

The Efficient Computation of Field-Dependent Molecular Properties in the Frequency and Time Domains

Benjamin Gilbert Peyton

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

Daniel Crawford, Chair

Nicholas Mayhall

Diego Troya

John Morris

May 11, 2022

Blacksburg, Virginia

Keywords: Electronic structure theory, machine learning, coupled cluster, local correlation

Copyright 2022, Benjamin Gilbert Peyton

The Efficient Computation of Field-Dependent Molecular Properties in the Frequency and Time Domains

Benjamin Gilbert Peyton

(ABSTRACT)

The efficient computation of dynamic (time-dependent) molecular properties is a broad field with numerous applications in aiding molecular synthesis and design, with a particular prevalence in spectroscopic predictions. Typical methods for computing the response of a molecular system to an electromagnetic field (EMF) considers a quantum mechanical description of the molecule and a classical approximation for the EMF. Methods for describing light-matter interactions with high-accuracy electronic structure methods, such as coupled cluster (CC), are discussed, with a focus on improving the efficiency of such methods.

The CC method suffers from high-degree polynomial scaling. In addition to the ground-state calculation, computing dynamic properties requires the description of sensitive excited-state effects. The cost of such methods often prohibits the accurate calculation of response properties for systems of significant importance, such as large-molecule drug candidates or chiral species present in biological systems. While the literature is ripe with reduced-scaling methods for CC ground-state calculations, considerably fewer approaches have been applied to excited-state properties, with even fewer still providing adequate results for realistic systems. This work presents three studies on the reduction of the cost of molecular property evaluations, in the hopes of closing this gap in the literature and widening the scope of current theoretical methods.

There are two main ways of simulating time-dependent light-matter interactions: one may consider these effects in the frequency domain, where the response of the system to an EMF

is computed directly; or, the response may be considered explicitly in the time domain, where wave function (or density) parameters can be propagated in time and examined in detail. Each methodology has unique advantages and computational bottlenecks. The first two studies focus on frequency-domain calculations, and employ fragmentation and machine-learning techniques to reduce the cost of single-molecule calculations or sets of calculations across a series of geometric conformations. The third study presents a novel application of the local correlation technique to real-time CC calculations, and highlights deficiencies and possible solutions to the approach.

The Efficient Computation of Field-Dependent Molecular Properties in the Frequency and Time Domains

Benjamin Gilbert Peyton

(GENERAL AUDIENCE ABSTRACT)

Theoretical chemistry plays a key role in connecting experimental results with physical interpretation. Paramount to the success of theoretical methods is the ability to predict molecular properties without the need for costly high-throughput synthesis, aiding in the determination of molecular structure and the design of new materials. Light-matter interactions, which govern spectroscopic techniques, are particularly complicated, and sensitive to the theoretical tools employed in their prediction. Compounding the issue of accuracy is one of efficiency — accurate theoretical methods typically incur steep scaling of computational cost (memory and processor time) with respect to the size of the system.

An important aspect in improving the efficiency of these methods is understanding the nature of light-matter interactions at a quantum level. Many unanswered questions still remain, such as, “Can light-matter interactions be thought of as a sum of interactions between smaller fragments of the system?” and “Can conventional methods of accelerating ground-state calculations be expected to perform well for spectroscopic properties?” The present work seeks to answer these questions through three studies, focusing on improving the efficiency of these techniques, while simultaneously addressing their fundamental flaws and providing reasonable alternatives.

Acknowledgments

I would like to acknowledge several people without whom this work would not have been possible:

- My brother, for never giving up on me, and assuring me that the sun will shine on us again.
- My parents, for their unconditional love and support.
- My grandparents, for keeping me on this path and being there for me from start to finish. I miss you dearly.
- Ruhee, for improving my science and my life in every aspect.
- Josh and Mike, for being the greatest friends and allowing me to become part of the family.
- Memaw and WC, for providing an example and a warm home. We all love you.
- The Mage and the CCG, for cultivating my love of science and teaching me the importance of camaraderie.
- Doc Ozz, for showing me The Way.
- Conrad and Amber, for challenging me to do better, and putting up with me for so long.
- Vibin and Shannon, for jumping every grad school hurdle together with me.

- Past and present members of the Crawford group: Coleman, Ashutosh, AJ, Kirk, Zhe, Roberto, Alex, Monika, Susana, Jose, and Brendan, for being the best coworkers I could have asked for.
- Mayhall group members: Daniel and Harper, for reminding me CI and quantum computing are, in fact, useful.
- Valeev group members: Kalyani, Karl, Nakul, Marjory, and Varun, for providing outside input and being a great source of knowledge in programming and electronic structure.
- My committee members: Profs. Mayhall, Troya, and Morris, for seeing me through each milestone.
- Prof. Crawford, for everything I could have asked for as a mentor, an advisor, and a lifelong colleague.
- And to all the children, congratulations!

Contents

List of Figures	xi
List of Tables	xix
1 Introduction	1
2 Theoretical background	5
2.1 Ground state electronic structure theory	5
2.1.1 Many-Body Perturbation Theory	6
2.1.2 Coupled Cluster	8
2.1.3 Reduced Density Matrices	10
2.2 Response theory	12
2.2.1 Exact Response	13
2.2.2 Approximate Response	18
3 Basis Set Superposition Errors in the Many-Body Expansion of Molecular Properties	21
3.1 Introduction	21
3.2 Computational Details	25
3.3 Results and Discussion	28

3.3.1	Interaction Energies and Electric Dipole Moments	28
3.3.2	Dynamic Dipole Polarizabilities	29
3.3.3	Specific Rotations	31
3.4	Conclusions	43
4	Machine-Learning Coupled Cluster Properties through a Density Tensor Representation	45
4.1	Introduction	45
4.2	Theory	48
4.2.1	Electronic Structure Theory	48
4.2.2	Machine-Learning	50
4.3	Computational Details	54
4.4	Results and Discussion	57
4.4.1	Truncation considerations	57
4.4.2	Energies	59
4.4.3	Dipole moments	61
4.4.4	Extensions	65
4.5	Conclusions	70
5	Locally Correlated Real-Time Coupled Cluster Theory	72
5.1	Introduction	72

5.2	Theoretical Background	75
5.2.1	Real-Time Coupled Cluster Theory	76
5.2.2	Properties	78
5.2.3	Local Correlation	79
5.3	Computational Details	82
5.4	Results and Discussion	84
5.4.1	Absorption and ECD Spectra	85
5.4.2	Amplitude Dynamics	89
5.5	Conclusions	95
6	Conclusions	97
Bibliography		101
Appendices		125
Appendix A	Supporting Information for Basis Set Superposition Errors in the Many-Body Expansion of Molecular Properties	126
A.1	Additional Data	126
A.1.1	Methyloxirane	126
A.1.2	Methylthiirane	135
A.1.3	Dimethylallene	140

A.2 Atomic Coordinates	142
Appendix B Supporting Information for Machine-Learning Coupled Cluster Properties through a Density Tensor Representation	150
Appendix C Supporting Information for Locally Correlated Real-Time Coupled Cluster Theory	186

List of Figures

3.1 MBE and SSFC of (a) interaction energies and (b) dipole moments for (<i>S</i>)-methyloxirane in a 13-water solvent shell.	29
3.2 MBE and SSFC of dynamic polarizabilities for (a) (<i>S</i>)-methyloxirane in a seven-water solvent shell, (b) (<i>S</i>)-methylthiirane in a six-water solvent shell, and (c) (<i>M</i>)-dimethylallene in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.	30
3.3 MBE and SSFC of dynamic polarizabilities for (<i>S</i>)-methyloxirane in a 13-water solvent shell. Computed with CAM-B3LYP/aDZ.	32
3.4 MBE and SSFC of specific rotation for (a) (<i>S</i>)-methyloxirane in a seven-water solvent shell, (b) (<i>S</i>)-methylthiirane in a six-water solvent shell, and (c) (<i>M</i>)-dimethylallene in a seven-water solvent shell. Computed with B3LYP/aDZ.	34
3.5 MBE and SSFC of specific rotation for (a) (<i>S</i>)-methyloxirane in a seven-water solvent shell and (b) (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with B3LYP/aTZ.	36
3.6 MBE and SSFC of specific rotation for (a) (<i>S</i>)-methylthiirane in a six-water solvent shell and (b) (<i>M</i>)-dimethylallene in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.	37
3.7 MBE and SSFC of specific rotation for additional snapshots of (a) (<i>S</i>)-methyloxirane in a seven-water solvent shell and (b) (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.	38

3.8 MBE and SSFC of specific rotation for (<i>S</i>)-methyloxirane in a 13-water solvent shell. Computed with CAM-B3LYP/aDZ.	40
3.9 MBE and SSFC of solute-fragment-only specific rotation for (a) (<i>S</i>)-methyloxirane in a seven-water solvent shell and (b) (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.	42
3.10 MBE and SSFC of solute-fragment-only specific rotation for (<i>S</i>)-methyloxirane in a 13-water solvent shell. Computed with CAM-B3LYP/aDZ.	43
4.1 (a) KRR for the carbon monoxide potential energy curve with and without molecular symmetry considered and (b) their respective amplitude distributions. TATR models used $M = 12$ training points. All amplitudes with magnitude $< 10^{-8}$ are set to 0.	58
4.2 DTR vs TATR errors in mE_h for small molecule datasets: (a) H_2O , (b) CH_3OH , (c) (<i>S</i>)-methyloxirane, and (d) (<i>R</i>)-methylthiirane. Red lines indicate $2 mE_h$.	60
4.3 DTR vs TATR errors in mE_h for all datasets. (* = methyl)	62
4.4 DTR vs TATR errors in <i>milliDebye</i> for all datasets. (* = methyl)	64
4.5 DTR and TATR validation curves for (<i>S</i>)-methyloxirane correlation energy.	69
5.1 Reference and PNO absorption spectra of $(H_2)_4$ for five cutoffs: $[1 \times 10^{-10}, 1 \times 10^{-9}, 1 \times 10^{-8}, 1 \times 10^{-7}, 2 \times 10^{-6}]$ corresponding to [93%, 82%, 63%, 44%, 24%] of the MO virtual space, respectively.	85

5.2	Reference and PAO absorption spectra of $(H_2)_4$ for five cutoffs: $[1 \times 10^{-4}, 1 \times 10^{-3}, 1 \times 10^{-2}, 5 \times 10^{-2}, 1 \times 10^{-1}]$ corresponding to [95%, 86%, 63%, 46%, 23%] of the MO virtual space, respectively.	86
5.3	Reference and PNO ECD spectra of $(H_2)_4$ for five cutoffs: $[1 \times 10^{-10}, 1 \times 10^{-9}, 1 \times 10^{-8}, 1 \times 10^{-7}, 2 \times 10^{-6}]$ corresponding to [93%, 82%, 63%, 44%, 24%] of the MO virtual space, respectively.	88
5.4	Reference and PAO ECD spectra of $(H_2)_4$ for five cutoffs: $[1 \times 10^{-4}, 1 \times 10^{-3}, 1 \times 10^{-2}, 5 \times 10^{-2}, 1 \times 10^{-1}]$ corresponding to [95%, 86%, 63%, 46%, 23%] of the MO virtual space, respectively.	89
5.5	Time-dependent change in the norm of the amplitude tensors relative to the ground-state amplitudes. (Field and step parameters remain unchanged, and the amplitude norm is taken at every 1 a.u.)	90
5.6	MO-basis t_1 amplitude deviations from $t = 0$ after (a) 1 a.u., (b) 50 a.u., and (c) 100 a.u. of time propagation. Each row contains the same four occupied orbital indices and a subset of virtual indices as indicated by the x-axis labels.	91
5.7	PNO-basis t_1 amplitude deviations from $t = 0$ after (a) 1 a.u., (b) 50 a.u., and (c) 100 a.u. of time propagation. Each row contains the same four occupied orbital indices and a subset of virtual indices as indicated by the x-axis labels.	93
5.8	Virtual MO energy ϵ_a and the occupation number n_a (plotted on a log scale) for unique PNO spaces i_1 and i_2 versus orbital extent. Virtual MOs 7 and 15 are denoted by a solid + and \times , respectively. The horizontal line denotes a PNO cutoff of 1×10^{-7}	94

A.1 MBE and SSFC of interaction energy for (<i>S</i>)-methyloxirane in a seven-water solvent shell. Computed with B3LYP/aDZ.	127
A.2 MBE and SSFC of dipole moment for (<i>S</i>)-methyloxirane in a seven-water solvent shell. Computed with B3LYP/aDZ.	127
A.3 MBE and SSFC of interaction energy for (<i>S</i>)-methyloxirane in a seven-water solvent shell. Computed with B3LYP/aTZ.	128
A.4 MBE and SSFC of dipole moment for (<i>S</i>)-methyloxirane in a seven-water solvent shell. Computed with B3LYP/aTZ.	128
A.5 MBE and SSFC of interaction energy for (<i>S</i>)-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.	129
A.6 MBE and SSFC of dipole moment for (<i>S</i>)-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.	129
A.7 MBE and SSFC of specific rotation for (<i>S</i>)-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.	130
A.8 MBE and SSFC of interaction energy for snapshot #2 of (<i>S</i>)-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.	130
A.9 MBE and SSFC of dipole moment for snapshot #2 of (<i>S</i>)-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.	131
A.10 (<i>S</i>)-methyloxirane	131
A.11 MBE and SSFC of dynamic polarizabilities for snapshot #2 of (<i>S</i>)-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.	131

A.12 MBE and SSFC of specific rotation of (<i>S</i>)-methyloxirane in a seven-water solvent shell and two manually reduced H-bonds. Computed with CAM-B3LYP/aDZ.	132
A.13 MBE and SSFC of specific rotation of (<i>S</i>)-methyloxirane in a seven-water solvent shell using “tight” convergence criteria and “fine” grids. Computed with CAM-B3LYP/aDZ.	133
A.14 MBE and SSFC of specific rotation of (<i>S</i>)-methyloxirane in a seven-water solvent shell and unpruned “fine” grids. Computed with CAM-B3LYP/aDZ.	134
A.15 MBE and SSFC of interaction energy for (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with B3LYP/aDZ.	135
A.16 MBE and SSFC of dipole moment for (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with B3LYP/aDZ.	135
A.17 MBE and SSFC of interaction energy for (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with B3LYP/aTZ.	136
A.18 MBE and SSFC of dipole moment for (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with B3LYP/aTZ.	136
A.19 MBE and SSFC of interaction energy for (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.	137
A.20 MBE and SSFC of dipole moment for (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.	137
A.21 MBE and SSFC of interaction energy for snapshot #2 of (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.	138

A.22 MBE and SSFC of dipole moment for snapshot #2 of (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.	138
A.23 MBE and SSFC of dynamic polarizabilities for snapshot #2 of (<i>S</i>)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.	139
A.24 MBE and SSFC of interaction energy for (<i>M</i>)-dimethylallene in a seven-water solvent shell. Computed with B3LYP/aDZ.	140
A.25 MBE and SSFC of dipole moment for (<i>M</i>)-dimethylallene in a seven-water solvent shell. Computed with B3LYP/aDZ.	140
A.26 MBE and SSFC of interaction energy for (<i>M</i>)-dimethylallene in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.	141
A.27 MBE and SSFC of dipole moment for (<i>M</i>)-dimethylallene in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.	141
B.1 DTR vs TATR errors in mE_h for diatomic datasets: (a) H ₂ , (b) HF, (c) CO, (d) LiF, and (e) N ₂ . Red lines indicate 2 mE_h	151
B.2 DTR vs TATR errors in Debye for diatomic datasets: (a) HF, (b) CO, and (c) LiF. Red lines indicate 2 milliDebye.	152
B.3 DTR vs TATR errors in Debye for small molecule datasets: (a) H ₂ O, (b) CH ₃ OH, (c) (<i>S</i>)-methyloxirane, and (d) (<i>R</i>)-methylthiirane. Red lines indicate 2 milliDebye.	153
B.4 DTR and TATR learning curves for H ₂ correlation energy.	154
B.5 DTR and TATR learning curves for HF correlation energy.	155
B.6 DTR and TATR learning curves for CO correlation energy.	156

B.7 DTR and TATR learning curves for LiF correlation energy.	157
B.8 DTR and TATR learning curves for N ₂ correlation energy.	158
B.9 DTR and TATR learning curves for H ₂ O correlation energy.	159
B.10 DTR and TATR learning curves for CH ₃ OH correlation energy.	160
B.11 DTR and TATR learning curves for (<i>S</i>)-methyloxirane correlation energy. . .	161
B.12 DTR and TATR learning curves for (<i>R</i>)-methylthiirane correlation energy. .	162
B.13 DTR and TATR learning curves for HF correlated dipole.	163
B.14 DTR and TATR learning curves for CO correlated dipole.	164
B.15 DTR and TATR learning curves for LiF correlated dipole.	165
B.16 DTR and TATR learning curves for H ₂ O correlated dipole.	166
B.17 DTR and TATR learning curves for CH ₃ OH correlated dipole.	167
B.18 DTR and TATR learning curves for (<i>S</i>)-methyloxirane correlated dipole. .	168
B.19 DTR and TATR learning curves for (<i>R</i>)-methylthiirane correlated dipole. .	169
B.20 DTR and TATR validation curves for H ₂ correlation energy.	170
B.21 DTR and TATR validation curves for HF correlation energy.	171
B.22 DTR and TATR validation curves for CO correlation energy.	172
B.23 DTR and TATR validation curves for LiF correlation energy.	173
B.24 DTR and TATR validation curves for N ₂ correlation energy.	174
B.25 DTR and TATR validation curves for H ₂ O correlation energy.	175
B.26 DTR and TATR validation curves for CH ₃ OH correlation energy.	176

B.27 DTR and TATR validation curves for (<i>S</i>)-methyloxirane correlation energy. . .	177
B.28 DTR and TATR validation curves for (<i>R</i>)-methylthiirane correlation energy. .	178
B.29 DTR and TATR validation curves for HF correlated dipole.	179
B.30 DTR and TATR validation curves for CO correlated dipole.	180
B.31 DTR and TATR validation curves for LiF correlated dipole.	181
B.32 DTR and TATR validation curves for H ₂ O correlated dipole.	182
B.33 DTR and TATR validation curves for CH ₃ OH correlated dipole.	183
B.34 DTR and TATR validation curves for (<i>S</i>)-methyloxirane correlated dipole. .	184
B.35 DTR and TATR validation curves for (<i>R</i>)-methylthiirane correlated dipole. .	185
 C.1 Time-dependent change in the norm of the amplitude tensors relative to the ground-state amplitudes, using a PNO cut-off of 1×10^{-9} . (Amplitudes have been back-transformed to the MO space for comparison to MO-basis ampli- tude norms. Field and step parameters remain unchanged, and the amplitude norm is taken at every 1 a.u.)	187
 C.2 Time-dependent change in the norm of the amplitude tensors relative to the ground-state amplitudes, using a PAO cut-off of 1×10^{-3} . (Amplitudes have been back-transformed to the MO space for comparison to MO-basis ampli- tude norms. Field and step parameters remain unchanged, and the amplitude norm is taken at every 1 a.u.)	188

List of Tables

4.1	Mean absolute errors (in Debye) relative to CCSD of four machine-learning datasets utilizing full, electronic, and correlated dipole moments as training targets. MP2 dipoles given for comparison. (* = methyl)	63
A.1	Atomic coordinates of (<i>S</i>)-methyloxirane in a seven-water solvent shell (Ångstroms)	142
A.2	Atomic coordinates of (<i>S</i>)-methyloxirane in a seven-water solvent shell (Snapshot #2) (Ångstroms)	143
A.3	Atomic coordinates of (<i>S</i>)-methyloxirane in a 7-water solvent shell and two manually reduced H-bonds (Ångstroms)	144
A.4	Atomic coordinates of (<i>S</i>)-methyloxirane in a 13-water solvent shell (Ångstroms)	145
A.5	Atomic coordinates of (<i>S</i>)-methylthiirane in a six-water solvent shell (Ångstroms)	147
A.6	Atomic coordinates of (<i>S</i>)-methylthiirane in a six-water solvent shell (Snapshot #2) (Ångstroms)	148
A.7	Atomic coordinates of (<i>M</i>)-dimethylallene in a seven-water solvent shell (Ångstroms)	149
C.1	Atomic coordinates of Helical H ₂ tetramer (Ångstroms)	187

Attribution

Chapters 3, 4, and 5 are coauthored by Prof. T. Daniel Crawford. Chapter 4 is additionally coauthored by Connor Briggs, Dr. Ruhee D'Cunha, and Dr. Johannes T. Margraf. Chapter 5 is additionally coauthored by Zhe Wang. Calculations, figures, and a majority of the text for these chapters was contributed by myself. Prof. Crawford served as a mentor for all projects and edited these chapters. For Chapter 4, Connor Briggs contributed machine learning code, Dr. D'Cunha provided molecular geometries, and Dr. Margraf provided TATR comparison data and editing support. For Chapter 5, Zhe Wang contributed to the Python packaged used, PyCC.

Chapter 1

Introduction

Modern synthetic organic chemistry employs a vast array of sophisticated instrumentation. Principal among these are probes of light-matter interactions, which reveal rich structural and electro-magnetic characteristics. These instruments work by measuring the absorption, scattering, or refraction of an electromagnetic field (EMF) interacting with a target system.¹ Experimental techniques which measure these interactions with respect to the energy or frequency of the incident EMF are known as spectroscopy. Many staple experimental apparatus probe these relationships, including (but certainly not limited to) ultraviolet-visible absorption (UV-Vis), nuclear magnetic resonance (NMR), electronic and vibrational circular dichroism (ECD and VCD), and optical rotatory dispersion (ORD). These data can be used for characterizing synthetic products or establishing structure-property relationships which aid in the development of novel systems for applications in materials, bio-organics, and more. Interpreting or predicting the results of such experiments requires a knowledge of fundamental light-matter interactions on a quantum level.

A quantum description of any molecular interaction may be viewed as the effects of a perturbation on a quantum-mechanical system. In the case of EMF, this perturbation may be static (fixed) or dynamic (varying in frequency or time). A fully quantum-mechanical description of the light-matter system would require quantum electrodynamics (QED); however, it is often sufficient to treat the EMF from a classical perspective, treating only the molecular response using quantum mechanics.^{1,2} This allows us to utilize many well-established methods

in the field of theoretical chemistry. These generally provide a representation of the system (a wave function or density) with which to take the expectation value of a given operator or, in some cases, predict the expectation value directly.³ Methods which utilize only mathematical techniques (that is, no experimental or phenomenological parameters) are said to be *ab initio* methods. These methods often (but not always) simplify or ignore the quantum effects of the nuclei and their motion, an approximation known as the Born-Oppenheimer approximation.⁴ Wave function-based methods are a subclass of *ab initio* methods which center around solving the time-dependent or time-independent Schrödinger wave equation to obtain an explicit form of the wave function. Finally, methods that go beyond the mean-field or Hartree-Fock approximation, allowing electrons to interact through more than an average self-consistent field, are classified as correlated methods. It is these correlated wave function-based methods under the Born-Oppenheimer approximation upon which the bulk of the current work is built. As such, Chapter 2 is a primer in the theoretical underpinnings necessary to understand this work. The focus will be on fundamentals which are not explicitly presented in the publications that follow in Chapters 3–5. Chapter 2 is roughly divided into details of electronic structure theory (Section 2.1) and molecular response properties (Section 2.2).

The first topic discussed is many-body perturbation theory (MBPT).⁵ This framework allows us to separate the quantum mechanical properties of the system from the effects of some perturbing force, which is expanded in orders and truncated under the assumption that higher-order terms become negligible. This force is often taken to be electron correlation when describing the electronic ground state wave function;⁶ however, it may also be a static or dynamic EMF. In Section (2.1.1) we present some basic characteristics of general MBPT, in preparation for its application in the chapters that follow.

Second, in Section (2.1.2) we present the wave function-based method which the bulk of this

work is either based upon or seeks to approximate: coupled cluster (CC) theory.^{7–10} This method, like perturbation theory, is most commonly used for computing electron correlation (though its roots are in nuclear physics). Unlike perturbation theory, we do not perform an order-by-order expansion of the perturbation; instead, a cluster operator folds in contributions based on a physical intuition, that is the substitution of electrons in occupied orbitals into unoccupied or virtual orbitals, which gives rise to electron correlation. Truncation is then based on the number of simultaneous electron substitutions allowed – singles, doubles, *etc.* This method, while accurate and systematically improvable, is notoriously expensive, suffering from high-order polynomial scaling. Extensions to excited states and molecular response properties only compound this issue;^{2,11–13} as such, Chapters 3 and 4 are focused on ways to circumvent using CC on any but a small subset of systems. It should be noted that, while Chapter 3 utilizes a different correlated method based on a density-functional theory approach, the primary goal was to ascertain the effects of fragmentation of a system through the many-body expansion on the computation of molecular properties. This is a benchmark study, whose conclusions are understood to be applicable to more expensive methods such as CC, where such benchmark studies could not be performed. Additional considerations of reduced density matrices,^{14–16} which are used throughout the work to refer to correlated wave functions, follows in Section 2.1.3.

Finally, we describe the method by which we couple the classical perturbing field to the quantum-mechanical correlated wave function. By applying time-dependent perturbation theory¹⁷ we derive general expressions for tensors which describe the molecular response to an EMF. In the exact theory, these tensors are functions of the excited electronic states of the system; however, an approach to generalize these expressions to approximate theories such as truncated coupled cluster is discussed,^{18–21} which also avoids the costly evaluation of excited-state wave function parameters. For dynamic EMF, these properties may be described as

a function of frequency, so the most common approach is to derive the working equations using a Fourier transform of expressions obtained using the time-dependent Schrödinger equation (more generally known as response theory); however, these expressions may also be evaluated explicitly in the time-domain.^{22,23} In Section 2.2 we focus on the frequency-domain formulation, while Chapter 5 explores the recently revived prospect of explicit time-propagation, as well as a technique to make this process cheaper through a well-established concept known as local correlation.²⁴

Chapter 2

Theoretical background

2.1 Ground state electronic structure theory

Electron correlation is defined in the present work to mean any additional effect beyond what is included in the ground-state Hartree-Fock (HF) wave function $|0\rangle$. In ground-state electronic structure theory, this is typically taken to be an anti-symmetrized, single-determinant wave function, generally composed of linear combinations of atomic Gaussian basis functions. In solving the HF equations, the electronic energy is variationally minimized, under the constraint that the orbital basis functions of $|0\rangle$ remain orthonormal.³ The resulting orbitals will be referred to as the canonical molecular orbital (MO) basis.

This section will briefly describe a selection of methods for recovering the correlation energy which are pertinent to this work. Correlation is usually included by considering the effects of substituted determinants in which electrons in ground-state occupied orbitals are substituted (often referred to as “excited”) into higher-energy unoccupied (or virtual) orbitals. These occupied and virtual orbitals are often chosen to be the canonical MOs of the reference wave function, which in turn are linear combinations of atomic orbital basis functions. However, as discussed in Chapter 5, some schemes apply a different approach – namely, localization (and potentially truncation) of the canonical MOs prior to computing the correlation energy. Therefore, this chapter will focus on *general* expressions, without assuming canonical or orthogonal orbital spaces except where mentioned. Einstein summation notation (implied

summation over repeated indices) will be used throughout this chapter, except in cases where an explicit sum is instructive.

2.1.1 Many-Body Perturbation Theory

Correlation energy corrections to the HF energy may be obtained through many-body perturbation theory (MBPT). We may begin with either the time-dependent or time-independent Schrödinger equation – notation for time-dependence will be suppressed here, but the same rules apply as mentioned in Section 2.2. By partitioning the electronic Hamiltonian \hat{H} into a zeroth-order (HF) term $\hat{H}^{(0)}$ and a perturbation $\hat{H}^{(1)}$, which is first-order in an undetermined coefficient λ ,

$$\hat{H} = \hat{H}^{(0)} + \lambda \hat{H}^{(1)} \quad (2.1)$$

and expanding the electronic energy E and wave function $|\Psi\rangle$ similarly in orders of the perturbation (denoted by superscript),

$$|\Psi\rangle = \sum_{i=0}^{\infty} \lambda^i |\Psi^{(i)}\rangle \quad (2.2a)$$

$$E = \sum_{i=0}^{\infty} \lambda^i E^{(i)}, \quad (2.2b)$$

contributions to the correlation energy are obtained order-by-order. $|\Psi^{(0)}\rangle$ is taken to be $|0\rangle$, and λ^0 to be one. Implicit here is the assumption that the perturbation is small relative to the reference – *i.e.*, the correlation energy is several orders of magnitude smaller than the HF energy.

For a perturbation containing correlation operators, inserting Eqs. (2.1) and (2.2) into the time-independent Schrödinger equation and truncating at the second order in λ yields the

first non-zero contribution to the energy:³

$$E^{(2)} = \langle 0 | \hat{H}^{(1)} | \Psi^{(1)} \rangle. \quad (2.3)$$

Second-order MBPT is equivalent to the minimization of Eq. (2.3). Using the Hylleraas partitioning of the Hamiltonian,²⁵

$$\hat{H}^{(0)} = E^{(0)} + \hat{F} \quad (2.4a)$$

$$\hat{H}^{(1)} = \hat{W} \quad (2.4b)$$

where \hat{F} and \hat{W} are the one- and two-particle components of the electronic Hamiltonian, respectively, we may solve Eq. (2.3). This is achieved through minimization of a residual expression defined by the first-order wave function equation,

$$\hat{W}|0\rangle + \hat{F}|\Psi^{(1)}\rangle = 0. \quad (2.5)$$

To generate programmable expressions, we may choose to represent our first-order perturbed wave function as a linear combination of substituted HF determinants with coefficients t_{ij}^{ab}

$$|\Psi^{(1)}\rangle = \sum_{ijab} \frac{1}{4} t_{ij}^{ab} |0\rangle. \quad (2.6)$$

Since the Hamiltonian does not connect $|0\rangle$ to singly-substituted determinants due to the Brillouin condition,³ it follows that only double substitutions are present in $|\Psi^{(1)}\rangle$. Projection of Eq. (2.5) onto a basis of doubly substituted determinants yields programmable expressions for t_{ij}^{ab} . In the limit of canonical HF orbitals, Eq. (2.3) reduces to the more commonly known

second-order Møller-Plesset perturbation theory (MP2)^{6,26} expressions,

$$E^{(2)} = \frac{1}{4} t_{ij}^{ab} \langle ij || ab \rangle \quad (2.7a)$$

$$t_{ij}^{ab} = -\frac{\langle ij || ab \rangle}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \quad (2.7b)$$

where $\langle ij || ab \rangle$ is the antisymmetrized two-electron integral between occupied orbitals $\{i, j\}$ and virtual orbitals $\{a, b\}$, and ϵ_p is the p th diagonal element of the Fock matrix. This requires only one $\mathcal{O}(N^5)$ step (where N is a measure of system size) in the rotation of the two-electron atomic orbital integrals into the MO basis followed by a one-step energy calculation. However, commonplace localizations of the occupied space^{27–29} require instead the iterative evaluation of Eq. (2.5). Furthermore, reduced nonorthogonal virtual spaces may be employed in the projection basis (as will be explored in Chapter 5),²⁴ which necessitates the inclusion of additional overlap matrices between nonorthogonal orbital spaces in these equations. While this increases the cost by adding iterations and additional expressions to evaluate, these additional costs are heavily outweighed by the computational savings due to the reduction of N in large systems.

2.1.2 Coupled Cluster

The coupled-cluster family of electronic structure methods^{7–10} employs an exponentiated cluster operator \hat{T} to include the effects of substituted determinants,

$$\hat{H}e^{\hat{T}}|0\rangle = Ee^{\hat{T}}|0\rangle \quad (2.8)$$

with

$$\hat{T} = \sum_{n=1}^M \hat{T}_n \quad (2.9a)$$

$$\hat{T}_n = \left(\frac{1}{n!} \right)^2 t_{ij\dots}^{ab\dots} a_a^\dagger a_b^\dagger \dots a_j a_i \quad (2.9b)$$

where M is the number of electrons in the system, and a_p and a_p^\dagger are the standard second-quantized excitation and de-excitation operators, respectively, for a general orbital p . Rather than perturbative expansion, the cluster operator is expanded in orders of substitutions (singles, doubles, *etc.* in Eq. (2.9a)), then truncated to a given substitution level to produce tractable equations. Additional simplification is achieved by using the similarity-transformed Hamiltonian

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}} \quad (2.10)$$

which has the same eigenspectrum as \hat{H} , but naturally truncates at four nested commutators in a Campbell-Baker-Hausdorff expansion

$$\bar{H} = \hat{H} + [\hat{H}, \hat{T}] + \frac{1}{2!} [[\hat{H}, \hat{T}], \hat{T}] + \frac{1}{3!} [[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}]] + \frac{1}{4!} [[[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}], \hat{T}]]. \quad (2.11)$$

Projecting Eq. (2.8) on the left by $\langle 0 | e^{-\hat{T}}$ yields the coupled cluster energy expression and, as in Section 2.1.1, expressions for the amplitudes $t_{ij\dots}^{ab\dots}$ are obtained by projection onto substituted determinants $\langle \mu |$:

$$\langle 0 | \bar{H} | 0 \rangle = E_{CC} \quad (2.12a)$$

$$\langle \mu | \bar{H} | 0 \rangle = 0. \quad (2.12b)$$

Here Eq. (2.12a) follows from the normalization condition of $|0\rangle$, and Eq. (2.12b) follows from the orthogonality condition. In the case of localized orbitals, such as those used in Chapter 5, additional terms involving the overlap of nonorthogonal orbitals will appear in the final expressions.

Untruncated, Eqs. (2.12a) and (2.12b) produce exact solutions to the time-independent Schrödinger equation for the electronic Hamiltonian in a given basis set. Perhaps the most

common truncation of Eq. (2.9a) is to single and double substitutions, defining the CCSD method. The computational cost of this method scales as $\mathcal{O}(N^6)$. In terms of MBPT, CCSD may be viewed as correct through infinite order, but only within the space of single and double substitutions. A perturbative correction for triples produces the “gold standard” in quantum chemistry, the CCSD(T) method,^{30,31} which scales as $\mathcal{O}(N^7)$. Aside from these, additional truncation schemes exist for a variety of purposes. For example, the CCD method includes only double substitutions, but the complexity of the resulting equations is far simpler. Another method, CC2,³² is tuned for the calculation of molecular response properties, such as dipoles and polarizabilities. This method truncates the doubles expression at first order by considering the \hat{T}_2 operator as first order. The full singles expressions from CCSD are retained by considering \hat{T}_1 to be first order, and the one- and two-electron terms of the Hamiltonian are still considered to be zeroth and first order, respectively, as in Section 2.1.1. Solving the resulting iterative amplitude expressions for CC2 scales as $\mathcal{O}(N^5)$ but, unlike the similarly scaling MP2 method, includes the effects of singles, which have been shown to be crucial for computing accurate molecular properties.^{32,33} It is important to note that while the studies in Chapters 4 and 5 utilize CCSD, many other flavors of CC may be used as drop-in replacements, both in theory and in the implementation of their codes.

2.1.3 Reduced Density Matrices

A general expression for the ground-state energy of an arbitrary wave function $|\Psi\rangle$ prepared in a basis $\{p, q, r, s\}$ can be written in terms of one- and two-particle reduced density matrices (1-RDM D and 2-RDM Γ):^{15,34}

$$\begin{aligned} E &= \langle \Psi | \hat{H} | \Psi \rangle \\ &= D_{pq} h_p^q + \Gamma_{pqrs} g_{pqrs} \end{aligned} \tag{2.13}$$

where h_p^q and g_{pqrs} are the one- and two-electron integrals of the electronic Hamiltonian, and D and Γ are defined according to the form of $|\Psi\rangle$ and the excitation operators of the second quantized Hamiltonian:

$$D_{pq} = \langle \Psi | a_p^\dagger a_q | \Psi \rangle \quad (2.14a)$$

$$\Gamma_{pqrs} = \langle \Psi | a_p^\dagger a_q^\dagger a_s a_r | \Psi \rangle. \quad (2.14b)$$

Clearly, this requires the left-hand wave function $\langle \Psi |$. For the electronic Hamiltonian, which is Hermitian, the left- and right-hand wave functions are identical, as is the case for MP2. However, this Hermiticity is destroyed for standard CC methods due to the similarity transformation of the Hamiltonian in Eq. (2.10).¹⁰ It is therefore necessary to solve for the CC left-hand wave function,

$$\langle \mathcal{L} | = \langle \Psi | \hat{\mathcal{L}} \quad (2.15)$$

where $\hat{\mathcal{L}}$ is a left-hand cluster operator analogous to Eq. (2.9b), with analogous amplitudes $\lambda_{ij\dots}^{ab\dots}$. This additional set of coupled equations can be solved in the same manner as the right-hand amplitude expressions. We may then rewrite the coupled cluster energy expression using Eq. (2.8):

$$\langle \Psi | \hat{\mathcal{L}} \bar{H} | \Psi \rangle = E. \quad (2.16)$$

As will be explored in Section 2.2, molecular properties may be expressed as derivatives of the electronic energy.² Given the left- and right-hand wave functions and assuming the Hellmann-Feynman theorem^{35,36} holds (and thus $|\Psi\rangle$ carries no dependence on external perturbations), the derivative of the energy with respect to an arbitrary perturbation operator $\hat{\Omega}$ can be conveniently expressed

$$\begin{aligned} \frac{\partial E}{\partial \hat{\Omega}} &= \langle \Psi | \frac{\partial \hat{H}}{\partial \hat{\Omega}} | \Psi \rangle \\ &= D_{pq} \frac{\partial h_{pq}}{\partial \hat{\Omega}} + \Gamma_{pqrs} \frac{\partial g_{pqrs}}{\partial \hat{\Omega}} \end{aligned} \quad (2.17)$$

without further specification of the wave function or Hamiltonian. Thus, time-independent first-order properties may be expressed without differentiation of the RDMs. This was a driving force behind the study in Chapter 3, and is important for the machine-learning methods discussed in Chapter 4. The cases of higher-order and time-dependent properties, which are the focus of Chapters 3 and 5, will be handled in the following section.

2.2 Response theory

Response theory combines adiabatic perturbation theory with a Fourier transform of the time-dependent equations into the frequency domain.^{1,37} The prediction of the response of molecular systems to an electromagnetic field has important applications in organic synthesis and characterization. To explore these effects, we may relate properties to a perturbative expansion of a total electric or magnetic multipole moment, from which response tensors related to a host of properties may be defined. These response tensors generally take the form of products of ground- and excited-state transition moments of the electric or magnetic dipole operator. In this chapter we will summarize the derivation of these response tensors and their relationships to field-induced molecular phenomena. We then present a method for obtaining these tensors through Fourier transform of a time-dependent quasi-energy^{2,20,21} which is rigorously defined in the context of electronic structure theory, providing working equations for predicting molecular responses to EMF using approximate wave function-based theory.

2.2.1 Exact Response

Under classical electrodynamics, the potential energy V of a system of charges in a static electric field may be written in a multipole expansion:¹

$$V = q(\phi)_O - \mu_\alpha(E_\alpha)_O - \frac{1}{3}\Theta_{\alpha\beta}(E_{\alpha\beta})_O + \dots \quad (2.18)$$

with the scalar potential ϕ , the Cartesian component α of the electric field E_α , and the electric field gradient with respect to two Cartesian coordinates $E_{\alpha\beta}$. Here we have defined the multipoles as the electric monopole or charge q , the electric dipole μ_α , and the electric quadrupole $\Theta_{\alpha\beta}$. The subscript O indicates that the potential, field, or gradient is taken at the system origin. We may also expand the potential in a Taylor series about the potential evaluated at zero field:

$$V[E_O] = V_O + (E_\alpha)_O \frac{\partial V}{\partial (E_\alpha)_O} \Big|_0 + \frac{1}{2}(E_\alpha)_O (E_\beta)_O \frac{\partial V^2}{\partial (E_\alpha)_O \partial (E_\beta)_O} \Big|_0 + \dots \quad (2.19)$$

From Eq. (2.18), we see that a Cartesian component of the electric dipole may be written as the derivative of the potential with respect to the field,

$$\mu_\alpha = -\frac{\partial V}{\partial (E_\alpha)_O} \Big|_0. \quad (2.20)$$

By comparing Eqs. (2.19) and (2.20), we see that the total electric dipole may be re-written

$$\mu_\alpha = \mu_{0\alpha} + \alpha_{\alpha\beta}(E_\beta)_O + \frac{1}{2}\beta_{\alpha\beta\gamma}(E_\beta)_O (E_\gamma)_O + \dots \quad (2.21)$$

where we have defined the response tensors for the permanent (field-independent) dipole moment $\mu_{0\alpha}$, electric polarizability $\alpha_{\alpha\beta}$, first electric hyperpolarizability $\beta_{\alpha\beta\gamma}$, and higher

order terms as higher-order derivatives of the potential, *viz.*

$$\mu_{0\alpha} = -\frac{\partial V}{\partial(E_\alpha)_O}\Big|_0 \quad (2.22a)$$

$$\alpha_{\alpha\beta} = -\frac{\partial^2 V}{\partial(E_\alpha)_O \partial(E_\beta)_O}\Big|_0 \quad (2.22b)$$

$$\beta_{\alpha\beta\gamma} = -\frac{\partial^3 V}{\partial(E_\alpha)_O \partial(E_\beta)_O \partial(E_\gamma)_O}\Big|_0. \quad (2.22c)$$

Property tensors for magnetic and mixed electric-magnetic properties, as well as higher-order multipoles in Eq. (2.18), are defined in an analogous manner.¹

Quantum mechanical expressions for the property tensors may be obtained through perturbation theory. For static-field properties, this process is straightforward, and expressions similar to those found for correlation energy corrections in Eq. (2.7a) arise in the perturbative expansion of the time-independent Schrödinger equation, e.g., for the static polarizability tensor:

$$\alpha_{\alpha\beta} = -2 \sum_{j \neq 0} \frac{\langle n | \mu_\alpha | j \rangle \langle j | \mu_\beta | n \rangle}{V_n - V_j} \quad (2.23)$$

where the sum runs over excited-states j .

Of more practical application is the prediction of dynamic-field properties, such as dynamic polarizabilities, optical rotation, and circular dichroism (which are the primary targets of Chapters 3 and 5). For these and other time-dependent response properties, some description of the time evolution of the wave function is required. These properties may be solved directly from the time-dependent analogue of Eq. (2.21). To compute the now-time-dependent dipole, one must compute the dipole moment expectation value of a system in the pres-

ence of an explicitly time-dependent EMF. Fourier transformation of the dynamic dipole (and the corresponding EMF terms) into the frequency domain produces a broadband spectrum. However, explicit time propagation of the wave function is exceedingly expensive.²² This is the topic of Chapter 5. For now, we will consider an alternative approach utilizing time-dependent perturbation theory. This approach avoids explicit propagation of the wave function, but limits our application to individual frequencies, rather than broadband spectra. Nevertheless, some properties (such as optical rotation) are only experimentally measured at a select few frequencies, and spectra may still be approximated by computing values at multiple frequencies.

Beginning from the time-dependent Schrödinger equation,

$$\hat{H}(t)|\Psi(t)\rangle = i\hbar \frac{\partial}{\partial t}|\Psi(t)\rangle \quad (2.24)$$

we will, without loss of generality, assume a time-dependent wave function expansion of the form

$$|\Psi(t)\rangle = d_n(t)|\psi_n(t)\rangle \quad (2.25a)$$

$$|\psi_n(t)\rangle = e^{-iE_n t/\hbar}|n\rangle \quad (2.25b)$$

where $|\psi(t)\rangle$ is an approximate wave function consisting of a weighted sum of known, time-independent functions $|n\rangle$ and an exponentiated phase. As in Eq. (2.1), we will express the Hamiltonian as a sum of the time-independent molecular Hamiltonian $\hat{H}^{(0)}$ with a time-dependent perturbation $\hat{V}(t)$ (generally taken to be an EMF) and inserting this and the wave function expansion into Eq. (2.24) yields

$$(\hat{H}^{(0)} + \hat{V}(t))d_n(t)e^{-iE_n t/\hbar}|n\rangle = i\hbar \frac{\partial}{\partial t} [d_n(t)e^{-iE_n t/\hbar}]|n\rangle. \quad (2.26)$$

It should be noted that, in the absence of a perturbation (time $t = 0$), this reduces to the time-independent Schrödinger equation ($d_n(0) = 1$). Using the chain rule, Eq. (2.26) becomes

$$(\hat{H}^{(0)} + \hat{V}(t))d_n(t)e^{-iE_nt/\hbar}|n\rangle = \left[E_n d_n(t) + i\hbar \frac{\partial d_n(t)}{\partial t} \right] e^{-iE_nt/\hbar}|n\rangle. \quad (2.27)$$

If we assume we have *exact* eigenstates $|n\rangle$ of the molecular Hamiltonian, *i.e.* exact ground- and excited-state wave functions, then the first terms of both sides of Eq. (2.27) cancel. Projecting on the left by the excited-state $\langle m|$ and rearranging, we arrive at the differential equation

$$\frac{\partial d_m(t)}{\partial t} = \frac{1}{i\hbar} \sum_n \langle m | \hat{V}(t) | n \rangle d_n(t) e^{i\omega_{mn}t} \quad (2.28)$$

which describes the time evolution of the wave function coefficient $d_m(t)$. We have also introduced the angular frequency $\omega_{mn} = (E_m - E_n)/\hbar$ and included the explicit sum over excited states $|n\rangle$ for clarity.

Integration of Eq. (2.28) from time $t' = 0$ to $t' = t$ yields a recursive equation for the wave function parameter $d_m(t)$, due to the presence of $d_n(t)$ on the right-hand side. However, as a consequence of our assumption that the system begins in its unperturbed ground-state, all coefficients at $t' = 0$ reduce to $\delta_{n,0}$. To linear order in \hat{V} , Eq. (2.28) may be written

$$d_m(t) = \frac{1}{i\hbar} \int_0^t \langle m | \hat{V}(t') | 0 \rangle e^{i\omega_{m0}t'} dt'. \quad (2.29)$$

To continue, we choose our perturbation to be a simple, time-dependent field at a given frequency ω

$$\hat{V}(t) = \hat{v}_\alpha^\omega F_\alpha^\omega e^{-i\omega t} \quad (2.30)$$

with field strength F_α^ω and arbitrary field operator \hat{v}_α^ω , such as the electric or magnetic dipole operator. Finally, inserting Eq. (2.30) into Eq. (2.29) and integrating, we arrive at the final

expression for the wave function parameters

$$d_m(t) = -\frac{\langle m | \hat{v}_\alpha^\omega | 0 \rangle e^{i(\omega_{m0} - \omega)t}}{\hbar(\omega_{m0} - \omega)} F_\alpha^\omega. \quad (2.31)$$

To obtain a dipole expression in the form of Eq. (2.21), we may compute the expectation value of the electric dipole operator $\hat{\mu}$. After rearrangement, the expectation value can be written:

$$\begin{aligned} \langle \Psi(t) | \hat{\mu} | \Psi(t) \rangle &= \langle 0 | \hat{\mu} | 0 \rangle \\ &- \sum_{m \neq 0} \left[\frac{\langle m | \hat{\mu} | 0 \rangle \langle 0 | \hat{v}_\alpha^\omega | m \rangle}{\hbar(\omega_{m0} + \omega)} + \frac{\langle 0 | \hat{\mu} | m \rangle \langle m | \hat{v}_\alpha^\omega | 0 \rangle}{\hbar(\omega_{m0} - \omega)} \right] e^{-i\omega t} F_\alpha^\omega \end{aligned} \quad (2.32)$$

where we have truncated at the linear response function of $\hat{\mu}$ perturbed by $\hat{V}(t)$. This is known as the *exact* linear response function. In general, response functions can be identified by the expansion of an operator $\hat{\Omega}$:

$$\begin{aligned} \langle \Psi(t) | \hat{\Omega} | \Psi(t) \rangle &= \langle 0 | \hat{\Omega} | 0 \rangle \\ &+ \langle \langle \hat{\Omega}; \hat{v}_\alpha^{\omega_1} \rangle \rangle e^{-i\omega_1 t} F_\alpha^{\omega_1} \\ &+ \langle \langle \hat{\Omega}; \hat{v}_\alpha^{\omega_1}, \hat{v}_\beta^{\omega_2} \rangle \rangle e^{-i(\omega_1 + \omega_2)t} F_\alpha^{\omega_1} F_\beta^{\omega_2} \\ &+ \dots \end{aligned} \quad (2.33)$$

where we have introduced notation for the linear $\langle \langle \hat{\Omega}; \hat{v}_\alpha^{\omega_1} \rangle \rangle$ and quadratic $\langle \langle \hat{\Omega}; \hat{v}_\alpha^{\omega_1}, \hat{v}_\beta^{\omega_2} \rangle \rangle$ response functions.

