

Electronic Theses and Dissertations, 2020-

2021

Utilizing High Throughput Computing Techniques for the Predictions of Spectroscopic Properties of Astrophysically Relevant Molecules

Brian Ferrari
University of Central Florida

 Part of the [Astrophysics and Astronomy Commons](#), and the [Chemistry Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd2020>

University of Central Florida Libraries <http://library.ucf.edu>

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2020- by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Ferrari, Brian, "Utilizing High Throughput Computing Techniques for the Predictions of Spectroscopic Properties of Astrophysically Relevant Molecules" (2021). *Electronic Theses and Dissertations, 2020-*. 862.

<https://stars.library.ucf.edu/etd2020/862>

**UTILIZING HIGH THROUGHPUT COMPUTING
TECHNIQUES FOR THE PREDICTIONS OF
SPECTROSCOPIC PROPERTIES OF ASTROPHYSICALLY
RELEVANT MOLECULES**

by

BRIAN CARMONA FERRARI
B.S. University of Central Florida, 2018

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Physics
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2021

Major Professor: Christopher J. Bennett

ABSTRACT

Here, we utilize Quantum Chemistry (QC) approaches to predict the structures, vibrational frequencies, infrared intensities and Raman activities of unusual molecular species using the General Atomic and Molecular Structure System (GAMESS(US)) package. A Python-based software, AutoGAMESS, was developed to automate the workflow and take advantage of High Throughput Computing (HTC) techniques enabling the automated generation of spectroscopic data from hundreds of calculations. This approach was utilized to determine these properties for a series of carbon oxides (C_2O_n ; $n = 3$ to 4), anticipated to be produced during the radiation of pure carbon dioxide ices, under conditions relevant to the interstellar medium. Beyond generating predicted spectroscopic results, we additionally performed a benchmark study of 70 different basis sets across multiple levels of theory (including Density Functional Theory, Møller–Plesset, and Coupled Cluster calculations), in QC to identify the method with the best balance between obtaining the lowest error in predictions while being mindful of the computation resources required.

*Dedicated to my mother, Leda P. Carmona,
for her unbelievably valuable support.*

ACKNOWLEDGMENTS

I would first and foremost like to thank Katie Slavicinska for always supporting, encouraging and helping me with my academic career.

I am enormously grateful for all the love and encouragement I have received from the *Friends Family*, both from the *Tias & Tios* as well as from the *kids*.

I would like to thank Dr. Efthimiou for mentoring me during the second half of my undergraduate career, without his mentorship I am certain I would not be where I am today.

Last but most certainly not least, I would like to extend a massive thanks to Esperanza Soto whose impeccable organizational skills and meticulousness kept me ahead of the bureaucracy of the department.

I should also acknowledge the NASA Florida Space Grant Consortium (FSGC) for funding through the NASA FSGC Dissertation Improvement Fellowship (DIF). I also acknowledge the funding received from NASA Minority University Research and Education Project(MUREP) Fellowship.

TABLE OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES	ix
I INTRODUCTION	1
1. BACKGROUND INFORMATION	3
1.1 Cometary/Interstellar Ice Chemistry	5
1.2 Observational Studies of Cometary/Interstellar Ices	6
1.3 Laboratory Simulations of Cometary/Interstellar Ice Chemistry	8
1.4 The Potential Role of Higher Order Carbon Oxides	9
2. MOTIVATION	12
2.1 Importance of Raman Activities	12
2.1.1 Overlapping/Inactive Infrared Bands	12
2.2 Importance of High Throughput Computing	15
2.2.1 Benchmarking Quantum Chemistry Methods	16
3. THESIS SCOPE	18
II RESULTS	20
4. AUTOGAMESS: A PYTHON PACKAGE FOR AUTOMATION OF GAMESS(US) RAMAN CALCULATIONS	21
4.1 Summary	21
4.2 Capabilities	23
4.3 Use of AutoGAMESS	26
4.4 Availability	26
4.5 Dependencies	26
5. A COMPARISON OF MEDIUM-SIZED BASIS SETS FOR THE PREDICTION OF GEOMETRIES, VIBRATIONAL FREQUENCIES, INFRARED INTENSITIES AND RAMAN ACTIVITIES FOR WATER	27
5.1 Introduction	27

5.2	Methods	30
5.3	Results and Discussion	31
5.3.1	Calculations Removed from Study	36
5.3.2	Molecular Geometries	37
5.3.3	Vibrational Frequencies	38
5.3.4	Infrared Intensities	40
5.3.5	Raman Activities	41
5.4	Conclusion	42
6.	A COMPUTATIONAL INVESTIGATION OF THE EQUILIBRIUM GEOMETRIES, ENERGETICS, VIBRATIONAL FREQUENCIES, INFRARED INTENSITIES AND RAMAN ACTIVITIES OF C_2O_y ($y = 3, 4$) SPECIES	44
6.1	Introduction	44
6.1.1	C_2O_3	46
6.1.2	C_2O_4	48
6.2	Computational Details	49
6.3	Results and Discussion	53
6.3.1	C_2O_3a (C_{2v}) [CO + CO ₂ Cyclic Co-Oligomer]	55
6.3.2	C_2O_3b (C_{2v}) [Isomer]	57
6.3.3	C_2O_4a (C_{2v}) [CO ₂ Dimer]	58
6.3.4	C_2O_4b (C_{2v}) [Isomer]	61
6.3.5	C_2O_4c (D_{2h}) [CO ₂ Cyclic Dimer]	61
6.3.6	C_2O_4d (D_{2h}) [CO ₂ Dimer]	64
6.3.7	Energetics	67
6.3.8	Evaluation of Theoretical Approaches	70
6.3.9	Summary of Results	72
6.4	Conclusions	74
III	CONCLUDING REMARKS	77
7.	CONCLUSION AND OUTLOOKS	78
7.1	Software Development	78
7.1.1	Major Conclusions	78
7.1.2	Future Directions	79
7.2	Basis Set Benchmarking Study	80
7.2.1	Major Conclusions	80
7.2.2	Future Directions	80
7.3	Carbon Oxides Study	81
7.3.1	Major Conclusions	81
7.3.2	Future Directions	82
APPENDIX A.	QUANTUM CHEMISTRY	84
A.1	Molecular Orbital Theory	85
A.2	Basis Set Expansion	87

A.2.1	Slater and Gaussian Type Atomic Orbitals	89
A.2.2	Pople Style Basis Sets	90
A.2.3	Atomic Natural Orbital Basis Sets	90
A.2.4	Correlation Consistent Basis Sets	91
A.2.5	Polarization Consistent Basis Sets	92
A.3	Hartree-Fock Theory	92
A.3.1	Closed-Shell Systems	92
A.3.2	Open-Shell Systems	94
A.4	Electron Correlation Methods	95
A.4.1	Excited Slater Determinants	96
A.4.2	Møller-Plesset Perturbation Theory	97
A.4.3	Couple Cluster Theory	100
A.5	Density Functional Theory	103
A.5.1	Kohn-Sham Theory	103
A.5.2	Exchange-Correlation Functionals	104
APPENDIX B. MOLECULAR VIBRATIONS		109
B.1	Normal Coordinates	110
B.2	Selection Rules in Wave Mechanics	111
B.2.1	Infrared Intensities	112
B.2.2	Raman Activities	113
APPENDIX C. DIAGNOSTICS		115
C.1	Coupled Cluster Diagnostics	116
C.1.1	%TAE[(T)]	116
C.1.2	T ₁ and T ₂ Diagnostics	117
C.2	DFT Diagnostics	117
C.2.1	A _λ	117

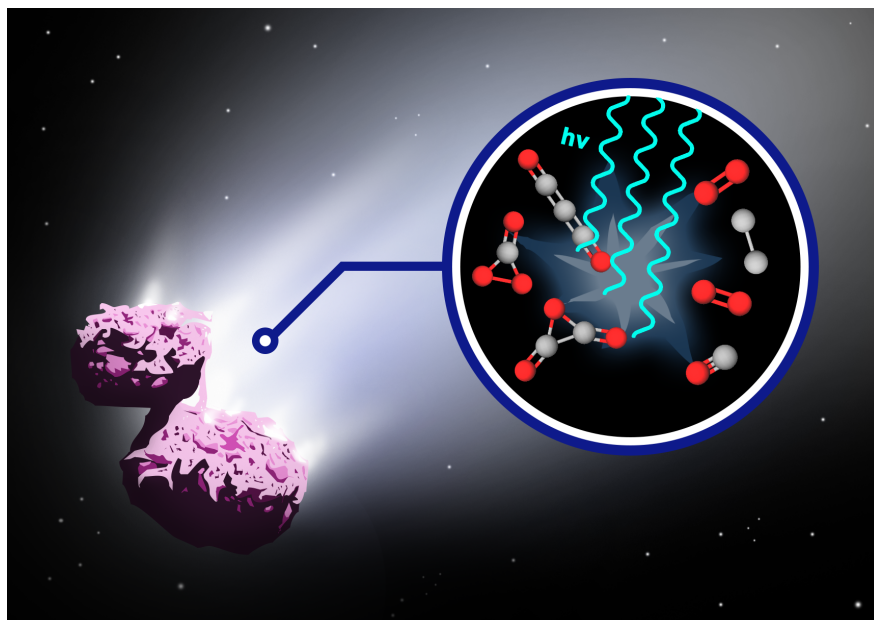
LIST OF FIGURES

1.1	Radiative processing of CH ₄ , NH ₃ and CO ₂ ice mixture to form glycine. . .	6
2.1	Comparison of the fundamental frequencies along with their infrared intensities (km mol ⁻¹) and Raman activities (Å ⁴ amu ⁻¹) for carbon oxides CO ₃ (C _{2v}), CO ₃ (D _{3h}), CO ₄ (C _{2v}), CO ₅ (C ₂), as well as C ₂ O ₃ (C _{2v}), C ₂ O ₄ (C _{2v}), and C ₂ O ₄ (D _{2h}) predicted with the B3LYP/aug-cc-pVQZ level of theory under the harmonic approximation in vacuo.	13
2.2	A crude plot of the relationship between the level of theory and basis set selected for quantum chemical calculations.	15
4.1	Line Plots Example	25
5.1	Unsigned Error of CCSD(T) Bond Angle Calculations Compared to Experimental Data	37
5.2	Mean Unsigned Error of CCSD(T) Frequency Calculations Compared to Experimental Data	38
5.3	Mean Unsigned Error of B3LYP Infrared Intensity Calculations Compared to Experimental Data	40
5.4	Mean Unsigned Error of MP2 Raman Activities Calculations Compared to Experimental Data	41
6.1	Optimized geometries of all reported molecules. Bond angles and bond lengths reported here are calculated at the CCSD(2) _T /cc-pVTZ level of theory. . . .	54
6.2	Plots showing how the B3LYP level of theory compares to CCSD(2) _T /cc-pVTZ in the prediction of molecular geometries (top left) and harmonic frequencies (top right) and to SCS-MP2/aug-cc-pVQZ for infrared intensities (bottom left) and Raman activities (bottom right). T-test, p-test, and r ² values are indicated in each plot. <i>Note: ACCQ stands for aug-cc-pVQZ.</i> . .	72
6.3	A Kernel Density Estimate (KDE) plot of the difference in predicted bond lengths (top) and harmonic vibrational frequencies (bottom) between each individual theory versus those predicted by CCSD(2) _T /cc-pVTZ. <i>Note: ACCQ stands for Aug-cc-pVQZ.</i>	74
A.1	Example electron configuration diagram	86
A.2	An example of a doublet state electronic configuration in RHF theory.	95
A.3	An example of a doublet state electronic configuration in UHF theory.	95
A.4	An example plot of the oscillatory behavior of MP theory as a function of n.	100

LIST OF TABLES

1.1	Summary of results of several experiments that produced higher order carbon oxides.	11
2.1	Various levels of theory used to calculate the frequencies and intensities for CO ₃ (D _{3h}) with the cc-pVTZ basis set (unless otherwise stated); intensities in parentheses are in km mol ⁻¹	14
5.1	Molecular Geometry Predictions, Bond Lengths are in Angstroms (Å), Bond Angles are in Degrees (°), {1} is fastest time rank	31
5.2	Harmonic Vibrational Frequencies in Wavenumbers (cm⁻¹)	32
5.3	Infrared Intensities in Km · mol⁻¹ and Raman Activities in Å⁴amu⁻¹	34
5.4	VSCF Calculations for Anharmonic Frequencies in cm⁻¹ and Infrared Intensities in Km · mol⁻¹	36
6.1	%TAE(T), T1 diagnostic, Largest T2 amplitudes, and several A _λ DFT-based diagnostics for the molecules of interest here as well as several well-studied species that exhibit varying degrees of multireference character (see text for details).	54
6.2	Harmonic Frequencies, Anharmonic Frequencies, Infrared Intensities, Raman Activities and Isotopic Shifts for C ₂ O ₃ a (C _{2v}).	57
6.3	Harmonic Frequencies, Anharmonic Frequencies, Infrared Intensities, Raman Activities and Isotopic Shifts for C ₂ O ₄ a (C _{2v}).	60
6.4	Harmonic Frequencies, Anharmonic Frequencies, Infrared Intensities, Raman Activities and Isotopic Shifts for C ₂ O ₄ c (D _{2h}).	63
6.5	Harmonic Frequencies, Anharmonic Frequencies, Infrared Intensities, Raman Activities and Isotopic Shifts for C ₂ O ₄ d (D _{2h}).	66
6.6	Heats of Formation (kcal/mol)	67
6.7	Energy Differences for Bond Dissociation Reactions (kcal/mol), Detonation Velocities (km/s) and Detonation Pressures (GPa)	69
6.8	Comparison of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Signed Error (MSE) and Mean Signed Percentage Error (MSPE) for Bond Lengths (in Å), Vibrational Frequencies (in cm ⁻¹), Infrared Intensities (in km mol ⁻¹), and Raman Activities (in Å ⁴ amu ⁻¹).	71

INTRODUCTION



“If you can think of a molecule, there is a finite probability that over the immensity of space and eons of time it has existed somewhere at some time.”

-Ryan Fortenberry

This part (Part I: Introduction) contains previously published content. The usage of this content is approved by the journal's copyright policy, which states authors can include their publications in full or in part in a thesis or dissertation.

REF: Brian C Ferrari et al. “Role of Suprathermal Chemistry on the Evolution of Carbon Oxides and Organics within Interstellar and Cometary Ices”. In: *Accounts of Chemical Research* 54.5 (2021), pp. 1067–1079

1 BACKGROUND INFORMATION

Infrared telescopes became increasingly capable during the 60s and 70s. Subsequently, water ice was detected in the direction of the Orion constellation in 1973 [2]. More recent surveys of low- and high-mass protostars show that H₂O is unanimously the most abundant species identified, with the following volatile molecules found in decreasing abundances (relative to water, rtw): carbon monoxide (CO; 7-61%), carbon dioxide (CO₂; 7-35%), ammonia (NH₃; 4-17%), methanol (CH₃OH; 5-17.7%), and methane (CH₄; 2-7%), with a few other species (e.g., OCN⁻) at ~1-3.5% abundance [3]. Carbon dust, along with silicate grains, serve as the solid cores on which these volatile molecules in interstellar clouds condense. These icy mantels are expected to have a layered structure [4], this is because different volatiles condense under different conditions. For instance, CO forms in the gas phase and condenses on already existing H₂O ice mantles [5]. The layers of condensed volatiles surrounding these carbon/silicate grains are commonly referred to as interstellar ices.

Over the lifetime of a typical interstellar cloud, interstellar ices are subject to radiation from UV photons as well as galactic cosmic rays (GCRs). The GCRs are comprised of 87% protons, 12% helium nuclei, and 1% heavier nuclei and often bear high energies (>GeV/nucleon), where an estimated flux of 10 cm⁻² s⁻¹ for 1-10 MeV protons is typically adopted [6]. An external ultraviolet (UV) photon field also exists, predominantly composed of Lyman- α (10.2 eV) photons with a flux of around 10⁸ cm⁻² s⁻¹, but this is unable to penetrate dense clouds, which instead have an internally induced UV field generated by the GCR that holds a lower flux ~10³ cm⁻² s⁻¹ [6]. The chemical composition of these ices evolves over time as they are exposed to UV and GCRs (this is discussed in more detail in Section 1.1).

A key question is how this radiation alters the composition of interstellar ices. Moll *et al.* studied the irradiation of solid CO₂ held at 77 K with vacuum-ultraviolet (VUV) photons [7]. They assigned eight infrared absorptions to a novel carbon trioxide species (CO₃) and, based upon observed isotopic shifts, deduced a structure consistent with C_{2v} symmetry. This finding represents one of the first efforts demonstrating the production of novel molecular species during the irradiation of a pure CO₂ ice. More recent studies have shown that interstellar ices exposed to ionizing radiation can produce a wide variety of complex organic molecules (COMs) [8]. It has also been shown that astrophysical ices exposed to ionizing radiation often form an organic rich refractory residue with similar spectral features to those of the ISM [9, 10]. The gravitational collapse of an interstellar cloud can eventually lead to the formation of a planetary system, where the COMs from the interstellar ices and refractory residues can be incorporated into asteroids, comets, or planetesimals. It is unlikely that any COMs would have survived the harsh conditions of planetary formation; however, COMs incorporated into asteroids or comets in the outskirts of the collapsing cloud may have survived [11].

Currently, a large part of the scientific community agrees that radiation induced chemistries within interstellar ices play a major role in the enriching of the chemical inventory of the early Solar System [12, 13]. The early Earth is thought to have been hit by a large number of comets and asteroids, which could have brought extraterrestrial organics to the surface of the primordial Earth. As a result, a more diverse chemistry may have been available to facilitate the origin of life on Earth. This makes understanding the chemical dynamics within interstellar ices essential to understanding how life originated on Earth. However, interstellar ices can only be studied by observational astronomy which is limited to only spectroscopy-based techniques. Comets, which are formed from interstellar ices, can be studied with vast suites of instruments onboard a spacecraft to supplement observational studies. For instance, the European Space Agency's ROSETTA spacecraft detected the amino acid (building blocks of proteins) glycine with a mass spectrometer (discussed in

more detail in Section 1.2). Unfortunately, spacecraft-based missions tasked with studying cometary ices are expensive, thus making them rare. This results in the majority of our understanding of this topic coming from either by laboratory simulations (Section 1.3), or observational astronomy (Section 1.2). Equally important is understanding the chemistries that these ices undergo, and how those chemistries can result in COMs, such as glycine.

1.1 Cometary/Interstellar Ice Chemistry

The extremely low temperatures of the ISM make it highly inefficient for chemical reactions to occur via traditional means, thus making radiation chemistry one of the more efficient methods of inducing chemical reactions. As a primary source of ionizing radiation (energetic particles or photons) traverses through matter, it produces secondary sources of ionizing radiation (secondary electrons). The cascade of secondary electrons is the most abundant species formed along a radiation track in condensed matter [14]. Pimblott *et al.* showed that the mean energy of secondary electrons from ions was 50-60eV while for electrons it was roughly 9eV [15]. At such an electron energy range, the predominant interaction with molecules is no longer ionization, but homolytic bond dissociation [16]. These bond dissociation interactions result in radical species, which subsequently recombine with other radicals in what are thought to be barrier-less radical-radical recombination reactions [1]. As a result, strange meta-stable molecules which under standard conditions would quickly dissociate can be formed and preserved by the low temperatures in the ISM.

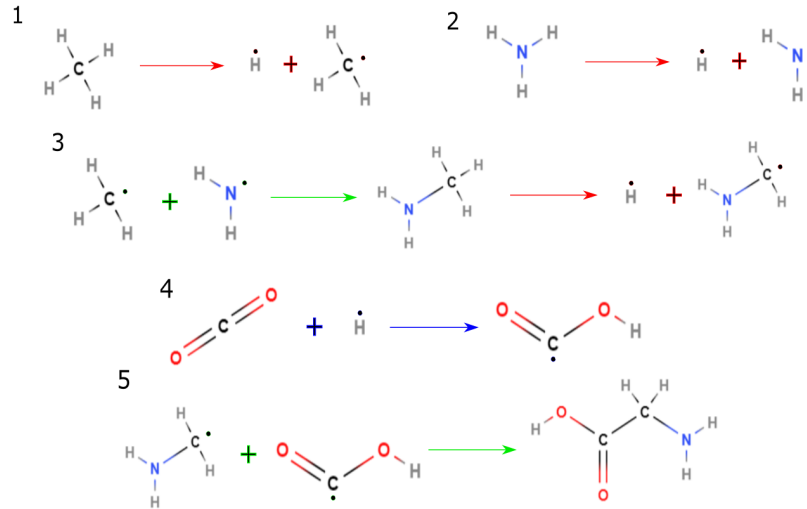


Figure 1.1: Radiative processing of CH_4 , NH_3 and CO_2 ice mixture to form glycine.

Figure 1.1 demonstrates how radiation chemistry can form complex organic molecules, such as amino acids. The reactions with red arrows are radiation-induced bond cleavages that result in radical species and suprathermal hydrogen atoms (atoms with excess kinetic energy compared to its surroundings). The reactions with green arrows are radical-radical recombination reactions that don't have an energy barrier, whereas the reaction with a blue arrow (reaction 4) is a radical-radical recombination reaction with an energy barrier of ≈ 1.1 eV [17]. This energy barrier is not a constraint since the suprathermal hydrogen atoms produced during the bond cleavages have sufficient energy to overcome it [1]. This implies that ice mixtures of CH_4 , NH_3 and CO_2 such as those on interstellar ices are able to form glycine through abiotic processes such as radiation chemistry. As such, understanding these chemistries is essential to understanding the chemical evolution that may have resulted in life on Earth. However, in the pursuit of untangling the chemical dynamics we are limited to studying these processes via observational astronomy and laboratory simulations.

1.2 Observational Studies of Cometary/Interstellar Ices

Gas phase molecules can be detected via their rotational transitions (microwave and radio wavelengths), or through their fundamental vibrational modes (infrared wavelengths). Inter-

stellar ices can also be detected through their fundamental vibrational modes [18], although this can be difficult due to overlapping modes. When telescopes collect spectra of the ISM, they do so by collecting incoming photons from a certain direction. Typically these observations are done by first collecting a spectrum of a proto-star or star (“black body” spectrum), then subtracting it from the ISM spectrum to obtain absorption bands of the matter in between the origin source and the telescope [19]. The wavelengths absorbed are dependent on the molecular species present in the matter. Absorption bands can then be identified, either by referencing experimental studies or theoretical calculations of infrared activities. These same methods can be applied to studying cometary ices, however; due to their large size, comet ice spectra cannot be observed directly. Instead, spectra of the gaseous species within the comet’s coma are collected [20]. Although this method provides some insight into the composition of comet ices, it cannot discern if observed species were present in the ice or a fragment of some less stable molecule originally within the comet ice. Another issue with this method is that the most well preserved interstellar ices are within the cometary nuclei [20]. This is a result of the exterior shell of the comet being exposed to the harsh environment (radiation, impact, etc.), which induces chemical changes within the ice. As such, the comet coma is predominantly comprised of altered material and only gives us a partial understanding of interstellar ice compositions. This makes direct links between a comet coma composition and interstellar ices more challenging. However, studying both comet comas and interstellar ices gives us a broader understanding of the chemical evolution. Our understanding of cometary composition and evolution has undergone several changes over the past decade, thanks in part to the European Space Agency’s *Rosetta* spacecraft which made detailed *in-situ* analyses of comet 67P/Churyumov–Gerasimenko (67P). The double-focus magnetic mass spectrometer (DFMS) had sufficient resolution that it could unambiguously identify the simplest amino acid, glycine [21]. The DFMS was also able to detect molecular oxygen (O₂), a hypervolatile species not anticipated to be present, let alone in considerable abundance, at comet 67P [22]. This detection has brought forth a

tremendous amount of discourse within the scientific community (this is discussed in more detail in Section 1.4). Deciphering the origin of the molecular oxygen would greatly improve our understanding of the chemical evolution of our Solar System.

1.3 Laboratory Simulations of Cometary/Interstellar Ice Chemistry

Experimental set-ups are tasked with accurately simulating the conditions of space, where extremely low pressures, temperatures, and a harsh radiation environment can persist, along with a means of both introducing and analyzing chemical species. Several recent reviews have provided extensive coverage of experimental approaches [23, 24, 25].

Various analytical techniques for identifying irradiation products both *in situ* and *ex situ* have been applied to studying ices subject to irradiation [24]. FTIR spectroscopy remains one of the most popular *in situ* techniques in laboratory astrochemistry of analyzing the ices during the radiation process and provides a non-destructive, highly sensitive (sub-monolayer detection), and highly quantitative means to determine chemical species from their quantized vibrational modes and, in transmission, is directly comparable to remote observations of interstellar ices. Raman instrumentation is more difficult to implement, and quantification is not routine. Also, the cross section for Raman scattering is inherently weaker than that of infrared absorptions, and often dwarfed by a competing fluorescence signal. Bennett *et al.* recently introduced a gated-Raman set-up to overcome the latter problem [26]. A more detailed explanation of the benefit and importance of Raman spectroscopy for studying these irradiated ices is given in Chapter 2.1.

In many astrochemistry experiments, after irradiation is performed, the substrate is returned to room temperature by controlled temperature programmed desorption (TPD), where a heater within the sample holder ramps the temperature at a rate typically ranging between 0.5 K min^{-1} and $>10 \text{ K s}^{-1}$. During this process, volatile species are released into the gaseous state from the ice, which can itself undergo crystallization or other phase changes. Such processes crudely mimic the sublimation of ices surrounding a hot molecular core (HMC;

a region surrounding a young star, where the majority of >200 interstellar molecules have been detected) or as comets approach perihelion [23]. Gas-phase species can be detected *in situ* by mass spectrometry, which is typically (although this depends on the equipment being used) a few orders of magnitude more sensitive than FTIR spectroscopy.

The general methodology for identifying novel species within laboratory simulations is to reference QC calculations to make tentative assignments within the FTIR spectra. Subsequent mass spectrometry during TPD can then corroborate the tentative assignments to give a more conclusive identification of the novel species. However, this method is not effective at identifying meta-stable molecules, since they are likely to dissociate during TPD. Recently, astrochemists have begun using amorphous solid phase *ab initio* molecular dynamics (AIMD) calculations to produce more reliable infrared intensities, in an attempt to circumnavigate this issue [27]. These calculations provide more accurate solid phase frequencies, however, they come with an increased computational cost. This is talked about more in Chapter 2.2, with an emphasis on how benchmarking studies for molecules in vacuo can help increase accuracy while reducing computational cost.

1.4 The Potential Role of Higher Order Carbon Oxides

Comet 67P has sparked a lot of interest recently, this is due to the detection of hypervolatiles O_2 [22] and N_2 [28]. The N_2 is detected at very low levels and its codetection of O_2 at high levels is difficult to explain since these molecules have very similar physical properties. The current consensus of the scientific community is that the O_2 detected at comet 67P had a primordial (interstellar) origin [29]. This is consistent with several recent models that have demonstrated significant levels of interstellar ices can survive Solar System formation and are incorporated into comets [30]. Several arguments have been made for the source of this codetection, some have suggested the decomposition of oxalic anhydride (C_2O_3) is responsible [31]. Fortenberry *et al.* suggest that the source of the oxalic anhydride be from irradiation of CO_2 ice, which has been shown to produce several higher order carbon oxide

species [32]. Bennett *et al.* showed that carbon suboxide (C_3O_2) and C_6O produced within CO-irradiated ices served as less volatile species which could produce distributed sources of carbon monoxide (CO) [33].

Similarly, Huntress suggested in 1991 that carbon suboxide (C_3O_2) may explain the extended/distributed source of CO observations within comet Halley's coma by acting as a parent species that, upon photodissociation, gave rise to the observed CO (and C_2O) daughter fragment(s) [34, 35]. However, the claim was quickly refuted as infrared spectra of comet Halley taken by the VEGA 1 probe placed upper limits on the levels of C_3O_2 below those required to explain the observations [36]. Alternatively, formaldehyde (H_2CO , or related polymers) could also act as a parent species to release extended CO [35]. In fact, the radiolysis of just pure CO ices alone gives rise to a plethora of additional candidate parent species that could contribute to CO (and C_2O) formation [37].

As an example, let us consider two of the carbon oxide species, C_3O_2 and C_6O , that are formed when pure CO ices are subjected to electron irradiation [37]. When the ices are subsequently warmed during TPD, these species sublime at ~ 97 and ~ 120 K, respectively (compared to CO, which itself sublimates around 25-30 K) [38]. Bennett *et al.* found that during the TPD of irradiated CO ice besides observation of the parent ions ($C_3O_2^+$ and C_6O^+), they also see fragment daughter ions (C_2O^+ and C_5^+) which are formed during the electron-impact while generating neutral CO [33].

Though the origin of O_2 in comet 67P's coma is still under discussion, decomposition of higher order carbon oxides could yield CO and O_2 . For instance, CO_3 could dissociate into $CO + O_2$ [39], while C_2O_4 may dissociate into $2CO + O_2$. Thus, the majority of higher order carbon oxide species could contribute to the O_2 abundance. The possibility of these molecules existing and remaining stable within comet ices requires more investigation. However, many higher order carbon oxides have been observed in irradiated ices CO_2 rich ices. Table 1.1 gives a summary of the higher order carbon oxides detected in experimental studies.

Table 1.1: Summary of results of several experiments that produced higher order carbon oxides.

Ice Mixture (ratio)	Detection Methods	Products Observed	Ref
CO ₂	FTIR & QMS	O ₃ , CO, CO ₃ (C _{2v})	[32]
CO ₂	FTIR	CO ₃ (D _{3h})	[40]
CO ₂	FTIR	O ₃ , CO, CO ₃ (C _{2v}), CO ₃ (D _{3h}), CO ₄ (C _{2v})	[41]
CO ₂	FTIR	O ₃ , CO, CO ₃ (C _{2v}), CO ₃ (D _{3h}), CO ₄ (C _{2v}), CO ₅ (C ₂)	[42]
CO ₂	FTIR	O ₃ , CO, CO ₃ (C _{2v}), CO ₃ (D _{3h}), CO ₄ (C _{2v}), CO ₅ (C ₂), CO ₆ (C _s)	[43]
CO ₂	FTIR & Raman	O ₂ , O ₃ , CO, CO ₃ (C _{2v}), CO ₄ (C _{2v}), CO ₅ (C ₂)	[26]
CO ₂	FTIR	O ₃ , CO, CO ₃ (C _{2v})	[44]
CO ₂	FTIR & QMS	O ₂ , O ₃ , CO, CO ₃ (C _{2v})	[45]
CO ₂	FTIR & QMS	O ₃ , CO, CO ₃ (C _{2v})	[46]
CO ₂	FTIR & QMS	O ₃ , D ₂ O, D ₂ CO ₃ , D ₂ CO, CD ₃ OD, DCOOD, CO ₃ (C _{2v}), CO ₃ (D _{3h}), CO ₄ (C _{2v}), CO ₅ (C ₂), CO ₆ (C _s)	[47]
CO ₂	FTIR & QMS	O ₃ , CO, CO ₃ (C _{2v})	[46]
CO ₂ :O ₂ (~ 1 : 4)	FTIR & QMS	O ₃ , CO, CO ₃ (C _{2v}), CO ₃ (D _{3h}), CO ₄ (C _{2v}), CO ₅ (C ₂), CO ₆ (C _s)	[48]
CO ₂ :CO (~ 1 : 1)	FTIR & QMS	O ₂ , O ₃ , C ₂ O, C ₃ O ₂ , CO ₃ (C _{2v}), CO ₃ (D _{3h}), CO ₄ (C _{2v}), CO ₅ (C ₂), CO ₆ (C _s)	[48]
CO: ¹⁸ O ₂ (~ 4 : 1)	FTIR & QMS	O ₃ , CO, CO ₃ (C _{2v}), CO ₃ (D _{3h}), CO ₄ (C _{2v}), CO ₅ (C ₂), CO ₆ (C _s)	[49]

Considering how abundant carbon dioxide is in the ISM, it is plausible these higher order carbon oxides formed within interstellar ices. Despite the large number of detected species, there still remains to be any species with more than one carbon atom detected. The formation of carbon suboxide (C₃O₂) and C₆O within irradiated CO ices indicates that higher order carbon oxides with multiple carbon atoms are also likely to be produced. One possibility for why these species have not yet been detected is that their IR bands are being masked by other carbon oxides (this is discussed in more detail in Section 2.1). To circumvent this issue, more comprehensive studies utilizing a combination of FTIR and Raman spectroscopy would need to be done. However, there is no existing literature on the Raman activities of higher order carbon oxides. The work reported in this thesis fills this void, and Chapter 2 explains in more detail the relevance it.

2 MOTIVATION

2.1 Importance of Raman Activities

Few experimental astrochemistry groups utilize Raman spectroscopy to probe irradiated ices. This is partly due to the inability to directly compare results with observational data, but also due to the difficulty of collecting Raman spectra of irradiated ices. The cross section for Raman scattering is inherently weaker than that of infrared absorptions, and often dwarfed by a competing fluorescence signal. Despite these difficulties, certain groups have utilized Raman spectroscopy to study irradiated ices [50, 51]. Recently, Bennett *et al.* introduced a gated-Raman set-up to overcome the fluorescence issue [26]. Since then more groups have begun incorporating Raman spectroscopy within their laboratory studies [52, 53].

One of the key reasons why astrochemists have been eager to adopt Raman spectroscopy is that it gives information on the structural properties of the sample. Beyond structural information, Raman spectroscopy can also be used to probe infrared inactive molecular species or resolve species with overlapping IR bands (this is discussed in more detail in Section 2.1.1). As such, using Raman spectroscopy in conjunction with IR spectroscopy gives a more complete picture of the molecular properties of the sample.

2.1.1 Overlapping/Inactive Infrared Bands

Raman spectroscopy is complementary to FTIR, where the same vibrational modes are probed, but it is dictated by different selection rules (change in polarizability for Raman *vs.* change in dipole moment for infrared spectroscopy). The strongest observable functional

groups in Raman spectroscopy are those that are more covalent in character (e.g., C=C), whereas in FTIR it is bonds that are more ionic (e.g., O-H, C=O, N-H). Furthermore, homonuclear diatomic molecules (such as O₂, N₂, etc.) are uniquely observable with Raman spectroscopy [26].

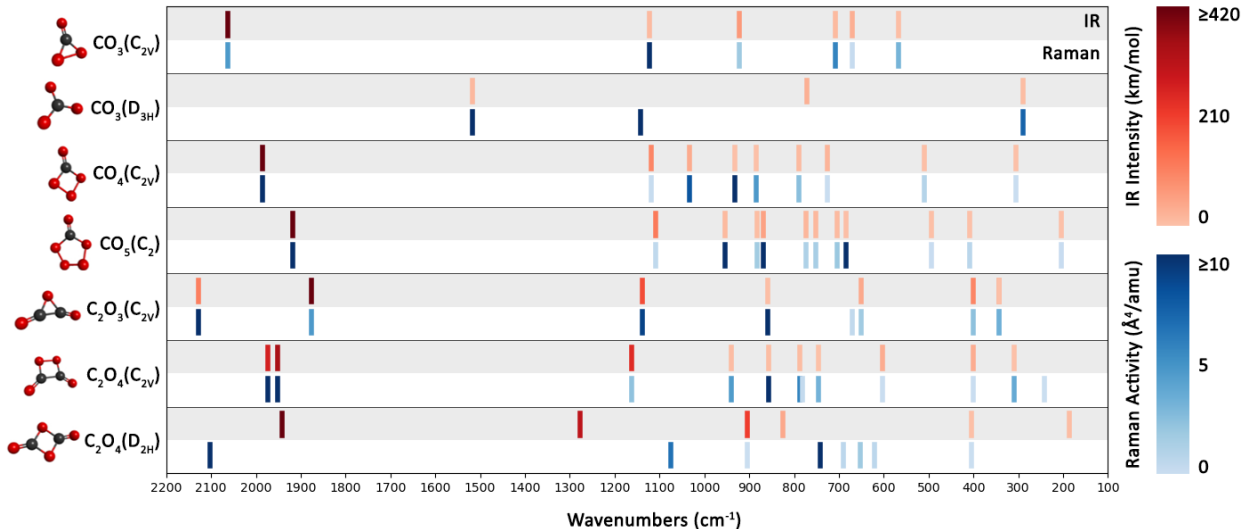


Figure 2.1: Comparison of the fundamental frequencies along with their infrared intensities (km mol⁻¹) and Raman activities (Å⁴ amu⁻¹) for carbon oxides CO₃(C_{2v}), CO₃(D_{3h}), CO₄(C_{2v}), CO₅(C₂), as well as C₂O₃(C_{2v}), C₂O₄(C_{2v}), and C₂O₄(D_{2h}) predicted with the B3LYP/aug-cc-pVQZ level of theory under the harmonic approximation in vacuo.

Figure 2.1 reveals the importance of Raman spectroscopy in the identification of novel higher order carbon oxides. Note that CO₃(D_{3h}) and many of the other carbon oxides of the form CO_x (x = 3-6) only have one relatively strong infrared absorption, while there are several strong Raman absorptions that could be used to confirm their assignments. Furthermore, while the C₂O_y (y = 3,4) species have overlapping IR absorptions particularly in the 1900-2000 cm⁻¹ with those from CO_x (x = 3-6) species, their Raman signatures are sufficiently unique that they could be distinguished. An example of the importance of utilizing Raman spectroscopy to identify higher order carbon oxides is the contested identification of CO₃(D_{3h}) isomer.

The CO₃(D_{3h}) isomer was identified as a product formed during radiolysis of CO₂ ices

by Jamieson *et al.* [40]. A single absorption at 1165 cm^{-1} in the IR spectrum was assigned to the degenerate ν_1/ν_2 fundamental band of this species, based upon *ab initio* calculations which predicted an absorption at 1168 cm^{-1} . Additional experiments were performed using $^{13}\text{CO}_2$, C^{18}O_2 , and $^{13}\text{C}^{18}\text{O}_2$ isotopes, and the corresponding shift of the ν_1/ν_2 band was found to be consistent with this assignment.

However, Kowalczyk & Krylov disputed that the $\text{D}_{3\text{h}}$ isomer was a true minimum on the CO_3 potential energy surface (PES) [54]. Arguments have been made that both studies utilized inadequate levels of theory to correctly describe this species [1]. Table 2.1 shows the results from these calculations alongside other existing results found in literature. While these calculations demonstrate that the $\text{D}_{3\text{h}}$ isomer is likely a real species, the frequencies for this absorption are ambiguous enough to cast doubt on the assignment of the carrier at 1165 cm^{-1} to the $\text{D}_{3\text{h}}$ isomer. The ambiguity of the assignment is demonstrated in Table 2.1, where the row for the band used to identify the isomer is highlighted in grey, and cells highlighted in red indicate the structure is unstable.

Table 2.1: Various levels of theory used to calculate the frequencies and intensities for $\text{CO}_3(\text{D}_{3\text{h}})$ with the cc-pVTZ basis set (unless otherwise stated); intensities in parentheses are in km mol^{-1} .

Mode/Symmetry	Level of Theory						
	CASSCF ^{a,b}	MP2 ^c	CCSD(T) ^c	BCCD(T) ^d	MRCI+Q ^d	MRMP2 ^{a,e}	CCSD(2) _T ^e
Bend/ e'	447 (2.3)	672	-402	455	460	437	321
OPLA/ a_2''	763 (31.5)	746	776	762	769	732	761
CO stretch/ e'	1204 (72.3)	3501	1093	1316	1647	1432	1267
CO stretch/ a_1'	1089 (0)	1022	1105	1094	1124	1065	1110

^a Calculation uses 6-311G(d) basis set.

