbf{z}^{\text{val-vir}}$ are solved by considering the (valence-virtual)-(valence-virtual) part of Eq. 5.18 yielding

$$(1 - \mathcal{P}_{ab})(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{D}[\mathbf{z}^{\text{core}}] + \mathbf{a}\big[\mathbf{z}^{\text{val-occ}}\big] + \mathbf{a}\big[\mathbf{z}^{\text{val-vir}}\big] + \mathbf{D}[\boldsymbol{\lambda}])_{ab} = 0, \tag{5.25}$$

which further simplifies to

$$E_{ab} - E_{ba} + \sum_{c>d} C_{abcd} z_{cd}^{\text{val-vir}} = 0, \tag{5.26}$$

where the 4-dimensional tensor C is expanded in Appendix 5.7. These are the Z-CPL equations which are used to solve for $\mathbf{z}^{\text{val-vir}}$. Subsequently, Eq. 5.14 can be used to compute $\mathbf{a}\big[\mathbf{z}^{\text{val-vir}}\big]$.

Next we show that the Lagrange multipliers $\boldsymbol{\lambda}$ are solved by considering the (non valence-virtual)-(valence-virtual) part of Eq. 5.18, yielding

$$(1 - \mathcal{P}_{wa})(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{D}[\mathbf{z}^{\text{core}}] + \mathbf{a}\left[\mathbf{z}^{\text{val-occ}}\right] + \mathbf{a}\left[\mathbf{z}^{\text{val-vir}}\right] + \mathbf{D}[\boldsymbol{\lambda}])_{wa} = 0, \quad (5.27)$$

which further simplifies to

$$E_{wa} - E_{aw} + \mathbf{a}\left[\mathbf{z}^{\text{val-vir}}\right]_{wa} - (\mathbf{P}\boldsymbol{\lambda})_{aw} = 0. \quad (5.28)$$

These are the Z-vector equations used to solve for $\boldsymbol{\lambda}$. Subsequently, Eq. 5.15 can be used to compute $\mathbf{D}[\boldsymbol{\lambda}]$.

Next we show that the Lagrange multipliers $\mathbf{z}$ are solved by considering the virtual-occupied part of Eq. 5.18, yielding

$$(1 - \mathcal{P}_{ai})(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{D}[\mathbf{z}^{\text{core}}] + \mathbf{a}\left[\mathbf{z}^{\text{val-occ}}\right] + \mathbf{a}\left[\mathbf{z}^{\text{val-vir}}\right] + \mathbf{D}[\boldsymbol{\lambda}])_{ai}\Big|_{a\in\text{vir},i\in\text{occ}} = 0, \quad (5.29)$$

which further simplifies to

$$\begin{aligned} E_{ai} - E_{ia} + (2\mathbf{g}[\bar{\mathbf{z}}^{\text{core}}])_{ai} + \left(\mathbf{a}\left[\mathbf{z}^{\text{val-occ}}\right]\right)_{ai} - \left(\mathbf{a}\left[\mathbf{z}^{\text{val-vir}}\right]\right)_{ia} \\ - (\mathbf{P}\boldsymbol{\lambda})_{ia} + (\mathbf{F}\mathbf{z} - \mathbf{z}\mathbf{F} + 2\mathbf{g}[\bar{\mathbf{z}}])_{ai} = 0. \end{aligned} \quad (5.30)$$

Here, the MO indices a and i refer to the full virtual and occupied spaces, respectively. These are the Z-vector coupled perturbed Hartree–Fock (Z-CPHF) equations. With the solutions to all Z-vector equations we can return to Eq. 5.19 to solve for $\boldsymbol{x}$.

Incorporating localization strategy

Finally, to provide the working expression of Eq. 5.8 in terms of derivative AO integrals we must choose specific localization methods. For this derivation we choose the Foster-Boys and IBO localization methods to localize the valence-occupied and valence-virtual orbitals respectively, yielding

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial q} = \text{tr}\left[\mathbf{d}_{\text{a}}\mathbf{h}^{(q)}\right] + \text{tr}\left[\mathbf{X}_1\mathbf{S}_1^{(q)}\right] + \text{tr}\left[\mathbf{X}_2\mathbf{S}_2^{(q)}\right] + \text{tr}\left[\mathbf{X}_{12}\mathbf{S}_{12}^{(q)}\right] \\ + \sum_n \text{tr}\left[\mathbf{W}_n(\mathbf{R}^n)^{(q)}\right] + \tfrac{1}{2}\sum_{\mu\nu\lambda\sigma} D_{\mu\nu\kappa\sigma}(\mu\nu|\kappa\sigma)^{(q)}, \end{aligned} \quad (5.31)$$

where $\mathbf{h}$ is the standard one-electron Hamiltonian, μ, ν, κ and σ label AO basis functions in the original basis, $(\mu\nu|\kappa\sigma)$ are the two-electron repulsion integrals, $\mathbf{S}_2$ is the overlap matrix of the minimal AO basis (MINAO) used in the IBO procedure,

and $\mathbf{S}_{12}$ is the overlap matrix between the original AO and MINAO basis sets. The effective one-particle density $\mathbf{d}_\mathrm{a}$ is defined as

$$\mathbf{d}_\mathrm{a} = \boldsymbol{\gamma} + \tfrac{1}{2}\mathbf{C}\bar{\mathbf{D}}^\mathrm{F}\mathbf{C}^\dagger + \tfrac{1}{2}\mathbf{C}\bar{\mathbf{z}}\mathbf{C}^\dagger + \tfrac{1}{2}\mathbf{C}\bar{\mathbf{z}}^\mathrm{core}\mathbf{C}^\dagger, \tag{5.32}$$

where $\boldsymbol{\gamma}$ is the full system HF density. The effective two-particle density $\mathbf{D}$ is defined as

$$\begin{aligned} D_{\mu\nu\kappa\sigma} = (\mathbf{d}_\mathrm{b})_{\mu\nu}\gamma_{\kappa\sigma} - \tfrac{1}{2}(\mathbf{d}_\mathrm{b})_{\mu\kappa}\gamma_{\nu\sigma} + 2\sum_{pq} D^\mathrm{J}_{pq} C_{\mu p} C_{\nu p} C_{\kappa q} C_{\sigma q} \\ + 2\sum_{pq} D^\mathrm{K}_{pq} C_{\mu p} C_{\kappa p} C_{\nu q} C_{\sigma q}, \end{aligned} \tag{5.33}$$

where the effective one-particle density $\mathbf{d}_\mathrm{b}$ is defined as

$$\mathbf{d}_\mathrm{b} = \boldsymbol{\gamma} + \mathbf{C}\bar{\mathbf{D}}^\mathrm{F}\mathbf{C}^\dagger + \mathbf{C}\bar{\mathbf{z}}\mathbf{C}^\dagger + \mathbf{C}\bar{\mathbf{z}}^\mathrm{core}\mathbf{C}^\dagger. \tag{5.34}$$

The matrices $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_{12}$, and $\mathbf{W}_n$ are defined as

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{C}\mathbf{x}\mathbf{C}^\dagger + \sum_{a>b} \frac{\partial r_{ab}}{\partial \mathbf{S}_1} z_{ab}^\text{val-vir} + \sum_{wa} \frac{\partial P_{wa}}{\partial \mathbf{S}_1}\lambda_{wa}, \\ \mathbf{X}_2 &= \sum_{a>b} \frac{\partial r_{ab}}{\partial \mathbf{S}_2} z_{ab}^\text{val-vir} + \sum_{wa} \frac{\partial P_{wa}}{\partial \mathbf{S}_2}\lambda_{wa}, \\ \mathbf{X}_{12} &= \sum_{a>b} \frac{\partial r_{ab}}{\partial \mathbf{S}_{12}} z_{ab}^\text{val-vir} + \sum_{wa} \frac{\partial P_{wa}}{\partial \mathbf{S}_{12}}\lambda_{wa} \end{aligned} \tag{5.35}$$

and

$$\mathbf{W}_n = \sum_{i>j} \frac{\partial r_{ij}}{\partial \mathbf{R}^n} z_{ij}^\text{val-occ} + \mathbf{C}\mathbf{D}^{\mathrm{R},n}\mathbf{C}^\dagger, \tag{5.36}$$

where Eq. 5.35 is expanded in Appendix 5.7 and Eq. 5.36 is expanded in Appendix 5.7.

5.3 Computational Details

In this work, we perform calculations on four different data sets: (i) the thermalized water data set published in Ref. 71, (ii) a thermalized set of organic molecules featuring up to seven heavy atoms (QM7b-T) [71], and (iii) the ISO17 set of short trajectories for constitutional isomers with the chemical formula $C_7O_2H_{10}$ [57].

All MOB-ML energy and analytical gradient are implemented in and performed with ENTOS QCORE [150]. The DF-HF calculations for the QM7b-T set [71], the ISO17 set [57] are performed with a cc-pVTZ [80] basis set and a cc-pVTZ-JKFIT density

fitting basis [183]. The DF-HF calculations for the water calculations are performed with a aug-cc-pVTZ [191] and a aug-cc-pVTZ-JKFIT [183] basis set. We employ a molecular orbital convergence threshold of `orbital_grad_threshold` = 1×10^{-8} a.u. In all MOB-ML calculations, the Foster–Boys [159] localization method is used to localize the valence-occupied MOs. The valence-virtual space is either localized with Foster–Boys localization (QM7b-T, ISO17) or the IBO localization method [130] (water). The diagonal and off-diagonal feature vectors are constructed following the procedure outlined in Ref. 190. For all Z-CPHF calculations a convergence threshold of 1×10^{-8} a.u. is specified.

All WF calculations are performed in Molpro [192] with the frozen-core approximation, and with density fitting. All WF pair energy calculations employ the non-canonical MP2 [131, 167, 193, 194] or non-canonical coupled-cluster singles, doubles, and perturbative triples [CCSD(T)] [44, 195–200] correlation treatments with the cc-pVTZ, cc-pVTZ-MP2FIT, [185] aug-cc-pVTZ and aug-cc-pVTZ-MP2FIT [185] basis sets. An interface between Molpro and ENTOS QCORE is used such that WF pair energies are calculated using the DF-HF LMOs produced by ENTOS QCORE. All WF gradient calculations employ the canonical MP2 or CCSD(T) correlation treatments with the aug-cc-pVTZ, aug-cc-pVTZ-JKFIT and aug-cc-pVTZ-MP2FIT basis sets. For all Z-CPHF calculations needed for the WF gradient an iterative solver with a convergence threshold of 1×10^{-9} a.u. is used.

The MOB-ML models for water are trained on non-canonical CCSD(T)/aug-cc-pVTZ pair correlation energies. When constructing the feature vector all non-zero elements from the Fock and $\boldsymbol{K}$ matrices are used. All linear regression (LR) models are trained using Scikit-Learn [201]. All Gaussian process regression (GPR) [202] models use the Matern 5/2 kernel [202, 203] and are optimized using the scaled conjugate gradient option in GPy [204]. All regression clustering models are trained following the framework outlined in Ref. 72 using a GPR within each cluster.

The MOB-ML models for the QM7b-T data set, and the ISO17 data set are trained on non-canonical MP2/cc-pVTZ pair correlation energies. Feature selection is performed using random forest regression [205] with the mean decrease of accuracy criterion, which is sometimes referred to as permutation importance [206]. All GPR models use the Matern 5/2 kernel and are optimized using the scaled conjugate gradient option in GPy.

Table 5.1: Mean absolute error (MAE) of the MOB-ML analytical nuclear gradient with respect to the MOB-ML numerical nuclear gradient for a non-equilibrium geometry of water. The numerical nuclear gradients in were obtained with a two-step central difference formula with a step size of 5×10^{-4} bohr. The non-equilibrium geometry of water has bond lengths of 0.986Å and 0.958Å, and a bond angle of 94.5°. All MOB-ML models are trained on data for 100 water geometries.

Regression technique	MAE (hartree/bohr)
Linear regression	1.45×10^{-8}
Gaussian process regression	3.75×10^{-8}
Clustered Gaussian process regression	2.28×10^{-8}

5.4 Results and Discussion

First, we compare the MOB-ML analytical gradient to the numerical gradient for an exemplary molecule to ascertain the correctness of our derivation and implementation in Table 5.1. Table 5.1 shows that the mean absolute errors (MAE) of the analytical MOB-ML gradients of a distorted water molecule with respect to the numerical ones are on the order of 10^{-8} hartree/bohr for all MOB-ML models. A similar MAE is commonly found when comparing analytical and numerical gradients of pure electronic structure methods [28, 160, 189, 197]. Additionally, Table 5.1 shows that the difference of the numerical and analytical gradient is largely independent of the regression technique (linear regression, Gaussian process regression, or a clustered Gaussian process regression) applied within the MOB-ML model. More generally, this illustrates (as also pointed out in Section 5.2) that any desired regression technique can be applied within MOB-ML without changes to the gradient framework provided that the regression prediction is differentiable with respect to the features.