2.2.2 Approximate Response

Sum-over-states expressions like those in Eq. (2.32) can be solved by computing the excited-state wave functions and summing their individual contributions to the property. This poses a unique challenge in that the response tensor is not a product of only one targeted excited-state, but of all possible excited-states. While it is expected many high-energy states will have negligible contributions, the number of states required to accurately model some properties such as optical rotation makes this approach prohibitively expensive for excited-state extensions to ground-state methods, such as equation-of-motion (EOM) CC.³⁸ We can avoid this problem by building response equations based on approximate ground-state wave functions, such as Eq. (2.8). This eliminates the requirement of computing excited-state wave functions and allows us to replace the sum-over-states expression with a set of coupled linear equations which are far more computationally tractable.^{2,21}

Beginning with a time-dependent wave function $|\Psi(t)\rangle$, we may separate it into two time-dependent pieces - the exponential of a phase factor $\phi(t)$, and a phase-isolated wave function $|\bar{\psi}(t)\rangle$:

$$|\Psi(t)\rangle = e^{-i\phi(t)}|\bar{\psi}(t)\rangle. \quad (2.34)$$

We may require that the phase of the projection of $|\bar{\psi}(t)\rangle$ onto the ground-state wave function be zero - in other words, in the limit of zero time-dependent perturbation, $|\bar{\psi}(t)\rangle$ reduces to the ground-state wave function. This is analogous to our approach in Eq. (2.25), where our focus has now shifted from the wave function parameters to the phase factor. Inserting these definitions into the time-dependent Schrödinger equation, we arrive at Eq. (2.35):

$$(\hat{H} - i\hbar \frac{\partial}{\partial t})|\bar{\psi}(t)\rangle = Q(t)|\bar{\psi}(t)\rangle \quad (2.35)$$

where we have defined the time-dependent quasi-energy as the reduced Planck constant times

the derivative of the phase factor,

$$Q(t) = \hbar \frac{\partial \phi(t)}{\partial t}. \quad (2.36)$$

It is important to note that, just as the phase-isolated wave function reduces to the ground-state wave function in the absence of a perturbation, the quasi-energy also reduces to the ground-state energy in this case.

It can be shown that the quasi-energy is manifestly real, and both a time-dependent variational principle and Hellmann-Feynman theorem apply.²¹ To obtain a time-independent quantity to perturbatively expand, the quasi-energy is integrated over time to form the time-averaged quasi-energy Q_T . The variational and Hellmann-Feynman theorems can then be written

$$\partial Q_T = 0 \quad (2.37)$$

and

$$\frac{dQ_T}{d\hat{\Omega}} = \frac{1}{T} \int_{t'}^{t'+T} \langle \bar{\psi}(t) | \frac{\partial \hat{H}}{\partial \hat{\Omega}} | \bar{\psi}(t) \rangle dt, \quad (2.38)$$

respectively, for an arbitrary perturbation $\hat{\Omega}$ (usually taken to be an electromagnetic field). Thus, the quasi-energy is a well-defined dynamic analogue to the electronic energy.

In the preceding derivation, at no point was it assumed that we have *exact* eigenfunctions of the time-independent Hamiltonian. Thus, the above equations are valid for approximate ground-state theories, such as CC. We may now define our approximate response functions as we did in Eq. (2.22) as derivatives of this time-averaged quasi-energy,

$$\langle \langle \hat{\Omega}; \hat{v}_\alpha^{\omega_1} \rangle \rangle = \frac{d^2 Q_T}{d\hat{\Omega} dF_\alpha^{\omega_1}} \quad (2.39a)$$

$$\langle\langle \hat{\Omega}; \hat{v}_\alpha^{\omega_1}, \hat{v}_\beta^{\omega_2} \rangle\rangle = \frac{d^3 Q_T}{d\hat{\Omega} dF_\alpha^{\omega_1} dF_\beta^{\omega_2}} \quad (2.39b)$$

and working equations may be derived by inserting the specific approximate wave function ansatz.

The response functions of interest to the present work are the linear response functions between the electric dipole and an electric field $\langle\langle \hat{\mu}; \hat{\mu} \rangle\rangle$, and the magnetic dipole in an electric field $\langle\langle \hat{m}; \hat{\mu} \rangle\rangle$. These are responsible for the dynamic electric polarizability and chiroptical response (optical rotation and circular dichroism) respectively.² The latter constitutes a long-standing challenge for response theory,^{13,39} requiring mixed electric- and magnetic-field derivatives and leaving minimal room for error, but is also a prime candidate for comparisons to experiment. Some limitations of and alternatives to response theory will be explored in Chapters 3 and 5.

Chapter 3

Basis Set Superposition Errors in the Many-Body Expansion of Molecular Properties

Reprinted with permission from Peyton, B. G.; Crawford, T. D. *J. Phys. Chem. A* **2019**, *123*, 4500–4511. Copyright 2019 American Chemical Society.

3.1 Introduction

The accurate simulation of solvent effects on molecular optical response is exceptionally challenging because of the numerous dynamical factors required for a physically faithful model.^{40–48} Whereas many properties may be viewed as intrinsic to the solute subjected to a perturbation by the solvent environment (perhaps as represented by a dielectric continuum), others — particularly mixed electric-/magnetic-field responses^{1,12,39,49–51} — are, in fact, correctly viewed as inherent to the combined solute/solvent system. Schemes based on implicit solvent^{44,52–54} and frozen density embedding (FDE)^{55–59} have been used, but with limited success. Concomitant non-additivity effects, dynamic configurational sampling, and molecule-specific interactions serve to exacerbate the computational and theoretical demands on robust simulations.^{45,46} Thus, reducing the cost of these calculations for explicitly

solvated systems has become a major concern.^{54,60–62}

The many-body expansion (MBE) formalism has seen widespread success in predicting the energies of large molecular systems at a reduced cost, though recently the limits of these successes have been tested with clusters of increasing size and complexity.^{63–66} These limits are bounded by well-explored challenges like loss of precision^{64,66,67} and basis set superposition error (BSSE).^{63,66–75} Electrostatic embedding (EE) has been applied to MBE treatments of water clusters alongside BSSE corrections to improve convergence⁷³, as well as the N -body:Many-body QM:QM technique⁷⁶ to reduce the cost of including higher-order effects. The latter has also been applied to small water clusters for predicting vibrational frequencies.⁷⁷

The fundamental concept behind the MBE is the decomposition of the energy (or other properties) into a sum of smaller contributions from sub-components of a complex, interacting system.^{78,79} In the MBE, the energy (or energy derivative) of n interacting fragments may be expanded in orders of interaction energies:

$$E_{1,2,\dots,n} = \sum_{i \subseteq N_1} E_i + \sum_{ij \subseteq N_2} \epsilon_{ij} + \sum_{ijk \subseteq N_3} \epsilon_{ijk} + \dots, \quad (3.1a)$$

where N_1 , N_2 , N_3 , etc., denote the sets of unique monomers, dimers, trimers, etc. The interaction energies are defined as, for example, for dimers,

$$\epsilon_{ij} = E_{ij} - E_i - E_j, \quad (3.1b)$$

and for trimers,

$$\epsilon_{ijk} = E_{ijk} - (E_{ij} + E_{ik} + E_{jk}) + (E_i + E_j + E_k), \quad (3.1c)$$

where the subtraction of energies of sub-components is necessary to avoid overcounting of

energy contributions. While the untruncated MBE scheme is formally exact, large computational savings result from the truncation of Eq. (3.1a) to k -body terms where $k < n$. Truncation to two- or three-body terms only has proved sufficient for interaction energies in many examples. However, limiting cases exist for increasingly complex systems.^{63–66}.

While interaction energies are the popular target of MBE applications, properties computed using the MBE have received comparatively little attention. Though some work has been carried out concerning induced electronic properties in linear species,^{80–84} solvated systems such as those studied in Ref. 85 are of particular interest for chiroptical property prediction. A study by Mach and Crawford⁸⁶ showed that the MBE for a number of solvated chiral systems exhibited oscillatory convergence in specific rotation with respect to n -body truncation, while interaction energies, dipole moments, and dynamic polarizabilities still converged well by the three-body approximation as is typical for interaction energies with the MBE for systems with fewer than 12 monomers^{63,66}.

The successive sign-flips of k -body terms, such as E_i changing from positive to negative in Eqs. (3.1b) and (3.1c), have been suggested to be the cause of oscillations in the MBE for large clusters,⁶³ and BSSE has been identified as a major contributing factor to these oscillating errors. BSSE is a result of an imbalance of basis functions and thus will be present in any electronic structure calculation using a finite number of functions. The BSSE of a dimer interaction term in Eq. (3.1b) is relatively straightforward to correct using the Boys and Bernardi Counterpoise Method (BBCP):⁶⁸

$$\epsilon_{ij} = E_{ij} - E_i^{(ij)} - E_j^{(ij)}, \quad (3.2)$$

where the superscript denotes the basis set used, and any terms without a superscript are calculated in their own basis. When E_i and E_j are calculated in just their own basis sets, each

monomer does not benefit from the nearby basis functions placed on the partner monomer. However, when calculating E_{ij} , the basis functions are shared between the monomers, which increases flexibility in the wavefunction and changes the energy despite having little bearing on the interaction between monomers i and j , only their (incomplete) basis sets. By using the same basis (that of the dimer ij) for all three calculations, there is no imbalance in the wavefunction, hence the term “counterpoise (CP) correction”.

The BBCP method was originally intended for correcting dimer interaction energies, but two fundamentally different generalizations of the BBCP correction scheme for an n -body interaction term are currently in use: the site-site function counterpoise (SSFC) and Valiron-Mayer Function Counterpoise (VMFC) methods.^{69,70} The former, also referred to as the “full cluster basis,” simply uses the basis functions of the full n -body cluster for each fragment calculation,

$$E_{1,2,\dots,n} = \sum_{i \subseteq N_1} E_i^{(n)} + \sum_{ij \subseteq N_2} \epsilon_{ij}^{(n)} + \sum_{ijk \subseteq N_3} \epsilon_{ijk}^{(n)} + \dots \quad (3.3)$$

with similar generalizations for Eq. (3.1b) and (3.1c). While often prohibitively expensive, it has been shown to be effective in eliminating oscillations in the MBE.^{63,66}

The VMFC method is based on correcting for BSSE at each k -body level such that the k -mer basis is used for each k -body interaction term,

$$E_{1,2,\dots,n} \approx \sum_{i \subseteq N_1} E_i^{(i)} + \sum_{ij \subseteq N_2} \epsilon_{ij}^{(ij)} + \sum_{ijk \subseteq N_3} \epsilon_{ijk}^{(ijk)} + \dots \quad (3.4)$$

By correcting at each k -body term, the VMFC method prevents spurious “ghost dipoles” from appearing in the k -body approximations, which can occur with the SSFC method: asymmetrically distributed basis functions placed around a real fragment can cause electron density to move to those locations, causing net dipoles (and possibly other properties) which are solely dependent on the placement of ghost functions, though such effects should only

appear significantly for close-packed fragments. The VMFC method requires that each calculation be performed in multiple basis sets, resulting in many more calculations (an additional binomial coefficient, in fact) than the classic MBE. The exact nature of the MBE at n -body is also lost, e.g. the monomer energies in Eq. (3.1a) no longer cancel exactly with the monomer energies in Eq. (3.1b), due to the difference in basis sets. The resulting energy is a “counterpoise corrected” energy, which makes benchmarking relative to complete cluster calculations difficult. In practice, results from the VMFC method vary only slightly relative to those from the SSFC method and other correction schemes^{66,71,73}.

The goal of the present study is to understand the cause of the oscillations in the MBE for specific rotation observed in the earlier work by Mach and Crawford⁸⁶ and to determine whether or not they can be removed using a BSSE correction scheme. We will focus on the SSFC method due to its convergence to the correct n -body limit. We have chosen to exclude the VMFC method at present because of its high computational cost and ambiguity in benchmarking calculations. We further extend the previous work by employing larger solvation clusters and testing additional configurations of explicitly solvated systems.

3.2 Computational Details

The dynamic polarizability at an external field frequency, ω , is computed as the isotropic average of the dynamic polarizability property tensor:^{1,12}

$$\alpha_{\alpha\beta}(\omega) = \frac{2}{\hbar} \sum_{j \neq n} \frac{\omega}{\omega_{jn}^2 - \omega^2} \text{Re}(\langle n | \mu_\alpha | j \rangle \langle j | \mu_\beta | n \rangle), \quad (3.5)$$

where n is the electronic ground state, j is an electronic excited state with excitation frequency ω_{jn} , and μ is the electric dipole operator. Similarly, the specific rotation (in

deg dm^{-1} (g/mL^{-1}) is related to the isotropic average of the electric-dipole/magnetic-dipole property tensor (also called the Rosenfeld tensor):⁸⁷

$$G'_{\alpha\beta}(\omega) = -\frac{2}{\hbar} \sum_{j \neq n} \frac{\omega}{\omega_{jn}^2 - \omega^2} \text{Im}(\langle n|\mu_\alpha|j\rangle\langle j|m_\beta|n\rangle). \quad (3.6)$$

The $\boldsymbol{\alpha}$ and \mathbf{G}' property tensors were computed in this work using the linear response formalism²¹, the latter using gauge-including atomic orbitals (GIAOs)⁸⁸. Dynamic response properties such as polarizabilities or specific rotations may be formulated in terms of the time-averaged quasi-energy,

$$\mathcal{Q} = \frac{1}{T} \int_0^T \langle \bar{\Psi} | \left(\hat{H} - i \frac{\partial}{\partial t} \right) | \bar{\Psi} \rangle dt, \quad (3.7)$$

where $\bar{\Psi}$ is the regular part of the phase-isolated wave function⁸⁹ and \hat{H} includes the time-dependent external field. For periodic potentials, the time-averaged quasi-energy is uniquely defined, and both variational and Hellmann-Feynman theorems apply.⁹⁰ Indeed, in the limit of a time-independent Hamiltonian, the quasi-energy reduces to the energy of the stationary state. Thus, just as the many-body expansion applies to the time-independent energy, it also applies to the time-averaged quasi-energy. Furthermore, since response functions are computed as derivatives of \mathcal{Q} with respect to the perturbation coefficients (e.g., the polarizability is the negative of the second derivative of the time-averaged quasi-energy with respect to the strength parameter of the external electric field), response functions should also be subject to the many-body expansion, just as in the time-independent case. This was the driving force behind the Mach paper, which also showed that this was correct for one linear response property, the polarizability⁸⁶. In the following, it will be shown that while this expansion is valid, its effective truncation depends highly on the additivity of the subjected property. CP corrections should remove BSSE errors which do not cancel in the MBE equations, and

other potential issues will be explored for the non-convergent specific rotation.

Geometries for (*S*)-methyloxirane in a cage of seven water molecules, (*S*)-methylthiirane surrounded by six water molecules, and (*M*)-dimethylallene with seven water molecules were re-used from the previous study⁸⁶ for consistency. These geometries were generated by Gromacs⁹¹ simulations of each solute in water, and snapshots with 5.5 Å solvent shells were extracted from the resulting trajectories. Additional geometries for (*S*)-methyloxirane in seven- and 13-water-molecule cages and (*S*)-methylthiirane in a six water-molecule cage were generated to test the effects of geometry and solvent shell size on MBE convergence and BSSE. All geometries are available in the Supporting Information.

Interaction energies, dipole moments, and specific rotations were calculated using the B3LYP^{92,93} functional in the aug-cc-pVDZ (aDZ) basis^{94,95} as in the previous study. Additionally, the aug-cc-pVTZ (aTZ) basis set was explored in selected examples to determine the effects of basis set size on the MBE convergence and the BSSE. The CAM-B3LYP⁹⁶ functional was also employed to explore the effects of long-range interactions on the \mathbf{G}' tensor. Dipole polarizabilities for all systems (and all properties for the 13-water/*S*-methyloxirane system) were computed using only CAM-B3LYP/aug-cc-pVDZ. Specific rotations and dynamic polarizabilities were computed at four common wavelengths: 355, 436, 589, and 633 nm. All calculations were performed using Gaussian 09⁹⁷ using inputs generated by a Psi4⁹⁸ plugin, which also carried out the subsequent data collection and analysis. This plugin gathers the data from formatted checkpoint files generated by Gaussian 09 to address precision issues brought about by propagation of error as noted in recent studies.^{64,66,67} Default SCF and CPHF convergence criteria (10^{-7} and 10^{-10} , respectively) were used with standard pruned integration grids, which include 75 and 35 radial shells, and 302 and 110 angular points per shell for SCF and CPHF, respectively. When compared to tighter convergence (10^{-12}) and “fine” grids (pruned 75 shell, 302 nodes) for both SCF and CPHF, no appreciable difference

was found for the (*S*)-methyloxirane system in a seven-water solvent shell (compare Figs. 7 and 13 of the Supporting Information). Additionally, unpruned “fine” grids were also tested with the same system; similarly, this had little effect on the convergence of the expansion (compare Figs. 7 and 14 of the Supporting Information).

We will use both graphical evaluations and standard deviations of the error relative to the converged results to assess the oscillations present in the MBE. We calculated standard deviations only for the two-body and higher approximations, due to the much larger, non-representative errors associated with the one-body approximation. We report standard deviations and percent errors for dynamic polarizabilities and specific rotations at 633 nm unless otherwise noted. Plots of data not discussed, such as the interaction energies and electric dipole moments of smaller solvent cages, are available in the Supporting Information.

3.3 Results and Discussion

3.3.1 Interaction Energies and Electric Dipole Moments

Interaction energies and electric dipole moments computed with the MBE were reported in the previous study⁸⁶ and extensive reviews of the general behavior of the MBE for such properties are available elsewhere (e.g. Ref. 66 and references therein). We found that the SSFC correction was insignificant for all cases (see the Supporting Information), except for methyloxirane in a 13-water solvent shell, Fig. 3.1. As noted in previous studies of \sim ten-body expansions^{63,66}, oscillating errors in the convergence of the interaction energy in the milli-Hartree range were observed for the methyloxirane system considered. While these oscillations are not as significant as those seen for other properties, the improvement in convergence by the SSFC correction as evidenced by Fig. 3.1a is still noteworthy. Note

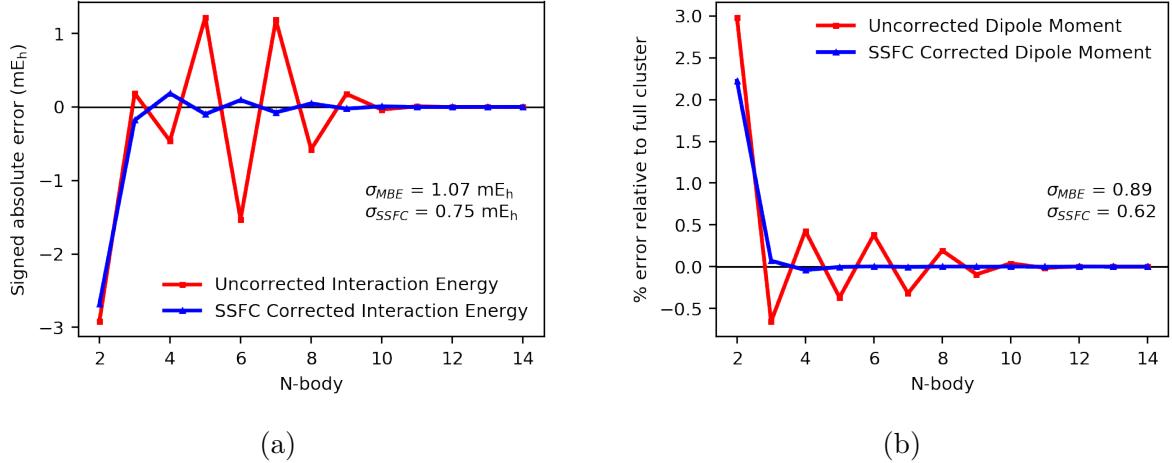


Figure 3.1: MBE and SSFC of (a) interaction energies and (b) dipole moments for (*S*)-methyloxirane in a 13-water solvent shell.

that, by definition, the MBE and SSFC interaction energies are different due to the different basis sets used for the monomer terms, so the absolute errors relative to the MBE or SSFC interaction energy are reported. The dipole moment in Fig. 3.1b also exhibits oscillations uncharacteristic of the smaller solvent shells (nearly 0.5% error even at six-body), and again the SSFC correction dampens the oscillations dramatically. This suggests that the BSSE inherent in the MBE for interaction energies — and its characteristic increase relative to system size — is present for other properties as well. If the oscillations observed for higher-order properties in the previous study are indeed indicative of BSSE, then CP corrections should play a major role in damping them.

3.3.2 Dynamic Dipole Polarizabilities

Dipole polarizabilities calculated with the MBE were reported for solvated (*S*)-methyloxirane, (*M*)-dimethylallene, and (*S*)-methylthiirane in the previous study⁸⁶ using the B3LYP functional. Their oscillatory behavior has been reproduced here with the CAM-B3LYP functional and CP-corrected with the SSFC method, as shown in Fig. 3.2. For all three systems, the

CHAPTER 3. BASIS SET SUPERPOSITION ERRORS IN THE MANY-BODY EXPANSION OF MOLECULAR PROPERTIES
30

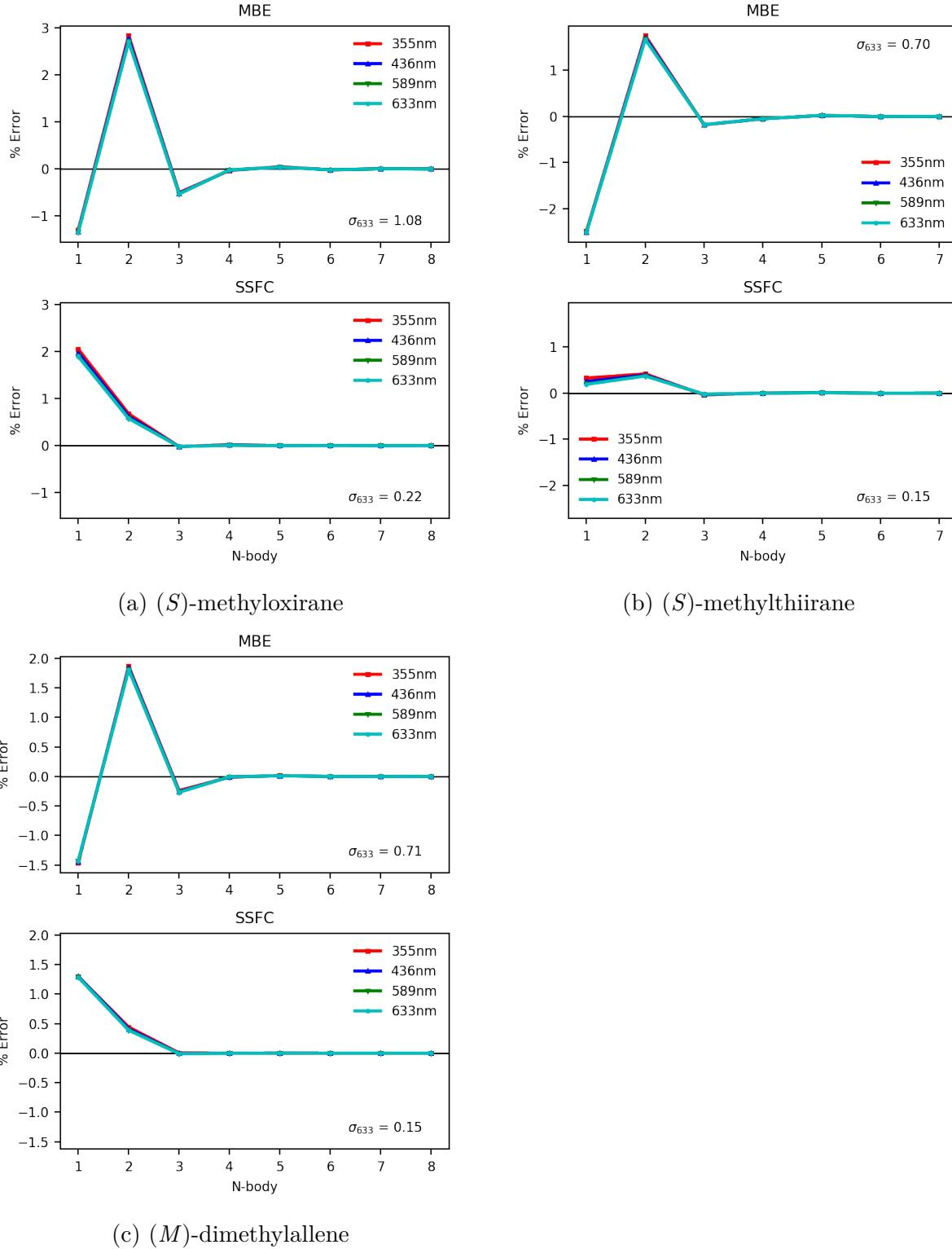


Figure 3.2: MBE and SSFC of dynamic polarizabilities for (a) (S)-methyloxirane in a seven-water solvent shell, (b) (S)-methylthiirane in a six-water solvent shell, and (c) (M)-dimethylallene in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.

relatively large oscillations of the MBE for one- to four-body contributions were almost completely removed by SSFC, replacing the convergence trend with a nearly monotonic decaying function. Analysis of a second snapshot taken from a molecular dynamics trajectory for (*S*)-methyloxirane and (*S*)-methylthiirane produce similar results (see the Supporting Information), with only that of (*S*)-methyloxirane exhibiting a slightly negative three-body term before reaching convergence. In all three systems, the SSFC correction yields substantial improvement in the convergence of the MBE as compared to the uncorrected results, e.g. the standard deviation for (*S*)-methyloxirane decreases from 1.08 to only 0.22.

Dynamic polarizabilities were also calculated for the larger 13-water (*S*)-methyloxirane system, as shown in Fig. 3.3. Oscillations to a higher error (up to 6%) resulted in a higher standard deviation in this case than for the smaller solvent systems considered for the MBE. However, the SSFC still damped these oscillations significantly, with some small oscillations remaining. As with the electric dipole moment data considered earlier, BSSE is shown to be present and perhaps more important for properties than for interaction energies of the same system, with SSFC being a consistent method of reducing oscillations and speeding up convergence regardless of system size.

3.3.3 Specific Rotations

As stated previously, the properties considered thus far can be thought of as intrinsic to the solute being perturbed by the solvent. MBEs for such properties, even beyond the energy, still converge, and oscillations in the convergence for large clusters are mostly due to BSSE and therefore can be corrected with CP-corrections such as SSFC or truncated schemes (such as VMFC(n)⁷² or MBCP(n)⁷³). It is shown that BSSE is perhaps more important in response property calculations; however, CP corrections still recover convergence within

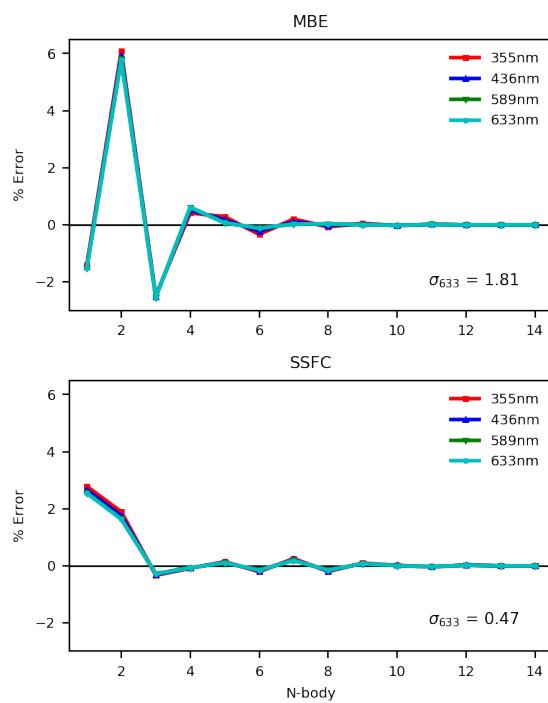


Figure 3.3: MBE and SSFC of dynamic polarizabilities for (*S*)-methyloxirane in a 13-water solvent shell. Computed with CAM-B3LYP/aDZ.

a three body approximation for even the largest solvent cluster considered. Next, the oscillatory convergence of the highly non-additive specific rotation of these clusters will be investigated, with the goal of deciding whether an MBE is an appropriate approximation even with a CP correction.

The earlier work by Mach and Crawford⁸⁶ identified specific rotations as a particular challenge for the MBE and its failure to converge with even small solvent shells suggested that BSSE was a potentially significant source of error. However, a number of other potential parameters exist that could also cause or exacerbate the observed erratic behavior for this property. As Ouyang *et al.* pointed out⁶³, oscillations in the MBE can occur due to successive sign-flips of terms with large errors; as a result, oscillations could potentially occur due to any large bias in the subsystem calculations. To that end, we selected (*S*)-methyloxirane and (*S*)-methylthiirane as representative cases for investigating various strategies for addressing these oscillations, though we also report results for (*M*)-dimethylallene at the B3LYP/aDZ and CAM-B3LYP/aDZ levels of theory.

One possible source of convergence problems for the MBE is noise due to the precision of the collected \mathbf{G}' property tensor elements. Recent publications^{64,66,67} have taken notice of the effects of subsystem calculation convergence criteria and precision on higher-order terms in the MBE. Richard, Lao, and Herbert demonstrated in Ref. 64 that precision errors in the energy rise rapidly for four- and five-body calculations as the total system size increases. We accounted for this by using Gaussian's formatted checkpoint files for the data collection to ensure that we maintain full machine precision until the final result. Nevertheless, significant oscillations in the MBE results for specific rotations remained with little to no difference from the previous results⁸⁶, as shown in Fig. 3.4. (NB the significant difference in the vertical scale for the three test cases.) Indeed, the specific rotation of (*S*)-methylthiirane approaches 500% error at the three-body contributions as compared to the converged results for the

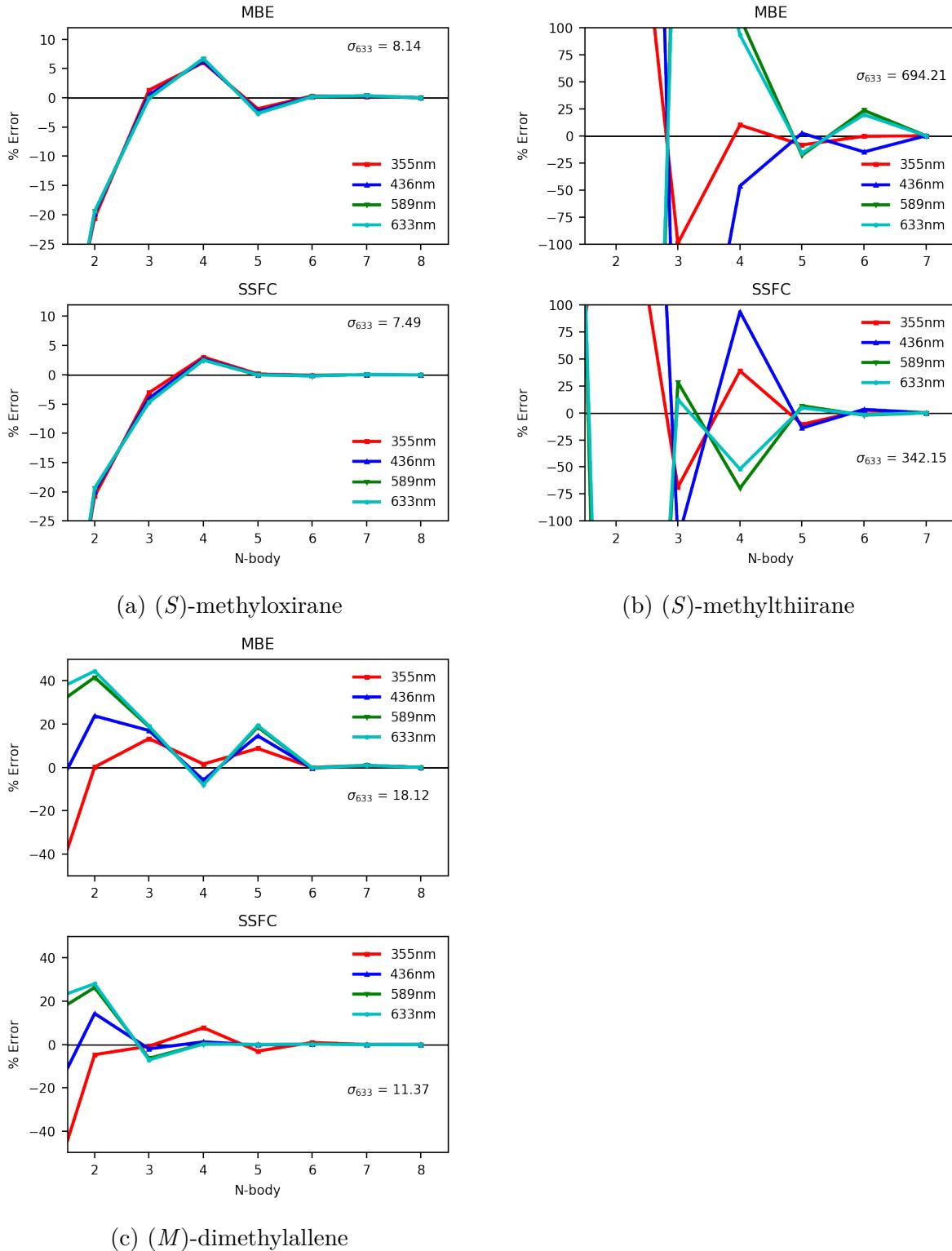


Figure 3.4: MBE and SSFC of specific rotation for (a) (S)-methyloxirane in a seven-water solvent shell, (b) (S)-methylthiirane in a six-water solvent shell, and (c) (M)-dimethylallene in a seven-water solvent shell. Computed with B3LYP/aDZ.

three longer wavelengths considered, in agreement with Table 8 of Ref. 86, and thus the precision of the individual terms is ruled out as a contributor to this behavior.

A second potential source of MBE convergence error is the use of the relatively small aDZ basis set. Although this basis set has been found to be adequate for many applications in studies of optical activity, for some systems (notably, methylthiirane), larger basis sets are needed.^{54,99} To investigate this issue, we carried out B3LYP/aTZ specific rotation calculations for both (*S*)-methyloxirane and (*S*)-methylthiirane, the results of which are presented in Fig. 3.5. For (*S*)-methyloxirane, the four- and five-body errors with the aDZ basis (Fig. 3.4a) exhibit a significant oscillation from ~7% error to ~-3% error; the aTZ basis (Fig. 3.5a) reduces this somewhat to ~5% to -0.7% error, and the two- to *n*-body standard deviation slightly decreases from 8.1 to 6.8% error. Comparing Figs. 3.4b and 3.5b for (*S*)-methylthiirane, increasing the basis set from aDZ to aTZ similarly offers no significant qualitative improvement, as errors vary wildly and are greater than 15% for all wavelengths even at five-body contributions (and worse than the aDZ basis in some cases). The standard deviation at 633 nm does decrease from 694 to 159% error with the improved basis set, though this is still well beyond acceptable limits and convergence with respect to *n*-body truncation remains strongly dependent on wavelength.

What about the choice of density functional? While most studies to date have utilized the popular B3LYP functional, CAM-B3LYP has been proven useful to reduce errors in the near-resonance regions^{47,100} and to produce more consistent results with respect to wavelength, as shown in Fig. 3.6b. While CAM-B3LYP for (*S*)-methyloxirane does not yield significant differences in the MBE (cf. Fig. 3.4a with Fig. 7 of the Supporting Information), the change in functional makes a considerable difference for (*S*)-methylthiirane (Fig. 3.6a), as convergence trends were more consistent across all wavelengths tested. Nevertheless, the MBE still never converges to below 5% error until the full expansion is reached, and a standard deviation

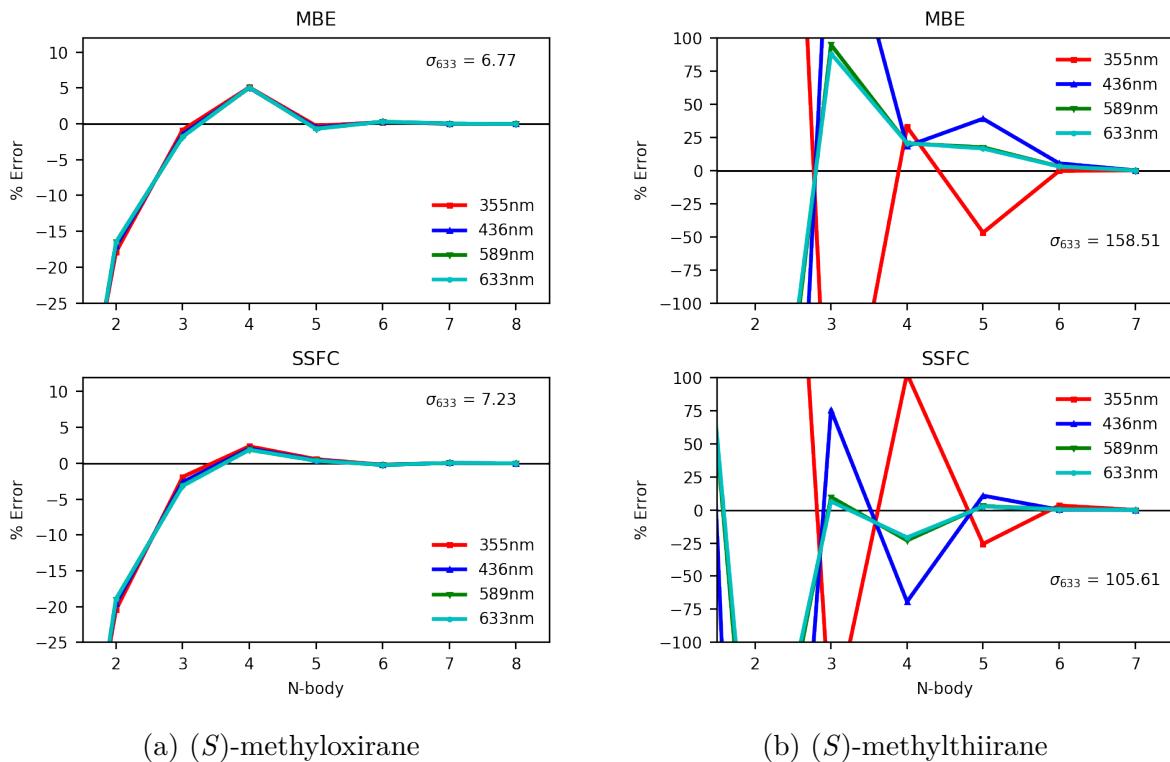


Figure 3.5: MBE and SSFC of specific rotation for (a) (*S*)-methyloxirane in a seven-water solvent shell and (b) (*S*)-methylthiirane in a six-water solvent shell. Computed with B3LYP/aTZ.

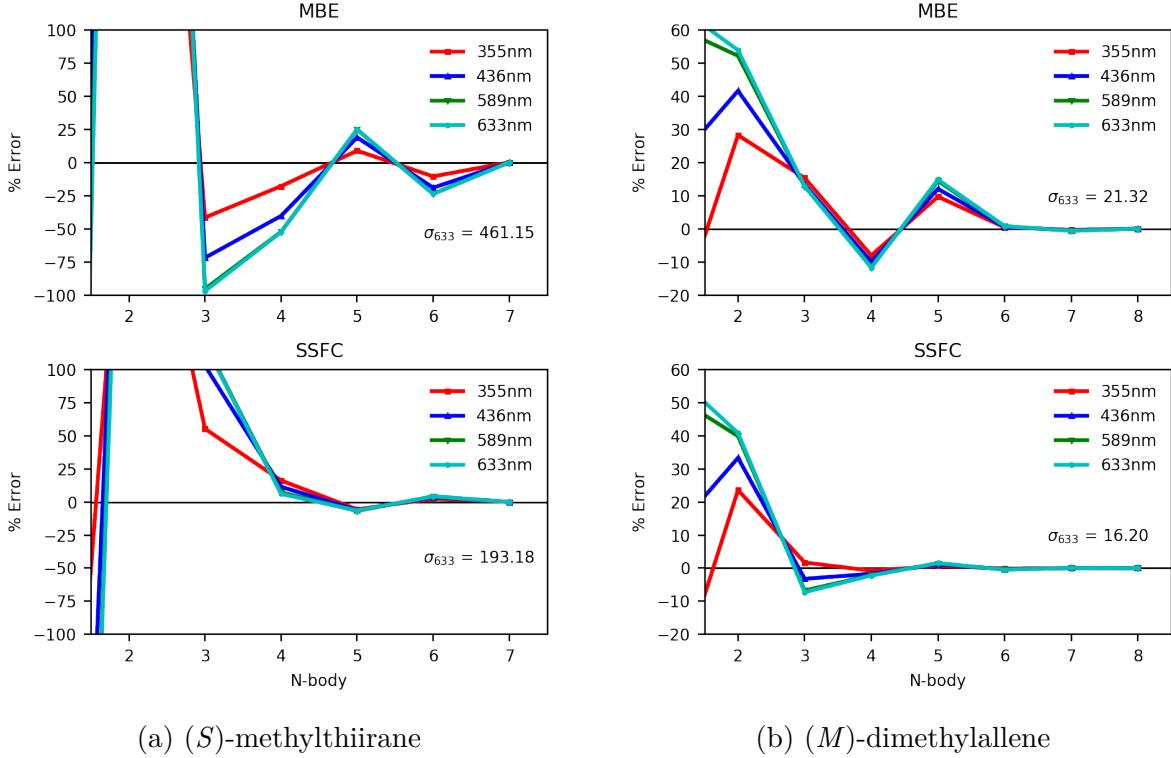


Figure 3.6: MBE and SSFC of specific rotation for (a) (*S*)-methylthiirane in a six-water solvent shell and (b) (*M*)-dimethylallene in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.

of 461% error was still observed (primarily due to the two-body error which is over 1000% error). (*M*)-dimethylallene also exhibits more consistent convergence trends with CAM-B3LYP, with only a slightly increased standard deviation (18% to 21% error).

The final non-BSSE issue considered is the choice of snapshot geometry. Calculations of solvent-phase specific rotation based on MD trajectories typically employ a large number of such snapshots, each with a distinct geometry that may be far from an energetic local minimum. Thus, we considered an additional snapshot for both (*S*)-methyloxirane and (*S*)-methylthiirane in their seven- and six-water solvation shells respectively, using CAM-B3LYP/aDZ (Fig. 3.7). For (*S*)-methyloxirane, the MBE performs somewhat worse for the new snapshot, with errors approaching from far below zero and changing sign at the six-

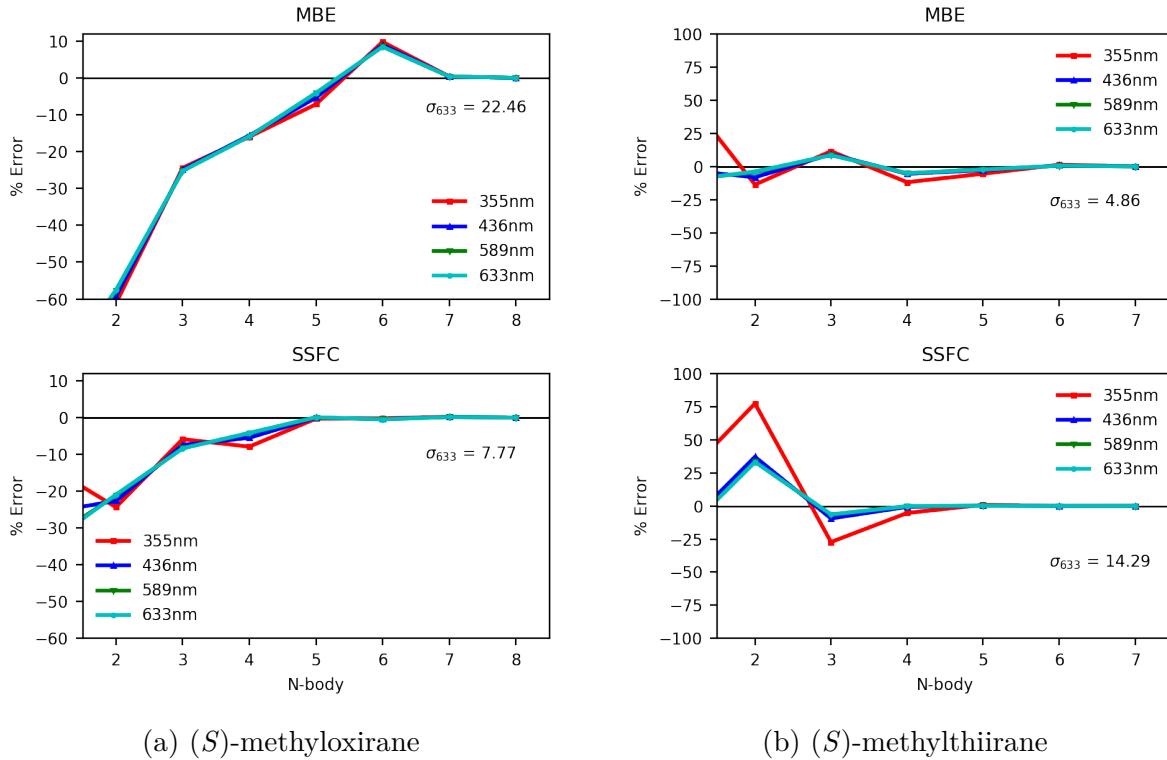


Figure 3.7: MBE and SSFC of specific rotation for additional snapshots of (a) (*S*)-methyloxirane in a seven-water solvent shell and (b) (*S*)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.

body level with errors of nearly 10% still remaining, and a considerably larger 22% error standard deviation, though the same qualitative trends appear. The new snapshot for (*S*)-methylthiirane, on the other hand, yields somewhat better results, but again with the same qualitative trends. The large errors near (and over) 100% seen previously are not present for this snapshot, though oscillations are still clearly visible with respect to *n*-body truncation, as evidenced by the 4.9% error standard deviation. Additionally, the original 7-water (*S*)-methyloxirane snapshot was edited to contain two short O...H bonds to test the effects of hydrogen bonding distance on the convergence (compare Figs. 7 and 12 of the Supporting Information). As with the second snapshot, the same qualitative trends were observed, with very little difference in percent errors.

Next, in order to focus on the impact of BSSE on the MBE, we employed the SSFC CP-correction scheme in each of the test cases discussed above. In all cases, the corrections reduce — but do not eliminate — the oscillations in the relative errors of the MBE, as evidenced by the lower frame of every sub-figure of Figs. 3.4-3.7. Standard deviations also generally agree with this trend, with exceptions being Fig. 3.5a and 3.7b. In those two cases, the slightly higher two-body error in the SSFC case inflates the standard deviation, but from three-body onwards the oscillations are reduced. For the smaller solvation shells of (*S*)-methyloxirane, (*M*)-dimethylallene, and (*S*)-methylthiirane, the convergence was also accelerated: by three- or four-body contributions, errors reasonable enough for predicting the sign of the optical rotation (<5%) were achieved for most wavelengths with CAM-B3LYP/aDZ, though (*S*)-methylthiirane remains a challenge (16% error at the four-body truncation for 355 nm), as well as 355 nm and 436 nm for the second methyloxirane snapshot (-8% and -6% at four-body). However, despite the relative improvements compared to the uncorrected MBE, these results do not imply that rapid convergence of even a CP-corrected MBE should be expected for highly non-additive properties such as specific rotation. Furthermore, a qualitative description of the specific rotation will not suffice: proper simulations of the property should be averaged over possibly hundreds of snapshots from an MD trajectory, and for complex systems with multiple chiral centers even a small error in the specific rotation can predict the wrong stereoisomer¹⁰¹. The fact that these trends also hold for (*M*)-dimethylallene speaks to the generality of the conclusions made thus far. Chirality in this compound is induced by a stereogenic axis, rather than a stereogenic center such as that in methyloxirane. In addition, the lowest energy excited states for dimethylallene are $\pi \rightarrow \pi^*$ valence transitions localized near the double bonds, as opposed to methyloxirane whose lowest excitation energies are Rydberg states¹⁰².