^b Calculations from Ref. [32].

^c Calculations from Ref. [54].

^d Calculations from Ref. [55].

^e Calculations from Ref. [1].

The validity of the $\text{D}_{3\text{h}}$ assignment is an area of current ongoing research, in particular investigating how different levels of theory handle dynamic correlation in an effort to better

characterize the D_{3h} isomer. It's important to note that Raman spectroscopy could provide a means to unambiguously detect this species, or identify the true source of the 1165 cm^{-1} absorption. The work reported in this thesis supplements these two points, in particular by reporting Raman activities for other molecular species which could be the source of the absorption and providing software tools to aid in diagnostic calculations (this is discussed in more detail in Section 2.2).

2.2 Importance of High Throughput Computing

Spectroscopic property predictions are highly dependent on the level of theory (ie: Density Functional Theory, Møller-Plesset perturbation theory, ...) and the basis set used in the calculation (see Figure 2.2)¹. Note, that although the ideal calculations would be done with the largest possible basis set and full Configuration Interaction (CI), it would be extremely computationally demanding. As such, chemists need to play a “balancing act”, where sacrificing accuracy in the calculation reduces the computational demand. Since computational resources are often constrained, smaller Pople-style basis sets are often employed since they are often not demanding.

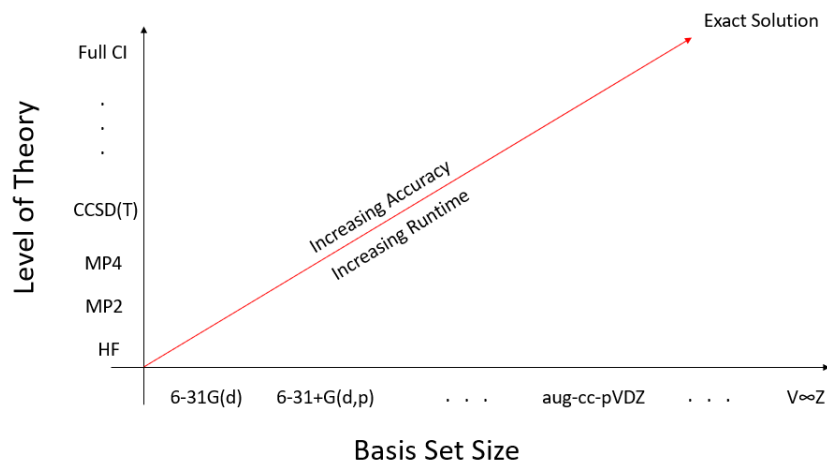


Figure 2.2: A crude plot of the relationship between the level of theory and basis set selected for quantum chemical calculations.

¹ A more detailed discussion of Quantum Chemistry methods can be found in Appendix Chapter A

The rise of massive computer clusters has brought forth the ability for High Throughput Computing (HTC). HTC is the strategic use of many computing resources over a period of time to complete the largest number of jobs possible. Utilizing HTC strategies it is possible to generate comprehensive sets of data on molecules likely to form in irradiated interstellar ices, these large sets of data on one molecule help reduce uncertainty when assigning a detection.

2.2.1 Benchmarking Quantum Chemistry Methods

While coupled-cluster with single, double, and perturbative triple excitations [CCSD(T)] remains the “gold standard” offering exceptionally accurate results, it has a high dependency on basis set size and scales $\sim N^7$ with the basis set size (denoted N). Second-order Møller-Plesset (MP2) perturbation theory scales with N^5 , is often of high accuracy, and is particularly well-suited to the calculation of Raman activities, but is also highly dependent on basis set size. Therefore, the most popular approach adopted within the astrochemistry community has been to utilize density functional theory (DFT) methods, such as B3LYP, which only scale as N^3 . Often, DFT methods are utilized alongside Pople family of basis sets, since results typically converge readily with smaller basis sets, particularly prevalent within the astrochemistry community [32, 40, 42, 41, 43]. Of note, the vast majority of calculations are currently performed *in vacuo* and are not representative of the solid state where the matrix induces shifts in vibrational frequencies (typically $<40 \text{ cm}^{-1}$) and can influence the IR intensities significantly [56]. There are several approaches currently being considered to help improve predictions in the solid state, for instance direct simulations of the condensed phase itself [27]. However, these come at an additional computational cost and thus are not suitable for large basis sets and high levels of theory. As such, benchmarking studies provide the necessary analysis of the computational cost versus performance such that an ideal level of theory and basis set can be selected. Since the benchmarking can be done on a single molecule *in vacuo*, comparisons to higher levels of theory can be made and then the ideal

candidate can be used for the condensed phase simulation.

3 THESIS SCOPE

This thesis focuses on the development of software tools to facilitate quantum chemistry calculations, and the subsequent use of this software to calculate the spectroscopic properties of astrophysically relevant molecules. This work has filled three major voids within the community:

- (i) A lack of software tools designed to automate the process of calculating Raman activities
- (ii) A wide-ranging assessment of commonly available basis sets in their ability to accurately reproduce Raman activities
- (iii) A lack of complete spectroscopic data (IR and Raman) on astrophysically relevant higher order carbon oxides

The first void filled by this work will allow the scientific community to implement HTC techniques in the calculation of Raman activities. This will allow for more diagnostics and assessments of theoretical methods, which in turn will result in more accurate calculations. The second void filled by this work will give experimental astrochemists the ability to utilize Raman spectroscopy to detect the molecules studied.

In Chapter 4 the capabilities and summary of AutoGAMESS, a Python based software for automating calculations of Raman activities, are explained. This software was essential to completing the work in subsequent chapters, in particular, the software's ability to compile data from calculations into spreadsheets automatically allowed for increased amounts of

calculations to be done during the benchmarking done in Chapter 5. The software is open source and is under continual improvements with aims at being intuitive to all users regardless of Python ability.

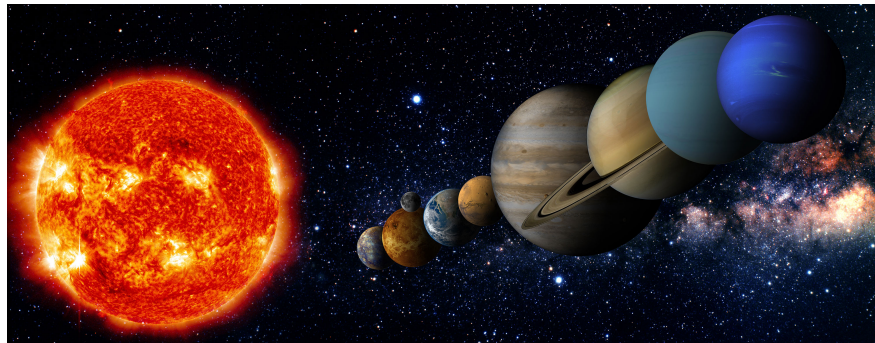
In Chapter 5 we systematically compare ~ 70 basis sets to reveal trends that will aid the selection of basis sets to be used in the study described in Chapter 6. The focus of this study was to find a basis set that would best suit Raman calculations, and the Def-2 basis family was found to be an ideal candidate due to slightly faster runtimes than those of aug-cc-pVQZ. We also noted that diffuse functions and extensive polarization functions were beneficial to the performance of Raman calculations. However, preliminary tests revealed that the quadruple zeta Def-2 basis set was not performing as well as the aug-cc-pVQZ basis set. As such, we continued the study with the augmented correlation consistent basis set, which contained the diffuse function we previously noted to be important.

In Chapter 6 calculations of the geometries, harmonic frequencies, infrared intensities and Raman activities of higher order carbon oxides were done across several levels of theory. Several diagnostic methods were used to discern the multi-reference character of the carbon oxide molecules being studied. Our results indicated moderate to mild levels, as such, we included levels of theory more capable of handling multi-reference character. Finally, we did a comprehensive assessment of the performance of the each level of theory. These assessments should aid future studies on these molecules which may not be able to take advantage of high levels of theory. For instance, calculations of the spectroscopic properties of amorphous solid phase carbon oxides would be restricted to DFT based methods, of which we outline ideal candidates.

PART

II

RESULTS



*“If I have seen farther than others, it is by standing on the shoulders of giants.”
-Sir Issac Newton*

4 AUTOGAMESS: A PYTHON PACKAGE FOR AUTOMATION OF GAMESS(US) RAMAN CALCULATIONS

This chapter (Chapter 4) contains previously published content. The usage of this content is approved by the journal's copyright policy, which states authors can include their publications in full or in part in a thesis or dissertation.

REF: Brian C Ferrari. “AutoGAMESS: A Python package for automation of GAMESS (US) Raman calculations”. In: *Journal of Open Source Software* 4.41 (2019), p. 1612

4.1 Summary

The *ab initio* quantum chemistry software GAMESS(US)[58, 59] is capable of calculating a variety of molecular properties. One of the many popular uses of GAMESS(US) is the prediction of properties of volatile and unstable species that have not been experimentally characterized or quantified before by chemists, physicists, astro-chemists and astro-physicists [60, 61, 62, 63]. Applications of, and research done using, GAMESS(US) is not limited to uncharacterized species; it's also widely used in material characterization or material property prediction research. Research utilizing these types of *ab initio* calculations typically require calculations with multiple steps required to achieve each final result. For instance, a Raman

activity prediction first requires a geometry optimization and Hessian calculation be performed on the molecule, making automation extremely beneficial. This leads to complicated and tedious workflows slowing a user’s research.

Oftentimes single calculations of molecular properties are not reliable, resulting in publications requiring several calculations, each implementing either a different level of theory or basis set, for each property be done on each molecule. This has brought about a demand for high throughput data calculation packages which automate these *ab initio* calculations [64, 65, 66]; as well as workflow management systems [67] and data parsers [68]. Specific packages have also been developed to compliment GAMESS(US) [69, 70, 71, 72, 73]. However, because these programs are largely visualization and graphical programs, there is still a need for packages that automate GAMESS(US) Raman calculations. Automation is essential to generate large databases of Raman data, which could have further applications for machine learning of Raman data. As it stands, the automation of Raman calculations is either not being done, or being implemented individually by each research group utilizing the GAMESS(US) software. This slows scientific progress down, and an automation software written in a language extremely simple and well adopted by scientists, such as Python, is an attractive solution to the problem.

AutoGAMESS provides an open source, Python-based software for automating conversion between optimization calculations to Hessian calculations and then to Raman calculations. It also offers automation of data collection from the output files, for quick tabular data readouts of each calculation. AutoGAMESS has currently been used for a study presented at the 30th Annual Conference on Computation Physics CCP2018 that has been published in the conference proceedings. AutoGAMESS is also currently being used in multiple other computational chemistry projects, soon to be published by scientists at the University of Central Florida. Finally, it has been used for a publication (reported within this thesis) studying the spectroscopic properties of higher order carbon oxides.

AutoGAMESS provides the following public interfaces:

- **new_project**: Builds a directory tree for housing input/output files with spreadsheets for collected data.
- **input_builder**: Builds optimization input files based on text file specifications
- **opt2hes**: Converts optimization input files into Hessian input files
- **hes2Raman**: Converts Hessians input files into Raman input files
- **sort_logs**: Sorts GAMESS(US) output files
- **fill_spreadsheet**: Fills in Excel Spreadsheets with data collected from log files
- **get_data**: Collects data from output files
- **make_plot**: Makes a vibrational frequency vs. IR/Raman intensity line plot

4.2 Capabilities

AutoGAMESS is capable of initializing an entire directory with well-organized subdirectories for housing all input and output files. This main directory will also contain spreadsheets that AutoGAMESS is later capable of filling with the data collected from parsing the output files. Once a main directory is initialized, input files can be generated for GAMESS(US) optimization calculations. Requiring only a simple CSV file as input, AutoGAMESS' **input_builder** function can generate thousands of files with any form of internal GAMESS(US) level of theory and both internal and external basis sets. External basis sets must be a part of EMSL Basis Set Exchange [74, 75, 76]. This requirement is due to the integration of the EMSL basis set exchange Python package into AutoGAMESS. After the user has run a geometry optimization calculation, AutoGAMESS is able to quickly get the required data from the output to modify the geometry optimization input file into a Hessian calculation input file. Similarly, after a Hessian calculation AutoGAMESS can use the output to quickly generate a Raman calculation input file. Once all calculations a user desires to run have been

completed, AutoGAMESS can sort the output files, then parse files for specific molecular properties and fill in the spreadsheets that had been generated initially. All Hessian and Raman data is pulled directly from output files, while geometry optimization properties, such as bond lengths and angles, are calculated by AutoGAMESS. Bond lengths are calculated by using the simple Euclidean distance formula,

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (4.1)$$

while bond angles are calculated by first performing a translation on the Cartesian coordinates, of the general form, $P(x, y, z) \rightarrow P'(x-a, y-b, z-c)$, where (a, b, c) are the coordinates of a central atom. Then the angle between two atoms with a third as the origin is found using the equation,

$$A = \arccos(\hat{V}_1 \cdot \hat{V}_2) \quad (4.2)$$

where A is the bond angle and \hat{V}_1 and \hat{V}_2 are the normalized position vectors for each atoms location in relation to the modified origin. AutoGAMESS is also capable of generating line plots of the vibrational frequency vs. IR/Raman intensity/activity. Generated plots will be titled with the molecule name in the file and the theory/basis set used for the calculation. Each symmetry group will be plotted in a different color, from either a default or user-specified color list. The spectral line (sum of line broadening) will also be plotted in red with 50% transparency. Figure 1 shows an example line plot of H₂O using B3LYP/CCD for the calculation.

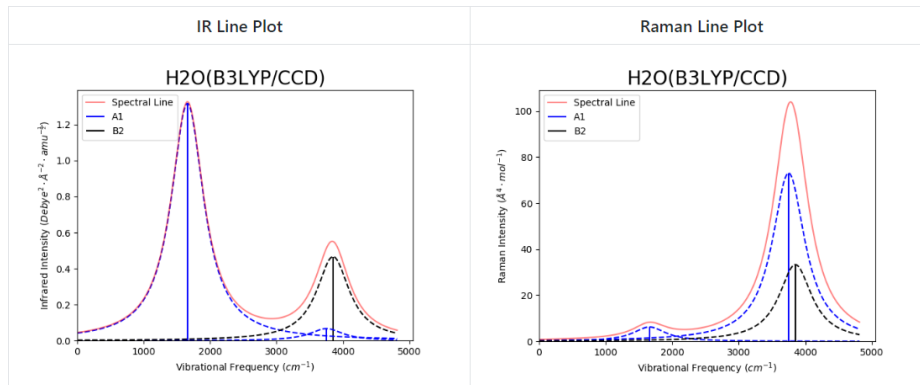


Figure 4.1: Line Plots Example

AutoGAMESS can also generate scaling factors for vibrational frequencies using the least squares method by Scott & Radom (1996)[77]. This method involves minimizing the residuals, Δ , given by

$$\Delta = \sum_i^{\text{all}} (\lambda \omega_i^{\text{theory}} - \nu_i^{\text{expt}})^2 \quad (4.3)$$

resulting in

$$\lambda = \frac{\sum_i^{\text{all}} \omega_i^{\text{theory}} \nu_i^{\text{expt}}}{\sum_i^{\text{all}} (\omega_i^{\text{theory}})^2} \quad (4.4)$$

where ω_i^{theory} and ν_i^{expt} are the i th theoretical harmonic and i th experimental fundamental frequencies (in cm^{-1}) and λ is the scaling factor. The root mean square error is then given by

$$\text{rms} = \sqrt{\frac{\sum_i^n \Delta_{\text{min}}}{n}} \quad (4.5)$$

where n is the number of frequency modes for the molecule and Δ_{min} is the minimized residual for each particular mode.

4.3 Use of AutoGAMESS

AutoGAMESS was developed to be versatile in its usability, several examples (in both shell and Python scripts) can be found in the software's GitHub repository

4.4 Availability

This software is distributed under the MIT License and can be installed through Python's pip install command.

```
python -m pip install autogamess --user
```

4.5 Dependencies

AutoGAMESS requires all the following Python packages:

- Python-3.x
- NumPy [78, 79]
- SciPy [80]
- Pandas [81]
- `basis_set_exchange`
- PeriodicElements
- openpyxl
- Matplotlib [82]

5 A COMPARISON OF MEDIUM-SIZED BASIS SETS FOR THE PREDICTION OF GEOMETRIES, VIBRATIONAL FREQUENCIES, INFRARED INTENSITIES AND RAMAN ACTIVITIES FOR WATER

This chapter (Chapter 5) contains previously published content. The usage of this content is approved by the journal's copyright policy, which states authors can include their publications in full or in part in a thesis or dissertation.

REF: Brian C Ferrari and Chris J Bennett. “A Comparison of Medium-Sized Basis Sets for the Prediction of Geometries, Vibrational Frequencies, Infrared Intensities and Raman Activities for Water”. In: *Journal of Physics: Conference Series*. Vol. 1290. 1. IOP Publishing. 2019, p. 012013

5.1 Introduction

Predictions of molecular properties, such as geometries, vibrational frequencies, infrared intensities and Raman activities have become powerful tools for chemists and physicists performing spectroscopic analysis of substances. Particularly involving the characterization or quantification of unstable species, or radicals which are produced *in situ* during experiments, where experimental values are often not available[63, 84]. Commonly employed quantum me-

chemical approaches to calculation these properties include density functional theory (DFT) and *ab initio* methods, of which extensive studies have been done investigating accuracy of their predictions[85, 86, 87, 88, 89, 90, 91, 92, 93]. However, only a few studies exist systematically comparing the performance of a limited number of differing atomic orbital basis sets to accurately calculate these properties[94]. Moreover, careful selection from the plethora of basis set choices now easily accessible (for example through the EMSL database, <https://bse.pnl.gov/bse/portal>) can lead to optimization of the approach implemented based upon computational resources and required accuracy levels[94, 95, 96]. Since computational resources are often constrained, Pople-style basis sets are often employed since they are often not demanding, and widely available in commercial packages, however, they are often out-performed by other basis sets at similar computational cost. On the other hand, correlation consistent basis sets are widely available and - although computationally heavy - they do systematically converge on accurate results when either 5/6 zeta basis sets are used, or complete basis set extrapolation schemes are incorporated (where each zeta, ζ , typically approximately doubles the number of atomic orbitals). Since computational costs for these calculations scale with N^4 to N^7 , the runtimes for even medium-sized molecules under 5/6 zeta become unfeasible. However, these large calculations don't necessarily produce the most accurate predictions. The results from several basis set families can be extrapolated to complete basis set (CBS) limit via a series of calculations performed at lower zeta level[97, 98, 99]. This study compares the results of several medium-sized basis sets across three levels of theory, in order to identify which basis sets can be reliably utilized to predict molecular properties (such as geometries, frequencies, IR intensities and Raman activities) maximizing the precision/accuracy for a given set of computational resources. Here, we only focus on preliminary results for water, but the findings are in line with those observed for other molecules we have investigated. Both DFT and *ab initio* calculations were performed, using Becke's three parameter exchange functional[100] along with Lee-Yang-Parr's correlation functional[101] (B3LYP), second-order Møller-Plesset (MP2) perturbation theory[102] and coupled-cluster

with single, double and perturbative triple excitations [CCSD(T)][103, 104]. For each level of theory 69 different atomic orbital basis sets were used varying from Pople-style (6-31G & 6-311G)[105], Dunning’s correlation consistent (cc-pV(n+d)Z & aug-cc-pV(n+d)Z)[106], as well as Truhlar’s calendar variations[107, 108, 109, 110, 111, 112], Jensen’s polarization consistent (pc-n & aug-pc-n)[113, 114, 115], Alrich’s (def2-...)[116], Sapporo’s and Karlsruhe’s, as well as atomic natural orbitals (ANOs) such as NASA Ames (ANOn), Neese-style[117], and Roos-style. We also compare several basis sets specifically designed for calculating vibrational and electronic properties, including the Sadlej-pVTZ (and LPol-X families)[118, 119], as well as SNS & NO7 families of Barone[120, 121] and a new generation of Thakkar DZ basis sets, NLO-1[122]. Calculated values are compared to both experimental values[123, 124, 125] where available, as well as various convergences found either running large zeta calculations or through extrapolation. Though many medium-sized basis sets may converge on the solutions provided by extrapolation schemes and high-zeta results, this does not guarantee, and in fact rarely results in the prediction of molecular properties being accurately calculated due to inherent flaws in the level of theory taken. Therefore, minor improvements at the cost of significantly larger runtimes are rarely worthwhile. It is of note that this study does not account or attempt to correct for errors arising from inadequacies in the levels of theory described here, although it is possible to do so (e.g., by correcting for the truncation of the theoretical approaches, core-correlation effects, or relativistic effects which need to be taken into account to reproduce frequencies to within 1 cm^{-1}). However, due to fortuitous cancellation of errors, the CBS limits, or cc-pV6Z results are often within $5\text{-}10 \text{ cm}^{-1}$ of the corrected values[126]. Here, we report on which basis sets can be reliably utilized to predict molecular properties (such as geometries, frequencies, IR intensities and Raman activities) maximizing the precision/accuracy for a given set of computational resources. The need for high zeta basis sets, or extrapolated results can often be avoided, if the basis sets are more carefully considered for the property being evaluated, and more accurate results reliably obtained at a lower computational cost. We briefly address the benefits of using anharmonic

corrections to determine vibrational frequencies versus the use of scaling factors for a few select cases.

5.2 Methods

Calculations were done using the General Atomic and Molecular Electronic Structure System, GAMESS(US)[127, 128, 129], using the 14 FEB 2018 (R1) version of Gamess(US). All calculations were run using spherical harmonics; Pople-style basis sets were typically ran using cartesian functionals unless they incorporated f-type polarization functions, where spherical harmonics were used. The segmented versions of polarization consistent basis sets were used, it should also be noted that within the Def2 family QZVPD is equivalent to QVZPPD and that QZVP is equivalent to QVZPP. Tight convergence criteria were used throughout (e.g., integral cutoff 10^{-12} , primitive cutoff 10^{-25} , 10^{-7} SCF convergence, gradient convergence 10^{-6} for geometries). Hessian and Raman calculations were run with semi-numerical (B3LYP and MP2) or fully-numerical gradients [CCSD(T)] under the double harmonic approximation with projected frequencies and two displacements in each cartesian direction during force field computations. The frozen-core approximation was used for both MP2 and CCSD(T) calculations. B3LYP calculations were run using an “army grade” pruned grid (JANS=2) with roughly 71,000 grid points and 155 radial points per atom. In general calculations were run on four processors, with only a few exceptions. Larger basis sets such cc-pV5Z, cc-pV6Z and Pcseg-4 were run with fully numerical gradients for all theory methods, and required calculations to be run on a single processor. For a few special cases, the anharmonic frequencies were calculated using Vibrational self-consistent field (VSCF)[130]. These calculations were run with third order mode couplings, ten grid points, with second order degenerate perturbation theory corrections. Extrapolated complete basis set (CBS) limits were obtained using the following two-point linear extrapolation formula

generated by Halkier *et al*[131].

$$E_{X,Y} = \frac{E_X^{\text{corr}}X^3 - E_Y^{\text{corr}}Y^3}{X^3 - Y^3} \quad (5.1)$$

Where X and Y are the basis set’s zeta value, with $Y = X - 1$, CBS limits were found using the highest zeta basis sets calculated per basis “family”. Error analysis was done by taking the average error between the calculated values and either experimental data or CBS limits.

5.3 Results and Discussion

Table 5.1: Molecular Geometry Predictions, Bond Lengths are in Angstroms (Å), Bond Angles are in Degrees (°), {1} is fastest time rank

	Optimization	Bond Length{Time Rank}			Bond Angle		
	Experimental[124]	0.9578			104.4776		
#	Basis Set	B3LYP	MP2	CCSD(T)	B3LYP	MP2	CCSD(T)
1	cc-pV6Z	0.9606{67}	0.9582{68}	0.9582{69}	105.120	104.338	104.422
2	cc-pV5Z	0.9605{63}	0.9580{63}	0.9580{61}	105.091	104.286	104.373
3	Aug-cc-pVQZ	0.9608{55}	0.9589{55}	0.9589{55}	105.110	104.269	104.365
4	Jul-cc-pVQZ	0.9608{53}	0.9588{52}	0.9589{52}	105.109	104.263	104.360
5	Jun-cc-pVQZ	0.9608{50}	0.9586{48}	0.9587{49}	105.114	104.209	104.301
6	May-cc-pVQZ	0.9608{48}	0.9584{45}	0.9585{45}	105.122	104.219	104.306
7	cc-pVQZ	0.9605{45}	0.9577{43}	0.9579{43}	104.877	104.018	104.116
8	Aug-cc-pVTZ	0.9621{41}	0.9614{39}	0.9616{39}	105.082	104.109	104.180
9	Jul-cc-pVTZ	0.9620{39}	0.9614{37}	0.9616{37}	105.100	104.125	104.188
10	Jun-cc-pVTZ	0.9620{35}	0.9609{34}	0.9611{35}	105.110	104.063	104.118
11	May-cc-pVTZ	0.9620{32}	0.9603{31}	0.9604{31}	105.052	103.893	103.920
12	cc-pVTZ	0.9616{29}	0.9591{27}	0.9594{29}	104.517	103.518	103.582
13	Aug-cc-pVDZ	0.9651{21}	0.9659{20}	0.9665{21}	104.741	103.873	103.937
14	Jul-cc-pVDZ	0.9650{14}	0.9653{14}	0.9658{14}	105.035	104.232	104.256
15	Jun-cc-pVDZ	0.9686{8}	0.9665{8}	0.9672{8}	104.045	102.835	102.798
16	May-cc-pVDZ	0.9686{10}	0.9665{8}	0.9672{9}	104.045	102.835	102.798
17	cc-pVDZ	0.9689{4}	0.9649{5}	0.9663{6}	102.720	101.902	101.912
18	6-311++G(3df,3pd)	0.9614{34}	0.9589{33}	0.9592{34}	105.040	104.037	104.086
19	6-311G(2df,2pd)	0.9610{30}	0.9569{26}	0.9575{23}	103.936	103.109	103.211
20	6-311++G(2d,2p)	0.9612{14}	0.9583{17}	0.9584{20}	105.138	104.279	104.272
21	6-311G(2d,2p)	0.9617{13}	0.9573{13}	0.9578{12}	103.902	103.246	103.313
22	6-311++G(d,p)	0.9624{11}	0.9600{8}	0.9607{13}	105.043	103.358	103.299
23	6-311+G(d,p)	0.9624{6}	0.9600{5}	0.9607{10}	105.047	103.374	103.314
24	6-311G(d,p)	0.9624{6}	0.9586{5}	0.9596{4}	103.752	102.300	102.306
25	6-31+G(d,p)	0.9653{4}	0.9632{1}	0.9633{5}	105.744	105.421	105.338
26	pc4	0.9606{64}	0.9576{66}	0.9577{63}	105.128	104.347	104.430
27	aug-pc3	0.9606{57}	0.9581{57}	0.9583{59}	105.137	104.370	104.449
28	pc3	0.9606{49}	0.9578{49}	0.9580{48}	105.139	104.341	104.415
29	aug-pc2	0.9610{40}	0.9590{38}	0.9593{38}	105.145	104.306	104.340
30	pc2	0.9606{23}	0.9580{30}	0.9582{30}	105.149	104.177	104.165

#	Optimization	Bond Length{Time Rank}			Bond Angle		
	Experimental[124]	0.9578			104.4776		
	Basis Set	B3LYP	MP2	CCSD(T)	B3LYP	MP2	CCSD(T)
31	aug-pc1	0.9661{19}	0.9662{17}	0.9662{18}	104.831	103.922	104.000
32	pc1	0.9686{3}	0.9650{1}	0.9654{3}	103.865	103.553	103.518
33	Def2-QZVPD	0.9607{53}	0.9585{51}	0.9586{47}	105.121	104.237	104.321
34	Def2-QZVP	0.9604{46}	0.9580{46}	0.9581{44}	105.099	104.223	104.301
35	Def2-TZVPPD	0.9613{37}	0.9598{36}	0.9600{34}	105.127	104.070	104.118
36	Def2-TZVPP	0.9610{26}	0.9586{27}	0.9589{26}	104.893	103.823	103.859
37	Def2-TZVPD	0.9636{31}	0.9644{31}	0.9650{32}	105.111	104.377	104.401
38	Def2-TZVP	0.9630{17}	0.9622{21}	0.9630{19}	105.236	104.597	104.548
39	Def2-SVPD	0.9644{18}	0.9643{14}	0.9649{15}	105.370	104.444	104.476
40	Def2-SVP	0.9670{2}	0.9625{1}	0.9638{2}	103.078	102.438	102.465
41	SPK-AQZP	0.9607{56}	0.9585{56}	0.9586{57}	105.157	104.351	104.440
42	SPK-QZP	0.9604{47}	0.9580{47}	0.9582{46}	105.082	104.234	104.321
43	SPK-ATZP	0.9612{42}	0.9598{40}	0.9602{40}	105.102	104.227	104.266
44	SPK-TZP	0.9612{33}	0.9588{35}	0.9593{33}	104.355	103.493	103.565
45	SPK-ADZP	0.9645{22}	0.9640{23}	0.9645{22}	104.790	103.996	104.047
46	SPK-DZP	0.9674{8}	0.9637{8}	0.9647{7}	102.772	102.267	102.287
47	KTZVPP	0.9610{24}	0.9586{27}	0.9589{26}	104.893	103.823	103.859
48	KTZV	0.9750{1}	0.9757{1}	0.9781{1}	109.965	110.567	110.162
49	Sadlej-LPolX-fl	0.9616{69}	— ^a	0.9621{69}	105.038	— ^a	104.364
50	Sadlej-LPolX-fs	0.9613{60}	0.9615{60}	0.9622{60}	105.048	104.431	104.489
51	Sadlej-LPolX-dl	0.9622{54}	0.9607{54}	0.9611{54}	105.055	104.420	104.454
52	Sadlej-LPolX-ds	0.9616{44}	0.9614{50}	0.9621{51}	105.038	104.265	104.296
53	Sadlej-pVTZ	0.9675{28}	0.9689{25}	0.9697{28}	104.592	103.541	103.610
54	N07T	0.9644{27}	0.9665{22}	0.9676{25}	104.754	103.609	103.632
55	N07D	0.9644{12}	0.9641{8}	0.9645{11}	104.643	103.679	103.733
56	SNST	0.9649{24}	0.9653{24}	0.9659{24}	104.702	103.853	103.896
57	SNSD	0.9644{20}	0.9643{17}	0.9648{17}	104.662	103.748	103.802
58	NLO-1	0.9751{16}	0.9840{14}	0.9847{16}	106.514	105.024	105.088
59	Roos-Aug-TZ-ANO	0.9607{61}	0.9581{61}	0.9583{62}	105.128	104.251	104.310
60	Roos-Aug-DZ-ANO	0.9616{43}	0.9609{44}	0.9613{50}	105.054	104.333	104.371
61	NASA-Ames-ANO2	0.9606{65}	0.9587{64}	0.9588{65}	105.042	104.249	104.340
62	NASA-Ames-ANO1	0.9629{49}	0.9614{58}	0.9617{56}	104.923	103.928	104.000
63	NASA-Ames-ANO0	0.9647{36}	0.9638{41}	0.9642{41}	104.393	103.662	103.602
64	Neese-ANO-AVQZ	0.9606{68}	0.9588{67}	0.9588{68}	105.144	104.330	104.414
65	Neese-ANO-VQZ	0.9608{66}	0.9589{65}	0.9589{66}	105.127	104.303	104.391
66	Neese-ANO-AVTZ	0.9609{62}	0.9581{62}	0.9583{64}	105.085	104.213	104.286
67	Neese-ANO-VTZ	0.9628{59}	0.9617{59}	0.9619{58}	105.087	104.069	104.130
68	Neese-ANO-AVDZ	0.9633{52}	0.9605{53}	0.9606{53}	105.053	104.275	104.335
69	Neese-ANO-VDZ	0.9648{38}	0.9645{42}	0.9646{42}	104.584	103.861	103.794

^aRemoved due to convergence issues

Table 5.2: Harmonic Vibrational Frequencies in Wavenumbers (cm^{-1})

Hes	ω_1 {Time Rank}			ω_2			ω_3		
Exp [87]	3832.17			1648.47			3942.53		
The [126]	3835.89			1649.39			3946.65		
#	B3LYP	MP2	CCSD(T)	"	"	"	"	"	"
1	3806.0{67}	3845.0{68}	3837.3{ ^a }	1631.4	1633.0	1651.3 ^a	3907.4	3971.2	3947.2 ^a
2	3805.6{62}	3848.9{62}	3840.1{ ^a }	1630.0	1635.6	1653.4 ^a	3906.9	3974.3	3949.3 ^a
3	3801.6{55}	3839.8{55}	3831.2{54}	1628.4	1632.2	1649.6	3903.3	3965.6	3940.8

Hes	ω_1 {Time Rank}			ω_2			ω_3		
Exp [87]	3832.17			1648.47			3942.53		
The [126]	3835.89			1649.39			3946.65		
#	B3LYP	MP2	CCSD(T)	"	"	"	"	"	"
4	3802.2{53}	3840.5{52}	3831.8{51}	1628.8	1632.5	1649.9	3903.8	3966.4	3941.5
5	3801.7{51}	3843.4{50}	3834.4{49}	1628.4	1635.6	1652.5	3903.3	3968.8	3943.5
6	3801.9{47}	3845.0{47}	3836.1{45}	1628.3	1636.5	1653.4	3903.5	3970.3	3945.1
7	3803.1{45}	3855.2{43}	3844.5{42}	1633.9	1642.5	1658.8	3903.4	3977.8	3951.5
8	3793.8{40}	3822.0{39}	3810.9{38}	1626.4	1628.1	1645.4	3896.4	3947.8	3920.0
9	3798.2{38}	3824.6{37}	3813.7{36}	1627.3	1629.9	1647.4	3901.1	3952.0	3924.4
10	3799.6{34}	3832.6{33}	3821.9{32}	1626.1	1633.8	1650.8	3903.3	3960.0	3932.7
11	3799.9{33}	3843.4{32}	3833.6{31}	1625.9	1642.0	1661.0	3902.8	3967.0	3940.1
12	3797.8{27}	3855.5{29}	3841.0{29}	1638.8	1651.9	1668.4	3898.0	3975.8	3945.6
13	3791.7{19}	3803.4{19}	3787.0{20}	1617.8	1622.1	1637.7	3901.4	3937.6	3904.9
14	3794.9{14}	3818.9{13}	3804.2{13}	1624.2	1632.4	1648.8	3904.6	3952.7	3922.1
15	3772.7{8}	3842.6{8}	3824.1{8}	1617.5	1638.4	1655.8	3875.2	3964.2	3930.4
16	3772.7{8}	3842.6{8}	3824.1{9}	1617.5	1638.4	1655.8	3875.2	3964.2	3930.4
17	3748.4{5}	3851.8{5}	3821.7{5}	1658.1	1677.6	1689.8	3849.8	3971.1	3927.6
18	3810.8{39}	3869.8{34}	3855.9{34}	1625.7	1623.7	1640.5	3909.7	3989.1	3959.7
19	3805.3{27}	3890.6{26}	3871.5{25}	1655.0	1670.3	1685.6	3901.1	4005.0	3971.4
20	3816.4{21}	3861.2{19}	3848.9{21}	1637.6	1659.7	1678.6	3917.8	3982.7	3953.9
21	3803.7{16}	3872.5{16}	3852.7{15}	1666.0	1686.1	1701.1	3900.1	3988.6	3953.7
22	3813.3{13}	3880.1{11}	3860.8{12}	1601.6	1629.7	1648.0	3918.7	3999.6	3965.1
23	3813.2{11}	3879.6{7}	3860.2{11}	1601.8	1629.7	1647.9	3918.7	3999.4	3964.9
24	3804.3{7}	3896.6{5}	3868.9{6}	1636.8	1667.2	1681.7	3902.3	4007.4	3966.7
25	3806.8{4}	3867.1{3}	3855.8{4}	1602.5	1622.8	1639.4	3928.7	4013.9	3987.5
26	3805.5{64}	— ^b {64}	— ^b {61}	1630.8	1654.2	1677.8	3907.1	3968.6	3945.0
27	3802.6{68}	3822.4{58}	3811.5{58}	1628.0	1632.9	1649.5	3904.4	3948.2	3921.1
28	3802.9{49}	3822.7{49}	3811.5{46}	1627.6	1635.1	1651.7	3904.6	3948.7	3921.2
29	3805.3{41}	3846.2{38}	3832.8{37}	1626.3	1632.5	1649.8	3908.9	3972.2	3942.6
30	3806.6{27}	3859.3{28}	3846.8{30}	1623.5	1637.7	1656.3	3910.9	3982.8	3954.2
31	3780.2{18}	3805.5{18}	3797.2{18}	1615.1	1630.7	1648.5	3887.6	3940.3	3916.1
32	3762.0{7}	3854.7{2}	3840.2{3}	1627.5	1648.7	1664.7	3873.5	3991.1	3962.4
33	3805.0{50}	3849.3{51}	3840.0{48}	1628.4	1635.3	1652.2	3906.8	3975.6	3949.9
34	3807.2{46}	3855.9{45}	3846.5{43}	1629.6	1637.0	1654.1	3908.3	3980.7	3955.0
35	3808.2{37}	3838.1{35}	3826.0{35}	1627.2	1636.3	1653.3	3911.2	3963.5	3935.2
36	3807.3{31}	3853.7{29}	3841.1{28}	1631.9	1646.5	1664.1	3910.2	3976.3	3947.7
37	3783.4{26}	3798.0{27}	3783.1{26}	1622.9	1620.4	1636.9	3884.8	3925.2	3893.5
38	3782.9{19}	3819.0{21}	3801.3{19}	1616.2	1613.3	1629.4	3888.1	3946.1	3912.4
39	3816.8{15}	3843.4{14}	3828.0{14}	1584.2	1596.4	1612.3	3923.1	3974.9	3944.5
40	3788.9{3}	3895.3{3}	3867.3{2}	1638.2	1654.5	1666.9	3884.0	4011.3	3971.1
41	3805.6{57}	3845.2{56}	3836.5{56}	1627.5	1631.4	1648.7	3907.1	3971.4	3946.5
42	3807.1{47}	3852.9{46}	3843.2{44}	1629.8	1635.4	1652.5	3908.7	3978.2	3952.5
43	3799.9{42}	3835.5{42}	3821.3{41}	1625.4	1626.4	1643.6	3901.5	3961.3	3930.8
44	3799.1{36}	3852.4{36}	3835.5{33}	1648.0	1655.2	1670.8	3895.5	3971.9	3939.3
45	3778.4{22}	3797.3{22}	3780.4{22}	1604.9	1617.9	1634.3	3888.1	3931.8	3899.3
46	3759.5{6}	3844.6{12}	3817.3{7}	1653.6	1665.2	1679.3	3864.5	3972.5	3932.2
47	3807.3{30}	3853.7{29}	3841.1{27}	1631.9	1646.5	1664.1	3910.2	3976.3	3947.7
48	3610.8{1}	3598.2{1}	3551.4{1}	1567.0	1601.4	1610.4	3775.9	3787.2	3725.3
49	3798.6{69}	— ^c	4031.4{62}	1634.4	— ^c	2126.1	3898.0	— ^c	4126.8
50	3803.0{60}	3811.3{60}	3793.1{59}	1638.6	1642.1	1658.6	3904.4	3943.8	3910.5
51	3796.7{54}	3820.2{54}	3805.0{53}	1630.4	1629.3	1645.5	3899.0	3949.7	3918.3
52	3799.1{44}	3811.9{48}	3793.4{49}	1629.8	1640.1	1655.9	3900.3	3937.6	3902.9
53	3794.7{24}	3797.9{25}	3779.9{24}	1633.8	1642.9	1658.3	3908.4	3935.8	3900.8
54	3770.5{25}	3775.4{24}	— ^b	1631.4	1622.9	— ^b	3871.7	3901.6	— ^b
55	3812.6{8}	3841.5{8}	3829.4{10}	1640.9	1650.4	1665.3	3921.1	3975.9	3948.0