In a next step, we now consider a large part of the potential energy surface of a single water molecule, following our previous work [71]. Fig. 5.1 shows the MAE for the energy predictions and for the associated analytical gradients we obtained with MOB-ML models trained on CCSD(T) energies for an increasing number of water geometries. As already highlighted in Ref. 72, the MAE for the energy prediction decreases steeply with the number of training geometries and we reach an MAE of 2×10^{-4} kcal/mol when training on correlation energies of 100 training geometries. Additionally, we can now see that the MAE of the analytical MOB-ML gradients with respect to the analytical CCSD(T) gradients strictly decreases with an increasing amount of training data although the training data in this context are correlation energy labels and not gradients. The MAE of the MOB-ML analytical gradient is

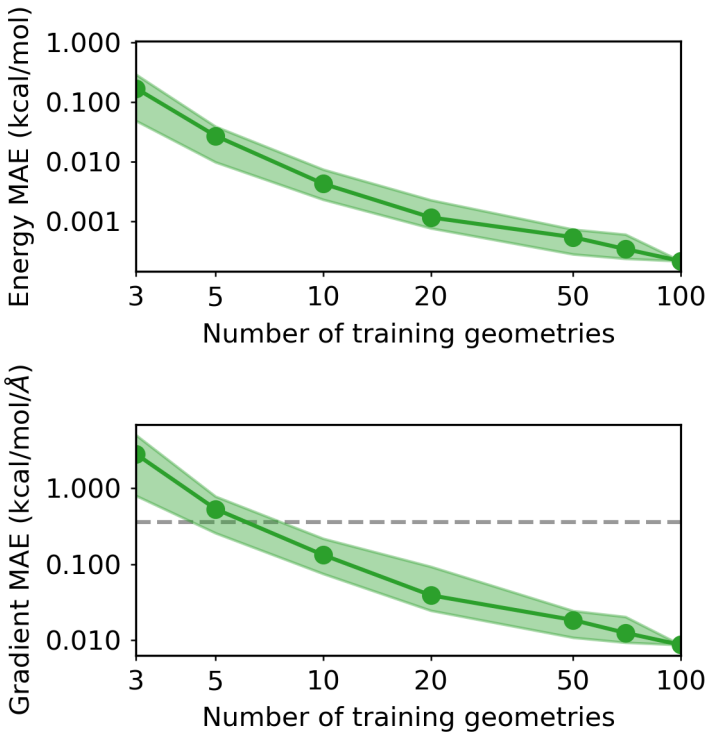

Figure 5.1: Mean absolute errors (MAE) of the total correlation energy prediction (top panel) and of the associated analytical gradient (bottom panel) as a function of the number CCSD(T) reference calculations for water geometries chosen for training MOB-ML models. The green circles correspond to the mean MAE obtained from 50 random samples of the training data, the green shaded area corresponds to the 90% confidence interval for the predictions and for the gradients obtained from 50 random samples. The black horizontal line at 0.3 mH/bohr in the bottom panel indicates the commonly used threshold to determine geometry optimization convergence.

9×10^{-3} kcal/mol/Å when training on correlation energies for 100 water geometries. We can contextualize this result by considering that the threshold commonly used to determine if a structure optimization is converged is 0.36 kcal/mol/Å. The MAE for the gradient drops below this threshold when training on as few as three to nine water geometries. This demonstrates that MOB-ML is able to describe potential energy surfaces to a high accuracy and with a high data efficiency.

We now examine if this result generalizes to a diverse set of molecules. To this end, we first study the QM7b-T data set which is comprised of a thermalized set of 7211 organic molecules with 7 or fewer heavy atoms [207]. Fig. 5.2 shows the MAE for the MOB-ML energy prediction and for the associated analytical gradient with respect to the corresponding MP2 quantities as a function of the number of MP2 reference energy calculations.

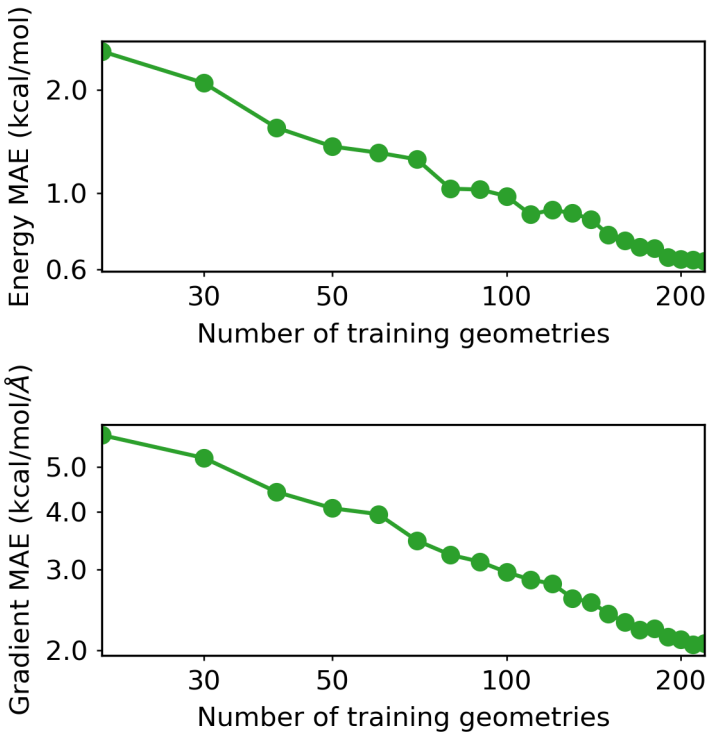

Figure 5.2: Mean absolute errors (MAE) of the total correlation energy prediction (top panel) and of the associated analytical gradient (bottom panel) as a function of the number MP2 reference calculations for QM7b-T geometries randomly chosen for training MOB-ML models.

As already reported in Ref. 190, the learning curve for the energy decreases steeply and we obtain an MAE of 1.0 kcal/mol when training on about 70 structures. The decrease in the MAE for the energy prediction is accompanied by a decrease in the MAE for the analytical MOB-ML gradient with respect to the analytical MP2 gradient. We reach a MAE of 2.08 kcal/mol/A when training on 220 structures. To put this number into context and to compare with other machine learning methods, we now also examine the ISO17 data set [57].

The ISO17 data set consists of short MD trajectories for constitutional isomers with the composition $C_7O_2H_{10}$. Table 5.2 shows the performance of two MOB-ML models, one trained on 220 QM7b-T structures and one trained on 100 ISO17 structures, and summarizes the MAEs obtained with other ML models in the literature, i.e., SchNet [57], FHCL [64], PhysNet [65], the shared-weight neural network (SWNN) [66], GM-sNN [68] and GNNFF [69]. The MOB-ML models are the only ML models which are on average chemically accurate although the MOB-ML models were only trained on energies for 100 ISO17 molecules and 220 QM7b-T molecules, respectively. The fact that our model trained on a small set of the seven-heavy atom

molecules which are smaller in size than ISO17 and which are chemically more diverse (QM7b-T additionally contains the elements N, S, Cl) showcases again how transferable and data efficient MOB-ML models are. The next best model in terms of the energy MAE is GM-sNN which was trained on energies and gradients for 400k ISO17 structures and achieves an MAE of 1.97 kcal/mol. The force MAE of the MOB-ML models (1.63 and 1.64 kcal/mol/Å, respectively) is comparable to that of GM-sNN (1.66 kcal/mol/Å) while employing only 0.025% of the training data. MOB-ML is significantly more accurate in the forces than other models trained on energies alone, i.e., SchNet which obtained an MAE of 5.71 kcal/mol/Å and SWNN which obtained an MAE of 6.61 kcal/mol/Å. The only model which is more accurate in terms of the force MAE is PhysNet which is trained on energies and forces for 400k ISO17 structures. PhysNet obtains a force MAE of 1.38 kcal/mol/A. Given the demonstrated learnability of forces, it is very likely that MOB-ML could be trained to be more accurate by including more training data. Furthermore, analytical gradients have not been derived for all reference theories which considerably limits the scope of these machine learning methodologies. For example, the popular local coupled cluster methods [187, 208, 209] do not currently have derived analytical gradient theories

Despite comparing favorably to other ML methods, it is still an open question if a force MAE of 1.63 kcal/mol/Å is sufficient for an actual application. Therefore, we now use our model for structure optimizations which is a very frequent task for quantum chemical methods. To be able to gauge how a certain force MAE translates into optimized structures, we optimize the constitutional isomers in ISO17 with MP2 and with MOB-ML and compare the resulting structures via the root mean square deviation (RMSD) of the atoms positions in Figure 5.3.

Figure 5.3 shows that the MOB-ML optimized structures are very similar to the MP2 optimized structures with a mean RMSD of 0.01 Å. Reassuringly, the structures are significantly and systematically more similar to the MP2 structures than HF structures which exhibit an average RMSD of 0.03 Å, where HF is the theory we correct with MOB-ML. Additionally, we can ascertain that MOB-ML structures are more similar to MP2 structures than a B3LYP-D3 (as a typical DFT functional) structure is to an MP2 structure. In fact, B3LYP-D3 structures exhibit an average

Table 5.2: Comparison of the mean absolute error for the prediction of energies and atomic forces for the unknown test set of the ISO17 data set obtained with different ML methods. The different ML methods applied different training sizes and drew on different labels to train the models on. Energy and force errors are reported in kcal/mol and kcal/mol/Å, respectively.

Method	Training Size	Trained on energies		Trained on energies+gradients	
		Energy MAE	Force MAE	Energy MAE	Force MAE
SchNet [57]	400,000	3.11	5.71	2.40	2.18
FCHL [64]	1,000	—	—	3.70	3.50
PhysNet [65]	400,000	—	—	2.94	1.38
SWNN [66]	400,000	3.72	6.61	8.57	6.74
GM-sNN[68]	400,000	—	—	1.97	1.66
GNNFF [69]	400,000	—	—	—	2.02
MOB-ML	**100**	**0.84**	**1.64**	—	—
MOB-ML	**220***	**0.76**	**1.63**	—	—

*This MOB-ML model was trained on 220 randomly selected structures from the QM7b-T data set.

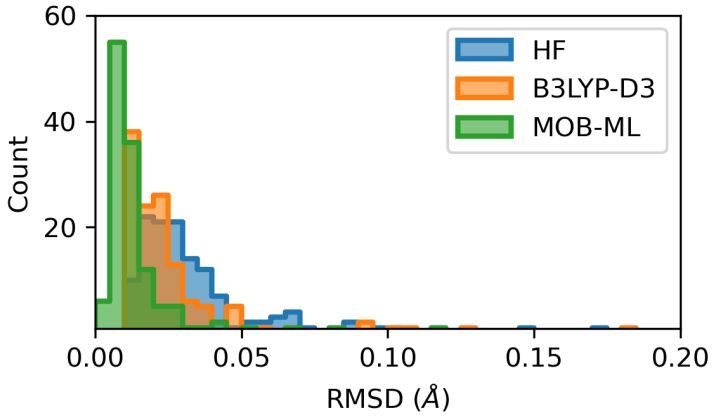

Figure 5.3: Histogrammed root mean square deviations (RMSD) of HF structures (blue), B3LYP-D3 structures (orange), and MOB-ML structures (green) with respect to MP2 structures for the unique isomers in the ISO17 data set. The MOB-ML was trained on 220 randomly selected QM7b-T structures.

RMSD of 0.03 Å with respect to MP2 structures confirming this hypothesis. We can take this as an indication that we can systematically approach structures of any desired reference theory.

5.5 Conclusions

In this work, we have presented the derivation and implementation of the formally complete MOB-ML analytical nuclear gradient theory within a general Lagrangian framework. We have validated our derivation and implementation by comparison of numerical and analytical gradients. The MOB-ML gradient framework can be applied in conjunction with any desired fitting technique (e.g., Gaussian process regression or neural networks) and any desired recipe for assembling the MOB-ML feature information. Furthermore, the framework for evaluating the gradient of a predicted high-accuracy wave function energy is independent of the wave function method MOB-ML was trained to predict. Hence, we can take the analytical gradient of a MOB-ML method trained to predict an arbitrary accurate wave function theory.

MOB-ML was previously shown to predict high-accuracy wave function energies at the cost of a molecular orbital evaluation. We now have shown that a MOB-ML model trained on correlation energies alone also yields highly accurate gradients for potential energy surfaces of a single molecule and for sets of diverse molecules. Specifically, we presented a MOB-ML model which obtains a force MAE of 1.64 kcal/mol/Å for the ISO17 set when only trained on reference energies for 100 molecules beating out the next best model only trained on energies in the literature, SchNet (5.71 kcal/mol/Å) which was trained on 400k molecules [57]. The transferability and data efficiency becomes even clearer when considering that we obtain an MAE of 1.63 kcal/mol/Å for the ISO17 set when training on 220 QM7b-T molecules which are smaller in size (seven versus nine heavy atoms) and which are more diverse in terms of chemical composition. The accuracy of a MOB-ML model trained on energies for 220 QM7b-T molecules for the forces is on par with some of the best ML models trained on energies and forces for hundreds of thousands of ISO17 molecules. Furthermore, we have demonstrated that a force MAE of this magnitude translates into structures which are very close to reference structures. Specifically, we obtain a mean RMSD of 0.01 Å with respect to MP2 optimized structures for the ISO17 data set. This deviation is three times smaller than for HF or B3LYP-D3 optimized structures.

The presented method can be applied out-of-the-box whenever hybrid DFT methods

are applied while being systematically improvable and offering a controlled accuracy. Natural objectives for future work include (i) the inclusion of gradients in the training process to boost the performance in the very low data regime, (ii) the extension to an open-shell framework, (iii) the adaptation of the Lagrangian framework to derive the analytical gradients of the MOB-ML energy with respect to quantities such as electric and magnetic fields.

5.6 Supporting Information

The data set used in Fig. 5.1 is available from Ref. 207. Geometries used in all other tables and figures are available in the supporting information of Ref. 210.