The 13-water solvated (*S*)-methyloxirane in Fig. 3.8, however, presents an even larger chal-

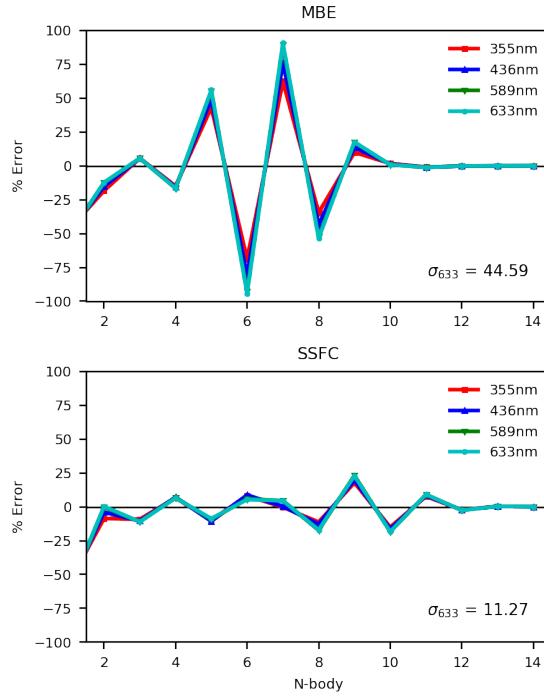


Figure 3.8: MBE and SSFC of specific rotation for (*S*)-methyloxirane in a 13-water solvent shell. Computed with CAM-B3LYP/aDZ.

length, despite including the SSFC correction. Encouragingly, errors slightly above 5% were found at the four-body truncation, but then the error continues to oscillate without converging until the 12-body truncation. While the SSFC greatly outperforms the uncorrected MBE in the five- to seven-body range (the MBE presents errors between 50 and 100%, while SSFC presents <10% in this range), it is actually worse from the nine-body truncation onwards. The eight-, nine-, and ten-body truncations increase substantially to roughly +/- 20% error, while the MBE converged to <10% by nine-body. This is one of the largest solvated property calculations performed with the MBE to date, and the results undermine its usefulness. This, combined with the substantial cost of the SSFC correction, greatly hinders the feasibility of the approach for production-level calculations of specific rotation for such large clusters, which may be necessary to properly model the effects of the solvent.

In cases of achiral solvents, such as water, one way to reduce the expense of the MBE for

larger solvent shells is to include only contributions due to the solute and solute-solvent interactions. This is equivalent to restricting the sums in Eq. (3.1a) to only include fragments that contain the chiral solute. In principle, this should be advantageous for specific rotation calculations, as a pure achiral solvent (such as water) should give an optical rotation of zero upon averaging over a large number of snapshots and/or sufficiently large solvation clusters. Thus, neglecting solvent-solvent contributions should be of little consequence in the context of a sufficient MD trajectory while perhaps reducing the cost of the MBE through the removal of roughly half of the calculations needed for the full expansion of a solvated system (and much more for truncated expansions). For the smaller solvent shells of (*S*)-methyloxirane and (*S*)-methylthiirane (Fig. 3.9), the removal of solvent-solvent interactions has little impact from five-body interactions onwards. However, the four-body approximation is slightly worsened even with the SSFC correction for (*S*)-methylthiirane (though the standard deviation decreases). Using this approach, the total number of calculations necessary for the full expansion decreases from 255 to 128 unique electronic structure calculations for (*S*)-methyloxirane, and, at any truncated *n*-body approximation, this number would decrease even more. For large systems in which only a qualitative description of the specific rotation is necessary, these errors may be acceptable for such an improvement in the computational cost.

In the case of the 13-water solvent shell for (*S*)-methyloxirane, excluding solvent-solvent interactions has a significant positive impact on the magnitude of oscillations in the MBE (Fig. 3.10). While the standard deviation increases for the SSFC, the convergence in the troublesome eight- to ten-body range is improved, as compared to Fig. 3.8. Furthermore, the number of unique electronic structure calculations for the full expansion reduces by roughly a factor of two — 16383 to 8192 — with massive savings possible for truncated expansions. Still, due to the significant non-BSSE oscillations in the MBE, the possibility

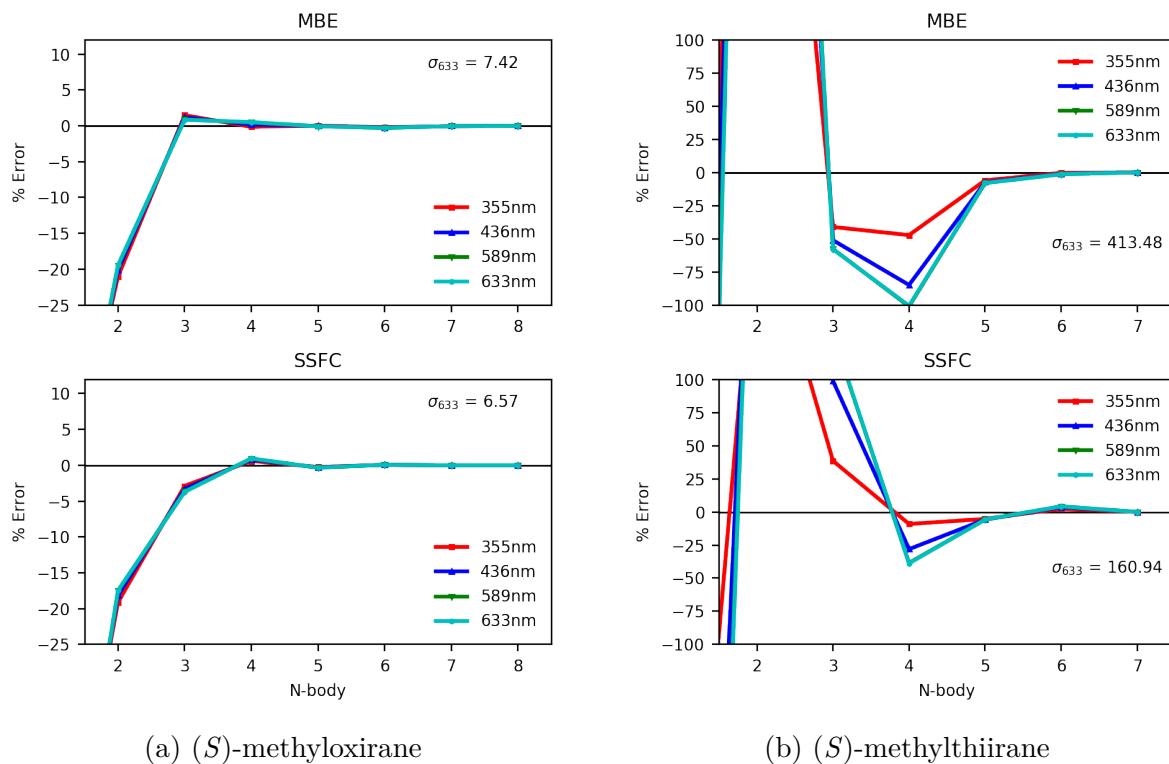


Figure 3.9: MBE and SSFC of solute-fragment-only specific rotation for (a) (*S*)-methyloxirane in a seven-water solvent shell and (b) (*S*)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.

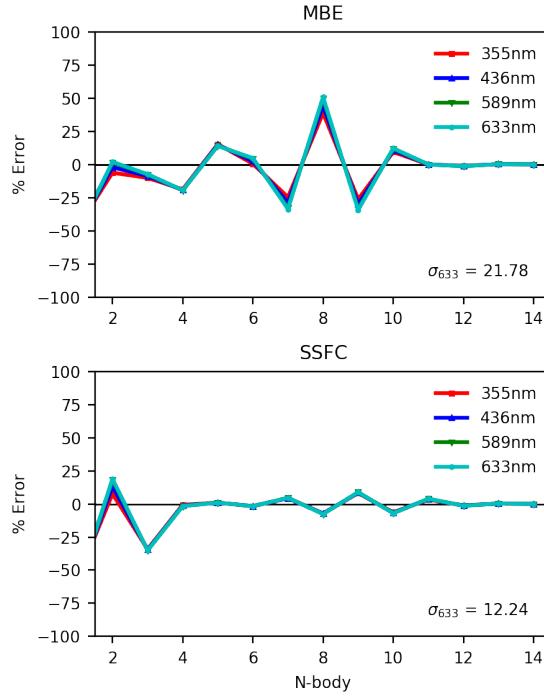


Figure 3.10: MBE and SSFC of solute-fragment-only specific rotation for (*S*)-methyloxirane in a 13-water solvent shell. Computed with CAM-B3LYP/aDZ.

of false convergence, and the prohibitive cost of the SSFC correction, a CP-corrected MBE cannot be recommended as a reliable low-cost quantitative approximation to the specific rotation of the full-cluster.

3.4 Conclusions

We find that the SSFC serves to reduce significantly the BSSE in MBE expansions of higher-order properties such as dipole polarizabilities and specific rotations, resulting in somewhat dampened oscillations and thus smoother convergence in most cases. Computed standard deviations decrease on average with inclusion of SSFC corrections, with the only exceptions being (*S*)-methylthiirane in a six-water solvent shell computed at the CAM-B3LYP/aDZ level of theory and (*S*)-methyloxirane in a seven-water solvent shell computed at the B3LYP/aTZ

level of theory due to the erratic behavior of the MBE at the two-body correction. Furthermore, these trends held for variations in the choice of MD snapshot, density functional, and basis set. We thus conclude that, in agreement with previous studies on interaction energies,^{63,66,67} counterpoise corrections are essential for the MBE calculation of response properties, whose magnitude (and sign) often depend heavily on the diffuse regions of electron density most prone to BSSE. However, the convergence difficulties characteristic of properties such as specific rotations remain a challenge, exhibiting oscillations in the MBE that cannot be remedied by a counterpoise correction alone. As illustrated clearly by the example of (*S*)-methyloxirane in a solvent shell of 13 water molecules, significant oscillations remain even after BSSE corrections, and these are inherent to non-additive, non-perturbative properties such as mixed electric-/magnetic-field responses. Significant changes in the subsystem calculations as n increases can cause sign flips of successive terms, and small changes in the subsystem can cause large changes in the computed specific rotation contribution. In short, the rapid convergence of the MBE inherently assumes that the property in question is fundamentally additive, in this case that the property may be viewed as that of the solute with relatively small perturbations arising from the nearby solvent molecules. This requirement does not hold for specific rotations (and related chiroptical responses), leading to the conclusion that the MBE is of very limited utility for such cases.

Chapter 4

Machine-Learning Coupled Cluster Properties through a Density Tensor Representation

Reprinted with permission from Peyton, B. G.; Briggs, C.; D'Cunha, R.; Margraf, J. T.; Crawford, T. D. *J. Phys. Chem. A* **2020**, *124*, 4861–4871. Copyright 2020 American Chemical Society.

4.1 Introduction

Machine-learning (ML) has been applied to a number of chemical applications, with excellent (and current) surveys of the field published in 2019^{103,104}. While applications in high-throughput chemical screening using computationally “free” descriptors of molecular systems (e.g. the molecular formula or graph) have been fruitful, recently much work has been focused on machine-learning quantum mechanics (MLQM). These methods generally apply machine-learning to either replace or bolster the often expensive algorithms used in quantum mechanics for solving the electronic Schrödinger equation. Schemes for ML-assisted models include learning force-fields or other parameterizations from *ab initio* data^{105,106}, improved guesses for e.g. important contributions to configuration-interaction space¹⁰⁷ or amplitudes

in coupled cluster (CC) theory¹⁰⁸, and direct energy or property prediction using mean-field, correlated, or density-functional-based approaches^{109–113}.

The focus of this work will be on the bolstering of wave function-based methods, such as CC theory¹⁰ and many-body perturbation theory (MBPT)^{114,115}, for predicting energies and properties across potential energy surfaces or molecular dynamics (MD) trajectories. This is an application also covered by recent work facilitating ML for forces in *ab initio* MD¹¹⁶, fast numerical gradients for geometry optimization¹¹⁷, and general-purpose local potential energy surfaces (PESs)¹¹⁸. Note, however, that these methodologies may be extended to other applications, such as learning of the wave function itself and predicting properties across large chemical datasets as described in Refs. 119 and 120, respectively.

There has also been some recent interest in alternative, wave function-based representations (as opposed to the more common representations based on the molecular structure), which in principle have several advantages. In particular, we define three desirable properties: first, representations derived from increasingly accurate electronic structure methods should provide more faithful representations of the wave function, resulting in more accurate and efficient ML. This rules out features that are not systematically improvable, such as the Hartree-Fock wave function parameters used by Miller and co.^{109,110} Second, the representation should be of reasonable size to be stored for thousands of molecules, with the possibility of reducing their size for extremely large datasets. This rules out a real space representation of the electron density on a grid¹²¹. Third, a solid theoretical motivation for the representation should be given, with simple mappings from representation to target preferred. This third point is critical, as we will show that the relationship between the electronic wave function (specifically, the corresponding density matrix) and properties expressed as expectation values of one- and two-electron operators suggests a clear recipe for a wave function representation which is appropriate for machine-learning any molecular response property.

One such wave function-based representation is the t-amplitude tensor representation (TATR) introduced by Margraf and Reuter¹²². This representation is based on CC or MP2 amplitudes, paired with the many-body tensor representation (MBTR) formalism¹²³. Using the TATR, it was shown that the CC correlation energies of a series of diatomics can be predicted with chemical accuracy (~ 1 kcal/mol) across the full potential energy surface with only 12 training points. This representation adheres to the three criteria above: amplitudes from increasingly accurate electronic structure methods should provide increasingly accurate TATRs, these amplitudes can be cut-off and stored based solely on their magnitude (as evidenced by retaining only 150 amplitudes for the diatomic representations in Ref. 122), and the relationship between wave function amplitudes and the target function (the correlation energy) is well-defined.

Indeed, the electronic wave function, in principle, gives access to any molecular property. The derivation of arbitrary properties in terms of wave function amplitudes is far from straightforward, however. In order to generalize the mapping from representation to target function (i.e. arbitrary molecular properties), we propose the density tensor representation (DTR), an analogous extension of the MBTR using one- and two-particle reduced density matrices rather than wave function amplitudes. In the following sections we demonstrate that careful choice of the underlying raw wave function features, as laid out in section 4.2.1, results in a more compact representation with a simple theoretical motivation. A proof-of-principle ML model using this representation is detailed in section 4.2.2. Results from this initial implementation for electronic energies and dipole moments of several small molecules are given in sections 4.4.1 - 4.4.3, and details of their possible extensions for molecular transferability and arbitrary property calculations given in section 4.4.4.

4.2 Theory

4.2.1 Electronic Structure Theory

The time-independent electronic Schrödinger equation (and, therefore, the energy) can be written generally in terms of one- and two-particle reduced density matrices (1-RDM D and 2-RDM Γ)¹⁵:

$$\begin{aligned} E &= \langle \Psi | \hat{H} | \Psi \rangle \\ &= D_{pq} h_{pq} + \Gamma_{pqrs} g_{pqrs} \end{aligned} \tag{4.1}$$

where implicit sums in Einstein summation notation are included over arbitrary orbitals p, q, r, s for the one- and two-electron integrals of the electronic Hamiltonian, h and g . D and Γ are defined according to the wave function ansatz

$$D_{pq} = \langle \Psi | a_p^\dagger a_q | \Psi \rangle \tag{4.2a}$$

$$\Gamma_{pqrs} = \langle \Psi | a_p^\dagger a_q^\dagger a_s a_r | \Psi \rangle \tag{4.2b}$$

where the excitation operators of the second-quantized electronic Hamiltonian are included in the reduced densities. As molecular properties can be expressed in terms of energy derivatives, they can also be expressed in terms of these reduced density matrices. Differentiating Eq. (4.1) with respect to an arbitrary parameter Ω , noting that $|\Psi\rangle$ carries no dependence on Ω (assuming the Hellmann-Feynman theorem^{35,36} holds),

$$\begin{aligned} \frac{\partial E}{\partial \Omega} &= \langle \Psi | \frac{\partial \hat{H}}{\partial \Omega} | \Psi \rangle \\ &= D_{pq} \frac{\partial h_{pq}}{\partial \Omega} + \Gamma_{pqrs} \frac{\partial g_{pqrs}}{\partial \Omega} \end{aligned} \tag{4.3}$$

thus first-order molecular properties can also be described by the RDM's without their differentiation. This can be said even for time-dependent properties, in which a time-dependent Hellmann-Feynman theorem applies.²¹ Furthermore, one-electron properties (such as the electronic dipole moment) can be expressed solely in terms of the 1-RDM and the relevant property integrals,

$$\Omega = D_{pq} \langle p | \hat{\Omega} | q \rangle. \quad (4.4)$$

The 1-RDM is a matrix with as many elements as the number of basis functions squared, and is diagonally dominant for correlated methods in the molecular orbital (MO) basis. Matrices of this size are amenable to large-scale storage, but still hold a wealth of relevant wave function information which can be improved by simply improving the correlated method used to generate them. Additionally, the additivity of densities dictates an additivity of the property contributions, e.g. for the total CC dipole moment in one direction α :

$$\mu_\alpha = (D_{pq}^{SCF} + D_{pq}^{CC}) \langle p | \hat{\mu}_\alpha | q \rangle + \mu_\alpha^{nuc} \quad (4.5)$$

where we have separated the contributions of the uncorrelated (e.g. self-consistent field (SCF)), correlated (e.g. CC theory), and the nuclear (μ_α^{nuc}) components. This separability of uncorrelated, correlated, and nuclear contributions to the property is particularly desirable for machine-learning, as we minimize the contributions learned by the algorithm. In order to probe the efficacy of using RDMs as “raw features” in a machine-learning context, we will outline and employ a method used previously by Margraf and Reuter¹²² for using wave function amplitudes as wave function features. The authors also acknowledge a related work¹²⁴, published on arXiv during manuscript revision, using the eigenvalues of a localized (Hartree-Fock or density functional theory) 1-RDM as the starting point for engineering features in a neural network approach to computing the correlation energy, dubbed the

Deep Post-Hartree-Fock (DeePHF) method.

4.2.2 Machine-Learning

Kernel Ridge Regression

We will focus on Kernel Ridge Regression (KRR), as a prototypical ML method.^{125,126} In KRR, a target property, y , is approximated as:

$$y(\boldsymbol{\nu}) = \sum_i^M k(\boldsymbol{\nu}'_i, \boldsymbol{\nu}) \alpha_i. \quad (4.6)$$

Here, $\boldsymbol{\nu}$ is the representation of the system of interest, $k(\boldsymbol{\nu}'_i, \boldsymbol{\nu})$ is a kernel function which measures the similarity between two systems (with representations $\boldsymbol{\nu}$ and $\boldsymbol{\nu}'$), and α_i is a regression coefficient corresponding to the training point with the representation $\boldsymbol{\nu}'_i$. The sum runs over M training points, so that the target property can be computed as a dot-product between a kernel vector \mathbf{k} and a coefficient vector $\boldsymbol{\alpha}$.

An advantageous feature of KRR is that the coefficients $\boldsymbol{\alpha}$ can be obtained from a closed-form linear algebra expression:

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}. \quad (4.7)$$

Here, the regularization hyperparameter λ is used to prevent overfitting. \mathbf{K} is the Kernel matrix with elements containing the kernel function of all pairs of training points. Herein, we use the radial basis function (or Gaussian) kernel to measure the similarity of representations:

$$k(\boldsymbol{\nu}', \boldsymbol{\nu}) = \exp\left(-\frac{\|\boldsymbol{\nu}' - \boldsymbol{\nu}\|^2}{2\sigma_m^2}\right) \quad (4.8)$$

where σ_m is another hyperparameter of the model which adds flexibility to how the kernel

measures system similarity. Both hyperparameters are optimized empirically to give the best performance in leave-one-out cross-validation.

Overall, it should be emphasized that the KRR model bears clear similarities to the equations we wish to approximate— Eqs. (4.1) and (4.3) — which contain a density-related term multiplied by an additional, property-specific term.

Representation

The representations used herein follow the original idea of the many-body tensor representation (MBTR) of Rupp¹²³, a geometrical representation proposed as an extension to the Coulomb matrix¹²⁷. Margraf and Reuter recently proposed a modification of the MBTR that uses wave function amplitudes from Møller-Plesset perturbation theory (MP2) or CC theory instead of geometrical features, termed the t-Amplitude Tensor Representation (TATR)¹²². Simply put, these representations can be understood as discretized, broadened histograms. This is achieved via a sum of Gaussian functions (with some fixed width σ), centered on the raw features t_i (i.e. wave function amplitudes). This function is evaluated for a discretized range of values $x \in \chi$:

$$\nu(x) = \sum_{i < N} g(t_i, \sigma, x) \quad (4.9)$$

where the sum runs over the N raw features. The representation is thus an l -dimensional vector ν , where χ is discretized into l points with even step-size Δx . Thus the N raw features are represented in a vector form which now contains l new features. Though it could in theory be optimized, we will use $\sigma = 0.05$ for consistency throughout this work for simplicity. (This value was empirically determined to be optimal for diatomics in Ref. 122).

In Ref. 122 the highest N amplitudes (by magnitude) were used for each excitation level m to give a TATR of fixed length l times the number of excitation levels considered. This number

of amplitudes N could be chosen in a number of ways, but Margraf and Reuter chose the highest 150 for simplicity in their paper. This proved sufficient for diatomics in a reasonable basis set, but is not a universally applicable choice for larger cases (see section 4.4.1). The discretization range χ was also fixed as [-1:1] with $l = 150$ and $\Delta x = \frac{2}{150}$, which we will also use for comparison.

The main novelty of the current work is to use density matrix elements in place of t_i . This “density tensor representation” (DTR) has several advantages. For simplicity we will focus on the 1-RDM — while this is sufficient for one-electron properties, the information stored in the 1-RDM is also necessary (and in some cases, as discussed in section 4.3, perhaps sufficient) for describing the total wave function. The 1-RDM will thus provide much of the wave function character needed to describe the correlation energy. This is the first advantage of using a density matrix based representation: one-electron properties are exactly described by a product of the 1-RDM with property integrals, as seen in Eq. (4.4). In analogy, Eq. (4.6) describes the target function as the product of a (now density-dependent) kernel and the regression coefficients. These coefficients thus describe the property integrals and all approximations in the model (e.g. correlation beyond the level of the density matrix used to build the representation). Thus, a simple mapping from representation to target function is achieved in strict analogy to theory. The 1-RDM is diagonally dominant in the MO basis, meaning the number of significant ($> 10^{-8}$) elements is scarcely larger than the number of basis functions. This is a further advantage of using 1-RDM elements instead of t-amplitudes: even considering the entire matrix, there will never be more elements than the number of basis functions squared, far fewer than the doubles amplitudes of MP2 or CC theory. These few elements are amenable to large-scale storage, and the number of elements retained could even be reduced in the case of large systems. Finally, the density matrix can be easily defined for most electronic structure methods, unlike other descriptors

such as the Hartree-Fock-level descriptors (Fock, Coulomb, and exchange matrix elements) employed in Refs. 109 and 110 or wave function amplitudes which are only available for some correlated levels of theory (and depend on the level of “excitation” considered). The DTR is thus systematically improvable, by building it from improved wave functions. Importantly, TATR and DTR based ML models will always use two levels of theory: a lower level method (e.g. MP2) to build the representation and a higher level method (e.g. CCSD) to provide the training data.

Algorithm

Rather than spanning chemical compound space, wave function representations span the “wave function space” — that is, the space of possible wave function parameters — of the systems we wish to describe. The space covered by a KRR model is defined by its training set, meaning that the model will be unsuitable for predicting properties of systems that are very different from the training set. Since the high-level *ab initio* calculations for the construction of the training set are by far the most computationally demanding part of the process, we want to ensure that the training set efficiently spans the wave function space of interest in as few points as possible.

To this end, we apply the k-means clustering algorithm as implemented in the Scikit-Learn ML package¹²⁸ to an initial “grand set” of relevant structures. This grand set could, e.g., consist of all grid points that will be computed for a final PES of interest or to a representative MD trajectory. k-means is a multi-dimensional partitioning scheme— N initial points are partitioned into M clusters by taking the norm difference of the representations and “binning” them based on randomly selected initial cluster centroids. The cluster centers are then updated and the procedure is iterated to self-consistency. Due to the stochastic nature of choosing an initial guess, the algorithm is run repeatedly (30 times for all datasets

in this manuscript) and the “best” set of clusters is chosen (as determined by minimizing the standard deviation of the distance between each point and its cluster centroid).

This approach requires that representations (TATRs, DTRs, *etc.*) for each point in the “grand set” of N structures be calculated first (at the low level of theory). To ensure that the model is truly predictive (i.e. able to reliably predict properties of systems for which no information known during training), a test set of systems is held back from the grand dataset prior to clustering. For the M representative data points chosen in this manner (i.e., the training set), the target function (energy, dipole, *etc.*) is computed with the high-level method. Once this is complete, the model is trained and subsequent predictions at approximately high-level accuracy and low-level cost are possible.

It should be noted here that the steps outlined in section 4.2.2 are completely general – no assumption was made about the level of theory besides that a density matrix can be formed. Furthermore, while KRR and k-means are used herein, alternative clustering schemes, hyperparameter optimization routines, and regression techniques can be substituted. The optimization of these choices will be the subject of a future study.

For benchmarking purposes, wave function parameters and high-level target values were computed for all datasets. The performance of a model is generally measured by the mean absolute error (MAE) of the predicted test set; however, the issues of truncation, separability, and learning rate are also considered.

4.3 Computational Details

Second-order MBPT using a restricted closed-shell Hartree Fock reference wave function (i.e. Møller-Plesset perturbation theory, MP2)^{26,129}, was used to generate both TATRs and DTRs

for all systems. The TATR is generated from Eq. (4.9) using the highest 150 (by magnitude) amplitudes of double “excitations” t_{ij}^{ab} . For DTRs, the elements of the MP2 1-RDM are used. This choice is advantageous not just for comparison to Ref. 122, but also due to the special form of the MP2 reduced density matrices³⁴:

$$D_{pq} = \langle 0 | \hat{T}_2^\dagger a_p^\dagger a_q \hat{T}_2 | 0 \rangle \quad (4.10a)$$

$$\Gamma_{pqrs} = 2 \langle 0 | a_p^\dagger a_q^\dagger a_s a_r \hat{T}_2 | 0 \rangle. \quad (4.10b)$$

where the “excitation” operators are defined as $\hat{T}_n = (\frac{1}{n!})^2 t_{ij\dots}^{ab\dots} a_a^\dagger a_b^\dagger \dots a_j a_i$, and we have again implied Einstein summation notation this time over occupied i, j, \dots and virtual a, b, \dots orbital spaces. The form of Eqs. (4.10a) and (4.10b) indicates that the reduced density matrices are fully defined by the doubles amplitudes and products thereof. In this light, it is reasonable to suggest that all of the information necessary to reproduce the wave function is coded into either of these and, to reduce the size of the representation, the 1-RDM should suffice for describing the MP2-level correlated wave function. Thus, any correlated one-electron property at the MP2 level can be described by this matrix.

Coupled cluster with single and double excitations (CCSD) was used to generate the high-level reference data, i.e. correlated energies and electronic dipole moments, for all systems. This method represents a marked improvement over MP2 in systems where correlation is strong, such as stretched diatomics, without incurring the additional cost of the gold-standard CCSD with perturbative triples method, CCSD(T). However, the CCSD(T) method could be used to provide higher quality targets (e.g. energies and dipoles) with no additional modifications — as noted in section 4.2.2, the algorithm as presented is completely general.

All electronic structure calculations were performed with the Psi4⁹⁸ electronic structure

package. Molecular symmetry was used to maximize TATR performance with as few amplitudes as possible (except in section 4.4.1). All data (energies, dipoles, amplitudes, and densities) were harvested in full machine precision through either the Psi4-JSON interface or the PsiAPI infrastructure, with the exception of wave function amplitudes at the MP2 level, which were printed in the output file at ten decimal places. The def2-TZVP¹³⁰ basis was employed for diatomic calculations for comparison with Ref. 122, and the aug-cc-pVDZ (aDZ) basis^{94,95} was used for all remaining calculations. Density-fitted integrals were used in the construction of MP2 densities using the default auxiliary basis (the so-called “RI” basis sets for def2-TZVP¹³¹ and aDZ¹³², obtained from the New Basis Set Exchange¹³³).

As outlined above, the general procedure from Margraf and Reuter¹²² was followed for machine-learning. The hyperparameters σ_m and λ were optimized on a uniform grid from $10^{-8} : 10^8$ for all models unless otherwise specified, using the “negative-mean-squared” loss function. All regressions were performed using 20 training points, except in section 4.4.1 where 12 training points were used to emphasize the difference between the two strategies considered. k-means clustering was repeated 30 times to account for stochastic deviations in the algorithm. Test sets consisting of $N/4$ points are held back before the clustering step. The radial basis function kernel (sometimes called the Gaussian kernel), Eq. (4.8), was used for all representations. Machine-learning algorithms were performed using Scikit-Learn (skl)¹²⁸ and the Machine-Learning Quantum Mechanics (MLQM) python package¹³⁴, which generates a number of molecular representations, provides options for generating Psi4 input files and harvesting their results, and wraps some skl functions. Both of these codes are open-source and freely available on GitHub.

Geometries for diatomics were constructed by taking $N = 200$ uniform increments from 0.5-2.0 Å. All other geometries were randomly selected from MD simulations ($N = 150$) carried out in the Gromacs⁹¹ software package. Simulations of (*S*)-methyloxirane and (*R*)-

methylthiirane surrounded by a 5 Å box of water molecules were executed using an all-atom OPLS/AA forcefield¹³⁵ for the solute and the TIP3P model for water¹³⁶. Each 5 ns trajectory was carried out in the NVT ensemble, with the solute and solvent coupled separately to a temperature bath at 300 K using a modified Berendsen thermostat and a coupling time of 0.1 ps. Geometries were selected from a set of 250 evenly-spaced snapshots along the trajectory. Geometries, Psi4 input and output files, and their accompanying optimized (hyper)parameter values in JSON format can be found at the Virginia Tech Data Repository¹³⁷ to permit reproducibility of our results. Scripts for harvesting and manipulating data using MLQM are also included in the form of Jupyter Notebooks.

4.4 Results and Discussion

4.4.1 Truncation considerations

As stated previously, the “raw” wave function features used to build machine-learning representations must be carefully selected. To illustrate this point, the TATR was used to model a carbon monoxide potential energy curve using the highest 150 amplitudes (by magnitude) and $M = 12$ training points as in Ref. 122. Figure 4.1a shows that, when molecular symmetry is considered (C_{2v} , the largest Abelian subgroup of the full $C_{\infty v}$ point group), excellent regression results are produced using only 150 amplitudes.

However, to model the effect of decreased sparsity in the amplitudes on the representation, molecular symmetry was then dropped to C_1 . When molecular symmetry is ignored, the number of “significant” ($> 10^{-8}$) amplitudes greatly increases, as shown in Figure 4.1b. In this case, taking only the highest 150 amplitudes no longer creates an effective representation of the wave function, resulting in poorer performance across the curve, also shown in

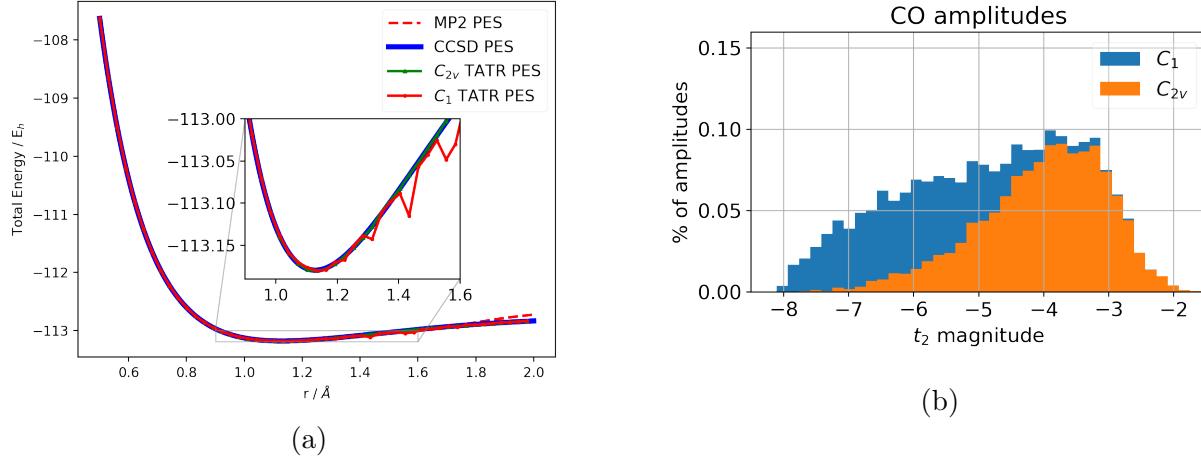


Figure 4.1: (a) KRR for the carbon monoxide potential energy curve with and without molecular symmetry considered and (b) their respective amplitude distributions. TATR models used $M = 12$ training points. All amplitudes with magnitude $< 10^{-8}$ are set to 0.

Figure 4.1a.

This example shows a considerable disadvantage of the TATR, namely that the T2-amplitude tensor is in general very large (equal to the number of occupied orbitals squared times the number of virtual orbitals squared). It is, of course, in principle possible to build the TATR with an arbitrary number of amplitudes, as the size of the representation is independent of the number of amplitudes. However, using all amplitudes is not advisable, as a large number of negligible amplitudes would lead to large (but chemically meaningless) TATR values around zero. Meanwhile, without extensive testing, it is in general unclear how many amplitudes to retain for an effective representation. For this reason, a large number of amplitudes must in general be stored, at least for initial testing and hyperparameter optimization. Here cutoffs based on the amplitude magnitude may be employed, but this can still lead to significant storage requirements for larger molecules and basis sets. Figure 4.1 also indicates that the TATR will likely not be applicable for larger systems that have no symmetry.

Storing the 1-RDM, on the other hand, only requires storing the number of basis functions

squared floating point numbers. Furthermore, it is also known that the majority of significant elements in the MO-basis 1-RDM will be along the diagonal — with some number of off-diagonal contributions. This gives a useful rule-of-thumb for how it could be truncated in extreme cases. For the remainder of this manuscript, the TATR will be computed using the highest 150 amplitudes (by magnitude) computed in the highest symmetry available. The DTR will be computed using all 1-RDM elements.

4.4.2 Energies

In the spirit of comparing to previous results, the same diatomic potential energy curves from Ref. 122 were computed using the DTR. The prediction errors on the test set are shown in the Supporting Information, along with TATR results for comparison. Overall, DTR results do not vary significantly from the C_{2v} -symmetry TATR results for these systems; however, when molecular symmetry is not considered, a far greater number of amplitudes are required in the TATR to achieve the same accuracy.

To test the applicability of the DTR and TATR models to larger systems which may not benefit from molecular symmetry, geometries for water, methanol, (*S*)-methyloxirane, and (*R*)-methylthiirane near equilibrium were sampled randomly from a set of molecular dynamics (MD) trajectories and examined using both the TATR and DTR models. Correlation energy data are presented in Figure 4.2. The prediction errors on the test set are plotted against the true CCSD correlation energy, ordered by increasing energy to evaluate how the model error varies with respect to the magnitude of the correlation energy. Each point represents the prediction error of a single geometry along the MD trajectory. As with the diatomics, most predictions lie well within the bounds of “chemical accuracy” (1.6mE_h). However, for the TATR, linear error centered around zero and which changes sign near the

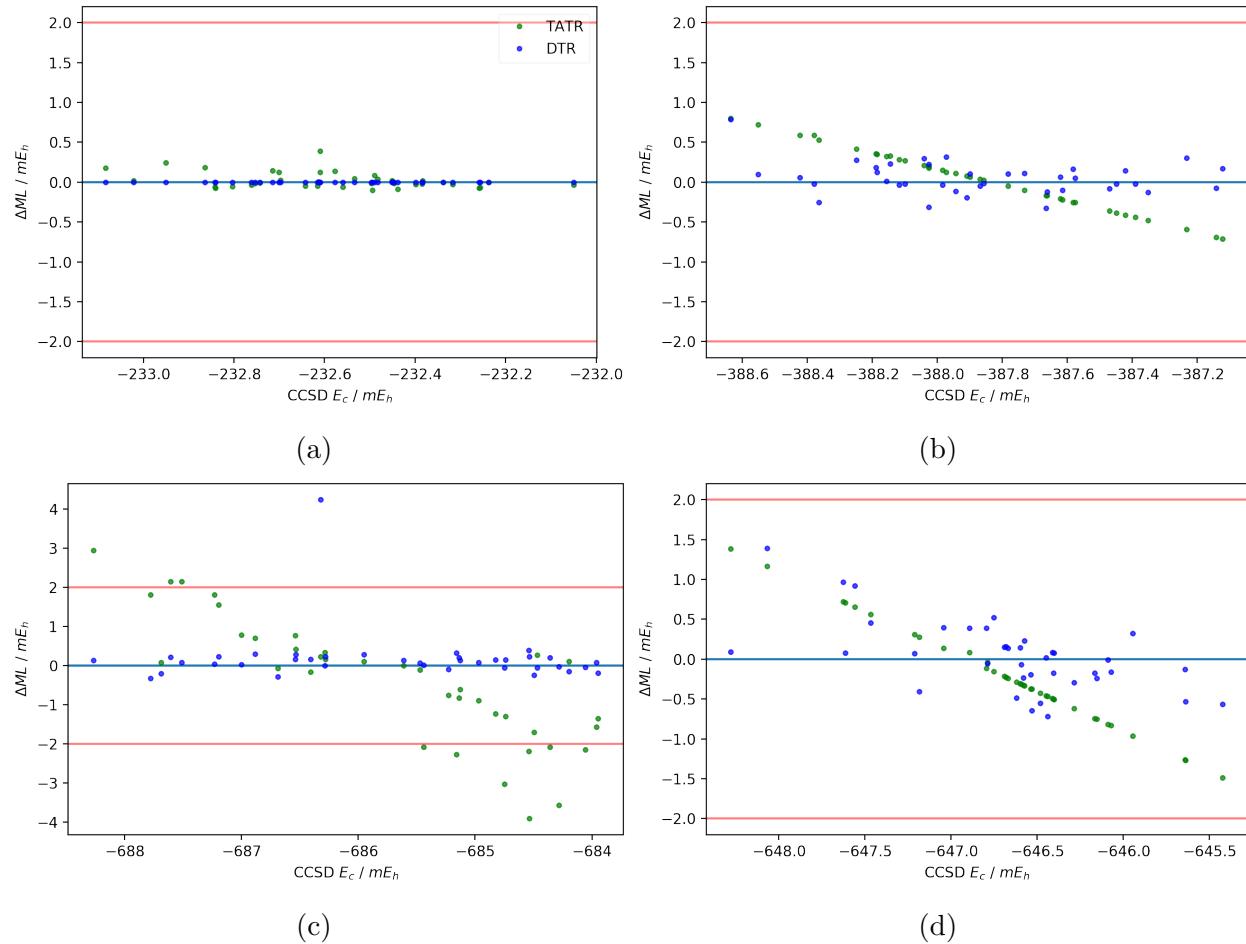


Figure 4.2: DTR vs TATR errors in mE_h for small molecule datasets: (a) H_2O , (b) CH_3OH , (c) (*S*)-methylloxirane, and (d) (*R*)-methylthiirane. Red lines indicate $2\ mE_h$.

mean correlation energy value suggest that this model is biased towards the mean of the training set. While this still gives a rather agreeable mean-absolute-error across this particular test set, predictions for geometries far from the energetic minimum (i.e., sufficiently different from the training set) may not perform as admirably. This issue is explored further in section 4.4.4.

A summary of energy results is shown for both the TATR and DTR methods in Figure 4.3. The trained models reproduce the mean average correlation energy of the test set to within two milli-Hartree for every system considered. Furthermore, DTR errors are kept below 0.5

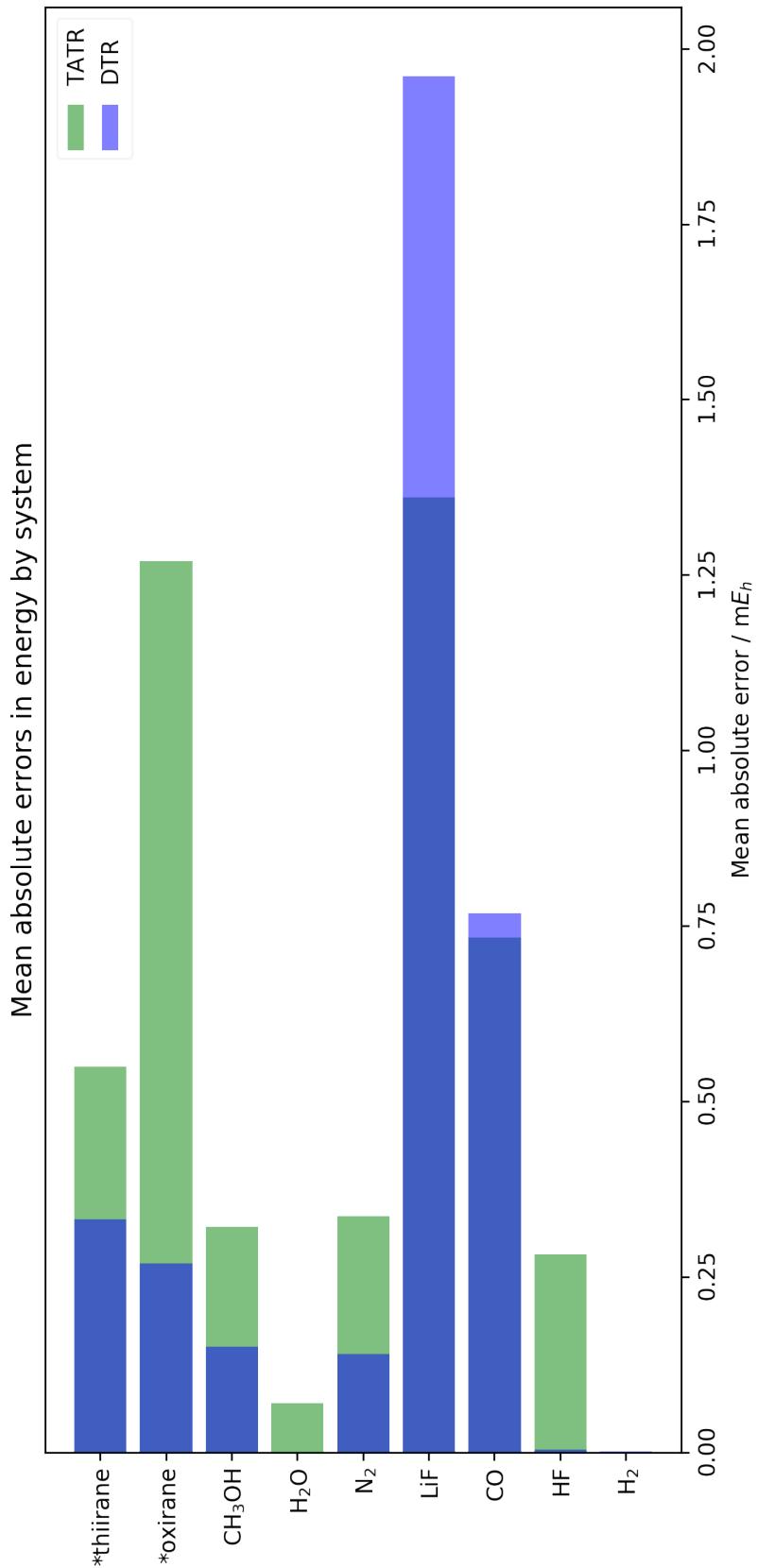
milli-Hartree for all systems except for CO and LiF. Inspection of their individual model performance (see the Supporting Information) reveals some difficulty in modeling the extreme ends of the curve (i.e. points far from the energetic average across the test set). These errors are also dominated by a relatively small number of outliers with increased errors. These examples also cover large ranges of correlation energies (100 and 60 milli-Hartree, respectively), which test both the regression capability and the validity of the underlying MP2 wave function used for the representation.

The energy data suggests that the improvements made by the DTR method result in greater accuracy in the molecular representation, in particular for larger molecules. As shown in Ref. 122, the electronic wave function can be well approximated using this relatively simple functional form when applied to energies. If this functional form is truly representative of the total wave function, rather than simply the parts which are important to the energy, then it should also be possible to compute molecular properties with a similar approach.

4.4.3 Dipole moments

Several schemes for computing dipole moments were employed to emphasize the importance of Eq. (4.5). Full density matrices ($D^{SCF} + D^{MP2}$) were used as raw features to build DTRs for learning total ($\mu^T = \mu^e + \mu^{nuc}$), electronic ($\mu^e = \mu^{SCF} + \mu^{CC}$), and correlated (μ^{CC}) dipole moments. MAEs (in Debye) for the machine-learned dipole moments, using each of the three possible target values, are given in Table 4.1. The MAEs for the MP2-level correlated dipole (relative to the CCSD result) for the same predicted sets are given for comparison.

As expected, minimizing the extrapolation necessary for the model reduces the error drastically. The error in the electronic component of the dipole moment in e.g. (*S*)-methyloxirane is reduced by an order of magnitude relative to using the full dipole moment. Furthermore,

Figure 4.3: DTR vs TATR errors in mE_h for all datasets. (* = methyl)

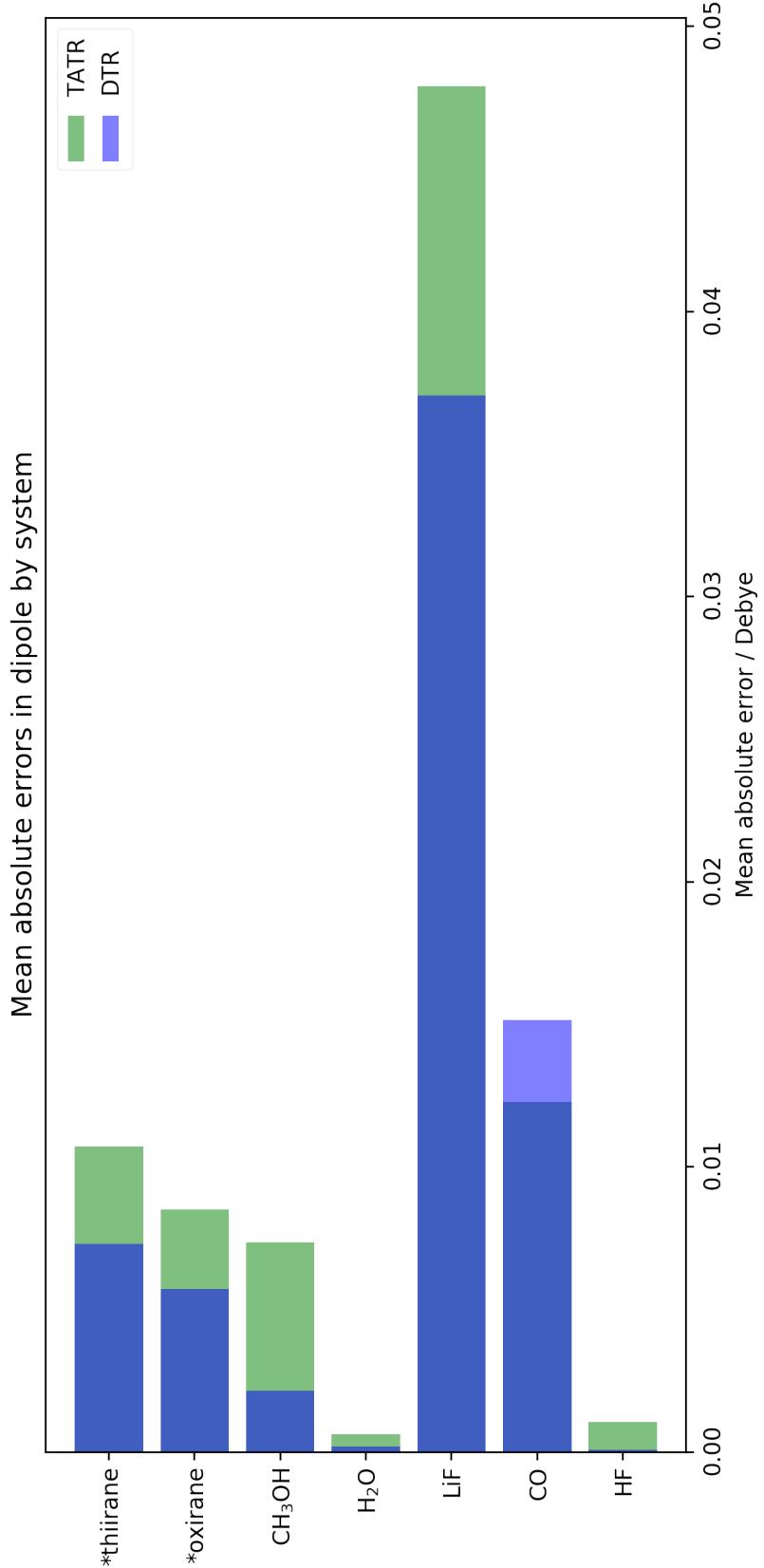
Molecule	μ^T	μ^e	μ^{CC}	MP2
H ₂ O	1.5×10^{-3}	1.3×10^{-3}	1.8×10^{-4}	1.8×10^{-2}
CH ₃ OH	3.8×10^{-2}	1.7×10^{-1}	2.1×10^{-3}	1.8×10^{-2}
*oxirane	1.0×10^{-1}	4.3×10^{-1}	5.7×10^{-3}	2.0×10^{-2}
*thiirane	5.4×10^{-2}	6.5×10^{-1}	7.3×10^{-3}	3.5×10^{-2}

Table 4.1: Mean absolute errors (in Debye) relative to CCSD of four machine-learning datasets utilizing full, electronic, and correlated dipole moments as training targets. MP2 dipoles given for comparison. (* = methyl)

this error is also an order of magnitude lower than the MAE in the MP2 correlated dipole moments on the same predicted set. This is consistent with Ref. 122 where extremely accurate models for the *correlation* energy were built. The same is clearly the case for the (correlated) dipole moment.