Hes	ω_1 {Time Rank}			ω_2			ω_3		
Exp [87]	3832.17			1648.47			3942.53		
The [126]	3835.89			1649.39			3946.65		
#	B3LYP	MP2	CCSD(T)	"	"	"	"	"	"
56	3785.8{23}	3810.4{23}	3794.3{23}	1629.8	1632.6	1647.6	3896.4	3946.3	3914.3
57	3808.6{17}	3834.6{17}	3820.6{17}	1633.6	1640.4	1655.4	3918.0	3968.3	3938.1
58	3716.6{12}	3620.4{15}	3597.6{16}	1600.3	1605.1	1623.5	3862.4	3790.4	3750.2
59	3807.0{61}	3850.1{61}	3838.1{60}	1627.6	1639.3	1656.2	3906.6	3974.4	3946.0
60	3807.9{43}	3828.9{44}	3812.6{47}	1635.6	1644.9	1660.5	3908.3	3954.9	3921.7
61	3803.7{65}	3843.7{65}	3835.1{65}	1631.6	1636.0	1652.7	3906.6	3972.5	3947.5
62	3798.1{56}	3838.6{57}	3827.2{55}	1629.5	1637.9	1653.3	3901.7	3969.1	3942.3
63	3810.2{32}	3854.4{40}	3835.0{39}	1638.1	1642.0	1659.1	3927.6	3991.6	3958.2
64	3804.4{68}	3840.9{67}	3832.9{64}	1628.6	1631.3	1649.0	3904.8	3967.5	3943.2
65	3802.1{66}	3842.2{66}	— ^b	1628.5	1634.2	— ^b	3907.0	3973.2	— ^b
66	3803.6{63}	3851.9{63}	3842.0{63}	1629.3	1640.0	1656.8	3904.8	3977.6	3951.6
67	3800.2{59}	3833.9{59}	3823.8{57}	1624.3	1633.2	1649.3	3904.2	3965.2	3939.0
68	3796.7{52}	3845.0{53}	3832.9{52}	1628.8	1637.9	1653.1	3900.3	3972.3	3944.1
69	3811.5{35}	3847.8{41}	3833.0{40}	1633.8	1635.3	1654.0	3931.3	3989.1	3959.2

^a These results were taken from Tew *et al.* (2007)

^b An error in the calculation of this vibrational frequency caused this particular result to be invalid

^c Removed due to convergence issues

Table 5.3: Infrared Intensities in $\text{Km} \cdot \text{mol}^{-1}$ and Raman Activities in $\text{\AA}^4 \text{amu}^{-1}$

—	Infrared Intensities						Raman Intensities					
Exp [87]	2.93		62.5		41.7		111 ± 12		0.9 ± 0.2		19 ± 2	
—	IR ₁		IR ₂		IR ₃		R ₁		R ₂		R ₃	
#	B3LYP	MP2	"	"	"	"	"	"	"	"	"	"
3	4.63{57}	6.09{57}	76.10	73.17	63.27	78.36	98.26	109.50	0.81	0.83	25.4	22.1
4	5.20{52}	6.68{51}	76.19	73.04	63.65	77.85	87.63	97.56	0.71	0.75	25.0	21.9
5	5.21{55}	6.80{54}	76.03	73.50	63.67	78.52	87.75	96.98	0.72	0.75	25.0	21.8
6	5.16{49}	6.69{48}	76.56	72.64	63.16	77.23	88.55	97.72	0.71	0.79	25.4	22.2
7	3.89{45}	6.55{43}	72.79	69.11	51.46	68.30	78.51	78.28	2.75	2.78	26.0	23.5
8	4.59{41}	5.56{39}	75.71	71.68	62.83	75.46	96.82	107.62	1.02	1.09	25.6	22.8
9	5.86{36}	7.44{33}	77.62	72.50	62.87	74.21	80.43	87.64	0.86	0.97	26.1	23.6
10	5.91{39}	7.61{37}	77.20	73.81	63.08	75.77	79.33	87.53	0.86	0.91	25.8	23.1
11	6.53{34}	9.57{32}	69.58	63.43	60.65	73.11	78.84	81.66	3.79	4.00	28.5	26.2
12	3.16{29}	5.74{29}	69.46	64.52	40.54	55.20	71.40	69.55	4.34	4.25	26.1	24.8
13	4.00{19}	4.14{19}	71.28	67.46	60.41	67.02	96.28	104.28	1.86	1.96	25.5	24.2
14	11.75{8}	16.28{9}	65.74	60.84	59.66	66.22	80.40	81.55	4.52	4.82	37.5	36.3
15	6.33{14}	9.13{13}	81.06	79.23	55.36	65.70	72.34	74.73	1.35	1.39	29.9	27.6
16	11.75{9}	16.28{9}	65.74	60.84	59.66	66.22	80.40	81.55	4.52	4.82	37.5	36.3
17	2.84{4}	6.56{5}	55.64	56.98	19.70	32.64	73.03	68.32	6.24	5.83	33.3	33.8
18	4.45{37}	6.42{34}	72.10	67.09	59.30	72.70	94.11	96.06	2.20	2.28	26.5	23.5
19	2.72{26}	5.12{26}	68.08	64.04	33.76	49.29	69.05	65.45	4.11	4.13	23.8	22.8

	Infrared Intensities						Raman Intensities					
Exp [87]	2.93		62.5		41.7		111 ± 12		0.9 ± 0.2		19 ± 2	
	IR ₁		IR ₂		IR ₃		R ₁		R ₂		R ₃	
#	B3LYP	MP2	"	"	"	"	"	"	"	"	"	"
20	7.23{22}	9.84{19}	70.95	65.33	61.76	73.54	82.31	82.95	4.54	4.74	28.3	26.1
21	2.06{17}	3.60{15}	65.23	61.36	29.89	43.09	73.50	69.74	4.71	4.75	24.5	23.8
22	9.01{13}	12.66{11}	66.60	56.65	56.58	61.67	83.54	81.66	7.45	7.94	34.7	34.1
23	9.48{11}	12.96{7}	67.05	57.13	57.07	61.97	80.38	78.95	7.77	8.09	34.2	33.8
24	3.41{7}	5.82{6}	57.32	50.17	24.14	32.17	71.51	66.12	7.90	7.92	30.6	31.1
25	6.34{5}	10.08{3}	91.20	92.87	57.07	66.68	80.34	78.73	3.36	3.48	35.1	34.0
27	4.64{60}	6.09{60}	76.26	72.91	63.30	78.90	97.55	111.47	0.67	0.75	25.5	21.8
28	5.07{50}	7.04{50}	74.88	71.04	62.08	78.03	86.39	95.65	1.63	1.76	26.4	23.0
29	4.66{40}	5.60{38}	75.92	71.88	63.41	75.58	97.96	109.88	0.74	0.81	25.7	22.4
30	5.76{28}	9.42{28}	71.38	66.09	54.89	68.70	76.92	76.53	4.31	4.38	29.9	27.7
31	4.57{16}	4.79{18}	75.72	72.47	65.99	73.73	96.78	104.35	0.69	0.65	25.3	22.3
32	7.35{2}	13.81{3}	74.19	78.31	42.98	63.56	76.17	70.14	5.08	4.57	37.1	35.7
33	4.61{54}	6.08{52}	76.07	71.96	63.09	77.54	97.95	111.63	0.77	0.84	25.4	22.1
34	5.47{46}	8.14{45}	74.45	70.25	58.55	74.74	77.32	82.65	2.49	2.57	26.6	23.7
35	4.66{38}	5.59{36}	75.81	70.19	62.89	74.20	96.54	107.74	0.78	0.89	25.5	22.6
36	4.46{30}	7.23{29}	69.75	64.39	48.72	62.46	76.52	75.31	4.51	4.44	27.8	26.0
37	4.08{32}	5.13{27}	77.76	76.03	61.03	74.42	98.89	108.40	0.66	0.73	26.1	23.3
38	4.38{21}	8.83{21}	80.64	80.97	43.76	60.96	78.80	74.43	5.09	4.98	32.9	31.5
39	7.05{15}	7.22{14}	77.96	73.99	75.28	82.02	93.91	103.84	0.92	0.90	24.5	22.5
40	4.79{2}	9.87{2}	55.38	58.08	26.65	42.65	74.23	69.19	7.01	6.44	32.7	33.1
41	4.74{59}	6.25{58}	76.38	73.41	63.66	79.15	94.50	109.18	0.81	0.83	25.3	22.0
42	4.97{48}	7.96{47}	74.87	71.14	56.06	73.35	78.04	80.52	2.58	2.61	26.8	23.9
43	4.73{42}	5.95{42}	75.66	71.85	63.25	76.44	90.30	110.88	0.87	0.89	25.6	22.4
44	3.08{35}	4.94{35}	68.75	64.64	44.39	59.51	75.62	78.28	3.78	3.88	25.1	23.2
45	4.03{20}	4.36{22}	71.41	68.02	60.28	68.26	92.34	103.12	1.28	1.27	25.8	23.6
46	4.40{6}	8.40{11}	60.07	62.63	28.94	44.97	72.22	69.20	5.38	5.10	31.1	31.4
47	4.46{31}	7.23{31}	69.75	64.39	48.72	62.46	76.52	75.31	4.51	4.44	27.8	26.0
48	6.91{1}	3.27{1}	94.34	103.74	19.53	29.17	98.06	94.75	9.97	9.17	37.2	39.0
50	4.55{62}	5.15{62}	76.26	72.75	62.69	75.00	106.73	121.62	0.64	0.83	25.7	22.6
51	4.57{56}	4.93{56}	76.56	72.06	63.43	73.14	103.55	106.04	0.85	0.95	26.0	23.2
52	4.79{44}	5.13{49}	76.05	71.78	63.30	72.97	95.17	118.33	0.76	0.91	25.9	23.1
53	4.66{23}	4.68{25}	73.71	69.16	61.79	68.04	102.57	110.91	1.08	1.00	24.8	22.8
54	4.63{24}	4.49{24}	75.96	69.42	60.95	67.08	98.22	109.62	0.86	0.95	25.7	23.8
55	5.65{10}	5.39{7}	78.46	74.25	62.45	69.30	72.52	76.56	0.87	0.81	21.1	20.3
56	5.09{25}	4.78{23}	76.43	71.70	62.76	69.52	95.51	108.23	0.88	0.92	25.5	23.4
57	4.56{17}	4.53{17}	74.88	69.79	61.51	67.60	97.11	105.17	0.81	0.83	25.7	23.8
58	3.95{11}	3.73{15}	81.39	73.79	69.13	70.28	92.47	99.53	2.86	3.26	24.2	24.6
59	4.65{63}	5.57{63}	76.27	72.09	63.68	76.90	90.80	111.12	0.96	0.96	25.3	22.1
60	3.99{43}	4.48{44}	73.68	69.35	59.44	69.17	95.38	113.06	1.06	1.30	24.7	22.5
61	4.36{65}	6.95{65}	73.82	71.56	56.02	73.41	72.29	86.85	2.44	2.24	24.4	22.8
62	4.51{58}	6.24{59}	73.82	69.50	47.62	61.96	66.56	72.05	3.74	3.51	21.8	21.2
63	7.78{27}	10.17{40}	77.07	76.05	47.81	61.06	68.46	65.61	5.11	4.96	25.9	27.3
64	6.37{67}	8.55{67}	75.87	73.08	63.32	78.99	73.72	87.14	2.25	2.17	25.1	22.6
65	5.00{66}	7.89{66}	74.08	71.86	58.17	76.13	74.70	85.16	2.57	2.38	25.0	23.2
66	5.12{64}	7.51{64}	73.25	70.42	59.13	74.38	78.62	83.35	2.71	2.71	25.6	23.4
67	5.10{61}	7.16{61}	74.30	69.80	51.45	65.63	65.97	73.44	3.69	3.49	22.0	21.5
68	6.32{51}	7.70{55}	71.70	67.09	53.11	64.38	68.66	73.80	4.15	3.93	23.5	23.0
69	9.09{33}	12.02{41}	78.73	77.57	51.62	65.64	65.21	67.87	5.25	5.15	26.7	27.9

**Table 5.4: VSCF Calculations
for Anharmonic Frequencies in cm^{-1} and Infrared Intensities in
 $\text{Km} \cdot \text{mol}^{-1}$**

VSCF	ω_1	ω_2	ω_3	IR ₁	IR ₂	IR ₃
Exp [132]	3657	1595	3756	2.93	62.5	41.7
B3LYP/cc-pV6Z	3639.4	1546.2	3713.5	4.39	76.68	61.01
B3LYP/cc-pV5Z	3639.8	1546.9	3713.7	4.17	76.16	58.07
B3LYP/Aug-cc-pVQZ	3637.1	1545.8	3711.3	3.92	77.46	61.74
B3LYP/May-cc-pVTZ	3635.2	1542.4	3710.7	5.61	71.47	59.03
B3LYP/aug-pc3	3638.5	1545.5	3712.9	3.92	77.63	61.76
B3LYP/Def2-QZVPD	3639.5	1545.8	3713.8	3.90	77.46	61.57
MP2/cc-pV5Z	3680.3	1553.0	3779.6	6.53	72.86	74.81
MP2/Aug-cc-pVQZ	3671.5	1550.1	3771.2	5.32	74.24	76.47
MP2/May-cc-pVTZ	3677.3	1559.5	3775.3	8.49	65.02	71.19
MP2/aug-pc3	3669.3	1551.0	3769.4	5.33	74.05	76.98
MP2/Def2-QZVPD	3679.6	1552.5	3779.6	5.30	73.16	75.67
SCS-MP2/Def2-QZVPD	3673.7	1569.2	3768.0	5.29	73.31	76.07
CCSD(T)/Def2-QZVPD	3664.6	1569.4	3746.8	13.03	93.92	79.31
CCSD(T)/Def2-TZVPPD	3650.2	1569.5	3731.9	13.11	93.55	78.27
CCSD(2) _T /Def2-QZVPD	3670.9	1571.1	3752.2	3.64	73.08	58.12
CCSD(2) _T /Def2-TZVPPD	3656.4	1571.0	3737.2	3.04	70.65	54.36

5.3.1 Calculations Removed from Study

After careful consideration, certain combinations of basis set and theory methods were dropped from the study due to issues discussed in this section. The version of Gamess(US) used for this study, does not support basis sets with high angular momentum i-functions (5/6 zeta basis sets) to be utilized for predicting IR intensities or Raman activities. Zeta 5/6

Hessian calculations for CCSD(T) were too computationally intensive and did not provide significant improvements on molecular property predictions to justify their extreme run-times. Sadlej-LPolX-fl was also dropped from most calculations (MP2 within optimization and hessian calculations, across all theories within Raman calculations) due to convergence issues.

5.3.2 Molecular Geometries

The floating integer values within Figures 1-4 throughout the following sections are the zeta values for the basis sets. As expected, CCSD(T) performs best for calculating molecular geometries yielding consistent results with few errors. B3LYP converges faster than MP2 and CCSD(T), with the latter providing only slightly more accurate bond lengths and angles, for water.

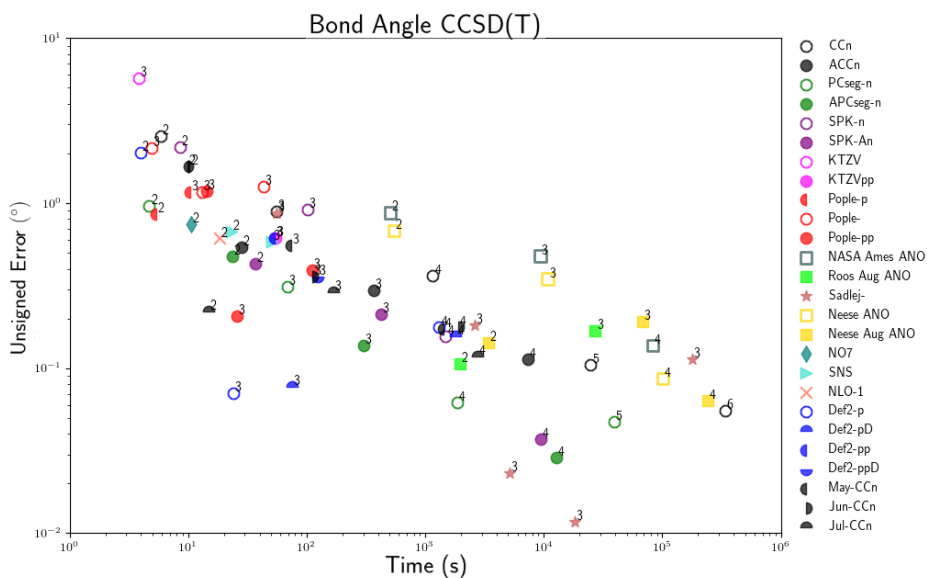


Figure 5.1: Unsigned Error of CCSD(T) Bond Angle Calculations Compared to Experimental Data

It is notable to mention that the Pople-style basis sets provide reasonably accurate geometries, and it requires considerably high zeta basis sets to converge on the accurate geometry; particularly for the bond angle of water. While bond length converges quickly, for all levels

of theory studied here, the bond angle converges more slowly, as shown for CCSD(T) in Figure 1. In general, most basis families show linear trend lines of slower convergence as either diffuseness or zeta values increase; making DFT with pople basis sets particularly well-suited for optimizing the geometries of large molecules. The Def2 family performs well in both runtime and accuracy, similarly both the PCseg and APCseg yield high accuracy predictions with lower runtimes, particularly when compared to their CC counterparts.

5.3.3 Vibrational Frequencies

When comparing vibrational frequencies to experimental data CCSD(T) calculations yielded the most accurate results, as expected. Interestingly, the NO7 and SNS basis set families yield results within ten wavenumbers (here, we consider this a reasonable threshold for the characterization or discrimination of most molecular species) with minimal runtimes[93].

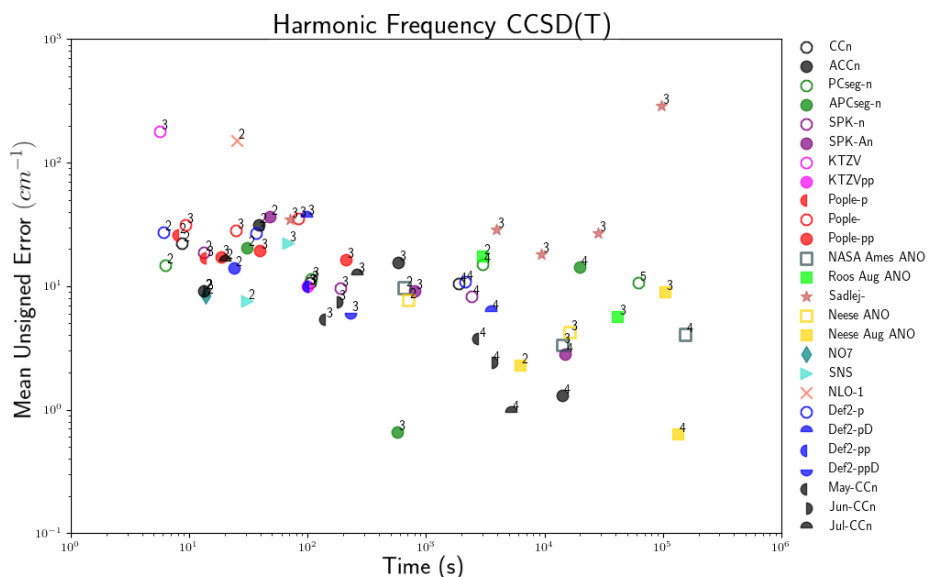


Figure 5.2: Mean Unsigned Error of CCSD(T) Frequency Calculations Compared to Experimental Data

The Def2 basis family consistently outperforms both correlation consistent and polarization consistent basis sets, at significantly reduced computational expense. Both the augmented and non-augmented SPK- basis sets also perform very well in calculating frequencies, pro-

ducing similar accuracy and runtimes as Def2 sets. While, for B3LYP, the commonly used Pople family, converges above the 10 cm^{-1} threshold with similar runtimes. It is noteworthy to mention that fortuitous cancellation of errors often plays a large role here, particularly with smaller basis sets. However, if these errors consistently off-set a deficiency in the level of theory encountered at the basis set limit, or that is not overcome unless much larger basis sets are used, this may prove useful. As an example, for the prediction of vibrational frequencies, the ‘may’ calendar basis sets consistently out-perform even their fully augmented counterparts at only a minor additional computational expense compared to the CCn basis at all levels of theory studied. B3LYP is known to over-estimate bond-lengths, and under-estimate harmonic frequencies. Often, a scaling factor is applied to reduce the errors in predicting anharmonic values from calculations performed under the harmonic approximation [133, 134, 135]. In Table 4 we can see clearly that the anharmonic frequencies for B3LYP are all below the experimental values. Again in Table 4, the anharmonic frequencies are shown for a few select levels of theory and basis sets. Here, we can see that B3LYP consistently underestimates the experimental frequencies, and that even the may-cc-pVTZ basis set values lie within 5 cm^{-1} of the cc-pV6Z values. Similarly, we see for MP2 that the def2, may and pc families provide excellent agreement with the cc-pV5Z values. For each level of theory, the def2-QZVPD basis set approaches the accuracy of cc-pV5Z/cc-pV6Z much faster than most other basis sets of similar quality. Since anharmonic calculations are computationally expensive, CCSD(T) calculations were only performed using two of the def2 basis sets, where it could be seen that although an improvement over MP2 is observed, there is still not overall agreement with the experimental harmonic frequencies. This is likely because MP2 and CCSD(T) do not perform as well when used to calculate energetics at geometries far from the equilibrium position. To emphasize this point, calculations were additionally ran using the SCS-MP2 and CCSD(2)_T levels of theory, which in each case show a marked improvement in the agreement with experimental anharmonic frequencies. All of the levels of theory considered here, struggle to reproduce the ν_2/ω_2 frequency. We note that the

CCSD(2)_T/def2-QVZPD level of theory only has a MUE of $\approx 13 \text{ cm}^{-1}$.

5.3.4 Infrared Intensities

From Figure 3, we can see that though the Pople and other small basis sets can outperform larger basis sets at predicting some infrared intensities, however, caution is urged since this is not always the case. Both the B3LYP and MP2 levels of theory have significant errors when calculating infrared intensities under the harmonic approximation, even when extrapolated to their CBS limits. The B3LYP theory level convergence line appears to lie just above the ten percent error margin, some basis sets - particularly small basis sets, such as Pople - deviate from this producing more accurate results. In some cases where small basis sets are utilized, the errors arising actually compensates for the deficiencies in B3LYP, producing seemingly more accurate results. However, these results may occur sporadically, and caution is urged using these results, which could easily be taken out of context[56].

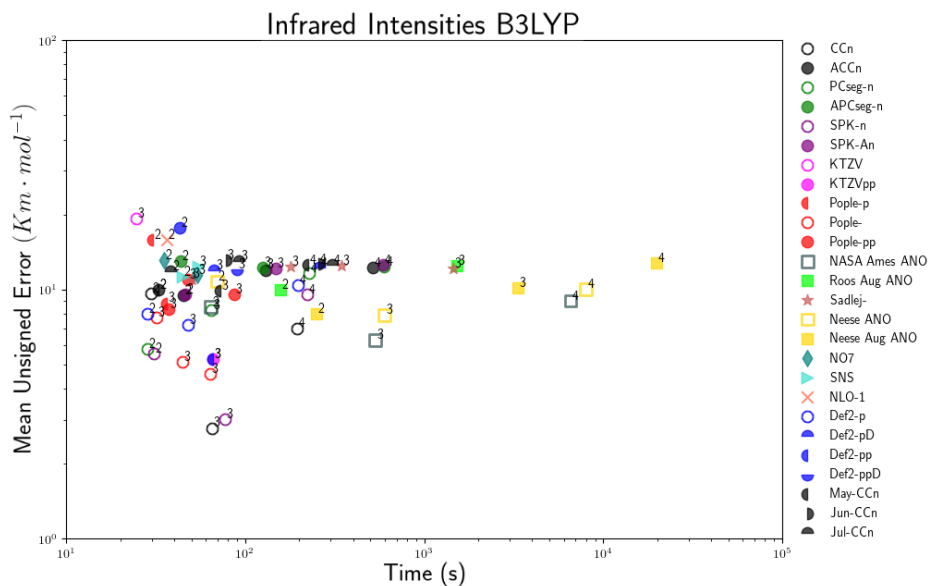


Figure 5.3: Mean Unsigned Error of B3LYP Infrared Intensity Calculations Compared to Experimental Data

Both the harmonic and anharmonic calculation of the infrared frequencies is subject to the accuracy of the initial Hessian used to project the geometries, as well as the accuracy of the

level of theory and basis set used to calculate the energetics and dipole moment. As explained previously, the MP2 and CCSD(T) levels of theory are not as reliable at geometries required for anharmonic, or even harmonic calculations. As a result, infrared intensities are not as dependent upon diffuse and polar functions as Raman activities, but rather how accurately the relative energies are calculated at each geometry, which explains why some basis sets that do not incorporate large polarization and diffuse functions are only necessary to the extent that the energetics and dipole moment are calculated with sufficient accuracy. Table 4 shows that by utilizing SCS-MP2[136] or CCSD(2)_T[137] levels of theory (the former at no additional cost, the latter at twice the cost of regular CCSD(T) calculations), further improvements in the frequencies and IR intensities can be obtained.

5.3.5 Raman Activities

In contrast to the IR intensities, the Raman activities can be very accurately calculated under the harmonic approximation, in particular when the MP2 level of theory is used (in agreement with previous studies[85, 86]) and are extremely sensitive to the degree of polarization and diffusivity incorporated into the basis sets.

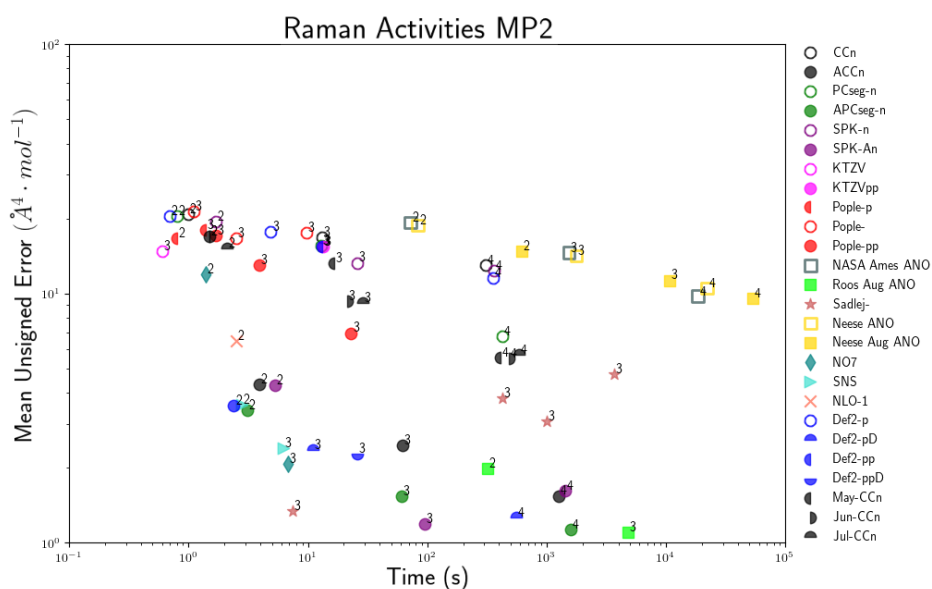


Figure 5.4: Mean Unsigned Error of MP2 Raman Activities Calculations Compared to Experimental Data

It is clear from Figure 4 that while increasing the zeta-level or overall size of basis set is helpful, those basis sets that incorporate higher polarization levels and more diffuse functions are generally more accurate at reproducing Raman activities. In particular, the spectroscopic basis sets all perform well here (e.g., Sadlej, N07, SNS). The (PPD versions of) Def2 and (aug) Roos basis sets perform very well here also. The NLO-1 basis set also performs fairly well in calculating Raman activities, comparable to many quadruple-zeta basis sets. The dependence on the level of diffusivity is well demonstrated from the correlation consistent and calendar basis sets (cc→ may→ jun→ jul→ aug).

5.4 Conclusion

From our investigation we have identified various trends for the predictions of water’s molecular properties, of which the most notable is the Def2-n basis set family’s overall performance. Pople-style basis sets tend to be very consistent with their predictions, typically producing minimal changes between each individual basis set’s calculation. Pople basis sets are a reasonable choice for exploration, but there is little systematic improvement within their family, and are not recommended when results require high accuracy or quantitative values. Although they are occasionally somewhat accurate for infrared intensities when used with the B3LYP level of theory, they are not recommended to predict Raman activities. The Aug-CCn family is good for predicting Raman activities, particularly with the MP2 level of theory, however, they are computationally heavy, and outperformed by less computationally heavy options, such as aug-PCn, Def2-PD, and SPK augmented versions, as well as the majority of spectroscopic basis sets in this area. These basis sets have been shown to produce results comparable to six zeta basis sets, with minimal runtimes for the calculation of many optical properties. Another notable trend is the NO7 and SNS basis families success with calculating vibrational frequencies with harmonic approximation methods. Regarding the ANO series of basis sets, the NASA and Neese series offer high accuracy for geometries, as well as vibrational frequencies and infrared intensities, but lack sufficient diffuse functions to

be able to accurately determine Raman activities (even the Aug-Neese sets). Only the Roos augmented ANO basis sets were able to accurately reproduce Raman activities, and represent an excellent all-round choice. We noticed that for the calculation of Raman activities basis sets which incorporated large numbers of diffuse functions always provided superior accuracy, however our scope was limited to the water molecule and further investigations need to be conducted. We are currently increasing the scope of this study to include various other molecular species in the hopes of validating trends found in this study.

6 A COMPUTATIONAL INVESTIGATION OF THE EQUILIBRIUM GEOMETRIES, ENERGETICS, VIBRATIONAL FREQUENCIES, INFRARED INTENSITIES AND RAMAN ACTIVITIES OF C_2O_y ($y = 3, 4$) SPECIES

This chapter (Chapter 6) contains previously published content. The usage of this content is approved by the journal's copyright policy, which states authors can include their publications in full or in part in a thesis or dissertation.

REF: Brian C Ferrari and Chris J Bennett. “A computational investigation of the equilibrium geometries, energetics, vibrational frequencies, infrared intensities and Raman activities of C_2O_y ($y= 3, 4$) species”. In: *Molecular Physics* (2020), e1837404

6.1 Introduction

High energy density materials (HEDMs) are highly sought after by defense and space researchers for their potential as explosives and propellants. A material is considered an HEDM if it exhibits a high “thrust-to-weight ratio”, i.e. if the chemical potential energy to specific weight ratio is large [139]. Recently, cyclic C_xO_y compounds have received a lot of attention as potential HEDMs [140, 141, 142, 143, 144, 145], particularly in the form of dimers or tetramers of carbon monoxide (CO) and carbon dioxide (CO_2). The strength of the CO

triple bond (255.79 kcal/mol) surpasses that of the N_2 triple bond (225.06 kcal/mol) [146], which means CO is one of the strongest bonded neutral diatomic molecules. As such, any meta-stable molecule that dissociates to form multiple CO molecules could in principle be an HEDM. Similarly, the double bonds are also fairly strong within CO_2 (125.75 kcal/mol) and molecular oxygen (O_2 ; 119 kcal/mol) [146]. More generally, a molecule that fragments into any combination of O_2 , CO and CO_2 species are therefore likely to be potential HEDMs.

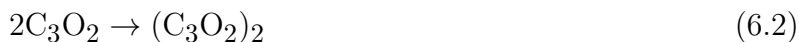
In addition, the possibility of meta-stable C_2O_3 or C_2O_4 species existing has recently become of interest to astrochemists and planetary scientists due to the recent detection of O_2 in the coma of comet 67P/Churyumov-Gerasimenko (hereafter 67P) [22] and subsequent confirmation of this species also present in the coma of comet 1P/Halley [147]. The origin of O_2 in these in cometary comae is still under discussion. Several explanations to its origin include formation either via photolysis or radiolysis of water ice [148, 22], the dismutation of a parent species [149] or Eley–Rideal reaction mechanism [150]. Recently, an alternative hypothesis was suggested whereby the detected O_2 originates from the dissociation of a parent carbon oxide species released into the comae, which could include C_2O_3 [31], or C_2O_4 species. The origin of a parent carbon oxide species likely comes from the irradiation of CO/ CO_2 ices present in both interstellar and cometary ices. Carbon oxide species are commonly formed at a few-% level abundances when CO and CO_2 ices are exposed to radiation. Irradiation of CO_2 ice has been shown to yield several higher-order carbon oxides: CO_3 (C_{2v}) [32], CO_3 (D_{3h}) [40], CO_4 (C_{2v}) [41], CO_5 (C_2) [42] and CO_6 (C_s) [43]. The presence of carbon oxides involving multiple carbon atoms could be present within radiation ices, but spectroscopic data to verify or discount their formation has not been available. Thus, both C_2O_3 and C_2O_4 could reside within the irradiated ices that constitute comet 67P, which may then subsequently sublime and form the O_2 observed upon dissociation by energetic particles or photons [31].

Here we outline a brief history of the studies on C_2O_3 and C_2O_4 compounds, detailing reasons for interest in them beyond their value as HEDMs and also possibilities for their

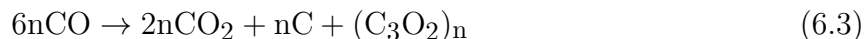
experimental detection. We present tabular data on their harmonic and anharmonic vibrational frequencies, infrared intensities, Raman activities and energetic studies on the heats of formation, as well as bond dissociation reactions. We examine the effects of different exchange correlation methods on these meta-stable molecules. The T1 diagnostic, largest T2 amplitudes, %TAE(T), and A_λ for each molecule is presented to evaluate the reliability of calculations relying on single-reference electronic configurations.

6.1.1 C_2O_3

Oxalic anhydride (C_2O_3) was initially suggested as an intermediate in the decomposition process of 2-chloro-2-oxo ethanoic acid (C_2O_2ClOH) by Staudinger in 1908 [151], although he was unsuccessful in observing it. In 1948 Mills *et al* [152] speculated that, under high pressures, solid carbon monoxide would polymerize into carbon suboxide (C_3O_2) while also producing oxalic anhydride. This was expected to take place via the following reaction:



Lipp *et al.* [153] investigated this phenomenon in 1998 and argued to the contrary by presenting the first Raman and IR absorption spectra of the compound, which indicated the CO formed a mixture of carbon dioxide, graphitic carbon, and polymerized carbon suboxide via the following reaction:



Oxalic anhydride has also been suggested to be an intermediate in the thermal decomposition of oxalates ($C_2O_4^{2-}$) [154] and in chemoluminescent reactions of oxalyl chloride ($C_2O_2Cl_2$) [155]. To date, neutral C_2O_3 has yet to be experimentally confirmed, although there was

fleeting evidence a neutral C_2O_3 species may have been generated from ion-neutralization reactions, though if so, it had a lifetime below 10^{-6} seconds [156, 157]. The cyclic dimer dioxane tetraketone (C_4O_6) has been observed by Strazzolini *et al.* in 1998 via NMR [158]. Its anion has also been experimentally observed by Zhou *et al.* in 2000 via IR spectroscopy [159]. Giving significant merit to the existence of the C_2O_3 molecule, yet it still remains undetected perhaps in part due to a lack of spectroscopic data on it. There has been some attention to the properties of C_2O_3 by computational chemists attempting to deduce the structure and stability of isomers/co-oligomers. Gambi *et al.* [144] investigated the cyclic co-oligomer of CO and CO_2 with C_{2v} symmetry (herein C_2O_3a). It was shown that C_2O_3a was kinetically stable at low temperatures, though it is susceptible to thermal decomposition, which seems supported by the low 0.1 kcal/mol activation energy calculated for its barrier to decomposition ($C_2O_3 \rightarrow CO + CO_2$) by Peppe *et al.* [157]. Gambi *et al.* [144] also provided the optimized geometry of the co-oligomer and the two largest vibrational frequencies and infrared intensities of the molecule. However, they left much of the molecule poorly characterized, particularly the components required for predicting the molecular spectra. In the recent study by Fortenberry *et al.* [31] where C_2O_3a was suggested as a potential source for O_2 in comet 67P, they also characterized a significant amount of properties for this species, including force and rotational constants, harmonic and anharmonic frequencies, infrared intensities and its dipole moment. One of the few remaining properties of the co-oligomer that remains to be characterized is the Raman activities, presented here. A second C_2O_3 isomer identified (herein C_2O_3b) has received relatively little attention from computational chemists and thus remains poorly characterized. Here we find that C_2O_3b has an imaginary frequency and thus is assumed to not be a real isolatable compound, at least on the singlet potential energy surface.

6.1.2 C₂O₄

Our focus in this paper is primarily on bonded oligomers of carbon dioxide and isomers of C₂O₄. Studies on the properties of charged and neutral carbon dioxide clusters held together by intermolecular forces, which slightly diverge from the topic of this paper can be found in Bukowski *et al.* [160].

Dioxetanedione (C₂O₄), originally suggested as an intermediate in a chemiluminescence reaction by Rauhut *et al.* in 1967 [155], and was experimentally observed by Cords *et al.* in 1969 [161]. Detection of a C₂O₄ molecule (exact structure unknown) was performed using a mass spectrometer, where the mass-to-charge-ratio (m/q) of 88 (C₂O₄) peak was correlated to a $m/q = 44$ peak twice the size. This was interpreted as the decomposition of dioxetanedione (C₂O₄) into two molecules of carbon dioxide (CO₂). Despite an experimental observation of the intermediate, the molecule remained poorly characterized, having only one frequency predicted by Rauhut *et al.* occurring around 1885cm⁻¹. In 1996, Lewars presented the first *ab initio* investigation of the carbon dioxide cyclic dimer of D_{2h} symmetry (herein C₂O_{4c}) [141]. They identified that C₂O_{4c} is a minima on the potential energy hypersurface and obtained optimized geometries, vibrational frequencies and relative infrared intensities for this structure. Lewars also calculated the barrier of fragmentation at the HF/6-31G level of 50kJ mol⁻¹ for C₂O_{4c}; which is relatively low for an average molecule but quite high for a metastable species. Following this, Zhang *et al.* [162] published a complete basis set extrapolation energetic study on two possible carbon dioxide dimers they posed could be the intermediate. Investigating the C_{2v} dimer (herein C₂O_{4a}) and the D_{2h} isomer (herein C₂O_{4d}) they presented optimized geometries, vibrational frequencies and relative infrared intensities. They concluded that the C₂O_{4a} structure, which was favored by both CASSCF-MP2 and CBS-QB3 calculations, was the intermediate involved in the peroxyoxalate chemiluminescent system. Frapper & Saillard (2000) calculated the optimized geometry of C₂O_{4c} using DFT (B3LYP) [163]. It was also found that although the dimer was less stable than its monomer (having 24 kcal mol⁻¹ higher energy per CO₂ than the

monomer) the dimer’s existence could be feasible. In 2004, Bos *et al.* experimentally investigated the chemiluminescence intermediate using ^{13}C NMR, where they were able to confirm its presence during the reaction [164]. A follow-up to this study was performed by Tonkin *et al.* in 2008 where they varied the amount of reagent molar ratios in the reaction while monitoring with ^{13}C NMR [165]. They also found that excess amounts of hydrogen peroxide favored formation of 1,2-dioxetanedione, $\text{C}_2\text{O}_4\text{a}$. In 2010 Howart *et al.* studied the $\text{C}_2\text{O}_4\text{d}$ structure as well as the C_2O_4 C_{2v} isomer (herein $\text{C}_2\text{O}_4\text{b}$), where they found both molecules to be meta-stable relative to their dissociation limits [166]. They also found the optimized geometries and harmonic frequencies of the two molecules (finding that the $\text{C}_2\text{O}_4\text{b}$ isomer had one imaginary frequency, and thus is not an isolatable stable species), and investigated the possible low-lying excited states for both molecules. Recently Dunlap *et al.* (2013) extensively studied the $\text{C}_2\text{O}_4\text{c}$ and $\text{C}_2\text{O}_4\text{a}$ isomers, calculating optimized geometries, vibrational frequencies, infrared intensities and Raman activities [145]. They inferred that the structural stability of $\text{C}_2\text{O}_4\text{a}$ may be more dependent on pressure, rather than temperature. Despite the significant amount of work done on the various C_2O_4 species the molecular spectroscopic properties of these species are insufficient for experimental verification, and almost completely lacking for $\text{C}_2\text{O}_4\text{b}$. Here we extend the characterization of C_2O_4 species, providing information pertinent to their detection via infrared or Raman spectroscopy.