5.7 Appendix

Foster-Boys Localization

This appendix provides additional details for the Boys-related terms in Eqs. 5.13, 5.22 and 5.36 of the main text. The localization conditions for Foster-Boys are [189]

$$r_{ij} = \sum_n R^n_{ij}\left(R^n_{ii} - R^n_{jj}\right) = 0 \text{ for all } i > j, \tag{5.37}$$

where n corresponds to the x, y, z-coordinates of the position operator. The matrices R^n are defined as

$$R^n_{ij} = \sum_\mu (i|n|j), \tag{5.38}$$

where $|i)$ and $|j)$ are valence-occupied MOs. The orbital derivative contributions from the Foster-Boys localization conditions shown in Eq. 5.13 are

$$\left(\mathbf{a}\left[\mathbf{z}^{\text{val-occ}}\right]\right)_{pq} = \sum_{i>j} \mathcal{B}_{pqij} z^{\text{loc}}_{ij}\Big|_{q\in\text{val-occ}}, \tag{5.39}$$

where

$$\mathcal{B}_{pqkl} = \sum_n \left[\left(2R^n_{pk}\delta_{kq} - 2R^n_{pl}\delta_{lq}\right)R^n_{kl} + \left(R^n_{kk} - R^n_{ll}\right)\left(R^n_{pl}\delta_{kq} + R^n_{pk}\delta_{lq}\right)\right]. \tag{5.40}$$

Next, the dipole derivative contribution from the localization conditions from the first term on the RHS of Eq. 5.36 is

$$(\mathbf{W}_n)_{\mu\nu} = \tfrac{1}{2}\sum_{ij} C_{\mu i}C_{\nu j} z^{\text{val-occ}}_{ij}\left(R^n_{ii} - R^n_{jj}\right) + \sum_i C_{\mu i}C_{\nu i}\sum_j z^{\text{val-occ}}_{ij} R^n_{ij}. \tag{5.41}$$

For a full derivation of the orbital and dipole derivatives of Foster-Boys localization conditions please refer to Ref. 189.

IBO Localization

This appendix provides additional details for the terms IBO-related terms in Eqs. 5.7, 5.14, 5.26 and 5.35 of the main text. The localization conditions for IBO are [211]

$$r_{ab} = 4 \sum_{A} Q^{A}_{ab} \left(\left(Q^{A}_{aa} \right)^{3} - \left(Q^{A}_{bb} \right)^{3} \right) = 0 \text{ for all } a > b, \tag{5.42}$$

where A corresponds to an atom in the molecule. The matrices Q^A are defined as

$$Q^{A}_{ab} = \sum_{\mu \in A} C^{\mathrm{IAO}}_{\mu a} C^{\mathrm{IAO}}_{\mu b}, \tag{5.43}$$

where the summation over μ is restricted to basis functions at atom A. The matrix $\mathbf{C}^{\mathrm{IAO}}$ is the MO coefficient matrix represented in the intrinsic atomic orbital (IAO) basis which is defined as

$$\mathbf{C}^{\mathrm{IAO}} = \mathbf{X}^{\mathrm{IAO},\dagger} \mathbf{S}_1 \mathbf{C}, \tag{5.44}$$

where $\mathbf{S}_1$ is the overlap matrix in the original atomic orbital (AO) basis and $\mathbf{C}$ is the MO coefficient matrix in the original AO basis. The matrix that transforms from the AOs to the IAOs shown in Eqs. 5.44 and 5.7 is

$$\mathbf{X}^{\mathrm{IAO}} = \bar{\mathbf{X}}^{\mathrm{IAO}} \left(\bar{\mathbf{X}}^{\mathrm{IAO},\dagger} \mathbf{S}_1 \bar{\mathbf{X}}^{\mathrm{IAO}} \right)^{-1/2}, \tag{5.45}$$

where

$$\bar{\mathbf{X}}^{\mathrm{IAO}} = \left(\mathbf{S}_1^{-1} + \mathbf{L}\mathbf{L}^{\dagger} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\dagger} \right) \mathbf{S}_{12}. \tag{5.46}$$

The matrix $\mathbf{L}$ is the subset of the MO coefficient matrix being localized. The matrix $\tilde{\mathbf{L}}$ is

$$\tilde{\mathbf{L}} = \bar{\mathbf{L}} \left(\bar{\mathbf{L}}^{\dagger} \mathbf{S}_1 \bar{\mathbf{L}} \right)^{-1/2}, \tag{5.47}$$

where

$$\bar{\mathbf{L}} = \mathbf{S}_1^{-1} \mathbf{S}_{12} \mathbf{S}_2^{-1} \mathbf{S}_{12}^{\dagger} \mathbf{L}. \tag{5.48}$$

The orbital derivative contributions from the IBO localization conditions shown in Eq. 5.14 corresponds to Eq. 60 in Ref. 211. The tensor C from Eq. 5.26 corresponds to Eq. 37 in Ref. 211. The overlap derivative contributions from the IBO localization conditions shown in Eq. 5.35 correspond to Eqs. 50 - 52 in Ref. 211.

Valence Virtual Conditions

This appendix provides additional details for the terms in Eqs. 5.7 and 5.35 of the main text. In Eq. 5.7 the matrix $\mathbf{X}^{\mathrm{IAO}}_{\mathrm{occ}}$ is calculated using Eqs. 5.45 - 5.48

where the matrix $\mathbf{L}$ corresponds to all occupied MOs. Next, the overlap derivative contributions from the valence virtual conditions in Eq. 5.35 is

$$\begin{aligned}\sum_{wa}\frac{\partial P_{wa}}{\partial \mathbf{S}_1}\lambda_{wa} &= \mathbf{X}_{\mathrm{occ}}^{\mathrm{IAO}}\mathbf{X}_{\mathrm{occ}}^{\mathrm{IAO},\dagger}\mathbf{S}_1\mathbf{C}_{\mathrm{vv}}\boldsymbol{\lambda}^{\dagger}\mathbf{C}_{\mathrm{nvv}}^{\dagger} \\ &+ \left(\mathbf{X}_{\mathrm{occ}}^{\mathrm{IAO}}\mathbf{X}_{\mathrm{occ}}^{\mathrm{IAO},\dagger}\mathbf{S}_1\mathbf{C}_{\mathrm{vv}}\boldsymbol{\lambda}^{\dagger}\mathbf{C}_{\mathrm{nvv}}^{\dagger}\right)^{\dagger} \\ &+ \tfrac{1}{2}\left(\tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\dagger}\bar{\mathbf{X}}\tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\dagger} - \mathbf{S}_1^{-1}\bar{\mathbf{X}}\mathbf{S}_1^{-1} - \tilde{\mathbf{X}}\right) - \bar{\mathbf{X}}^{\mathrm{IAO}}\mathbf{H}\bar{\mathbf{X}}^{\mathrm{IAO},\dagger}.\end{aligned} \tag{5.49}$$

The matrices $\bar{\mathbf{X}}$, $\tilde{\mathbf{X}}$, and $\mathbf{H}$ are the same as the Eqs. 56, 57 and 48, respectively, shown in Ref. 211. The evaluation of these matrices differ here by redefining the matrix $\mathbf{G}$ (Eq. 42 in Ref. 211), to be

$$\mathbf{G} = \mathbf{S}_1\mathbf{C}_{\mathrm{nvv}}\boldsymbol{\lambda}\mathbf{C}_{\mathrm{vv}}^{\dagger}\mathbf{S}_1\mathbf{X}_{\mathrm{occ}}^{\mathrm{IAO}}, \tag{5.50}$$

the matrix $\mathbf{B}$ to be

$$\mathbf{B} = \mathbf{X}_{\mathrm{occ}}^{\mathrm{IAO}}, \tag{5.51}$$

and the matrix $\mathbf{L}$ to span all occupied MOs. The overlap derivative contributions from the valence virtual conditions in Eq. 5.35 are

$$\sum_{wa}\frac{\partial P_{wa}}{\partial \mathbf{S}_2}\lambda_{wa} = -\tfrac{1}{2}\mathbf{S}_2^{-1}\mathbf{S}_{12}^{\dagger}\check{\mathbf{X}}\mathbf{S}_{12}\mathbf{S}_2^{-1}, \tag{5.52}$$

and

$$\sum_{wa}\frac{\partial P_{wa}}{\partial \mathbf{S}_{12}}\lambda_{wa} = \left(\mathbf{S}_1^{-1} + \mathbf{L}\mathbf{L}^{\dagger} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\dagger}\right)\mathbf{V} + \check{\mathbf{X}}\mathbf{S}_{12}\mathbf{S}_2^{-1}. \tag{5.53}$$

The matrices $\check{\mathbf{X}}$, and $\mathbf{V}$ are evaluated by Eqs. 58 and 54, respectively, in Ref. 211 with the same modifications to $\mathbf{G}$, $\mathbf{B}$ and $\mathbf{L}$.

Density Fitting Approximation

This appendix provides details on how the density fitting approximation can be used to approximate the four-center AO integral derivatives in Eq. 5.31. The AO integral derivatives are approximated by

$$\begin{aligned}(\mu\nu|\kappa\sigma)^{(q)} \approx (\mu\nu|\kappa\sigma)_{\mathrm{DF}}^{(q)} &= \sum_P (\mu\nu|P)^{(q)}c_{\kappa\sigma}^P + \sum_P c_{\mu\nu}^P(P|\kappa\sigma)^{(q)} \\ &- \sum_{PQ} c_{\mu\nu}^P J_{PQ}^{(q)} c_{\kappa\sigma}^Q,\end{aligned} \tag{5.54}$$

where P and Q label density fitting basis functions, $(\mu\nu|P)^{(q)}$ are three-center AO integrals, and J_{PQ} are two-center AO integrals. The matrix $\mathbf{c}^P$ is

$$c_{\kappa\sigma}^P = \sum_Q \left[\mathbf{J}^{-1}\right]_{PQ}(Q|\kappa\sigma). \tag{5.55}$$

Substituting Eq. 5.54 into Eq. 5.31 yields

$$\sum_{\mu\nu\kappa\sigma} (\mu\nu|\kappa\sigma)^{(q)}_{\mathrm{DF}} \sum_{pq} D_{\mu\nu\kappa\sigma} = 2 \sum_{P\mu\nu} (\mu\nu|P)^{(q)} \Lambda^P_{\mu\nu} - \sum_{PQ} J^{(q)}_{PQ} \Gamma_{PQ}, \tag{5.56}$$

where

$$c^P_{pq} = \sum_Q \left[\mathbf{J}^{-1}\right]_{PQ} (Q|pq), \tag{5.57}$$

$$\begin{aligned} \Lambda^P_{\mu\nu} = (\mathbf{d}_\mathrm{b})_{\mu\nu} \sum_{\kappa\sigma} \gamma_{\sigma\kappa} c^P_{\sigma\kappa} - \tfrac{1}{2} \sum_{\kappa\sigma} \gamma_{\sigma\mu} (\mathbf{d}_\mathrm{b})_{\mu\kappa} c^P_{\sigma\kappa} \\ + 2 \sum_q C_{\mu q} C_{\nu q} \sum_p c^P_{pp} D^\mathrm{J}_{pq} + 2 \sum_q C_{\nu q} \sum_p c^P_{pq} D^\mathrm{K}_{pq} C_{\mu p}, \end{aligned} \tag{5.58}$$

and

$$\begin{aligned} \Gamma_{PQ} = \sum_{\mu\nu} (\mathbf{d}_\mathrm{b})_{\mu\nu} c^P_{\mu\nu} \sum_{\kappa\sigma} \gamma_{\sigma\kappa} c^Q_{\sigma\kappa} - \tfrac{1}{2} \sum_{\mu\nu\kappa\sigma} (\mathbf{d}_\mathrm{b})_{\mu\kappa} \gamma_{\nu\sigma} c^P_{\mu\nu} c^Q_{\kappa\sigma} \\ + 2 \sum_{pq} c^P_{pp} c^Q_{qq} D^\mathrm{J}_{pq} + 2 \sum_{pq} c^P_{pq} c^Q_{pq} D^\mathrm{K}_{pq}. \end{aligned} \tag{5.59}$$

Feature derivatives

Tables 5.3–5.5 specify how the partial derivative of the diagonal feature vector with respect to the Fock matrix and two center molecular orbital integrals are computed. Tables 5.6–5.8 specify how the partial derivative of the off-diagonal feature vector with respect to the Fock matrix and two center molecular orbital integrals are computed.

Table 5.3: Partial derivative of the diagonal feature vector of FS 3 with respect to Fock matrix elements $\partial \mathbf{f}_i / \partial F_{pq}$.

Partial derivatives	Expression
$\partial F_{ii} / \partial F_{pq}$	$\delta_{pi}\delta_{qi}$
$\partial \mathbf{F}_{ik} / \partial F_{pq}$	$\delta_{pi}\delta_{qk} \frac{F_{ik}}{\lvert F_{ik}\rvert}$
$\partial \mathbf{F}_{aa} / \partial F_{pq}$	$\delta_{pa}\delta_{qa} \langle ii\rangle aa\rangle^3$
$\partial \mathbf{F}_{ab} / \partial F_{pq}$	$\langle aa\rangle bb\rangle \frac{\langle ii\rangle aa\rangle}{2(F_{aa}-F_{ii})} \frac{\langle ii\rangle bb\rangle}{2(F_{bb}-F_{ii})} \Big(\delta_{pa}\delta_{qb} \frac{F_{ab}}{\lvert F_{ab}\rvert} - \lvert F_{ab}\rvert \frac{(\delta_{pa}\delta_{qa}-\delta_{pi}\delta_{qi})}{(F_{aa}-F_{ii})}$ $-\lvert F_{ab}\rvert \frac{(\delta_{pb}\delta_{qb}-\delta_{pi}\delta_{qi})}{(F_{bb}-F_{ii})} \Big)$
$\partial [\boldsymbol{K}^{aa}]_{bb} / \partial F_{pq}$	$-\langle aa\rangle bb\rangle^4 \frac{\langle ii\rangle aa\rangle}{2(F_{aa}-F_{ii})} \frac{\langle ii\rangle bb\rangle}{2(F_{bb}-F_{ii})} \left(\frac{(\delta_{pa}\delta_{qa}-\delta_{pi}\delta_{qi})}{(F_{aa}-F_{ii})} + \frac{(\delta_{pb}\delta_{qb}-\delta_{pi}\delta_{qi})}{(F_{bb}-F_{ii})} \right)$
$\partial [\boldsymbol{K}^{ab}]_{ab} / \partial F_{pq}$	$-\langle ab\rangle ab\rangle^2 \frac{\langle ia\rangle ia\rangle}{2(F_{aa}-F_{ii})} \frac{\langle ib\rangle ib\rangle}{2(F_{bb}-F_{ii})} \left(\frac{(\delta_{pa}\delta_{qa}-\delta_{pi}\delta_{qi})}{(F_{aa}-F_{ii})} + \frac{(\delta_{pb}\delta_{qb}-\delta_{pi}\delta_{qi})}{(F_{bb}-F_{ii})} \right)$

For the off-diagonal feature vector, the damping function is defined as

$$G_{ij} = \frac{1}{1 + \frac{1}{6}(R_{ij}/R_0)^6} \tag{5.60}$$

Table 5.4: Partial derivative of the diagonal feature vector of FS 3 with respect to Coulomb-type two-center molecular orbital integrals $\partial \mathbf{f}_i / \partial [\boldsymbol{K}^{pp}]_{qq}$.