A summary of dipole results is shown for both the TATR and DTR methods in Figure 4.4. The same datasets as for correlation energy learning were considered, except for the homonuclear diatomics. As with the energy, the trained models reproduce the correlated dipole moment of the test sets to reasonable precision. Once again, the challenging cases of CO and LiF produce the maximum errors; however, it is encouraging that even here, errors for states near equilibrium (those modeled by molecular dynamics) are of milliDebye magnitude.

Indeed, it has been shown (e.g. in Ref. 120) that predicting dipole moments with ML is significantly more challenging than predicting energies. Still, our results support the notion that, as long as the reference wave function is sound, the molecular representation will “correctly represent the physics of the problem”¹²² as desired with a favorably small number of training points. While these results represent only a marginal improvement over the already sound MP2 approximation, they demonstrate that simple ML models effectively capture the characteristics of the wave function without the need for a complex functional form. They also show that this generalizes beyond the energy to molecular properties as well.

Figure 4.4: DTR vs TATR errors in *milliDebye* for all datasets. (* = methyl)

4.4.4 Extensions

The present model can be used to solve a number of potential problems, mostly involving local PESs or dipole moment surfaces of small molecules which are amenable to MP2-level calculations (and a few CCSD training calculations). This is especially useful for generating accurate force field parameters^{105,106} and evaluating numerical gradients¹¹⁷. Furthermore, the DTR representation has been shown to provide high accuracy with very few training points, resulting in a highly data-efficient method. While this covers a significant space of applications, there are some areas which cannot currently be addressed, but could be well within reach of the proposed algorithm.

Transferability

The DTR-based machine-learning model as described above shows that a mapping from the MP2 1-RDM to CCSD electronic properties is feasible using a simple ML model. This allows avoiding the high-cost evaluation of the CC wave function for multiple geometries of the same molecule. However, if the method is to be generalized to different molecular systems (rather than just different geometries as we've shown here), additional modifications must be considered. Specifically, two (related) types of transferability are desirable: on one hand, a model can be transferable between different chemical compositions (i.e. trained on water and methane, applied to methanol)¹⁰⁹. On the other hand, a model can be transferable among different system sizes (i.e. trained on water clusters and applied to bulk water)¹³⁸.

The main impediment for both the TATR and the DTR to be transferable is the use of a molecular orbital basis, which is strongly system dependent. An obvious route towards more transferability would be to use the atomic orbital (AO) basis instead. In this representation, the molecular dependence of the representation would stem only from the identity of the basis

functions placed on each atom. While this still precludes transferability between different elements, this is not very problematic in practice, as one can always train on the elements of interest. Regarding transferability to different sizes, this essentially boils down to the requirement that the ML model be size extensive. The lack of extensivity is a common weakness of ML models that use global representations such as the DTR or TATR. As recently discussed by Jung et al., global KRR models can be made size extensive, however, if the Kernel function is properly normalized¹³⁸. While that study focused on representations of the molecular geometry, it would certainly be worthwhile exploring how extensive models based on wave function representations would perform.

Finally, a theoretical consideration can also be made with respect to transferability. As mentioned above, one-electron properties can be described by the general form of Eq. (4.4). Comparing this to the KRR equations (see Eq. (4.6)), and considering that the kernel only includes information from the 1-RDM, it can be inferred that the coefficients α are modelling the contributions arising from the property integrals, $\langle p|\Omega|q \rangle$. In the present model, however, the coefficients are also doing the job of mapping the low-level (MP2) representation onto the high-level (CCSD) target function. To truly separate the molecule-specific terms from the correlation contributions, the property integrals (in a transferable basis, such as the AO basis) should thus also be included in the representation of the molecule. Modifications such as using product kernels to fold in the contribution from the integrals can be explored to correct this shortcoming.

Additional properties

Properties beyond dipole moments are also accessible, but not without modifications. While first-order one-electron properties are available through Eq. (4.4), considerations of both the one- and two-electron densities is generally necessary for two-electron properties. Fortu-

nately, we have seen that the 1-RDM suffices to predict at least one property which depends on both the 1- and 2-RDM, namely the energy. This may be related to the special formulation of 1- and 2-RDMs in MP2, as discussed above. It could also be a particular property of the ground-state energy which, according to the Hohenberg and Kohn theorem¹³⁹, formally only depends on the ground state density (and thus the 1-RDM). However, this is not generally the case, for example in CC response theory. Higher-order properties in CC response depend on both the 1- and 2-RDMs. Extensions to include both density matrices could be made, as with a singles-and-doubles TATR, though considerable truncation procedures would need to be implemented in order to keep storage costs reasonable across realistic datasets.

Alternatively, second-order properties which depend only on one-electron operators, such as polarizabilities and chiroptical response, could also be considered using a reformulated CC linear response (CC-LR). It can be shown that, similarly to first-order properties, second-order response properties of this type can be expressed using the 1-RDM only. This is achieved by computing the first-order perturbed left-hand wave function parameters, $\lambda^{(1)}$. Using these, second-order properties are computed in CC-LR as products of property integrals with *perturbed* one-electron reduced density matrices¹³. Contributions from one operator are folded into the 1-RDM, and contributions from the other are described by the property integrals. In the context of the DTR, the density representation would now require only this perturbed 1-RDM with no additional changes. Contributions from property integrals may also be included in the representation, perhaps through the same sort of generalizations made to allow for transferability. In this fashion, many linear response properties could be made accessible.

We will acknowledge one final potential oversight in property predictions done in the preceding manner: a new model must still be trained for each property. To allow for a “one-shot” model prediction of arbitrary properties, the wave function itself (or some usable representa-

tion thereof) must be learned. This has recently been explored by Schütt and coworkers¹¹⁹, which has proved some efficacy of this strategy. Additional work on the usability of such representations will be considered while algorithms to learn the representation itself are formulated. In practice, however, once the ground state equations (e.g. the amplitudes in CC theory) are solved, it is not necessary to recompute them for additional properties. Instead, the (linear) response function can be evaluated at this time for a multitude of properties. Because of this, we suggest that many properties are computed and stored in a database when proposing and gathering reference or training data for ML algorithms, such as those provided in QCArchive¹⁴⁰.

ML algorithm optimization

Aside from coding additional physics into the representation and returning more varied properties, the ML workflow itself may also be optimized to be more data-efficient, accurate, and precise across varied molecular representations. To further evaluate model performance and perhaps diagnose potential problems, learning curves were generated for every system considered. These curves plot the error in the test (the “validation curve”) and training (“training curve”) sets with respect to the number of training points used, with values ranging from two to one-hundred training points. The validation curve exposes the limitations of the current model by testing whether the model can be made more accurate using more training points. The training curve, on the other hand, describes how well the model is adhering to the training set, and so reveals problems of under- or over-training. Together, the curves indicate the practical limitations of the model’s accuracy and use of data, suggesting an irreducible error that can only be addressed by changing the model or the representation itself. Learning curves for all systems considered are available in the Supporting Information.

The data discussed within the manuscript so far (excluding section 4.4.1) has been collected

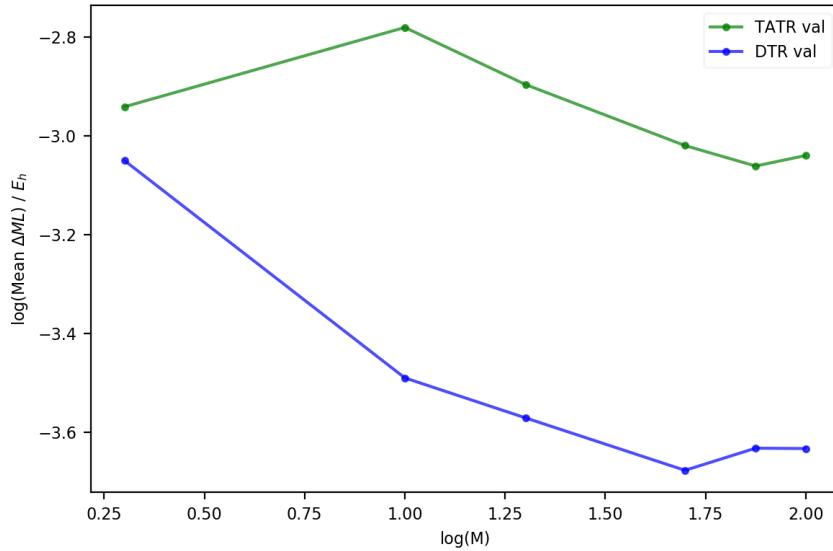


Figure 4.5: DTR and TATR validation curves for (S) -methyloxirane correlation energy.

using $M = 20$ training points, which result in the data points at that value in every validation curve. In a log-log plot, the validation curve should be linear such that an increase in the number of training points results in a commensurate decrease in error. This is well exhibited by, e.g., the validation curve for predicting the correlation energy for (S) -methyloxirane, Figure 4.5. Generally, the DTR validation curve starts lower than the TATR. The DTR model also tends to benefit more from additional training points in the two- to fifty-point range. It is important to note that the most significant outliers to this trend are the dipole moment models for (S) -methyloxirane and (R) -methylthiirane (see SI), the largest molecules considered and undoubtedly the most complex electronic environments. These systems benefit from larger training sets and are more prone to fortuitously low errors with small training sets, requiring a careful analysis of the validation curve before deciding on an optimal training set size. This is, of course, subject to the constraint that the training points are still taken from near-equilibrium values. Improving the quality of the training set by including more distinct electronic states away from equilibrium may improve performance.

In considering the training curves (see SI), it is revealed that the TATR suffers from increased training error as compared to the DTR, referred to as *bias*. Bias (sometimes referred to as *inductive bias*)¹²⁶ refers to error arising from simplifying assumptions made in the model: in this case, the small number of amplitudes retained is the most obvious source of additional bias. As this bias causes high training errors, it suggests that the TATR model is not fitting the training data as well as the DTR. This is likely the cause of the linear errors discussed in section 4.4.2, as the model is simply learning toward the mean of the training set, rather than properly reproducing each point. These errors can be reduced by improving the model, e.g. by including more amplitudes in the TATR or by the DTR modifications described above. Additional features or a reformulation of the representation may also be explored using automated model-selection routines, such as those employed in Ref. 124. These routines would, in theory, result in a more complex model accessing a higher dimensional feature space, so the challenges of extension become ensuring that this model still fits the training data (does not under-train as a result of bias) and generalizes to unseen data (does not over-train as a result of variance) when compared to a simpler approach.

4.5 Conclusions

Here we introduce the density tensor representation (DTR) for machine-learning quantum mechanics applications. The representation is based on the previous t-amplitude tensor representation (TATR), with improvements made through strictly theoretical considerations of three categories: systematic improvement, storage, and simplified representation-target mapping. Investigating the limits of these categories on small test sets show a number of favorable properties. The DTR can be easily defined for any electronic structure method in which a density can be defined. When compared to the TATR for MP2, it achieves superior

accuracy across most test cases when the MP2 wave function is expected to produce reasonable results. This accuracy is in the sub-mE_h range for correlation energies. Furthermore, applications to molecular properties are both theoretically and operationally justified for representations utilizing electronic densities as raw wave function features. Roughly milliDebye error was achieved for correlated electronic dipole moments of several small molecules near equilibrium. Extensions to include additional properties and molecular transferability are also considered, with the data-efficient DTR model providing a vital stepping stone to these generalizing improvements.

Chapter 5

Locally Correlated Real-Time Coupled Cluster Theory

5.1 Introduction

Dynamic molecular properties induced by the absorption, scattering, or refraction of an electromagnetic field (EMF) give rise to a number of experimental techniques for the detailed investigation and characterization of molecular light-matter interactions and structure.¹ Among these properties are absorbance, circular dichroism (CD), birefringence, Raman scattering, and many more. These techniques are essential for modern synthetic chemistry in both research and industrial settings.

Theoretical chemistry has become a “full partner with experiment”¹⁴¹ in this regard, providing high-quality benchmark calculations for affirming or even predicting many molecular properties, saving time, and increasing certainty in spectral assignments and molecular structure determination. Computing dynamic properties with current *ab initio* methods generally involves frequency-domain perturbation theory, referred to as response theory, to directly calculate the quantity of interest.^{2,12,21} Coupled cluster (CC) response theory^{18,19} has emerged as a robust solution to frequency-domain property calculations, when its cost is not prohibitive.^{13,51} Many techniques for circumventing the high-degree polynomial scaling of coupled cluster methods exist, with extensions of these to property calculations and

response theory providing promising results.^{61,62,142,143}

However, there are several drawbacks to the response formalism.^{17,22,23} First and foremost, the perturbations must be “small” relative to the intramolecular forces present in the system. This immediately precludes the possibility of simulating high-energy experiments such as X-ray spectroscopy, which have numerous applications in materials science and beyond. Second, only broadband excitations can be modeled straightforwardly: response theory typically assumes a single, uniform “kick” perturbation across all frequencies. Experimental apparatus, on the other hand, can make use of complex, multi-phase procedures involving tuned laser pulses, pump-probe analysis, *etc.*¹⁴⁴ Finally, temporally controlled multi-photon events such as high harmonic generation (HHG)^{145,146} lie outside the realm of the response formalism. Together, these drawbacks mean a wide variety of experiments cannot be predicted or supplemented with response theory calculations. To overcome this, we must move to non-perturbative, time-domain electronic structure theory,^{13,22,23} where there are fewer limitations on the form of the perturbing EMF.

The alternative of real-time CC (RTCC) methods has been discussed nearly as far back as the origins of CC itself in the realm of nuclear physics.^{11,147,148} More recently, a renewed interest in real-time coupled cluster has developed for the reasons discussed above. In the past 10 years, several implementations have been reported,^{149–153} with new insights into the aspects of numerical integration^{152,154} and interpretation^{37,152} as well as applications for a number of spectral properties.^{151,153,155–157} Orbital adaptive¹⁵⁰ and orbital optimized¹⁵⁸ variants have also explored the limitations of unrelaxed canonical Hartree Fock orbitals, and the effects of alternative reference orbitals on the propagation of unphysical imaginary components to energetics and electric dipole moments. Notably absent are studies on the ability to *reduce* the cost of real-time coupled cluster methods.

Real-time time-dependent density functional theory (RT-TDDFT) calculations, a cheaper

alternative introduced in the 1990s (then called the *time-dependent local-density approximation*),^{159–162} have become routine.^{22,23,163–168} Efforts to reduce the cost of RT-TDDFT have largely focused on reducing simulation time, utilizing techniques such as Padé approximants to accelerate the convergence of the Fourier transform,¹⁶⁷ and fitting schemes to avoid the Fourier transform all together, eliminating the problem of short trajectories resulting in low-resolution spectra.¹⁶⁹ Repisky *et al.* introduced the concept of dipole pair contributions,^{170,171} which are typically less complicated than the total electric dipole, and so these may be individually approximated efficiently using the techniques mentioned above.¹⁶⁷ However, the problems of frequency-domain DFT carry over directly to the time domain, such as the underestimation of excited state energies¹⁰⁰ and difficulties arising from the adiabatic approximation.^{167,172,173} We refer the reader to a recent, comprehensive review article²³ and citations therein for a more complete discussion of these challenges. Regardless, the success of RT-TDDFT under most conditions combined with its drastically reduced computational cost make it the only viable method for large molecules at present.

Borrowing from the vast literature of reduced-scaling ground-state or frequency-domain CC, there are numerous potential candidates for reducing the cost of RTCC, besides adapting the successful approaches implemented for RT-TDDFT. First, the standard non-perturbative truncated approaches used for properties such as CC2³² and CC3³³ are immediately possible, as are property-optimized basis sets.^{82,142,174–180} Further, details of implementation such as choice of intermediates, the effects of single- or mixed-precision, or the use of graphical processing units have only just begun to be explored.¹⁸¹ An alternative formulation developed by DePrince and Bartlett, dubbed the time-dependent equation-of-motion CC (TD-EOM-CC)^{151,153,155–157} method, reduces the cost by targeting the difficulty of numerical integration of multiple “stiff” coupled differential equations. By selecting a given moment function to propagate in time, the coupled sets of t - and λ -amplitude expressions do not have to be

propagated, reducing both the number and difficulty of numerical integrations required.

Absent from this list is the family of local correlation methods,²⁴ which have been wildly successful for reduced-scaling approaches to ground state energies for correlated methods and selected properties.^{13,143,180,182} These methods seek to build a reduced virtual orbital space based on lower-cost criteria, such as (pair) energies from low-order perturbation theory or atomic orbital charge analysis. While still only routine for ground-state calculations, these methods and variants thereof have shown promise in the calculation of selected response properties.

In this work, we report the first application of local correlation to RTCC. This is achieved through a simulation approach,¹⁸³ which forgoes computational savings in favor of algorithmic simplicity, for the purposes of rapid exploration and development. The effects of occupied and virtual space localization are considered for the simulations of small hydrogen clusters in the presence of electric field perturbations. Absorption cross sections as well as electronic CD (ECD) spectra are computed using successively smaller fractions of the canonical orbital space using the popular projected atomic orbital (PAO)^{184–187} and pair natural orbital (PNO)^{188,189} schemes. The results are analyzed with respect to full-space simulations. Finally, wave function amplitude dynamics are investigated in order to determine the extent to which these schemes suppress or cause large amplitude deviations, which cause instabilities in numerical integration and spurious oscillations in the dipole trajectory.

5.2 Theoretical Background

5.2.1 Real-Time Coupled Cluster Theory

Conventional RTCC implementations begin by computing the ground-state right- and left-hand CC wave function amplitudes t_μ and λ_μ via the residual expressions

$$\langle \mu | \bar{H} | \Phi \rangle = 0 \quad (5.1a)$$

$$\langle \Phi | (1 + \hat{\Lambda})[\bar{H}, \tau_\mu] | \Phi \rangle = 0 \quad (5.1b)$$

where $|\Phi\rangle$ is the Hartree-Fock ground state determinant and $|\mu\rangle$ are substituted determinants (singles, doubles, *etc.*) obtained by the action of the second-quantized excitation and de-excitation operators τ_μ , and \bar{H} is the similarity transformed electronic Hamiltonian

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}} \quad (5.2)$$

with right-hand cluster operators

$$\hat{T} = \sum_{\mu}^N \hat{T}_\mu \quad (5.3a)$$

$$\hat{T}_\mu = \sum_{\mu} \tau_\mu t_\mu \quad (5.3b)$$

and left-hand cluster operators

$$\hat{\Lambda} = \sum_{\mu}^N \hat{\Lambda}_\mu \quad (5.4a)$$

$$\hat{\Lambda}_\mu = \sum_{\mu} \tau_\mu^\dagger \lambda_\mu. \quad (5.4b)$$

The time evolution of the amplitudes is governed by the nonlinear differential equations

obtained through the time-dependent Schrödinger equation (in atomic units)

$$i \frac{\partial t_\mu}{\partial t} = \langle \mu | \bar{H}(t) | \Phi \rangle \quad (5.5a)$$

$$-i \frac{\partial \lambda_\mu(t)}{\partial t} = \langle \Phi | (1 + \hat{\Lambda}(t)) [\bar{H}(t), \tau_\mu] | \Phi \rangle. \quad (5.5b)$$

The right-hand sides of Eqs. (5.5a) and (5.5b) are simply the residual expressions, where we have replaced the Hamiltonian with a time-dependent Hamiltonian. By including a field perturbation as a time-dependent addition to the Fock operator, the right-hand sides may be computed by updating the Hamiltonian from time t to time $t' = t + h$ and recomputing the residual expressions. This is achieved using a numerical integrator, which produces solutions to equations of the form

$$\frac{dy(t)}{dt} = f(y, t). \quad (5.6)$$

Here, y is the amplitude tensor, and the function f is the residual expression. Multiple integration schemes are possible; for simplicity, we adopt the popular fourth-order Runge-Kutta integrator,¹⁹⁰ defined by

$$\begin{aligned} k_1 &= f(y, t) \\ k_2 &= f\left(y + \frac{1}{2}hk_1, t + \frac{1}{2}h\right) \\ k_3 &= f\left(y + \frac{1}{2}hk_2, t + \frac{1}{2}h\right) \\ k_4 &= f(y + hk_3, t + h) \end{aligned} \quad (5.7)$$

with time step h , and the propagated tensor is computed as

$$y(t + h) = y(t) + \frac{1}{6}h(k_1 + 2k_2 + 2k_3 + k_4). \quad (5.8)$$

5.2.2 Properties

To within the dipole approximation, the complex time-domain response tensors $\boldsymbol{\alpha}$ and \mathbf{G}' can be defined by low-order expansions of the time-dependent electric and magnetic dipole moment expectation values in an electric field \mathbf{E} with frequency ω , *viz.*

$$\langle \mu \rangle_i = \mu_0 + \alpha_{ij} E_j \quad (5.9a)$$

$$\langle m \rangle_i = -\frac{1}{\omega} \frac{\partial E_j}{\partial t} G'_{ij} \quad (5.9b)$$

where i and j are Cartesian coordinates, and we have suppressed the time dependence for clarity. The dipole strength function S is related to the imaginary component of $\boldsymbol{\alpha}$ by

$$S(t) \propto \text{Tr} [\text{Im} (\boldsymbol{\alpha}(t))] . \quad (5.10)$$

Fourier transformation of S from the time to frequency domain yields the broadband absorption spectrum. The differential molar extinction coefficient is proportional to the imaginary part of the Rosenfeld \mathbf{G}' tensor¹⁹¹ by

$$\eta_i(t) \propto -\text{Tr}[\text{Im}(\mathbf{G}'(t))]. \quad (5.11)$$

The Fourier transform of Eq. (5.11) yields the ECD spectrum.

We note here two important points. First, we could just as easily define both $\boldsymbol{\alpha}$ and \mathbf{G}' with respect to the electric dipole expectation value; however, by expanding both moments in an electric field, we may recover both properties by computing expectation values of both the electric and magnetic dipole operators along the same electric field-perturbed trajectory. In principle, we may compute *any* electric-field-perturbed expectation value from a single

propagation – this is in contrast to the RT-EOM-CC method, which propagates a single moment function. Additional expectation values would require additional moment function propagations.

Second, we note that the low-order expansions in Eq. (5.9) are an approximation. The total dipole moments will contain many higher-order terms; however, at the field strengths used in this work, these effects are expected to be negligible. These terms can be separated and have been examined in the context of real-time simulations of X-ray absorption spectroscopy (XAS).¹⁵⁷ While very important to the advantages of the RTCC method, these effects are beyond the scope of the current work.

5.2.3 Local Correlation

Projected Atomic Orbitals

In the PAO method, the virtual space is localized using a linear combination of N atom-centered atomic orbital (AO) basis functions. For every occupied orbital i , the contribution to the Mulliken charge q from every atom j is computed as:

$$q_j = \sum_{\mu \in j} \sum_{\lambda} S_{\mu\lambda} C_{\lambda i} C_{\mu i} \quad (5.12)$$

where μ is the set of atomic orbitals centered on atom j , and λ runs over all AOs. \mathbf{S} and \mathbf{C} are the AO overlap matrix and the Hartree-Fock molecular orbital (MO) coefficients, respectively. Atoms (and their AO basis functions) are added to the domain of orbital i one at a time based on decreasing charge contribution. Each time the domain is altered, new

coefficients \mathbf{C}' are computed by solving

$$\sum_{\nu \in i} S_{\mu\nu} C'_{\nu i} = \sum_{\lambda} S_{\mu\lambda} C_{\lambda i} \quad (5.13)$$

where μ and ν are atomic orbitals belonging to domain i , and λ runs over all AOs. The Boughton-Pulay completeness value b_i

$$b_i = 1 - \sum_{\mu} \sum_{\lambda} C'_{\mu i} S_{\mu\lambda} C_{\lambda i} \quad (5.14)$$

is then compared to a chosen cutoff, δ_{PAO} . If b_i is above the cutoff, the next atom is added to the domain of i , and the process is repeated until the cutoff value is met. Orbitals with negligible norms less than another chosen parameter δ_{norm} (usually corresponding to contributions from core orbitals prone to linear dependencies) are also removed from each domain.

Once the domains are assigned, occupied *pairs* ij are then assigned *pair* domains, based on the union of the domains of the two occupied orbitals. An occupied-space projector $\tilde{\mathbf{C}}'$ is computed

$$\tilde{\mathbf{C}}' = \mathbf{1} - \mathbf{D}\mathbf{S} \quad (5.15)$$

where \mathbf{D} is (half of) the closed-shell Hartree-Fock one-particle density matrix. $\tilde{\mathbf{C}}'$ serves to remove contributions from the occupied orbital space, creating the redundant PAO space. Linear dependencies are removed by diagonalizing the projected overlap matrix

$$\tilde{\mathbf{S}}_{ij} = \tilde{\mathbf{C}}_{ij}^{T} \mathbf{S} \tilde{\mathbf{C}}_{ij}' \quad (5.16)$$

where $\tilde{\mathbf{C}}_{ij}'$ contains only the columns for atomic orbitals belonging to the domain of pair ij . PAOs which correspond to eigenvalues of $\tilde{\mathbf{S}}_{ij}$ below one final parameter, δ_{lin} , are then

removed. The orbitals are normalized to yield the non-redundant PAO basis for a given pair, $\tilde{\mathbf{C}}_{ij}$.

Two matrices are required to transform MO-basis quantities into the PAO basis. The first, which rotates from the MO to the redundant PAO basis for a given pair, is computed as

$$\mathbf{Q}_{ij}^{PAO} = \mathbf{C}_{ij}^\dagger \mathbf{S} \tilde{\mathbf{C}}'_{ij}. \quad (5.17)$$

To facilitate the use of the usual orbital energy denominator terms during the update of the amplitude equations at every iteration, a semi-canonical virtual basis for pair ij , $L_{ij}^{PAO} = \chi_{ij}\tilde{C}$, is found by diagonalizing the Fock matrix in the space of non-redundant PAOs \tilde{F} :

$$\tilde{F}\chi_{ij} = \epsilon_{ij}\chi_{ij} \quad (5.18)$$

where ϵ_{ij} are the semi-canonical orbital energies for the virtual space of occupied pair ij .

Pair Natural Orbitals

PNOs are obtained by diagonalizing a pair density built in a space of localized occupied orbital pairs. The MP2-level pair density is defined as

$$\mathbf{D}_{ij} = 2 \frac{\mathbf{T}_{ij}\tilde{\mathbf{T}}_{ij}^\dagger + \mathbf{T}_{ij}^\dagger\tilde{\mathbf{T}}_{ij}}{1 + \delta_{ij}} \quad (5.19)$$

with $\tilde{\mathbf{T}}_{ij} = 2\mathbf{T}_{ij} - \mathbf{T}_{ij}^\dagger$. Diagonalizing \mathbf{D}_{ij} yields the transformation matrix from the MO to the PNO basis \mathbf{Q}^{PNO} :

$$\mathbf{D}_{ij}\mathbf{Q}_{ij}^{PNO} = \mathbf{D}_{ij}n_{ij} \quad (5.20)$$

with occupation numbers n_{ij} . Truncation of the space is done by removing PNOs which correspond to occupation numbers below a cutoff, δ_{PNO} . As before, the transformation matrix from the PNO to a semi-canonical PNO basis is found by diagonalizing the Fock matrix in the space of PNOs:

$$\tilde{\mathbf{F}}\mathbf{L}_{ij}^{PNO} = \epsilon_{ij}\mathbf{L}_{ij}^{PNO}. \quad (5.21)$$

5.3 Computational Details

The CC wave function for a helical H₂ tetramer was propagated for 1000 a.u., with a time step of 0.02 a.u., in the presence of an explicit electric field. A short pulse approximating a Dirac delta pulse was applied to generate all possible excited states. Atomic coordinates are found in the supplementary information (SI). Nuclear coordinates were frozen throughout these simulations.

To approximate a Dirac delta pulse, we apply a narrow time-dependent Gaussian field of the form

$$E(t) = F e^{-\frac{(t-\nu)^2}{2\sigma^2}} \quad (5.22)$$

with field strength F , center ν , and standard deviation σ . The field is propagated in the y-direction, which is along the helical axis of the system. All calculations in this work use a field defined by $F = 1 \times 10^{-3}$, $\nu = 0.05$, and $\sigma = 0.01$, all in atomic units. Electric and magnetic dipole moments were damped using a damping function of the form $e^{-t\tau}$, with $\tau = 150$.

The reference simulation was performed in the MO space following a localization of the occupied orbitals using the Pipek-Mezey procedure.¹⁹² All PNO and PAO spaces were also

built following the same occupied orbital localization. These simulations were then repeated in both the PNO and PAO virtual spaces, with cutoffs corresponding to average virtual orbital domains containing roughly 20%, 40%, 60%, 80%, and 90% of the untruncated MO virtual space, as well as untruncated PNO- and PAO-basis simulations to ascertain the effects of the virtual space localization on the amplitude dynamics of the wave function.

The effect of local correlation was computed using a simulation approach,¹⁸³ in which all tensor contractions are done in the MO basis. In every CC iteration, before the energy denominator is applied when computing an update to the amplitude tensors, the residuals are transformed into the local basis using \mathbf{Q} and into the semi-canonical basis using \mathbf{L} . These matrices are computed from either Eqs. (5.20) and (5.21), respectively, for PNOs, or Eqs. (5.17) and (5.18), respectively, for PAOs. The residuals in the localized basis \tilde{r}_μ are computed by

$$\tilde{\mathbf{r}}_i = \mathbf{L}_{ii}^T \mathbf{Q}_{ii}^T \mathbf{r}_i \quad (5.23a)$$

$$\tilde{\mathbf{r}}_{ij} = \mathbf{L}_{ij}^T \mathbf{Q}_{in}^T \mathbf{r}_{ij} \mathbf{Q}_{in} \mathbf{L}_{ij} \quad (5.23b)$$

where r_μ are the residuals from Eqs. (5.1a) or (5.1b). Once the energy denominator is applied, the resulting amplitude step is back-transformed into the MO basis, effectively filtering out non-local components. The residual is also filtered in this way, without the application of the energy denominator, in every step of the time propagation. The cc-pVDZ basis set augmented with diffuse functions^{94,95} was used throughout.

For absorption spectra, the imaginary component of the Fourier transform of the *induced* electric dipole $\tilde{\mu} = (\langle \mu \rangle - \mu_0)$ following an electric-field kick may be directly divided by the Fourier transform of the field strength to yield the spectrum. In the case of circular dichroism spectra, however, it is advantageous to first analyze the Fourier transform of the derivative of the field. For a Dirac delta pulse $E_\delta(t) = \kappa\delta(t)$, the Fourier transform of the

derivative yields

$$\text{FFT}[E_\delta] = i\omega\kappa. \quad (5.24)$$

Therefore, for such a field, the CD is proportional to the negative of the *real* part of the Fourier transform of the induced magnetic dipole. In practice, the assumption of a Dirac delta pulse is sufficient, provided a thin Gaussian or Lorentzian pulse is used.

Discrete Fourier transformation was done using a wrapper to the `fft` submodule of the SciPy python library.¹⁹³ All methods were implemented in the Python-based coupled cluster package, PyCC¹⁹⁴, a NumPy-based¹⁹⁵ open-source code developed in the Crawford group for the testing and implementation of novel coupled cluster methods. The code utilizes the Psi4 electronic structure package¹⁹⁶ for integral generation and computing reference wave functions. The RTCC code makes use of the `opt_einsum` package¹⁹⁷ for tensor contractions, and time propagation is performed using an in-house suite of integrators. The integrator used throughout this work was the fourth-order Runge-Kutta method.¹⁹⁰

5.4 Results and Discussion

Here we present results from the first applications of local correlation to RTCC. Results are examined by the convergence of absorption and ECD spectra to the reference results in Section 5.4.1, followed by an analysis of the amplitude dynamics in Section 5.4.2. Also explored is the effect of localization and truncation on the orbital extent, in an attempt to explain the deficiencies of the PNO space, in light of variations in the time-dependent deviations in the amplitudes from the ground-state.

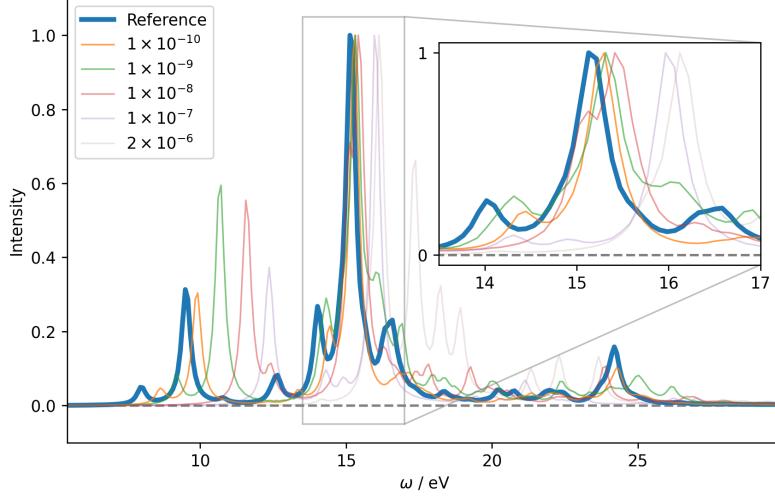


Figure 5.1: Reference and PNO absorption spectra of $(H_2)_4$) for five cutoffs: [1×10^{-10} , 1×10^{-9} , 1×10^{-8} , 1×10^{-7} , 2×10^{-6}] corresponding to [93%, 82%, 63%, 44%, 24%] of the MO virtual space, respectively.

5.4.1 Absorption and ECD Spectra

Absorption

Absorption spectra are obtained from the Fourier transform of Eq. (5.10). Figure 5.1 shows the normalized absorption spectrum obtained from a reference propagation along with five PNO cutoffs. The average truncated virtual orbital spaces are from roughly 20% to 90% of the MO virtual space (see caption). Overall, truncated PNO virtual spaces approximate the position of the base peak well, with the smallest space predicting a base peak within 1.5 eV of the reference, and the two largest spaces predict this peak to within 0.2 eV of the reference. Convergence to the reference base peak occurs from the right, indicating a lowering of excited state energies as the size of the virtual space increases. This trend can also be seen for the smaller peak near 10 eV. However, convergence of the shoulder peaks on either side of the base peak, indicated by the inset of Figure 5.1, is less predictable. Even the largest spaces

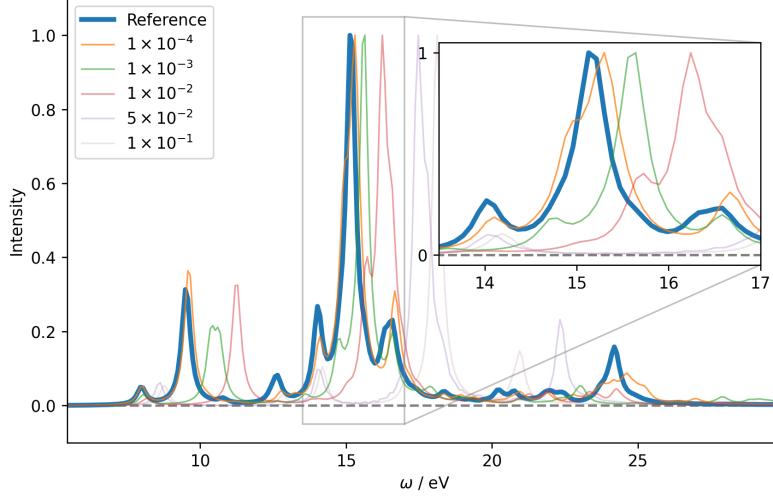


Figure 5.2: Reference and PAO absorption spectra of $(H_2)_4$ for five cutoffs: [1×10^{-4} , 1×10^{-3} , 1×10^{-2} , 5×10^{-2} , 1×10^{-1}] corresponding to [95%, 86%, 63%, 46%, 23%] of the MO virtual space, respectively.

considered do not correctly predict the excitation energy, with no clear advantage to having 93% of the virtual space as compared to just 83% for predicting these peaks. This trend continues into the higher-energy range of the spectrum, with the performance of each cutoff being nearly indistinguishable.

Performance of the PAO space is shown in Figure 5.2. The largest truncated PAO virtual space, on average 95% of the MO space, accurately predicts the excitation energies for each major peak below 17 eV. Particularly around 10 eV, this is noticeably improved performance relative to the largest PNO space tested, with only a 2% difference in the average size of the virtual space. However, accuracy rapidly declines even at 86% of the virtual space, where the base peak position is already worse than what was predicted with a PNO space of just 63% of the MO space. Performance continues to degrade as energy increases and the average size of the PAO space decreases. For the final two cutoffs, at averages of 46% and 23% of the MO space, the base peaks are 3 eV or more away from the reference, and no peak is

exhibited near 25 eV. These spaces also fail to predict the second largest peak, the excitation just below 10 eV.

ECD

Overall, neither scheme produced adequate results upon truncation of the virtual space. This result is not entirely surprising – in studies of local correlation applied to response theory by Crawford *et al.*,^{13,61,62,143} traditional schemes proved inaccurate for another electric dipole–electric dipole property, the electric polarizability. In terms of response theory, the polarizability (and the refractive index) is related to the *real* part of the electric dipole–electric dipole linear response tensor (α_{ij} in Eq. (5.9a)), while absorption is related to the *imaginary* part. Indeed, all linear absorptive properties such as absorption and CD are related to the imaginary component of a linear response tensor, while dispersive properties such as refractive index and circular birefringence (also known as optical rotation) are related to the real component.^{1,21} To continue, we will look at another absorptive property which is related to the mixed electric dipole–magnetic dipole linear response tensor – ECD.

The ECD spectrum is obtained from the Fourier transform of Eq. (5.11). Being a bisignate, mixed-response property, ECD is a considerable computational challenge, similar to its dispersive counterpart circular birefringence. Figure 5.3 shows the results for an ECD spectrum in the same PNO orbital spaces used in the previous section. The dynamic response of the magnetic dipole to the electric field in this frequency range is considerably more complicated than that of the electric dipole. Below 60% of the MO space, virtually all distinguishing characteristics of the reference spectrum are unidentifiable. Further, at 82%, the base peak appears to be a pair of peaks, more resembling the pair of peaks appearing just above 15 eV in the reference spectrum, with the major peak just below 15 eV being the second strongest. At an average of 93%, the overall *shape* of the spectrum in the 10 eV to 20 eV range more

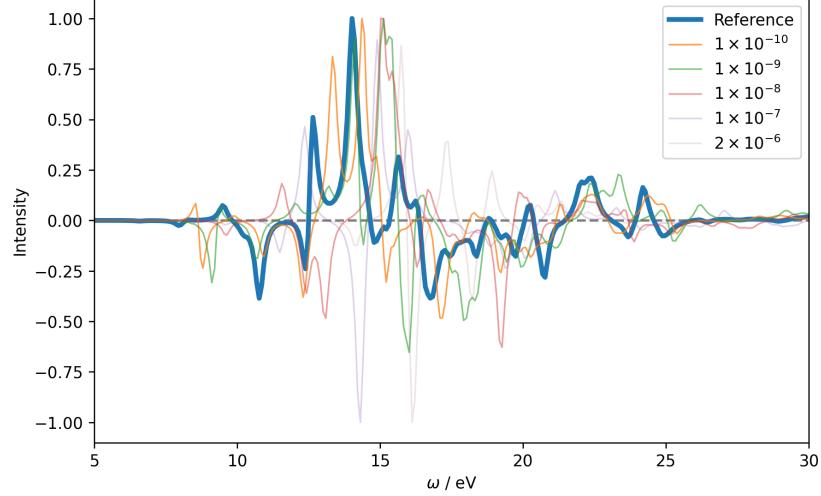


Figure 5.3: Reference and PNO ECD spectra of $(H_2)_4$ for five cutoffs: [1×10^{-10} , 1×10^{-9} , 1×10^{-8} , 1×10^{-7} , 2×10^{-6}] corresponding to [93%, 82%, 63%, 44%, 24%] of the MO virtual space, respectively.

closely resembles that of the reference; however, the excitation energies are, in some cases, even less accurate than those of smaller PNO spaces. The trend of lowering excited state energies with increased virtual space seen in Section 5.4.1 is no longer discernible.

As in the case of absorption, the PAO basis is not noticeably more efficient at approximating the full MO space than the PNO space. Figure 5.4 shows the results using the same truncated PAO spaces as in Section 5.4.1. The performance of the PAO basis near the base peak varies wildly with truncation, as in the PNO case. In the low-frequency region, the PAO results are considerably worse – see the two negative peaks between 10 eV and 15 eV. Curiously, the largest PAO spaces considered predict significant peaks above 25 eV that are not present in the reference, any of the PNO spaces tested, or the smaller PAO spaces. This suggests a strong sensitivity of the response of the wave function to the completeness threshold used for determining the occupied domains.

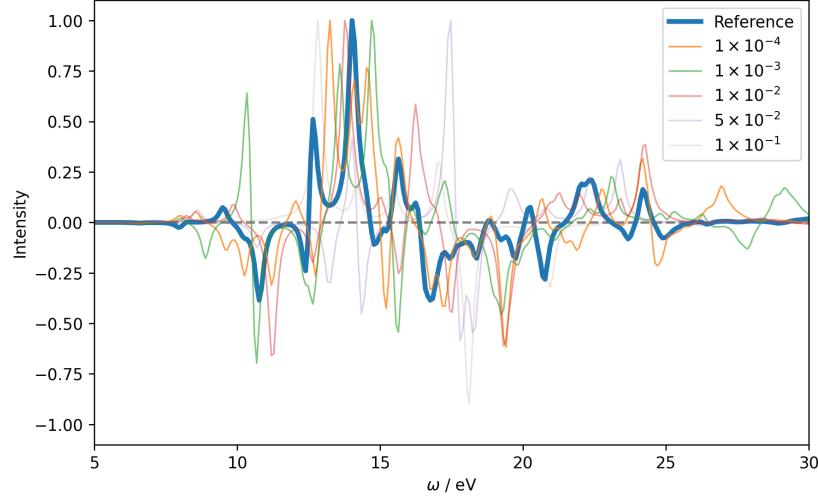


Figure 5.4: Reference and PAO ECD spectra of $(H_2)_4$ for five cutoffs: $[1 \times 10^{-4}, 1 \times 10^{-3}, 1 \times 10^{-2}, 5 \times 10^{-2}, 1 \times 10^{-1}]$ corresponding to $[95\%, 86\%, 63\%, 46\%, 23\%]$ of the MO virtual space, respectively.

5.4.2 Amplitude Dynamics

As evidenced by the preceding data, the truncated PNO and PAO virtual spaces do not efficiently model the wave function in the presence of a perturbing EMF. As noted in Section 5.4.1, these shortcomings are well-documented in the case of response theory. However, a real-time formalism offers the opportunity to analyze the wave function in great detail over time, perhaps shedding light on *where* and *how* the locally correlated wave functions are deficient. The following section will scrutinize the t_μ and λ_μ amplitudes of Eqs. (5.3b) and (5.4b), respectively, in hopes of determining the important fluctuations in the wave function and whether these spaces sufficiently capture these changes.

Response to external perturbations by the CC amplitudes give rise to dynamic energetics and properties. In the past, distributions of perturbed amplitudes (relative to their ground-state counterparts) have been used to justify the difficulty in computing response functions with local correlation methods in the frequency domain.^{13,61,143} However, initial findings

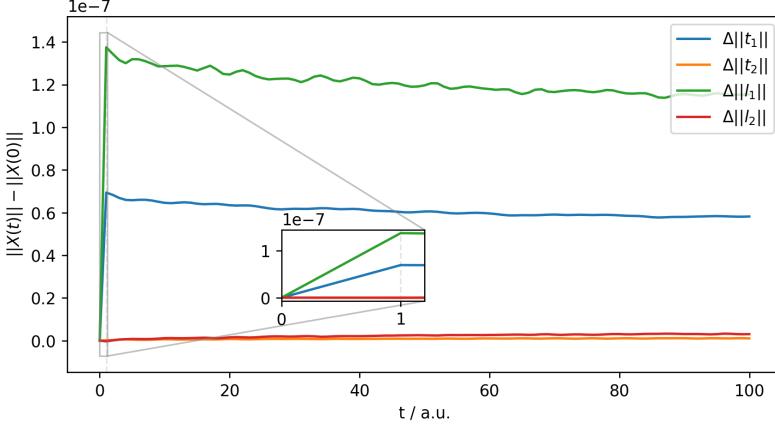


Figure 5.5: Time-dependent change in the norm of the amplitude tensors relative to the ground-state amplitudes. (Field and step parameters remain unchanged, and the amplitude norm is taken at every 1 a.u.)

show that in RTCC, the relative distribution of amplitudes by magnitude is not significantly impacted.¹³ Despite this, typical means of exploiting amplitude sparsity have been shown to be inefficient by the preceding sections. First, to understand the response of the amplitudes to the external perturbation, we plot the change in the norm of the amplitude tensors relative to the ground-state amplitudes as a function of time in Figure 5.5. Results for the untruncated PNO and PAO spaces are identical to those for the MO space, as the unitary transformations resulting from untruncated localized virtual spaces in Eqs. (5.23) preserves the tensor norm. Amplitude norms from propagations carried out in truncated PNO and PAO spaces are nearly indistinguishable (see SI).