6.2 Computational Details

Calculations were performed using the 14 FEB 2018 (R1) version of the General Atomic and Molecular Electronic Structure System, GAMESS(US) package [129, 104, 128, 127]. Calculation progress was assisted by use of MacMolPlt [71], Multiwfn [167] and considerable automation of workflow was achieved through AutoGAMESS [168].

Geometry optimization and harmonic frequency calculations generally considered to be the most accurate in this study were performed using the completely renormalized coupled-cluster with a perturbative treatment of the triples contribution referred to as CR-CC(2,3),A,

or CCSD(2)_T. Compared to the ‘gold-standard’, CCSD(T), this level of theory is better suited for situations where multi-reference character or bond-breaking may be present [137]. The correct treatment of the potential energy surface utilized to generate vibrational frequencies relies on coordinates away from the equilibrium geometry which are anticipated to have increased multi-reference character, and therefore should be more accurately treated by this approach [83]. The more traditional CCSD and CCSD(T) levels of theory were also performed for comparison [169, 103]. In a similar vein, we also adopted both the regular and spin-component-scaled variants of 2nd order Møller-Plesset perturbation theory (MP2 [102] and SCS-MP2 [136], respectively). The latter has been shown to provide slightly increased accuracy when determining energies along potential energy surfaces, and predicting vibrational frequencies [170].

We also performed Density Functional Theory (DFT) type functionals from rungs 2 and 4 of Jacob’s ladder [171] which included: **i**) two generalized gradient approximation (GGA) functionals, PBE [172] and BLYP [173], **ii**) their global hybrid GGA counterparts, PBE0 (*a.k.a.*, PBE1PBE) [174, 175] and B3LYP [101, 100], and **iii**) the long-range corrected hybrid functional ω B97X-D [176]. Some additional calculations were performed using the M06-L[177] and M11-L[178] meta-GGA (rung 3) as well as the M06[179], M06-HF[180], and M11[181] hybrid meta-GGA (rung 4) Minnesota functionals which can be found in the *Supporting Information*. DFT approaches (particularly GGA and local functionals which do not incorporate Hartree-Fock exchange) should suffer less from the failings of single-reference wavefunction-based methods when accounting for dynamic and non-dynamic electron correlation effects [182, 183].

In general, calculations were carried out using Dunning’s Correlation Consistent augmented quadruple- ζ aug-cc-pV(Q+d)Z basis set [110], except for the coupled cluster calculations performed using the smaller cc-pV(T+d)Z basis set [108]; note that the “+d” notation indicates an additional tight d-function included in GAMESS(US) by default, and is used throughout, though we typically omit specifically stating it. CCSD(T)/cc-pVTZ calculations

are commonly utilized to benchmark other calculations [184]. Some calculations have been performed using additional basis sets, including Pople’s 6-311G(d) [185] to compare with previous works, Sadlej-pVTZ [186] due to its accuracy in predicting Raman activities, and the calendar basis set may-cc-pV(T+d)Z [112] which was found to predict vibrational frequencies close to those with the much larger cc-pV5Z basis set (most likely due to a cancellation of errors) in our previous study [83]. Further details on the basis set dependence and the performance of those selected can be found in the *Supporting Information*.

All calculations were run using spherical harmonics (with the exclusion of Pople style basis sets) with tight convergence criteria used throughout (e.g., gradient convergence 10^{-6} for geometries, integral cutoff 10^{-12} , primitive cutoff 10^{-25} , and 10^{-7} SCF convergence). DFT calculations were run using an “army grade” pruned grid (JANS=2) consisting of 71,000 grid points and 155 radial points per atom. Hessian calculations were run with semi-numerical (DFT and MP2) or fully numerical gradients [CCSD(T)] under the double harmonic approximation with the residual translations/rotations removed by projection. Raman calculations were run with similar semi-numerical approximations, but were not performed for coupled cluster calculations, since these are not currently supported in the version of GAMESS(US) used. The frozen-core approximation was implemented for all MP2 and CCSD(T) calculations, but is thought to not affect our results dramatically [187]. Basis sets used for molecular predictions were selected from a benchmark study done by our group prior to this work [83] and are discussed further in the *Supporting Information*.

Anharmonic calculations were performed using the Vibrational Self-Consistent Field (VSCF) [130] approach using mostly the default settings (direct, NCOUP=2, and NGRID=12) at the B3LYP/6-311G(d) level of theory, which gave satisfactory results.

Couple Cluster methods could not be used to calculate IR intensities or Raman activities in the version of GAMESS(US) used for this study. Therefore, DFT and MP2 methods were used, previously shown to produce reliable results [83].

The Raman activities (S_i) reported can be converted into Raman intensities (I_i) by using

equation 6.4 [188]; where ν_0 is the excitation laser wavenumber, ν_i is the normal mode wavenumber, T is the absolute temperature, h and k_B are Planck and Boltzmann’s constants, and c is the speed of light.

$$I_i = \frac{(\nu_0 - \nu_i)^4 S_i}{\nu_i (1 - e^{-\frac{hc\nu_i}{k_B T}})} \quad (6.4)$$

In order to assess the multi-reference character of our carbon oxide species, we performed several different diagnostics. The T1 diagnostic [189], the largest T2 amplitudes, the %TAE(T) and A_λ values are reported [190]. For each of these diagnostics, the optimized geometry at the B3LYP/cc-pVTZ was used. The T1 diagnostic, largest T2 amplitudes, and %TAE(T) were all taken from CCSD and CCSD(T) calculations with the cc-pVTZ basis set. The A_λ diagnostic is based on differences in the total atomization energies between local (GGA) and hybrid functionals that utilize some percentage of HF-exchange, and is described by Fogueri *et al.* [190]. It is calculated using equation 6.5, where XC is a pure-DFT exchange-correlation functional, and $X_\lambda C$ represents the corresponding hybrid with 100 λ % Hartree-Fock-type exchange (*i.e.*, BLYP \rightarrow B3LYP would be $A_{20\%}$ since B3LYP has 20% Hartree-Fock Exchange). We use a similar methodology to calculate $A_{25\%}$ for PBE (local GGA) vs. PBE0 (hybrid GGA with 25% HF exchange) and $A_{27\%}$ for M06-L (local meta-GGA) vs. M06 (hybrid meta-GGA with 27% HF exchange). To ensure consistency, all DFT based diagnostics were calculated with the cc-pVTZ basis set.

$$A_\lambda = \frac{1}{\lambda} \left(1 - \frac{\text{TAE}[X_\lambda C]}{\text{TAE}[XC]} \right) \quad (6.5)$$

Heats of formation were calculated using several composite method levels of theory, including G3MP2 [191, 192], CCCA-S4 and CCCA-CCL [193, 194], with the latter based upon the energetics calculated with the CR-CC(2,3),D level of theory which should provide high accuracy even for species which have significant multi-reference character.

6.3 Results and Discussion

The optimized geometries at the CCSD(2)_T/cc-pVTZ level of theory are shown in Figure 6.1 for all molecules considered within the current study (geometries for each species obtained at other levels of theory can be found in the *Supporting Information*). Table 6.1 shows the %TAE(T) and T1 diagnostics for all these species along with several other molecular species of interest, for comparison (notably, to [190]). Note that the %TAE(T) values in Table 6.1 are reported in percentage, while the $A_{20\%}$ [B3LYP], $A_{25\%}$ [PBE0] and $A_{27\%}$ [M06] values are reported as the decimal representation of the percentage. A T1 diagnostic value of greater than 0.02 is a general indicator of significant levels of non-dynamical correlation, and while it hovers just below this value for the carbon oxides, this by itself is not sufficient to indicate an issue. Similarly, multi-reference character is generally indicated by largest T2 amplitudes above 0.2, whereas again, this diagnostic only indicates a mild level of multi-reference character. However, ozone (O₃; also shown in Table 1) also does not raise any major concerns from such diagnostics, despite it being notorious example of species with multi-reference character [190]. In such cases, the %TAE(T) often offers a more reliable diagnostic whereby O₃ has one of the largest values. However, this diagnostic only reveals mild multi-reference character for the carbon oxides, again. Similarly, the $A_{25\%}$ [PBE0] values of around 0.3 are indicative of moderate non-dynamical correlation [190]. The $A_{20\%}$ [B3LYP] scale appears to tell a similar story, only the values seem to be about 20% lower. Lastly, the $A_{27\%}$ [M06-L] values seem to be consistently 3-4 times lower than the $A_{25\%}$ [PBE0] values, where a value of 0.3 would indicate a severe problems, but again none of the calculated values implicate anything more than a mild level of non-dynamical character for the carbon oxide species. Overall, the C₂O_{4c} molecule seems to have the least multi-reference character, whereas C₂O_{4d} may present the greatest concern. To conclude, these diagnostics indicate that the multi-reference character is sufficiently low/mild in most of these molecules that single-reference methods should provide reasonable accuracy; a topic that will be revisited

later.

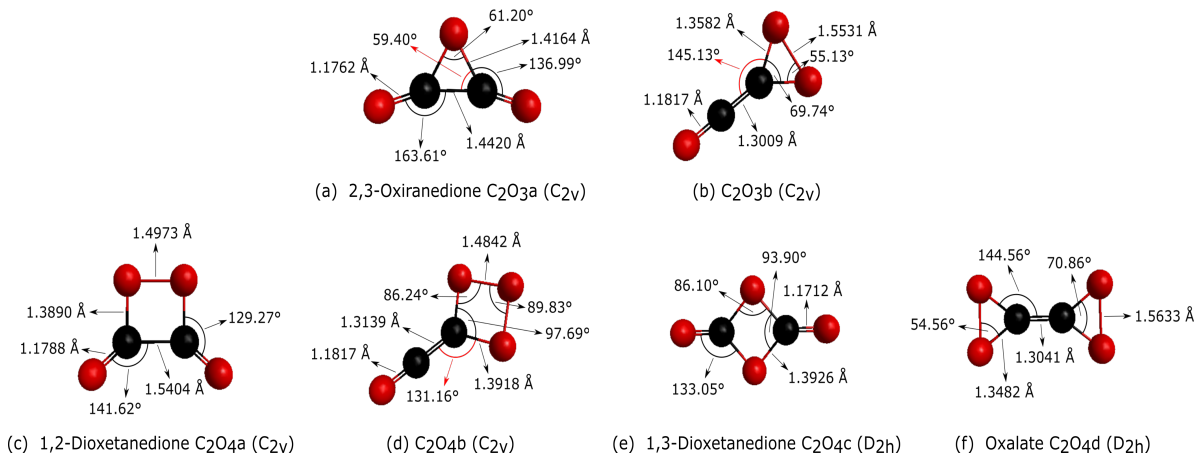


Figure 6.1: Optimized geometries of all reported molecules. Bond angles and bond lengths reported here are calculated at the CCSD(2)_T/cc-pVTZ level of theory.

Table 6.1: %TAE(T), T1 diagnostic, Largest T2 amplitudes, and several A_λ DFT-based diagnostics for the molecules of interest here as well as several well-studied species that exhibit varying degrees of multireference character (see text for details).

Species	%TAE(T)	T1 Diagnostic	Largest T2 Amplitude	$A_{20\%}$ [B3LYP]	$A_{25\%}$ [PBE0]	$A_{27\%}$ [M06]	Ref
H ₂ O	01.50	0.0070	-0.048	—	0.1160	—	b
H ₂ O	01.37	0.0066	-0.048	0.3960	0.1169	0.0369	a
CH ₄	00.67	0.0070	-0.035	-0.0428	0.0226	0.0354	a
HCN	02.87	0.0143	-0.062	0.1303	0.1856	0.1637	a
CO	03.10	0.0190	-0.067	—	0.2060	—	b
CO ₂	03.60	0.0180	-0.063	—	0.2440	—	b
O ₂	07.70	0.0070	-0.101	—	0.5510	—	b
O ₃	17.40	0.0270	-0.192	—	1.0070	—	b
O ₃	17.88	0.0259	-0.188	0.8838	0.9611	1.3768	a
FOOF	16.90	0.0260	-0.069	—	1.0980	—	b
C ₂ O ₃ (C _{2v})a	04.17	0.0194	-0.065	0.1917	0.2664	0.5543	a
C ₂ O ₃ (C _{2v})b	05.24	0.0175	-0.134	0.2628	0.3434	0.6869	a
C ₂ O ₄ (C _{2v})a	05.69	0.0191	-0.076	0.2202	0.2987	0.5888	a
C ₂ O ₄ (C _{2v})b	03.27	0.0171	-0.093	0.2918	0.3786	0.7267	a
C ₂ O ₄ (D _{2h})c	03.67	0.0168	-0.054	0.1559	0.2352	0.5038	a
C ₂ O ₄ (D _{2h})d	06.21	0.0162	-0.101	0.3084	0.3958	0.7738	a

^a Present Data

^b Fogueri *et al.* 2013 [190]

In the succeeding subsections, we present the tabulated data on harmonic and anharmonic vibrational frequencies, infrared intensities, and Raman activities.

6.3.1 C₂O₃a (C_{2v}) [CO + CO₂ Cyclic Co-Oligomer]

Table 6.2 summarizes the harmonic frequencies, anharmonic frequencies, infrared intensities, Raman activities and isotopic shifts for C₂O₃a (C_{2v}).

The harmonic frequencies and infrared intensities presented here show good agreement with previous studies on this species. In particular, CCSD(2)_T/cc-pVTZ strongly agrees with the recent high-level calculations from Fortenberry *et al.*[31]. Interesting to note is the strong agreement between harmonic frequencies predicted by B3LYP/6-311G(d) and B3LYP/aug-cc-pVQZ despite the large size difference in basis sets. This implies the B3LYP level of theory has essentially converged for this molecule at the 6-311G(d) basis set size, therefore we assume it can be used to calculate reliable anharmonic corrections. The observed basis set convergence of B3LYP and other DFT methods over MP2 and couple cluster calculations has already been well established [195]. In contrast, the Coupled Cluster type calculations (ie: CCSD, CCSD(T), CCSD(2)_T) show only slight changes in the predicted frequencies when the basis set is kept constant but level of theory increased (*i.e.*, CCSD/Sadlej-pVTZ → CCSD(T)/Sadlej-pVTZ → CCSD(2)_T/Sadlej-pVTZ). A similar trend is seen with the cc-pVTZ basis set, implying that basis set dependence has a greater effect than the degree of dynamical correlation, at least for this species. Further details on the basis set dependence for MP2 are provided in the *Supporting Information*, as well as evidence for why the cc-pVTZ basis set should provide reliable results. The vibrational frequencies predicted by B3LYP/aug-cc-pVQZ results are in good agreement with those from CCSD(2)_T/cc-pVTZ. SCS-MP2 results generally are more accurate than those from MP2. Both GGA functionals (PBE and BLYP) consistently underestimate harmonic frequencies, while two of the hybrid functionals (PBE0 and ωB97X-D) repeatedly overestimate them. These results can be tied to their inability to accurately describe the bond strengths, and is also apparent from their overestimation and underestimation of bond lengths, respectfully (see *Supporting Information*).

The infrared intensities reported here for C₂O₃a show moderate agreement between DFT and

MP2 level calculations, particularly the $\omega_2, \omega_3, \omega_7$ and ω_9 values hold larger discrepancies. We refer to our prior study [83], which argued MP2 level calculations were more accurate at predicting infrared intensities. The ω_2, ω_7 and ω_9 modes exhibit larger discrepancies in both infrared intensities and Raman activities over different levels of theory, including between SCS-MP2 and MP2 values. The hybrid functionals (B3LYP, PBE0, and ω B97X-D) each predicted Raman intensities that differed considerably from the SCS-MP2 and MP2 levels, whereas the GGA functionals predict closer values, particularly PBE. In contrast to the vibrational frequencies, the infrared intensities and Raman activities predicted by B3LYP do not converge at the 6-311G(d) basis set size, evidenced by larger discrepancies between basis sets. The aug-cc-pVQZ infrared intensities and Raman activities should be close to their basis set limit values. For the anharmonic calculations, we assume that the proportion of intensity change when transitioning from harmonic to anharmonic values is conserved regardless of the basis set used.

According to Table 6.1, only the T1 diagnostic indicates that non-dynamical character may be of concern, with the rest indicating only a mild-level, which should be handled reasonably well by all levels of theory in this study. Despite this, the DFT levels of theory struggled to reproduce the IR intensities predicted by MP2 or SCS-MP2 for this species considerably with the hybrid functionals, with ω B97X-D performing the worst.

Here, we briefly comment on the prospects for identification of this species where it may be produced by radiolysis alongside other carbon oxide species. In the infrared, the strongest band, ν_7 , occurs at 1842 cm^{-1} which should be distinct from nearby absorptions assigned to CO_3 (C_{2v}) and CO_6 (C_s) [32, 43, 47]. The ν_2 band at 1183 cm^{-1} is also sufficiently strong that it may be observable, but overlaps with bands assigned to CO_3 (D_{3h}) [40, 47]. The ν_1 band at 2091 cm^{-1} may also be detectable, but overlaps with the fundamental bands of ^{13}CO and C^{18}O . In Raman, the ν_3 band at 844 cm^{-1} is strongest, and the ν_1 and ν_2 bands are approximately half as active.

Table 6.2: Harmonic Frequencies, Anharmonic Frequencies, Infrared Intensities, Raman Activities and Isotopic Shifts for C₂O₃a (C_{2v}).

Level of theory	Harmonic Frequency, cm ⁻¹									Ref
	$\omega_1(A_1)$	$\omega_2(A_1)$	$\omega_3(A_1)$	$\omega_4(A_1)$	$\omega_5(A_2)$	$\omega_6(B_1)$	$\omega_7(B_2)$	$\omega_8(B_2)$	$\omega_9(B_2)$	
G2MP2-HF/6-31G(d)	2105	—	—	—	—	—	1865	—	—	b
CCSD(T)-F12/cc-pVTZ-F12	2124.5	1148.4	854.6	342.6	645.6	436.3	1868.0	662.9	343.1	c
BLYP/aug-cc-pVQZ	2037.1	1043.8	821.21	320.94	637.26	320.82	1790.14	612.06	239.65	a
B3LYP/6-311G(d)	2146.80	1144.81	858.74	345.66	669.90	346.73	1888.57	651.56	396.92	a
B3LYP/aug-cc-pVQZ	2128.07	1137.54	857.58	342.13	669.70	342.72	1876.58	650.16	400.63	a
PBE/aug-cc-pVQZ	2074.73	1077.79	834.10	325.87	644.69	320.16	1818.29	615.59	330.53	a
PBE0/aug-cc-pVQZ	2175.75	1184.75	875.05	351.98	682.39	348.41	1915.18	663.13	479.98	a
ω B97X-D/aug-cc-pVQZ	2174.32	1194.55	870.91	352.24	681.45	352.34	1921.01	666.17	485.21	a
MP2/aug-cc-pVQZ	2106.81	1099.68	856.66	341.76	671.82	346.35	1838.94	641.64	340.11	a
SCS-MP2/aug-cc-pVQZ	2113.32	1121.77	858.94	344.08	671.71	349.21	1850.72	646.64	381.00	a
CCSD/Sadlej-pVTZ	2153.78	1182.51	856.67	343.37	649.25	338.9	1899.23	642.44	464.48	a
CCSD(T)/Sadlej-pVTZ	2103.89	1122.47	833.77	329.5	624.8	323.35	1847.86	619.06	373.02	a
CCSD(T)/cc-pVTZ	2126.43	1151.82	852.34	343.45	658.22	345.35	1865.05	644.24	435.81	a
CCSD(2) _T /Sadlej-pVTZ	2109.05	1127.99	835.32	331.05	628.88	324.84	1853.80	621.49	383.04	a
CCSD(2) _T /cc-pVTZ	2133.17	1158.84	854.69	345.26	662.27	347.23	1872.42	647.19	446.08	a
	Infrared Intensity, km mol ⁻¹									
G2MP2-HF/6-31G(d)	114	—	—	—	—	—	943	—	—	b
MP2/6-31+G*	124	164	1	2	0	31	581	28	147	c
BLYP/aug-cc-pVQZ	105.09	164.98	2.91	0.78	0.0	27.73	609.44	26.40	111.89	a
B3LYP/6-311G(d)	89.69	208.96	7.23	1.35	0.0	30.36	663.07	40.00	83.87	a
B3LYP/aug-cc-pVQZ	112.11	191.02	5.52	1.73	0.0	33.22	707.43	41.08	101.01	a
PBE/aug-cc-pVQZ	99.75	156.58	4.31	0.98	0.0	25.89	609.22	33.59	99.62	a
PBE0/aug-cc-pVQZ	109.15	187.74	6.90	2.21	0.0	32.39	724.24	54.74	82.02	a
ω B97X-D/aug-cc-pVQZ	110.44	197.87	6.91	2.30	0.0	33.44	755.53	55.89	83.57	a
MP2/aug-cc-pVQZ	115.22	157.80	0.70	1.41	0.0	28.89	597.72	26.45	140.21	a
SCS-MP2/aug-cc-pVQZ	112.31	171.86	1.11	1.62	0.0	30.27	615.24	29.65	129.70	a
	Raman Activity, Å ⁴ amu ⁻¹									
BLYP/aug-cc-pVQZ	19.297	9.563	18.733	3.655	0.352	0.043	5.537	1.625	2.975	a
B3LYP/6-311G(d)	10.843	11.083	10.695	3.492	0.720	0.006	4.081	2.912	1.044	a
B3LYP/aug-cc-pVQZ	15.682	9.069	18.544	3.394	0.454	0.017	4.692	1.757	2.407	a
PBE/aug-cc-pVQZ	18.067	9.291	19.894	3.542	0.334	0.043	4.921	1.794	2.666	a
PBE0/aug-cc-pVQZ	14.812	8.893	17.885	3.237	0.461	0.014	4.162	2.055	1.901	a
ω B97X-D/aug-cc-pVQZ	15.147	9.862	15.352	3.225	0.483	0.009	4.061	1.998	1.979	a
MP2/aug-cc-pVQZ	13.542	15.700	25.045	4.193	0.117	0.051	6.801	1.607	2.754	a
SCS-MP2/aug-cc-pVQZ	16.952	13.641	20.749	4.058	0.219	0.030	7.152	1.676	2.554	a
Level of theory	$\nu_1(A_1)$	$\nu_2(A_1)$	$\nu_3(A_1)$	$\nu_4(A_1)$	$\nu_5(A_2)$	$\nu_6(B_1)$	$\nu_7(B_2)$	$\nu_8(B_2)$	$\nu_9(B_2)$	Ref
	VSCF Anharmonic Frequency, cm ⁻¹									
CCSD(T)-F12/cc-pVTZ-F12	2084.4	1155.6	839.4	337.2	630.5	369.1	1857.8	642.2	339.5	c
B3LYP/6-311G(d)	2104.34	1168.1	847.52	340.37	662.32	344.53	1857.23	643.44	290.41	a
B3LYP/cc-pVTZ	2094.47	1167.33	848.01	338.68	662.16	342.96	1849.55	642.13	292	a
CCSD(2) _T /cc-pVTZ	2091.04	1182.86	843.71	340.10	655.11	345.28	1841.45	639.19	326.40	a
	¹³ C ₂ O ₃ a VSCF Anharmonic Frequency, cm ⁻¹									
CCSD(2) _T /cc-pVTZ	2030.42	1159.74	833.31	338.40	632.03	336.48	1800.89	631.09	317.41	a
$\Delta\nu$	60.63	23.12	10.39	1.69	23.09	8.80	40.56	8.10	8.99	a
	¹² C ₂ ¹⁸ O ₃ a VSCF Anharmonic Frequency, cm ⁻¹									
CCSD(2) _T /cc-pVTZ	2061.30	1150.35	808.36	322.88	651.37	338.54	1796.41	615.31	319.95	a
$\Delta\nu$	29.74	32.51	35.35	17.22	3.74	6.74	45.04	23.87	6.45	a
	VSCF Infrared Intensity, km mol ⁻¹									
B3LYP/6-311G(d)	88.88	204.89	7.22	1.39	0.0	29.77	656.64	39.61	70.38	a
B3LYP/cc-pVTZ	94.03	194.45	6.17	1.58	0.0	30.41	657.59	38.68	78.46	a
SCS-MP2/aug-cc-pVQZ	111.30	168.77	1.11	1.67	0.0	29.65	609.31	29.37	108.63	a

[†] The light blue shaded rows indicate approximate anharmonic corrections to the level of theory indicated after multiplying by the ratio of the anharmonic to harmonic values calculated at the B3LYP/6-311G(d) level of theory.

^a Present Data

^b Gambi *et al.* 2001 [144]

^c Fortenberry *et al.* 2019 [31]

6.3.2 C₂O₃b (C_{2v}) [Isomer]

The C₂O₃b (C_{2v}) molecule was not found to be a stable minimum on the singlet potential energy surface, whereby an imaginary vibrational frequency was found at every level of theory investigated. At the highest level considered, CCSD(2)_T/cc-pVTZ, the harmonic vibrational frequencies ω_1 - ω_9 were found to be 2277.6 (A₁), 1580.48 (A₁), 883.73 (A₁), 603.66 (A₁), 1074.93 (B₁), 632.37 (B₁), 219.26 (B₁), 327.12 (B₂), and -311.96 (B₂) cm⁻¹,

respectively. A full description of the harmonic frequencies, the infrared intensities and the Raman activities for C_2O_3b (C_{2v}) can be found in the *Supplementary Information* section. We briefly investigated the triplet state, but found it to also have one or more imaginary frequencies; as such, it will not be discussed further.

The largest variations are observed for the ω_9 mode, with the softest imaginary frequency predicted by $\omega B97X-D$ and the hardest frequency from MP2 and CCSD variants, particularly when the Sadlej-pVTZ basis set is employed. For the remaining modes, similar trends are observed as with all other species. Compared to the CCSD(2)_T/cc-pVTZ calculations, BLYP and PBE strongly underestimate vibrational frequencies, while PBE0 and $\omega B97X-D$ strongly overestimate them. Although B3LYP provides more accurate estimates, it still does not reach the accuracy of MP2, or SCS-MP2 which both perform well. The coupled cluster results in general show basis set dependence, and would benefit further study.

Regarding the infrared intensities, the two strongest bands (ω_1 and ω_2) were predicted to be considerably stronger by DFT methods, in particular, the hybrid GGA functionals. A similar trend is observed for the most active Raman modes (ω_1, ω_3 and ω_4), whose activities are generally overestimated by DFT.

Table 6.1, predicts that C_2O_3b (C_{2v}) should have a moderate level of non-dynamical character, yet most levels of theory were able to provide reasonably accurate descriptions of this molecule. The geometries had the 2nd largest discrepancies of the six species studied here.

6.3.3 C_2O_4a (C_{2v}) [CO_2 Dimer]

Table 6.3 summarizes the harmonic frequencies, anharmonic frequencies, infrared intensities, Raman activities and isotopic shifts for C_2O_4a (C_{2v}). According to table 6.1, this species lies fairly low on the anticipated level of non-dynamic character by most diagnostics, with the exception of %TAE(T). The geometries and vibrational frequencies were all calculated with reasonable agreement between different levels of theory, while the infrared intensities and Raman activities showed more disparity.

Our predicted vibrational frequencies are in general agreement with those from previous works [162, 145]. Most levels of theory considered are in good agreement with those from CCSD(2)_T/cc-pVTZ with the notable exceptions of PBE and BLYP, the latter often underestimating frequencies by $> 100 \text{ cm}^{-1}$. Here, the SCS-MP2/aug-cc-pVQZ frequencies are considerably closer than MP2/aug-cc-pVQZ to those predicted by CCSD(2)_T/cc-pVTZ

For this species, there were often large discrepancies between the infrared intensities and Raman activities between different levels of theory compared to values predicted at the SCS-MP2/aug-cc-pVQZ level. Here, BLYP and PBE often underestimate infrared intensities and overestimate Raman activities for C₂O_{4a} (C_{2v}) more drastically than their hybrid GGA counterparts.

Detection of this species via infrared spectroscopy would be most likely from the three approximately equally intense bands at 1962 (ν_1), 1179 (ν_2), and 1913 (ν_9) cm^{-1} . Attempts to detect this species in the presence of other carbon oxides may be difficult due to their overlap with bands from CO₄(C_{2v}) [41], CO₃(D_{3h}) [40], and CO₅(C₂) [42], respectively. In Raman spectroscopy, ν_1 is very strong and the ν_9 could also be detectable alongside ν_3 at 817 cm^{-1} .

Table 6.3: Harmonic Frequencies, Anharmonic Frequencies, Infrared Intensities, Raman Activities and Isotopic Shifts for C₂O₄a (C_{2v}).

Level of theory	$\omega_1(A_1)$	$\omega_2(A_1)$	$\omega_3(A_1)$	$\omega_4(A_1)$	$\omega_5(A_1)$	$\omega_6(A_2)$	$\omega_7(A_2)$	$\omega_8(B_1)$	$\omega_9(B_2)$	$\omega_{10}(B_2)$	$\omega_{11}(B_2)$	$\omega_{12}(B_2)$	Ref
Harmonic Frequency, cm ⁻¹													
B3LYP/6-311G(d,p)	1987	1165	838	786	308	779	238	401	1965	939	743	603	^b
CCSD/cc-pVTZ	2039	1236	907	826	323	798	235	415	1993	1001	759	625	^c
BLYP/aug-cc-pVQZ	1861.81	1087.45	758.45	669.06	291.21	741.77	234.55	381.52	1852.64	889.5	704.05	557.29	^a
B3LYP/6-311G(d)	1986.49	1165.08	837.92	785.94	307.98	778.75	238.67	399.95	1964.55	938.93	743.09	602.81	^a
B3LYP/aug-cc-pVQZ	1973.75	1162.59	855.83	786.56	308.31	781.71	240.78	400.18	1951.13	939.31	744.68	602.09	^a
PBE/aug-cc-pVQZ	1897.70	1121.86	783.57	716.239	294.33	746.67	238.67	380.23	1881.17	929.16	714.03	567.35	^a
PBE0/aug-cc-pVQZ	2018.63	1191.54	925.47	811.63	311.56	793.04	243.53	402.82	1989.38	964.36	758.49	615.47	^a
ω B97X-D/aug-cc-pVQZ	2030.62	1207.24	951.60	813.09	316.26	794.84	245.08	407.94	1998.64	981.81	764.11	621.38	^a
MP2/aug-cc-pVQZ	1941.43	1178.27	821.02	785.94	310.64	773.76	228.63	400.89	1911.34	962.87	731.1	599.07	^a
SCS-MP2/aug-cc-pVQZ	1955.5	1187.31	809.62	793.94	314.14	776.83	227.9	405.35	1921.66	970.15	735.5	604.43	^a
CCSD(2) _T /cc-pVTZ	1986.42	1203.11	840.23	800.57	313.81	777.60	231.37	404.63	1945.12	972.40	738.11	606.50	^a
Infrared Intensity, km mol ⁻¹													
CCSD/cc-pVTZ	250.08	282.08	7.48	7.22	2.40	0.0	0.0	35.82	329.57	11.78	0.24	38.72	^c
BLYP/aug-cc-pVQZ	216.81	189.11	0.47	13.14	1.25	0.0	0.0	29.05	277.75	2.93	0.15	25.15	^a
B3LYP/6-311G(d)	244.92	260.79	5.30	9.18	1.50	0.0	0.0	33.20	333.19	7.87	0.22	32.93	^a
B3LYP/aug-cc-pVQZ	270.07	247.77	7.76	9.34	1.68	0.0	0.0	35.21	347.11	8.22	0.30	35.25	^a
PBE/aug-cc-pVQZ	218.47	188.82	0.01	12.30	1.30	0.0	0.0	26.96	276.23	3.14	0.17	24.70	^a
PBE0/aug-cc-pVQZ	278.49	254.78	8.63	4.19	1.62	0.0	0.0	34.44	356.06	9.78	0.42	34.61	^a
ω B97X-D/aug-cc-pVQZ	290.09	267.00	11.11	4.66	1.91	0.0	0.0	35.52	380.84	9.96	0.46	38.23	^a
MP2/aug-cc-pVQZ	256.23	260.41	2.51	19.14	1.75	0.0	0.0	32.38	266.79	3.01	0.05	36.63	^a
SCS-MP2/aug-cc-pVQZ	254.73	270.12	0.44	20.05	2.02	0.0	0.0	34.01	275.07	3.86	0.09	38.50	^a
Raman Activity, Å ⁴ amu ⁻¹													
CCSD/cc-pVTZ	28.4	2.3	13.4	10.5	3.5	0.3	0.0	0.1	14.5	3.8	3.4	0.0	^c
BLYP/aug-cc-pVQZ	46.839	2.291	24.424	14.618	4.555	0.251	0.001	0.022	20.458	6.517	3.655	0.695	^a
B3LYP/6-311G(d)	24.407	2.277	19.588	3.63	4.023	0.881	0.009	0.178	11.898	4.796	3.856	0.037	^a
B3LYP/aug-cc-pVQZ	36.783	2.326	21.127	5.787	3.722	0.375	0.002	0.056	15.292	4.556	3.254	0.078	^a
PBE/aug-cc-pVQZ	45.808	2.353	26.68	7.889	4.427	0.214	0.0	0.02	18.493	5.357	3.738	0.393	^a
PBE0/aug-cc-pVQZ	36.987	2.727	17.428	8.152	3.550	0.370	0.0	0.059	14.226	3.663	3.299	0.007	^a
ω B97X-D/aug-cc-pVQZ	34.091	3.144	14.848	7.565	3.606	0.395	0.0	0.073	13.950	3.439	3.212	0.005	^a
MP2/aug-cc-pVQZ	28.142	3.111	11.335	10.768	4.419	0.069	0.003	0.016	15.542	3.966	3.962	0.014	^a
SCS-MP2/aug-cc-pVQZ	29.52	2.675	12.223	5.009	4.077	0.152	0.002	0.037	16.687	3.976	3.919	0.012	^a
Level of theory	$\nu_1(A_1)$	$\nu_2(A_1)$	$\nu_3(A_1)$	$\nu_4(A_1)$	$\nu_5(A_1)$	$\nu_6(A_2)$	$\nu_7(A_2)$	$\nu_8(B_1)$	$\nu_9(B_2)$	$\nu_{10}(B_2)$	$\nu_{11}(B_2)$	$\nu_{12}(B_2)$	Ref
VSCF Anharmonic Frequency, cm ⁻¹													
B3LYP/6-311G(d)	1962.22	1141.44	815.04	767.26	304.85	769.79	236.79	397.85	1932.34	912.52	732.33	596.29	^a
CCSD(2) _T /cc-pVTZ	1962.15	1178.70	817.29	781.54	310.62	768.66	229.55	402.52	1913.23	945.05	727.42	599.94	^a
¹³ C ₂ O ₄ a VSCF Anharmonic Frequency, cm ⁻¹													
CCSD(2) _T /cc-pVTZ	1911.70	1144.40	816.34	774.21	309.73	742.48	229.38	391.34	1866.18	925.93	719.23	594.51	^a
$\Delta\nu$	50.45	34.30	0.94	7.33	0.89	26.19	0.17	11.17	47.05	19.12	8.19	5.43	^a
¹⁸ O ₄ a VSCF Anharmonic Frequency, cm ⁻¹													
CCSD(2) _T /cc-pVTZ	1924.98	1162.28	771.95	745.53	293.97	763.00	216.59	395.94	1873.76	921.12	695.64	572.55	^a
$\Delta\nu$	37.17	16.42	45.34	36.01	16.65	5.66	12.96	6.58	39.47	23.93	31.78	27.39	^a
VSCF Infrared Intensity, km mol ⁻¹													
B3LYP/6-311G(d)	242.08	258.01	4.52	9.21	1.48	0.0	0.0	32.75	330.55	7.74	0.23	32.54	^a
SCS-MP2/aug-cc-pVQZ	251.77	267.23	0.37	20.11	1.98	0.0	0.0	33.55	272.89	3.80	0.09	38.03	^a

[†] The light blue shaded rows indicate approximate anharmonic corrections to the level of theory indicated after multiplying by the ratio of the anharmonic to harmonic values calculated at the B3LYP/6-311G(d) level of theory.

- ^a Present Data
- ^b Zhang *et al.* 2000 [162]
- ^c Dunlap *et al.* 2013 [145]

6.3.4 C₂O₄b (C_{2v}) [Isomer]

According to Table 6.1, traditional diagnostics for non-dynamical character do not predict that this molecule should be problematic. In contrast, the DFT-based diagnostics show that this species has moderate non-dynamical character. At every level of theory considered, this species was also found to have an imaginary frequency and is therefore not considered a stationary isolatable species (at least on the singlet potential energy surface). Preliminary investigations on the triplet potential energy surface also resulted in one or more imaginary frequencies. Tables of the geometries, harmonic frequencies, infrared intensities, and Raman activities for each level of theory considered here can be found in the *Supporting Information*. To summarize, these results demonstrate how poorly the DFT results - particularly for functionals with HF exchange - agree with those from other calculations of the geometries, harmonic frequencies, infrared intensities and Raman activities, whereby the predictions from PBE were found to be in the best agreement with Møller-Plesset and Coupled Cluster calculations.

The harmonic vibrational modes ω_1 - ω_{12} were predicted to lie at 2237.6 (A₁), 1461.06 (A₁), 930.28 (A₁), 878.46 (A₁), 601.34 (A₁), 1244.61 (B₁), 884.73 (B₁), 624.69 (B₁), 194.33 (B₁), 326.71 (B₂), 147.80 (B₂), and -179.99 (B₂) cm⁻¹ at the CCSD(2)_T/cc-pVTZ level of theory, respectively.