Partial derivatives	Expression
$\partial \mathbf{F}_{aa} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3\delta_{pi}\delta_{qa} F_{aa} \langle ii \vert aa \rangle^2$
$\partial \mathbf{F}_{ab} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$\vert F_{ab} \vert \Big(\frac{\delta_{pi}\delta_{qa}}{2(F_{aa}-F_{ii})} \langle aa \vert bb \rangle \frac{\langle ii \vert bb \rangle}{2(F_{bb}-F_{ii})}$
	$+ \frac{\langle ii \vert aa \rangle}{2(F_{aa}-F_{ii})} \delta_{pa}\delta_{qb} \frac{\langle ii \vert bb \rangle}{2(F_{bb}-F_{ii})} + \frac{\langle ii \vert aa \rangle}{2(F_{aa}-F_{ii})} \langle aa \vert bb \rangle \frac{\delta_{pi}\delta_{qb}}{2(F_{bb}-F_{ii})} \Big)$
$\partial [\boldsymbol{K}^{ii}]_{ii} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3\delta_{pi}\delta_{qi} \langle ii \vert ii \rangle^2$
$\partial [\boldsymbol{K}^{ii}]_{kk} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3\delta_{pi}\delta_{qk} \langle ii \vert kk \rangle^2$
$\partial [\boldsymbol{K}^{ii}]_{aa} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3\delta_{pi}\delta_{qa} \langle ii \vert aa \rangle^2$
$\partial [\boldsymbol{K}^{aa}]_{aa} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3\delta_{pa}\delta_{qa} \langle aa \vert aa \rangle^2 \langle ii \vert aa \rangle^3 + 3\delta_{pi}\delta_{qa} \langle aa \vert aa \rangle^3 \langle ii \vert aa \rangle^2$
$\partial [\boldsymbol{K}^{aa}]_{bb} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$4\delta_{pa}\delta_{qb} \langle aa \vert bb \rangle^3 \frac{\langle ii \vert aa \rangle}{2(F_{aa}-F_{ii})} \frac{\langle ii \vert bb \rangle}{2(F_{bb}-F_{ii})} + \langle aa \vert bb \rangle^4 \frac{\delta_{pi}\delta_{qa}}{2(F_{aa}-F_{ii})} \frac{\langle ii \vert bb \rangle}{2(F_{bb}-F_{ii})}$
	$+ \langle aa \vert bb \rangle^4 \frac{\langle ii \vert aa \rangle}{2(F_{aa}-F_{ii})} \frac{\delta_{pi}\delta_{qb}}{2(F_{bb}-F_{ii})}$

Table 5.5: Partial derivative of the diagonal feature vector of FS 3 with respect to exchange-type two-center molecular orbital integrals $\partial \mathbf{f}_i / \partial [\boldsymbol{K}^{pq}]_{pq}$.

Partial derivatives	Expression
$\partial [\boldsymbol{K}^{ik}]_{ik} / \partial [\boldsymbol{K}^{pq}]_{pq}$	$\delta_{pi}\delta_{qk}$
$\partial [\boldsymbol{K}^{ia}]_{ia} / \partial [\boldsymbol{K}^{pq}]_{pq}$	$\delta_{pi}\delta_{qa}$
$\partial [\boldsymbol{K}^{ab}]_{ab} / \partial [\boldsymbol{K}^{pq}]_{pq}$	$2\delta_{pa}\delta_{qb} \langle ab \vert ab \rangle \frac{\langle ia \vert ia \rangle}{2(F_{aa}-F_{ii})} \frac{\langle ib \vert ib \rangle}{2(F_{bb}-F_{ii})}$
	$+ \langle ab \vert ab \rangle^2 \frac{\delta_{pi}\delta_{qa}}{2(F_{aa}-F_{ii})} \frac{\langle ib \vert ib \rangle}{2(F_{bb}-F_{ii})} + \langle ab \vert ab \rangle^2 \frac{\langle ia \vert ia \rangle}{2(F_{aa}-F_{ii})} \frac{\delta_{pi}\delta_{qb}}{2(F_{bb}-F_{ii})}$

Table 5.6: Partial derivative of the off-diagonal feature vector of FS 3 with respect to Fock matrix elements $\partial \mathbf{f}_i / \partial F_{pq}$.

Partial derivatives	Expression
$\partial F_{\tilde{i}\tilde{i}} / \partial F_{pq}$	$G_{ij}\left(\frac{1}{2}\delta_{pi}\delta_{qi} + \frac{1}{2}\delta_{pj}\delta_{qj} + \delta_{pi}\delta_{qj}\frac{F_{ij}}{\lvert F_{ij}\rvert}\right)$
$\partial F_{\tilde{i}\tilde{j}} / \partial F_{pq}$	$G_{ij}\left(\frac{1}{2}\delta_{pi}\delta_{qi} - \frac{1}{2}\delta_{pj}\delta_{qj}\right)\frac{F_{ii}-F_{jj}}{\lvert F_{ii}-F_{jj}\rvert}$
$\partial F_{\tilde{j}\tilde{j}} / \partial F_{pq}$	$G_{ij}\left(\frac{1}{2}\delta_{pi}\delta_{qi} + \frac{1}{2}\delta_{pj}\delta_{qj} - \delta_{pi}\delta_{qj}\frac{F_{ij}}{\lvert F_{ij}\rvert}\right)$
$\partial \mathbf{F}_{\tilde{i}k} / \partial F_{pq}$	$G_{ij}\left(\frac{1}{\sqrt{2}}\delta_{pi}\delta_{qk}\frac{F_{ik}}{\lvert F_{ik}\rvert} + \frac{1}{\sqrt{2}}\delta_{pj}\delta_{qk}\frac{F_{jk}}{\lvert F_{jk}\rvert}\right)$
$\partial \mathbf{F}_{\tilde{j}k} / \partial F_{pq}$	$G_{ij}\left(\frac{1}{\sqrt{2}}\delta_{pi}\delta_{qk}\frac{F_{ik}}{\lvert F_{ik}\rvert} - \frac{1}{\sqrt{2}}\delta_{pj}\delta_{qk}\frac{F_{jk}}{\lvert F_{jk}\rvert}\right)\frac{F_{ik}-F_{jk}}{\lvert F_{ik}-F_{jk}\rvert}$
$\partial \mathbf{F}_{aa} / \partial F_{pq}$	$G_{ij}\left(\delta_{pa}\delta_{qa}\lvert \langle ii\lvert aa\rangle^3 - \langle jj\lvert aa\rangle^3 \rvert/2\right)$
$\partial \mathbf{F}_{ab} / \partial F_{pq}$	$G_{ij}\,\langle aa\lvert bb\rangle \frac{\frac{1}{2}\lvert\langle ii\lvert aa\rangle - \langle jj\lvert aa\rangle\rvert}{2F_{aa}-F_{ii}-F_{jj}} \frac{\frac{1}{2}\lvert\langle ii\lvert bb\rangle - \langle jj\lvert bb\rangle\rvert}{2F_{bb}-F_{ii}-F_{jj}}\left(\delta_{pa}\delta_{qb}\frac{F_{ab}}{\lvert F_{ab}\rvert}\right.$ $\left. -\lvert F_{ab}\rvert\frac{2\delta_{pa}\delta_{qa}-\delta_{pi}\delta_{qi}-\delta_{pj}\delta_{qj}}{2F_{aa}-F_{ii}-F_{jj}} - \lvert F_{ab}\rvert\frac{2\delta_{pb}\delta_{qb}-\delta_{pi}\delta_{qi}-\delta_{pj}\delta_{qj}}{2F_{bb}-F_{ii}-F_{jj}}\right)$
$\partial [\boldsymbol{K}^{aa}]_{bb} / \partial F_{pq}$	$-G_{ij}\,\langle aa\lvert bb\rangle^4 \frac{\frac{1}{2}\lvert\langle ii\lvert aa\rangle - \langle jj\lvert aa\rangle\rvert}{2F_{aa}-F_{ii}-F_{jj}} \frac{\frac{1}{2}\lvert\langle ii\lvert bb\rangle - \langle jj\lvert bb\rangle\rvert}{2F_{bb}-F_{ii}-F_{jj}}\left(\frac{2\delta_{pa}\delta_{qa}-\delta_{pi}\delta_{qi}-\delta_{pj}\delta_{qj}}{2F_{aa}-F_{ii}-F_{jj}}\right.$ $\left. +\frac{2\delta_{pb}\delta_{qb}-\delta_{pi}\delta_{qi}-\delta_{pj}\delta_{qj}}{2F_{bb}-F_{ii}-F_{jj}}\right)$
$\partial [\boldsymbol{K}^{ab}]_{ab} / \partial F_{pq}$	$-G_{ij}\,\langle ab\lvert ab\rangle^2 \frac{\frac{1}{2}\lvert\langle ia\lvert ia\rangle - \langle ja\lvert ja\rangle\rvert}{2F_{aa}-F_{ii}-F_{jj}} \frac{\frac{1}{2}\lvert\langle ib\lvert ib\rangle - \langle jb\lvert jb\rangle\rvert}{2F_{bb}-F_{ii}-F_{jj}}\left(\frac{2\delta_{pa}\delta_{qa}-\delta_{pi}\delta_{qi}-\delta_{pj}\delta_{qj}}{2F_{aa}-F_{ii}-F_{jj}}\right.$ $\left. +\frac{2\delta_{pb}\delta_{qb}-\delta_{pi}\delta_{qi}-\delta_{pj}\delta_{qj}}{2F_{bb}-F_{ii}-F_{jj}}\right)$

Table 5.7: Partial derivative of the off-diagonal feature vector of FS 3 with respect to Coulomb-type two-center molecular orbital integrals $\partial \mathbf{f}_i / \partial [\boldsymbol{K}^{pp}]_{qq}$.