Figure 5.5 shows that the magnitude of the response by the wave function is predominantly within the singles amplitudes t_1 and λ_1 . This is consistent with the notion that singles are paramount for the computation of response properties.^{32,33} However, the form of Eq. (5.19) does not include any contributions by singles, due to being built from MP2-level amplitudes where singles do not contribute until at least the second order in the wave function and fourth order in the energy. This suggests that even in schemes which seek to include the

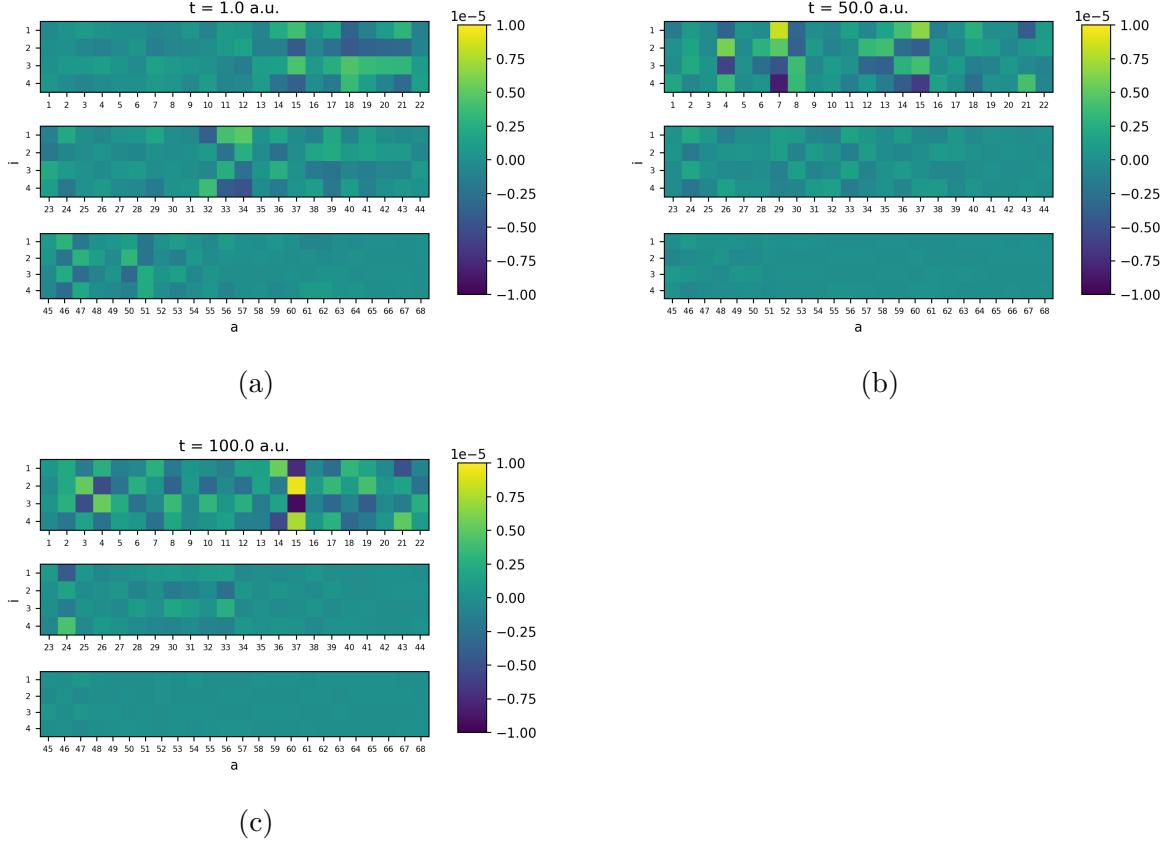


Figure 5.6: MO-basis t_1 amplitude deviations from $t = 0$ after (a) 1 a.u., (b) 50 a.u., and (c) 100 a.u. of time propagation. Each row contains the same four occupied orbital indices and a subset of virtual indices as indicated by the x-axis labels.

EMF perturbation in the construction of the reduced virtual space, such as PNO++,¹⁴³ response of the singles should be considered.

Aside from the matrix norm, we can also inspect the individual amplitudes to track their evolution in time. The heat maps in Figure 5.6 show the difference in t_1 amplitudes, relative to the ground state, for three time steps selected from the first 100 a.u. of the simulation. The amplitudes are ordered by the orbital energies of the associated MOs. The amplitudes which experience significant oscillations vary throughout the simulation, though there are several discernible trends. First, most large amplitude deviations are associated with all occupied orbitals simultaneously. This is due to the relatively small size of the system, with

only four occupied orbitals, all of which are likely important in the description of the ground- and excited-state wave functions. Secondly, at any given time during the propagation, a large number of amplitudes have not significantly deviated from their ground state values. This supports the notion that relative sparsity is maintained within the amplitudes throughout the simulation, but this sparsity is distributed differently throughout the amplitude tensors as the wave function is propagated.

A third trend is that amplitudes which respond strongly tend to be associated with low-energy virtual orbitals. Chemical intuition would suggest that energetically low-lying molecular orbitals will be the most involved in electronic excitations. However, while amplitude responses are indeed larger for lower-energy virtual orbitals, smaller amplitude deviations in Figure 5.6 extend far into the virtual space. This explains the difficulty of simply truncating with respect to orbital energy: the high-energy MOs are still important to the time-evolution of the wave function in the presence of an EMF.

Figure 5.7 shows the t_1 amplitudes for the same simulation, rotated into the untruncated PNO basis using Q_{ii} as defined in Eq. (5.20). (It should be noted that, due to redundancy in the AO-based virtual spaces for each pair, PAO-basis amplitudes cannot be compared directly in this manner.) It can be immediately seen that the amplitude deviations are less sparse in the PNO basis after the application of the EMF. Many more amplitudes exhibit perceivable differences, and strong deviations (magnitudes approaching 1×10^{-5}) are no longer present. This is a clear demonstration of the issue with truncating orbital spaces based on the present criterion — rather than exploiting sparsity, the amplitude tensors have become less sparse.

Virtual orbital extent has been used previously^{62,143} to estimate the ability of locally correlated spaces to describe the diffuse regions of electron density which are important for response properties. Figure 5.8 shows the virtual MO energy ϵ_a and the PNO occupation

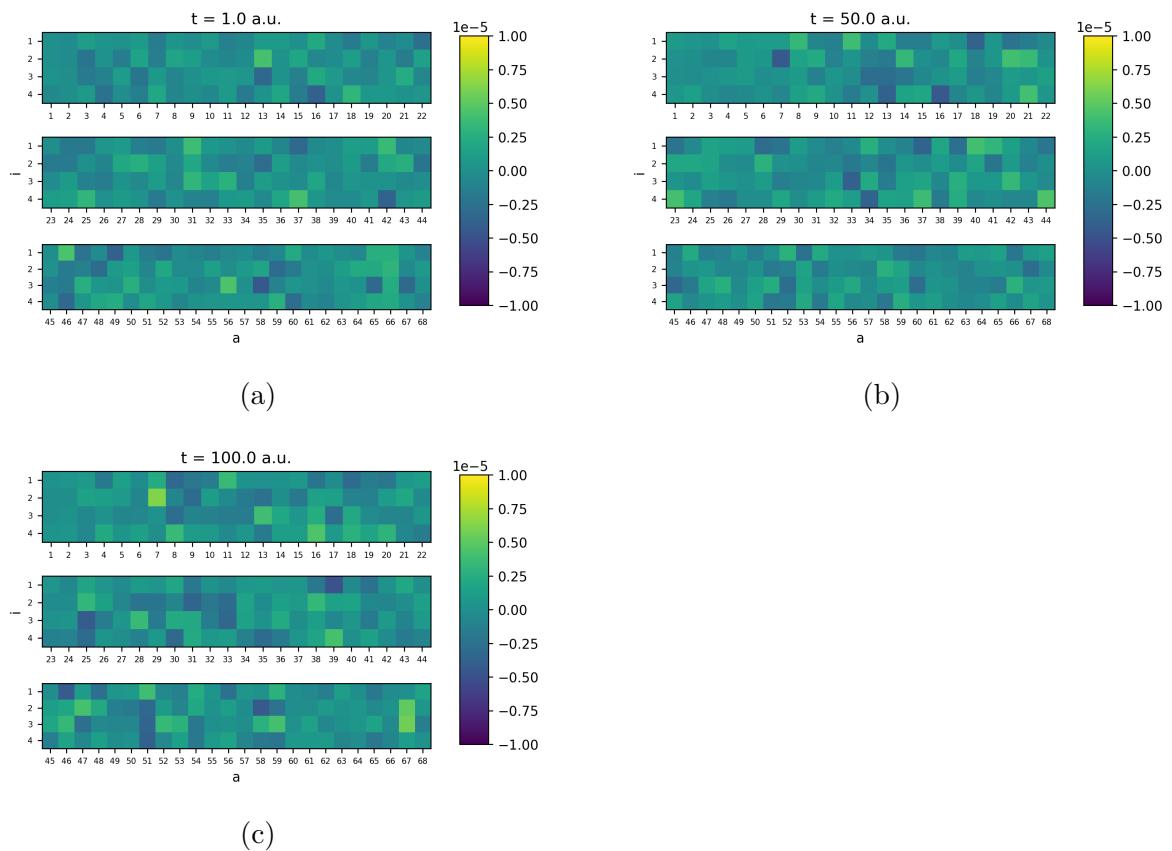


Figure 5.7: PNO-basis t_1 amplitude deviations from $t = 0$ after (a) 1 a.u., (b) 50 a.u., and (c) 100 a.u. of time propagation. Each row contains the same four occupied orbital indices and a subset of virtual indices as indicated by the x-axis labels.

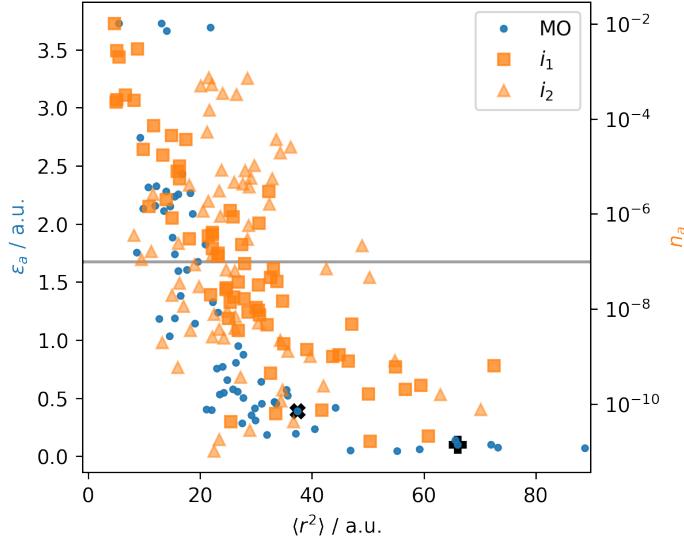


Figure 5.8: Virtual MO energy ϵ_a and the occupation number n_a (plotted on a log scale) for unique PNO spaces i_1 and i_2 versus orbital extent. Virtual MOs 7 and 15 are denoted by a solid + and \times , respectively. The horizontal line denotes a PNO cutoff of 1×10^{-7} .

number n_a plotted against the orbital extent $\langle r^2 \rangle$ in a.u. In the PNO basis, a unique virtual space is prepared for every occupied pair, resulting in 16 unique spaces for the four occupied spatial orbitals i . However, for transforming a single orbital index, we only require the diagonal rotation matrices, *i.e.*, Q_{ii} . There are four such spaces; however, by symmetry, only two are unique. Both are included in Figure 5.8.

Truncation of the PNO space begins from the bottom of Figure 5.8. At an occupation number cutoff of 1×10^{-7} (indicated by a horizontal line), all orbitals below this line are neglected in the PNO space. Roughly 66% of the virtual space lies in this region. From these data, it is clear that even modest truncation of the virtual space neglects the diffuse regions of the wave function, which are important for excited-state properties in systems with significantly delocalized characteristics, such as systems containing Rydberg-type excitations.

Spatial extent alone may not be a suitable criterion for truncation — this would have a negative impact on the accuracy of the correlation energy, which is inherently local in nature.

Additionally, contracted orbitals may very well contribute significantly to the response of the wave function. Figure 5.8 highlights virtual MOs 7 (+) and 15 (\times). These orbitals correspond to the strongest deviations in Figures 5.6b and 5.6c, respectively. These deviations represent a strong response of the wave function to the perturbing field, implying they are of particular importance when computing dynamic properties. Examination of the spatial extents of these orbitals in particular may shed light on the nature of the spatial distribution of orbitals necessary to describe the wave function response. Virtual MO 7 appears at 66 a.u., while MO 15 is nearly half that at 37 a.u. That these orbitals are of such varying extent demonstrates that both diffuse and contracted orbitals play a role in the wave function dynamics.

5.5 Conclusions

Here we present the first application of local correlation to RTCC simulations. The popular PNO and PAO virtual space localization schemes are applied to the calculation of dynamic electric and magnetic dipole moments in the presence of an explicit electric field, providing absorption and ECD spectra, respectively. For a helical H₂ tetramer, truncation of the localized virtual space to successively larger fractions of the canonical virtual space resulted in convergence to the canonical result; however, this convergence is slow, and errors in excitation energies and intensity are present even in some of the largest spaces tested, especially for ECD. This corroborates the results of recent studies applying locally correlated methods to the prediction of dynamic properties in the frequency domain using response theory.

Examining the amplitude dynamics during the propagation, it is shown that the t_1 and λ_1 amplitudes respond most strongly to the field – a large increase in the norm of these matrices is observed upon application of the field, followed by a steady oscillation. The t_2

and λ_2 tensors, by comparison, remain relatively static throughout. These oscillations are largely, but not completely, localized to a selection of only a few orbitals, as evidenced by consideration of time-dependent deviations in the t_1 amplitudes from the ground-state. In the localized virtual spaces tested, these oscillations are delocalized throughout the t_1 and λ_1 matrices.

Orbital extent alone cannot explain the shortcomings of the PNO space; however, its effect is significant. These results provide an insight into the importance of singly-substituted determinants in the time-dependent wave function in the presence of an electric field, as well as a potential metric to gauge the performance of new localization schemes for frequency- or time-domain calculations of dynamic response properties. In order to attain a balanced description of wave function components important for both energy and property calculations, the combination of appropriately determined spaces such as the combined PNO++ approach has been fruitful.¹⁴³ Still neglected in this approach are the singles amplitudes, which are absent in the MP2 wave functions used to approximate the occupied pair domains. Schemes to include these effects, such as approximate CC2-level t_1 guess amplitudes, may further improve the space and allow greater flexibility for truncation. The prospect of utilizing these methodologies within the current framework is promising, and work is underway to explore their efficiency.

Chapter 6

Conclusions

The time- or frequency-dependent response of a system to electromagnetic fields gives rise to many different dynamic properties which can be experimentally probed by various forms of spectroscopy. Theoretical models of electronic structure can provide assistance in the design and characterization of novel materials through *ab initio* computation of the light-matter interactions which govern this response. The accurate computation of these properties on modern computing hardware with correlated methods, such as coupled cluster theory, is often prohibitively expensive for all but the smallest systems. These computations typically employ a semi-classical approach, which models the response of a system, typically treated as a quantum-mechanical wave function or density, to an external electromagnetic field, represented as a classical potential. The current work seeks to improve the efficiency of these methods through considerations of both frequency- and time-domain formulations.

After discussing the basics of correlated electronic structure treatments and the foundations of semi-classical approaches to property prediction, three studies were presented. First, we examine the effects of basis set superposition error on the many-body expansion of molecular properties, computed using the frequency-domain response theory approach. Previous efforts to apply this divide-and-conquer approach to dynamic molecular properties proved difficult – specifically, slow, oscillatory convergence to the analytic result was observed, especially in the case of optical rotation calculations. Literature precedent suggests that the correction of basis set superposition errors, along with careful considerations of other methodological

parameters such as numerical precision, alleviate oscillatory convergence issues commonly encountered when calculating binding energies of large water clusters through a many-body approach. While these corrections proved to dampen the oscillations, the troublesome case of optical rotation for systems in modest solvent shells could still not be reproduced with computationally tractable expansions. Thus, fragmentation-type approaches such as the many-body expansion are of limited utility in such cases.

Moving beyond simple fragmentation of the system, a condensed wave function representation of the entire system could allow for the rapid evaluation of target properties. The second study utilized a machine-learning approach. The input is a representation of molecular systems, developed using wave function-based descriptors. This representation, dubbed the density tensor representation, was derivative of the recently published *t*-amplitude tensor representation, but unique in its construction by exploiting the naturally dense representation of the wave function by the reduced density matrix. A simple machine-learning method, kernel ridge regression, was applied to compute the ground-state energies and electric dipole moments of several small organic molecules across molecular dynamics trajectories. MP2-level densities were used to build representations, and the training set consisted of CCSD-level ground-state correlated energies and electric dipoles. Utilizing new implementations of both *t*-amplitude and density tensor representations in an open-source, python-based code proved the density tensor representation to be the more efficient representation of the two, requiring less than a dozen training calculations to provide results within chemical accuracy across short trajectories for the systems considered. Proposed extensions of the procedure to dynamic properties were explored, and the implementation of these methods is ongoing.

Using frequency-domain response theory to compute training set values results in certain limitations. This perturbative approach excludes the possibility of strong electromagnetic fields, such as those used in ultrafast spectroscopy. These properties may be computed using

a non-perturbative, explicitly time-dependent approach, referred to as real-time electronic structure theory. These methods require the expensive (and possibly unstable) propagation of the wave function in time. For real-time coupled cluster, this means repeatedly evaluating the high-degree polynomial-scaling residual expressions for long periods of time. To alleviate the cost, the third and final study focused on the application of the local correlation family of reduced-scaling methods to real-time coupled cluster.

An open-source coupled cluster development code based on the Psi4 electronic structure package was developed which can simulate the effects of various local correlation schemes, namely PAO and PNO, with minimal modifications of the underlying expressions. The code is capable of modeling explicit electric fields, and simulations of both absorption and electronic circular dichroism spectra are used to test the efficiency of the PAO and PNO approaches. These tests revealed deficiencies in the virtual orbital spaces produced by truncating the wave function in these locally correlated bases, which are corroborated by recent results of their application in frequency-domain response theory. The explicitly time-dependent nature of the wave function allowed for the detailed investigation of separate wave function elements as they evolved in time, revealing a strong dependence on single substitutions, particularly in diffuse orbital spaces. These substitutions are not considered in the typical construction of PAOs or PNOs. These bases are also intentionally spatially localized, since they are designed to capture the correlation energy, which is inherently local for insulators and other weakly-correlated systems. Alternative approaches, such as the recently-proposed PNO++ method, may address these problems, allowing for more aggressive truncation of the virtual orbital space, leading to far cheaper evaluation of the expensive residual expressions required for time propagation.

The work presented here comprises several different methods for improving the efficiency of accurate, systematically improvable computations of field-dependent molecular properties.

All solutions have been implemented as open-source software, which can be freely accessed, used, and extended. Continued development on the methods presented in Chapters 4 and 5, in particular, are encouraging. In the near term, these and other improvements will continue to close the gap between experimental and theoretical understandings of light-matter interactions. In the far term, as this gap is closed and sufficient computational resources become more widely available, the theoretical prediction of field-induced properties will become an indispensable tool for future synthetic chemists, whose interpretations of spectroscopic results will be improved by a fundamental understanding of these interactions at a quantum level.

Bibliography

- [1] Barron, L. D. *Cambridge Univ. Press*; Cambridge University Press, 2004.
- [2] Helgaker, T.; Coriani, S.; Jørgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. Recent advances in wave function-based methods of molecular-property calculations. *Chem. Rev.* **2012**, *112*, 543–631.
- [3] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications, 1996.
- [4] Born, M.; Oppenheimer, R. Zur Quantentheorie der Moleküle. *Ann. Phys.* **1927**, *389*, 457–484.
- [5] Shavitt, I.; Bartlett, R. J. *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*; Cambridge Molecular Science; Cambridge University Press, 2009.
- [6] Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618–622.
- [7] Sinanoğlu, O. *Adv. Chem. Phys.*; John Wiley & Sons, Ltd, 1964; Vol. 6; pp 315–412.
- [8] Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, *45*, 4256–4266.
- [9] Čížek, J. *On the Use of the Cluster Expansion and the Technique of Diagrams in Calculations of Correlation Effects in Atoms and Molecules*; John Wiley & Sons, Ltd, 1969; pp 35–89.

- [10] Crawford, T.; Schaefer III, H. An introduction to coupled cluster theory for computational chemists. *Rev. Comput. Chem.* Vol. 14 **2000**, 14, 33–136.
- [11] Hoodbhoy, P.; Negele, J. W. Time-dependent coupled-cluster approximation to nuclear dynamics. II. General formulation. *Phys. Rev. C* **1979**, 19, 1971–1982.
- [12] Crawford, T. D. Ab initio calculation of molecular chiroptical properties. *Theor. Chem. Acc.* **2006**, 115, 227–245.
- [13] Crawford, T. D.; Kumar, A.; Bazanté, A. P.; Di Remigio, R. Reduced-scaling coupled cluster response theory: Challenges and opportunities. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2019**, 9, 1–25.
- [14] Davidson, E. R. In *Reduced Density Matrices in Quantum Chemistry*; Davidson, E. R., Ed.; Theoretical Chemistry; Elsevier, 1976; Vol. 6.
- [15] Harris, F. E.; Monkhorst, H. J.; Freeman, D. L. *Algebraic and diagrammatic methods in many-fermion theory*; Oxford University Press, 1992.
- [16] Helgaker, T.; Jørgensen, P.; Olsen, J. *Mol. Electron. Theory*; John Wiley & Sons, Ltd, 2000; Chapter 2, pp 34–79.
- [17] Langhoff, P. W.; Epstein, S. T.; Karplus, M. Aspects of time-dependent perturbation theory. *Rev. Mod. Phys.* **1972**, 44, 602–644.
- [18] Koch, H.; Jørgensen, P. Coupled cluster response functions. *J. Chem. Phys.* **1990**, 93, 3333–3344.
- [19] Pedersen, T. B.; Koch, H. Coupled cluster response functions revisited. *J. Chem. Phys.* **1997**, 106, 8059–8072.

- [20] Christiansen, O.; Jorgensen, P.; Hättig, C. Response functions from Fourier component variational perturbation theory applied to a time-averaged quasienergy. *Int. J. Quantum Chem.* **1998**, *68*, 1–52.
- [21] Norman, P. A perspective on nonresonant and resonant electronic response theory for time-dependent molecular properties. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20519.
- [22] Goings, J. J.; Lestrange, P. J.; Li, X. Real-time time-dependent electronic structure theory. 2018.
- [23] Li, X.; Govind, N.; Isborn, C.; Deprince, A. E.; Lopata, K. Real-Time Time-Dependent Electronic Structure Theory. *Chem. Rev.* **2020**, *120*, 9951–9993.
- [24] Werner, H. J.; Pflüger, K. On the Selection of Domains and Orbital Pairs in Local Correlation Treatments. *Annu. Rep. Comput. Chem.* **2006**, *2*, 53–80.
- [25] Pulay, P.; Saebø, S. Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory. *Theor. Chim. Acta* **1986**, *69*, 357–368.
- [26] Bartlett, R. J.; Silver, D. M. Pair-correlation energies in sodium hydride with many-body perturbation theory. *Phys. Rev. A* **1974**, *10*, 1927–1931.
- [27] Pulay, P.; Saebø, S. Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory. *Theor. Chim. Acta* **1986**, *69*, 357–368.
- [28] Surján, P. R. *Second Quantized Approach to Quantum Chem.*; Springer-Verlag Berlin Heidelberg, 1989; pp 103 – 113.
- [29] Boughton, J. W.; Pulay, P. Comparison of the Boys and Pipek-Mezey Localizations in the Local Correlation Approach and Automatic Virtual Basis Selectiion. *J. Comput. Chem.* **1993**, *14*, 736–740.

- [30] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- [31] Bartlett, R. J. In *Theory Appl. Comput. Chem.*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 1191–1221.
- [32] Christiansen, O.; Koch, H.; Jørgensen, P. The second-order approximate coupled cluster singles and doubles model CC2. *Chem. Phys. Lett.* **1995**, *243*, 409–418.
- [33] Koch, H.; Christiansen, O.; Jørgensen, P.; Sanchez De Merás, A. M.; Helgaker, T. The CC3 model: An iterative coupled cluster approach including connected triples. *J. Chem. Phys.* **1997**, *106*, 1808–1818.
- [34] Trucks, G. W.; Salter, E. A.; Sosa, C.; Bartlett, R. J. Theory and implementation of the MBPT density matrix. An application to one-electron properties. *Chem. Phys. Lett.* **1988**, *147*, 359–366.
- [35] Hellmann, H. *Einführung in die Quantenchemie*; Deuticke: Leipzig, 1937.
- [36] Feynman, R. P. Forces in molecules. *Phys. Rev.* **1939**, *56*, 340–343.
- [37] Pedersen, T. B.; Kristiansen, H. E.; Bodenstein, T.; Kvaal, S.; Schøyen, Ø. S. Interpretation of Coupled-Cluster Many-Electron Dynamics in Terms of Stationary States. *J. Chem. Theory Comput.* **2021**, *17*, 388–404.
- [38] Wiberg, K. B.; Wang, Y. G.; Wilson, S. M.; Vaccaro, P. H.; Cheeseman, J. R. Sum-over-states calculation of the specific rotations of some substituted oxiranes, chloropropionitrile, ethane, and norbornenone. *J. Phys. Chem. A* **2006**, *110*, 13995–14002.

- [39] Crawford, T. D.; Tam, M. C.; Abrams, M. L. The current state of ab initio calculations of optical rotation and electronic circular dichroism spectra. *J. Phys. Chem. A* **2007**, *111*, 12057–12068.
- [40] J.Tomasi,; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027–2094.
- [41] Gao, J. Hybrid QM/MM Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. *Acc. Chem. Res.* **1996**, *29*, 298–305.
- [42] Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.
- [43] Tomasi, J.; Cammi, R.; Mennucci, B.; Cappelli, C.; Corni, S. Molecular Properties in Solution Described with a Continuum Solvation Model. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5697–5712.
- [44] Mennucci, B.; Tomasi, J.; Cammi, R.; Cheeseman, J. R.; Frisch, M. J.; Devlin, F. J.; Gabriel, S.; Stephens, P. J. Polarizable Continuum Model (PCM) Calculations of Solvent Effects on Optical Rotation of Chiral Molecules. *J. Phys. Chem. A* **2002**, *106*, 6102–6113.
- [45] Mukhopadhyay, P.; Zuber, G.; Goldsmith, M.-R.; Wipf, P.; Beratan, D. N. Solvent Effect on Optical Rotation: A Case Study of Methyloxirane in Water. *ChemPhysChem* **2006**, *7*, 2483–2486.
- [46] Mukhopadhyay, P.; Zuber, G.; Wipf, P.; Beratan, D. N. Contribution of a Solute's Chiral Colvent Imprint to Optical Rotation. *Angew. Chemie - Int. Ed.* **2007**, *46*, 6450–6452.

- [47] Lipparini, F.; Egidi, F.; Cappelli, C.; Barone, V. The Optical Rotation of Methyloxirane in Aqueous Solution: A Never Ending Story? *J. Chem. Theory Comput.* **2013**, *9*, 1880–1884.
- [48] Egidi, F.; Giovannini, T.; Del Frate, C.; Lemler, P. M.; Vaccaro, P. H.; Cappelli, C. A Combined Experimental and Theoretical Study of Optical Rotatory Dispersion for (*R*)-glycidyl Methyl Ether in Aqueous Solution. *Phys. Chem. Chem. Phys.* **2019**, *21*, 3644–3655.
- [49] Pecul, M.; Ruud, K. The Ab Initio Calculation of Optical Rotation and Electronic Circular Dichroism. *Adv. Quantum Chem.* **2005**, *50*, 185–212.
- [50] Crawford, T. D. In *Comprehensive Chiroptical Spectroscopy*; Berova, N., Nakanishi, K., Woody, R. W., Polavarapu, P., Eds.; Wiley: Hoboken, New Jersey, 2012; Vol. 1; Chapter 23, pp 675–697.
- [51] Daniel Crawford, T. In *Front. Quantum Chem.*; Wójcik, M. J., Nakatsuji, H., Kirtman, B., Ozaki, Y., Eds.; Springer Singapore: Singapore, 2018; pp 49–68.
- [52] Kongsted, J.; Pedersen, T. B.; Strange, M.; Osted, A.; Hansen, A. E.; Mikkelsen, K. V.; Pawłowski, F.; Jørgensen, P.; Hättig, C. Coupled Cluster Calculations of the Optical Rotation of *S*-propylene Oxide in Gas Phase and Solution. *Chem. Phys. Lett.* **2005**, *401*, 385–392.
- [53] Kongsted, J.; Pedersen, T. B.; Jensen, L.; Hansen, A. E.; Mikkelsen, K. V. Coupled Cluster and Density Functional Theory Study of the Vibrational Contribution to the Optical Rotation of (*S*)-propylene Oxide. *J. Am. Chem. Soc.* **2006**, *128*, 976–982.
- [54] Howard, J. C.; Sowndarya, S.; Ansari, I.; Mach, T. J.; Baranowska-Alaczkowska, A.;

- Crawford, T. D. On the Performance of Property-Optimized Basis Sets for Optical Rotation With Coupled Cluster Theory. *J. Phys. Chem. A* **2018**, *122*, 5962–5969.
- [55] Govind, N.; Wang, Y. A.; da Silva, A. J. R.; Carter, E. A. Accurate Ab Initio Energetics of Extended Systems via Explicit Correlation Embedded in a Density Functional Environment. *Chem. Phys. Lett.* **1998**, *295*, 129–134.
- [56] Neugebauer, J.; Jacob, C. R.; Wesolowski, T. A.; Baerends, E. J. An Explicit Quantum Chemical Method for Modeling Large Solvation Shells Applied to Aminocoumarin C151. *J. Phys. Chem. A* **2005**, *109*, 7805–7814.
- [57] Jacob, C. R.; Neugebauer, J.; Visscher, L. A Flexible Implementation of Frozen-Density Embedding for Use in Multilevel Simulations. *J. Comp. Chem.* **2007**, *29*, 1011–1018.
- [58] Gomes, A. S. P.; Jacob, C. R.; Visscher, L. Calculation of local excitations in large systems by embedding wave-function theory in density-functional theory. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5353–5362.
- [59] Crawford, T. D.; Kumar, A.; Hannon, K. P.; Höfener, S.; Visscher, L. Frozen-Density Embedding Potentials and Chiroptical Properties. *J. Chem. Theory Comput.* **2015**, *11*, 5305–5315.
- [60] Crawford, T. D. In *Recent Prog. Coupled Clust. Methods Theory Appl.*; Cársky, P., Paldus, J., Pittner, J., Eds.; Springer Netherlands: Dordrecht, 2010; pp 37–55.
- [61] McAlexander, H. R.; Crawford, T. D. A Comparison of Three Approaches to the Reduced-Scaling Coupled Cluster Treatment of Non-Resonant Molecular Response Properties. *J. Chem. Theory Comput.* **2016**, *12*, 209–222.
- [62] Kumar, A.; Crawford, T. D. Frozen Virtual Natural Orbitals for Coupled-Cluster Linear-Response Theory. *J. Phys. Chem. A* **2017**, *121*, 708–716.

- [63] Ouyang, J. F.; Cvitkovic, M. W.; Bettens, R. P. A. Trouble with the Many-Body Expansion. *J. Chem. Theory Comput.* **2014**, *10*, 3699–3707.
- [64] Richard, R. M.; Un Lao, K.; Herbert, J. M. Understanding the Many-Body Expansion for Large Systems. I. Precision Considerations. *J. Chem. Phys.* **2014**, *141*, 14108.
- [65] Ouyang, J. F.; Bettens, R. P. A. When are Many-Body Effects Significant? *J. Chem. Theory Comput.* **2016**, *12*, 5860–5867.
- [66] Richard, R. M.; Bakr, B. W.; Sherrill, C. D. Understanding the Many-Body Basis Set Superposition Error: Beyond Boys and Bernardi. *J. Chem. Theory Comput.* **2018**, *14*, 2386–2400.
- [67] Liu, K.-Y.; Herbert, J. M. Understanding the Many-Body Expansion for Large Systems. III. Critical Role of Four-Body Terms, Counterpoise Corrections, and Cutoffs. *J. Chem. Phys.* **2017**, *12*, 5860–5867.
- [68] Boys, S.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553–566.
- [69] Wells, B. H.; Wilson, S. Van der Waals Interaction Potentials: Many-Body Basis Set Superposition Effects. *Chem. Phys. Lett.* **1983**, *101*, 429–434.
- [70] Valiron, P.; Mayer, I. Hierarchy of Counterpoise Corrections for N-body Clusters: Generalization of the Boys-Bernardi Scheme. *Chem. Phys. Lett.* **1997**, *275*, 46–55.
- [71] Salvador, P.; Szczesniak, M. M. Counterpoise-Corrected Geometries and Harmonic Frequencies of N-body Clusters: Application to (HF)_n (n=3,4). *J. Chem. Phys.* **2003**, *118*, 537–549.

- [72] Kamiya, M.; Hirata, S.; Valiev, M.; Kamiya, M.; Hirata, S.; Valiev, M. Fast Electron Correlation Methods for Molecular Clusters without Basis Set Superposition Errors. *J. Chem. Phys.* **2008**, *128*, 074103.
- [73] Richard, R. M.; Lao, K. U.; Herbert, J. M. Achieving the CCSD(T) Basis-Set Limit in Sizable Molecular Clusters: Counterpoise Corrections for the Many-Body Expansion. *J. Phys. Chem. Lett.* **2013**, *2680*, 4–10.
- [74] Ouyang, J. F.; Bettens, R. P. A. Many-Body Basis Set Superposition Effect. *J. Chem. Theory Comput.* **2015**, *11*, 5132–5143.
- [75] Mayer, I.; Bakó, I. Many-Body Energy Decomposition with Basis Set Superposition Error Corrections. *J. Chem. Theory Comput.* **2017**, *13*, 1883–1886.
- [76] Hopkins, B. W.; Tschumper, G. S. A Multicentered Approach to Integrated QM/QM Calculations. Applications to Multiply Hydrogen Bonded Systems. *J. Comput. Chem.* **2003**, *24*, 1563–1568.
- [77] Howard, J. C.; Tschumper, G. S. N-body:Many-body QM:QM Vibrational Frequencies: Application to Small Hydrogen-bonded Clusters. *J. Chem. Phys.* **2013**, *139*, 144305.
- [78] Hankins, D.; Moskowitz, J. W.; Stillinger, F. H. Water Molecule Interactions. *J. Chem. Phys.* **1970**, *53*, 4544–4554.
- [79] Kaplan, I. G. *Theory of Molecular Interactions*; Elsevier: Amsterdam, 1986.
- [80] Skwara, B.; Zawada, A.; Bartkowiak, W. On the Many-Body Components of Interaction-Induced Electric Properties: Linear Fluoroacetylene Trimer as a Case Study. *Comput. Lett.* **2007**, *3*, 175–182.

- [81] Skwara, B.; Bartkowiak, W.; Da Silva, D. L. On the Basis Set Superposition Error in Supermolecular Calculations of Interaction-induced Electric Properties: Many-Body Components. *Theor. Chem. Acc.* **2009**, *122*, 127–136.
- [82] Baranowska, A.; Zawada, A.; Fernández, B.; Bartkowiak, W.; Kedziera, D.; Kaczmarek-Kedziera, A. Interaction-Induced Electric Properties and Cooperative Effects in Model Systems. *Phys. Chem. Chem. Phys.* **2010**, *12*, 852–62.
- [83] Zawada, A.; Kaczmarek-Kēdziera, A.; Bartkowiak, W. On the Potential Application of DFT Methods in Predicting the Interaction-Induced Electric Properties of Molecular Complexes. Molecular H-bonded Chains as a Case of Study. *J. Mol. Model.* **2011**, *18*, 3073–3086.
- [84] Zawada, A.; Bartkowiak, W. Many-Body Interactions and the Electric Response of Hydrogen-Bonded Molecular Chains. *Comput. Theor. Chem.* **2011**, *967*, 120–128.
- [85] Leverentz, H. R.; Maerzke, K. A.; Keasler, S. J.; Siepmann, J. I.; Truhlar, D. G. Electrostatically Embedded Many-Body Method for Dipole Moments, Partial Atomic Charges, and Charge Transfer. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7669–7678.
- [86] Mach, T. J.; Crawford, T. D. Computing Optical Rotation via an N-Body Approach. *Theor. Chem. Acc.* **2014**, *133*, 1–9.
- [87] Rosenfeld, L. Quantenmechanische Theorie der Naturlichen Optischen Aktivitaet von Fluessigkeiten und Gasen. *Z. Phys.* **1929**, *52*, 161–174.
- [88] Wolinski, K.; Hilton, J. F.; Pulay, P. Efficient Implementation of the Gauge-Independent Atomic Orbital Method for NMR Chemical Shift Calculations. *J. Am. Chem. Soc.* **1990**, *112*, 8251–60.

- [89] Langhoff, P. W.; Epstein, S. T.; Karplus, M. Aspects of Time-Dependent Perturbation Theory. *Rev. Mod. Phys.* **1972**, *44*, 602–644.
- [90] Helgaker, T.; Coriani, S.; Jørgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. Recent Advances in Wave Function-Based Methods of Molecular -Property Calculations. *Chem. Rev.* **2012**, *112*, 543–631.
- [91] Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.
- [92] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.
- [93] Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [94] Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- [95] Woon, D. E.; Dunning, T. H. J. Gaussian Basis Sets for Use in Correlated Molecular Calculations. IV. Calculation of Static Electrical Response Properties. *J. Chem. Phys.* **1994**, *100*, 2975–2988.
- [96] Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- [97] Frisch, M. J. et al. Gaussian 16 Revision A.03. 2016.

- [98] Parrish, R. M. et al. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185–3197.
- [99] Mach, T. J.; Crawford, T. D. Basis Set Dependence of Coupled Cluster Optical Rotation Computations. *J. Phys. Chem. A* **2011**, *115*, 10045–10051.
- [100] Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation Energies in Density Functional Theory: An Evaluation and a Diagnostic Test. *J. Chem. Phys.* **2008**, *128*, 044118.
- [101] Fuentes, R.; Pearce, K.; Du, Y.; Rakotondrafara, A.; Valenciano, A. L.; Cassera, M. B.; Rasamison, V. E.; Crawford, T. D.; Kingston, D. G. I. Phloroglucinols from the Roots of Garcinia dauphinensis and Their Antiproliferative and Antiplasmodial Activities. *J. Nat. Prod.* **2019**, *82*, 431–439.
- [102] Tam, M. C.; Russ, N. J.; Crawford, T. D. Coupled Cluster Calculations of Optical Rotatory Dispersion of (S)-methyloxirane. *J. Chem. Phys.* **2004**, *121*, 3550–7.
- [103] Haghighatlari, M.; Hachmann, J. Advances of Machine Learning in Molecular Modeling and Simulation. *Curr. Opin. Chem. Eng.* **2019**, *23*, 51–57.
- [104] Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep Learning for Molecular Design - A Review of the State of the Art. *Mol. Syst. Des. Eng.* **2019**, *4*, 828–849.
- [105] Saucedo, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K. R.; Tkatchenko, A. Molecular Force Fields With Gradient-Domain Machine Learning: Construction and Application to Dynamics of Small Molecules With Coupled Cluster Forces. *J. Chem. Phys.* **2019**, *150*, 114102.

- [106] Galvelis, R.; Doerr, S.; Damas, J. M.; Harvey, M. J.; De Fabritiis, G. A Scalable Molecular Force Field Parameterization Method Based on Density Functional Theory and Quantum-Level Machine Learning. *J. Chem. Inf. Model.* **2019**, *59*, 3485–3493.
- [107] Coe, J. P. Machine Learning Configuration Interaction. *J. Chem. Theory Comput.* **2018**, *14*, 5739–5749.
- [108] Townsend, J.; Vogiatzis, K. D. Data-Driven Acceleration of the Coupled-Cluster Singles and Doubles Iterative Solver. *J. Phys. Chem. Lett.* **2019**, *10*, 4129–4135.
- [109] Welborn, M.; Cheng, L.; Miller, T. F. Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis. *J. Chem. Theory Comput.* **2018**, *14*, 4772–4779.
- [110] Cheng, L.; Welborn, M.; Christensen, A. S.; Miller, T. F. A Universal Density Matrix Functional From Molecular Orbital-Based Machine Learning: Transferability Across Organic Molecules. *J. Chem. Phys.* **2019**, *150*, 131103.
- [111] Yuan, S.; Jiao, Z.; Quddus, N.; Kwon, J. S. I.; Mashuga, C. V. Developing Quantitative Structure-Property Relationship Models to Predict the Upper Flammability Limit Using Machine Learning. *Ind. Eng. Chem. Res.* **2019**, *58*, 3531–3537.
- [112] Wilkins, D. M.; Grisafi, A.; Yang, Y.; Lao, K. U.; DiStasio, R. A.; Ceriotti, M. Accurate Molecular Polarizabilities With Coupled Cluster Theory and Machine Learning. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 3401–3406.
- [113] Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy With a General-Purpose Neural Network Potential Through Transfer Learning. *Nat. Commun.* **2019**, *10*, 2903.

- [114] Coester, F.; Kümmel, H. Short-Range Correlations in Nuclear Wave Functions. *Nucl. Phys.* **1960**, *17*, 477–485.
- [115] Bartlett, R. J. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359–401.
- [116] Chmiela, S.; Sauceda, H. E.; Müller, K. R.; Tkatchenko, A. Towards Exact Molecular Dynamics Simulations With Machine-Learned Force Fields. *Nat. Commun.* **2018**, *9*, 3887.
- [117] Schmitz, G.; Christiansen, O. Gaussian Process Regression to Accelerate Geometry Optimizations Relying on Numerical Differentiation. *J. Chem. Phys.* **2018**, *148*, 241704.
- [118] Abbott, A. S.; Turney, J. M.; Zhang, B.; Smith, D. G.; Altarawy, D.; Schaefer, H. F. PES-Learn: An Open-Source Software Package for the Automated Generation of Machine Learning Models of Molecular Potential Energy Surfaces. *J. Chem. Theory Comput.* **2019**, *15*, 4386–4398.
- [119] Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K. R.; Maurer, R. J. Unifying Machine Learning and Quantum Chemistry – A Deep Neural Network for Molecular Wavefunctions. *Nat. Commun.* **2019**, *10*, 5024.
- [120] Christensen, A. S.; Faber, F. A.; Von Lilienfeld, O. A. Operators in Quantum Machine Learning: Response Properties in Chemical Space. *J. Chem. Phys.* **2019**, *150*, 64105.
- [121] Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K. R.; Burke, K. Finding Density Functionals With Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 253002.
- [122] Margraf, J. T.; Reuter, K. Making the Coupled Cluster Correlation Energy Machine-Learnable. *J. Phys. Chem. A* **2018**, *122*, 6343–6348.

- [123] Huo, H.; Rupp, M. Unified Representation of Molecules and Crystals for Machine Learning. 2018; arXiv:1704.06439v3.
- [124] Chen, Y.; Zhang, L.; Wang, H.; E, W. Ground State Energy Functional With Hartree-Fock Efficiency and Chemical Accuracy. 2020; arXiv:2005.00169.
- [125] Murphy, K. P. *Machine Learning: A Probabilistic Perspective*; The MIT Press, 2012.
- [126] Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*; The MIT Press, 2006.
- [127] Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies With Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
- [128] Pedregosa, F. et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- [129] Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
- [130] Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- [131] Hättig, C. Optimization of Auxiliary Basis Sets for RI-MP2 and RI-CC2 Calculations: Core-Valence and Quintuple- ζ Basis Sets for H to Ar and QZVPP Basis Sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.
- [132] Weigend, F.; Köhn, A.; Hättig, C. Efficient Use of the Correlation Consistent Basis Sets in Resolution of the Identity MP2 Calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.

- [133] Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New Basis Set Exchange: An Open, Up-To-Date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.* **2019**, *59*, 4814–4820.
- [134] Machine-Learning Quantum Mechanics. github.com/CrawfordGroup/MLQM.
- [135] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- [136] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- [137] Crawford, D.; Peyton, B.; Margraf, J.; Briggs, C.; D'Cunha, R. Machine-Learning Coupled Cluster Properties through a Density Tensor Representation. University Libraries, Virginia Tech. doi:0.7294/A822-4M33.
- [138] Jung, H.; Stocker, S.; Kunkel, C.; Oberhofer, H.; Han, B.; Reuter, K.; Margraf, J. T. Size-Extensive Molecular Machine Learning With Global Representations. *ChemSystemsChem* **2020**, *2*, e1900052.
- [139] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864.
- [140] Smith, D. G. A.; Altarawy, D.; Burns, L. A.; Welborn, M.; Naden, L. N.; Ward, L.; Ellis, S.; Crawford, T. D. The MolSSI QCArchive Project: An Open-Source Platform to Compute, Organize, and Share Quantum Chemistry Data. 2020; ChemRxiv:22566e14d96e43f7611a.
- [141] Goddard, W. A. Theoretical chemistry comes alive: Full partner with experiment. *Science* **1985**, *227*, 917–923.

- [142] Howard, J. C.; V, S. S. S.; Ansari, I. M.; Mach, T. J.; Baranowska-Łączkowska, A.; Crawford, T. D. Performance of Property-Optimized Basis Sets for Optical Rotation with Coupled Cluster Theory. *J. Phys. Chem. A* **2018**, *122*, 5962–5969.
- [143] D'Cunha, R.; Crawford, T. D. PNO++: Perturbed Pair Natural Orbitals for Coupled Cluster Linear Response Theory. *J. Chem. Theory Comput.* **2021**, *17*, 290–301.
- [144] Maiuri, M.; Garavelli, M.; Cerullo, G. Ultrafast Spectroscopy: State of the Art and Open Challenges. *J. Am. Chem. Soc.* **2020**, *142*, 3–15.
- [145] Lewenstein, M.; Balcou, P.; Ivanov, M. Y.; L'Huillier, A.; Corkum, P. B. Theory of high-harmonic generation by low-frequency laser fields. *Phys. Rev. A* **1994**, *49*, 2117–2132.
- [146] Gorlach, A.; Neufeld, O.; Rivera, N.; Cohen, O.; Kaminer, I. The quantum-optical nature of high harmonic generation. *Nat. Commun.* **2020**, *11*.
- [147] Hoodbhoy, P.; Nehele, J. W. Time-dependent coupled-cluster approximation to nuclear dynamics. I. Application to a solvable model. *Phys. Rev. C* **1978**, *18*, 2380 – 2394.
- [148] Schönhammer, K.; Gunnarsson, O. Time-dependent approach to the calculation of spectral functions. *Phys. Rev. B* **1978**, *18*, 6606 – 6614.
- [149] Huber, C.; Klamroth, T. Explicitly time-dependent coupled cluster singles doubles calculations of laser-driven many-electron dynamics. *J. Chem. Phys.* **2011**, *134*, 54113.
- [150] Kvaal, S. Ab initio quantum dynamics using coupled-cluster. *J. Chem. Phys.* **2012**, *136*, 194109.
- [151] Nascimento, D. R.; Deprince, A. E. A general time-domain formulation of equation-of-motion coupled-cluster theory for linear spectroscopy. *J. Chem. Phys.* **2019**, *151*, 160901.

- [152] Pedersen, T. B.; Kvaal, S. Symplectic integration and physical interpretation of time-dependent coupled-cluster theory. *J. Chem. Phys.* **2019**, *150*, 144106.
- [153] Park, Y. C.; Perera, A.; Bartlett, R. J. Equation of motion coupled-cluster for core excitation spectra: Two complementary approaches. *J. Chem. Phys.* **2019**, *151*.
- [154] Kristiansen, H. E.; Schøyen, Ø. S.; Kvaal, S.; Pedersen, T. B. Numerical stability of time-dependent coupled-cluster methods for many-electron dynamics in intense laser pulses. *J. Chem. Phys.* **2020**, *152*.
- [155] Nascimento, D. R.; DePrince, A. E. Linear Absorption Spectra from Explicitly Time-Dependent Equation-of-Motion Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2016**, *12*, 5834–5840.
- [156] Nascimento, D. R.; DePrince, A. E. Simulation of Near-Edge X-ray Absorption Fine Structure with Time-Dependent Equation-of-Motion Coupled-Cluster Theory. *J. Phys. Chem. Lett.* **2017**, *8*, 2951–2957.
- [157] Park, Y. C.; Perera, A.; Bartlett, R. J. Equation of motion coupled-cluster study of core excitation spectra II: Beyond the dipole approximation. *J. Chem. Phys.* **2021**, *155*.
- [158] Sato, T.; Pathak, H.; Orimo, Y.; Ishikawa, K. L. Communication: Time-dependent optimized coupled-cluster method for multielectron dynamics. *J. Chem. Phys.* **2018**, *148*.
- [159] Yabana, K.; Bertsch, G. Time-dependent local-density approximation in real time. *Phys. Rev. B - Condens. Matter Mater. Phys.* **1996**, *54*, 4484–4487.
- [160] Yabana, K.; Bertsch, G. F. Optical response of small carbon clusters. *Zeitschrift fur Phys. D-Atoms Mol. Clust.* **1997**, *42*, 219–225.

- [161] Yabana, K.; Bertsch, G. F. Time-dependent local-density approximation in real time: Application to conjugated molecules. *Int. J. Quantum Chem.* **1999**, *75*, 55–66.
- [162] Bertsch, G. F.; Iwata, J. I.; Rubio, A.; Yabana, K. Real-space, real-time method for the dielectric function. *Phys. Rev. B - Condens. Matter Mater. Phys.* **2000**, *62*, 7998–8002.
- [163] Lopata, K.; Govind, N. Modeling fast electron dynamics with real-time time-dependent density functional theory: Application to small molecules and chromophores. *J. Chem. Theory Comput.* **2011**, *7*, 1344–1355.
- [164] Castro, A.; Rubio, A.; Gross, E. K. Enhancing and controlling single-atom high-harmonic generation spectra: a time-dependent density-functional scheme. *Eur. Phys. J. B* **2015**, *88*, 191.
- [165] Tussupbayev, S.; Govind, N.; Lopata, K.; Cramer, C. J. Comparison of real-time and linear-response time-dependent density functional theories for molecular chromophores ranging from sparse to high densities of states. *J. Chem. Theory Comput.* **2015**, *11*, 1102–1109.
- [166] Goings, J. J.; Li, X. An atomic orbital based real-time time-dependent density functional theory for computing electronic circular dichroism band spectra. *J. Chem. Phys.* **2016**, *144*, 234102.
- [167] Bruner, A.; Lamaster, D.; Lopata, K. Accelerated Broadband Spectra Using Transition Dipole Decomposition and Pad{\backslash'e} Approximants. *J. Chem. Theory Comput.* **2016**, *12*, 3741–3750.
- [168] Sun, S.; Beck, R. A.; Williams-Young, D.; Li, X. Simulating Magnetic Circular Dichroism Spectra with Real-Time Time-Dependent Density Functional Theory in Gauge including Atomic Orbitals. *J. Chem. Theory Comput.* **2019**, *15*, 6824–6831.