6.3.5 C₂O₄c (D_{2h}) [CO₂ Cyclic Dimer]

Table 6.4 summarizes the harmonic frequencies, anharmonic frequencies, infrared intensities, Raman activities and isotopic shifts for C₂O₄c (D_{2h}). Overall, this species was predicted to have one of the lowest levels of non-dynamical character considered here (Table 6.1), and accordingly was found to have the least discrepancies between the calculations of the geometries, harmonic frequencies, infrared intensities and Raman activities overall. This would seem to indicate that at the very least, the A_λ DFT diagnostic approach can help determine whether hybrid GGAs can reliably produce accurate results when other methodologies

are prohibitively computationally expensive where they generally provided far superior results to their GGA counterparts, particularly in the prediction of geometries and harmonic frequencies.

For the reported vibrational frequencies for C_2O_4c , we see the $CCSD(T)/cc-pVTZ \rightarrow CCSD(2)_T/cc-pVTZ$ has only slight changes in predicted values, whereas from $CCSD/cc-pVTZ \rightarrow CCSD(T)/cc-pVTZ$ has much larger discrepancies. This indicates the inclusion of a perturbative treatment of triples is necessary to accurately describe this species. The vibrational frequencies typically showed only a slight increase in vibrational frequency from $MP2 \rightarrow SCS-MP2 \rightarrow CCSD(T) \rightarrow CCSD(2)_T$. Considering $MP2$ and $SCS-MP2$ show good agreement with $CCSD(2)_T$, we believe that the $MP2$ level calculations should provide reliable infrared and Raman predictions. It is also interesting to note the high level of agreement between $B3LYP$ and $SCS-MP2$ levels of theory.

In the infrared intensities we see strong agreement between $MP2$ and $SCS-MP2$, with $SCS-MP2$ expected to be slightly more reliable based on geometries and frequencies being closer to those determined at the higher level of $CCSD(2)_T$. We note that previous $CCSD/cc-pVTZ$ calculations performed by Dunlap *et al.* [145] predict significantly different infrared intensities and Raman activities than our $SCS-MP2$ values, where the former are likely due to deficiencies in the level of theory (*i.e.*, not taking into account the triples), the latter differences can be attributed to the small basis set used in their calculations (see *Supporting Information*).

Table 6.4: Harmonic Frequencies, Anharmonic Frequencies, Infrared Intensities, Raman Activities and Isotopic Shifts for C₂O₄c (D_{2h}).

Level of theory	$\omega_1(A_g)$	$\omega_2(A_g)$	$\omega_3(A_g)$	$\omega_4(B_1u)$	$\omega_5(B_1u)$	$\omega_6(B_2u)$	$\omega_7(B_2u)$	$\omega_8(B_3u)$	$\omega_9(B_3u)$	$\omega_{10}(B_{2g})$	$\omega_{11}(B_{3g})$	$\omega_{12}(B_{3g})$	Ref
Harmonic Frequency, cm ⁻¹													
MP2(FC)/6-31G*	2101	1055	732	1953	905	1298	403	801	180	673	669	641	b
CCSD/cc-pVTZ	2158	1113	765	1992	948	1352	422	853	188	708	770	668	c
BLYP/aug-cc-pVQZ	2010.74	1001.76	705.03	1857.53	836.63	1191.92	380.55	775.06	182.04	657.87	618.58	481.32	a
B3LYP/6-311G(d)	2120.68	1078.65	745.29	1962.43	910	1287.32	407.05	825.35	186.16	685.95	651.72	612.12	a
B3LYP/aug-cc-pVQZ	2102.98	1073.82	741.84	1940.71	904.64	1277.16	405	825.58	186.87	690.51	652.86	620.52	a
PBE/aug-cc-pVQZ	2039.02	1025.98	713.8	1887.17	866.04	1232.71	388.48	784.31	180.16	667.08	623.49	549.19	a
PBE0/aug-cc-pVQZ	2142.58	1105.26	754.7	1979.86	940.09	1327.8	416.52	841.79	189.01	704.65	697.1	659.56	a
ω B97X-D/aug-cc-pVQZ	2147.71	1110.97	756.57	1982.91	945.86	1334.64	419.44	846.08	188.24	707.04	713.17	664.63	a
MP2/aug-cc-pVQZ	2085.8	1056.07	736.17	1931.15	904.76	1277.27	407.78	817.19	183.1	696.01	656.36	647.05	a
SCS-MP2/aug-cc-pVQZ	2095.72	1064.42	741.45	1939.58	908.96	1287.68	409.58	821.43	184.3	694.91	683.37	651.23	a
CCSD(T)/cc-pVTZ	2110.21	1075.48	746.11	1949.08	913.32	1312.22	409.82	826.61	184.96	691.04	712.08	650.61	a
CCSD(2) _T /cc-pVTZ	2115.92	1080.28	748.44	1954.49	917.54	1316.88	411.33	829.92	185.35	693.85	719.16	652.91	a
Infrared Intensity, km mol ⁻¹													
MP2(FC)/6-31G*	[0] ¹	[0]	[0]	[100]	[22]	[32]	[0.3]	[4]	[0.5]	[0]	[0]	[0]	b
CCSD/cc-pVTZ	0.0	0.0	0.0	1256.2	223.2	324.1	4.5	48.7	6.0	0.0	0.0	0.0	c
BLYP/aug-cc-pVQZ	0.0	0.0	0.0	1143.66	200.76	279.42	1.59	32.48	5.46	0.0	0.0	0.0	a
B3LYP/6-311G(d)	0.0	0.0	0.0	1216.04	196.76	352.89	2.86	48.24	5.63	0.0	0.0	0.0	a
B3LYP/aug-cc-pVQZ	0.0	0.0	0.0	1284.19	214.48	312.41	3.14	42.89	6.48	0.0	0.0	0.0	a
PBE/aug-cc-pVQZ	0.0	0.0	0.0	1140.68	191.12	273.06	2.19	30.34	5.66	0.0	0.0	0.0	a
PBE0/aug-cc-pVQZ	0.0	0.0	0.0	1306.56	206.68	312.04	4.12	42.59	6.79	0.0	0.0	0.0	a
ω B97X-D/aug-cc-pVQZ	0.0	0.0	0.0	1355.66	218.00	318.85	4.41	44.34	6.85	0.0	0.0	0.0	a
MP2/aug-cc-pVQZ	0.0	0.0	0.0	1196.60	247.43	298.51	2.42	37.28	5.35	0.0	0.0	0.0	a
SCS-MP2/aug-cc-pVQZ	0.0	0.0	0.0	1213.37	247.68	303.32	2.77	39.42	5.48	0.0	0.0	0.0	a
Raman Activity, Å ⁴ amu ⁻¹													
CCSD/cc-pVTZ	43.3	8.2	12.8	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.1	2.4	c
BLYP/aug-cc-pVQZ	56.692	5.947	11.952	0.001	0.0	0.0	0.0	0.0	0.0	0.433	2.257	0.556	a
B3LYP/6-311G(d)	31.559	7.949	12.785	0.0	0.0	0.0	0.0	0.0	0.0	1.091	4.226	0.016	a
B3LYP/aug-cc-pVQZ	48.098	6.706	11.943	0.0	0.001	0.0	0.001	0.0	0.0	0.641	1.861	0.643	a
PBE/aug-cc-pVQZ	55.797	6.198	11.221	0.0	0.001	0.0	0.0	0.0	0.0	0.427	2.088	0.517	a
PBE0/aug-cc-pVQZ	48.683	6.146	9.852	0.002	0.0	0.0	0.0	0.0	0.0	0.670	0.004	2.367	a
ω B97X-D/aug-cc-pVQZ	47.628	6.505	9.529	0.0	0.0	0.0	0.0	0.0	0.0	0.733	0.0	2.37	a
MP2/aug-cc-pVQZ	51.185	11.03	17.421	0.0	0.0	0.0	0.0	0.0	0.0	0.37	0.007	2.551	a
SCS-MP2/aug-cc-pVQZ	54.231	9.024	16.552	0.0	0.0	0.0	0.0	0.0	0.0	0.504	0.0	2.586	a
VSCF Anharmonic Frequency, cm ⁻¹													
B3LYP/6-311G(d)	2086.82	1032.69	732.42	1928.36	901.53	1273.54	405.64	815.99	186.01	681.79	644.91	579.01	a
CCSD(2) _T /cc-pVTZ	2082.14	1034.25	735.52	1920.56	909.00	1302.78	409.91	820.51	185.20	689.64	711.65	617.59	a
¹³ C ₂ O ₄ c VSCF Anharmonic Frequency, cm ⁻¹													
CCSD(2) _T /cc-pVTZ	2024.17	1034.09	726.92	1874.09	905.03	1268.61	408.97	798.03	185.00	666.12	693.35	611.72	a
$\Delta\nu$	57.97	0.16	8.60	46.47	3.97	34.17	0.94	22.48	0.20	23.53	18.30	5.87	a
¹³ C ₂ O ₄ c VSCF Anharmonic Frequency, cm ⁻¹													
CCSD(2) _T /cc-pVTZ	2049.11	975.18	704.37	1879.84	862.06	1278.75	387.64	806.72	174.85	684.59	702.35	586.29	a
$\Delta\nu$	33.02	59.07	31.14	40.72	46.94	24.03	22.26	13.79	10.35	5.05	9.29	31.30	a
VSCF Infrared Intensity, km mol ⁻¹													
B3LYP/6-311G(d)	0.0	0.0	0.0	1204.23	194.85	351.92	2.86	47.86	5.55	0.0	0.0	0.0	a
SCS-MP2/aug-cc-pVQZ	0.0	0.0	0.0	1201.59	245.28	302.49	2.77	39.11	5.40	0.0	0.0	0.0	a

[†] The light blue shaded rows indicate approximate anharmonic corrections to the level of theory indicated after multiplying by the ratio of the anharmonic to harmonic values calculated at the B3LYP/6-311G(d) level of theory.

¹ These values are presented in relative intensity

^a Present Data

^b Lewars 1996 [141]

^c Dunlap *et al.* 2013 [145]

There is a very strong infrared absorption band characteristic of this species ν_4 around 1921 cm^{-1} , with two less intense bands at 909 (ν_9) and 1303 (ν_9) cm^{-1} . The first band is in a similar location to that from $\text{C}_2\text{O}_4\text{a}$ (ν_9) at 1913 cm^{-1} , and similarly it overlaps with an absorption previously assigned to $\text{CO}_5(\text{C}_2)$. In Raman, the ν_1 mode at 2082 cm^{-1} is predicted to be particularly active, but may overlap with the ν_1 band of $\text{C}_2\text{O}_3\text{a}$ at 2091 cm^{-1} as well as ^{13}CO and C^{18}O species, as before. There are also two other less active Raman bands at 1034 (ν_2) and 735 (ν_3) cm^{-1} which could be diagnostic for the presence of this species.

6.3.6 $\text{C}_2\text{O}_4\text{d}$ ($\text{D}_{2\text{h}}$) [CO_2 Dimer]

Table 6.5 summarizes the harmonic frequencies, anharmonic frequencies, infrared intensities, Raman activities and isotopic shifts for $\text{C}_2\text{O}_4\text{d}$ ($\text{D}_{2\text{h}}$). Table 6.1 shows that the majority of diagnostics implicate moderate levels of non-dynamical correlation are anticipated, with four of the diagnostics identifying this species as one of the most problematic of our study (alongside $\text{C}_2\text{O}_4\text{b}$). There were considerable discrepancies between different levels of theory for the calculated geometries, vibrational frequencies, and Raman activities.

The hybrid GGAs PBE0 and $\omega\text{B97X-D}$ typically overestimate the harmonic frequencies significantly, while the GGAs PBE and BLYP only slightly underestimate them, even outperforming B3LYP overall. We note that the calculations of Zhang *et al.* [162] predicted an imaginary frequency, and very different harmonic frequencies from those calculated here and in the later study of Howart *et al.* [166]. This is because their frequencies are actually calculated for a slightly different structure, with no bond between the terminal oxygen atoms. The structure shown here was identified as a stable minimum for all levels of theory considered here, with no imaginary frequencies.

The strong agreement between the MP2, SCS-MP2, and the $\text{CCSD}(2)_\text{T}$ calculations indicate that these should be reliable for the prediction of infrared intensities and Raman activities. Surprisingly, the MP2/cc-pVDZ calculations of Howart *et al.* [166] are in reasonable agree-

ment to our calculations of the harmonic frequencies and the infrared intensities. Again, the B3LYP calculations performed with the smaller Pople basis set are in good agreement with those calculated with the larger aug-cc-pVQZ basis set, and should offer reliable anharmonic VSCF calculations. The Raman activities are in general agreement, however, the DFT methods predict significantly stronger Raman activity for the ω_2 band, where even MP2 values deviate significantly from those predicted by SCS-MP2 here. We note that based upon the calculated geometries and harmonic frequencies, this species is one of the few cases where the MP2 calculations may be more accurate than those from SCS-MP2.

Detection of $\text{C}_2\text{O}_4\text{d}$ ($\text{D}_{2\text{h}}$) by infrared spectroscopy will be challenging since this species exhibits only one moderately intense absorption, ν_8 at 1394 cm^{-1} and a weaker band, ν_6 at 1043 cm^{-1} . Similarly, there is only one moderately active Raman band, ν_2 at 862 cm^{-1} , with a few other bands potentially detectable at 522 (ν_3) and 182 (ν_5) cm^{-1} .

Table 6.5: Harmonic Frequencies, Anharmonic Frequencies, Infrared Intensities, Raman Activities and Isotopic Shifts for C₂O₄d (D_{2h}).

Level of theory	$\omega_1(A_g)$	$\omega_2(A_g)$	$\omega_3(A_g)$	$\omega_4(A_u)$	$\omega_5(B_{1g})$	$\omega_6(B_{1u})$	$\omega_7(B_{1u})$	$\omega_8(B_{2u})$	$\omega_9(B_{2u})$	$\omega_{10}(B_{3g})$	$\omega_{11}(B_{3g})$	$\omega_{12}(B_{3u})$	Ref
	Harmonic Frequency, cm ⁻¹												
CBS-QB3*	1660	876	360	170	—	1455	482	1497	-190	1552	514	398	b
MP2/cc-pVDZ	2095	865	512	263	184	1096	184	1399	686	1169	557	311	c
BLYP/aug-cc-pVQZ	1977.96	867.02	502.74	256.52	208.35	974.72	184.05	1335.16	698.95	1037.03	544.32	277.38	a
B3LYP/6-311G(d)	2096.43	915.3	533.13	272.39	234.37	1028.09	185.72	1418.83	742.31	1083.34	564.53	299.99	a
B3LYP/aug-cc-pVQZ	2081.66	927.72	537.37	274.29	280.95	1026.88	187.54	1416.8	761.4	1087.93	569.88	291.41	a
PBE/aug-cc-pVQZ	2020.85	900.46	524.46	263.37	205.3	1017.98	183.68	1382.65	739.83	1073.63	549.76	282.11	a
PBE0/aug-cc-pVQZ	2135.65	972.32	563.67	286.81	294.49	1074.9	193.58	1471.41	813.74	1130.38	580.47	301.79	a
ω B97X-D/aug-cc-pVQZ	2149.01	978.93	565.78	289.29	307.19	1068.78	191.27	1477.58	818.14	1129.81	581.71	303.19	a
MP2/aug-cc-pVQZ	2076.97	877.65	520.35	268.07	232.59	1085.52	180.72	1399.02	712.88	1154.92	560.28	292.86	a
SCS-MP2/aug-cc-pVQZ	2077.83	865.16	509.9	268.53	239.2	1095.07	183.33	1394.87	689.2	1168.1	563	293.88	a
CCSD(2) _T /Sadlej-pVTZ	2073.22	840.77	485.68	257.6	—	1057.89	182.77	1379.66	646.61	1132.08	563.38	271.78	a
CCSD(2) _T /cc-pVTZ	2099.99	874.42	510.28	268.95	184.98	1066.71	186.48	1410.12	690.36	1133.03	565.98	295.04	a
	Infrared Intensity, km mol ⁻¹												
MP2/cc-pVDZ	0	0	0	0	0	44	9	230	25	0	0	5	c
BLYP/aug-cc-pVQZ	0.0	0.0	0.0	0.0	0.0	27.22	7.15	247.24	13.60	0.0	0.0	2.96	a
B3LYP/6-311G(d)	0.0	0.0	0.0	0.0	0.0	27.02	9.06	287.99	16.18	0.0	0.0	6.58	a
B3LYP/aug-cc-pVQZ	0.0	0.0	0.0	0.0	0.0	29.04	7.94	276.23	15.88	0.0	0.0	4.04	a
PBE/aug-cc-pVQZ	0.0	0.0	0.0	0.0	0.0	29.79	6.56	241.52	15.23	0.0	0.0	2.34	a
PBE0/aug-cc-pVQZ	0.0	0.0	0.0	0.0	0.0	32.05	7.55	274.21	17.72	0.0	0.0	3.59	a
ω B97X-D/aug-cc-pVQZ	0.0	0.0	0.0	0.0	0.0	32.97	7.89	284.78	19.22	0.0	0.0	3.93	a
MP2/aug-cc-pVQZ	0.0	0.0	0.0	0.0	0.0	46.01	8.87	236.25	16.14	0.0	0.0	3.44	a
SCS-MP2/aug-cc-pVQZ	0.0	0.0	0.0	0.0	0.0	43.26	9.22	238.86	14.69	0.0	0.0	4.14	a
	Raman Activity, Å ⁴ amu ⁻¹												
BLYP/aug-cc-pVQZ	2.663	39.374	11.999	0.0	6.718	0.001	0.0	0.007	0.0	3.561	3.952	0.0	a
B3LYP/6-311G(d)	2.049	28.074	12.638	0.0	9.253	0.0	0.0	0.0	0.0	3.19	4.658	0.0	a
B3LYP/aug-cc-pVQZ	2.871	43.758	13.318	0.0	7.305	0.0	0.0	0.009	0.0	3.45	3.75	0.0	a
PBE/aug-cc-pVQZ	2.643	36.678	13.314	0.0	6.628	0.0	0.0	0.016	0.0	3.499	3.379	0.0	a
PBE0/aug-cc-pVQZ	2.765	41.544	14.139	0.0	7.269	0.0	0.0	0.018	0.0	3.396	3.239	0.0	a
ω B97X-D/aug-cc-pVQZ	2.603	40.792	11.74	0.0	7.824	0.0	0.0	0.0	0.0	3.315	3.324	0.0	a
MP2/aug-cc-pVQZ	1.013	21.555	13.5	0.0	6.728	0.0	0.0	0.0	0.0	3.324	2.926	0.0	a
SCS-MP2/aug-cc-pVQZ	1.165	13.869	7.234	0.0	6.711	0.0	0.0	0.0	0.0	3.331	3.136	0.0	a
	VSCF Anharmonic Frequency, cm ⁻¹												
B3LYP/6-311G(d)	1992.19	902.45	545.58	267.98	230.32	1005.52	184.79	1402.2	730.83	1055.26	562.37	298.46	a
CCSD(2) _T /cc-pVTZ	1995.57	862.14	522.20	264.60	181.78	1043.29	185.55	1393.59	679.68	1103.66	563.81	293.54	a
	¹³ C ₂ O ₄ d VSCF Anharmonic Frequency, cm ⁻¹												
CCSD(2) _T /cc-pVTZ	1922.52	860.13	521.92	264.60	174.98	1024.51	183.57	1361.57	675.86	1074.46	557.31	285.19	a
$\Delta\nu$	73.06	2.01	0.28	0.00	6.80	18.78	1.98	32.02	3.82	29.20	6.51	8.35	a
	¹⁸ O ₂ C ₂ O ₄ d VSCF Anharmonic Frequency, cm ⁻¹												
CCSD(2) _T /cc-pVTZ	1988.21	815.18	492.60	249.44	181.28	1011.61	177.63	1361.70	645.70	1084.39	539.69	289.03	a
$\Delta\nu$	7.36	46.96	29.60	15.16	0.50	31.68	7.92	31.89	33.99	19.27	24.13	4.51	a
	VSCF Infrared Intensity, km mol ⁻¹												
B3LYP/6-311G(d)	0.0	0.0	0.0	0.0	0.0	26.73	8.89	286.95	15.92	0.0	0.0	6.57	a
SCS-MP2/aug-cc-pVQZ	0.0	0.0	0.0	0.0	0.0	42.80	9.04	238.00	14.45	0.0	0.0	4.13	a

† The light blue shaded rows indicate approximate anharmonic corrections to the level of theory indicated after multiplying by the ratio of the anharmonic to harmonic values calculated at the B3LYP/6-311G(d) level of theory.

* Zhang *et al.* also calculated a 812 B2G frequency while we had no B2G frequencies

^a Present Data

^b Zhang *et al.* 2000 [162]

^c Howart *et al.* 2010 [166]

6.3.7 Energetics

Table 6.6 shows the tabulated heats of formation at 0 K and 298 K for all of the molecules studied here calculated using the G3MP2, CCCA-S4, and CCCA-CCL composite approaches, and for some well-characterized chemical species, results from other composite methods are shown for comparison.

Table 6.6: Heats of Formation (kcal/mol)

Molecule	Theory Level	ΔH_f (0K)	ΔH_f (298K)	Ref
CO($X^1 \Sigma^+$)	Expt	-27.20 ± 0.04	-26.42 ± 0.04	b
	CBS-Q	-27.9	-27.1	c
	G2	-29.0	-28.2	c
	G3	-27.4	-26.6	c
	G3MP2	-27.88	-26.89	a
	CCCA-S4	-27.31	-26.32	a
	CCCA-CCL	-27.02	-26.03	a
CO ₂ ($X^1 \Sigma_g^+$)	Expt	-93.97 ± 0.01	-94.05 ± 0.01	b
	CBS-Q	-96.0	-96.1	c
	G2	-96.7	-96.8	c
	G3	-95.2	-95.3	c
	G3MP2	-95.19	-95.05	a
	CCCA-S4	-94.58	-94.42	a
	CCCA-CCL	-94.01	-93.84	a
O(3P)	NIST	—	59.55 ± 0.02	h
O ₂ ($X^3 \Sigma_g^-$)	NIST	0.0	0.0	h
	QCI	—	0.0	f
C(3P)	NIST	—	171.29 ± 0.10	h
C ₂ ($X^1 \Sigma_g^-$)	QCI	—	197.6	f
C ₂ ($X^3 \Sigma_g^-$)	QCI	—	199.4	f
O ₃ ($X^1 A_1$)	Expt	—	34.220 ± 0.180	g
	G3MP2	41.9	41.58	a
	CCCA-S4	36.91	36.63	a
	CCCA-CCL	38.55	38.26	a
C ₂ O($X^3 \Sigma^-$)	Expt	—	92.01 ± 4.61	d
	GVb-POL-CI	89	—	e
C ₂ O($^1 \Sigma^+$)	G3MP2	107.58	108.71	a
	CCCA-S4	109.76	110.99	a
	CCCA-CCL	110.77	112	a
C ₂ O ₃ a($X^1 A_1$)	G3MP2	-65.07	-65.14	a
	CCCA-S4	-64.19	-64.02	a
	CCCA-CCL	-62.16	-61.99	a
C ₂ O ₃ b($X^1 A_1$)	G3MP2	40.08	40.04	a
	CCCA-S4	42.06	42.13	a
	CCCA-CCL	44.62	44.69	a
C ₂ O ₄ a($X^1 A_1$)	G3MP2	-71.4	-72.09	a
	CCCA-S4	-71.09	-71.59	a
	CCCA-CCL	-68.09	-68.58	a
C ₂ O ₄ b($X^1 A_1$)	G3MP2	55.55	54.92	a
	CCCA-S4	56.36	55.93	a
	CCCA-CCL	60.04	59.62	a
C ₂ O ₄ c($X^1 A_g$)	G3MP2	-141.47	-142.31	a
	CCCA-S4	-140.98	-141.63	a
	CCCA-CCL	-138.22	-138.87	a
C ₂ O ₄ d($X^1 A_g$)	G3MP2	67.02	66.76	a
	CCCA-S4	69.32	69.26	a
	CCCA-CCL	73.32	73.25	a

^a Present Data

^b Chase *et al.* 1998 [196]

^c Feller *et al.* 2003 [197]

^d Zengin *et al.* 1996 [198]

^e Walch 1980 [199]

^f Goldstein *et al.* 2012 [200]

^g Streng 1961 [201]

^h Cox *et al.* 1984 [202]

Table 6.7 shows the energy differences for bond dissociation reactions from the calculated heats of formations, calculated via Equation 6.6.

$$\Delta E = \Delta H_f(\text{Products}) - \Delta H_f(\text{Reactants}) \quad (6.6)$$

Also shown in Table 6.7 is the detonation velocity (D) and detonation pressure (P) as calculated by the Kamlet-Jacobs equations[203].

$$D = \Phi^{0.5}(1.011 + 1.312\rho) \quad (6.7)$$

$$P = 1.558\Phi\rho^2 \quad (6.8)$$

$$\Phi = NM_{\text{ave}}^{0.5}Q^{0.5} \quad (6.9)$$

The detonation velocity (D) is given in km/s, detonation pressure (P) in GPa, Φ is the characteristic value of explosives, ρ is the packed density given in g/cm³, N is the moles of gas produced per gram of reactant, M_{ave} is the average molar weight of the detonation products, and Q is the estimated heat of detonation (here we use the bond dissociation energies calculated from the heat of formation reported in Table 6.6). These properties give information about the release of energy from a potential HEDM, which is of great relevance to those intending to use them as propellants or explosives.

Despite C₂O₃b and C₂O₄b having highly exothermic reactions they are not found to be suitable HEDMs since they were found to have imaginary frequencies and are not expected to be real meta-stable molecules. As noted in the introduction, reactions involving the production of CO and CO₂ (and to a lesser extent, O₂) are generally found to be exoergic/exothermic. For C₂O₃a, the most likely breakdown products would be a molecule each of CO and CO₂. For each of the C₂O₄ isomers, the most energy is released from the formation of two CO₂ molecules. Based on exothermicity considerations alone, the C₂O₄d isomer appears to have the most potential as a HEDM species, although C₂O₃a, C₂O₄a, and C₂O₄c should all be regarded as potential HEDM compounds. Note that in this preliminary work, we have not considered dissociation barriers, or inter-system crossing; both of which play a

considerable role in predicting how these species may behave as HEDM materials. When considering the detonation velocities and pressures C_2O_4d appears to be extremely promising, with predicted values comparable to military grade explosives such as CL-20 (hexanitrohexaazaisowurtzitane) which has a predicted detonation velocity of 9.73 km/s and predicted detonation pressure of 43.73 GPa [204]. The predicted values for C_2O_4d are close to those of CL-20 along with an identical C:O ratio of 1:2, we also note that these large value indicate that this isotope is more likely an explosives candidate rather than propellant. According to O’Sullivan *et al.* an ideal explosive will have an oxygen balance of zero [205]. As there are multiple equations for calculating the oxygen balance, depending on which equation is used we found the oxygen balance of C_2O_3 (-22 to 0%) and C_2O_4 (0 to 36%) to be zero, or close to zero. These further supports the predictions of these species being HEDM candidates.

Table 6.7: Energy Differences for Bond Dissociation Reactions (kcal/mol), Detonation Velocities (km/s) and Detonation Pressures (GPa)

Reaction	Theory Level	ΔE (0K)	ΔE (298K)	D (298K)	P (298K)
$C_2O_3a(X^1A_1) \rightarrow C_2O(X^1\Sigma^+) + O_2(X^3\Sigma_g^-)$	G3MP2	172.7	173.9	—	—
	CCCA-S4	174.0	175.0	—	—
	CCCA-CCL	172.9	174.0	—	—
$C_2O_3a(X^1A_1) \rightarrow 2CO(X^1\Sigma^+) + \frac{1}{2}O_2(X^3\Sigma_g^-)$	G3MP2	9.3	11.4	—	—
	CCCA-S4	9.6	11.4	—	—
	CCCA-CCL	8.1	9.9	—	—
$C_2O_3a(X^1A_1) \rightarrow CO(X^1\Sigma^+) + CO_2(X^1\Sigma_g^+)$	G3MP2	-58.0	-56.8	6.688	18.336
	CCCA-S4	-57.7	-56.7	6.686	18.323
	CCCA-CCL	-58.9	-57.9	6.720	18.510
$C_2O_4a(X^1A_1) \rightarrow C_2O(X^1\Sigma^+) + O_3(X^1A_1)$	G3MP2	220.9	222.4	—	—
	CCCA-S4	217.8	219.2	—	—
	CCCA-CCL	217.4	218.8	—	—
$C_2O_4c(X^1A_g) \rightarrow C_2O(X^1\Sigma^+) + O_3(X^1A_1)$	G3MP2	291.0	292.6	—	—
	CCCA-S4	287.7	289.3	—	—
	CCCA-CCL	287.5	289.1	—	—
$C_2O_4d(X^1A_g) \rightarrow C_2O(X^1\Sigma^+) + O_3(X^1A_1)$	G3MP2	82.4	83.5	—	—
	CCCA-S4	77.4	78.4	—	—
	CCCA-CCL	76.0	77.0	—	—
$C_2O_4a(X^1A_1) \rightarrow 2CO_2(X^1\Sigma_g^+)$	G3MP2	-119.0	-118.0	7.637	25.054
	CCCA-S4	-118.1	-117.3	7.624	24.974
	CCCA-CCL	-119.9	-119.1	7.654	25.170
$C_2O_4c(X^1A_g) \rightarrow 2CO_2(X^1\Sigma_g^+)$	G3MP2	-48.9	-47.8	6.168	16.528
	CCCA-S4	-48.2	-47.2	6.149	16.428
	CCCA-CCL	-49.8	-48.8	6.201	16.704
$C_2O_4d(X^1A_g) \rightarrow 2CO_2(X^1\Sigma_g^+)$	G3MP2	-257.4	-256.9	9.168	35.735
	CCCA-S4	-258.5	-258.1	9.180	35.821
	CCCA-CCL	-261.3	-260.9	9.205	36.017
$C_2O_4a(X^1A_1) \rightarrow 2CO(X^1\Sigma^+) + O_2(X^3\Sigma_g^-)$	G3MP2	15.6	18.3	—	—
	CCCA-S4	16.5	19.0	—	—
	CCCA-CCL	14.1	16.5	—	—
$C_2O_4c(X^1A_g) \rightarrow 2CO(X^1\Sigma^+) + O_2(X^3\Sigma_g^-)$	G3MP2	85.7	88.5	—	—
	CCCA-S4	86.4	89.0	—	—
	CCCA-CCL	84.2	86.8	—	—
$C_2O_4d(X^1A_g) \rightarrow 2CO(X^1\Sigma^+) + O_2(X^3\Sigma_g^-)$	G3MP2	-122.8	-120.5	9.398	37.543
	CCCA-S4	-123.9	-121.9	9.424	37.755
	CCCA-CCL	-127.4	-125.3	9.489	38.279
$C_2O_4a(X^1A_1) \rightarrow CO(X^1\Sigma^+) + CO_2(X^1\Sigma_g^+) + \frac{1}{2}O_2(X^3\Sigma_g^-)$	G3MP2	-51.7	-49.9	6.220	16.620
	CCCA-S4	-50.8	-49.2	6.198	16.503
	CCCA-CCL	-52.9	-51.3	6.264	16.858
$C_2O_4c(X^1A_g) \rightarrow CO(X^1\Sigma^+) + CO_2(X^1\Sigma_g^+) + \frac{1}{2}O_2(X^3\Sigma_g^-)$	G3MP2	18.4	20.4	—	—
	CCCA-S4	19.1	20.9	—	—
	CCCA-CCL	17.2	19.0	—	—
$C_2O_4d(X^1A_g) \rightarrow CO(X^1\Sigma^+) + CO_2(X^1\Sigma_g^+) + \frac{1}{2}O_2(X^3\Sigma_g^-)$	G3MP2	-190.1	-188.7	8.575	31.261
	CCCA-S4	-191.2	-190.0	8.590	31.368
	CCCA-CCL	-194.4	-193.1	8.625	31.625

6.3.8 Evaluation of Theoretical Approaches

Overall, the theoretical methods utilized to predict the geometries, vibrational frequencies, infrared intensities and Raman activities presented here are all reasonably well-suited to this task for the carbon oxides shown here. Though the two C_2O_3 and four C_2O_4 structures investigated here were susceptible to moderate levels of non-dynamical character (Table 6.1), this rarely presented a problem in describing these properties to a high-degree of accuracy. Here, we analyzed the discrepancies between results for 18 bond lengths, as well as harmonic frequencies, infrared intensities and Raman activities for 66 modes (see *Supporting Information for more details*). Table 6.8 shows a summary of the errors associated with each level of theory for the prediction of geometries and harmonic vibrational frequencies compared to those from $\text{CCSD}(2)_\text{T}/\text{cc-pVTZ}$ values, and for the prediction of infrared intensities and Raman activities compared to those from the $\text{SCS-MP2}/\text{aug-cc-pVQZ}$ level of theory. Table 6.8 provides a reasonable overview of the bias introduced by adopting particular levels of theory. For example, both GGA functionals (PBE and BLYP) consistently overestimate bond lengths and underestimate harmonic frequencies and infrared intensities, whereas two of the meta-GGA functionals (PBE0 and $\omega\text{B97X-D}$) perform in the opposite manner. Meanwhile, though B3LYP has comparable performance to PBE for predicting geometries, it does perform significantly better at estimating the harmonic vibrational frequencies. In general, the SCS-MP2 calculations offer slightly more accurate and precise values compared to its MP2 counterpart, in comparison with values calculated at the CCSD(T) and $\text{CCSD}(2)_\text{T}$ levels of theory. We note that $\omega\text{B97X-D}$ and (to a lesser extent) PBE0 functionals exhibited the worst overall performance, but provided Raman activities closer to those predicted by MP2 and SCS-MP2 values.

Table 6.8: Comparison of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Signed Error (MSE) and Mean Signed Percentage Error (MSPE) for Bond Lengths (in Å), Vibrational Frequencies (in cm^{-1}), Infrared Intensities (in km mol^{-1}), and Raman Activities (in $\text{Å}^4\text{amu}^{-1}$).

Level of theory	MAE	MAPE	MSE	MSPE
	Bond Lengths, Å			
BLYP/Aug-cc-pVQZ	0.0140 (± 0.0126)	0.98 (± 0.89)%	0.0140 (± 0.0126)	0.98 (± 0.89)%
PBE/Aug-cc-pVQZ	0.0072 (± 0.0059)	0.51 (± 0.42)%	0.0045 (± 0.0082)	0.34 (± 0.61)%
B3LYP/Aug-cc-pVQZ	0.0079 (± 0.0066)	0.58 (± 0.48)%	-0.0071 (± 0.0075)	-0.52 (± 0.55)%
PBE0/Aug-cc-pVQZ	0.0173 (± 0.0155)	1.23 (± 1.10)%	-0.0171 (± 0.0157)	-1.21 (± 1.11)%
ω B97X-D/Aug-cc-pVQZ	0.0185 (± 0.0156)	1.31 (± 1.11)%	-0.0179 (± 0.0163)	-1.27 (± 1.16)%
M06-L/Aug-cc-pVQZ ^a	0.0123 (± 0.0102)	0.87 (± 0.72)%	-0.0123 (± 0.0102)	-0.87 (± 0.72)%
M06/Aug-cc-pVQZ ^a	0.0201 (± 0.0179)	1.43 (± 1.27)%	-0.0199 (± 0.0182)	-1.41 (± 1.29)%
MP2/Aug-cc-pVQZ	0.0026 (± 0.0034)	0.19 (± 0.24)%	-0.0008 (± 0.0042)	-0.06 (± 0.32)%
SCS-MP2/Aug-cc-pVQZ	0.0035 (± 0.0028)	0.25 (± 0.20)%	0.0003 (± 0.0046)	0.01 (± 0.10)%
	Harmonic Vibrational Frequencies, cm^{-1}			
BLYP/Aug-cc-pVQZ	70.59 (± 45.18)	7.07 (± 6.59)%	-54.95 (± 47.51)	-5.95 (± 7.67)%
PBE/Aug-cc-pVQZ	48.48 (± 28.52)	5.23 (± 4.40)%	-33.51 (± 34.37)	-3.61 (± 5.79)%
B3LYP/Aug-cc-pVQZ	25.79 (± 22.19)	4.27 (± 8.86)%	2.91 (± 29.98)	0.89 (± 9.91)%
PBE0/Aug-cc-pVQZ	39.97 (± 30.38)	5.93 (± 10.49)%	30.85 (± 32.17)	4.13 (± 11.46)%
ω B97X-D/Aug-cc-pVQZ	46.01 (± 34.87)	6.98 (± 12.55)%	36.90 (± 35.33)	4.77 (± 13.70)%
MP2/Aug-cc-pVQZ	20.46 (± 18.67)	2.86 (± 4.97)%	-8.49 (± 23.14)	-0.91 (± 5.69)%
SCS-MP2/Aug-cc-pVQZ	16.91 (± 15.08)	2.60 (± 5.45)%	-4.91 (± 19.79)	-0.45 (± 6.09)%
	Infrared Intensities, km mol^{-1}			
BLYP/Aug-cc-pVQZ	10.29 (± 16.43)	22.93 (± 33.27)%	-4.88 (± 17.88)	-5.04 (± 39.66)%
PBE/Aug-cc-pVQZ	10.56 (± 16.60)	25.45 (± 40.40)%	-5.78 (± 17.92)	-5.82 (± 47.91)%
B3LYP/Aug-cc-pVQZ	13.10 (± 21.06)	52.81 (± 206.79)%	6.61 (± 23.02)	41.72 (± 212.24)%
PBE0/Aug-cc-pVQZ	15.23 (± 24.05)	68.48 (± 236.96)%	7.54 (± 26.44)	52.21 (± 244.32)%
ω B97X-D/Aug-cc-pVQZ	18.72 (± 31.53)	81.71 (± 304.54)%	11.47 (± 33.65)	67.59 (± 312.32)%
MP2/Aug-cc-pVQZ	1.18 (± 1.66)	21.41 (± 57.03)%	0.31 (± 2.05)	4.58 (± 59.41)%
	Raman Activity, $\text{Å}^4\text{amu}^{-1}$			
BLYP/Aug-cc-pVQZ	3.13 (± 4.32)	153.51 (± 712.15)%	1.58 (± 4.79)	134.52 (± 726.53)%
PBE/Aug-cc-pVQZ	2.90 (± 4.08)	109.91 (± 415.98)%	1.33 (± 4.55)	87.07 (± 427.40)%
B3LYP/Aug-cc-pVQZ	2.57 (± 4.20)	81.52 (± 288.19)%	1.00 (± 4.63)	64.25 (± 296.65)%
PBE0/Aug-cc-pVQZ	2.37 (± 3.87)	71.94 (± 259.67)%	0.80 (± 4.29)	50.91 (± 268.31)%
ω B97X-D/Aug-cc-pVQZ	2.20 (± 3.71)	75.16 (± 282.64)%	0.65 (± 4.11)	54.17 (± 291.48)%
MP2/Aug-cc-pVQZ	0.88 (± 1.53)	15.63 (± 22.97)%	0.30 (± 1.68)	3.05 (± 27.84)%

[†] Errors are determined from 18 bond lengths or 66 vibrational modes from 6 different structures/ Bond lengths and harmonic frequencies are compared to CCSD(2)_T/cc-pVTZ reference values whereas infrared intensities and Raman activities are compared to values calculated at the SCS-MP2/Aug-cc-pVQZ level of theory. \pm indicates one standard deviation in values obtained.

^a Some calculations were performed at the Aug-cc-pVTZ level, but the deviations from Aug-cc-pVQZ level were negligible.