Partial derivatives	Expression
$\partial \mathbf{F}_{aa} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3G_{ij}F_{aa}\left(\frac{1}{2}\delta_{pi}\delta_{qa}\langle ii\|aa\rangle^2 - \frac{1}{2}\delta_{pj}\delta_{qa}\langle jj\|aa\rangle^2\right)\frac{\langle ii\|aa\rangle^3 - \langle jj\|aa\rangle^3}{\|\langle ii\|aa\rangle^3 - \langle jj\|aa\rangle^3\|}$
$\partial \mathbf{F}_{ab} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$G_{ij}\frac{\frac{1}{2}\|\langle ii\|aa\rangle - \langle jj\|aa\rangle\|}{2F_{aa}-F_{ii}-F_{jj}}\frac{\frac{1}{2}\|\langle ii\|bb\rangle - \langle jj\|bb\rangle\|}{2F_{bb}-F_{ii}-F_{jj}}\|F_{ab}\|\Big(\delta_{pa}\delta_{qb}$ $+\frac{(\delta_{pi}\delta_{qa}-\delta_{pj}\delta_{qa})(\langle ii\|aa\rangle - \langle jj\|aa\rangle)}{\|\langle ii\|aa\rangle - \langle jj\|aa\rangle\|^2}\langle aa\|bb\rangle$ $+\langle aa\|bb\rangle\frac{(\delta_{pi}\delta_{qb}-\delta_{pj}\delta_{qb})(\langle ii\|bb\rangle - \langle jj\|bb\rangle)}{\|\langle ii\|bb\rangle - \langle jj\|bb\rangle\|^2}\Big)$
$\partial [\boldsymbol{K}^{\tilde{i}\tilde{i}}]_{\tilde{i}\tilde{i}} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3G_{ij}\left(\frac{1}{2}\delta_{pi}\delta_{qi}\langle ii\|ii\rangle^2 + \frac{1}{2}\delta_{pj}\delta_{qj}\langle jj\|jj\rangle^2 + \delta_{pi}\delta_{qj}\langle ii\|jj\rangle^2\right)$
$\partial [\boldsymbol{K}^{\tilde{i}\tilde{i}}]_{\tilde{j}\tilde{j}} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3G_{ij}\left(\frac{1}{2}\delta_{pi}\delta_{qi}\langle ii\|ii\rangle^2 - \frac{1}{2}\delta_{pj}\delta_{qj}\langle jj\|jj\rangle^2\right)\frac{\langle ii\|ii\rangle^3 - \langle jj\|jj\rangle^3}{\|\langle ii\|ii\rangle^3 - \langle jj\|jj\rangle^3\|}$
$\partial [\boldsymbol{K}^{\tilde{j}\tilde{j}}]_{\tilde{j}\tilde{j}} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3G_{ij}\left(\frac{1}{2}\delta_{pi}\delta_{qi}\langle ii\|ii\rangle^2 + \frac{1}{2}\delta_{pj}\delta_{qj}\langle jj\|jj\rangle^2 - \delta_{pi}\delta_{qj}\langle ii\|jj\rangle^2\right)$
$\partial [\boldsymbol{K}^{\tilde{i}\tilde{i}}]_{kk} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3G_{ij}\left(\frac{1}{\sqrt{2}}\delta_{pi}\delta_{qk}\langle ii\|kk\rangle^2 + \frac{1}{\sqrt{2}}\delta_{pj}\delta_{qk}\langle jj\|kk\rangle^2\right)$
$\partial [\boldsymbol{K}^{\tilde{j}\tilde{j}}]_{kk} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3G_{ij}\left(\frac{1}{\sqrt{2}}\delta_{pi}\delta_{qk}\langle ii\|kk\rangle^2 - \frac{1}{\sqrt{2}}\delta_{pj}\delta_{qk}\langle jj\|kk\rangle^2\right)\frac{\langle ii\|kk\rangle^3 - \langle jj\|kk\rangle^3}{\|\langle ii\|kk\rangle^3 - \langle jj\|kk\rangle^3\|}$
$\partial [\boldsymbol{K}^{\tilde{i}\tilde{i}}]_{aa} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3G_{ij}\left(\frac{1}{\sqrt{2}}\delta_{pi}\delta_{qa}\langle ii\|aa\rangle^2 + \frac{1}{\sqrt{2}}\delta_{pj}\delta_{qa}\langle jj\|aa\rangle^2\right)$
$\partial [\boldsymbol{K}^{\tilde{j}\tilde{j}}]_{aa} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3G_{ij}\left(\frac{1}{\sqrt{2}}\delta_{pi}\delta_{qa}\langle ii\|aa\rangle^2 - \frac{1}{\sqrt{2}}\delta_{pj}\delta_{qa}\langle jj\|aa\rangle^2\right)\frac{\langle ii\|aa\rangle^3 - \langle jj\|aa\rangle^3}{\|\langle ii\|aa\rangle^3 - \langle jj\|aa\rangle^3\|}$
$\partial [\boldsymbol{K}^{aa}]_{aa} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$3G_{ij}\Big(\delta_{pa}\delta_{qa}\langle aa\|aa\rangle^2\frac{1}{2}\|\langle ii\|aa\rangle^3 - \langle jj\|aa\rangle^3\|$ $+\langle aa\|aa\rangle^3\frac{1}{2}(\delta_{pi}\delta_{qa}\langle ii\|aa\rangle^2 - \delta_{pj}\delta_{qa}\langle jj\|aa\rangle^2)\frac{\langle ii\|aa\rangle^3 - \langle jj\|aa\rangle^3}{\|\langle ii\|aa\rangle^3 - \langle jj\|aa\rangle^3\|}\Big)$
$\partial [\boldsymbol{K}^{aa}]_{bb} / \partial [\boldsymbol{K}^{pp}]_{qq}$	$G_{ij}\frac{\frac{1}{2}\|\langle ii\|aa\rangle - \langle jj\|aa\rangle\|}{2F_{aa}-F_{ii}-F_{jj}}\frac{\frac{1}{2}\|\langle ii\|bb\rangle - \langle jj\|bb\rangle\|}{2F_{bb}-F_{ii}-F_{jj}}\Big(4\langle aa\|bb\rangle^3\delta_{pa}\delta_{qb}$ $+\langle aa\|bb\rangle^4\frac{(\delta_{pi}\delta_{qa}-\delta_{pj}\delta_{qa})(\langle ii\|aa\rangle - \langle jj\|aa\rangle)}{\|\langle ii\|aa\rangle - \langle jj\|aa\rangle\|^2}$ $+\langle aa\|bb\rangle^4\frac{(\delta_{pi}\delta_{qb}-\delta_{pj}\delta_{qb})(\langle ii\|bb\rangle - \langle jj\|bb\rangle)}{\|\langle ii\|bb\rangle - \langle jj\|bb\rangle\|^2}\Big)$

Table 5.8: Partial derivative of the off-diagonal feature vector of FS 3 with respect to exchange-type two-center molecular orbital integrals $\partial \mathbf{f}_i / \partial [\boldsymbol{K}^{pq}]_{pq}$.

Partial derivatives	Expression
$\partial [\boldsymbol{K}^{\tilde{ij}}]_{\tilde{ij}} / \partial [\boldsymbol{K}^{pq}]_{pq}$	$G_{ij} \left(\frac{1}{2}\delta_{pi}\delta_{qi} - \frac{1}{2}\delta_{pj}\delta_{qj} \right) \frac{\langle ii\vert ii\rangle - \langle jj\vert jj\rangle}{\vert\langle ii\vert ii\rangle - \langle jj\vert jj\rangle\vert}$
$\partial [\boldsymbol{K}^{\tilde{ik}}]_{\tilde{ik}} / \partial [\boldsymbol{K}^{pq}]_{pq}$	$G_{ij} \left(\frac{1}{\sqrt{2}}\delta_{pi}\delta_{qk} + \frac{1}{\sqrt{2}}\delta_{pj}\delta_{qk} \right)$
$\partial [\boldsymbol{K}^{\tilde{jk}}]_{\tilde{jk}} / \partial [\boldsymbol{K}^{pq}]_{pq}$	$G_{ij} \left(\frac{1}{\sqrt{2}}\delta_{pi}\delta_{qk} - \frac{1}{\sqrt{2}}\delta_{pj}\delta_{qk} \right) \frac{\langle ik\vert ik\rangle - \langle jk\vert jk\rangle}{\vert\langle ik\vert ik\rangle - \langle jk\vert jk\rangle\vert}$
$\partial [\boldsymbol{K}^{\tilde{ia}}]_{\tilde{ia}} / \partial [\boldsymbol{K}^{pq}]_{pq}$	$G_{ij} \left(\frac{1}{\sqrt{2}}\delta_{pi}\delta_{qa} + \frac{1}{\sqrt{2}}\delta_{pj}\delta_{qa} \right)$
$\partial [\boldsymbol{K}^{\tilde{ja}}]_{\tilde{ja}} / \partial [\boldsymbol{K}^{pq}]_{pq}$	$G_{ij} \left(\frac{1}{\sqrt{2}}\delta_{pi}\delta_{qa} - \frac{1}{\sqrt{2}}\delta_{pj}\delta_{qa} \right) \frac{\langle ia\vert ia\rangle - \langle ja\vert ja\rangle}{\vert\langle ia\vert ia\rangle - \langle ja\vert ja\rangle\vert}$
$\partial [\boldsymbol{K}^{ab}]_{ab} / \partial [\boldsymbol{K}^{pq}]_{pq}$	$G_{ij} \frac{\frac{1}{2}\vert\langle ia\vert ia\rangle - \langle ja\vert ja\rangle\vert}{2F_{aa} - F_{ii} - F_{jj}} \frac{\frac{1}{2}\vert\langle ib\vert ib\rangle - \langle jb\vert jb\rangle\vert}{2F_{bb} - F_{ii} - F_{jj}} \Big(2 \langle ab\vert ab\rangle \, \delta_{pa}\delta_{qb}$
	$+ \langle ab\vert ab\rangle^2 \frac{(\delta_{pi}\delta_{qa} - \delta_{pj}\delta_{qa})(\langle ia\vert ia\rangle - \langle ja\vert ja\rangle)}{\vert\langle ia\vert ia\rangle - \langle ja\vert ja\rangle\vert^2}$
	$+ \langle ab\vert ab\rangle^2 \frac{(\delta_{pi}\delta_{qb} - \delta_{pj}\delta_{qb})(\langle ib\vert ib\rangle - \langle jb\vert jb\rangle)}{\vert\langle ib\vert ib\rangle - \langle jb\vert jb\rangle\vert^2} \Big)$

Table 5.9: Partial derivative of the off-diagonal feature vector of FS 3 with respect to the centroid distance between orbital p and orbital q $\partial \mathbf{f}_{ij} / \partial R^n_{pq}$.

Partial derivatives	Expression
$\partial \mathbf{f}_{ij} / \partial R^n_{pq}$	$-\frac{R^4_{ij}}{R^6_0} G_{ij} \mathbf{f}_{ij} (R^n_{ii} - R^n_{jj})(\delta_{pi}\delta_{qi} - \delta_{pj}\delta_{qj})$

BIBLIOGRAPHY

[1]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, "Fragment molecular orbital method: an approximate computational method for large molecules", Chem. Phys. Lett. **313**, 701–706 (1999) 10.1016/S0009-2614(99)00874-X.

[2]V. Deev, and M. A. Collins, "Approximate ab initio energies by systematic molecular fragmentation", J. Chem. Phys. **122**, 154102 (2005) 10.1063/1.1879792.

[3]M. A. Collins, and V. A. Deev, "Accuracy and efficiency of electronic energies from systematic molecular fragmentation", J. Chem. Phys. **125** (2006) 10.1063/1.2347710.

[4]D. G. Fedorov, and K. Kitaura, "Extending the power of quantum chemistry to large systems with the fragment molecular orbital method", J. Phys. Chem. A **111**, 6904–6914 (2007) 10.1021/jp0716740.

[5]P. Elliott, M. H. Cohen, A. Wasserman, and K. Burke, "Density functional partition theory with fractional occupations", J. Chem. Theory Comput. **5**, 827–833 (2009) 10.1021/ct9000119.

[6]J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller III, "Exact nonadditive kinetic potentials for embedded density functional theory.", J. Chem. Phys. **133**, 084103 (2010) 10.1063/1.3474575.

[7]C. Huang, and E. A. Carter, "Potential-functional embedding theory for molecules and materials", J. Chem. Phys. **135**, 194104 (2011) 10.1063/1.3659293.

[8]J. D. Goodpaster, T. A. Barnes, and T. F. Miller III, "Embedded density functional theory for covalently bonded and strongly interacting subsystems.", J. Chem. Phys. **134**, 164108 (2011) 10.1063/1.3582913.

[9]F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller III, "A Simple, Exact Density-Functional-Theory Embedding Scheme", J. Chem. Theory Comput. **8**, 2564–2568 (2012) 10.1021/ct300544e.

[10]J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, "Density functional theory embedding for correlated wavefunctions: improved methods for open-shell systems and transition metal complexes.", J. Chem. Phys. **137**, 224113 (2012) 10.1063/1.4770226.

[11]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, "Fragmentation methods: a route to accurate calculations on large systems", Chem. Rev. **112**, 632–672 (2012) 10.1021/cr200093j.

[12]G. Knizia, and G. K.-L. Chan, "Density matrix embedding: a simple alternative to dynamical mean-field theory", Phys. Rev. Lett. **109**, 186404 (2012) 10.1103/PhysRevLett.109.186404.

[13]T. A. Barnes, J. D. Goodpaster, F. R. Manby, and T. F. Miller III, "Accurate basis set truncation for wavefunction embedding", J. Chem. Phys. **139**, 024103 (2013) 10.1063/1.4811112.

[14]J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, "Accurate and systematically improvable density functional theory embedding for correlated wavefunctions", J. Chem. Phys. **140**, 18A507 (2014) 10.1063/1.4864040.

[15]D. Neuhauser, R. Baer, and E. Rabani, "Communication: embedded fragment stochastic density functional theory", J. Chem. Phys. **141**, 041102 (2014) 10.1063/1.4890651.

[16]T. A. Barnes, J. W. Kaminski, O. Borodin, and T. F. Miller III, "Ab Initio Characterization of the Electrochemical Stability and Solvation Properties of Condensed-Phase Ethylene Carbonate and Dimethyl Carbonate Mixtures", J. Phys. Chem. C **119**, 3865–3880 (2015) 10.1021/jp510882g.

[17]M. E. Fornace, J. Lee, K. Miyamoto, F. R. Manby, and T. F. Miller III, "Embedded Mean-Field Theory", J. Chem. Theory Comput. **11**, 568–580 (2015) 10.1021/ct5011032.

[18]S. J. Bennie, M. Stella, T. F. Miller III, and F. R. Manby, "Accelerating wavefunction in density-functional-theory embedding by truncating the active basis set", J. Chem. Phys. **143**, 024105 (2015) 10.1063/1.4923367.

[19]M. Stella, S. J. Bennie, and F. R. Manby, "Computational study of adsorption of cobalt on benzene and coronene", Mol. Phys. **113**, 1858–1864 (2015) 10.1080/00268976.2015.1018359.

[20]P. Huo, C. Uyeda, J. D. Goodpaster, J. C. Peters, and T. F. Miller III, "Breaking the Correlation between Energy Costs and Kinetic Barriers in Hydrogen Evolution via a Cobalt Pyridine-Diimine-Dioxime Catalyst", ACS Catal. **6**, 6114–6123 (2016) 10.1021/acscatal.6b01387.

[21]S. J. Bennie, M. W. van der Kamp, R. C. Pennifold, M. Stella, F. R. Manby, and A. J. Mulholland, "A projector embedding approach for multiscale coupled-cluster calculations applied to citrate synthase.", J. Chem. Theory Comput. **12**, 2689–2697 (2016) 10.1021/acs.jctc.6b00285.

[22]R. C. R. Pennifold, S. J. Bennie, T. F. Miller III, and F. R. Manby, "Correcting density-driven errors in projection-based embedding", J. Chem. Phys. **146**, 084113 (2017) 10.1063/1.4974929.

[23]F. Ding, F. R. Manby, and T. F. Miller III, "Embedded Mean-Field Theory with Block-Orthogonalized Partitioning", J. Chem. Theory Comput. **13**, 1605–1615 (2017) 10.1021/acs.jctc.6b01065.

[24]X. Zhang, S. J. Bennie, M. W. van der Kamp, D. R. Glowacki, F. R. Manby, and A. J. Mulholland, "Multiscale analysis of enantioselectivity in enzyme-catalysed 'lethal synthesis' using projector-based embedding", R. Soc. Open Sci. **5**, 171390 (2018) 10.1098/rsos.171390.