- [169] Ding, F.; Van Kuiken, B. E.; Eichinger, B. E.; Li, X. An efficient method for calculating dynamical hyperpolarizabilities using real-time time-dependent density functional theory. *J. Chem. Phys.* **2013**, *138*.
- [170] Repisky, M.; Konecny, L.; Kadek, M.; Komorovsky, S.; Malkin, O. L.; Malkin, V. G.; Ruud, K. Excitation Energies from Real-Time Propagation of the Four-Component Dirac-Kohn-Sham Equation. *J. Chem. Theory Comput.* **2015**, *11*, 980–991.
- [171] Kadek, M.; Konecny, L.; Gao, B.; Repisky, M.; Ruud, K. X-ray absorption resonances near L_{2,3}-edges from real-time propagation of the Dirac-Kohn-Sham density matrix. *Phys. Chem. Chem. Phys.* **2015**, *17*, 22566–22570.
- [172] Fuks, J. I.; Elliott, P.; Rubio, A.; Maitra, N. T. Dynamics of charge-transfer processes with time-dependent density functional theory. *J. Phys. Chem. Lett.* **2013**, *4*, 735–739.
- [173] Fuks, J. I.; Luo, K.; Sandoval, E. D.; Maitra, N. T. Time-resolved spectroscopy in time-dependent density functional theory: An exact condition. *Phys. Rev. Lett.* **2015**, *114*, 1–6.
- [174] Wolinski, K.; Hinton, J. F.; Pulay, P. Efficient Implementation of the Gauge-Independent Atomic Orbital Method for NMR Chemical Shift Calculations. *J. Am. Chem. Soc.* **1990**, *112*, 8251–8260.
- [175] Sadlej, A. J. Molecular Electric Polarizabilities. Electric-Field-Variant (EFV) Gaussian Basis Set for Polarizability Calculations. *Chem. Phys. Lett.* **1977**, *47*, 50–54.
- [176] Roos, B. O.; Sadlej, A. J. Polarized basis sets for accurate predictions of molecular electric properties. Electric moments of the LiH molecule. *Chem. Phys.* **1985**, *94*, 43–53.

- [177] Sadlej, A. J. Medium-size polarized basis sets for high-level-correlated calculations of molecular electric properties - II. Second-row atoms: Si through Cl. *Theor. Chim. Acta* **1991**, *79*, 123–140.
- [178] Benkova, Z.; Sadlej, A. J.; Oakes, R. E.; Bell, S. E. Reduced-size polarized basis sets for calculations of molecular electric properties. I. the basis set generation. *J. Comput. Chem.* **2005**, *26*, 145–153.
- [179] Baranowska-Łc aczkowska, A.; Łc aczkowski, K. Z. The ORP basis set designed for optical rotation calculations. *J. Comput. Chem.* **2013**, *34*, 2006–2013.
- [180] Aharon, T.; Caricato, M. Compact Basis Sets for Optical Rotation Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 4408–4415.
- [181] Wang, Z.; Peyton, B.; Crawford, T. Accelerating Real-Time Coupled Cluster Methods with Numerical Integrators and Single-Precision Arithmetic. *[In Preparation]* **2022**,
- [182] Kodrycka, M.; Crawford, T. Local Correlation for Quadratic Coupled Cluster Response Properties. *[In Preparation]* **2022**,
- [183] Hampel, C.; Werner, H. J. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.* **1996**, *104*, 6286–6297.
- [184] Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151–154.
- [185] Sæbø, S.; Pulay, P. Local configuration interaction: An efficient approach for larger molecules. *Chem. Phys. Lett.* **1985**, *113*, 13–18.
- [186] Sæbø, S.; Pulay, P. Fourth-order Møller-Plessett perturbation theory in the local correlation treatment. I. Method. *J. Chem. Phys.* **1986**, *86*, 914–922.

- [187] Sæbø, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213–236.
- [188] Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *J. Chem. Phys.* **2009**, *130*.
- [189] Neese, F.; Hansen, A.; Liakos, D. G. Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *J. Chem. Phys.* **2009**, *131*, 064103.
- [190] Butcher, J. C. A history of Runge-Kutta methods. *Appl. Numer. Math.* **1996**, *20*, 247–260.
- [191] Rosenfeld, L. Quantenmechanische Theorie der natürlichen optischen Aktivitäten von Flüssigkeiten und Gasen. *Z. Phys.* **1929**, *52*, 161–174.
- [192] Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* **1989**, *90*, 4916–4926.
- [193] Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272.
- [194] Crawford, T.; Peyton, B. G.; Wang, Z. PyCC. <https://github.com/CrawfordGroup/pycc>.
- [195] Harris, C. R. et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362.
- [196] Smith, D. G. et al. Psi4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **2020**, *152*.

- [197] G. A. Smith, D.; Gray, J. opt__einsum - A Python package for optimizing contraction order for einsum-like expressions. *J. Open Source Softw.* **2018**, *3*, 753.

Appendices

Appendix A

Supporting Information for Basis Set Superposition Errors in the Many-Body Expansion of Molecular Properties

A.1 Additional Data

A.1.1 Methyloxirane

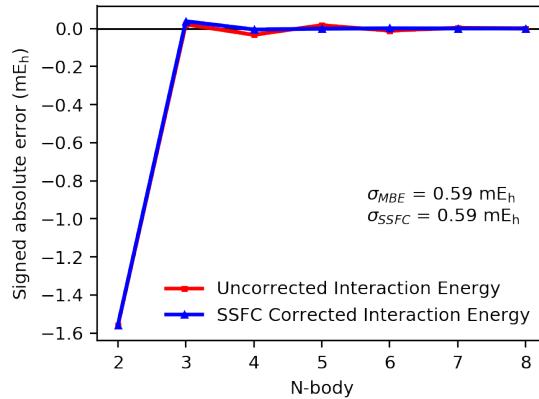


Figure A.1: MBE and SSFC of interaction energy for (*S*)-methyloxirane in a seven-water solvent shell. Computed with B3LYP/aDZ.

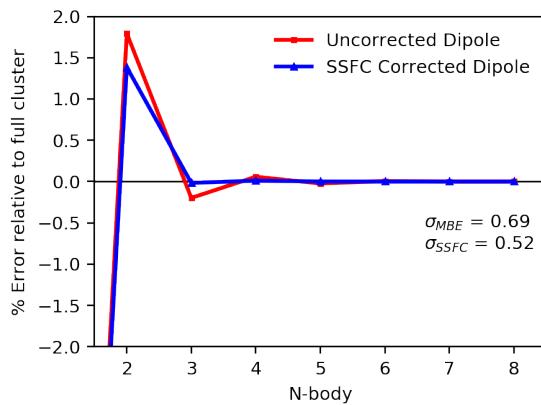


Figure A.2: MBE and SSFC of dipole moment for (*S*)-methyloxirane in a seven-water solvent shell. Computed with B3LYP/aDZ.

**APPENDIX A. SUPPORTING INFORMATION FOR BASIS SET SUPERPOSITION ERRORS IN THE
MANY-BODY EXPANSION OF MOLECULAR PROPERTIES**

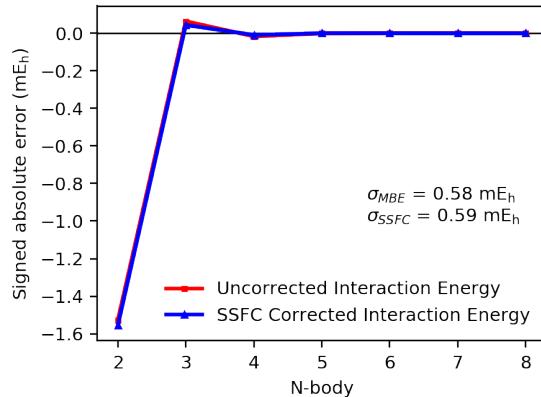


Figure A.3: MBE and SSFC of interaction energy for (*S*)-methyloxirane in a seven-water solvent shell. Computed with B3LYP/aTZ.

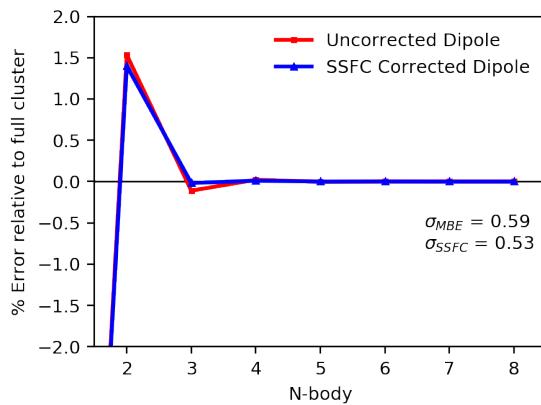


Figure A.4: MBE and SSFC of dipole moment for (*S*)-methyloxirane in a seven-water solvent shell. Computed with B3LYP/aTZ.

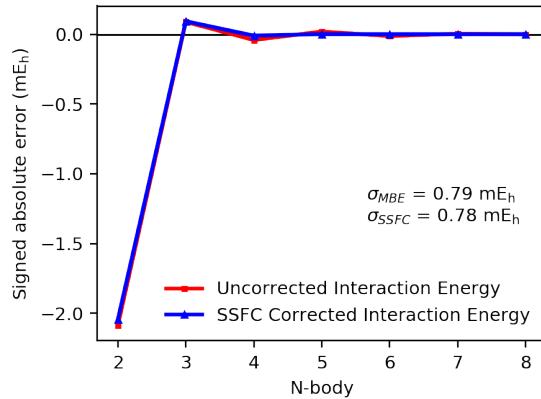


Figure A.5: MBE and SSFC of interaction energy for (*S*)-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.

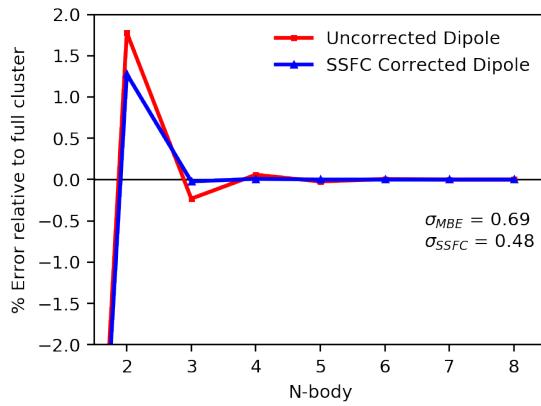


Figure A.6: MBE and SSFC of dipole moment for (*S*)-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.

**APPENDIX A. SUPPORTING INFORMATION FOR BASIS SET SUPERPOSITION ERRORS IN THE
MANY-BODY EXPANSION OF MOLECULAR PROPERTIES**

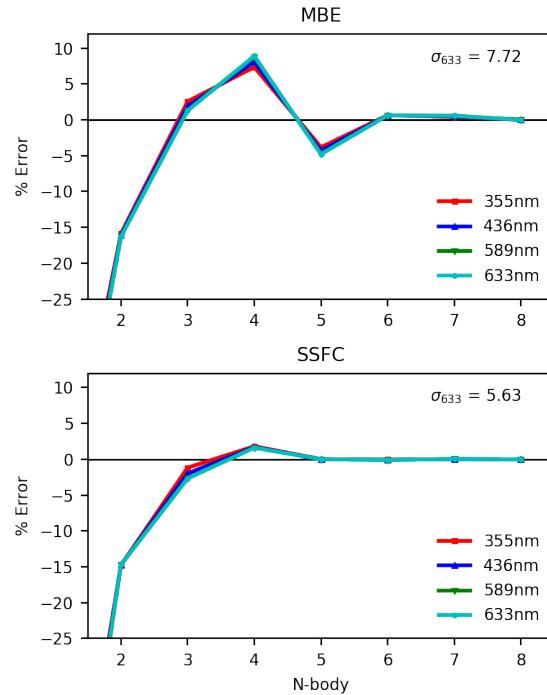


Figure A.7: MBE and SSFC of specific rotation for *(S)*-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.

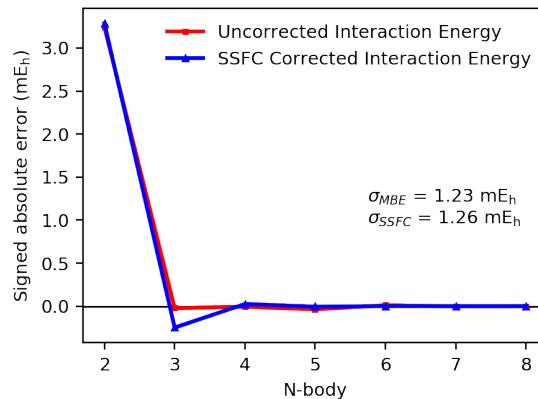


Figure A.8: MBE and SSFC of interaction energy for snapshot #2 of *(S)*-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.

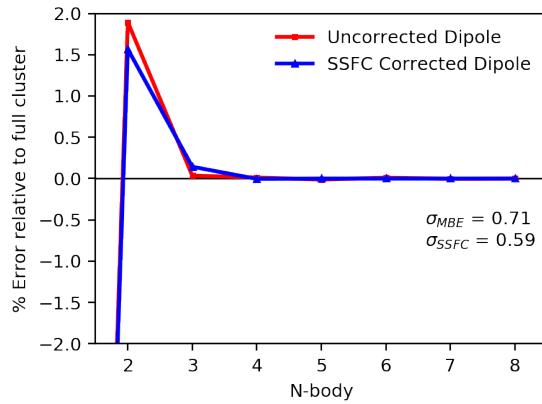


Figure A.9: MBE and SSFC of dipole moment for snapshot #2 of (*S*)-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.

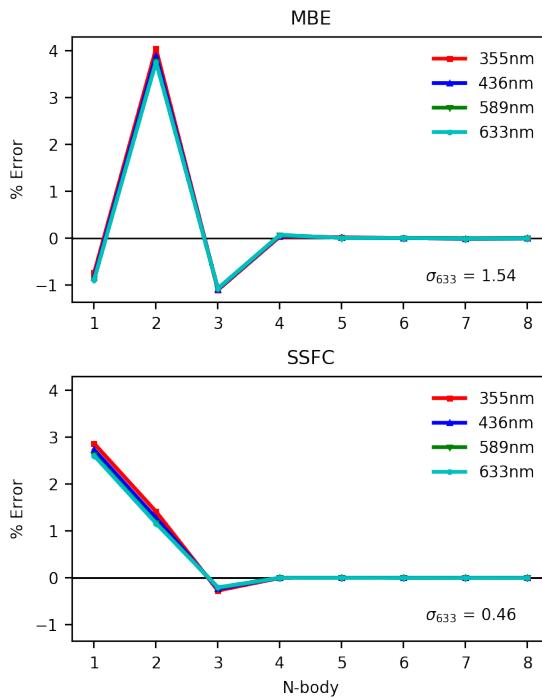


Figure A.10: (*S*)-methyloxirane

Figure A.11: MBE and SSFC of dynamic polarizabilities for snapshot #2 of (*S*)-methyloxirane in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.

APPENDIX A. SUPPORTING INFORMATION FOR BASIS SET SUPERPOSITION ERRORS IN THE MANY-BODY EXPANSION OF MOLECULAR PROPERTIES

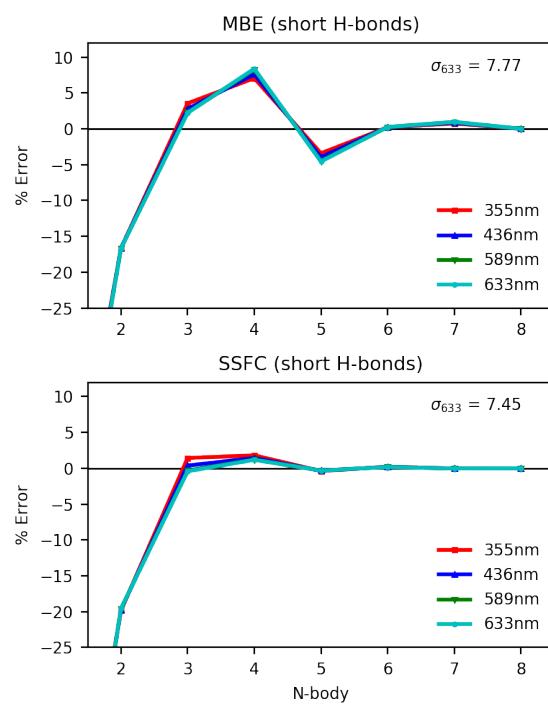


Figure A.12: MBE and SSFC of specific rotation of (*S*)-methyloxirane in a seven-water solvent shell and two manually reduced H-bonds. Computed with CAM-B3LYP/aDZ.

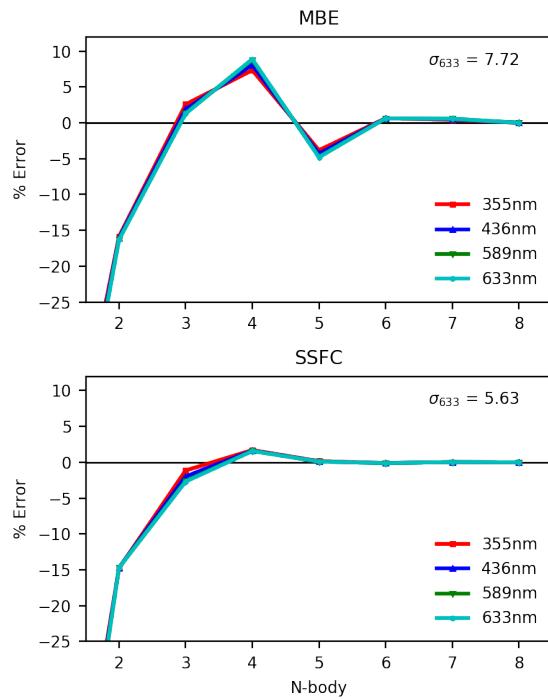


Figure A.13: MBE and SSFC of specific rotation of (*S*)-methyloxirane in a seven-water solvent shell using “tight” convergence criteria and “fine” grids. Computed with CAM-B3LYP/aDZ.

APPENDIX A. SUPPORTING INFORMATION FOR BASIS SET SUPERPOSITION ERRORS IN THE
MANY-BODY EXPANSION OF MOLECULAR PROPERTIES

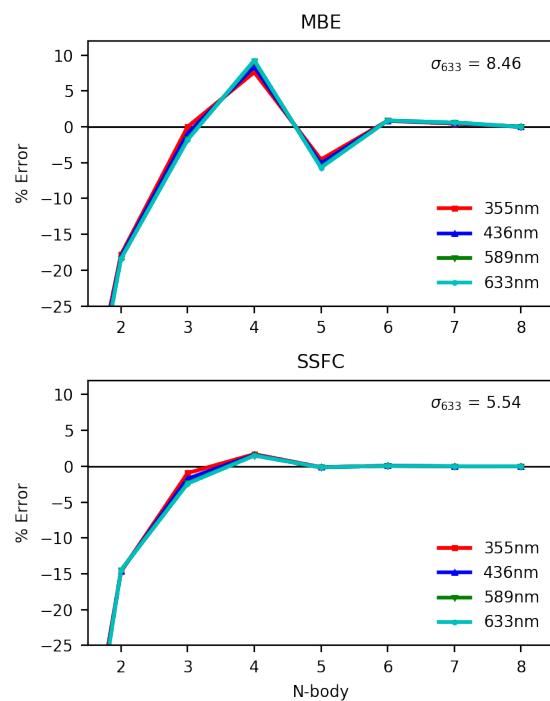


Figure A.14: MBE and SSFC of specific rotation of *(S)*-methyloxirane in a seven-water solvent shell and unpruned “fine” grids. Computed with CAM-B3LYP/aDZ.

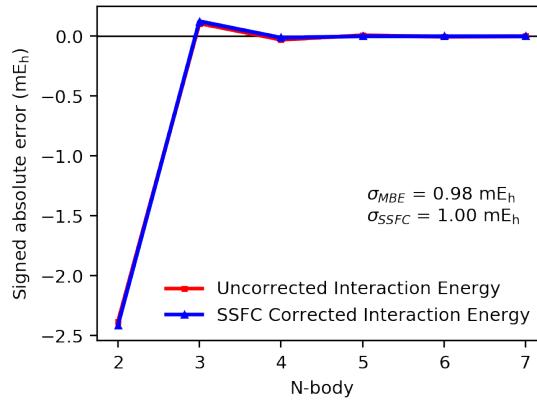


Figure A.15: MBE and SSFC of interaction energy for (*S*)-methylthiirane in a six-water solvent shell. Computed with B3LYP/aDZ.

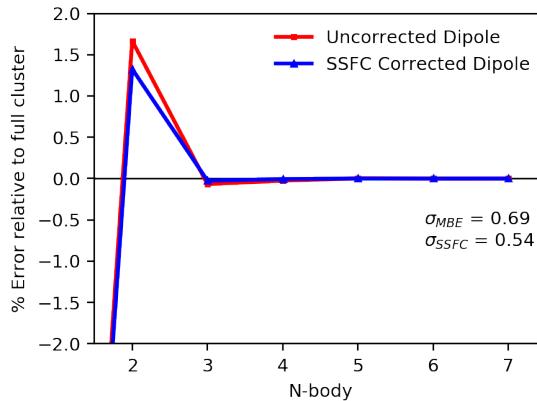


Figure A.16: MBE and SSFC of dipole moment for (*S*)-methylthiirane in a six-water solvent shell. Computed with B3LYP/aDZ.

A.1.2 Methylthiirane

**APPENDIX A. SUPPORTING INFORMATION FOR BASIS SET SUPERPOSITION ERRORS IN THE
MANY-BODY EXPANSION OF MOLECULAR PROPERTIES**

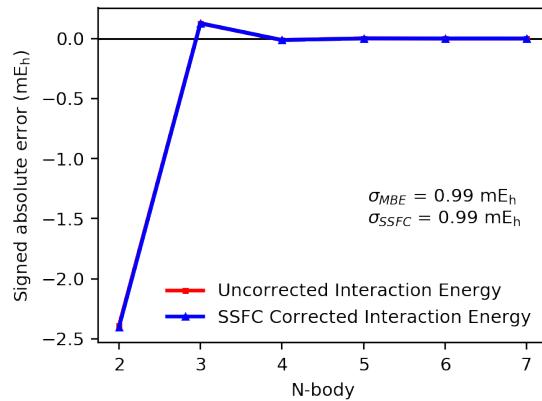


Figure A.17: MBE and SSFC of interaction energy for (*S*)-methylthiirane in a six-water solvent shell. Computed with B3LYP/aTZ.

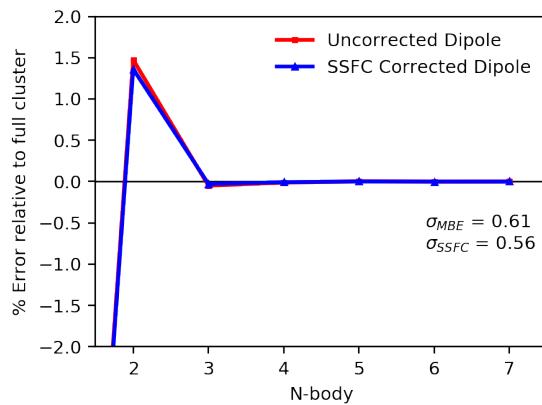


Figure A.18: MBE and SSFC of dipole moment for (*S*)-methylthiirane in a six-water solvent shell. Computed with B3LYP/aTZ.

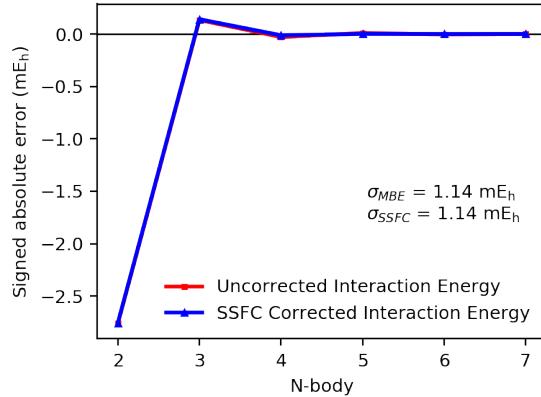


Figure A.19: MBE and SSFC of interaction energy for (*S*)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.

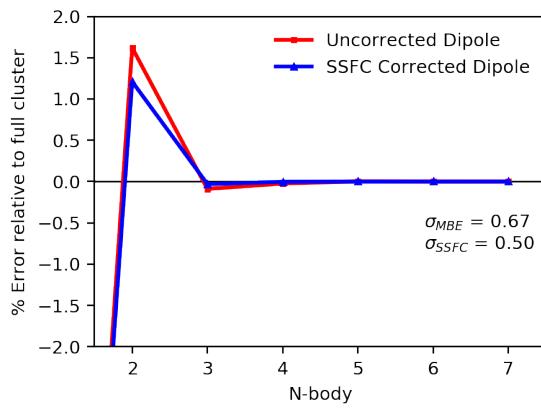


Figure A.20: MBE and SSFC of dipole moment for (*S*)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.

**APPENDIX A. SUPPORTING INFORMATION FOR BASIS SET SUPERPOSITION ERRORS IN THE
MANY-BODY EXPANSION OF MOLECULAR PROPERTIES**

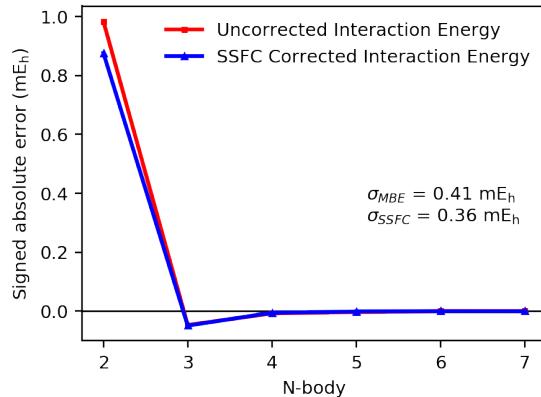


Figure A.21: MBE and SSFC of interaction energy for snapshot #2 of (*S*)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.

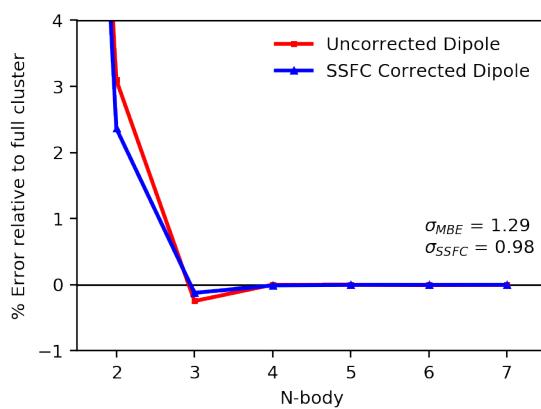


Figure A.22: MBE and SSFC of dipole moment for snapshot #2 of (*S*)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.

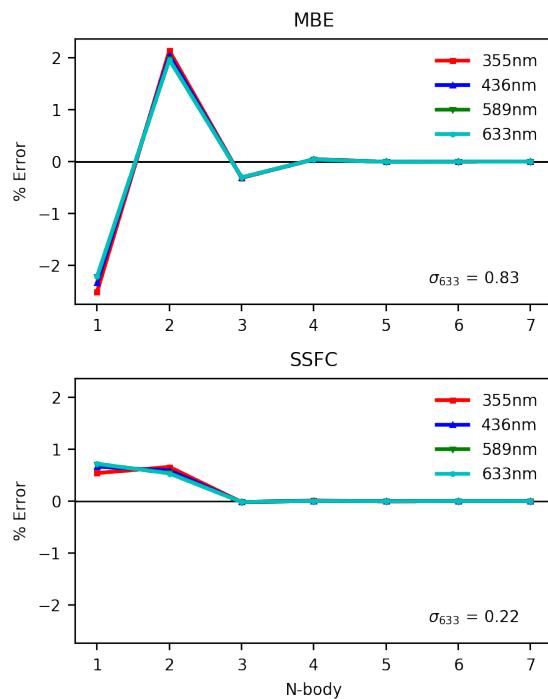


Figure A.23: MBE and SSFC of dynamic polarizabilities for snapshot #2 of (*S*)-methylthiirane in a six-water solvent shell. Computed with CAM-B3LYP/aDZ.

**APPENDIX A. SUPPORTING INFORMATION FOR BASIS SET SUPERPOSITION ERRORS IN THE
MANY-BODY EXPANSION OF MOLECULAR PROPERTIES**

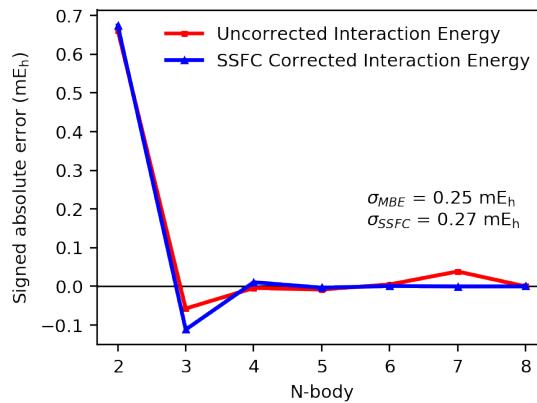


Figure A.24: MBE and SSFC of interaction energy for (*M*)-dimethylallene in a seven-water solvent shell. Computed with B3LYP/aDZ.

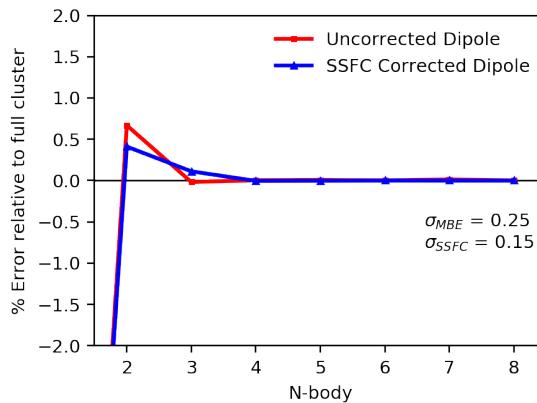


Figure A.25: MBE and SSFC of dipole moment for (*M*)-dimethylallene in a seven-water solvent shell. Computed with B3LYP/aDZ.

A.1.3 Dimethylallene

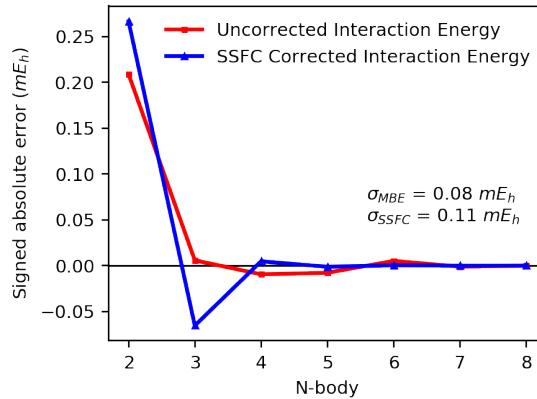


Figure A.26: MBE and SSFC of interaction energy for (*M*)-dimethylallene in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.

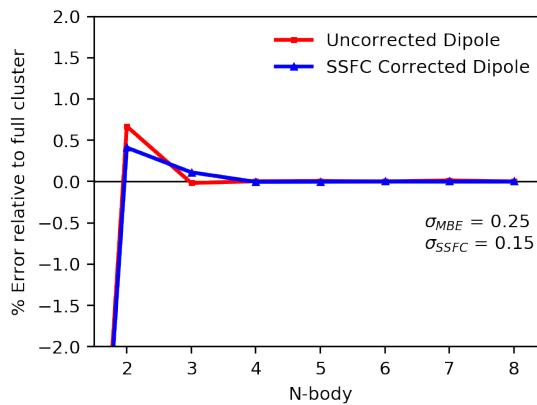


Figure A.27: MBE and SSFC of dipole moment for (*M*)-dimethylallene in a seven-water solvent shell. Computed with CAM-B3LYP/aDZ.

A.2 Atomic Coordinates

Table A.1: Atomic coordinates of (*S*)-methyloxirane in a seven-water solvent shell (Ångstroms)

Symbol	X	Y	Z
C	14.600000	14.530000	15.130000
O	14.600000	14.530000	16.530000
C	15.860000	14.530000	15.850000
C	14.520000	15.710000	14.300000
H	13.580000	15.710000	13.750000
H	14.580000	16.600000	14.920000
H	15.350000	15.710000	13.590000
H	14.090000	13.640000	14.770000
H	16.430000	13.640000	15.590000
H	16.430000	15.420000	15.590000
O	11.050000	14.880000	13.420000
H	11.070000	14.670000	14.350000
H	11.780000	15.490000	13.290000
O	18.580000	15.280000	15.690000
H	18.510000	14.460000	15.220000
H	19.390000	15.200000	16.200000
O	17.480000	16.470000	17.830000
H	17.700000	15.930000	17.070000
H	17.110000	15.850000	18.460000
O	11.290000	17.480000	15.180000
H	11.990000	16.870000	14.960000
H	11.360000	17.610000	16.130000
O	16.860000	14.160000	11.360000
H	16.530000	13.440000	10.810000
H	16.400000	14.040000	12.200000
O	12.960000	16.070000	16.940000
H	13.330000	15.780000	17.770000
H	13.710000	16.190000	16.360000
O	16.770000	14.240000	19.190000
H	17.490000	13.620000	19.050000
H	16.010000	13.820000	18.780000

Table A.2: Atomic coordinates of (*S*)-methyloxirane in a seven-water solvent shell (Snapshot #2) (Ångstroms)

Symbol	X	Y	Z
C	26.730001	26.240002	27.180002
O	27.889999	26.100002	26.400002
C	26.710003	26.310001	25.650002
C	26.830002	27.590000	27.900002
H	27.570002	27.500002	28.690002
H	27.030001	28.430002	27.240002
H	25.870001	27.820000	28.340000
H	26.700001	25.340002	27.790001
H	26.800001	25.380001	25.080002
H	26.900002	27.290001	25.220001
O	29.530001	24.260002	27.000000
H	29.930000	23.990002	26.120003
H	28.860001	24.990002	26.860001
O	29.590000	25.900002	24.580002
H	28.880003	25.980000	25.270002
H	29.820002	24.930000	24.450001
O	24.010000	25.500002	25.790003
H	23.310001	26.220001	25.760002
H	23.790001	24.799999	25.120001
O	27.560001	24.490002	30.270002
H	26.580002	24.660000	30.310001
H	27.790001	24.030001	29.410002
O	27.170002	22.970001	25.570002
H	27.790001	23.140001	26.350002
H	27.690001	23.050001	24.720001
O	24.600000	28.290003	23.990002
H	24.590000	29.290001	24.100000
H	23.710001	27.920000	24.260002
O	23.850002	23.740002	27.780001
H	24.180002	24.550001	27.299999
H	23.030003	23.390003	27.320002

**APPENDIX A. SUPPORTING INFORMATION FOR BASIS SET SUPERPOSITION ERRORS IN THE
MANY-BODY EXPANSION OF MOLECULAR PROPERTIES**

Table A.3: Atomic coordinates of (*S*)-methyloxirane in a 7-water solvent shell and two manually reduced H-bonds (Ångstroms)

Atomic Symbol	X	Y	Z
C	14.600000	14.530000	15.130000
O	14.600000	14.530000	16.530000
C	15.860000	14.530000	15.850000
C	14.520000	15.710000	14.300000
H	13.580000	15.710000	13.750000
H	14.580000	16.600000	14.920000
H	15.350000	15.710000	13.590000
H	14.090000	13.640000	14.770000
H	16.430000	13.640000	15.590000
H	16.430000	15.420000	15.590000
O	11.050000	14.880000	13.420000
H	11.070000	14.670000	14.350000
H	11.780000	15.490000	13.290000
O	18.601491	15.356686	16.138508
H	18.531491	14.536686	15.668508
H	19.411491	15.276686	16.648508
O	17.480000	16.470000	17.830000
H	17.700000	15.930000	17.070000
H	17.110000	15.850000	18.460000
O	11.805218	17.248086	15.788585
H	12.505218	16.638086	15.568585
H	11.875218	17.378086	16.738585
O	16.860000	14.160000	11.360000
H	16.530000	13.440000	10.810000
H	16.400000	14.040000	12.200000
O	12.960000	16.070000	16.940000
H	13.330000	15.780000	17.770000
H	13.710000	16.190000	16.360000
O	16.770000	14.240000	19.190000
H	17.490000	13.620000	19.050000
H	16.010000	13.820000	18.780000

Table A.4: Atomic coordinates of (*S*)-methyloxirane in a 13-water solvent shell (Ångstroms)

Atomic Symbol	X	Y	Z
C	14.600000	14.530000	15.130000
O	14.600000	14.530000	16.530000
C	15.860000	14.530000	15.850000
C	14.520000	15.710000	14.300000
H	13.580000	15.710000	13.750000
H	14.580000	16.600000	14.920000
H	15.350000	15.710000	13.590000
H	14.090000	13.640000	14.770000
H	16.430000	13.640000	15.590000
H	16.430000	15.420000	15.590000
O	15.600000	19.160000	17.660000
H	16.160000	19.800000	17.220000
H	15.980000	19.070000	18.540000
O	11.050000	14.880000	13.420000
H	11.070000	14.670000	14.350000
H	11.780000	15.490000	13.290000
O	19.150000	17.890000	13.810000
H	18.590000	17.110000	13.850000
H	19.280000	18.130000	14.730000
O	10.790000	14.420000	16.000000
H	10.100000	15.050000	16.210000
H	11.430000	14.510000	16.710000
O	14.750000	19.290000	13.750000
H	14.910000	18.510000	14.270000
H	15.410000	19.920000	14.040000
O	18.580000	15.280000	15.690000
H	18.510000	14.460000	15.220000
H	19.390000	15.200000	16.200000
O	13.370000	14.030000	19.010000
H	13.860000	13.220000	18.870000
H	12.470000	13.750000	19.180000
O	17.480000	16.470000	17.830000
H	17.700000	15.930000	17.070000
H	17.110000	15.850000	18.460000
O	11.290000	17.480000	15.180000
H	11.990000	16.870000	14.960000
H	11.360000	17.610000	16.130000
O	16.860000	14.160000	11.360000
H	16.530000	13.440000	10.810000
H	16.400000	14.040000	12.200000
O	15.050000	12.180000	18.300000
H	15.510000	11.370000	18.530000
H	15.000000	12.160000	17.340000
O	12.960000	16.070000	16.940000
H	13.330000	15.780000	17.770000
H	13.710000	16.190000	16.360000

APPENDIX A. SUPPORTING INFORMATION FOR BASIS SET SUPERPOSITION ERRORS IN THE
MANY-BODY EXPANSION OF MOLECULAR PROPERTIES

O	16.770000	14.240000	19.190000
H	17.490000	13.620000	19.050000
H	16.010000	13.820000	18.780000

Table A.5: Atomic coordinates of (*S*)-methylthiirane in a six-water solvent shell (Ångstroms)

Symbol	X	Y	Z
C	14.840000	14.890000	15.500000
S	13.220000	14.190000	14.950000
C	13.960000	15.880000	14.850000
C	16.110000	14.410000	14.820000
H	16.400000	13.420000	15.190000
H	15.960000	14.340000	13.740000
H	16.930000	15.110000	15.020000
H	14.910000	14.960000	16.590000
H	13.470000	16.630000	15.460000
H	14.210000	16.200000	13.840000
O	11.050000	14.880000	13.420000
H	11.070000	14.670000	14.350000
H	11.780000	15.490000	13.290000
O	18.580000	15.280000	15.690000
H	18.510000	14.460000	15.220000
H	19.390000	15.200000	16.200000
O	17.480000	16.470000	17.830000
H	17.700000	15.930000	17.070000
H	17.110000	15.850000	18.460000
O	11.290000	17.480000	15.180000
H	11.990000	16.870000	14.960000
H	11.360000	17.610000	16.130000
O	16.860000	14.160000	11.360000
H	16.530000	13.440000	10.810000
H	16.400000	14.040000	12.200000
O	16.770000	14.240000	19.190000
H	17.490000	13.620000	19.050000
H	16.010000	13.820000	18.780000

**APPENDIX A. SUPPORTING INFORMATION FOR BASIS SET SUPERPOSITION ERRORS IN THE
MANY-BODY EXPANSION OF MOLECULAR PROPERTIES**

Table A.6: Atomic coordinates of (*S*)-methylthiirane in a six-water solvent shell (Snapshot #2) (Ångstroms)

Atomic Symbol	X	Y	Z
C	26.350002	26.090002	-26.730001
S	28.010002	26.670000	-26.280001
C	26.410000	27.530001	-26.220001
C	25.530003	25.180000	-25.830002
H	24.570000	25.030001	-26.330002
H	26.040001	24.220001	-25.820000
H	25.470001	25.540001	-24.799999
H	26.360001	25.760002	-27.770002
H	26.350002	28.360003	-26.920002
H	25.990002	27.750000	-25.240002
O	25.600002	27.190001	-29.500000
H	24.640003	26.960001	-29.380001
H	26.160002	26.620003	-28.890001
O	23.630001	27.710001	-24.590000
H	22.740002	27.680000	-25.030001
H	24.300001	28.140001	-25.200003
O	24.650002	26.170002	-22.590000
H	24.300001	26.840000	-23.250000
H	25.450001	26.550003	-22.130001
O	24.870003	23.340000	-28.710001
H	24.490002	23.500002	-29.620001
H	25.840000	23.600002	-28.710001
O	28.270000	23.770000	-28.900002
H	28.500002	23.220001	-28.100002
H	28.570002	24.710003	-28.760002
O	29.250002	24.520000	-24.370001
H	30.040001	24.410002	-24.970001
H	28.660002	23.710001	-24.450001

Table A.7: Atomic coordinates of (*M*)-dimethylallene in a seven-water solvent shell (Ångstroms)

Symbol	Atomic		
	X	Y	Z
C	12.620000	11.350000	11.820000
H	12.140000	10.750000	12.590000
H	12.940000	10.700000	11.010000
H	11.960000	12.140000	11.460000
C	13.830000	12.070000	12.410000
H	14.840000	11.920000	12.030000
C	13.560000	12.880000	13.400000
C	13.310000	13.690000	14.390000
H	13.320000	13.360000	15.430000
C	13.000000	15.130000	14.000000
H	13.040000	15.230000	12.920000
H	13.730000	15.800000	14.460000
H	12.000000	15.320000	14.380000
O	11.270000	13.410000	16.900000
H	11.740000	13.410000	17.780000
H	10.790000	12.540000	16.790000
O	14.660000	14.170000	9.530000
H	14.070000	14.230000	10.330000
H	14.510000	13.290000	9.070000
O	10.640000	13.470000	13.400000
H	9.840000	13.240000	13.950000
H	11.470000	13.210000	13.890000
O	16.380000	10.990000	10.790000
H	16.430000	11.800000	11.380000
H	15.520000	11.010000	10.280000
O	12.520000	17.310000	11.280000
H	13.360000	17.360000	10.740000
H	12.290000	16.350000	11.460000
O	14.500000	12.200000	16.330000
H	14.410000	12.100000	17.320000
H	15.460000	12.130000	16.070000
O	17.050000	14.870000	11.040000
H	16.120000	14.620000	10.770000
H	17.310000	14.370000	11.860000

Appendix B

Supporting Information for Machine-Learning Coupled Cluster Properties through a Density Tensor Representation

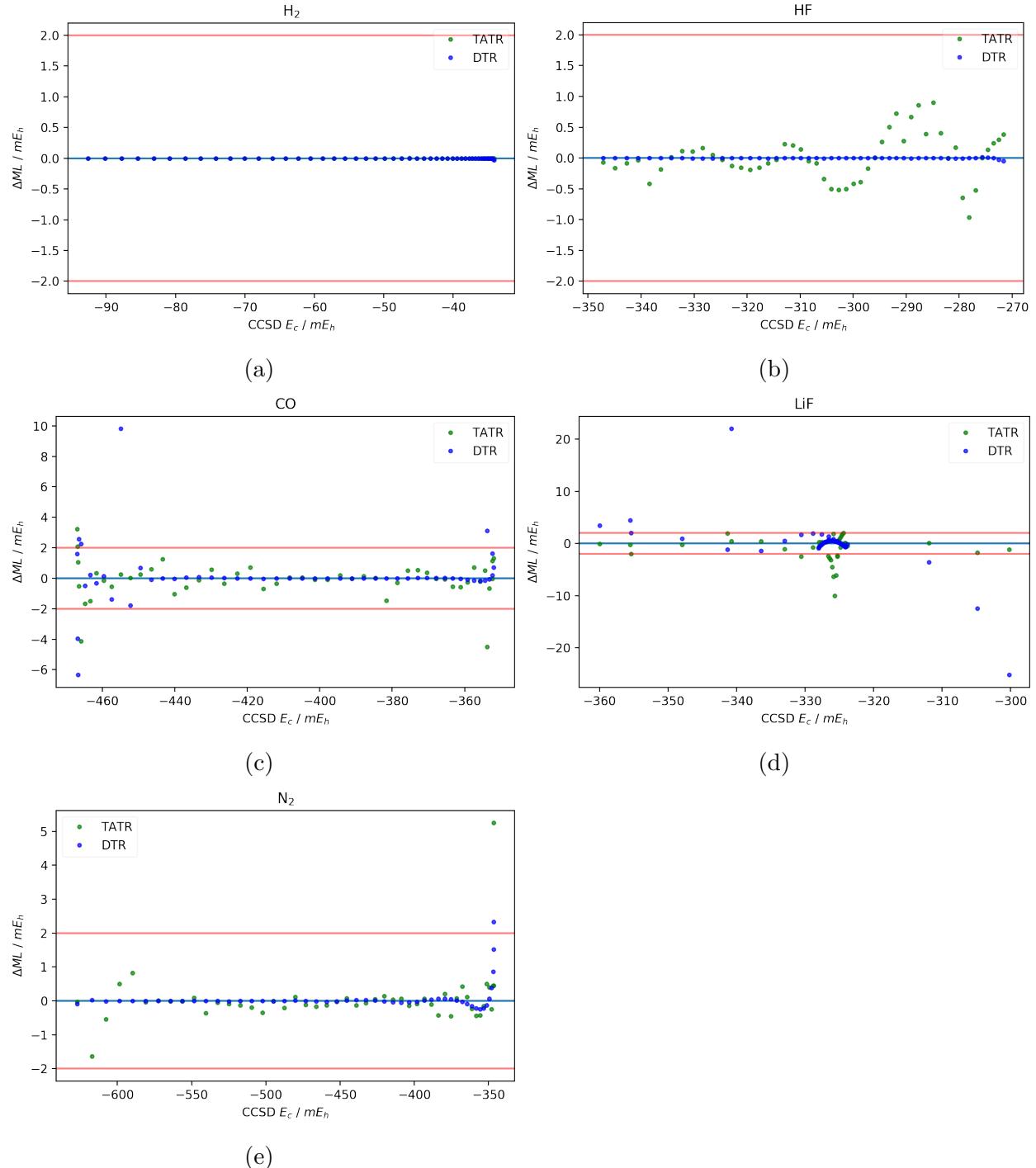


Figure B.1: DTR vs TATR errors in mE_h for diatomic datasets: (a) H₂, (b) HF, (c) CO, (d) LiF, and (e) N₂. Red lines indicate $\pm 2 mE_h$.