For each of the levels of theory considered, we have also produced several plots showing how well each level of theory performs against the CCSD(2)_T/cc-pVTZ (for geometries and harmonic frequencies) and SCS-MP2/aug-cc-pVQZ (for infrared intensities and Raman activities). Figure 6.2 shows an example of this for B3LYP; similar plots for each level of theory can be found in the *Supporting Information*.

Lastly, we note that the DFT-based A_λ diagnostic of Fogueri et al. [190] for assessing non-dynamical character was found to be in good agreement with the widely used %TAE(T) diagnostic. For example, the level of discrepancies between different levels of theory in predictions of geometries for these six species was found to better correlate with their A_λ values, than to the %TAE(T) diagnostic. This approach was found to be quite accurate at predicting which species would be problematic for calculations based upon a single reference wavefunction, which include the meta-GGA functionals that rely on part Hartree-Fock

exchange. Therefore, our assessment may be bias due to the fact that these calculations were utilized to determine these geometries. However, such diagnostics could still provide useful information on the degree to which DFT calculated values may be relied upon. More information on our statistical analysis of the various theoretical approaches can be found in the *Supporting Information*.

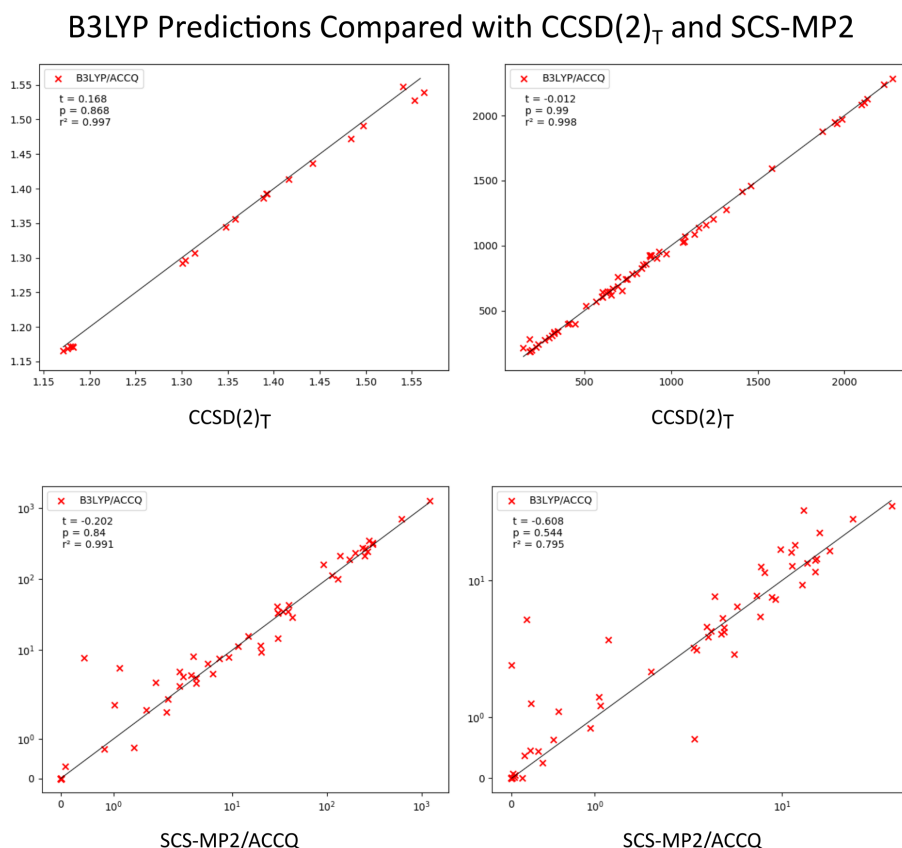


Figure 6.2: Plots showing how the B3LYP level of theory compares to CCSD(2)_T/cc-pVTZ in the prediction of molecular geometries (top left) and harmonic frequencies (top right) and to SCS-MP2/aug-cc-pVQZ for infrared intensities (bottom left) and Raman activities (bottom right). T-test, p-test, and r² values are indicated in each plot. *Note: ACCQ stands for aug-cc-pVQZ.*

6.3.9 Summary of Results

Figure 6.3 shows a Kernel Density Estimate of how each different level of theory compares to the CCSD(2)_T/cc-pVTZ level of theory at predicting the bond length and harmonic

frequencies. All coupled cluster theory calculations were using cc-pVTZ basis set, except for CCSD (where a dashed line was used to indicate calculations used the Sadlej-pVTZ basis set). The CCSD(T) values are found to be very close to the those from CCSD(2)_T/cc-pVTZ. Here, it can be clearly seen that the GGAs (PBE and BLYP) consistently underestimate the bond strengths of species, resulting in longer bonds and lower vibrational frequencies predicted, meanwhile the opposite trend is observed for the hybrid GGAs, in particular for PBE and ω B97X-D. The B3LYP hybrid GGA overall performs quite well but has a wider distribution than MP2 indicative of higher overall deviation from the CCSD(2)_T/cc-pVTZ values. MP2-SCS seems to offer a small improvement over MP2. CCSD overestimates bond strengths leading to shorter bonds and higher frequencies, whereas CCSD(T) over-corrects for this but improves on this significantly.

We note that frequencies and infrared intensities are highly dependent on the calculated optimized geometries [83], of which, the Coupled Cluster calculations typically produce the most accurate results¹. MP2 level calculations were found to present only slightly more accurate results when compared to DFT functionals. For Raman activities, our previous study found that MP2 level calculations typically resulted in the most reliable results and we found them to be extremely sensitive to the degree of polarization and diffusivity incorporated into the basis sets [83].

We summarize an estimation of the errors that arise from the performance of each level of theory considered, for each of the properties investigated here below:

For geometries: CCSD(2)_T \leq CCSD(T) < SCS-MP2 \sim MP2 < PBE < B3LYP < M06-L < BLYP < PBE0 < ω B97X-D < M06

For harmonic frequencies: CCSD(2)_T \leq CCSD(T) < SCS-MP2 < MP2 \leq B3LYP \ll PBE < PBE0 < ω B97X-D < BLYP

For infrared intensities: SCS-MP2 \sim MP2 < BLYP < PBE \ll B3LYP < PBE0 <

¹ All optimized geometry calculation data can be found in the Supporting Information

wB97X-D

For Raman activities: SCS-MP2 \sim MP2 \ll wB97X-D < PBE0 < B3LYP < PBE < BLYP

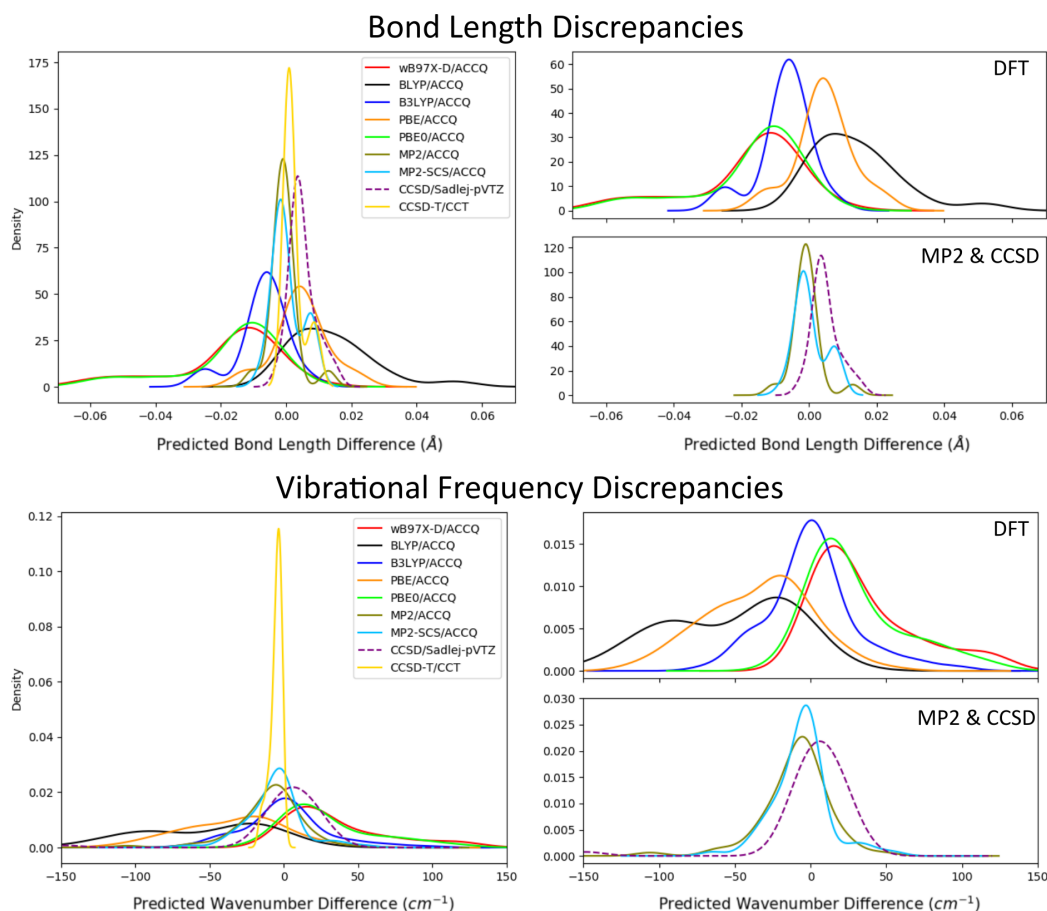


Figure 6.3: A Kernel Density Estimate (KDE) plot of the difference in predicted bond lengths (top) and harmonic vibrational frequencies (bottom) between each individual theory versus those predicted by CCSD(2)_T/cc-pVTZ. *Note:* ACCQ stands for Aug-cc-pVQZ.

6.4 Conclusions

We have presented a computational investigation of the geometries, vibrational frequencies, infrared intensities and Raman activities and energetics for two C₂O₃ and four C₂O₄ isomers. We have also provided a statistical assessment of how each level of theory performed compared to high-level calculations. We also interpreted their accuracy based on consider-

ations of their capability to handle non-dynamical correlation, as assessed by traditional as well as DFT-based diagnostics.

The potential for oligomers and co-oligomers of carbon oxides to be HEDMs lies within the bond strength of the carbon monoxide and carbon dioxide bonds which will form when these molecules decompose. Such an event could occur spontaneously when these compounds are heated or subjected to shock, but this will require further investigation into their properties. Though many of these oligomers still remain hypothetical molecules, here we have provided spectral information that could aid in their detection from calculations at several different levels of theory. For instance, the formation of C_2O_3 and C_2O_4 could be prevalent within Solar System and interstellar ices that were originally rich in CO or CO_2 and subsequently exposed to energetic radiation. If present in comets, for example, these *parent* species would likely dissociate to form CO, CO_2 , and O_2 upon sublimation, which would occur at higher abundances than their *daughter* fragments naturally occur in[35]. This work could help establish the presence of C_2O_3 and C_2O_4 parent species within laboratory experiments, where Raman spectroscopy is becoming more commonly employed [26]. In order to detect these species within the space environment (*e.g.*, in the comae of comets), if they are sufficiently stable, highly accurate calculations on their rovibrational levels will be required [31].

In addition, we found that, of the investigated oligomers, C_2O_4d may offer the most promising HEDM candidate, although each of the four stable molecules presented here are viable HEDM candidates. We hope this work will help to elucidate a broader view of higher order carbon oxide chemistry in several fields of research. In order to determine how feasible the synthesis of these C_2O_3 and C_2O_4 isomers are, and their relative stability, further work is required to determine what barriers to dissociation may be present. The spectral properties of the compounds reported here may also aid in the process of synthesizing these HEDMs from carbon dioxide oligomers which have already been experimentally observed, where the necessary information for quantitative measurements to be made on the yields of reactions that form these molecules is provided here. We also report that two of the potential oligomers

reported here, C_2O_3b and C_2O_4b , are not real isolatable chemical species (at least on the singlet potential energy surface), as revealed by the presence of imaginary vibrational frequencies. However, other than some exploratory work, our studies were essentially confined to the singlet potential energy surface and additional work is suggested to investigate the full global energy surfaces of these species.

PART

III

CONCLUDING REMARKS

7 CONCLUSION AND OUTLOOKS

This work has filled three major voids within the previously existing scientific literature.

- (i) A lack of software tools designed to automate the process of calculating Raman activities
- (ii) A wide-ranging assessment of commonly available basis sets in their ability to accurately reproduce Raman activities
- (iii) A lack of complete spectroscopic data (IR and Raman) on astrophysically relevant higher order carbon oxides

In filling these voids, we have provided the scientific community with the proper tools to implement HTC techniques in the calculation of Raman activities. We have also given experimental astrochemists with the necessary data for them to identify the higher order carbon oxides within irradiated ices. The major conclusions and future directions of all projects are discussed in more detail in the following sections.

7.1 Software Development

7.1.1 Major Conclusions

The python package I developed provides an excellent framework for computational chemists to utilize HTC techniques in Raman calculations. Recent increases in computing power available to scientists and recent increase in HTC techniques within computational chemistry brought forth a demand for more automation packages. The QC software package

GAMESS(US) was neglected by most workflow automation packages, and AutoGAMESS filled this void. Another important part of the AutoGAMESS package is that it gives experimental groups the ability to utilize QC techniques without overcoming the steep learning curve of most QC packages.

The major conclusions of this study are:

- AutoGAMESS provides an easy to use framework for automating QC workflow
- AutoGAMESS can automate the generation of scaling factors used to improve QC calculations

AutoGAMESS has been an essential tool for all subsequent studies. Without the development of AutoGAMESS the subsequent studies would either have not been done or would have been significantly less comprehensive.

7.1.2 Future Directions

Since AutoGAMESS is an open-source software licensed under the MIT License, it is freely available for any to use or modify. This allows other scientists or commercial users to clone and modify AutoGAMESS to improve their own user experience. Anyone is free to submit changes to the original AutoGAMESS package as well, as long as these changes pass the software tests these changes will be implemented in the next version of AutoGAMESS.

I personally have been working on including a Graphical User Interface (GUI) for AutoGAMESS. This would make the package more intuitive and user-friendly, which would hopefully increase its adoption in the computational chemistry community. Another feature I have been working on is functionality for automating diagnostic calculations such as %TAE[(T)] and similar DFT-based diagnostics (see appendix Section C). This would allow more comprehensive diagnostics to be performed prior to highly demanding calculations being performed.

7.2 Basis Set Benchmarking Study

7.2.1 Major Conclusions

In this study we identified various trends in the performance of different basis sets in predicting the molecular properties of water. An emphasis was placed on identifying trends in the predictions of Raman activities as these had often been overlooked in previous literature. In particular, our finding that diffuse functions play a major role in improving the Raman calculations will greatly help researchers aiming to calculate Raman activities.

The major conclusions of this study are:

- The Def2-n basis set family performs well overall
- Diffuse functions and extensive polarization functions are beneficial to the performance of Raman calculations
- B3LYP is better suited for infrared intensities than Raman activities
- MP2 is particularly well suited for Raman calculations

These conclusions influenced our selection of basis set and the decision to include more DFT functionals (since B3LYP performed poorly for Raman activities) within our next study which was described in Chapter 6.

7.2.2 Future Directions

Extending these calculations to a larger set of molecules would greatly improve the study. This would either validate or discredit the trends observed here for molecules with differing levels of multi-reference character. This process was started during my thesis work, although, there was not sufficient time for all calculations to finish running. I have been running calculations in the same methodology but across a larger set of molecules (C_2H_2 , C_2H_4 , C_2H_6 , CH_4 , H_2CO , HCN , H_2O , and NH_3) and have completed roughly three-quarters of all

calculations. I hope that this work is eventually completed by another student since it will be significantly more impactful to the community.

Incorporating more DFT functionals within future benchmarking studies would also provide added benefits to this study. As we have found that B3LYP performs rather poorly for Raman calculations, it would be beneficial to include more functionals in hopes of finding a better candidate. It would also be beneficial to the scientific community if a similar study was performed with a large set of DFT functionals but only one basis set. Specifically, assessing the performance of different DFT functionals in calculating Raman activities. Although, some previous studies have been done on this, none have been extensive and often are done with a basis set that is not sufficiently close to the basis set limit.

7.3 Carbon Oxides Study

7.3.1 Major Conclusions

In this study we presented the optimized geometries, vibrational frequencies, infrared intensities, Raman activities and energetics for astrophysically relevant higher order carbon oxides. We also provided a statistical assessment of how each level of theory performed compared to high-level calculations. Here, the Raman activities were one of the major contributions to the scientific community. The Raman activities for these molecules were not previously reported, as such, our work fills this void. We hope these calculated Raman activities will help experimental groups detect these higher order carbon oxides in irradiated carbon oxide rich ices. The statistical assessment will hopefully allow other computational chemists to extend our study to molecular dynamics calculations utilizing ideally suited DFT functionals.

The major conclusions of this study are:

- These carbon oxides could be ideal HEDMs, in particular C_2O_4d may be the most promising candidate
- C_2O_3b and C_2O_4b are not real isolatable chemical species on the singlet potential

energy surface

→ ω B97X-D performs well in calculating Raman activities

If I were to continue on with this work, the Raman activities calculated would have played a pivotal role in my ability to identify these higher order carbon oxides within laboratory studies. Our group has recently begun building a Raman spectrometer and previously I had been working on preliminary laboratory studies on electron irradiated CO₂-rich ices (only a poster presentation was done on this work). The conclusions drawn from this computational study would have served as the primary literature for identifying these novel carbon oxide species within those laboratory studies.

7.3.2 Future Directions

Highly accurate calculations on the rovibrational levels of these carbon oxides would more readily allow for their detection by astronomers. This has already been partially done through collaboration between our group and another group, in particular, it was done for the C₂O₃a isomer. These highly accurate rovibrational calculations still remain to be done on the C₂O₄ isomers. Without these rovibrational levels astronomers are unable to confidently identify these higher order carbon oxides in observations.

More exploratory work could be done on C₂O₃b and C₂O₄b across the full global energy surfaces of these species. Since our study was only on the singlet potential energy surface, we could not conclusively say that these isomers are not stable across the global energy surface. Similarly, more extensive studies on the stability of all isomers studied would greatly improve our understanding of these higher order carbon oxides. In our work, we found many isomer had exothermic bond dissociation reactions, this would imply the species are to some degree unstable. Investigating the decay time of these isomers would help researchers understand how long they could be expected to survive within irradiated ices.

Finally, our work was only done on single molecule in vacuo calculations of the infrared intensities and Raman activities, however, these higher order carbon oxides would be detected

within heterogeneous ice mixtures. There has been great work done lately to calculate the bulk-phase infrared and Raman spectra of molecules. This is primarily done using AIMD methods, an example of this is the TRajjectory Analyzer and VISualizer (TRAVIS) package [206]. These methods can be used to extend our studies to bulk-phase calculations of the infrared and Raman spectra. I had begun to learn these techniques and hope that the student which is replacing me will be able to continue this project. A future student could simulate the spectra of these higher order carbon oxides within heterogeneous ice mixtures dominated by CO or CO₂, which would be the most realistic case for laboratory studies.

APPENDIX A. QUANTUM CHEMISTRY

Should the reader want an in depth coverage of quantum chemistry, they are encouraged to read Introduction to Computational Chemistry by Frank Jensen [207]. Much of this section is written with said textbook as reference, and only a few other sources supply minor supplemental information.

A.1 Molecular Orbital Theory

Within molecular orbital theory the full wavefunction is approximated through many one-electron functions, termed molecular orbitals. A molecular orbital is a function of the Cartesian coordinates of a single electron, $\psi(x, y, z)$, and its square (or square modulus if its complex) is the probability distribution of the electron in space. A complete description of the distribution requires the spin component, ξ , of the electron also be taken into account. The spin component can take the value of $\pm\frac{1}{2}$, and is conventionally given along the z-axis with $\alpha(\xi)$ being along the positive axis and $\beta(\xi)$ being along the negative axis (shown below).

$$\begin{array}{l} +z : \quad \alpha\left(\frac{1}{2}\right) = 1 \quad \alpha\left(-\frac{1}{2}\right) = 0 \\ -z : \quad \beta\left(\frac{1}{2}\right) = 0 \quad \beta\left(-\frac{1}{2}\right) = 1 \end{array}$$

The complete wavefunction for a single electron is the product of the molecular orbital and the spin, termed the spin orbital $\chi(x, y, z, \xi)$. Note that the spin orbital can either be equal to $\psi(x, y, z)\alpha(\xi)$ or $\psi(x, y, z)\beta(\xi)$. Due to the anti-symmetry principle, a many electron wavefunction cannot simply be the product of spin orbitals since interchange of the i th and j th electrons would not be equivalent to multiplication of the wavefunction by -1 . As such, the wavefunction must be arranged as a determinantal wavefunction, where for simplicity I will use the shorthand notation $\chi_i(x_i, y_i, z_i, \xi_i) = \chi_i(i)$.

$$\Psi_{\text{det}} = \begin{vmatrix} \chi_1(1) & \chi_2(1) & \dots & \chi_n(1) \\ \chi_1(2) & \chi_2(2) & \dots & \chi_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(n) & \chi_2(n) & \dots & \chi_n(n) \end{vmatrix} \quad (\text{A.1})$$

The determinantal wavefunction can be normalized by multiplication of $(n!)^{-\frac{1}{2}}$. Building a determinantal wavefunction is usually done by first choosing a set of molecular orbitals (ψ_1, ψ_2, \dots) then assigning electrons of either spin α or β . This can be visualized through electron configuration diagrams, as is shown in Figure A.1, where each molecular orbital cannot contain two electrons with the same spin as is dictated by the Pauli exclusion principle.

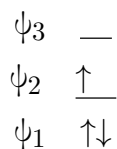


Figure A.1: Example electron configuration diagram

The example shown in Figure A.1 represents a configuration with $(\psi_1\alpha), (\psi_1\beta), (\psi_2\alpha)$ spin orbitals. Note that up-arrows represent α spin and down-arrows represent β spin. In proper terminology we could refer to the ψ_1 orbital as being doubly occupied, the ψ_2 orbital as being singly occupied and the ψ_3 orbital as empty. Wavefunctions where all orbitals are either doubly occupied or empty are referred to as closed-shell wavefunctions. The full normalized wavefunction for a system with n -electrons (even) doubly occupying $\frac{n}{2}$ orbitals is termed the Slater determinant (Equation A.2).

$$\Psi = (n!)^{-\frac{1}{2}} \begin{vmatrix} \psi_1(1)\alpha(1) & \psi_1(1)\beta(1) & \psi_2(1)\alpha(1) & \dots & \psi_n(1)\beta(1) \\ \psi_1(2)\alpha(2) & \psi_1(2)\beta(2) & \psi_2(2)\alpha(2) & \dots & \psi_n(2)\beta(2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_1(n)\alpha(n) & \psi_1(n)\beta(n) & \psi_2(n)\alpha(n) & \dots & \psi_n(n)\beta(n) \end{vmatrix} \quad (\text{A.2})$$

A.2 Basis Set Expansion

In molecular orbital theory it is often too challenging to find a complete description of the three dimensional wavefunction, as such, in practical applications the individual molecular orbitals (MOs) are expressed as linear combinations of finite basis functions. This basis set expansion can be thought of as considering the MOs as functions in a finite coordinate system where the basis is the prescribed functions. These basis functions are one-electron functions ($\varphi_1, \varphi_2, \dots$), and are weighted with a molecular orbital expansion coefficient (c_i). The coefficients are how the MOs are constrained on the finite coordinate system, as such, varying them will optimize the wavefunction.

$$\psi_i = \sum_{\mu=1}^N c_{\mu i} \varphi_{\mu} \quad (\text{A.3})$$

This reduces the problem of finding a complete description of the three dimensional wavefunction to finding only a finite set of linear coefficients for each orbital. A simple version of molecular orbital theory is linear combination of atomic orbitals (LCAO) theory, where atomic orbitals of constituent atoms are used as the basis functions for molecular orbitals. Any appropriately defined set of functions may be used for a basis expansion, however, it is convenient to define a particular set of basis functions associated with each nucleus, depending only on the charge of the nucleus. These functions may also have the symmetry properties of atomic orbitals and may be classified as s,p,d,f,... types according to their angular properties.

A basis set with the smallest number of functions (only enough functions to constrain the electrons of neutral atoms) is termed a minimum basis set. For hydrogen and helium, the minimum basis set is a single s-type function, and for first row species it is two s-type functions (1s, 2s) and one set of p-type functions ($2p_x, 2p_y, 2p_z$). The next step up from the minimum set is a double zeta (DZ) type basis, this doubles all basis function in the minimum set. Further improvements in size follow a similar scheme with terminology being;

triple zeta (TZ), quadruple zeta (QZ), pentuple zeta (5Z) and so on like that. The use of the word zeta comes from the exponent of the STO basis functions (see Section A.2.1).

The importance of basis sets larger than the minimum can be seen from the bonds within the HCN molecule. Here, the H–C bond will primarily consist of the hydrogen s-orbital and the carbon p_z -orbital. Whereas, the π -bond between carbon and nitrogen will consist mostly of p_x and p_y orbitals of C and N. As such, a minimum basis set would need to find a compromise between these bonds, while a DZ type basis can have one function primarily describe the H–C bond and the other function describe the $C\equiv N$ bond.

Since chemical bonding occurs between valence orbitals, doubling the core orbitals provides little improvement on molecular properties. Thus, doubling core orbitals is rarely considered in actual calculations, this is termed as a split valence basis. A split valence DZ type basis would explicitly be termed a VDZ type basis, however, in practice core orbital doubling is ignored and a DZ type basis can typically be assumed to be a VDZ type basis.

Polarization functions are functions with higher angular momentum than the largest angular momentum of the atomic orbital, for instance, the addition of a set of p-orbitals to hydrogen would polarize its basis (with the set of p-orbitals being the polarization functions). In principle, many sets of polarization functions may be added to a small sp-basis, but if an insufficient number of sp-functions have been selected the optimization procedure may try to compensate for the lack of sp-functions by using higher angular momentum functions, thus, creating artifacts. A general rule of thumb is to have at most one less function of higher angular momentum than the number of functions of one higher angular momentum (ie: for 2 p-type functions, you need 3 s-type functions). Adding a single set of polarization functions to a DZ basis forms a double zeta plus polarization (DZP) type basis. Similarly, adding two sets to a TZ basis makes triple zeta plus double polarization (TZ2P) type basis.

A.2.1 Slater and Gaussian Type Atomic Orbitals

One such set of basis functions is Slater-type atomic orbitals (STOs), which have exponential radial parts and are labeled like hydrogen atomic orbitals (1s, 2s, ...). A few examples of their normalized form are shown below, where ζ_i are constants determining the size of the orbitals.

$$\begin{aligned}\varphi_{1s} &= \left(\frac{\zeta_1^3}{\pi}\right)^{\frac{1}{2}} e^{-\zeta_1 r} \\ \varphi_{2s} &= \left(\frac{\zeta_2^5}{96\pi}\right)^{\frac{1}{2}} r e^{-\frac{\zeta_2 r}{2}}\end{aligned}$$

STOs are not well suited for numerical work, as such, Gaussian-type atomic functions were conceived. They are powers of x, y, z multiplied by $e^{-\alpha r^2}$, where α is a constant for the radial extent of the basis function. A few examples of their normalized form are shown below.

$$\begin{aligned}g_s(\alpha, r) &= \left(\frac{2\alpha}{\pi}\right)^{\frac{3}{4}} e^{-\alpha r^2} \\ g_x(\alpha, r) &= \left(\frac{128\alpha^5}{\pi^3}\right)^{\frac{1}{4}} x e^{-\alpha r^2}\end{aligned}$$

The g_s function has the angular symmetry of the s-type atomic orbital, while the g_x, g_y, g_z functions have the angular symmetries of the three p-type atomic orbitals. Linear combinations of higher order gaussian functions give d-type and higher atomic functions. Although, gaussian functions are better suited for numerical work, due to their integrals being able to be evaluated explicitly without numerical integration, they are worse representations of atomic orbitals since they do not have the cusp at the origin like STOs.

Another type of basis function is contracted gaussians, which use linear combinations of gaussian functions as basis functions. The gaussian function in the linear combination are termed primitive gaussians, and the constants are fixed.

A.2.2 Pople Style Basis Sets

These basis sets, designed by Pople and coworkers, are split valence type and are represented in the following format k-nlmG. The k is the number of primitive gaussian type orbitals (PGTOs) used for the core orbitals, and the nlm indicate how many PGTOs are used for each valence orbital split. Some examples of this notation style are explained below, note that any Pople basis set can also include diffuse and/or polarization functions which adds to the notation style.

3-21G: 3 PGTOs for the core orbitals, 2 PGTOs for the first valence orbital split, and 1 PGTO for the second valence orbital split (note the valence orbitals are split into two functions).

6-311G: 6 PGTOs for the core orbitals, 3 PGTOs for the first valence orbital split, 1 PGTO for the second valence orbital split, and 1 PGTO for the third valence orbital split (note the valence orbitals are split into three functions).

Diffuse functions are denoted by the + symbol before the G, with + meaning one set of diffuse s and p-functions on heavy atoms, and ++ meaning one set of diffuse s and p-functions on heavy atoms plus a diffuse s-function on hydrogen. Meanwhile, polarization functions are denoted by a parenthesis enclosed section after the G with a separate designation for functions added to heavy atoms and those added to hydrogen. For polarization functions, (2df, 2pd) would denote that there are 2 d-functions and 1 f-function added to heavy atoms, while hydrogen gets 2 p-functions and 1 d-function.

A.2.3 Atomic Natural Orbital Basis Sets

Atomic natural orbital (ANO) basis sets are large PGTOs contracted down to a small number of contracted gaussian type orbitals (CGTOs) by using natural orbitals from a correlated

calculations of the free atom, typically at the *configuration interaction with single and double excitations* (CISD) level of theory. Natural orbitals are those that diagonalize the density matrix (see Section A.3), and their eigenvalues are called orbital occupation numbers. The orbital occupation numbers are the number of electrons in the orbital, with a closed-shell system having all orbital occupation numbers equal to either 0 or 2. ANO basis sets have the unique property of smaller basis sets being true subsets of larger basis sets, that is, they use the same set of primitive functions.

A.2.4 Correlation Consistent Basis Sets

The primary disadvantage of ANO basis sets is that they require a large number of PGTs to converge toward the basis set limit. As such, Dunning and coworkers created the correlation consistent (cc) basis sets. The name comes from the basis sets being designed such that functions that contribute similar correlation energy are included at the same stage, independent of function type. The s and p-basis function exponents are optimized at the Hartree-Fock (HF) level for free atoms, while the polarization functions are optimized at CISD level. The primitive functions are contracted by a general contraction scheme using natural orbital coefficients.

These basis sets follow the notation convention of cc-pVxZ (where x = D, T, Z, 5, 6), with the p denoting polarization. For example, cc-pVDZ would be correlation consistent polarized valence double zeta. Each step up in terms of quality adds one more of each type of already existing function, plus adds a new type of higher order polarization function. Augmenting these basis sets with diffuse functions is denoted by the addition of a aug- prefix, for example, aug-cc-pVDZ would be cc-pVDZ with the addition of diffuse functions. They can also be augmented with additional tight functions (if core-core or core-valence electron correlation is of interest), producing acronyms of the style cc-pCVxZ (where x = D, T, Z, 5, 6).

A.2.5 Polarization Consistent Basis Sets

The polarization consistent (pc) basis sets are analogous to the cc basis sets except, as the name implies, they are geared toward describing the polarization of the electron density rather than describing the correlation energy. These basis sets are optimized for density functional theory (DFT), and they are denoted by the acronym pc-n where n indicates the level of polarization (ie. pc-0 \rightarrow unpolarized, and pc-1 \rightarrow single polarization function). Similar to the cc basis sets, adding diffuse functions produces the acronym convention aug-pc-n.

A.3 Hartree-Fock Theory

Hartree-Fock theory is based on the variational method, which is applied to determine optimum orbitals in the single determinant wavefunction. Here, a basis set is used for orbital expansion and the coefficients, $c_{\mu i}$, are varied to minimize the expectation value of the energy. This results in an energy as close to the exact energy as possible, given the limitations imposed by the single determinant wavefunction and the particular basis set employed. The variational equation used to minimize the energy, E' , is given by,

$$\frac{\partial E'}{\partial c_{\mu i}} = 0 \tag{A.4}$$

which is carried out across all μ and i .

A.3.1 Closed-Shell Systems

The variational equation (Equation A.4), leads to a set of algebraic equations for $c_{\mu i}$, which for the closed-shell wavefunction are called the Roothaan-Hall equations.

$$\sum_{\nu=1}^N (F_{\mu\nu} - \epsilon_i S_{\mu\nu}) c_{\nu i} = 0 \quad \mu = 1, 2, \dots, N \tag{A.5}$$

The normalization conditions for the above equation is given by,

$$\sum_{\mu=1}^N \sum_{\nu=1}^N c_{\mu i}^* S_{\mu\nu} c_{\nu i} = 1 \quad (\text{A.6})$$

where ε_i is the one-electron energy of molecular orbital ψ_i , $S_{\mu\nu}$ is the overlap matrix and $F_{\mu\nu}$ is the Fock matrix. The overlap matrix is given by,

$$S_{\mu\nu} = \int \varphi_{\mu}^*(1) \varphi_{\nu}(1) dx_1 dy_1 dz_1 \quad (\text{A.7})$$

and the Fock matrix by,

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \sum_{\lambda=1}^N \sum_{\sigma=1}^N P_{\lambda\sigma} [(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma)] \quad (\text{A.8})$$

where $H_{\mu\nu}^{\text{core}}$ is a matrix representing the energy of a single electron in a field of “bare” nuclei, $P_{\lambda\sigma}$ is the one-electron density matrix and $(\mu\nu|\lambda\sigma)$ & $(\mu\lambda|\nu\sigma)$ are two-electron repulsion integrals. The $H_{\mu\nu}^{\text{core}}$ matrix is given by,

$$H_{\mu\nu}^{\text{core}} = \int \varphi_{\mu}^*(1) \hat{H}^{\text{core}}(1) \varphi_{\nu}(1) dx_1 dy_1 dz_1$$

$$\hat{H}^{\text{core}}(1) = -\frac{1}{2} \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2} \right) - \sum_{A=1}^M \frac{Z_A}{r_{1,A}}$$

where Z_A is the atomic number of atom A, and M is the total number of atoms in the system.

The two-electron repulsion integrals are given by,

$$(\mu\nu|\lambda\sigma) = \int \int \varphi_{\mu}^*(1) \varphi_{\nu}(1) \left(\frac{1}{r_{1,2}} \right) \varphi_{\lambda}^*(2) \varphi_{\sigma}(2) dx_1 dy_1 dz_1 dx_2 dy_2 dz_2$$

$$(\mu\lambda|\nu\sigma) = \int \int \varphi_{\mu}^*(1) \varphi_{\lambda}(1) \left(\frac{1}{r_{1,2}} \right) \varphi_{\nu}^*(2) \varphi_{\sigma}(2) dx_1 dy_1 dz_1 dx_2 dy_2 dz_2$$

and the one-electron density matrix,

$$P_{\lambda\sigma} = 2 \sum_{i=1}^{\text{occ}} c_{\lambda i}^* c_{\sigma i} \quad (\text{A.9})$$

where the summation is over occupied molecular orbitals only, and the factor of 2 indicates that two electrons occupy each molecular orbital. The sum of the electronic energy, E^{ee} , and the internuclear repulsion energy, E^{nr} , gives the total energy of the system.

$$\begin{aligned} E &= E^{\text{ee}} + E^{\text{nr}} \\ E &= \frac{1}{2} \text{tr}(P_{\mu\nu} [F_{\mu\nu} + H_{\mu\nu}^{\text{core}}]) + E^{\text{nr}} \\ E &= \frac{1}{2} \text{tr}(P_{\mu\nu} [F_{\mu\nu} + H_{\mu\nu}^{\text{core}}]) + \sum_A^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \end{aligned}$$

Since the Fock matrix itself depends on the molecular orbital coefficients, through the density matrix, these equations are not linear. Thus the resulting molecular orbitals are derived from their own effective potential, hence this technique has been termed self-consistent-field (SCF) theory.

A.3.2 Open-Shell Systems

Molecular orbital theory can be extended to open-shell systems through two methods, spin-restricted Hartree-Fock (RHF) theory and spin-unrestricted Hartree-Fock (UHF) theory. RHF theory allows molecular orbitals to be singly occupied with only an α electron, and uses only a single set of molecular orbitals. Figure A.2 shows an example of a doublet state configuration in RHF theory.

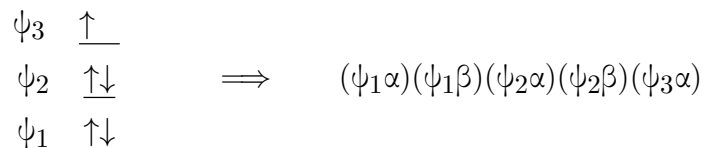


Figure A.2: An example of a doublet state electronic configuration in RHF theory.

On the other hand, UHF theory assigns separate orbitals to α and β electrons, thus forming two distinct sets of molecular orbitals. Figure A.3 shows an example of a doublet state configuration in UHF theory.

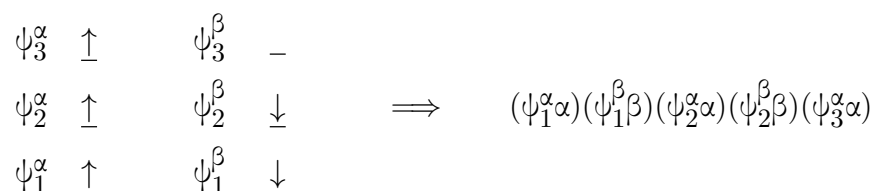


Figure A.3: An example of a doublet state electronic configuration in UHF theory.

Since the RHF function is a special case of the UHF function, it follows from the variational method that the UHF energy will always be lower than the RHF energy. However, UHF functions have the disadvantage of not being true eigenfunctions of the total spin operator. Thus, UHF wavefunctions designed for doublet states are contaminated by functions of higher spin multiplicities.

A.4 Electron Correlation Methods

With a sufficiently large basis set the HF wavefunction can account for 99% of the total energy. Although small, the remaining 1% is important for describing chemical phenomena. The missing 1% is due to HF methods replacing the real electron-electron interactions with an average interaction. The difference in energy between the HF energy and the lowest possible energy for a given basis is called the electron correlation (EC) energy. Physically, the EC energy corresponds to the motion of electrons being influenced by the presence of other electrons, hence the name electron correlation (because the motion is correlated).

There exists two types of EC, one is the correlation of electrons with opposite spin, termed Coulomb correlation. The other is correlation of electrons with the same spin, termed Fermi correlation. EC can also be split into dynamic and static, where the dynamic contribution is associated with “instant” correlation (that of electrons in the same orbital) and static contribution is associated with “permanent” correlation (that of electrons in different orbitals). The single-determinant wavefunction approximation used in HF methods does not take into account the Coulomb correlation, this is what results in the EC energy gap. To a certain degree, the Pauli exclusion principle takes into account the Fermi correlation. The UHF type wavefunctions can, to a certain degree, take into account Coulomb correlation effects. However, only by incorporating multiple determinants can EC energy be properly taken into account.

EC methods typically start with a HF wavefunction (Φ_{HF}), since it is 99% correct, and expand on it in a similar method to basis set expansions. A typical wavefunction expansion is shown below, where a_0 is usually close to 1.

$$\Psi = a_0 \Phi_{\text{HF}} + \sum_{i=1} a_i \Phi_i \quad (\text{A.10})$$

From this point, different EC methods use different approaches to solve for the coefficients a_i , with some of the most common methods being configuration interaction (CI), many-body perturbation theory (MBPT) and coupled cluster (CC).