[25]A. Chapovetsky, M. Welborn, J. M. Luna, R. Haiges, T. F. Miller III, and S. C. Marinescu, "Pendant hydrogen-bond donors in cobalt catalysts independently enhance CO_2 reduction", ACS Cent. Sci. **4**, 397–404 (2018) 10.1021/acscentsci.7b00607.

[26]A. H. Mühlbach, and M. Reiher, "Quantum system partitioning at the single-particle level", J. Chem. Phys. **149**, 184104 (2018) 10.1063/1.5055942.

[27]S. J. R. Lee, M. Welborn, F. R. Manby, and T. F. Miller III, "Projection-Based Wavefunction-in-DFT Embedding", Acc. Chem. Res. **52**, 1359–1368 (2019) 10.1021/acs.accounts.8b00672.

[28]S. J. R. Lee, F. Ding, F. R. Manby, and T. F. Miller III, "Analytical gradients for projection-based wavefunction-in-DFT embedding", en, J. Chem. Phys. **151**, 064112 (2019) 10.1063/1.5109882.

[29]K. Miyamoto, T. F. Miller III, and F. R. Manby, "Fock-Matrix Corrections in Density Functional Theory and Use in Embedded Mean-Field Theory", J. Chem. Theory Comput. **12**, 5811–5822 (2016) 10.1021/acs.jctc.6b00685.

[30]D. B. Chesnut, and K. D. Moore, "Locally dense basis sets for chemical shift calculations", J. Comput. Chem. **10**, 648–659 (1989) 10.1002/jcc.540100507.

[31]D. B. Chesnut, B. E. Rusiloski, K. D. Moore, and D. A. Egolf, "Use of locally dense basis sets for nuclear magnetic resonance shielding calculations", J. Comput. Chem. **14**, 1364–1375 (1993) 10.1002/jcc.540141113.

[32]D. B. Chesnut, and E. F. C. Byrd, "The use of locally dense basis sets in correlated NMR chemical shielding calculations", Chem. Phys. **213**, 153–158 (1996) 10.1016/S0301-0104(96)00281-9.

[33]J. F. Hinton, P. Guthrie, P. Pulay, and K. Wolinski, "Ab initio quantum mechanical calculation of the chemical shift anisotropy of the hydrogen atom in the (H2O)17 water cluster", J. Am. Chem. Soc. **114**, 1604–1605 (1992) 10.1021/ja00031a010.

[34]R. A. Kirby, and A. E. Hansen, "Study of locally dense and locally saturated basis sets in localized molecular orbital calculations of nuclear shielding: Ab initio LORG calculations for13C and17O in norbornenone", Int. J. Quantum Chem. **57**, 199–205 (1996) 10.1002/(SICI)1097-461X(1996)57:2<199::AID-QUA5>3.0.CO;2-T.

[35]S. Moon, and D. A. Case, "A comparison of quantum chemical models for calculating NMR shielding parameters in peptides: Mixed basis set and ONIOM methods combined with a complete basis set extrapolation", J. Comput. Chem. **27**, 825–836 (2006) 10.1002/jcc.20388.

[36]D. M. Reid, R. Kobayashi, and M. A. Collins, "Systematic Study of Locally Dense Basis Sets for NMR Shielding Constants", J. Chem. Theory Comput. **10**, 146–152 (2014) 10.1021/ct4007579.

[37]S. T. Holmes, R. J. Iuliucci, K. T. Mueller, and C. Dybowski, "Density functional investigation of intermolecular effects on 13 C NMR chemical-shielding tensors modeled with molecular clusters", J. Chem. Phys. **141**, 164121 (2014) 10.1063/1.4900158.

[38]J. D. Hartman, S. Monaco, B. Schatschneider, and G. J. O. Beran, "Fragment-based 13 C nuclear magnetic resonance chemical shift predictions in molecular crystals: An alternative to planewave methods", J. Chem. Phys. **143**, 102809 (2015) 10.1063/1.4922649.

[39]C. W. Bauschlicher, and P. S. Bagus, "Mixed basis set calculations for atomic hydrogen on beryllium(0001)", Chem. Phys. Lett. **90**, 355–358 (1982) 10.1016/0009-2614(82)83068-6.

[40]J. Wright, C. Rowley, and L. Chepelev, "A 'universal' B3LYP-based method for gas-phase molecular properties: bond dissociation enthalpy, ionization potential, electron and proton affinity and gas-phase acidity", Mol. Phys. **103**, 815–823 (2005) 10.1080/00268970412331333429.

[41]G. A. DiLabio, and J. S. Wright, "Calculation of bond dissociation energies for large molecules using locally dense basis sets", Chem. Phys. Lett. **297**, 181–186 (1998) 10.1016/S0009-2614(98)01146-4.

[42]G. A. DiLabio, "Using Locally Dense Basis Sets for the Determination of Molecular Properties", J. Phys. Chem. A **103**, 11414–11424 (1999) 10.1021/jp992492b.

[43]M. Orozco, and F. Luque, "On the use of mixed basis sets to compute accurate molecular electrostatic potentials", Chem. Phys. Lett. **160**, 305–310 (1989) 10.1016/0009-2614(89)87601-8.

[44]G. E. Scuseria, and T. J. Lee, "Comparison of coupled-cluster methods which include the effects of connected triple excitations", J. Chem. Phys. **93**, 5851–5855 (1990) 10.1063/1.459684.

[45]J. Olsen, P. Jørgensen, and J. Simons, "Passing the one-billion limit in full configuration-interaction (FCI) calculations", Chem. Phys. Lett. **169**, 463–472 (1990) 10.1016/0009-2614(90)85633-N.

[46]A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons", Phys. Rev. Lett. **104**, 136403 (2010) 10.1103/PhysRevLett.104.136403.

[47]M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning", Phys. Rev. Lett. **108**, 058301 (2012) 10.1103/PhysRevLett.108.058301.

[48]A. Lavecchia, "Machine-learning approaches in drug discovery: methods and applications", Drug Discov. Today **20**, 318–331 (2015) 10.1016/j.drudis.2014.10.012.

[49]E. Gawehn, J. A. Hiss, and G. Schneider, "Deep Learning in Drug Discovery", Mol. Inf. **35**, 3–14 (2016) 10.1002/minf.201501008.

[50]P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, "Machine-learning-assisted materials discovery using failed experiments", Nature **533**, 73–76 (2016) 10.1038/nature17439.

[51]J. N. Wei, D. Duvenaud, and A. Aspuru-Guzik, "Neural Networks for the Prediction of Organic Chemistry Reactions", ACS Cent. Sci. **2**, 725–732 (2016) 10.1021/acscentsci.6b00219.

[52]J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost", Chem. Sci. **8**, 3192–3203 (2017) 10.1039/C6SC05720A.

[53]S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, "Machine learning of accurate energy-conserving molecular force fields", Sci. Adv. **3**, Publisher: American Association for the Advancement of Science Section: Research Article, e1603015 (2017) 10.1126/sciadv.1603015.

[54]E. Kim, K. Huang, S. Jegelka, and E. Olivetti, "Virtual screening of inorganic materials synthesis parameters with deep learning", Npj Comput. Mater. **3**, Number: 1 Publisher: Nature Publishing Group, 1–9 (2017) 10.1038/s41524-017-0055-6.

[55]Z. W. Ulissi, A. J. Medford, T. Bligaard, and J. K. Nørskov, "To address surface reaction network complexity using scaling relations machine learning and DFT calculations", Nat. Commun. **8**, 14621 (2017) 10.1038/ncomms14621.

[56]M. H. S. Segler, and M. P. Waller, "Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction", Chem. Eur. J. **23**, 5966–5971 (2017) 10.1002/chem.201605499.

[57]K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, *Schnet: a continuous-filter convolutional neural network for modeling quantum interactions*, 2017.

[58]K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science", Nature **559**, 547–555 (2018) 10.1038/s41586-018-0337-2.

[59]N. Lubbers, J. S. Smith, and K. Barros, "Hierarchical modeling of molecular energies using a deep neural network", J. Chem. Phys. **148**, 241715 (2018) 10.1063/1.5011181.

[60]M. Popova, O. Isayev, and A. Tropsha, "Deep reinforcement learning for de novo drug design", Sci. Adv. **4**, eaap7885 (2018) 10.1126/sciadv.aap7885.

[61]S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, "Towards exact molecular dynamics simulations with machine-learned force fields", Nat. Commun. **9**, Number: 1 Publisher: Nature Publishing Group, 3887 (2018) 10.1038/s41467-018-06169-2.

[62]J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, "Less is more: Sampling chemical space with active learning", J. Chem. Phys. **148**, 241733 (2018) 10.1063/1.5023802.

[63]J. S Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. Roitberg, *Outsmarting quantum chemistry through transfer learning*, Publisher: ChemRxiv, July 2018, 10.26434/chemrxiv.6744440.v1.

[64]A. S. Christensen, F. A. Faber, and O. A. von Lilienfeld, "Operators in quantum machine learning: Response properties in chemical space", J. Chem. Phys. **150**, 064105 (2019) 10.1063/1.5053562.

[65]O. T. Unke, and M. Meuwly, "PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges", J. Chem. Theory Comput. **15**, 3678–3693 (2019) 10.1021/acs.jctc.9b00181.

[66]T. A. Profitt, and J. K. Pearson, "A shared-weight neural network architecture for predicting molecular properties", Phys. Chem. Chem. Phys. **21**, 26175–26183 (2019) 10.1039/C9CP03103K.

[67]A. S. Christensen, L. A. Bratholm, F. A. Faber, and O. Anatole von Lilienfeld, "FCHL revisited: Faster and more accurate quantum machine learning", J. Chem. Phys. **152**, Publisher: American Institute of Physics, 044107 (2020) 10.1063/1.5126701.

[68]V. Zaverkin, and J. Kästner, "Gaussian Moments as Physically Inspired Molecular Descriptors for Accurate and Scalable Machine Learning Potentials", J. Chem. Theory Comput., Publisher: American Chemical Society (2020) 10.1021/acs.jctc.0c00347.

[69]C. W. Park, M. Kornbluth, J. Vandermause, C. Wolverton, B. Kozinsky, and J. P. Mailoa, *Accurate and scalable multi-element graph neural network force field and molecular dynamics with direct force architecture*, 2020.

[70]M. Welborn, L. Cheng, and T. F. Miller III, "Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis", J. Chem. Theory Comput. **14**, 4772–4779 (2018) 10.1021/acs.jctc.8b00636.

[71]L. Cheng, M. Welborn, A. S. Christensen, and T. F. Miller III, "A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules", J. Chem. Phys. **150**, 131103 (2019) 10.1063/1.5088393.

[72]L. Cheng, N. B. Kovachki, M. Welborn, and T. F. Miller III, "Regression Clustering for Improved Accuracy and Training Costs with Molecular-Orbital-Based Machine Learning", J. Chem. Theory Comput. **15**, Publisher: American Chemical Society, 6668–6677 (2019) 10.1021/acs.jctc.9b00884.

[73]M.-C. Kim, E. Sim, and K. Burke, "Understanding and Reducing Errors in Density Functional Calculations", Phys. Rev. Lett. **111**, 073003 (2013) 10.1103/PhysRevLett.111.073003.

[74]A. Wasserman, J. Nafziger, K. Jiang, M.-C. Kim, E. Sim, and K. Burke, "The importance of being inconsistent", Annu. Rev. Phys. Chem. **68**, 555–581 (2017) 10.1146/annurev-physchem-052516-044957.

[75]S. Dapprich, I. Komáromi, K. Byun, K. Morokuma, and M. J. Frisch, "A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives", J. Mol. Struct. **461-462**, 1–21 (1999) 10.1016/S0166-1280(98)00475-8.

[76]T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery, K. Morokuma, and M. J. Frisch, "Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM", J. Chem. Theory Comput. **2**, edited by Intergovernmental Panel on Climate Change, 815–826 (2006) 10.1021/ct050289g.

[77]R. Sure, and S. Grimme, "Corrected small basis set Hartree-Fock method for large systems", J. Comput. Chem. **34**, 1672–1685 (2013) 10.1002/jcc.23317.

[78]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, 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, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, 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, and M. Wang, *Molpro, version 2015.1, a package of ab initio programs*, see http://www.molpro.net, Cardiff, UK, 2015.

[79]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple", Phys. Rev. Lett. **77**, 3865–3868 (1996) 10.1103/PhysRevLett.77.3865.

[80]T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen", J. Chem. Phys. **90**, 1007–1023 (1989) 10.1063/1.456153.

[81]W. J. Hehre, R. Ditchfield, and J. A. Pople, "Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules", J. Chem. Phys. **56**, 2257–2261 (1972) 10.1063/1.1677527.

[82]W. J. Hehre, R. F. Stewart, and J. A. Pople, "Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals", J. Chem. Phys. **51**, 2657–2664 (1969) 10.1063/1.1672392.

[83]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, "Molpro: a general-purpose quantum chemistry program package", Wiley Interdiscip. Rev. Comput. Mol. Sci. **2**, 242–253 (2012) 10.1002/wcms.82.

[84]K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, "Auxiliary basis sets to approximate Coulomb potentials", Chem. Phys. Lett. **240**, 283–290 (1995) 10.1016/0009-2614(95)00621-A.

[85]K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, "Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials", Theor. Chem. Acc. **97**, 119–124 (1997) 10.1007/s002140050244.

[86]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *Gaussian 09 Revision A.1*, Gaussian Inc. Wallingford CT 2009.

[87]A. Redondo, W. A. Goddard, C. A. Swarts, and T. C. McGill, "Oxidation of silicon surfaces", J. Vac. Sci. Technol. **19**, 498–501 (1981) 10.1116/1.571046.