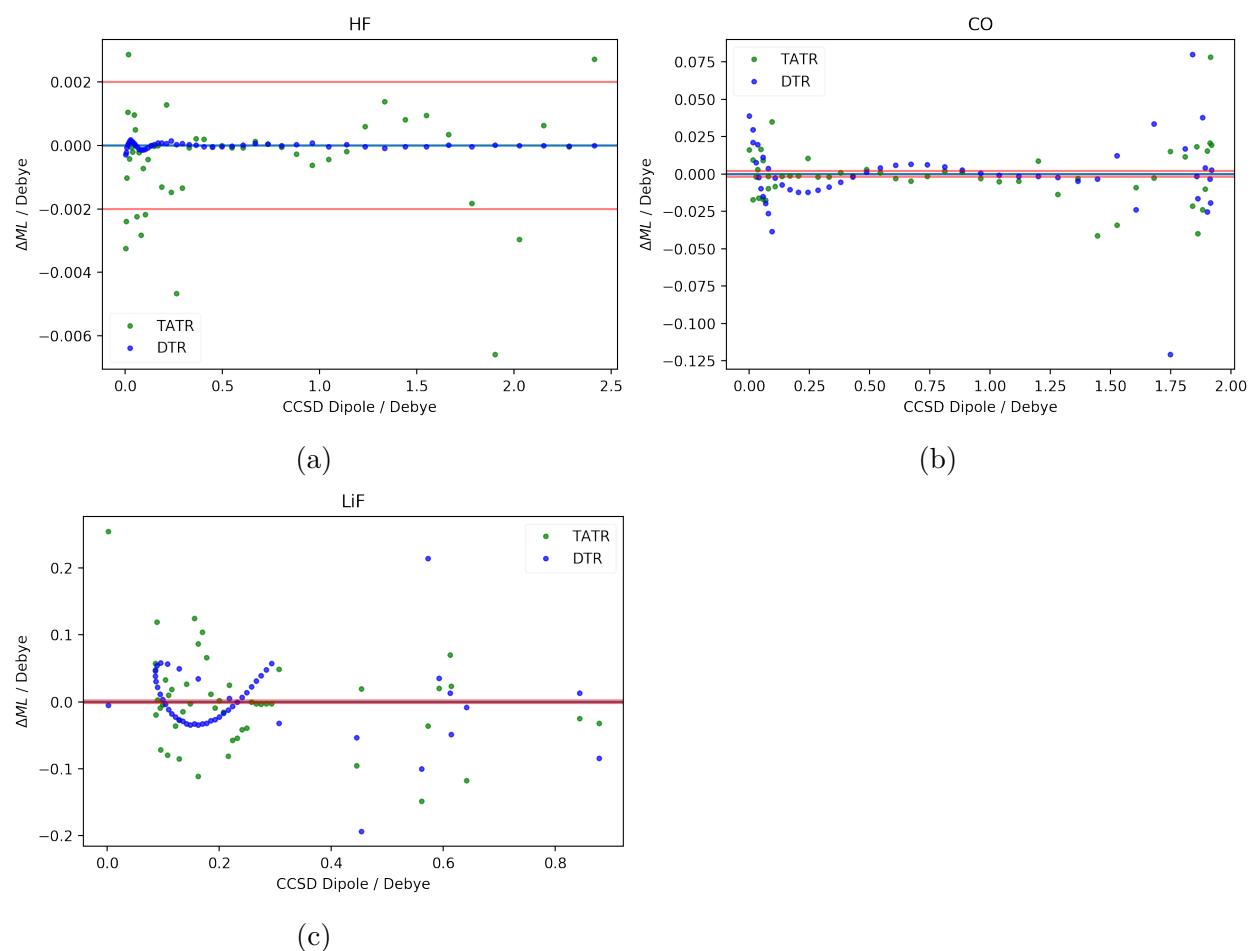


Figure B.2: DTR vs TATR errors in Debye for diatomic datasets: (a) HF, (b) CO, and (c) LiF. Red lines indicate 2 milliDebye.

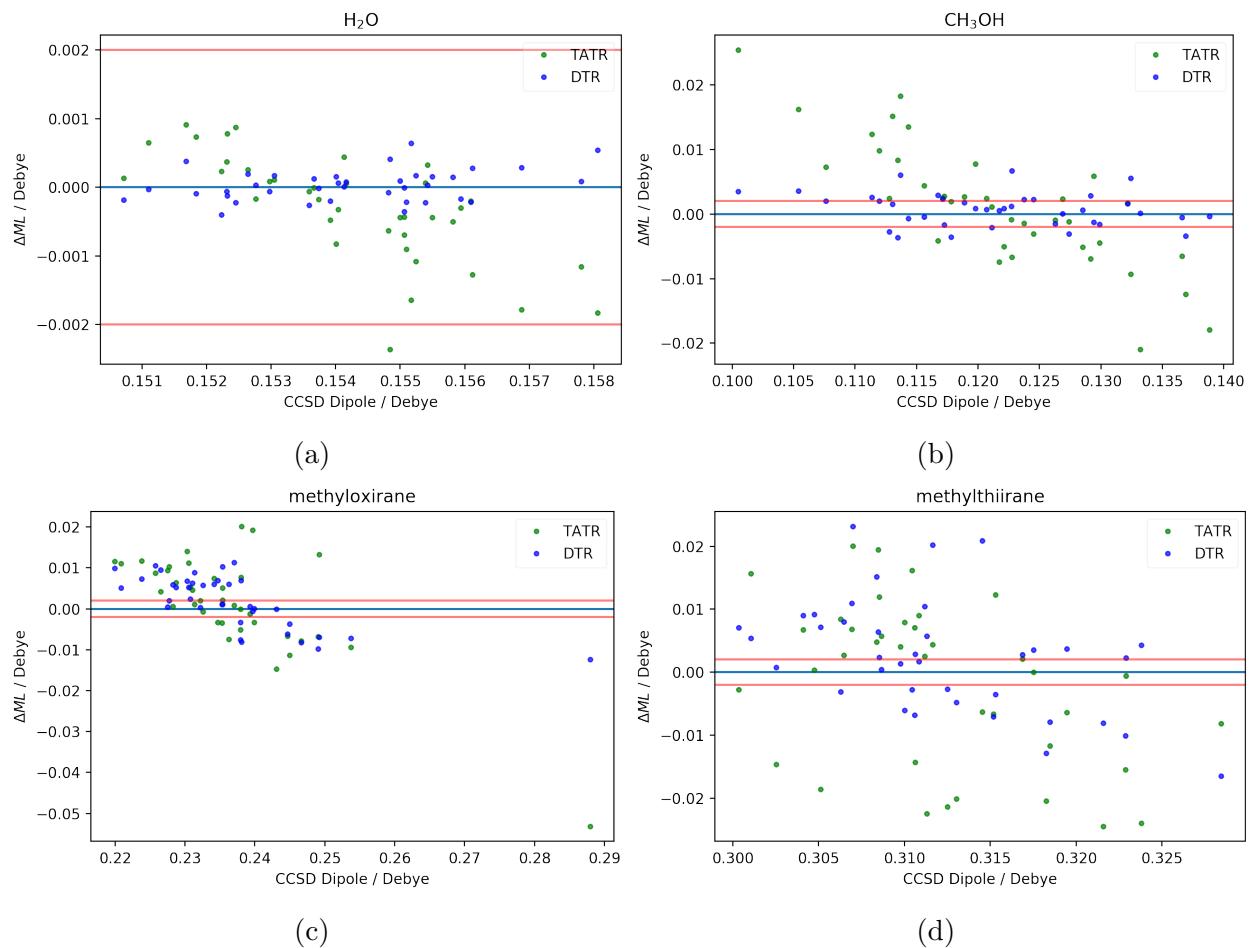


Figure B.3: DTR vs TATR errors in Debye for small molecule datasets: (a) H_2O , (b) CH_3OH , (c) (*S*)-methylloxirane, and (d) (*R*)-methylthiirane. Red lines indicate 2 milliDebye.

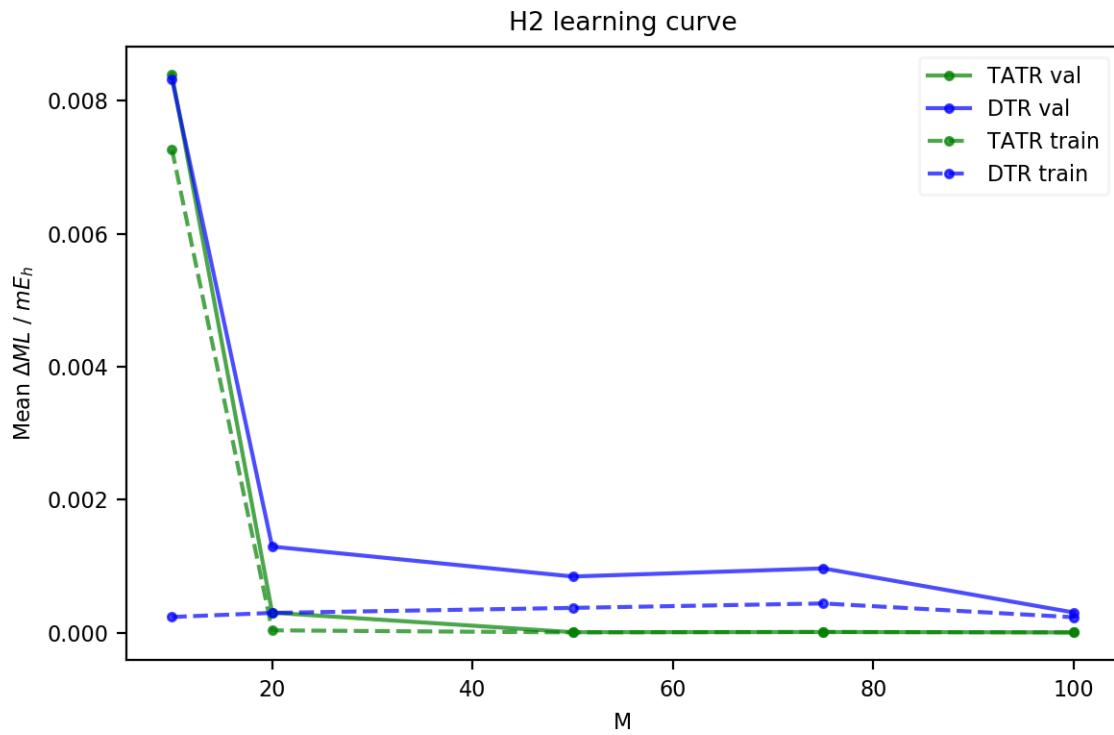


Figure B.4: DTR and TATR learning curves for H₂ correlation energy.

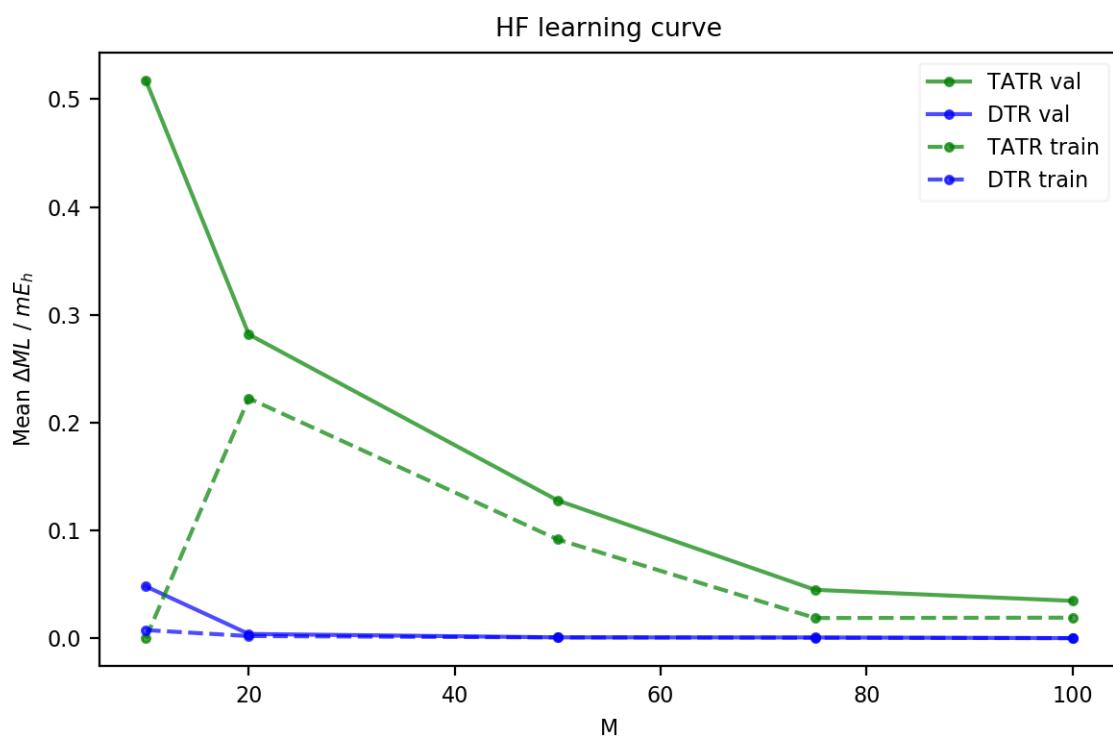


Figure B.5: DTR and TATR learning curves for HF correlation energy.

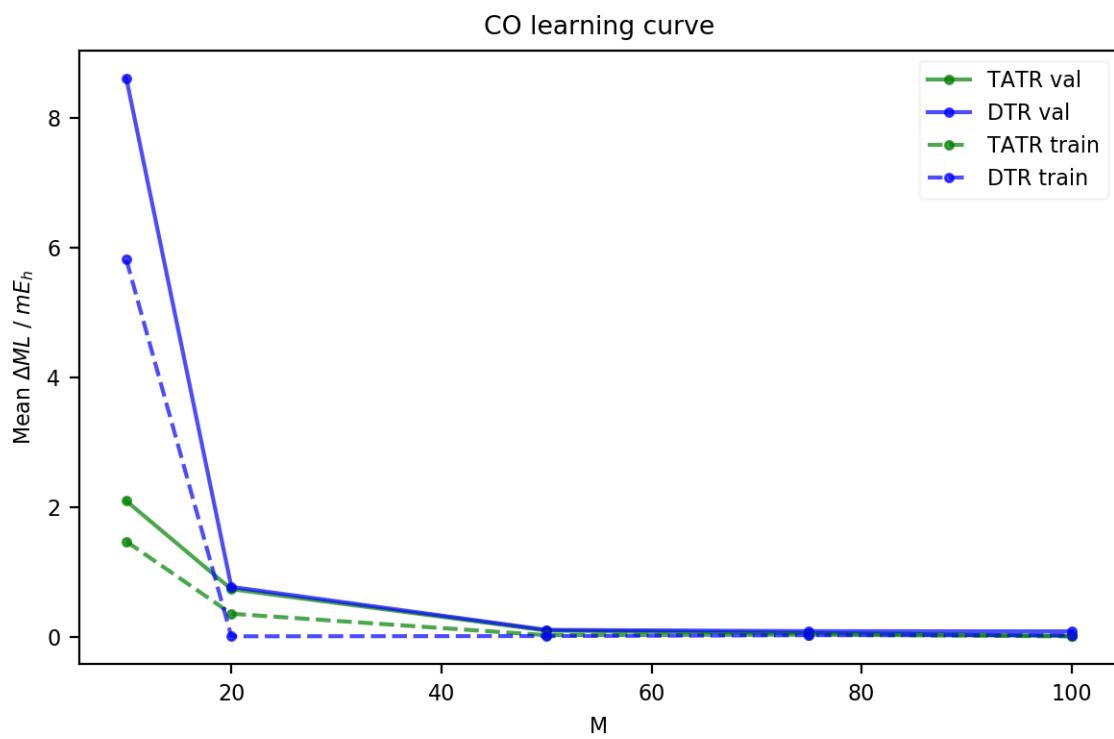


Figure B.6: DTR and TATR learning curves for CO correlation energy.

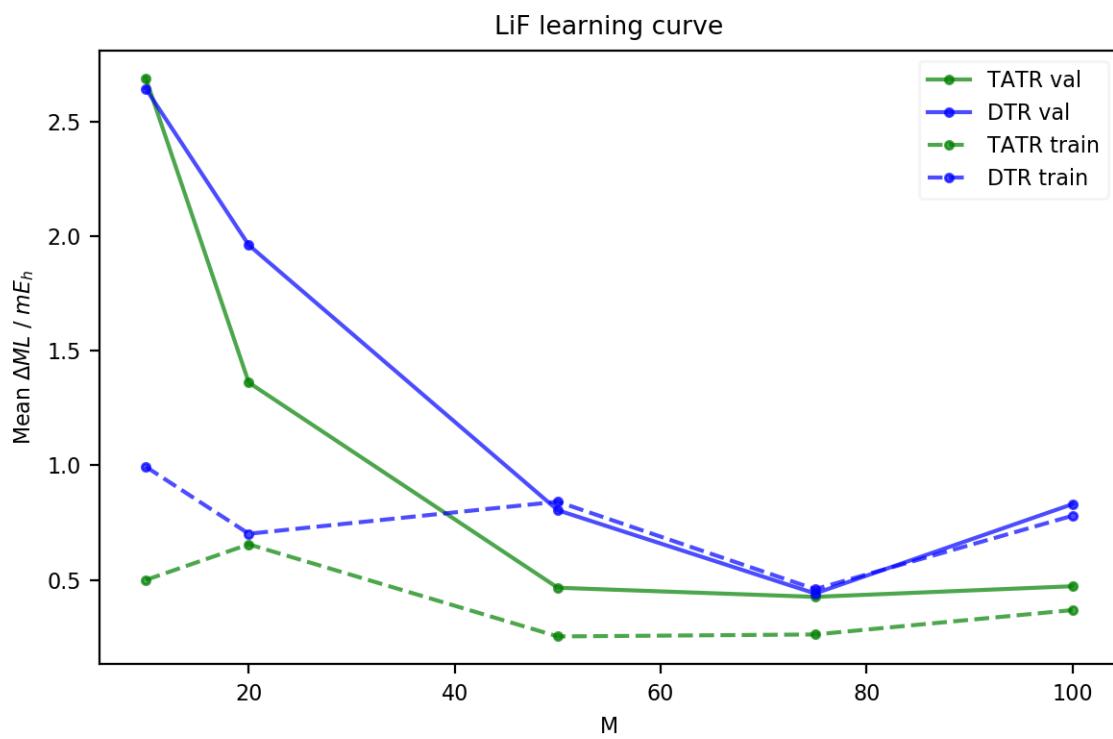


Figure B.7: DTR and TATR learning curves for LiF correlation energy.

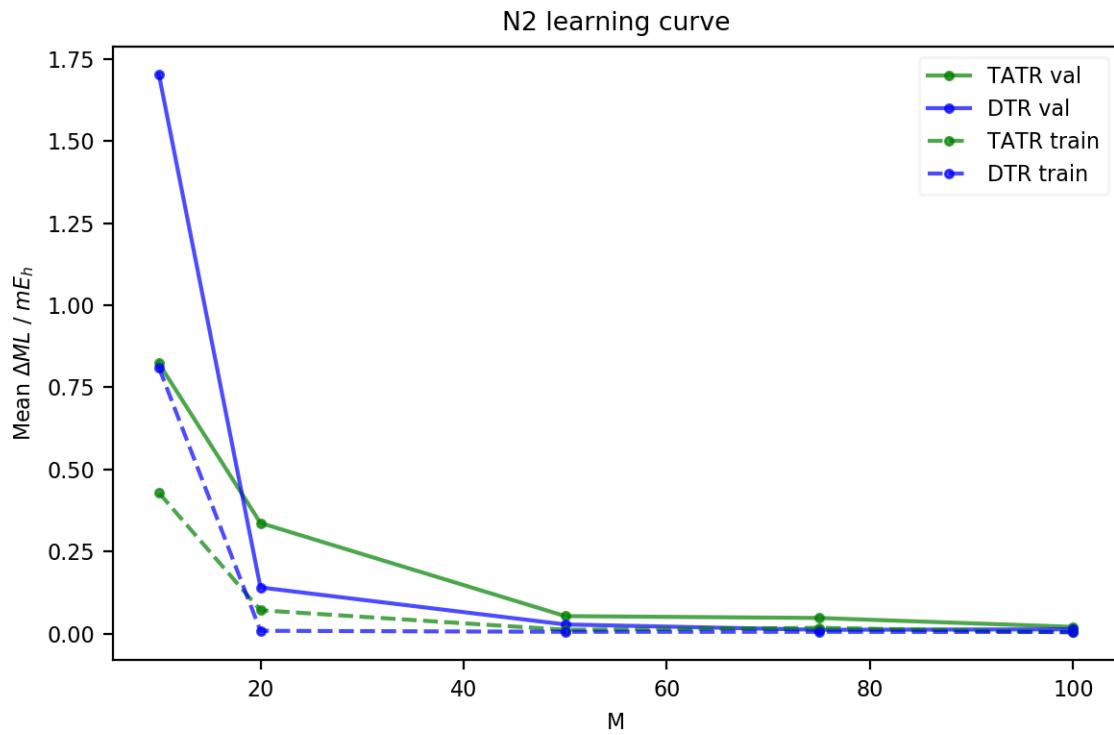


Figure B.8: DTR and TATR learning curves for N_2 correlation energy.

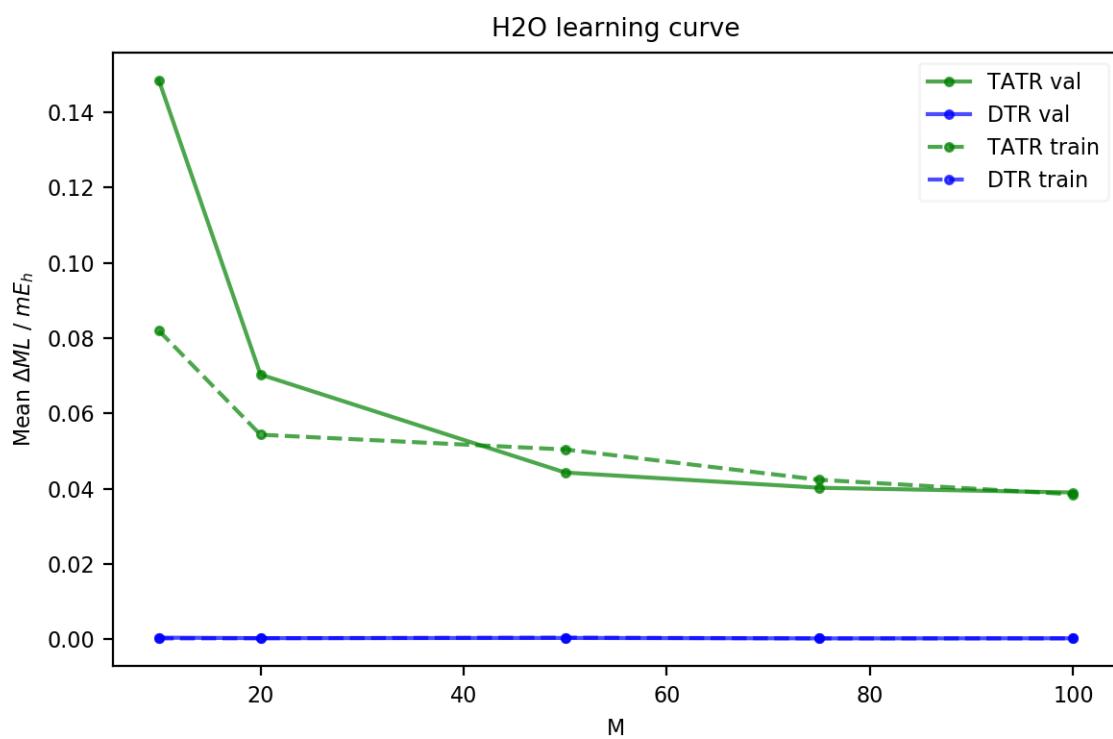


Figure B.9: DTR and TATR learning curves for H₂O correlation energy.

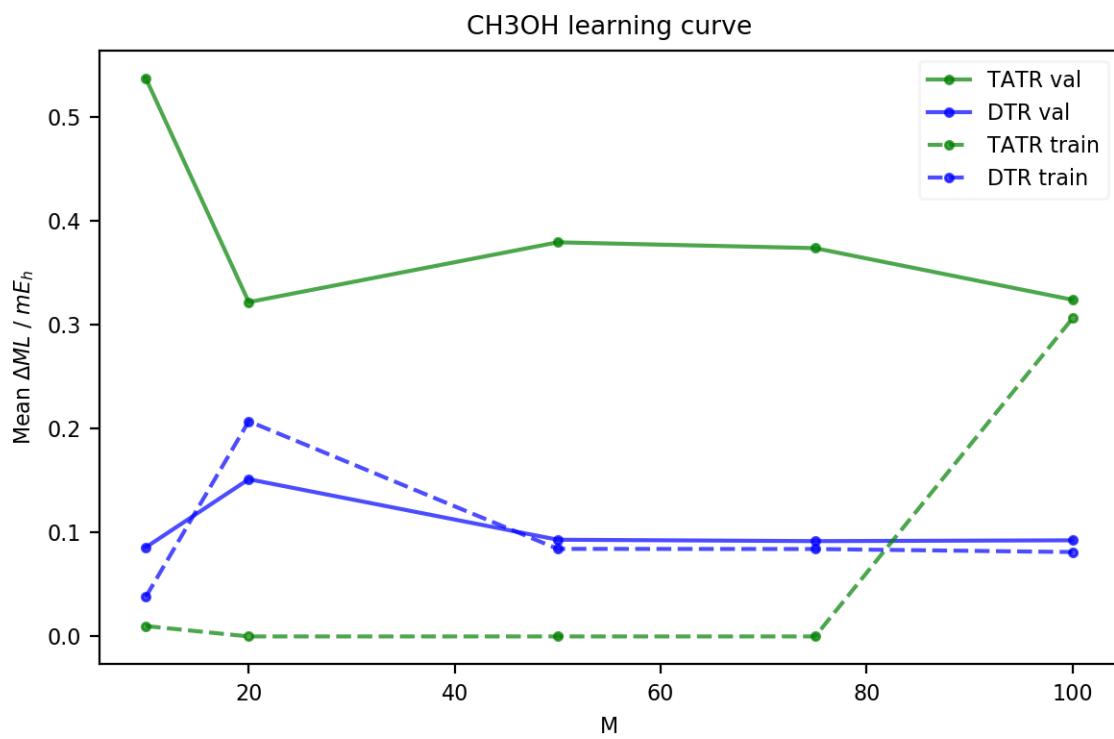


Figure B.10: DTR and TATR learning curves for CH₃OH correlation energy.

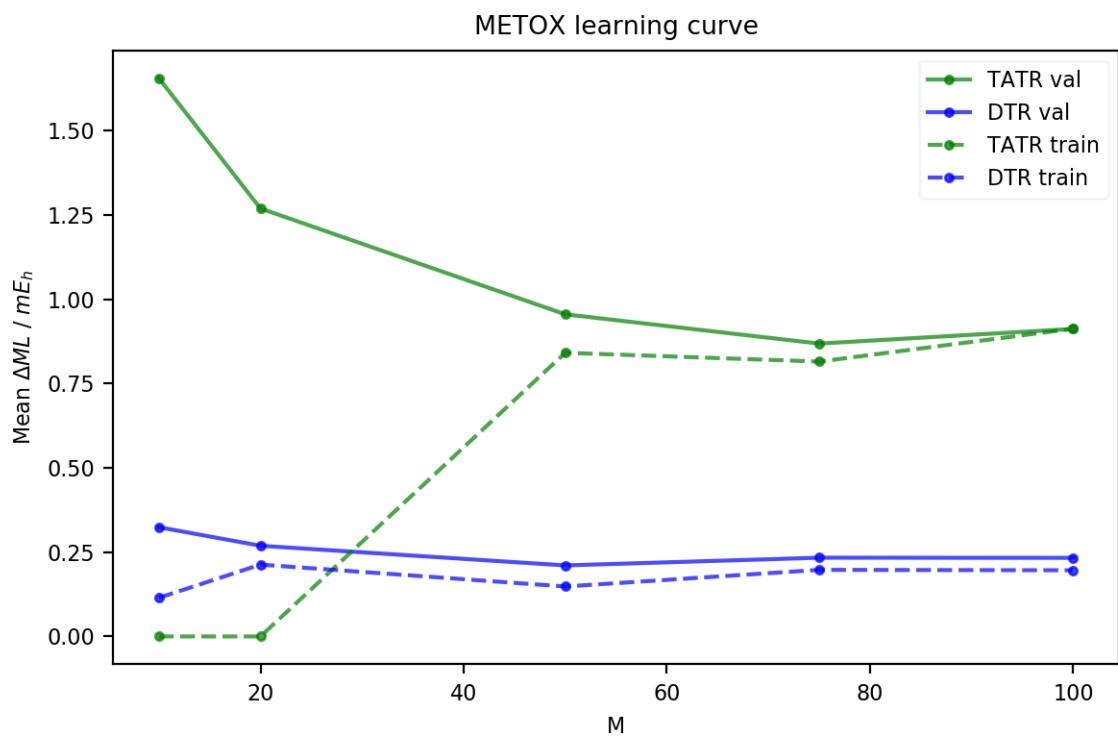


Figure B.11: DTR and TATR learning curves for (*S*)-methyloxirane correlation energy.

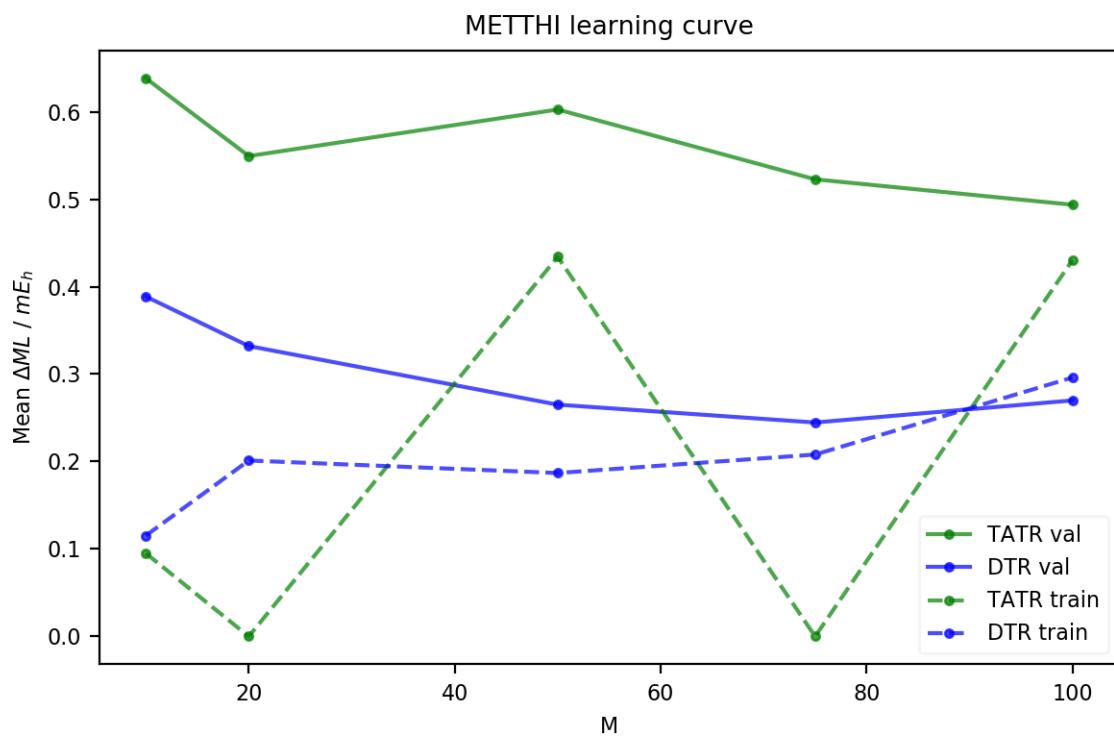


Figure B.12: DTR and TATR learning curves for (*R*)-methylthiirane correlation energy.

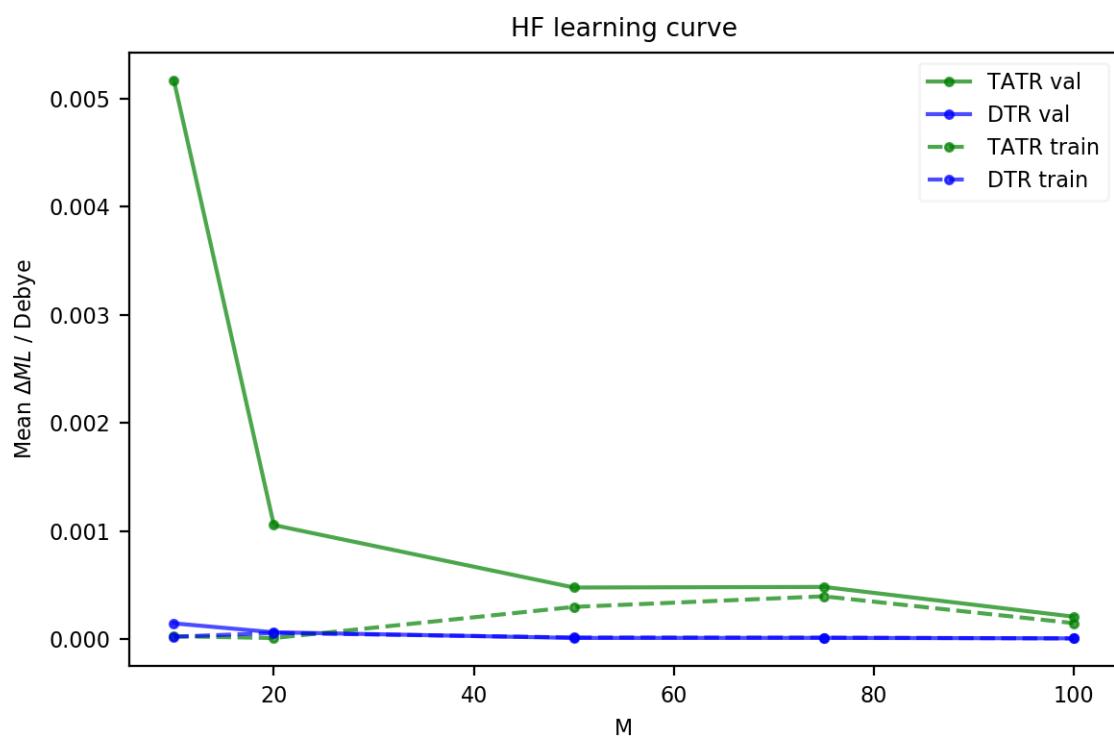


Figure B.13: DTR and TATR learning curves for HF correlated dipole.

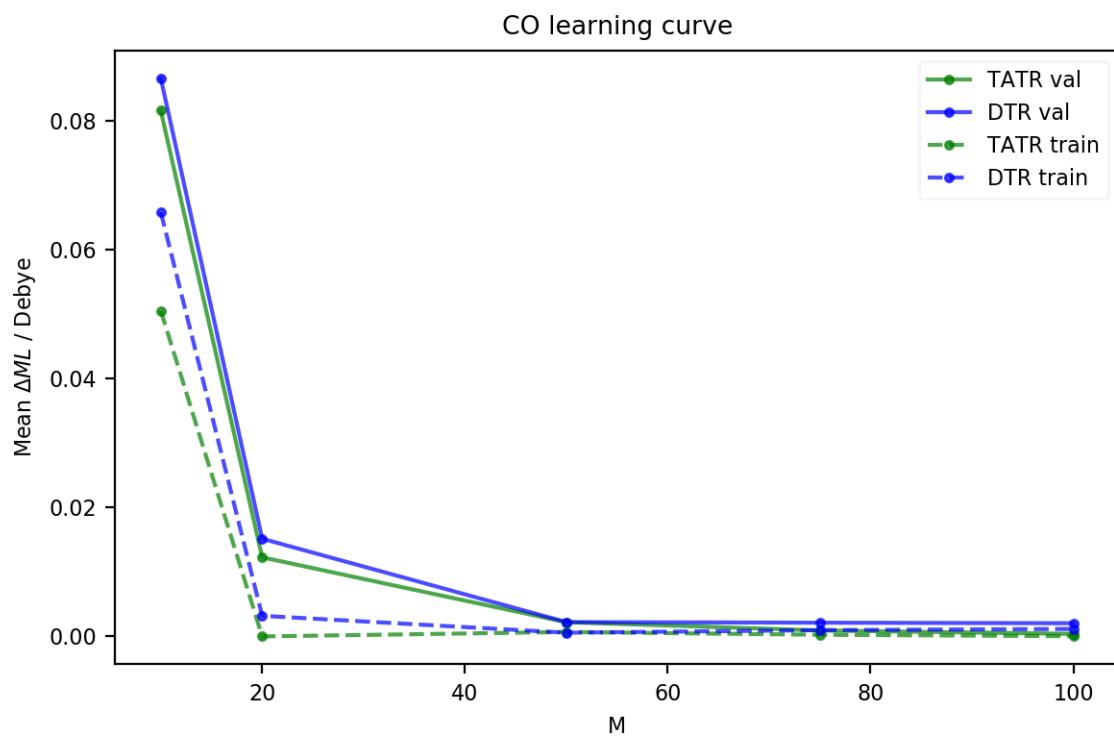


Figure B.14: DTR and TATR learning curves for CO correlated dipole.

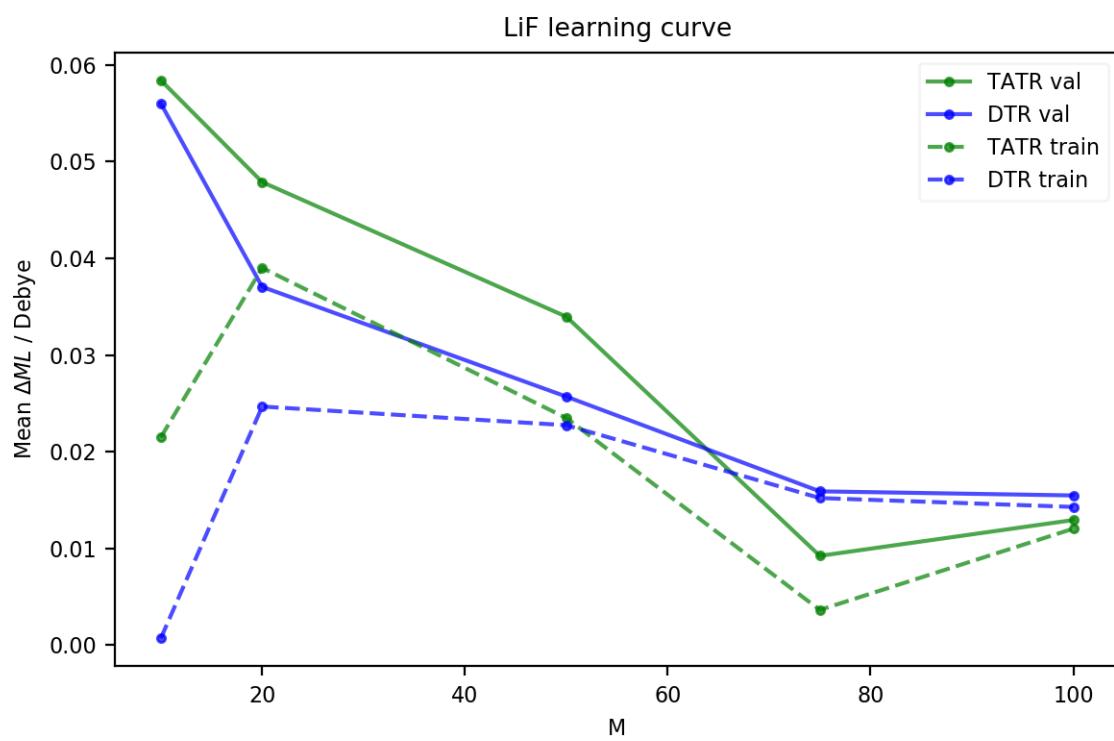


Figure B.15: DTR and TATR learning curves for LiF correlated dipole.

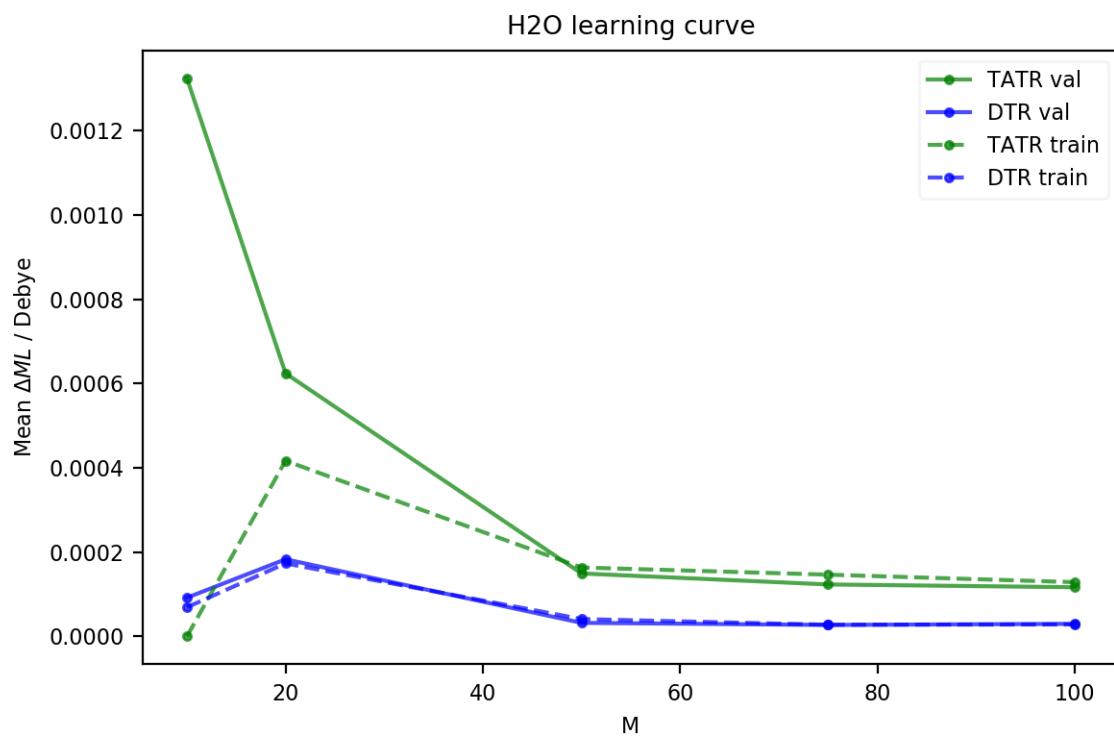


Figure B.16: DTR and TATR learning curves for H₂O correlated dipole.

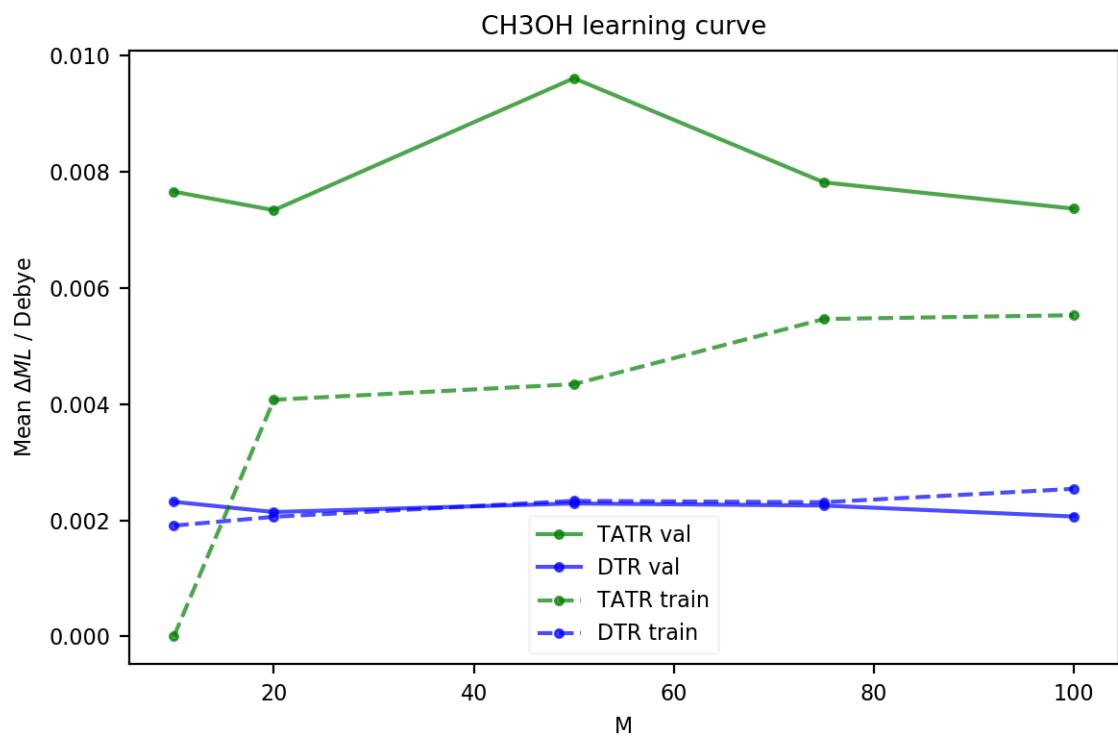


Figure B.17: DTR and TATR learning curves for CH₃OH correlated dipole.

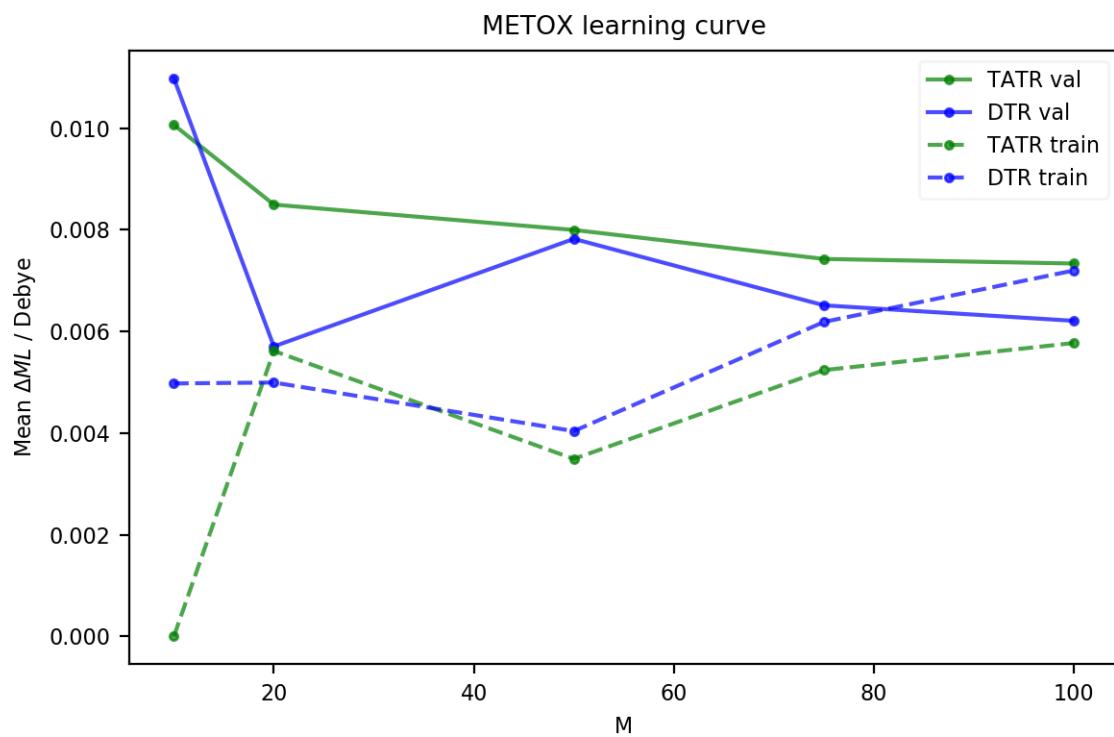


Figure B.18: DTR and TATR learning curves for (*S*)-methyloxirane correlated dipole.