A.4.1 Excited Slater Determinants

Given a system of N electrons and M basis functions, an RHF calculation will produce a solution to the Roothaan-Hall equations with $\frac{1}{2}N$ occupied MOs, and $M - \frac{1}{2}N$ unoccupied (virtual) MOs. With the exception of the minimum basis there will always be more virtual MOs than occupied MOs. A series of determinants can be made by replacing occupied MOs in the HF determinant with virtual MOs. These are appropriately named singly, doubly, triply, etc. excited determinants depending on how many MOs are replaced. All of the

EC energy for a given basis can be recovered if all excited determinants are included in the wavefunction.

In most cases only the EC energy of the valence electrons is of interest, limiting the number of excited determinants to only those that can be generated by exciting the valence electrons is called the frozen-core approximation. Another such approximation, is called the frozen virtuals approximation, where the anti-bonding contributions of the core orbitals are also removed from the correlation calculation. The justification for the frozen-core approximation is that the EC energy from the core is fixed, and so it drops out when relative energies are investigated.

A.4.2 Møller-Plesset Perturbation Theory

In Section A.3 concepts were explained in matrix form with explicit integrals, however, for simplicity it is wiser to continue in Dirac notation. As such, the Fock matrix will be replaced by the Fock operator (its constructor).

$$\hat{F}_i = \hat{h}_i \sum_j^N (\hat{J}_j - \hat{K}_j) \quad (\text{A.11})$$

Here, N is the number of electrons, \hat{h}_i is the one-electron Hamiltonian, \hat{J}_j is the Coulomb operator and \hat{K}_j is the exchange operator. Møller-Plesset (MP) theory considers the sum over Fock operators as the unperturbed Hamiltonian (see Section ??).

$$\hat{H}_0 = \sum_i^N \hat{F}_i \quad (\text{A.12})$$

The perturbation utilized by MP theory is contrary to conventional perturbation theory since it is not small compared with the unperturbed Hamiltonian.

$$\hat{H}' = \hat{V}_{ee} - 2 \langle \hat{V}_{ee} \rangle \quad (\text{A.13})$$

In the perturbation, \hat{V}_{ee} is the electron-electron repulsion potential. The zeroth-order wavefunction is the HF determinant and the zeroth-order energy is the sum of MO energies (ϵ_i), which overcounts the electron-electron repulsion energy.

$$W_0 = \langle \Phi_0 | \hat{H}_0 | \Phi_0 \rangle = \langle \Phi_0 | \sum_i^N \hat{F}_i | \Phi_0 \rangle = \sum_i^N \epsilon_i \quad (\text{A.14})$$

The first-order energy correction yields a correction to the overcounting in the zeroth-order energy. Meaning that the sum of the first two energy terms gives the HF energy ($W_0 + W_1 = E(\text{HF})$).

$$W_1 = \langle \Phi_0 | \hat{H}' | \Phi_0 \rangle = \langle \hat{V}_{ee} \rangle - 2 \langle \hat{V}_{ee} \rangle = - \langle \hat{V}_{ee} \rangle \quad (\text{A.15})$$

Conventional notation is to use MPn to denote the sum of MP energy corrections up to n.

$$\text{MP}_n = \sum_i^n W_i \quad \text{MP}_0 = \sum_i^N \epsilon_i \quad \text{MP}_1 = E(\text{HF}) = \sum_i^N \epsilon_i - \langle \hat{V}_{ee} \rangle \quad (\text{A.16})$$

The second-order energy correction is the first energy correction to consider EC energy. It involves all possible excited determinants, however, since the perturbation is a two-electron operator all elements involving triple or higher excited determinants will vanish. Similarly, all singly excited determinant elements also vanish, thus the second-order energy correction only involves a sum over doubly excited determinants. This is done by promoting electrons from occupied orbitals i and j to virtual orbitals a and b with restrictions on the summation to prevent double counting.

$$W_2 = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} \frac{\langle \Phi_0 | \hat{H}' | \Phi_{ij}^{ab} \rangle \langle \Phi_{ij}^{ab} | \hat{H}' | \Phi_0 \rangle}{E_0 - E_{ij}^{ab}} \quad (\text{A.17})$$

Invoking a similar principle to Koopman's theorem, which states that the ionization energy of a system is equal to the negative of the highest occupied molecular orbital energy, the second-order energy correction simplifies.

$$W_2 = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} \frac{(\langle \varphi_i \varphi_j | \varphi_a \varphi_b \rangle - \langle \varphi_i \varphi_j | \varphi_b \varphi_a \rangle)^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (\text{A.18})$$

The second-order energy correction is then just the sum over two-electron integrals over MOs. There are M_{basis}^4 integrals, however, the transformation from AO to MO grows as M_{basis}^5 , so MP2 is an M_{basis}^5 method. Since not all integrals need to be calculated the MP2 cost is similar to the HF cost, however, MP2 accounts for roughly 80-90% of the EC energy. The HF wavefunction already includes Fermi correlation, as such, MP2 describes the Fermi and electron correlation with different accuracies. To counter this, the spin component scaled MP2 (SCS-MP2) method was created. Here, both the same spin (Fermi correlation) and opposite spin (EC) components of the MP2 correlation energy are scaled by empirical constants.

MP2 describes the correlation between pairs of electrons, while MP3 describes the interaction between pairs of electrons. MP3 computationally scales as M_{basis}^6 and accounts for roughly 90-95% of the EC energy. It has been found that MPn methods oscillate around its limiting value (MP_∞) as shown in Figure A.4. Interestingly, it has been found that MP2 gives better results than MP3, despite it overshooting the correlation effect.

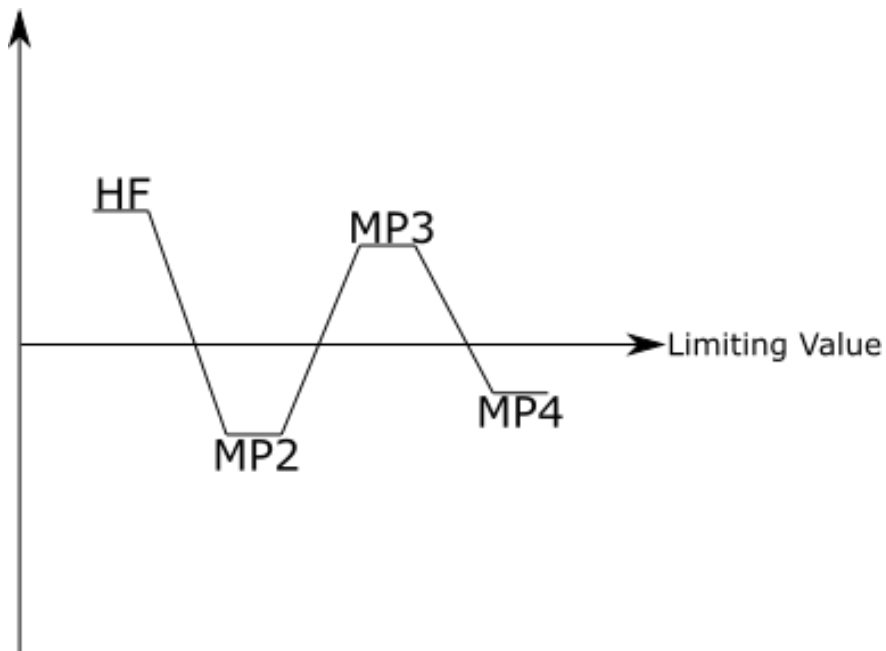


Figure A.4: An example plot of the oscillatory behavior of MP theory as a function of n .

The main limitation of perturbation theory methods is that they assume the zeroth-order wavefunction is a reasonable approximation of the real wavefunction. This means that as the HF wavefunction description of the system gets worse, the larger the correction terms become and the more are needed to reach a given level of accuracy.

A.4.3 Couple Cluster Theory

The idea in Coupled Cluster (CC) methods is to include all corrections of a given type to infinite order. For this, an excitation operator (\hat{T}) is constructed.

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_N \quad (\text{A.19})$$

Where N is the number of electrons, and the \hat{T}_i operator is constructed such that when it acts on the reference HF wavefunction (Φ_0) it generates all i th excited Slater determinants.

$$\hat{T}_1 = \sum_i \sum_a t_a^i \hat{a}^a \hat{a}_i \quad \hat{T}_2 = \frac{1}{4} \sum_{i,j} \sum_{a,b} t_{ij}^{ab} \hat{a}^a \hat{a}^b \hat{a}_i \hat{a}_j \quad (\text{A.20})$$

Here, \hat{T}_1 & \hat{T}_2 are described in the second quantization formalism where $\hat{a}^a = \hat{a}_a^\dagger$ and \hat{a}_i are creation and annihilation operators. The constants t_1^a & t_{ij}^{ab} are called amplitudes. When these operators act on the reference ket ($|\Phi_0\rangle$) it results in excited kets.

$$\hat{T}_1 |\Phi_0\rangle = \sum_i^{\text{occ}} \sum_a^{\text{vir}} t_1^a |\Phi_1^a\rangle \quad \hat{T}_2 |\Phi_0\rangle = \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{vir}} t_{ij}^{ab} |\Phi_{ij}^{ab}\rangle \quad (\text{A.21})$$

With the excitation operator a CI or CC wavefunction can be generated.

$$\Psi_{\text{CI}} = (1 + \hat{T}) |\Phi_0\rangle \quad \Psi_{\text{CC}} = e^{\hat{T}} |\Phi_0\rangle \quad (\text{A.22})$$

The exponential operator can be Taylor expanded, and then grouped by terms that produce singly, doubly and so on excited determinants.

$$e^{\hat{T}} \approx 1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{6}\hat{T}^3 + \dots = \sum_{k=0}^{\infty} \frac{1}{k!} \hat{T}^k$$

$$e^{\hat{T}} \approx 1 + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2}\hat{T}_1^2) + (\hat{T}_3 + \hat{T}_2\hat{T}_1 + \frac{1}{6}\hat{T}_1^3) + \dots$$

By comparison, the CC wavefunction includes all the same terms as the CI wavefunction plus the products of excitation operators which are not in the CI wavefunction. These product terms are what provide the increased accuracy in predicting the EC energy by CC versus CI methods, and physically these terms represent non-interacting pairs of interacting electrons. As an example, \hat{T}_4 represents four electrons simultaneously interacting with each other, and \hat{T}_2^2 represents two-pairs of two electrons interacting but the pairs are not interacting with the other pair. The conventional terminology is to refer to the product terms as *disconnected* and non-product terms as *connected*.

The CC Schrödinger equation is simply written as follows.

$$\hat{H}e^{\hat{T}} |\Phi_0\rangle = E_{\text{CC}} e^{\hat{T}} |\Phi_0\rangle \quad (\text{A.23})$$

Attempting to solve this via the variational method results in a series of non-vanishing terms all the way up to N . As such, it would be unfeasible to computationally solve it for any but the smallest of systems. It is for this reason that the standard method proceeds by projecting the CC Schrödinger equation onto the reference wavefunction. Multiplying from the left by $\langle \Phi_0 |$ allows for the equation to be solved for the energy.

$$\begin{aligned}\langle \Phi_0 | \hat{H} e^{\hat{T}} | \Phi_0 \rangle &= E_{CC} \langle \Phi_0 | e^{\hat{T}} | \Phi_0 \rangle \\ \langle \Phi_0 | \hat{H} e^{\hat{T}} | \Phi_0 \rangle &= E_{CC} \langle \Phi_0 | (1 + \hat{T}_1 + \dots) | \Phi_0 \rangle \\ \langle \Phi_0 | \hat{H} e^{\hat{T}} | \Phi_0 \rangle &= E_{CC} \langle \Phi_0 | \Phi_0 \rangle \\ \implies E_{CC} &= \langle \Phi_0 | \hat{H} e^{\hat{T}} | \Phi_0 \rangle\end{aligned}$$

Since the Hamiltonian only contains one and two-electron operators, the exponential operator can be reduced to only terms with singly or doubly excited determinants.

$$\begin{aligned}E_{CC} &= \langle \Phi_0 | \hat{H} (1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2} \hat{T}_1^2) | \Phi_0 \rangle \\ E_{CC} &= E_0 + \sum_i^{\text{occ}} \sum_a^{\text{vir}} t_i^a \langle \Phi_0 | \hat{H} | \Phi_i^a \rangle + \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} (t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a) \langle \Phi_0 | \hat{H} | \Phi_{ij}^{ab} \rangle \\ E_{CC} &= E_0 + \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} (t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a) \langle \Phi_0 | \hat{H} | \Phi_{ij}^{ab} \rangle \\ E_{CC} &= E_0 + \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} (t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a) (\langle \varphi_i \varphi_j | \varphi_a \varphi_b \rangle - \langle \varphi_i \varphi_j | \varphi_b \varphi_a \rangle)\end{aligned}$$

The final equation shows that the CC EC energy is entirely determined by the singles and doubles amplitudes and the two-electron MO integrals.

A.5 Density Functional Theory

Density functional theory (DFT) is built around the Hohenberg-Kohn theorem, which states that the ground state electronic structure is a unique functional dependent on the electron density. The importance of this is that the complexity of the wavefunction increases exponentially with the number of electrons, but the electron density has the same number of variables independent of the size of the system. Given that the correspondence between electron density and ground state structure is one-to-one, the aim of DFT is then to design a functional that maps these two properties.

The simplest model is the local density approximation (LDA), where the electron density is assumed to be slowly varying such that the EC energy can be calculated from formulas for a uniform electron density. The main disadvantage of DFT is its lack of ability to systematically improve the results. This has resulted in a large number of different functionals being designed with slightly different performances.

A.5.1 Kohn-Sham Theory

The main idea in the Kohn-Sham (KS) formalism is to split the kinetic energy functional into two parts, one that is solved exactly and a small correction term. The unfortunate part of this is that it returns the complexity of the problem back to $3N$ and re-separates the EC term. Although it may seem disadvantageous to do this, the alternative (orbital-free DFT) is typically outperformed by modern approaches.

The exact kinetic energy functional is given by,

$$T = \sum_{i=1}^N \langle \varphi_i | -\frac{1}{2} \nabla^2 | \varphi_i \rangle \tag{A.24}$$

where N is the number of electrons. Through natural orbitals (NOs) the exact kinetic energy can be calculated from the exact density matrix.

$$T[\rho] = \sum_{i=1}^{\infty} n_i \langle \varphi_i^{\text{NO}} | \frac{1}{2} \nabla^2 | \varphi_i^{\text{NO}} \rangle \quad \rho = \sum_{i=1}^{\infty} n_i |\varphi_i^{\text{NO}}|^2 \quad (\text{A.25})$$

Here, n_i are the orbital occupation numbers whose sum equal the total number of electrons ($N = \sum_{i=1}^{\infty} n_i$). Since the exact density isn't known, an approximation can be found from a set of auxiliary functions.

$$\rho \approx \sum_{i=1}^N |\varphi_i|^2 \quad (\text{A.26})$$

In the KS formalism the kinetic energy is calculated under the assumption of non-interacting electrons. A general expression for the DFT calculated energy is given by,

$$E_{\text{DFT}}[\rho] = T_{\text{S}}[\rho] + E_{\text{ne}}[\rho] + J[\rho] + E_{\text{xc}}[\rho] \quad (\text{A.27})$$

$$E_{\text{xc}}[\rho] = (T[\rho] - T_{\text{S}}[\rho]) + (E_{\text{ee}}[\rho] - J[\rho]) \quad (\text{A.28})$$

where only the EC functional is unknown. This is the task of KS theory, that is, it is necessary to derive approximations for the exchange-correlation energy functional only.

A.5.2 Exchange-Correlation Functionals

Although the exact exchange-correlation functional is unknown, many of its more important properties are well known. The first of which is that the functional should be self-interaction free, meaning that the exchange energy for a one-electron system should exactly cancel out the Coulomb energy. The second, is that when the density becomes constant, the uniform electron gas result should be recovered. The third, is that the coordinate scaling of the exchange energy should be linear, as is shown by the equation below.

$$\rho_{\lambda}(x, y, z) = \lambda^3 \rho(\lambda x, \lambda y, \lambda z) \quad E_{\text{x}}[\rho_{\lambda}] = \lambda E_{\text{x}}[\rho] \quad (\text{A.29})$$

A fourth property is that scaling the electron-coordinates by a factor larger than 1, should increase the magnitude of the correlation (the reverse is also applicable). The fifth property is that as the scaling parameter goes to infinity, the energy for a finite system should converge on a negative constant. Finally, the exchange potential and correlation potential should show an asymptotic $-r^{-1}$ and $-\frac{1}{2}\alpha r^{-4}$ behavior, respectively. The quality of an exchange-correlation functional is determined by comparison with experiments or high-level wave-mechanics calculations.

Local Density Approximation

In the local density approximation (LDA), it is assumed that the local density can be treated like a uniform electron gas. The exchange energy for a uniform electron gas is given by the Dirac formula.

$$E_x^{\text{LDA}}[\rho] = -C_x \int \rho^{\frac{4}{3}}(\vec{r}) d\vec{r} \quad \epsilon_x^{\text{LDA}} = -C_x \rho^{\frac{1}{3}} \quad (\text{A.30})$$

In more general cases, where the α and β spin electron densities are not equal, LDA is replaced by local spin density approximation (LSDA).

$$E_x^{\text{LSDA}}[\rho] = -2^{\frac{1}{3}} C_x \int (\rho_\alpha^{\frac{4}{3}} + \rho_\beta^{\frac{4}{3}}) d\vec{r} \quad (\text{A.31})$$

Naturally, for closed-shell species LSDA is equivalent to LDA. The analytical form of the correlation energy has been derived for low and high density uniform electron gas. Whereas, intermediate densities require the correlation energy by determined by quantum Monte Carlo methods. Analytic interpolation formulas are then implemented in DFT calculations to cover the correlation energy across all densities. Two popular interpolation formulas are the Vosko-Wilk-Nusair (VWM) and Perdew-Wang (PW) formulas. Typically, LDA methods produce results with similar accuracy to HF methods

Generalized Gradient Approximation

To improve on LDA methods, it is necessary to consider the density is not that of a uniform electron gas. Including a dependence on the gradient of the density in the exchange and correlation functionals would do that. However, this additional term would make it such the requirements of the Fermi hole integrating to -1 and the Coulomb hole integrating to 0 would not be fulfilled. The generalized gradient approximation (GGA) method includes the first derivative of the density, along with a requirement that the Fermi and Coulomb holes integrate to -1 and 0, respectively.

One of the most popular GGA exchange-functional is the Becke (B or B88) functional.

$$\epsilon_x^{\text{B88}} = \epsilon_x^{\text{LDA}} + \delta\epsilon_x^{\text{B88}} \quad \delta\epsilon_x^{\text{B88}} = -\beta\rho^{\frac{1}{3}} \frac{x^2}{1 + 6\beta x \sinh^{-1} x} \quad x = \frac{|\nabla\rho|}{\rho^{\frac{4}{3}}} \quad (\text{A.32})$$

Here, the β parameter is determined by fitting to known data. A popular GGA correlation-functional is the Lee-Yang-Parr (LYP) functional. The LYP functional is often combined with the B88 functional to make the GGA DFT method BLYP.

Meta-GGA Functionals

Logically, the next improvement to GGA methods is to include higher order derivatives in the exchange and correlation functions.

$$\tau(\vec{r}) = \frac{1}{2} \sum_i^{\text{occ}} \epsilon_i |\varphi_i(\vec{r})|^2 - v_{\text{eff}}(\vec{r})\rho(\vec{r}) + \frac{1}{2} \nabla^2 \rho(\vec{r}) \quad (\text{A.33})$$

Adding the Laplacian forms what is called a meta-GGA exchange-correlation functional.

Hybrid GGA Methods

Hybrid GGA methods use a mix of GGA, LDA, and HF energies to get a full system energy. The formulation for hybrid GGAs starts with considering the adiabatic connection formula.

$$E_{xc} = \int_0^1 \langle \psi_\lambda | \hat{V}_{xc}^{\text{hole}}(\lambda) | \psi_\lambda \rangle d\lambda \quad (\text{A.34})$$

Here, lambda is a parameter that activates the electron-electron interaction. This means that when $\lambda = 0$, the electrons are non-interacting and there is only exchange energy. Assuming $\hat{V}_{xc}^{\text{hole}}$ is linear with respect to λ , the exchange-correlation energy can be approximated as follows.

$$E_{xc} \approx \frac{1}{2} (\langle \psi_0 | \hat{V}_{xc}^{\text{hole}}(0) | \psi_0 \rangle + \langle \psi_1 | \hat{V}_{xc}^{\text{hole}}(1) | \psi_1 \rangle) \quad (\text{A.35})$$

Given that the $\lambda = 0$ case is non-interacting, then as long as the orbitals are those of the HF method, then the first term in the above expression is just the HF exchange energy.

$$\langle \psi_0 | \hat{V}_{xc}^{\text{hole}}(0) | \psi_0 \rangle = E_x^{\text{HF}} \quad (\text{A.36})$$

The second term is still unknown, however, it can be approximated by LDA type calculations.

$$\langle \psi_1 | \hat{V}_{xc}^{\text{hole}}(1) | \psi_1 \rangle = E_x^{\text{LDA}} + E_c^{\text{LDA}} \quad (\text{A.37})$$

Combining these expressions produces an expression for the exchange-correlation energy termed the half-and-half (H + H) method.

$$E_{xc}^{\text{H+H}} = \frac{1}{2} E_x^{\text{HF}} + \frac{1}{2} (E_x^{\text{LDA}} + E_c^{\text{LDA}}) \quad (\text{A.38})$$

Following a similar methodology, but with GGA methods rather than LDA results in what is termed a hybrid GGA functional. One popular hybrid GGA method is B3LYP, which uses B88 and HF exchange, LYP correlation and 3 parameters (hence B3LYP).

$$E_{xc}^{\text{B3LYP}} = (1-a)E_x^{\text{LDA}} + aE_x^{\text{HF}} + b\delta E_x^{\text{B88}} + (1-c)E_c^{\text{LDA}} + cE_c^{\text{LYP}} \quad (\text{A.39})$$

For B3LYP, the parameters a , b , and c are determined by fitting to experimental data.

APPENDIX B. MOLECULAR VIBRATIONS

B.1 Normal Coordinates

Quantum mechanical treatment of molecular vibrations is done in a new coordinate system, termed normal coordinates. The normal coordinates relate to the displacement of atoms from their equilibrium position, with respect to a normal mode of vibration (each normal mode is assigned one normal coordinate). The reason for adopting this coordinate system is that, under the assumption of small displacements, normal coordinates decouple the vibrational degrees of freedom from the rotational and translational degrees of freedom. The normal coordinates, Q_k , are defined in terms of the mass-weighted cartesian displacement coordinates q_i .

$$Q_k = \sum_{i=1}^{3N} l''_{ki} q_i \quad k = 1, 2, 3, \dots, 3N \quad (\text{B.1})$$

Here, N is the number of atoms and the coefficients l''_{ki} are chosen such that the kinetic energy and potential energy have the form

$$2T = \sum_{k=1}^{3N} \dot{Q}_k^2 \quad 2V = \sum_{k=1}^{3N} \lambda'_k Q_k^2 \quad (\text{B.2})$$

where the constants λ'_k are related to the normal frequencies. The equations of motion then become

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{Q}_k} + \frac{\partial V}{\partial Q_k} = \ddot{Q}_k + \lambda'_k Q_k = 0 \quad k = 1, 2, 3, \dots, 3N \quad (\text{B.3})$$

and the solutions are

$$Q_k = K'_k \cos(\lambda_k'^{\frac{1}{2}} t + \varepsilon'_k) \quad k = 1, 2, 3, \dots, 3N \quad (\text{B.4})$$

where K'_k and ε'_k are arbitrary constants.

B.2 Selection Rules in Wave Mechanics

The intensity of a spectral line is determined by the probability of the transition which gives rise to the line. The probability, $A_{n'n''}$, of a spontaneous transition from a higher state, n' , to a lower state, n'' , with the emission of light of frequency $\nu_{n'n''}$ is shown below.

$$A_{n'n''} = \frac{64\pi^4\nu_{n'n''}^3}{3hc^3} [|(\mu_x)_{n'n''}|^2 + |(\mu_y)_{n'n''}|^2 + |(\mu_z)_{n'n''}|^2] \quad (\text{B.5})$$

Here, h is Planks constant and c is the speed of light. The terms $(\mu_{x,y,z})_{n'n''}$ are given by the following integral,

$$(\mu_{x,y,z})_{n'n''} = \int \Psi_{n'}^* \mu_{x,y,z} \Psi_{n''} d\tau \quad (\text{B.6})$$

where $\Psi_{n'}^*$ is the complex conjugate of the complete wavefunction of state n' , $\mu_{x,y,z}$ is the x, y, or z component of the electric dipole moment, $\Psi_{n''}$ is the complete wavefunction of state n'' and $d\tau$ is the volume element of the configuration space. The coefficient of absorption equals the coefficient of induced emission and both are given by,

$$B_{n''n'} = B_{n'n''} = \frac{8\pi^3}{3h^2} [|(\mu_x)_{n'n''}|^2 + |(\mu_y)_{n'n''}|^2 + |(\mu_z)_{n'n''}|^2] \quad (\text{B.7})$$

The coefficients of absorption are directly related to the decrease in intensity ($-dI$) of a beam of light as it traverses through an attenuating species.

$$-dI = h\nu_{n'n''} B_{n'n''} \rho(\nu_{n'n''}) (N_{n''} - N_{n'}) dl \quad (\text{B.8})$$

The expression $B_{n'n''} \rho(\nu_{n'n''})$, is simply the probability of a transition from state n' to n'' per unit time in the presence of radiation of frequency $\nu_{n'n''}$ having a density $\rho(\nu_{n'n''})$, and $N_{n'}$ is the number of molecules per unit volume in the state n' (similarly for $N_{n''}$). This decrease in intensity can also be written in another form,

$$-dI = kI dl \quad (\text{B.9})$$

where k is the absorption coefficient. Solving for the absorption coefficient, then integrating it over the length gives an expression commonly referred to as the Beer-Lambert Law.

$$A = \ln\left(\frac{I_0}{I}\right) = \epsilon lc \quad (\text{B.10})$$

Here, A is the absorbance, I_0 is the initial intensity of the beam, I is the final intensity of the beam of light after traversing through the attenuating species, ϵ is the absorptivity of the attenuating species, l is the optical path length and c is the concentration of the attenuating species. Absorbance is a commonly measured quantity in laboratories, and is the principle by which spectrometers collect spectra. As such, it is this absorption coefficient, k , that theoretical calculations generate.

B.2.1 Infrared Intensities

For infrared spectroscopy, these absorption coefficients are commonly referred to as infrared intensities. Solving for these infrared intensities is done by first starting with equation B.8 and substituting in equation B.9.

$$kI dl = h\nu_{n'n''} B_{n'n''} \rho(\nu_{n'n''}) (N_{n''} - N_{n'}) dl \quad (\text{B.11})$$

The radiation flux intensity and density are related by $I = c\rho$, so we can replace density in the above expression.

$$kI dl = \frac{h\nu_{n'n''}}{c} B_{n'n''} (N_{n''} - N_{n'}) dl \quad (\text{B.12})$$

Dividing both sides by $I dl$, then substituting equation B.7.

$$k = \frac{8\pi^3}{3ch} \nu_{n'n''} (N_{n''} - N_{n'}) \sum_{\alpha} |(\mu_{\alpha})_{n'n''}|^2 \quad (\text{B.13})$$

Here, α is a dummy variable for iterating over the x, y, and z components of the dipole moment. In reality, the absorption coefficient would be measured as the sum of bands rather than that of a single band. This is done by integrating k over all line bands, which then gives the real infrared intensities (I_{IR}). Then applying the harmonic approximation to the above expression would give,

$$I_{\text{IR}} = \frac{N\pi}{3c^2} \sum_{\alpha} \left(\sum_{\mathbf{k}} \frac{\partial \mu_{\alpha}}{\partial R_{\mathbf{k}}} Q_{\mathbf{k}} \right)^2 \quad (\text{B.14})$$

where N is the number of molecules per unit volume, c is the speed of light, R is the atomic coordinates, Q is the normal coordinates, μ is the dipole moment. These are the infrared intensities calculated by traditional quantum chemistry packages.

B.2.2 Raman Activities

Unlike infrared intensities, Raman activities calculations are not trivially linked to experiments. This is in part due to Raman scattering being dependent on many experimental conditions. Calculated Raman activities can be converted into Raman intensities, which are more directly related to experimental results. The normal Raman scattering Raman activity for a particular normal mode, S_i , is given by,

$$S_i = g_i [45\bar{\alpha}_i^2 + 7\bar{\beta}_i^2] \quad (\text{B.15})$$

where g_i is the degeneracy of the i th normal mode, $\bar{\alpha}_i^2$ and $\bar{\beta}_i^2$ are given by the expressions below.

$$\bar{\alpha}_i^2 = \frac{1}{9} \left(\frac{\partial \alpha_{xx}}{\partial Q_i} + \frac{\partial \alpha_{yy}}{\partial Q_i} + \frac{\partial \alpha_{zz}}{\partial Q_i} \right)^2$$

$$\bar{\beta}_i^2 = \frac{1}{2} \left[\left(\frac{\partial \alpha_{xx}}{\partial Q_i} - \frac{\partial \alpha_{yy}}{\partial Q_i} \right)^2 + \left(\frac{\partial \alpha_{xx}}{\partial Q_i} - \frac{\partial \alpha_{zz}}{\partial Q_i} \right)^2 + \left(\frac{\partial \alpha_{yy}}{\partial Q_i} - \frac{\partial \alpha_{zz}}{\partial Q_i} \right)^2 + 6 \left[\left(\frac{\partial \alpha_{xy}}{\partial Q_i} \right)^2 + \left(\frac{\partial \alpha_{yz}}{\partial Q_i} \right)^2 + \left(\frac{\partial \alpha_{xz}}{\partial Q_i} \right)^2 \right] \right]$$

Here, α is the electric dipole–electric dipole polarizability tensor. Finally, the calculated Raman activity can be converted into a Raman intensity for direct comparison with experiment by the following equation.

$$I_i = \frac{(\nu_0 - \nu_i)^4 S_i}{\nu_i (1 - e^{-\frac{h c \nu_i}{k_B T}})} \quad (\text{B.16})$$

where S_i is the Raman activity of vibrational mode i , I_i is the Raman intensity of vibrational mode i , ν_0 is the frequency of the excitation laser, ν_i is the frequency of vibrational mode i , T is the absolute temperature, h is Planck constant, c is the speed of light, and k_B is the Boltzmann constant.

APPENDIX C. DIAGNOSTICS

C.1 Coupled Cluster Diagnostics

C.1.1 %TAE[(T)]

The “%TAE[(T)]” diagnostic is a method for measuring multi-reference character of molecular species. This is done by measuring the percent difference between the total atomization energy (TAE) at the CCSD(T) level and the TAE at the CCSD level.

$$\%TAE[(T)] = 100\left(\frac{TAE[CCSD(T)] - TAE[CCSD]}{TAE[CCSD(T)]}\right) \quad (C.1)$$

The TAE for a molecule, X_iY_j , at a specific level of theory, CCSD/cc-pVTZ, is given by the difference between the energies calculated for atoms versus the energy calculated for the molecule as a whole. For simplicity, let's consider the following notation: CCSD[X] means a CCSD/cc-pVTZ calculation of the atomic energy of X. The equation below then shows how to calculate the TAE[CCSD] required for calculating the %TAE[(T)] shown above (An identical formulation is used for CCSD(T), only the level of theory for energy calculations is changed).

$$TAE[CCSD] = i(CCSD[X]) + j(CCSD[Y]) - (CCSD[X_iY_j]) \quad (C.2)$$

The %TAE[(T)] value is an indicator of the expected contributions of post-CCSD(T) methods will introduce. Typically, values below 5% indicate that post-CCSD(T) contributions will be minimal and calculations at the CCSD(T) are sufficient. Whereas, values above 10% indicate significant post-CCSD(T) contributions and CCSD(T) is not sufficient. Values in between 5% and 10% indicate moderate post-CCSD(T) contributions and require caution when using CCSD(T).

C.1.2 T_1 and T_2 Diagnostics

The T_1 diagnostic is a measure of the reliability of CC calculations. This is done by calculating the norm of the vector of \hat{T}_1 amplitudes (see Section A.4.3). Generally, if the norm is less than 0.02 the CC calculation is considered reliable [208]. Another indicator of the reliability of CC calculations is the largest \hat{T}_2 amplitude (see Section A.4.3). Generally, if the largest amplitude is less than 0.05 the CC calculation is considered reliable [208].

The major difference between the two diagnostics is that the T_1 diagnostic is an averaged value, and therefore does not give indication of the non-homogeneity of the vector of \hat{T}_1 amplitudes. The two diagnostics used in conjunction give a more complete picture of the reliability of the CC calculation.

C.2 DFT Diagnostics

C.2.1 A_λ

The “ A_λ ” diagnostic is a DFT-based method for measuring multi-reference character of molecular species. This is done by measuring the scaled (scaling is dependent on degree of HF-type exchange) percent difference between the TAE of a pure-DFT functional, XC, and its corresponding hybrid functional $X_\lambda C$.

$$A_\lambda = \frac{1}{\lambda} \left(1 - \frac{\text{TAE}[X_\lambda C]}{\text{TAE}[XC]} \right) \quad (\text{C.3})$$

Here, TAE is calculated in the same methodology as was done for the CC diagnostic (Section C.1.1). A more comprehensive interpretation of the A_λ values is given in Fogueri *et al.* [190]. Briefly put, values approaching unity indicate severe non-dynamical correlation and values approaching zero indicate predominantly dynamical correlation.

REFERENCES

- [1] Brian C Ferrari, Katerina Slavicinska, and Christopher J Bennett. “Role of Suprathermal Chemistry on the Evolution of Carbon Oxides and Organics within Interstellar and Cometary Ices”. In: *Accounts of Chemical Research* 54.5 (2021), pp. 1067–1079.
- [2] FC Gillett and WJ Forrest. “Spectra of the Becklin-Neugebauer point source and the Kleinmann-Low nebula from 2.8 to 13.5 microns.” In: *The Astrophysical Journal* 179 (1973), pp. 483–491.
- [3] Karin I Öberg et al. “The Spitzer ice legacy: Ice evolution from cores to protostars”. In: *The Astrophysical Journal* 740.2 (2011), p. 109.
- [4] HM Cuppen and RT Garrod. “Modelling of surface chemistry on an inhomogeneous interstellar grain”. In: *Astronomy & Astrophysics* 529 (2011), A151.
- [5] KM Pontoppidan. “Spatial mapping of ices in the Ophiuchus-F core-A direct measurement of CO depletion and the formation of CO₂”. In: *Astronomy & Astrophysics* 453.3 (2006), pp. L47–L50.
- [6] Ralf I Kaiser and K Roessler. “Theoretical and laboratory studies on the interaction of cosmic-ray particles with interstellar ices. III. Suprathermal chemistry-induced formation of hydrocarbon molecules in solid methane (CH₄), ethylene (C₂H₄), and acetylene (C₂H₂)”. In: *The Astrophysical Journal* 503.2 (1998), p. 959.
- [7] Norman G Moll, Dale R Clutter, and Warren E Thompson. “Carbon trioxide: Its production, infrared spectrum, and structure studied in a matrix of solid CO₂”. In: *The Journal of Chemical Physics* 45.12 (1966), pp. 4469–4481.
- [8] Karin I Öberg et al. “Formation rates of complex organics in UV irradiated CH₃OH-rich ices-I. Experiments”. In: *Astronomy & Astrophysics* 504.3 (2009), pp. 891–913.
- [9] Carl Sagan and BN Khare. “Tholins: Organic chemistry of interstellar grains and gas”. In: *Nature* 277.5692 (1979), pp. 102–107.
- [10] Yvonne J Pendleton and Louis J Allamandola. “The organic refractory material in the diffuse interstellar medium: Mid-infrared spectroscopic constraints”. In: *The Astrophysical Journal Supplement Series* 138.1 (2002), p. 75.
- [11] Nicolas Biver et al. “Ethyl alcohol and sugar in comet C/2014 Q2 (Lovejoy)”. In: *Science Advances* 1.9 (2015), e1500863.
- [12] GM Muñoz Caro et al. “Amino acids from ultraviolet irradiation of interstellar ice analogues”. In: *Nature* 416.6879 (2002), pp. 403–406.

- [13] Michel Nuevo et al. “A detailed study of the amino acids produced from the vacuum UV irradiation of interstellar ice analogs”. In: *Origins of Life and Evolution of Biospheres* 38.1 (2008), pp. 37–56.
- [14] Yong-Ki Kim. “Energy distribution of secondary electrons”. In: *Radiation research* 64.1 (1975), pp. 96–105.
- [15] Simon M Pimblott and Jay A LaVerne. “Production of low-energy electrons by ionizing radiation”. In: *Radiation Physics and Chemistry* 76.8-9 (2007), pp. 1244–1247.
- [16] DA Alman, David N Ruzic, and JN Brooks. “A hydrocarbon reaction model for low temperature hydrogen plasmas and an application to the Joint European Torus”. In: *Physics of plasmas* 7.5 (2000), pp. 1421–1432.
- [17] Xinli Song et al. “Ab initio study of the potential energy surface for the $\text{OH} + \text{CO} \rightarrow \text{H} + \text{CO}_2$ reaction”. In: *The Journal of chemical physics* 125.9 (2006), p. 094301.
- [18] AC Adwin Boogert, Perry A Gerakines, and Douglas CB Whittet. “Observations of the icy universe”. In: *Annual Review of Astronomy and Astrophysics* 53 (2015).
- [19] SA Sandford et al. “The interstellar CH stretching band near 3.4 microns—Constraints on the composition of organic material in the diffuse interstellar medium”. In: *The Astrophysical Journal* 371 (1991), pp. 607–620.
- [20] Michael J Mumma and Steven B Charnley. “The chemical composition of comets—emerging taxonomies and natal heritage”. In: *Annual Review of Astronomy and Astrophysics* 49 (2011), pp. 471–524.
- [21] K Altwegg et al. “Prebiotic chemicals—amino acid and phosphorus—in the coma of comet 67P/Churyumov-Gerasimenko”. In: *Science Advances* 2.5 (2016), e1600285.
- [22] Andre Bieler et al. “Abundant molecular oxygen in the coma of comet 67P/Churyumov-Gerasimenko”. In: *Nature* 526.7575 (2015), p. 678.
- [23] Chris J Bennett, Claire Pirim, and Thomas M Orlando. “Space-weathering of solar system bodies: A laboratory perspective”. In: *Chemical reviews* 113.12 (2013), pp. 9086–9150.
- [24] Karin I Öberg. “Photochemistry and astrochemistry: Photochemical pathways to interstellar complex organic molecules”. In: *Chemical reviews* 116.17 (2016), pp. 9631–9663.
- [25] Scott A Sandford et al. “Prebiotic astrochemistry and the formation of molecules of astrobiological interest in interstellar clouds and protostellar disks”. In: *Chemical reviews* 120.11 (2020), pp. 4616–4659.
- [26] Chris J Bennett et al. “High-sensitivity Raman spectrometer to study pristine and irradiated interstellar ice analogs”. In: *Analytical chemistry* 85.12 (2013), pp. 5659–5665.
- [27] Vicente Timon et al. “Infrared spectra of amorphous and crystalline urea ices”. In: *Physical Chemistry Chemical Physics* (2021).
- [28] Martin Rubin et al. “Molecular nitrogen in comet 67P/Churyumov-Gerasimenko indicates a low formation temperature”. In: *Science* 348.6231 (2015), pp. 232–235.