[88]C. J. Wu, and E. A. Carter, "Structures and adsorption energetics for chemisorbed fluorine atoms on Si(100)-2×1", Phys. Rev. B **45**, 9065–9081 (1992) 10.1103/PhysRevB.45.9065.

[89]K. Jug, and T. Bredow, "Models for the treatment of crystalline solids and surfaces", J. Comput. Chem. **25**, 1551–1567 (2004) 10.1002/jcc.20080.

[90]T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular electronic-structure theory* (Wiley, 2002).

[91]D. Asturiol, M. Duran, and P. Salvador, "Intramolecular Basis Set Superposition Error Effects on the Planarity of DNA and RNA Nucleobases", J. Chem. Theory Comput. **5**, 2574–2581 (2009) 10.1021/ct900056u.

[92]H. Valdés, V. Klusák, M. Pitoňák, O. Exner, I. Starý, P. Hobza, and L. Rulíšek, "Evaluation of the intramolecular basis set superposition error in the calculations of larger molecules: [n]helicenes and Phe-Gly-Phe tripeptide", J. Comput. Chem. **29**, 861–870 (2008) 10.1002/jcc.20841.

[93]R. O. Jones, "Density functional theory: its origins, rise to prominence, and future", Rev. Mod. Phys. **87**, 897–923 (2015) 10.1103/RevModPhys.87.897.

[94]A. Warshel, and M. Levitt, "Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme", J. Mol. Biol. **103**, 227–249 (1976) 10.1016/0022-2836(76)90311-9.

[95]M. J. Field, P. A. Bash, and M. Karplus, "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations", J. Comput. Chem. **11**, 700–733 (1990) 10.1002/jcc.540110605.

[96]D. S. Lambrecht, "Generalizing energy decomposition analysis to response properties to inform expedited predictive models", Comput. Theor. Chem. **1149**, 24–30 (2019) 10.1016/j.comptc.2018.12.009.

[97]G. Senator, and K. Subbaswamy, "Density dependence of the dielectric constant of rare-gas crystals", Phys. Rev. B **34**, 5754–5757 (1986) 10.1103/PhysRevB.34.5754.

[98]P. Cortona, "Self-consistently determined properties of solids without band-structure calculations", Phys. Rev. B **44**, 8454–8458 (1991) 10.1103/PhysRevB.44.8454.

[99]T. A. Wesolowski, and A. Warshel, "Frozen density functional approach for ab initio calculations of solvated molecules", J. Phys. Chem. **97**, 8050–8053 (1993) 10.1021/j100132a040.

[100]M. Iannuzzi, B. Kirchner, and J. Hutter, "Density functional embedding for molecular systems", Chem. Phys. Lett. **421**, 16–20 (2006) 10.1016/j.cplett.2005.08.155.

[101]C. R. Jacob, J. Neugebauer, and L. Visscher, "Software news and update: a flexible implementation of frozen-density embedding for use in multilevel simulations", J. Comput. Chem. **29**, 1011–1018 (2008) 10.1002/jcc.20861.

[102]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", J. Chem. Phys. **129**, 184104 (2008) 10.1063/1.3007987.

[103]N. Govind, Y. A. Wang, A. J. R. da Silva, and E. A. Carter, "Accurate ab initio energetics of extended systems via explicit correlation embedded in a density functional environment", Chem. Phys. Lett. **295**, 129–134 (1998) 10.1016/S0009-2614(98)00939-7.

[104]C. Huang, M. Pavone, and E. A. Carter, "Quantum mechanical embedding theory based on a unique embedding potential", J. Chem. Phys. **134**, 154110 (2011) 10.1063/1.3577516.

[105]P. Elliott, K. Burke, M. H. Cohen, and A. Wasserman, "Partition density-functional theory", Phys. Rev. A **82**, 024501 (2010) 10.1103/PhysRevA.82.024501.

[106]G. Knizia, and G. K.-L. Chan, "Density matrix embedding: A strong-coupling quantum embedding theory", J. Chem. Theory Comput. **9**, 1428–1432 (2013) 10.1021/ct301044e.

[107]M. Welborn, T. Tsuchimochi, and T. Van Voorhis, "Bootstrap embedding: An internally consistent fragment-based method", J. Chem. Phys. **145** (2016) 10.1063/1.4960986.

[108]H. Lin, and D. G. Truhlar, "QM/MM: what have we learned, where are we, and where do we go from here?", Theor. Chem. Acc. **117**, 185–199 (2007) 10.1007/s00214-006-0143-z.

[109]C. R. Jacob, and J. Neugebauer, "Subsystem density functional theory", WIREs Comput. Mol. Sci. **4**, 325–362 (2014) 10.1002/wcms.1175.

[110]K. Yu, C. M. Krauter, J. M. Dieterich, and E. A. Carter, "Density and potential functional embedding: theory and practice", in *Fragmentation* (Wiley-Blackwell, 2017) Chap. 2, pp. 81–117, 10.1002/9781119129271.ch2.

[111]Q. Sun, and G. K.-L. Chan, "Quantum embedding theories", Acc. Chem. Res. **49**, 2705–2712 (2016) 10.1021/acs.accounts.6b00356.

[112]T. A. Wesolowski, "One-electron equations for embedded electron density: challenge for theory and practical payoffs in multi-level modeling of complex polyatomic systems", in *Computational chemistry: reviews of current trends*, Vol. 10 (World Scientific, Singapore, 2006), pp. 1–82.

[113]A. W. Götz, S. M. Beyhan, and L. Visscher, "Performance of Kinetic Energy Functionals for Interaction Energies in a Subsystem Formulation of Density Functional Theory", J. Chem. Theory Comput. **5**, 3161–3174 (2009) 10.1021/ct9001784.

[114]S. Fux, C. R. Jacob, J. Neugebauer, L. Visscher, and M. Reiher, "Accurate frozen-density embedding potentials as a first step towards a subsystem description of covalent bonds", J. Chem. Phys. **132**, 164101 (2010) 10.1063/1.3376251.

[115]J. Nafziger, Q. Wu, and A. Wasserman, "Molecular binding energies from partition density functional theory", J. Chem. Phys. **135**, 234101 (2011) 10.1063/1.3667198.

[116]J. P. Unsleber, J. Neugebauer, and C. R. Jacob, "No need for external orthogonality in subsystem density-functional theory", Phys. Chem. Chem. Phys. **18**, 21001–21009 (2016) 10.1039/C6CP00332J.

[117]I. G. Ryabinkin, A. A. Kananenka, and V. N. Staroverov, "Accurate and efficient approximation to the optimized effective potential for exchange", Phys. Rev. Lett. **111**, 013001 (2013) 10.1103/PhysRevLett.111.013001.

[118]W. Yang, and Q. Wu, "Direct method for optimized effective potentials in density-functional theory", Phys. Rev. Lett. **89**, 143002 (2002) 10.1103/PhysRevLett.89.143002.

[119]S. Kümmel, and J. P. Perdew, "Optimized effective potential made simple: Orbital functionals, orbital shifts, and the exact Kohn-Sham exchange potential", Phys. Rev. B **68**, 035103 (2003) 10.1103/PhysRevB.68.035103.

[120]D. S. Jensen, and A. Wasserman, "Numerical methods for the inverse problem of density functional theory", Int. J. Quantum Chem. **118**, e25425 (2018) 10.1002/qua.25425.

[121]J. C. Phillips, and L. Kleinman, "New method for calculating wave functions in crystals and molecules", Phys. Rev. **116**, 287–294 (1959) 10.1103/PhysRev.116.287.

[122]S. Huzinaga, and A. A. Cantu, "Theory of separability of many–electron systems", J. Chem. Phys. **55**, 5543–5549 (1971) 10.1063/1.1675720.

[123]B. Hégely, P. R. Nagy, G. G. Ferenczy, and M. Kállay, "Exact density functional and wave function embedding schemes based on orbital localization", J. Chem. Phys. **145**, 064107 (2016) 10.1063/1.4960177.

[124]T. Culpitt, K. R. Brorsen, and S. Hammes-Schiffer, "Communication: Density functional theory embedding with the orthogonality constrained basis set expansion procedure", J. Chem. Phys. **146**, 211101 (2017) 10.1063/1.4984777.

[125]T. M. Henderson, "Embedding wave function theory in density functional theory", J. Chem. Phys. **125**, 014105 (2006) 10.1063/1.2209688.

[126]Y. G. Khait, and M. R. Hoffmann, "Chapter three - on the orthogonality of orbitals in subsystem kohn-sham density functional theory", in *Annu. rep. comput. chem.* Vol. 8, edited by R. A. Wheeler, Annu. Rep. Comput. Chem. (Elsevier, 2012), pp. 53–70, https://doi.org/10.1016/B978-0-444-59440-2.00003-X.

[127]S. Sæther, T. Kjærgaard, H. Koch, and I.-M. Høyvik, "Density-based multilevel Hartree-Fock model", J. Chem. Theory Comput. **13**, 5282–5290 (2017) 10.1021/acs.jctc.7b00689.

[128]M. Welborn, F. R. Manby, and T. F. Miller III, "Even-handed subsystem selection in projection-based embedding", J. Chem. Phys. **149**, 144101 (2018) 10.1063/1.5050533.

[129]J. Pipek, and P. G. Mezey, "A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions", J. Chem. Phys. **90**, 4916–4926 (1989) 10.1063/1.456588.

[130]G. Knizia, "Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts", J. Chem. Theory Comput. **9**, Publisher: American Chemical Society, 4834–4843 (2013) 10.1021/ct400687b.

[131]M. Schütz, G. Hetzer, and H.-J. Werner, "Low-order scaling local electron correlation methods. I. Linear scaling local MP2", J. Chem. Phys. **111**, 5691–5705 (1999) 10.1063/1.479957.

[132]S. J. Bennie, B. F. Curchod, F. R. Manby, and D. R. Glowacki, "Pushing the limits of EOM-CCSD with projector-based embedding for excitation energies", J. Phys. Chem. Lett. **8**, 5559–5565 (2017) 10.1021/acs.jpclett.7b02500.

[133]R. M. Parrish, J. F. Gonthier, C. Corminbœuf, and C. D. Sherrill, "Communication: Practical intramolecular symmetry adapted perturbation theory via Hartree-Fock embedding", J. Chem. Phys. **143**, 051103 (2015) 10.1063/1.4927575.

[134]O. R. Meitei, and A. Heßelmann, "Intramolecular interactions in sterically crowded hydrocarbon molecules", J. Comput. Chem. **38**, 2500–2508 (2017) 10.1002/jcc.24908.

[135]A. P. de Lima Batista, A. G. S. de Oliveira-Filho, and S. E. Galembeck, "Photophysical properties and the NO photorelease mechanism of a ruthenium nitrosyl model complex investigated using the CASSCF-in-DFT embedding approach", Phys. Chem. Chem. Phys. **19**, 13860–13867 (2017) 10.1039/C7CP01642E.

[136]K. Yao, J. E. Herr, S. N. Brown, and J. Parkhill, "Intrinsic bond energies from a bonds-in-molecules neural network", J. Phys. Chem. Lett. **8**, 2689–2694 (2017) 10.1021/acs.jpclett.7b01072.

[137]L. Lin, and L. Zepeda-Núñez, "Projection based embedding theory for solving Kohn-Sham density functional theory", Multiscale Modeling & Simulation **17**, 1274–1300 (2019) 10.1137/18m1202670.

[138]M. Böckers, and J. Neugebauer, "Excitation energies of embedded open-shell systems: Unrestricted frozen-density-embedding time-dependent density-functional theory", J. Chem. Phys. **149**, 074102 (2018) 10.1063/1.5040233.

[139]X. Chen, and C. Franklin Goldsmith, "Predictive kinetics for the thermal decomposition of rdx", Proc. Combust. Inst. **37**, 3167–3173 (2019) 10.1016/j.proci.2018.06.036.

[140]F. Libisch, M. Marsman, J. Burgdörfer, and G. Kresse, "Embedding for bulk systems using localized atomic orbitals", J. Chem. Phys. **147**, 034110 (2017) 10.1063/1.4993795.

[141]D. V. Chulhai, and J. D. Goodpaster, "Projection-based correlated wave function in density functional theory embedding for periodic systems", J. Chem. Theory Comput. **14**, 1928–1942 (2018) 10.1021/acs.jctc.7b01154.

[142]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 2018.2, a package of ab initio programs*, see http://www.molpro.net, Cardiff, UK, 2018.

[143]A. Hesselmann, and O. R. Meitei, "Geometry optimizations with the incremental molecular fragmentation method", J. Theor. Comput. Chem. **17**, 1850037 (2018) 10.1142/S0219633618500372.

[144]T. Nagata, D. G. Fedorov, and K. Kitaura, "Analytic gradient for the embedding potential with approximations in the fragment molecular orbital method", Chem. Phys. Lett. **544**, 87–93 (2012) 10.1016/j.cplett.2012.07.004.

[145]H. Nakata, T. Nagata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, "Analytic second derivatives of the energy in the fragment molecular orbital method", J. Chem. Phys. **138**, 164103 (2013) 10.1063/1.4800990.

[146]K. R. Brorsen, F. Zahariev, H. Nakata, D. G. Fedorov, and M. S. Gordon, "Analytic Gradient for Density Functional Theory Based on the Fragment Molecular Orbital Method", J. Chem. Theory Comput. **10**, 5297–5307 (2014) 10.1021/ct500808p.

[147]H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch, and T. Vreven, "QM:QM electronic embedding using Mulliken atomic charges: Energies and analytic gradients in an ONIOM framework", J. Chem. Phys. **128**, 034107 (2008) 10.1063/1.2814164.