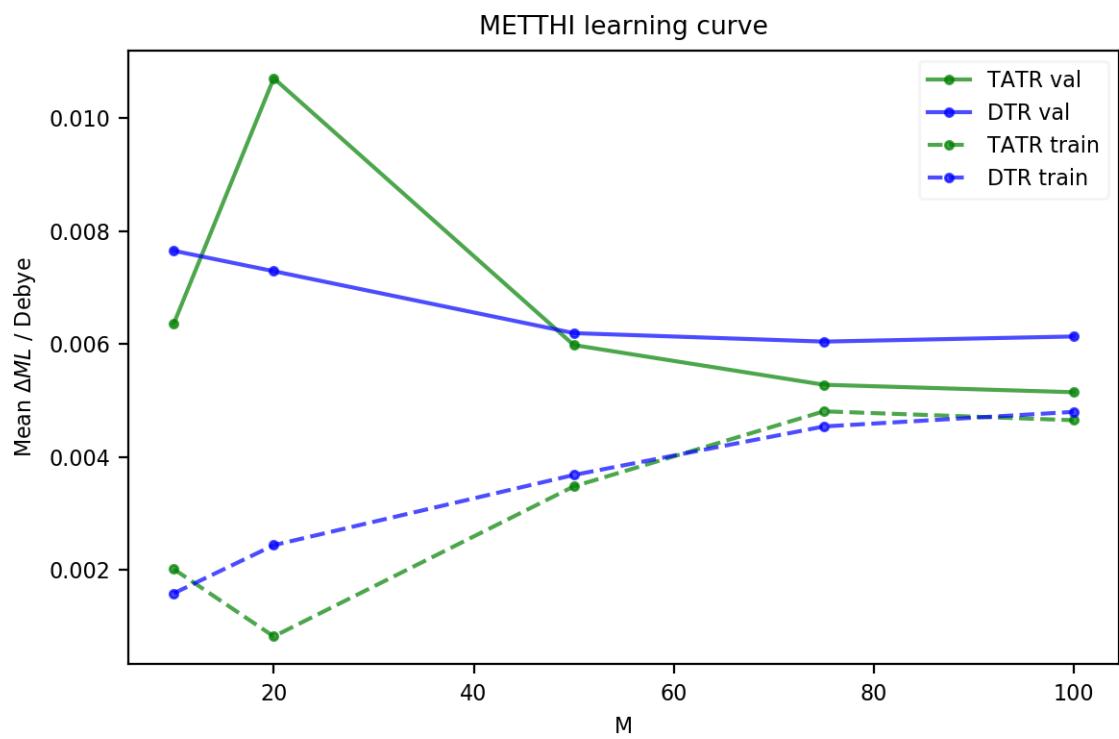


Figure B.19: DTR and TATR learning curves for (*R*)-methylthiirane correlated dipole.

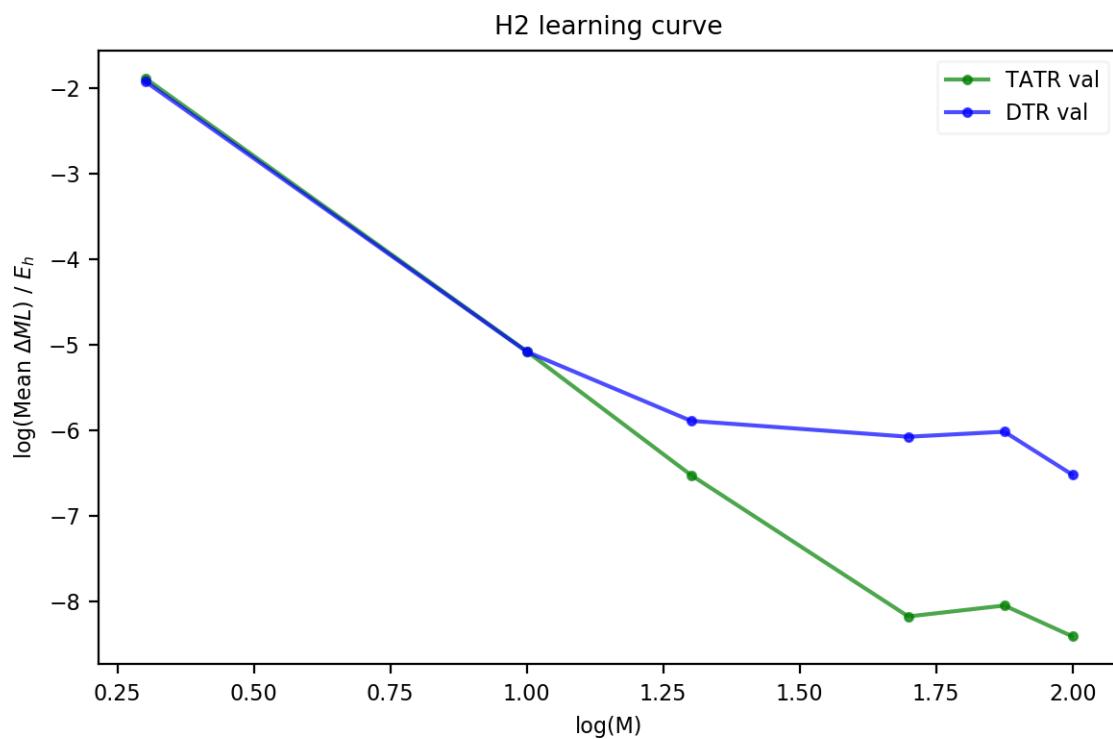


Figure B.20: DTR and TATR validation curves for H₂ correlation energy.

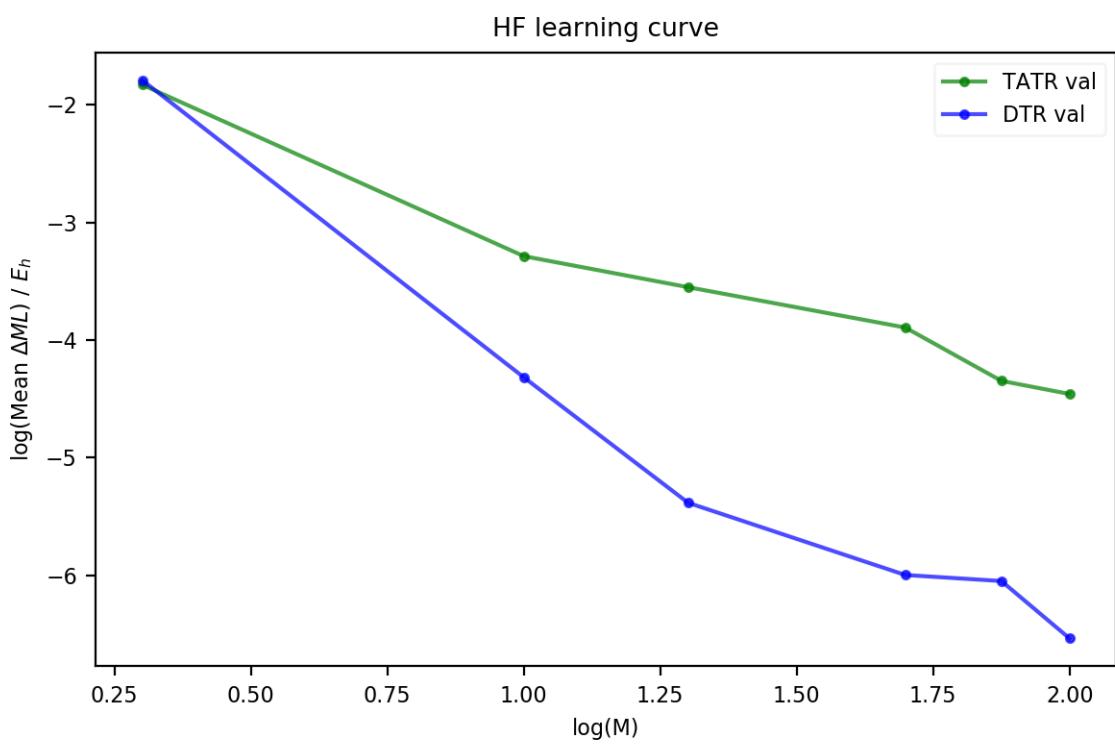


Figure B.21: DTR and TATR validation curves for HF correlation energy.

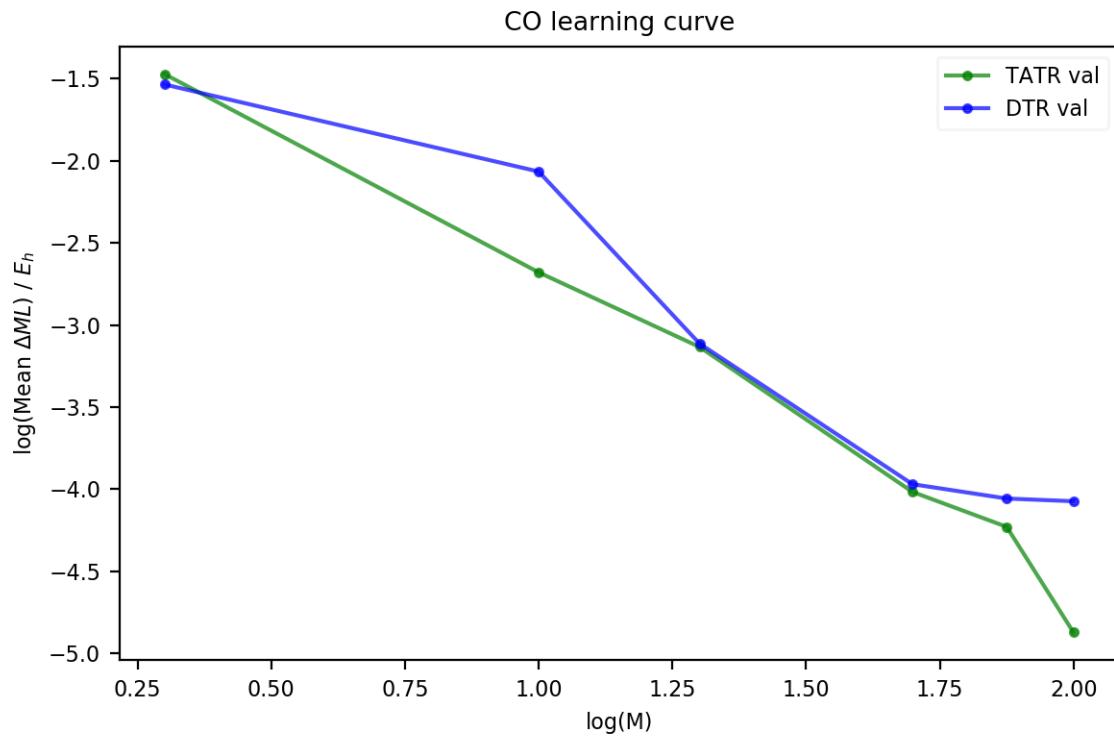


Figure B.22: DTR and TATR validation curves for CO correlation energy.

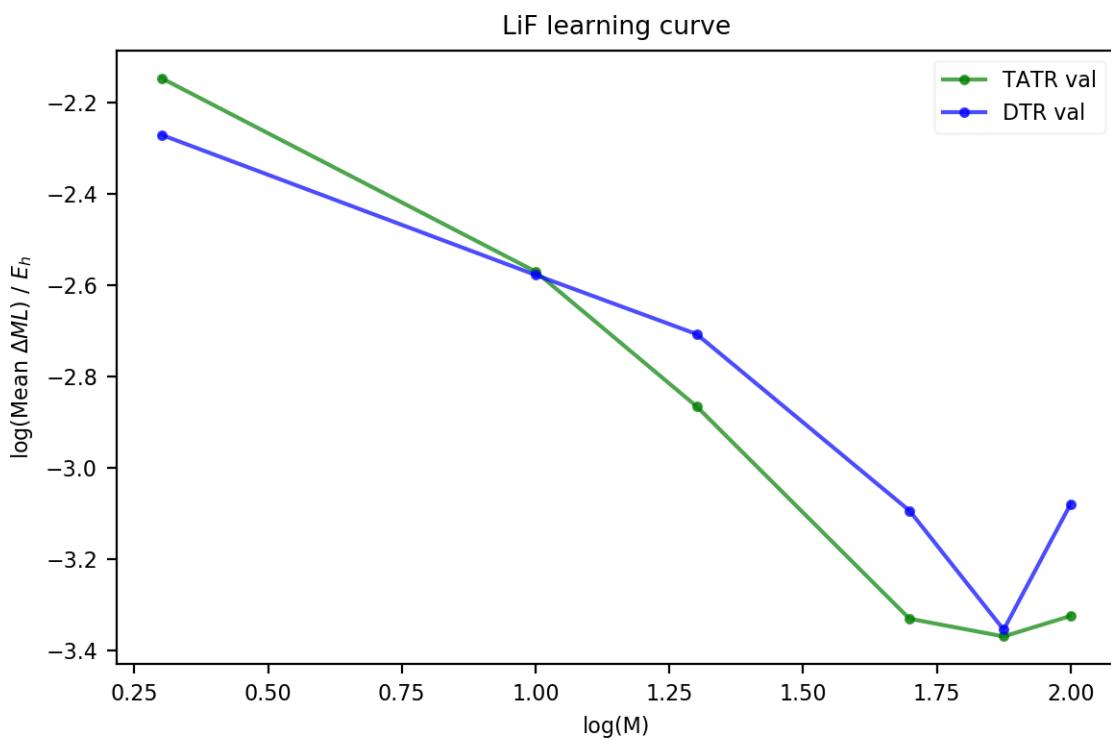


Figure B.23: DTR and TATR validation curves for LiF correlation energy.

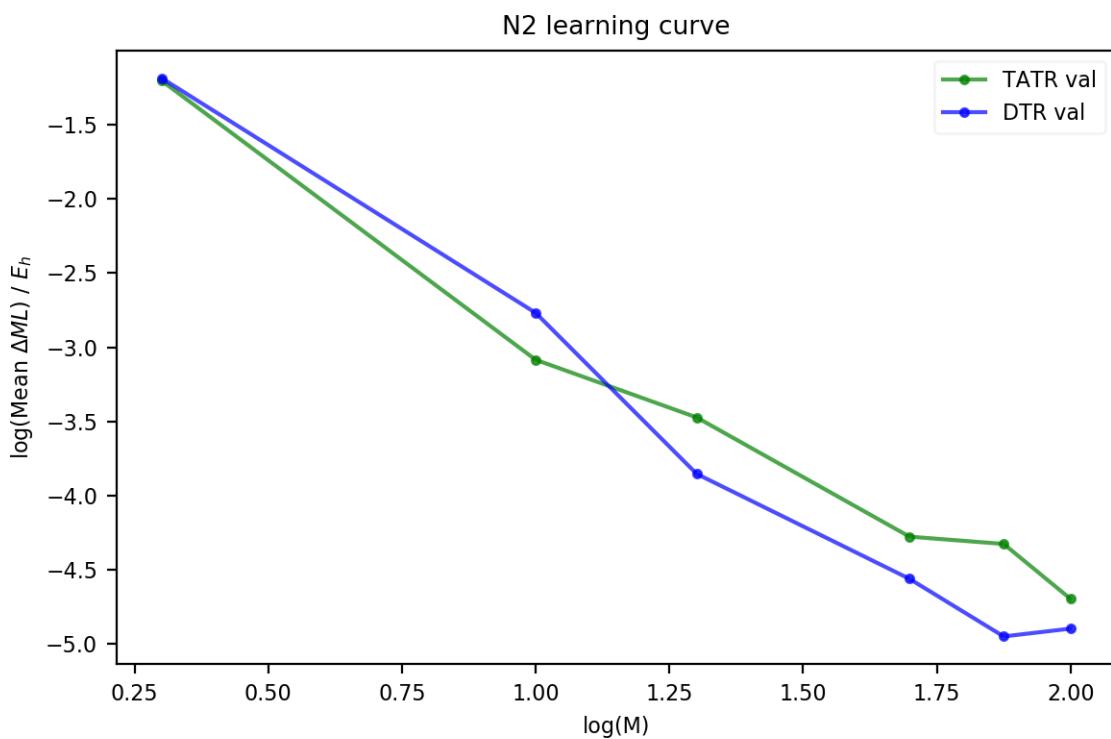


Figure B.24: DTR and TATR validation curves for N₂ correlation energy.

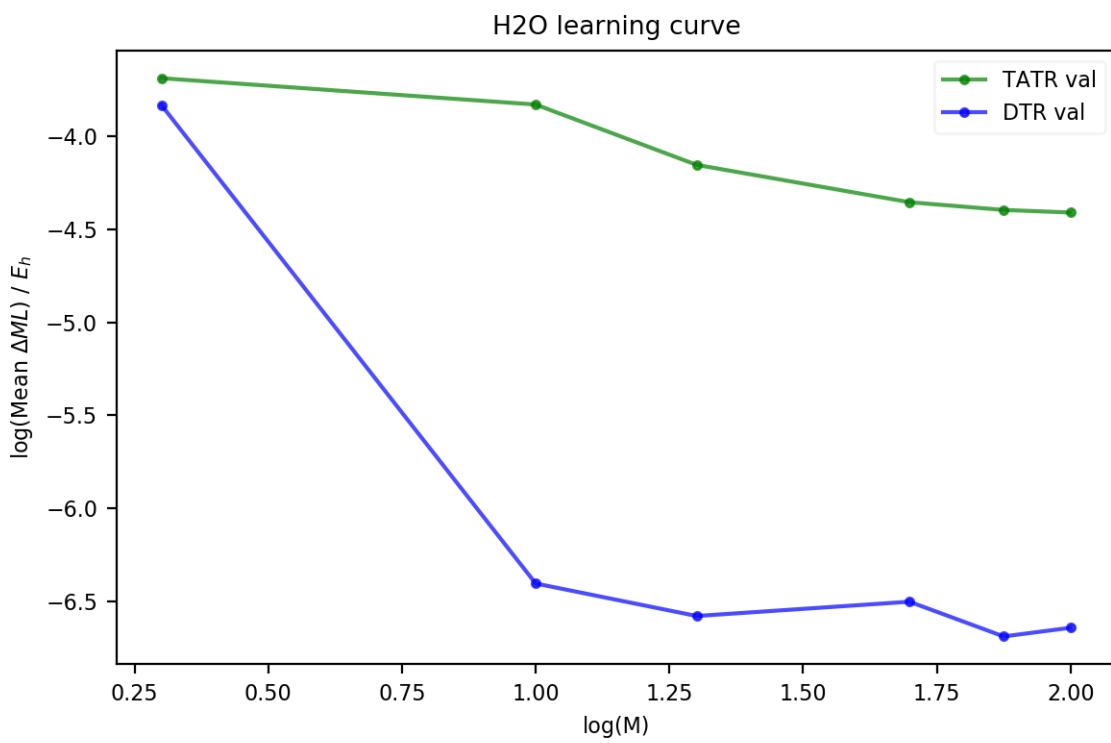


Figure B.25: DTR and TATR validation curves for H₂O correlation energy.

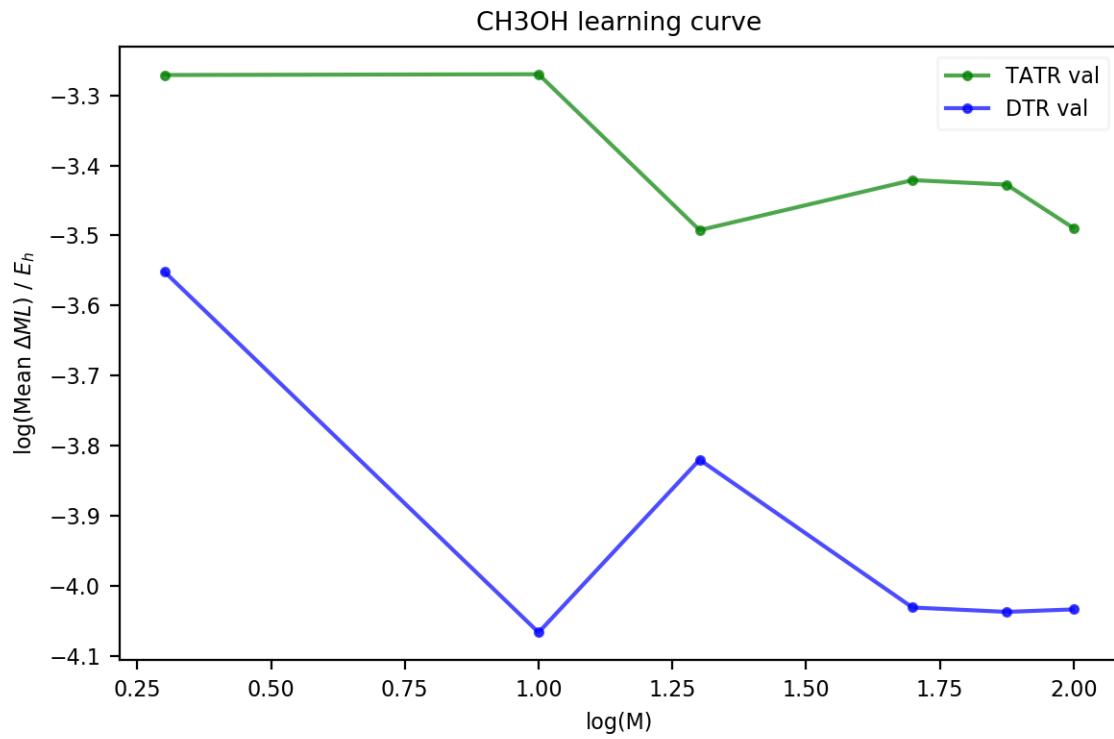


Figure B.26: DTR and TATR validation curves for CH_3OH correlation energy.

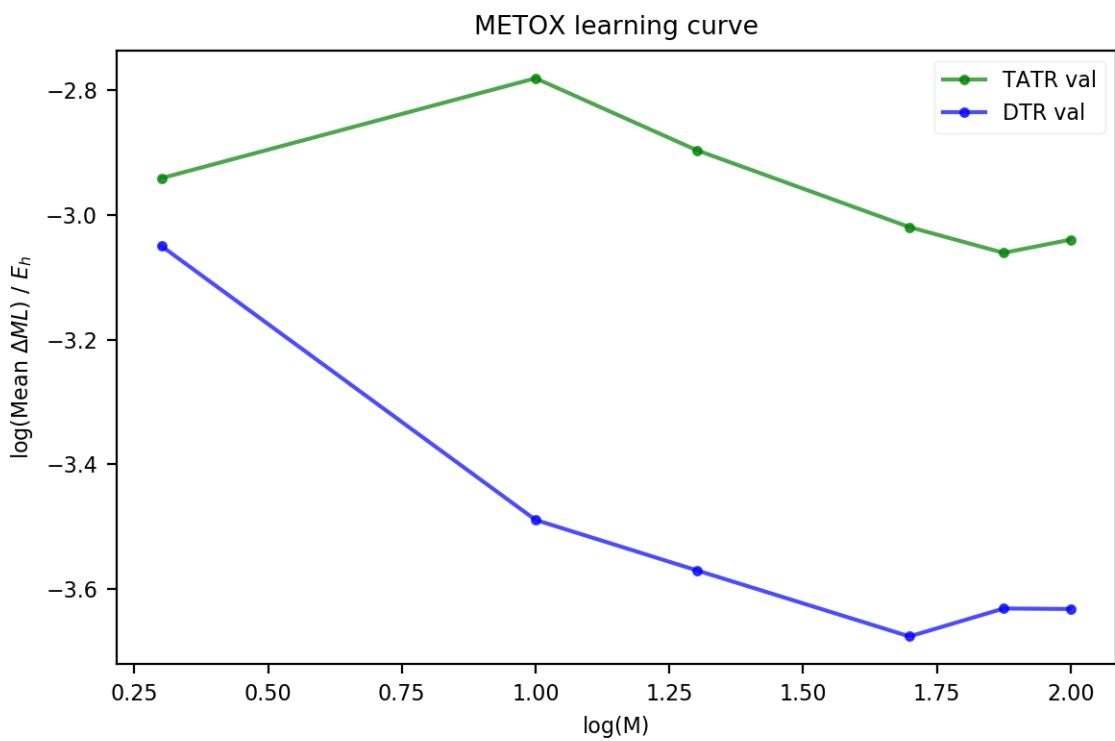


Figure B.27: DTR and TATR validation curves for (*S*)-methyloxirane correlation energy.

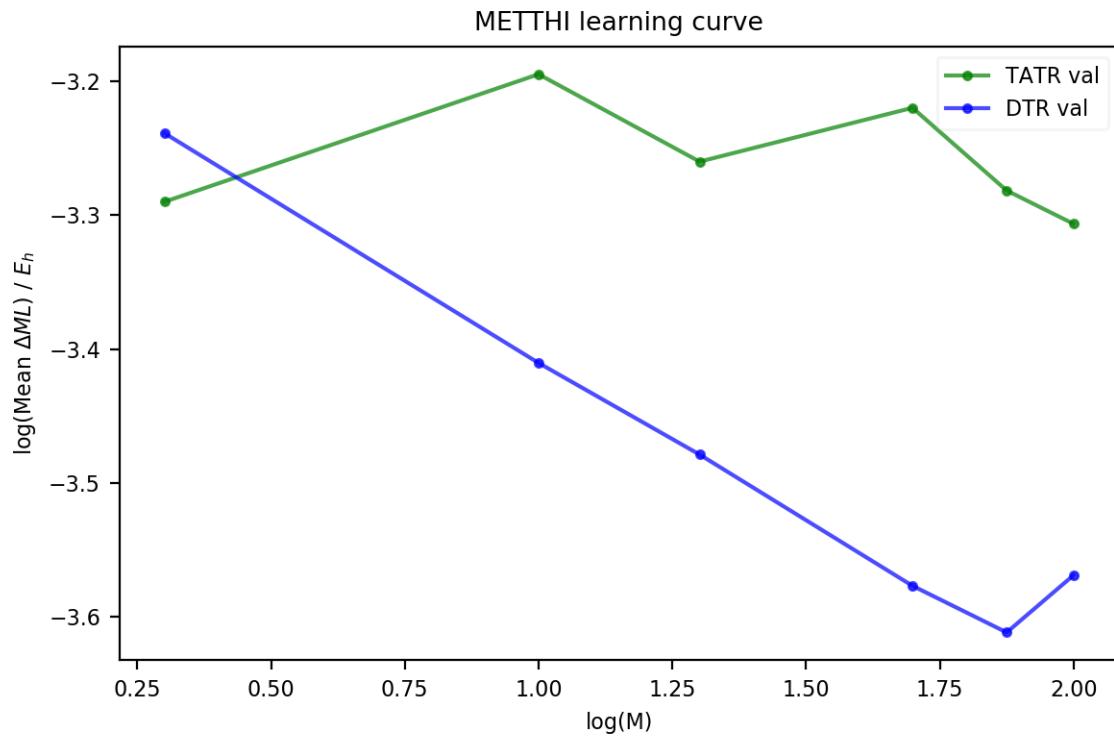


Figure B.28: DTR and TATR validation curves for (*R*)-methylthiirane correlation energy.

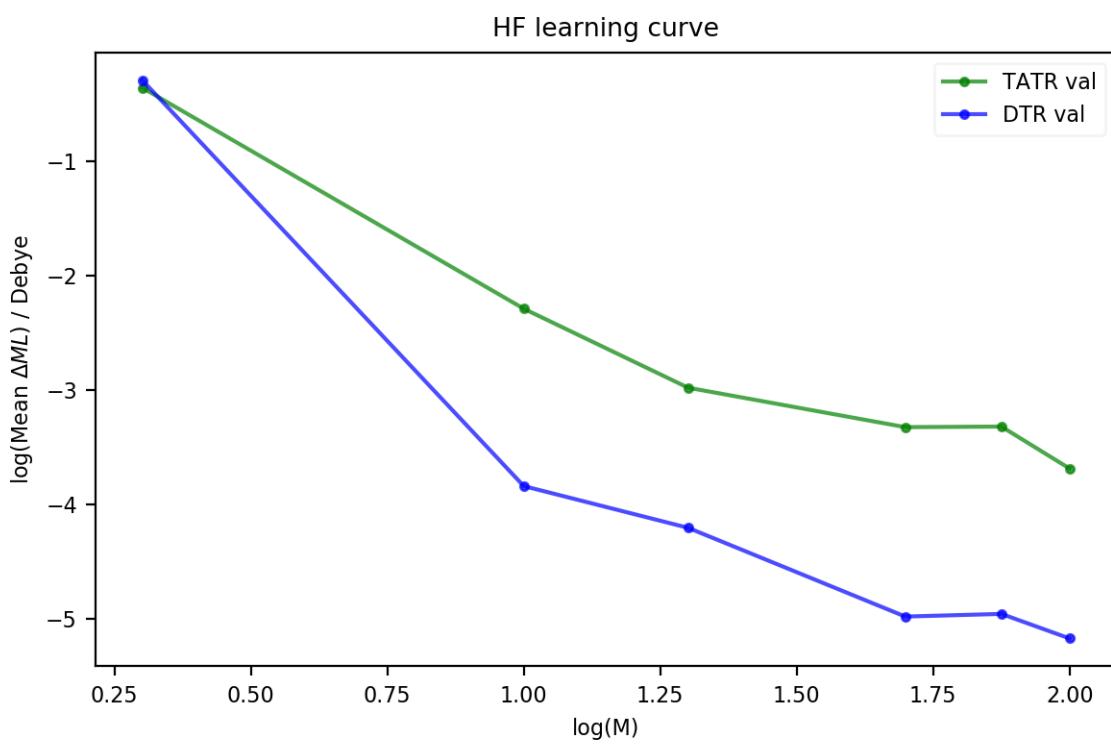


Figure B.29: DTR and TATR validation curves for HF correlated dipole.

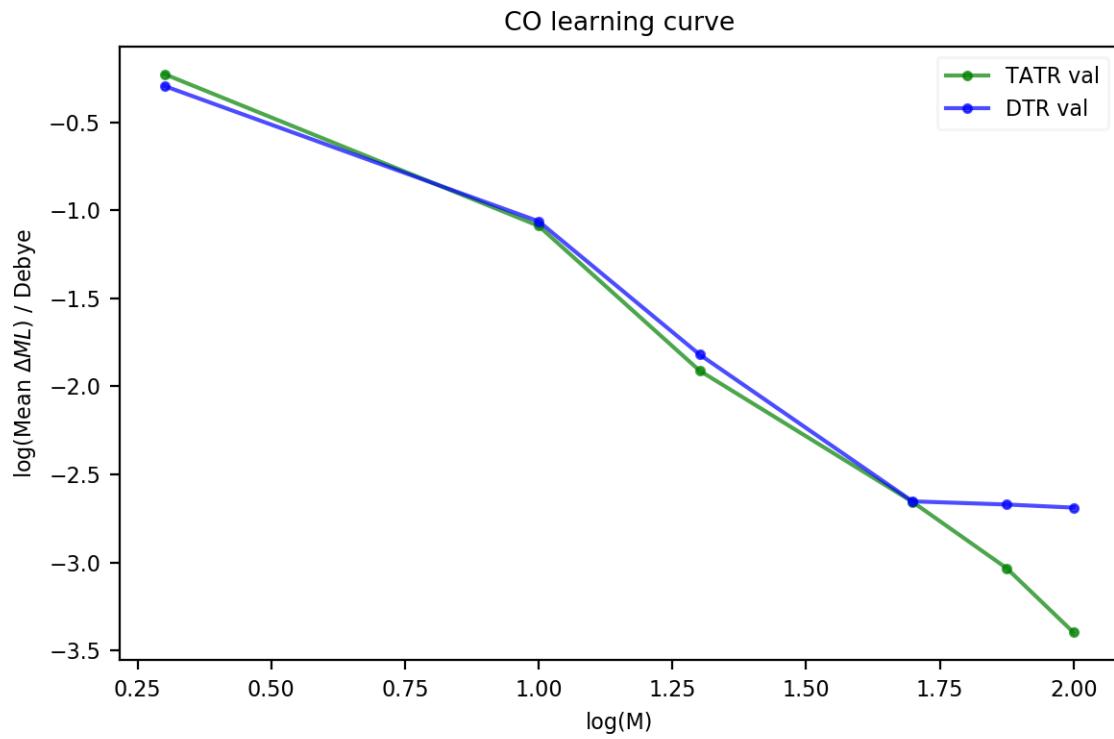


Figure B.30: DTR and TATR validation curves for CO correlated dipole.

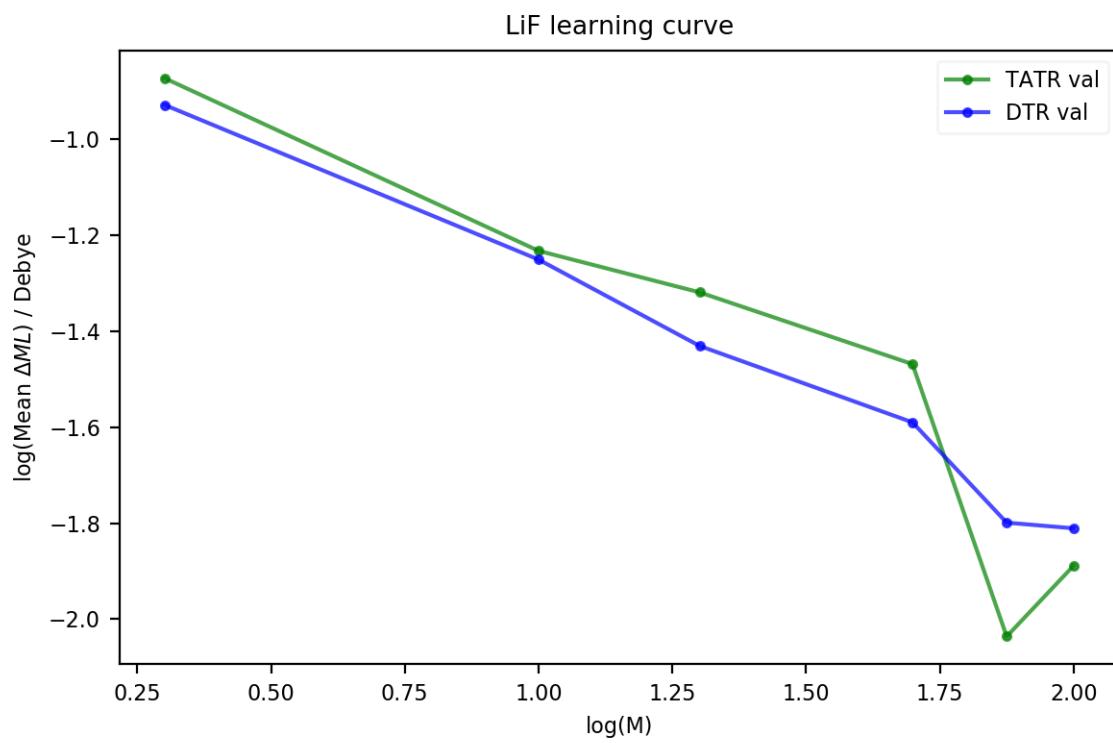


Figure B.31: DTR and TATR validation curves for LiF correlated dipole.

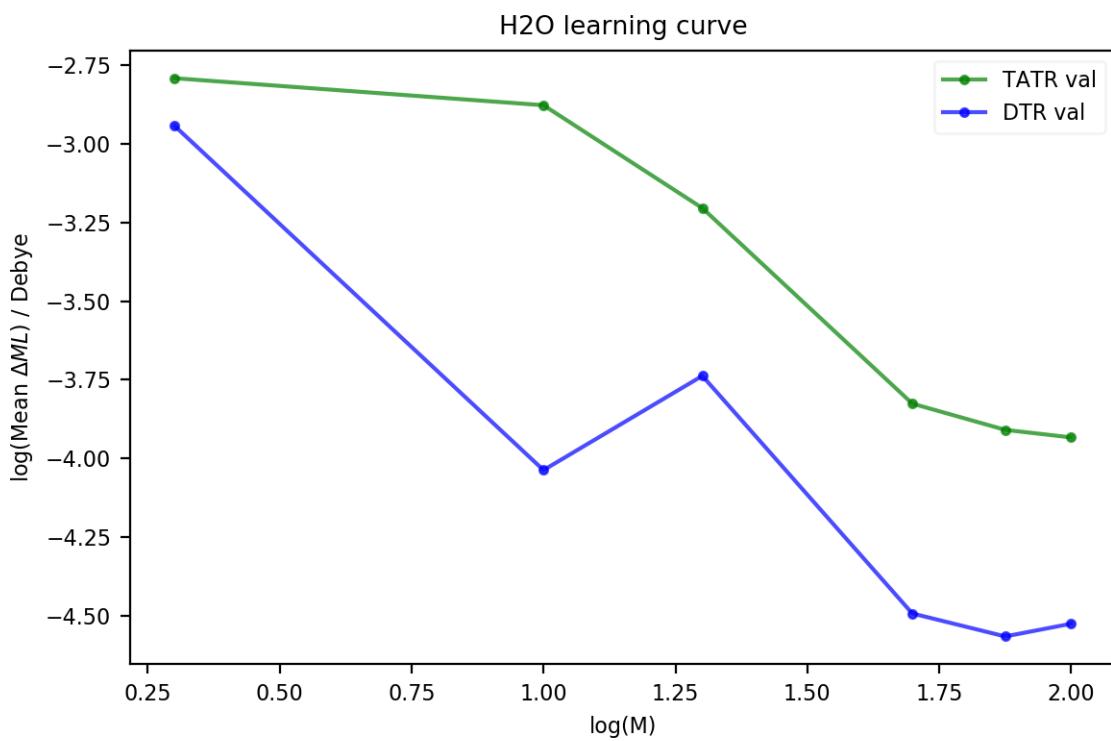


Figure B.32: DTR and TATR validation curves for H₂O correlated dipole.

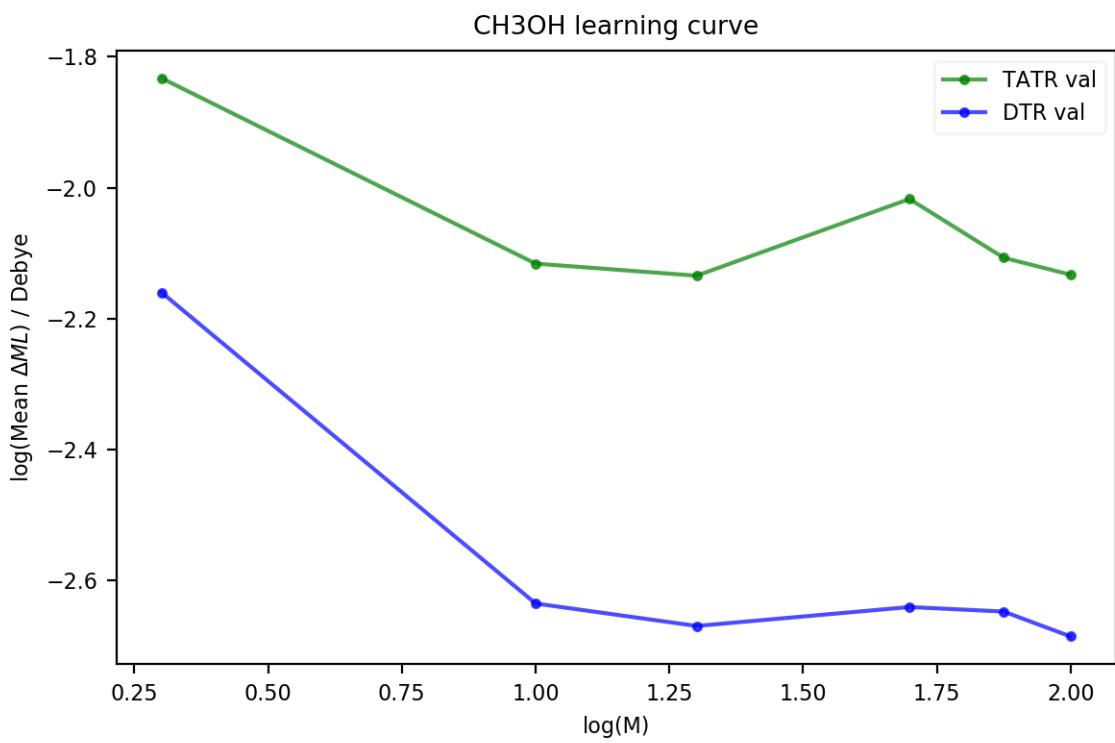


Figure B.33: DTR and TATR validation curves for CH₃OH correlated dipole.

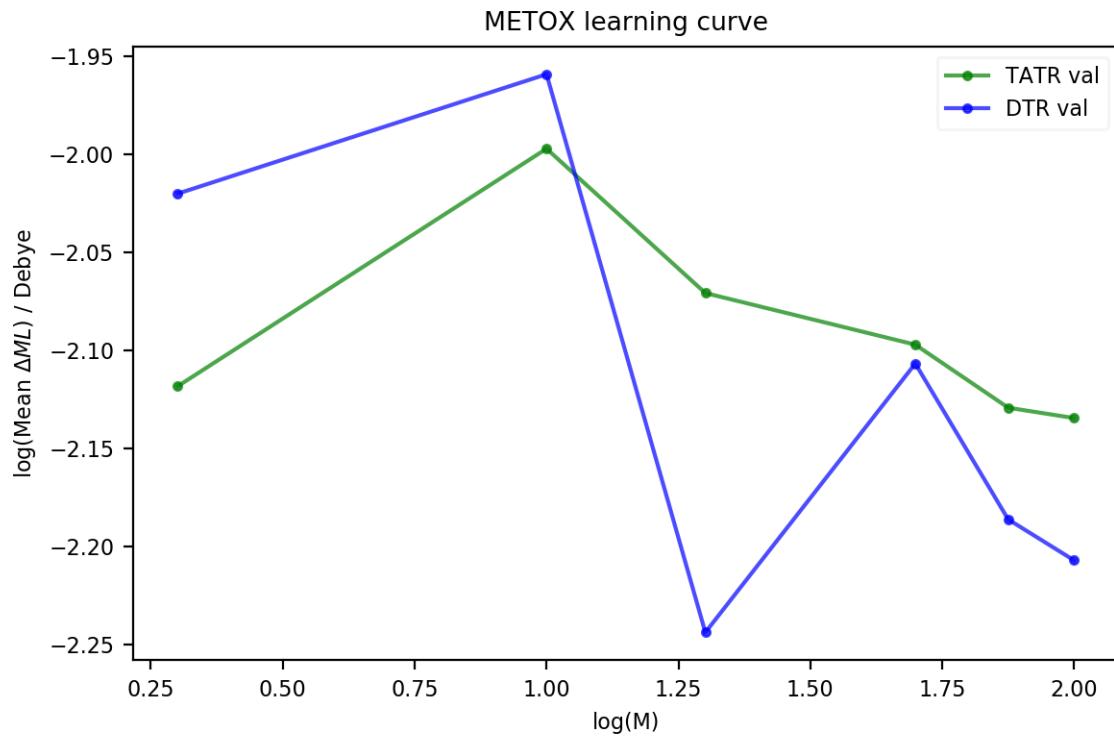


Figure B.34: DTR and TATR validation curves for (*S*)-methyloxirane correlated dipole.

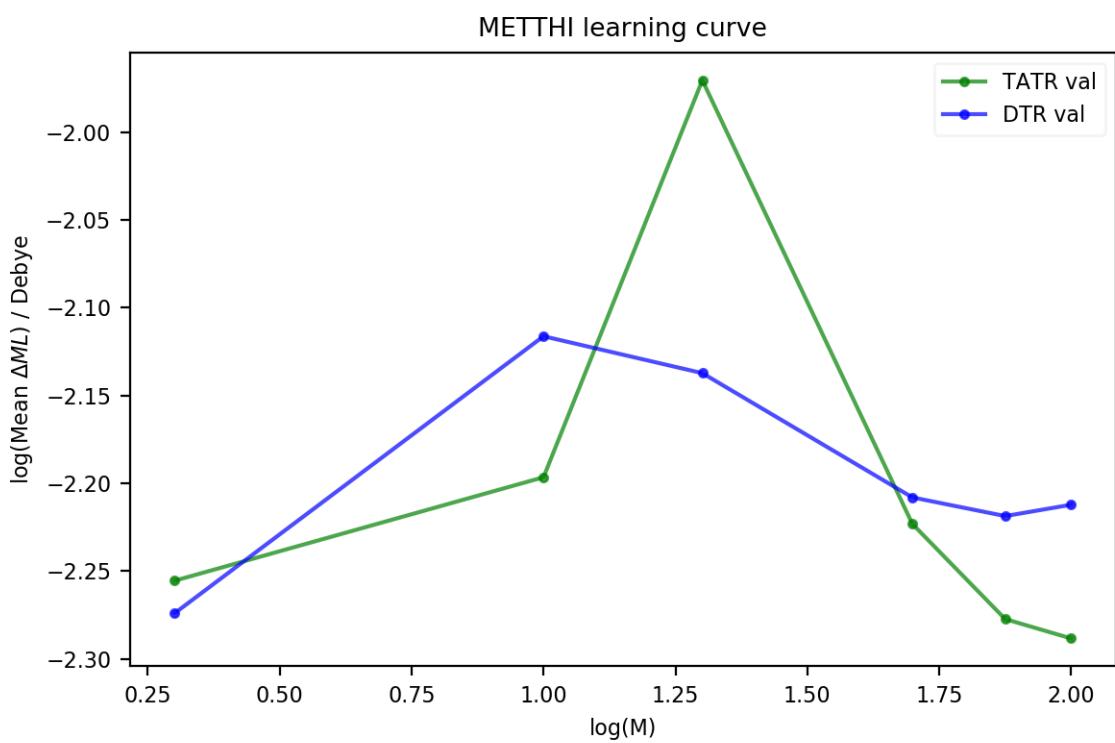


Figure B.35: DTR and TATR validation curves for (*R*)-methylthiirane correlated dipole.

Appendix C

Supporting Information for Locally Correlated Real-Time Coupled Cluster Theory

Table C.1: Atomic coordinates of Helical H₂ tetramer (Ångstroms)

Symbol	X	Y	Z
H	0.000000	0.000000	0.000000
H	0.750000	0.000000	0.000000
H	0.000000	1.500000	0.000000
H	0.375000	1.500000	-0.649520
H	0.000000	3.000000	0.000000
H	-0.375000	3.000000	-0.649520
H	0.000000	4.500000	-0.000000
H	-0.750000	4.500000	-0.000000

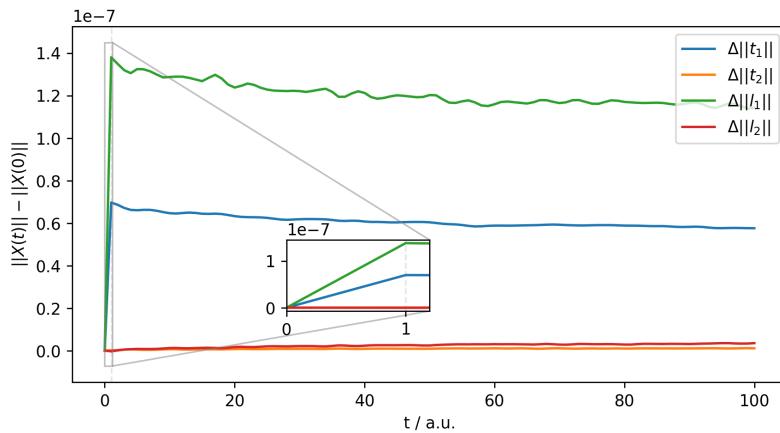


Figure C.1: Time-dependent change in the norm of the amplitude tensors relative to the ground-state amplitudes, using a PNO cut-off of 1×10^{-9} . (Amplitudes have been back-transformed to the MO space for comparison to MO-basis amplitude norms. Field and step parameters remain unchanged, and the amplitude norm is taken at every 1 a.u.)

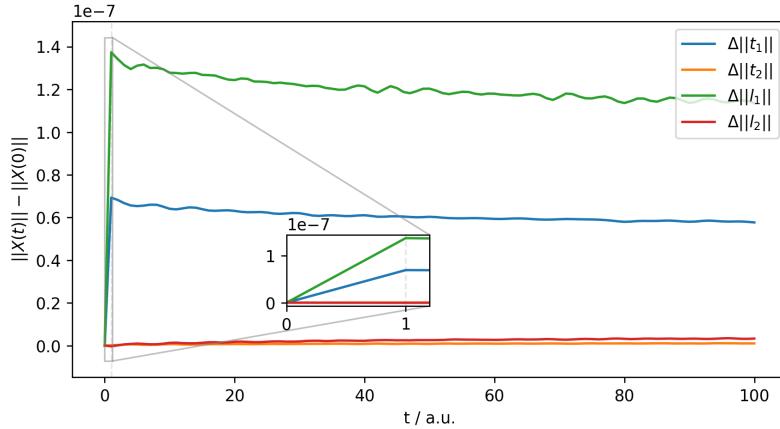


Figure C.2: Time-dependent change in the norm of the amplitude tensors relative to the ground-state amplitudes, using a PAO cut-off of 1×10^{-3} . (Amplitudes have been back-transformed to the MO space for comparison to MO-basis amplitude norms. Field and step parameters remain unchanged, and the amplitude norm is taken at every 1 a.u.)