- [29] Adrienn Luspay-Kuti et al. “Origin of Molecular Oxygen in Comets: Current Knowledge and Perspectives”. In: *Space science reviews* 214.8 (2018), p. 115.
- [30] Ugo Hincelin et al. “Survival of interstellar molecules to prestellar dense core collapse and early phases of disk formation”. In: *The Astrophysical Journal* 775.1 (2013), p. 44.
- [31] Ryan C Fortenberry et al. “Rovibrational Spectral Analysis of CO₃ and C₂O₃: Potential Sources for O₂ Observed in Comet 67P/Churyumov–Gerasimenko”. In: *The Astrophysical Journal Letters* 886.1 (2019), p. L10.
- [32] Chris J Bennett et al. “Untangling the formation of the cyclic carbon trioxide isomer in low temperature carbon dioxide ices”. In: *Physical Chemistry Chemical Physics* 6.4 (2004), pp. 735–746.
- [33] Chris J Bennett, Corey S Jamieson, and Ralf I Kaiser. “Mechanistic studies on the decomposition of carbon suboxide in a cometary ice analog”. In: *Planetary and Space Science* 56.9 (2008), pp. 1181–1189.
- [34] Wesley T Huntress Jr, Mark Alien, and Mona Delrtsky. “Carbon suboxide in comet Halley?” In: *Nature* 352.6333 (1991), p. 316.
- [35] Hervé Cottin and Nicolas Fray. “Distributed sources in comets”. In: *Space Science Reviews* 138.1-4 (2008), pp. 179–197.
- [36] Jacques Crovisier, Thérèse Encrenaz, and Michel Combes. “Carbon suboxide in comet Halley”. In: *Nature* 353.6345 (1991), pp. 610–610.
- [37] Corey S Jamieson, Alexander M Mebel, and Ralf I Kaiser. “Understanding the kinetics and dynamics of radiation-induced reaction pathways in carbon monoxide ice at 10 K”. In: *The Astrophysical Journal Supplement Series* 163.1 (2006), p. 184.
- [38] Chris J Bennett, Corey S Jamieson, and Ralf I Kaiser. “Mechanistic studies on the formation of carbon dioxide in extraterrestrial carbon monoxide ice analog samples”. In: *Physical Chemistry Chemical Physics* 11.21 (2009), pp. 4210–4218.
- [39] AM Mebel et al. “Theoretical study of oxygen isotope exchange and quenching in the O (1D) + CO₂ reaction”. In: *The Journal of Physical Chemistry A* 108.39 (2004), pp. 7983–7994.
- [40] Corey S Jamieson, Alexander M Mebel, and Ralf I Kaiser. “Identification of the D_{3h} isomer of carbon trioxide (CO₃) and its implications for atmospheric chemistry”. In: *ChemPhysChem* 7.12 (2006), pp. 2508–2513.
- [41] Corey S Jamieson, Alexander M Mebel, and Ralf I Kaiser. “Novel detection of the C_{2v} isomer of carbon tetraoxide (CO₄)”. In: *Chemical physics letters* 440.1-3 (2007), pp. 105–109.
- [42] Corey S Jamieson, Alexander M Mebel, and Ralf I Kaiser. “First detection of the C₂ symmetric isomer of carbon pentaoxide (CO₅) at 10 K”. In: *Chemical physics letters* 443.1-3 (2007), pp. 49–54.
- [43] Corey S Jamieson, Alexander M Mebel, and Ralf I Kaiser. “First detection of the C_s symmetric isomer of carbon hexaoxide (CO₆) at 10 K”. In: *Chemical Physics Letters* 450.4-6 (2008), pp. 312–317.

- [44] PA Gerakines, WA Schutte, and P Ehrenfreund. “Ultraviolet processing of interstellar ice analogs. I. Pure ices.” In: *Astronomy and Astrophysics* 312 (1996), pp. 289–305.
- [45] Rafael Martín-Doménech et al. “UV photoprocessing of CO₂ ice: a complete quantification of photochemistry and photon-induced desorption processes”. In: *Astronomy & Astrophysics* 584 (2015), A14.
- [46] Courtney P Ennis, Chris J Bennett, and Ralf I Kaiser. “On the formation of ozone in oxygen-rich solar system ices via ionizing radiation”. In: *Physical Chemistry Chemical Physics* 13.20 (2011), pp. 9469–9482.
- [47] Chris J Bennett, Courtney P Ennis, and Ralf I Kaiser. “IMPLANTATION OF ENERGETIC D+ IONS INTO CARBON DIOXIDE ICES AND IMPLICATIONS FOR OUR SOLAR SYSTEM: FORMATION OF D₂O AND D₂CO₃”. In: *The Astrophysical Journal* 794.1 (2014), p. 57.
- [48] Chris J Bennett, Corey S Jamieson, and Ralf I Kaiser. “Mechanistical studies on the formation and destruction of carbon monoxide (CO), carbon dioxide (CO₂), and carbon trioxide (CO₃) in interstellar ice analog samples”. In: *Physical Chemistry Chemical Physics* 12.16 (2010), pp. 4032–4050.
- [49] Chris J Bennett, Corey S Jamieson, and Ralf I Kaiser. “An experimental investigation of the decomposition of carbon monoxide and formation routes to carbon dioxide in interstellar ices”. In: *The Astrophysical Journal Supplement Series* 182.1 (2009), p. 1.
- [50] ME Palumbo et al. “Ion irradiation of astrophysical ices”. In: *Journal of Physics: Conference Series*. Vol. 101. 1. IOP Publishing. 2008, p. 012002.
- [51] GA Baratta et al. “Ion irradiation of ices relevant to astrophysics”. In: *Memorie della Societa Astronomica Italiana Supplementi* 5 (2004), p. 33.
- [52] RG Urso et al. “Combined infrared and Raman study of solid CO”. In: *Astronomy & Astrophysics* 594 (2016), A80.
- [53] Eva Mateo-Marti et al. “Characterizing interstellar medium, planetary surface and deep environments by spectroscopic techniques using unique simulation chambers at Centro de Astrobiologia (CAB)”. In: *Life* 9.3 (2019), p. 72.
- [54] Timothy Kowalczyk and Anna I Krylov. “Electronic Structure of Carbon Trioxide and Vibronic Interactions Involving Jahn- Teller States”. In: *The Journal of Physical Chemistry A* 111.33 (2007), pp. 8271–8276.
- [55] Yang Liu et al. “Combined Jahn- Teller and Pseudo-Jahn- Teller Effect in the CO₃ Molecule: A Seven-State Six-Mode Problem”. In: *Journal of chemical theory and computation* 5.10 (2009), pp. 2679–2686.
- [56] Reggie L Hudson, Perry A Gerakines, and Robert F Ferrante. “IR spectra and properties of solid acetone, an interstellar and cometary molecule”. In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 193 (2018), pp. 33–39.
- [57] Brian C Ferrari. “AutoGAMESS: A Python package for automation of GAMESS (US) Raman calculations”. In: *Journal of Open Source Software* 4.41 (2019), p. 1612.

- [58] Michael W Schmidt et al. “General atomic and molecular electronic structure system”. In: *Journal of Computational Chemistry* 14.11 (1993), pp. 1347–1363. DOI: 10.1002/jcc.540141112.
- [59] Mark S. Gordon and Michael W. Schmidt. “Chapter 41 - Advances in electronic structure theory: GAMESS a decade later”. In: *Theory and Applications of Computational Chemistry*. Amsterdam: Elsevier, 2005, pp. 1167–1189. DOI: 10.1016/B978-044451719-7/50084-6.
- [60] Jaroslav V Burda, Matěj Pavelka, and Milan Šimánek. “Theoretical model of copper Cu(I)/Cu(II) hydration. DFT and ab initio quantum chemical study”. In: *Journal of Molecular Structure: THEOCHEM* 683.1-3 (2004), pp. 183–193. DOI: 10.1016/j.theochem.2004.06.013.
- [61] Leonardo Pacifici et al. “A high-level ab initio study of the $N_2 + N_2$ reaction channel”. In: *Journal of Computational Chemistry* 34.31 (2013), pp. 2668–2676. DOI: 10.1002/jcc.23415.
- [62] AP Hickman et al. “Dissociative recombination of $e + HCNH^+$: Diabatic potential curves and dynamics calculations”. In: *Astronomy & Astrophysics* 438.1 (2005), pp. 31–37. DOI: 10.1051/0004-6361:20052658.
- [63] Chris J Bennett, Courtney P Ennis, and Ralf I Kaiser. “Experimental studies on the formation of D_2O and D_2O_2 by implantation of energetic D^+ ions into oxygen ices”. In: *The Astrophysical Journal* 782.2 (2014), p. 63. DOI: 10.1088/0004-637x/782/2/63.
- [64] Pierre L. Bhoorasingh et al. “Automated Transition State Theory Calculations for High-Throughput Kinetics”. In: *The Journal of Physical Chemistry A* 121.37 (2017), pp. 6896–6904. DOI: 10.1021/acs.jpca.7b07361.
- [65] Kiran Mathew et al. “Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows”. In: *Computational Materials Science* 139 (2017), pp. 140–152. DOI: 10.1016/j.commatsci.2017.07.030.
- [66] Ask Hjorth Larsen et al. “The atomic simulation environment—a Python library for working with atoms”. In: *Journal of Physics: Condensed Matter* 29.27 (2017), p. 273002. DOI: 10.1088/1361-648x/aa680e.
- [67] Jaron T. Krogel. “Nexus: A modular workflow management system for quantum simulation codes”. In: *Computer Physics Communications* 198 (2016), pp. 154–168. DOI: 10.1016/j.cpc.2015.08.012.
- [68] Noel M O’boyle, Adam L Tenderholt, and Karol M Langner. “Cclib: a library for package-independent computational chemistry algorithms”. In: *Journal of Computational Chemistry* 29.5 (2008), pp. 839–845. DOI: 10.1002/jcc.20823.
- [69] Gijs Schaftenaar and Jan H Noordik. “Molden: a pre-and post-processing program for molecular and electronic structures”. In: *Journal of Computer-Aided Molecular Design* 14.2 (2000), pp. 123–134. DOI: 10.1023/A:1008193805436.
- [70] JR Schmidt and WF Polik. “WebMO Enterprise, version 13.0”. In: *WebMO LLC* (2013). URL: <http://www.webmo.net>.

- [71] Brett M Bode and Mark S Gordon. “Macmolplt: a graphical user interface for GAMESS”. In: *Journal of Molecular Graphics and Modelling* 16.3 (1998), pp. 133–138. DOI: 10.1016/S1093-3263(99)00002-9.
- [72] Abdul-Rahman Allouche. “Gabedit—A graphical user interface for computational chemistry softwares”. In: *Journal of Computational Chemistry* 32.1 (2011), pp. 174–182. DOI: 10.1002/jcc.21600.
- [73] Céline DerMardirossian and Gary M Bokoch. “GDIs: central regulatory molecules in Rho GTPase activation”. In: *Trends in Cell Biology* 15.7 (2005), pp. 356–363. DOI: 10.1016/j.tcb.2005.05.001.
- [74] David Feller. “The role of databases in support of computational chemistry calculations”. In: *Journal of Computational Chemistry* 17.13 (1996), pp. 1571–1586. DOI: 10.1002/(SICI)1096-987X(199610)17:13<1571::AID-JCC9>3.0.CO;2-P.
- [75] Karen L Schuchardt et al. “Basis set exchange: a community database for computational sciences”. In: *Journal of Chemical Information and Modeling* 47.3 (2007), pp. 1045–1052. DOI: 10.1021/ci600510j.
- [76] Karen L Schuchardt et al. “Basis set exchange: a community database for computational sciences”. In: *Journal of chemical information and modeling* 47.3 (2007), pp. 1045–1052.
- [77] Anthony P. Scott and Leo Radom. “Harmonic Vibrational Frequencies: An Evaluation of Hartree–Fock, Møller–Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors”. In: *The Journal of Physical Chemistry* 100.41 (1996), pp. 16502–16513. DOI: 10.1021/jp960976r.
- [78] Travis E Oliphant. *A guide to NumPy*. Vol. 1. Trelgol Publishing USA, 2006.
- [79] Stéfan van der Walt, S. Chris Colbert, and Gaël Varoquaux. “The NumPy Array: A Structure for Efficient Numerical Computation”. In: *Computing in Science & Engineering* 13.2 (2011), pp. 22–30. DOI: 10.1109/MCSE.2011.37.
- [80] Eric Jones, Travis Oliphant, Pearu Peterson, et al. *SciPy: Open source scientific tools for Python*. [Online; accessed `today`]. 2001. URL: <http://www.scipy.org/>.
- [81] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 51–56.
- [82] J. D. Hunter. “Matplotlib: A 2D Graphics Environment”. In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.
- [83] Brian C Ferrari and Chris J Bennett. “A Comparison of Medium-Sized Basis Sets for the Prediction of Geometries, Vibrational Frequencies, Infrared Intensities and Raman Activities for Water”. In: *Journal of Physics: Conference Series*. Vol. 1290. 1. IOP Publishing. 2019, p. 012013.
- [84] Rainer Glaser et al. “Adenine synthesis in interstellar space: mechanisms of prebiotic pyrimidine-ring formation of monocyclic HCN-pentamers”. In: *Astrobiology* 7.3 (2007), pp. 455–470.

- [85] Elena E Zvereva, Artur R Shagidullin, and Sergey A Katsyuba. “Ab initio and DFT predictions of infrared intensities and Raman activities”. In: *The Journal of Physical Chemistry A* 115.1 (2010), pp. 63–69.
- [86] Carlos A Jiménez-Hoyos, Benjamin G Janesko, and Gustavo E Scuseria. “Evaluation of range-separated hybrid density functionals for the prediction of vibrational frequencies, infrared intensities, and Raman activities”. In: *Physical Chemistry Chemical Physics* 10.44 (2008), pp. 6621–6629.
- [87] Klaus-Peter Huber. *Molecular spectra and molecular structure: IV. Constants of diatomic molecules*. Springer Science & Business Media, 2013.
- [88] Trygve Helgaker, Poul Jorgensen, and Jeppe Olsen. *Molecular electronic-structure theory*. John Wiley & Sons, 2014.
- [89] Mathew D Halls and H Bernhard Schlegel. “Comparison of the performance of local, gradient-corrected, and hybrid density functional models in predicting infrared intensities”. In: *The Journal of chemical physics* 109.24 (1998), pp. 10587–10593.
- [90] Mathew D Halls and H Bernhard Schlegel. “Comparison study of the prediction of Raman intensities using electronic structure methods”. In: *The Journal of chemical physics* 111.19 (1999), pp. 8819–8824.
- [91] Anthony P Scott and Leo Radom. “Harmonic vibrational frequencies: an evaluation of Hartree-Fock, Møller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors”. In: *The Journal of Physical Chemistry* 100.41 (1996), pp. 16502–16513.
- [92] Aneta Buczek, Teobald Kupka, and Małgorzata A Broda. “Extrapolation of water and formaldehyde harmonic and anharmonic frequencies to the B3LYP/CBS limit using polarization consistent basis sets”. In: *Journal of molecular modeling* 17.8 (2011), pp. 2029–2040.
- [93] Boris Galabov et al. “High level ab initio quantum mechanical predictions of infrared intensities”. In: *The Journal of Physical Chemistry A* 106.5 (2002), pp. 819–832.
- [94] J Grant Hill. “Gaussian basis sets for molecular applications”. In: *International Journal of Quantum Chemistry* 113.1 (2013), pp. 21–34.
- [95] Frank Jensen. “Atomic orbital basis sets”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3.3 (2013), pp. 273–295.
- [96] Teobald Kupka and Carmay Lim. “Polarization-consistent versus correlation-consistent basis sets in predicting molecular and spectroscopic properties”. In: *The Journal of Physical Chemistry A* 111.10 (2007), pp. 1927–1932.
- [97] Trygve Helgaker et al. “Basis-set convergence of correlated calculations on water”. In: *The Journal of chemical physics* 106.23 (1997), pp. 9639–9646.
- [98] Donald G Truhlar. “Basis-set extrapolation”. In: *Chemical Physics Letters* 294.1-3 (1998), pp. 45–48.
- [99] GA Petersson and Mohammad A Al-Laham. “A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms”. In: *The Journal of chemical physics* 94.9 (1991), pp. 6081–6090.

- [100] Axel D Becke. “Density-functional thermochemistry. III. The role of exact exchange”. In: *The Journal of chemical physics* 98.7 (1993), pp. 5648–5652.
- [101] Chengteh Lee, Weitao Yang, and Robert G Parr. “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density”. In: *Physical review B* 37.2 (1988), p. 785.
- [102] Chr Møller and Milton S Plesset. “Note on an approximation treatment for many-electron systems”. In: *Physical Review* 46.7 (1934), p. 618.
- [103] Piotr Piecuch et al. “Efficient computer implementation of the renormalized coupled-cluster methods: the r-ccsd [t], r-ccsd (t), cr-ccsd [t], and cr-ccsd (t) approaches”. In: *Computer Physics Communications* 149.2 (2002), pp. 71–96.
- [104] Ryan M Olson et al. “A novel approach to parallel coupled cluster calculations: Combining distributed and shared memory techniques for modern cluster based systems”. In: *Journal of Chemical Theory and Computation* 3.4 (2007), pp. 1312–1328.
- [105] Praveen C Hariharan and John A Pople. “The influence of polarization functions on molecular orbital hydrogenation energies”. In: *Theoretica chimica acta* 28.3 (1973), pp. 213–222.
- [106] David E Woon and Thom H Dunning Jr. “Calculation of the electron affinities of the second row atoms: Al–Cl”. In: *The Journal of chemical physics* 99.5 (1993), pp. 3730–3737.
- [107] Ewa Papajak and Donald G Truhlar. “Convergent partially augmented basis sets for post-Hartree-Fock calculations of molecular properties and reaction barrier heights”. In: *Journal of chemical theory and computation* 7.1 (2010), pp. 10–18.
- [108] Thom H Dunning Jr. “Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen”. In: *The Journal of chemical physics* 90.2 (1989), pp. 1007–1023.
- [109] David E Woon and Thom H Dunning Jr. “Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties”. In: *The Journal of chemical physics* 100.4 (1994), pp. 2975–2988.
- [110] Rick A Kendall, Thom H Dunning Jr, and Robert J Harrison. “Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions”. In: *The Journal of chemical physics* 96.9 (1992), pp. 6796–6806.
- [111] Ewa Papajak and Donald G Truhlar. “Efficient diffuse basis sets for density functional theory”. In: *Journal of chemical theory and computation* 6.3 (2010), pp. 597–601.
- [112] Ewa Papajak et al. “Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions”. In: *Journal of chemical theory and computation* 7.10 (2011), pp. 3027–3034.
- [113] Frank Jensen. “Polarization consistent basis sets: Principles”. In: *The Journal of Chemical Physics* 115.20 (2001), pp. 9113–9125.
- [114] Frank Jensen. “Polarization consistent basis sets. II. Estimating the Kohn–Sham basis set limit”. In: *The Journal of chemical physics* 116.17 (2002), pp. 7372–7379.

- [115] Frank Jensen. “Unifying general and segmented contracted basis sets. Segmented polarization consistent basis sets”. In: *Journal of chemical theory and computation* 10.3 (2014), pp. 1074–1085.
- [116] Florian Weigend and Reinhart Ahlrichs. “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy”. In: *Physical Chemistry Chemical Physics* 7.18 (2005), pp. 3297–3305.
- [117] Frank Neese and Edward F Valeev. “Revisiting the atomic natural orbital approach for basis sets: Robust systematic basis sets for explicitly correlated and conventional correlated ab initio methods?” In: *Journal of chemical theory and computation* 7.1 (2010), pp. 33–43.
- [118] Angelika Baranowska and Andrzej J Sadlej. “Polarized basis sets for accurate calculations of static and dynamic electric properties of molecules”. In: *Journal of computational chemistry* 31.3 (2010), pp. 552–560.
- [119] Zuzana Benkova et al. “Reduced-size polarized basis sets for calculations of molecular electric properties. I. The basis set generation”. In: *Journal of computational chemistry* 26.2 (2005), pp. 145–153.
- [120] Vincenzo Barone, Paola Cimino, and Emiliano Stendardo. “Development and validation of the B3LYP/N07D computational model for structural parameter and magnetic tensors of large free radicals”. In: *Journal of chemical theory and computation* 4.5 (2008), pp. 751–764.
- [121] Vincenzo Barone, Malgorzata Biczysko, and Julien Bloino. “Fully anharmonic IR and Raman spectra of medium-size molecular systems: accuracy and interpretation”. In: *Physical Chemistry Chemical Physics* 16.5 (2014), pp. 1759–1787.
- [122] Diego Paschoal, Marcello F Costa, and Hélio F Dos Santos. “NLO-X (X= I–III): New Gaussian basis sets for prediction of linear and nonlinear electric properties”. In: *International Journal of Quantum Chemistry* 114.12 (2014), pp. 796–804.
- [123] Takehiko Shimanouchi. “Tables of molecular vibrational frequencies. Consolidated volume II”. In: *Journal of physical and chemical reference data* 6.3 (1977), pp. 993–1102.
- [124] AR Hoy and Po R Bunker. “A precise solution of the rotation bending Schrödinger equation for a triatomic molecule with application to the water molecule”. In: *Journal of Molecular Spectroscopy* 74.1 (1979), pp. 1–8.
- [125] F Culot and J Liévin. “Ab initio calculation of vibrational dipole moment matrix elements. II. the water molecule as a polyatomic test case”. In: *Physica Scripta* 46.6 (1992), p. 502.
- [126] David P Tew et al. “Basis set limit CCSD (T) harmonic vibrational frequencies”. In: *The Journal of Physical Chemistry A* 111.44 (2007), pp. 11242–11248.
- [127] M. S. Gordon and M. W. Schmidt. “Advances in electronic structure theory: GAMESS a decade later”. In: *Theory and Applications of Computational Chemistry: the first forty years*. Ed. by C. E. Dykstra et al. Amsterdam: Elsevier, 2005, pp. 1167–1189.

- [128] Aaron Sayvetz. “The kinetic energy of polyatomic molecules”. In: *The Journal of Chemical Physics* 7.6 (1939), pp. 383–389.
- [129] K Ishimura et al. “GAMESS-MW Schmidt, KK Baldrige, JA Boatz, ST Elbert, MS Gordon, JJ Jensen, S. Koseki, N. Matsunaga, KA Nguyen, S. Su, TL Windus, M. Dupuis, JA Montgomery J. Comput. Chem. 14, 1347-1363 (1993)”. In: *J. Comput. Chem* 14 (1993), pp. 1347–1363.
- [130] Sergey V Krasnoshchekov et al. “Comparing the accuracy of perturbative and variational calculations for predicting fundamental vibrational frequencies of dihalomethanes”. In: *The Journal of chemical physics* 148.8 (2018), p. 084102.
- [131] Asger Halkier et al. “Basis-set convergence in correlated calculations on Ne, N₂, and H₂O”. In: *Chemical Physics Letters* 286.3-4 (1998), pp. 243–252.
- [132] Benny G Johnson, Peter MW Gill, and John A Pople. “The performance of a family of density functional methods”. In: *The Journal of chemical physics* 98.7 (1993), pp. 5612–5626.
- [133] Jeffrey P Merrick, Damian Moran, and Leo Radom. “An evaluation of harmonic vibrational frequency scale factors”. In: *The Journal of Physical Chemistry A* 111.45 (2007), pp. 11683–11700.
- [134] Mattias P Andersson and Per Uvdal. “New scale factors for harmonic vibrational frequencies using the B3LYP density functional method with the triple- ζ basis set 6-311+ G (d, p)”. In: *The Journal of Physical Chemistry A* 109.12 (2005), pp. 2937–2941.
- [135] Mathew D Halls, Julia Velkovski, and H Bernhard Schlegel. “Harmonic frequency scaling factors for Hartree-Fock, S-VWN, B-LYP, B3-LYP, B3-PW91 and MP2 with the Sadlej pVTZ electric property basis set”. In: *Theoretical Chemistry Accounts* 105.6 (2001), pp. 413–421.
- [136] Stefan Grimme. “Improved second-order Møller–Plesset perturbation theory by separate scaling of parallel-and antiparallel-spin pair correlation energies”. In: *The Journal of chemical physics* 118.20 (2003), pp. 9095–9102.
- [137] Piotr Piecuch et al. “Single-reference, size-extensive, non-iterative coupled-cluster approaches to bond breaking and biradicals”. In: *Chemical physics letters* 418.4-6 (2006), pp. 467–474.
- [138] Brian C Ferrari and Chris J Bennett. “A computational investigation of the equilibrium geometries, energetics, vibrational frequencies, infrared intensities and Raman activities of C₂O_y (y= 3, 4) species”. In: *Molecular Physics* (2020), e1837404.
- [139] Thomas M Klapötke. *High energy density materials*. Vol. 125. Springer, 2007.
- [140] Rolf Gleiter, Isabella Hyla-Kryspin, and Karl-Heinz Pfeifer. “On the stability of the tetramers of carbon monoxide, hydrogen isocyanide, and vinylidene. A molecular orbital theoretical rationalization”. In: *The Journal of Organic Chemistry* 60.18 (1995), pp. 5878–5883.

- [141] Errol Lewars. “Polymers and oligomers of carbon dioxide: ab initio and semiempirical calculations”. In: *Journal of Molecular Structure: THEOCHEM* 363.1 (1996), pp. 1–15.
- [142] Stefano Evangelisti. “Ab initio study of C4O4 in Td symmetry”. In: *Chemical physics letters* 259.3-4 (1996), pp. 261–264.
- [143] Errol G Lewars. “Oligomers and Polymers of Carbon Dioxide and CO2/N2”. In: *Modeling Marvels*. Springer, 2008, pp. 165–184.
- [144] A Gambi, AG Giumanini, and P Strazzolini. “Theoretical investigations on (CO) n (CO2) m cyclic cooligomers”. In: *Journal of Molecular Structure: THEOCHEM* 536.1 (2001), pp. 9–16.
- [145] Brett I Dunlap et al. “Thermodynamic and kinetic stabilities of CO2 oligomers”. In: *The Journal of chemical physics* 138.13 (2013), p. 134304.
- [146] KP Huber and G Herzberg. “Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecules”. In: (1979).
- [147] Martin Rubin et al. “Molecular oxygen in Oort cloud comet 1P/Halley”. In: *The Astrophysical Journal Letters* 815.1 (2015), p. L11.
- [148] O Mousis et al. “Origin of molecular oxygen in comet 67P/Churyumov–Gerasimenko”. In: *The Astrophysical Journal Letters* 823.2 (2016), p. L41.
- [149] F Dulieu, M Minissale, and D Bockelée-Morvan. “Production of O2 through dismutation of H2O2 during water ice desorption: a key to understanding comet O2 abundances”. In: *Astronomy & Astrophysics* 597 (2017), A56.
- [150] Yunxi Yao and Konstantinos P Giapis. “Dynamic molecular oxygen production in cometary comae”. In: *Nature communications* 8 (2017), p. 15298.
- [151] H Staudinger. “Oxalylchlorid”. In: *Berichte der deutschen chemischen Gesellschaft* 41.3 (1908), pp. 3558–3566.
- [152] RL Mills et al. “New phases and chemical reactions in solid CO under pressure”. In: *Le Journal de Physique Colloques* 45.C8 (1984), pp. C8–187.
- [153] M Lipp et al. “Carbon Monoxide: Spectroscopic Characterization of the High-Pressure Polymerized Phase”. In: *Journal of low temperature physics* 111.3-4 (1998), pp. 247–256.
- [154] Ahmed A El-Sherif and Bakir JA Jeragh. “Mixed ligand complexes of Cu (II)-2-(2-pyridyl)-benzimidazole and aliphatic or aromatic dicarboxylic acids: synthesis, characterization and biological activity”. In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 68.3 (2007), pp. 877–882.
- [155] Michael M Rauhut et al. “Chemiluminescence from reactions of electronegatively substituted aryl oxalates with hydrogen peroxide and fluorescent compounds”. In: *Journal of the American Chemical Society* 89.25 (1967), pp. 6515–6522.
- [156] Suresh Dua, Salvatore Peppe, and John H Bowie. “The unusual neutral OCOCO and possible charged analogues. A theoretical investigation”. In: *Journal of the Chemical Society, Perkin Transactions 2* 12 (2001), pp. 2244–2247.

- [157] Salvatore Peppe, Suresh Dua, and John H Bowie. “Is the Elusive Trioxydehydroethene Neutral (O₂C- CO) Detectable in the Gas Phase?” In: *The Journal of Physical Chemistry A* 105.44 (2001), pp. 10139–10145.
- [158] Paolo Strazzolini et al. “The reaction between ethanedioyl (oxalyl) dihalides and Ag₂C₂O₄: A route to Staudinger’s elusive ethanedioic (oxalic) acid anhydride”. In: *Journal of the Chemical Society, Perkin Transactions 1* 998.16 (1998), pp. 2553–2558.
- [159] Mingfei Zhou et al. “Infrared spectra and theoretical studies of the C₂O₃- anion isolated in solid argon”. In: *The Journal of Chemical Physics* 112.16 (2000), pp. 7089–7094.
- [160] Robert Bukowski et al. “Intermolecular potential of carbon dioxide dimer from symmetry-adapted perturbation theory”. In: *The Journal of chemical physics* 110.8 (1999), pp. 3785–3803.
- [161] Herman F Cordes, Herbert P Richter, and Carl A Heller. “Mass spectrometric evidence for the existence of 1, 2-dioxetanedione (carbon dioxide dimer). Chemiluminescent intermediate”. In: *Journal of the American Chemical Society* 91.25 (1969), pp. 7209–7209.
- [162] Yong Zhang, Xi-Rui Zeng, and Xiao-Zeng You. “A new complete basis set model (CBS-QB3) study on the possible intermediates in chemiluminescence”. In: *The Journal of Chemical Physics* 113.18 (2000), pp. 7731–7734.
- [163] Gilles Frapper and Jean-Yves Saillard. “Search for new allotropic forms of carbon dioxide and carbon disulfide: a density functional study of CX₂-based oligomers (X= O, S)”. In: *Journal of the American Chemical Society* 122.22 (2000), pp. 5367–5370.
- [164] Richard Bos et al. “Studies on the mechanism of the peroxyoxalate chemiluminescence reaction: Part 1. Confirmation of 1, 2-dioxetanedione as an intermediate using ¹³C nuclear magnetic resonance spectroscopy”. In: *Analytica chimica acta* 502.2 (2004), pp. 141–147.
- [165] Sarah A Tonkin et al. “Studies on the mechanism of the peroxyoxalate chemiluminescence reaction: Part 2. Further identification of intermediates using 2D EXSY ¹³C nuclear magnetic resonance spectroscopy”. In: *Analytica chimica acta* 614.2 (2008), pp. 173–181.
- [166] Michael Howart, Brendan Yonke, and Jie Song. “THEORETICAL INVESTIGATION OF A NEW DIOXIRANE C₂O₄”. In: *Journal of Theoretical and Computational Chemistry* 9.02 (2010), pp. 531–542.
- [167] Tian Lu and Feiwu Chen. “Multiwfn: a multifunctional wavefunction analyzer”. In: *Journal of computational chemistry* 33.5 (2012), pp. 580–592.
- [168] Brian Ferrari. “AutoGAMESS: A Python package for automation of GAMESS(US) Raman calculations”. In: *Journal of Open Source Software* 4.41 (2019), p. 1612. DOI: 10.21105/joss.01612.
- [169] Krishnan Raghavachari et al. “A fifth-order perturbation comparison of electron correlation theories”. In: *Chemical Physics Letters* 157.6 (1989), pp. 479–483.

- [170] Manoj K Kesharwani, Brina Brauer, and Jan ML Martin. “Frequency and zero-point vibrational energy scale factors for double-hybrid density functionals (and other selected methods): can anharmonic force fields be avoided?” In: *The Journal of Physical Chemistry A* 119.9 (2015), pp. 1701–1714.
- [171] Sergio Filipe Sousa, Pedro Alexandrino Fernandes, and Maria Joao Ramos. “General performance of density functionals”. In: *The Journal of Physical Chemistry A* 111.42 (2007), pp. 10439–10452.
- [172] John P Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized gradient approximation made simple”. In: *Physical review letters* 77.18 (1996), p. 3865.
- [173] Axel D Becke. “Density-functional exchange-energy approximation with correct asymptotic behavior”. In: *Physical review A* 38.6 (1988), p. 3098.
- [174] John P Perdew, Matthias Ernzerhof, and Kieron Burke. “Rationale for mixing exact exchange with density functional approximations”. In: *The Journal of chemical physics* 105.22 (1996), pp. 9982–9985.
- [175] Carlo Adamo and Vincenzo Barone. “Toward reliable density functional methods without adjustable parameters: The PBE0 model”. In: *The Journal of chemical physics* 110.13 (1999), pp. 6158–6170.
- [176] Jeng-Da Chai and Martin Head-Gordon. “Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections”. In: *Physical Chemistry Chemical Physics* 10.44 (2008), pp. 6615–6620.
- [177] Yan Zhao and Donald G Truhlar. “A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions”. In: *The Journal of chemical physics* 125.19 (2006), p. 194101.
- [178] Roberto Peverati and Donald G Truhlar. “M11-L: A local density functional that provides improved accuracy for electronic structure calculations in chemistry and physics”. In: *The Journal of Physical Chemistry Letters* 3.1 (2011), pp. 117–124.
- [179] Yan Zhao and Donald G Truhlar. “The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals”. In: *Theoretical Chemistry Accounts* 120.1-3 (2008), pp. 215–241.
- [180] Yan Zhao and Donald G Truhlar. “Density functional for spectroscopy: no long-range self-interaction error, good performance for Rydberg and charge-transfer states, and better performance on average than B3LYP for ground states”. In: *The Journal of Physical Chemistry A* 110.49 (2006), pp. 13126–13130.
- [181] Roberto Peverati and Donald G Truhlar. “Improving the accuracy of hybrid meta-GGA density functionals by range separation”. In: *The Journal of Physical Chemistry Letters* 2.21 (2011), pp. 2810–2817.
- [182] Dieter Cremer. “Density functional theory: coverage of dynamic and non-dynamic electron correlation effects”. In: *Molecular Physics* 99.23 (2001), pp. 1899–1940.

- [183] Dieter Cremer et al. “Implicit and explicit coverage of multi-reference effects by density functional theory”. In: *International Journal of Molecular Sciences* 3.6 (2002), pp. 604–638.
- [184] Robert J Meier. “Calculating the vibrational spectra of molecules: An introduction for experimentalists with contemporary examples”. In: *Vibrational spectroscopy* 43.1 (2007), pp. 26–37.
- [185] RBSJ Krishnan et al. “Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions”. In: *The Journal of chemical physics* 72.1 (1980), pp. 650–654.
- [186] AJ Sadlej et al. “Collect. Czech. Chem. Commun.” In: (1988).
- [187] Wagner E Richter et al. “Core–valence correlation effects on IR calculations: the BF₃ and BCl₃ cases”. In: *Journal of molecular modeling* 20.7 (2014), p. 2333.
- [188] Derek Albert Long and DA Long. *Raman spectroscopy*. Vol. 276. McGraw-Hill New York, 1977.
- [189] Timothy J Lee and Peter R Taylor. “A diagnostic for determining the quality of single-reference electron correlation methods”. In: *International Journal of Quantum Chemistry* 36.S23 (1989), pp. 199–207.
- [190] Uma R Fogueri et al. “A simple DFT-based diagnostic for nondynamical correlation”. In: *Theoretical Chemistry Accounts* 132.1 (2013), p. 1291.
- [191] Larry A Curtiss, Krishnan Raghavachari, and John A Pople. “Gaussian-2 theory using reduced Mo/ller–Plesset orders”. In: *The Journal of chemical physics* 98.2 (1993), pp. 1293–1298.
- [192] Larry A Curtiss et al. “Gaussian-3 theory using reduced Mo/ller–Plesset order”. In: *The Journal of chemical physics* 110.10 (1999), pp. 4703–4709.
- [193] Piotr Piecuch and Marta Włoch. “Renormalized coupled-cluster methods exploiting left eigenstates of the similarity-transformed Hamiltonian”. In: *The Journal of chemical physics* 123.22 (2005), p. 224105.
- [194] Marta Włoch, Jeffrey R Gour, and Piotr Piecuch. “Extension of the renormalized coupled-cluster methods exploiting left eigenstates of the similarity-transformed Hamiltonian to open-shell systems: A benchmark study”. In: *The Journal of Physical Chemistry A* 111.44 (2007), pp. 11359–11382.
- [195] Peter MW Gill et al. “The performance of the Becke–Lee–Yang–Parr (B–LYP) density functional theory with various basis sets”. In: *Chemical Physics Letters* 197.4–5 (1992), pp. 499–505.
- [196] MW Chase et al. “J. Phys. Chem. Ref. Data”. In: *JANAF Thermochemical Tables* 4 (1998).
- [197] David Feller, David A Dixon, and Joseph S Francisco. “Coupled cluster theory determination of the heats of formation of combustion-related compounds: CO, HCO, CO₂, HCO₂, HOCO, HC (O) OH, and HC (O) OOH”. In: *The Journal of Physical Chemistry A* 107.10 (2003), pp. 1604–1617.

- [198] V Zengin et al. “Study of the low-lying electronic states of CCO by photoelectron spectroscopy of CCO- and ab initio calculations”. In: *The Journal of chemical physics* 105.22 (1996), pp. 9740–9747.
- [199] Stephen P Walch. “Calculated spectroscopic constants for the X $3\Sigma^-$, 1Δ , $1\Sigma^+$, 3Π , and 1Π states of CCO. The heat of formation of CCO”. In: *The Journal of Chemical Physics* 72.10 (1980), pp. 5679–5686.
- [200] C Franklin Goldsmith, Gregory R Magoon, and William H Green. “Database of small molecule thermochemistry for combustion”. In: *The Journal of Physical Chemistry A* 116.36 (2012), pp. 9033–9057.
- [201] AG Streng. “Tables of ozone properties”. In: *Journal of Chemical and Engineering Data* 6.3 (1961), pp. 431–436.
- [202] JD Cox, DD Wagman, and VA Medvedev. “CODATA Key Values for Thermodynamics”. In: (1984).
- [203] Mortimer J Kamlet and SJ Jacobs. “Chemistry of detonations. I. A simple method for calculating detonation properties of C–H–N–O explosives”. In: *The Journal of Chemical Physics* 48.1 (1968), pp. 23–35.
- [204] Jianying Zhang, Gangling Chen, and Xuedong Gong. “QSPR modeling of detonation parameters and sensitivity of some energetic materials: DFT vs. PM3 calculations”. In: *Journal of molecular modeling* 23.6 (2017), p. 193.
- [205] Owen T O’Sullivan and Michael J Zdilla. “Properties and Promise of Catenated Nitrogen Systems As High-Energy-Density Materials”. In: *Chemical Reviews* 120.12 (2020), pp. 5682–5744.
- [206] Martin Brehm and Barbara Kirchner. *TRAVIS-a free analyzer and visualizer for Monte Carlo and molecular dynamics trajectories*. 2011.
- [207] Frank Jensen. *Introduction to computational chemistry*. John wiley & sons, 2017.
- [208] Timothy J Lee and Gustavo E Scuseria. “Achieving chemical accuracy with coupled-cluster theory”. In: *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*. Springer, 1995, pp. 47–108.