[148]N. J. Mayhall, K. Raghavachari, and H. P. Hratchian, "ONIOM-based QM:QM electronic embedding method using Löwdin atomic charges: Energies and analytic gradients", J. Chem. Phys. **132**, 114107 (2010) 10.1063/1.3315417.

[149]H. P. Hratchian, A. V. Krukau, P. V. Parandekar, M. J. Frisch, and K. Raghavachari, "QM:QM embedding using electronic densities within an ONIOM framework: Energies and analytic gradients", J. Chem. Phys. **135**, 014105 (2011) 10.1063/1.3603450.

[150]F. R. Manby, T. F. Miller III, P. J. Bygrave, F. Ding, T. Dresselhaus, A. Batista-Romero, A. Buccheri, C. Bungey, S. J. R. Lee, R. Meli, C. Steinmann, T. Tsuchiya, M. Welborn, and T. Wiles, *Entos: a quantum molecular simulation package*, Feb. 2019, 10.26434/chemrxiv.7762646.v2.

[151]M. Dułak, J. W. Kamiński, and T. A. Wesołowski, "Equilibrium geometries of noncovalently bound intermolecular complexes derived from subsystem formulation of density functional theory", J. Chem. Theory Comput. **3**, 735–745 (2007) 10.1021/ct600367t.

[152]J. Heuser, and S. Höfener, "Wave-function frozen-density embedding: Approximate analytical nuclear ground-state gradients", J. Comput. Chem. **37**, 1092–1101 (2016) 10.1002/jcc.24301.

[153]J. Heuser, and S. Höfener, "Analytical nuclear excited-state gradients for the Tamm-Dancoff approximation using uncoupled frozen-density embedding", J. Comput. Chem. **38**, 2316–2325 (2017) 10.1002/jcc.24885.

[154]J. Heuser, and S. Hoefener, "Analytical nuclear excited-state gradients for the second-order approximate coupled-cluster singles and doubles (CC2) method employing uncoupled frozen-density embedding", J. Chem. Theory Comput. **14**, 4616–4628 (2018) 10.1021/acs.jctc.8b00369.

[155]A. Kovyrshin, and J. Neugebauer, "Analytical gradients for excitation energies from frozen-density embedding", Phys. Chem. Chem. Phys. **18**, 20955–20975 (2016) 10.1039/C6CP00392C.

[156]D. Schlüns, M. Franchini, A. W. Götz, J. Neugebauer, C. R. Jacob, and L. Visscher, "Analytical gradients for subsystem density functional theory within the slater-function-based amsterdam density functional program", J. Comput. Chem. **38**, 238–249 (2017) 10.1002/jcc.24670.

[157]K. Klahr, D. Schlüns, and J. Neugebauer, "Geometry Optimizations in a Subsystem Density Functional Theory Formalism: A Benchmark Study", J. Chem. Theory Comput. **14**, 5631–5644 (2018) 10.1021/acs.jctc.8b00475.

[158]P. Pulay, "Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules: I. Theory", Mol. Phys. **17**, 197–204 (1969) 10.1080/00268976900100941.

[159]J. M. Foster, and S. F. Boys, "Canonical Configurational Interaction Procedure", Rev. Mod. Phys. **32**, 300–302 (1960) 10.1103/RevModPhys.32.300.

[160]M. Schütz, H.-J. Werner, R. Lindh, and F. R. Manby, "Analytical energy gradients for local second-order Moller–Plesset perturbation theory using density fitting approximations", J. Chem. Phys. **121**, 737–750 (2004) 10.1063/1.1760747.

[161]P. Hohenberg, and W. Kohn, "Inhomogeneous Electron Gas", Phys. Rev. **136**, B864–B871 (1964) 10.1103/PhysRev.136.B864.

[162]S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis", Can. J. Phys. **58**, 1200–1211 (1980) 10.1139/p80-159.

[163]C. Adamo, and V. Barone, "Toward reliable density functional methods without adjustable parameters: The PBE0 model", J. Chem. Phys. **110**, 6158–6170 (1999) 10.1063/1.478522.

[164]F. Weigend, and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence", Phys. Chem. Chem. Phys. **7**, 3297–3305 (2005) 10.1039/B508541A.

[165]F. Weigend, "Accurate Coulomb-fitting basis sets for H to Rn", Phys. Chem. Chem. Phys. **8**, 1057 (2006) 10.1039/b515623h.

[166]F. Neese, "Software update: the ORCA program system, version 4.0", Wiley Interdiscip. Rev. Comput. Mol. Sci. **8**, e1327 (2018) 10.1002/wcms.1327.

[167]C. Møller, and M. S. Plesset, "Note on an Approximation Treatment for Many-Electron Systems", Phys. Rev. **46**, 618–622 (1934) 10.1103/PhysRev.46.618.

[168]G. E. Scuseria, C. L. Janssen, and H. F. Schaefer, "An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations", J. Chem. Phys. **89**, 7382–7387 (1988) 10.1063/1.455269.

[169]G. D. Purvis, and R. J. Bartlett, "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples", J. Chem. Phys. **76**, 1910–1918 (1982) 10.1063/1.443164.

[170]K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories", Chem. Phys. Lett. **157**, 479–483 (1989) 10.1016/S0009-2614(89)87395-6.

[171]U. Bozkaya, and C. D. Sherrill, "Analytic energy gradients for the coupled-cluster singles and doubles method with the density-fitting approximation", J. Chem. Phys. **144**, 174103 (2016) 10.1063/1.4948318.

[172]U. Bozkaya, and C. D. Sherrill, "Analytic energy gradients for the coupled-cluster singles and doubles with perturbative triples method with the density-fitting approximation", J. Chem. Phys. **147**, 044104 (2017) 10.1063/1.4994918.

[173]E. R. Davidson, "The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices", J. Comput. Phys. **17**, 87–94 (1975) 10.1016/0021-9991(75)90065-0.

[174]J. Kauczor, P. Jørgensen, and P. Norman, "On the Efficiency of Algorithms for Solving Hartree–Fock and Kohn–Sham Response Equations", J. Chem. Theory Comput. **7**, 1610–1630 (2011) 10.1021/ct100729t.

[175]L.-P. Wang, and C. Song, "Geometry optimization made simple with translation and rotation coordinates", J. Chem. Phys. **144**, 214108 (2016) 10.1063/1.4952956.

[176]L.-P. Wang, *Geometric*, https://github.com/leeping/geometric, 2019.

[177]G. Henkelman, and H. Jónsson, "Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points", J. Chem. Phys. **113**, 9978–9985 (2000) 10.1063/1.1323224.

[178]A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, "The atomic simulation environment—a Python library for working with atoms", J. Phys. Condens. Matter **29**, 273002 (2017).

[179]S. Smidstrup, A. Pedersen, K. Stokbro, and H. Jónsson, "Improved initial guess for minimum energy path calculations", J. Chem. Phys. **140**, 214106 (2014) 10.1063/1.4878664.

[180]G. Henkelman, B. P. Uberuaga, and H. Jónsson, "A climbing image nudged elastic band method for finding saddle points and minimum energy paths", J. Chem. Phys. **113**, 9901–9904 (2000) 10.1063/1.1329672.

[181]E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, "Structural Relaxation Made Simple", Phys. Rev. Lett. **97**, 170201 (2006) 10.1103/PhysRevLett.97.170201.

[182]M. Schwilk, D. Usvyat, and H.-J. Werner, "Communication: Improved pair approximations in local coupled-cluster methods", J. Chem. Phys. **142**, 121102 (2015) 10.1063/1.4916316.

[183]F. Weigend, "A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency", Phys. Chem. Chem. Phys. **4**, 4285–4291 (2002) 10.1039/b204199p.

[184]F. Weigend, "Hartree–Fock exchange fitting basis sets for H to Rn", J. Comput. Chem. **29**, 167–175 (2008) 10.1002/jcc.20702.

[185]F. Weigend, A. Köhn, and C. Hättig, "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations", J. Chem. Phys. **116**, 3175–3183 (2002) 10.1063/1.1445115.

[186]H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, and M. Dornbach, "Scalable electron correlation methods I.: PNO-LMP2 with linear scaling in the molecular size and near-inverse-linear scaling in the number of processors", J. Chem. Theory Comput. **11**, 484–507 (2015) 10.1021/ct500725e.

[187]M. Schwilk, Q. Ma, C. Köppl, and H.-J. Werner, "Scalable Electron Correlation Methods. 3. Efficient and Accurate Parallel Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD)", J. Chem. Theory Comput. **13**, Publisher: American Chemical Society, 3650–3675 (2017) 10.1021/acs.jctc.7b00554.

[188]Q. Ma, and H.-J. Werner, "Explicitly correlated local coupled-cluster methods using pair natural orbitals", Wiley Interdiscip. Rev. Comput. Mol. Sci. **8**, e1371 (2018) 10.1002/wcms.1371.

[189]P. Pinski, and F. Neese, "Analytical gradient for the domain-based local pair natural orbital second order Møller-Plesset perturbation theory method (DLPNO-MP2)", J. Chem. Phys. **150**, Publisher: American Institute of Physics, 164102 (2019) 10.1063/1.5086544.

[190]T. Husch, J. Sun, L. Cheng, S. J. R. Lee, and T. F. M. III, *Improved accuracy and transferability of molecular-orbital-based machine learning: organics, transition-metal complexes, non-covalent interactions, and transition states*, 2020.

[191]R. A. Kendall, T. H. Dunning, and R. J. Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions", J. Chem. Phys. **96**, 6796–6806 (1992) 10.1063/1.462569.

[192]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, et al., *MOLPRO, version 2019.2, a package of ab initio programs* (Cardiff, UK, 2019).

[193]G. Hetzer, M. Schütz, H. Stoll, and H.-J. Werner, "Low-order scaling local correlation methods II: Splitting the Coulomb operator in linear scaling local second-order Møller–Plesset perturbation theory", J. Chem. Phys. **113**, 9443–9455 (2000) 10.1063/1.1321295.

[194]H.-J. Werner, F. R. Manby, and P. J. Knowles, "Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations", J. Chem. Phys. **118**, 8149–8160 (2003) 10.1063/1.1564816.

[195]G. E. Scuseria, A. C. Scheiner, T. J. Lee, J. E. Rice, and H. F. Schaefer, "The closed-shell coupled cluster single and double excitation (CCSD) model for the description of electron correlation. A comparison with configuration interaction (CISD) results", J. Chem. Phys. **86**, 2881–2890 (1987) 10.1063/1.452039.

[196]A. C. Scheiner, G. E. Scuseria, J. E. Rice, T. J. Lee, and H. F. Schaefer, "Analytic evaluation of energy gradients for the single and double excitation coupled cluster (CCSD) wave function: Theory and application", J. Chem. Phys. **87**, 5361–5373 (1987) 10.1063/1.453655.

[197]T. J. Lee, and A. P. Rendell, "Analytic gradients for coupled-cluster energies that include noniterative connected triple excitations: Application to *c i s* - and *t r a n s* -HONO", J. Chem. Phys. **94**, 6229–6236 (1991) 10.1063/1.460411.

[198]M. Schütz, "Low-order scaling local electron correlation methods. III. Linear scaling local perturbative triples correction (*T*)", J. Chem. Phys. **113**, 9986–10001 (2000) 10.1063/1.1323265.

[199]M. Schütz, and H.-J. Werner, "Local perturbative triples correction (T) with linear cost scaling", Chem. Phys. Lett. **318**, 370–378 (2000) 10.1016/S0009-2614(00)00066-X.

[200]H.-J. Werner, and M. Schütz, "An efficient local coupled cluster method for accurate thermochemistry of large systems", J. Chem. Phys. **135**, 144116 (2011) 10.1063/1.3641642.

[201]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., "Scikit-learn: Machine learning in Python", J. Mach. Learn. Res. **12**, 2825–2830 (2011).

[202]C. E. Rasmussen, and C. K. I. Williams, *Gaussian processes for machine learning*, Adaptive computation and machine learning, OCLC: ocm61285753 (MIT Press, Cambridge, Mass, 2006).

[203]M. G. Genton, "Classes of kernels for machine learning: a statistics perspective", J. Mach. Learn. Res. **2**, 299–312 (2002).

[204] GPy, *GPy: a gaussian process framework in python*, http://github.com/SheffieldML/GPy, since 2012.

[205]L. Breiman, "Random forests", Mach. Learn. **45**, 5–32 (2001) 10.1023/A:1010933404324.

[206]L. Breiman, "Statistical Modeling: The Two Cultures", Statist. Sci. **16**, Publisher: Institute of Mathematical Statistics, 199–231 (2001) 10.1214/ss/1009213726.

[207]L. Cheng, M. Welborn, A. S. Christensen, and T. F. Miller III, *Thermalized (350K) QM7b, GDB-13, water, and short alkane quantum chemistry dataset including MOB-ML features*, Version Number: 1.1 type: dataset, Feb. 2019, 10.22002/D1.1177.

[208]Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo, and F. Neese, "Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]", J. Chem. Phys. **148**, Publisher: American Institute of Physics, 011101 (2018) 10.1063/1.5011798.

[209]P. R. Nagy, G. Samu, and M. Kállay, "Optimization of the Linear-Scaling Local Natural Orbital CCSD(T) Method: Improved Algorithm and Benchmark Applications", J. Chem. Theory Comput. **14**, Publisher: American Chemical Society, 4193–4215 (2018) 10.1021/acs.jctc.8b00442.

[210]S. J. R. Lee, T. Husch, F. Ding, and T. F. Miller III, "Analytical gradients for molecular-orbital-based machine learning", in preparation (2020).

[211]M. Dornbach, and H.-J. Werner, "Analytical energy gradients for local second-order Møller-Plesset perturbation theory using intrinsic bond orbitals", Mol. Phys. **117**, 1252–1263 (2019) 10.1080/00268976.2018.1537529.