Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Apurba Nandi	Date

Development of Machine-Learned Potential Energy Surfaces with Application to the Reactive and Vibrational Dynamics of High Dimensional Molecular Systems

by
Apurba Nandi
Doctor of Philosophy
Chemistry

Prof. Joel M. Bowman, Ph.D
Advisor

Prof. Francesco Evangelista, Ph.D
Committee Member

Prof. James T. Kindt, Ph.D
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Development of Machine-Learned Potential Energy Surfaces with Application to the Reactive and Vibrational Dynamics of High Dimensional Molecular Systems

by

Apurba Nandi

B.Sc., Jadavpur University, Kolkata, 2013

M.Sc., Indian Institute of Technology, Kanpur (IITK), 2015

Advisor: Prof. Joel M. Bowman, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2022

Abstract

Development of Machine-Learned Potential Energy Surfaces with Application to the Reactive and Vibrational Dynamics of High Dimensional Molecular Systems

By Apurba Nandi

The potential energy surfaces (PESs) play a significant role in quantum chemistry calculations; such as kinetics and dynamics of chemical reactions, reactive and non-reactive scattering, and molecular spectroscopy. Developing a robust mathematical model for the generation of high dimensional PESs is a challenging work in theoretical and computational research, mainly for large molecules and molecular clusters. There are several bottlenecks as the molecular size increases, such as dimensionality of PES, the number of appropriate basis functions to fit the dataset, and appropriate fitting method.

In this work, a software has been developed to incorporate both electronic energies and gradients generating high-dimensional PES using limited numbers of training configurations employing a permutationally invariant polynomial (PIP) basis. Our main goal of this work was to demonstrate a procedure that can produce a very good PES of a molecule without sacrificing quantitative accuracy using a small training dataset. We showed that a precisely fitted potential surface of CH_4 can be obtained using energies and gradients with only 100 or even just 50 widely scattered configurations and that was successfully applied to quantum calculations.

Next, a fragment-based PIP approach has been developed to extend our fitting method for more than 10-12 atom systems. This was the first time we were able to develop a PES of 12 atom peptide molecule, *N*-methyl acetamide, which described both cis and trans isomers and their two isomerization TS accurately, which was a great achievement. Rigorous diffusion Monte-Carlo (DMC) calculations were successfully performed to compute quantum ZPE of *N*-methyl acetamide.

Recently, a fascinating method has been developed based on Δ-Machine Learning approach to achieve CCSD(T) level accuracy (energies and gradients) from low level DFT or MP2 level of theory using a very limited number of data sets and successfully applied to 9-atom Ethanol, 12-atom *N*-methyl acetamide (NMA), and 15-atom Acetylacetone molecules. This Δ-Machine Learning approach provides a solid ground for future innovation in fields like materials science and computational biology by means of highly accurate simulations currently out of reach.

These analytical potential surfaces were also employed in chemical reaction dynamics study and anharmonic vibrational calculations. Quasi-classical trajectory (QCT) calculations were performed to investigate the "Zero-point energy leak" and "Isomerization" of *syn*-Criegee. The adiabatic switching method was applied to prepare the initial conditions for these quasi-classical trajectories. Quantum ZPEs of methanol and its all isotopologs and isotopomers and the D/H exchange probability in their zero-point state have been investigated which was great interest in astrochemical research. A fascinating work was done by establishing a vibration-facilitated roaming mechanism in the isomerization of CO molecule on NaCl(100) surface by considering a $(CO\text{-}NaCl)_n$ finite cluster models. The novelty of this work is the isomerization was seen for highly excited CO vibrational states, in excellent agreement with the experiment. Recently, quantum nuclear simulations have been performed on a newly developed CCSD(T) machine-learned PES and revealed the equivalence of gas-phase *trans* and *gauche* conformers of ethanol when accurate vibrational zero-point energies and dynamical effects were taken into consideration. This conclusion is drawn on the basis of agreeing on diffusion Monte-Carlo and semiclassical (SC) calculations of the two isomers' zero-point energies, and upon DMC determination of the ground-state vibrational wave-function.

Molecular vibrational properties of *N*-methyl acetamide (NMA), and ethanol have been investigated. Diffusion Monte Carlo calculations were applied to characterize the vibrational ground state properties and wave-function, while the vibrational eigenenergies and eigenstates were calculated employing the vibrational self-consistent field (VSCF) and virtual-state configuration interaction (VCI) method using our home built MULTIMODE software.

Additionally, I developed a software to train and predict bimolecular thermal rate constants over a large temperature range with the use of Machine Learning technique. Several quantum mechanical methods are there to compute tunneling rate constant, however, each method has some bottleneck, and a remedy is suggested by this new approach. The approach uses Gaussian process (GP) regression to predict the rate constant values. Clustering and training are done over 52 different reactions and predictions are made for the new reactions that are not included into the training data.

Development of Machine-Learned Potential Energy Surfaces with Application to the Reactive and Vibrational Dynamics of High Dimensional Molecular Systems

by

Apurba Nandi

B.Sc., Jadavpur University, Kolkata, 2013

M.Sc., Indian Institute of Technology, Kanpur (IITK), 2015

Advisor: Prof. Joel M. Bowman, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2022

Acknowledgements

Obtaining a Ph.D degree was my childhood dream, and of all the people to whom I want to deliver thanks, I would like to first express my deepest gratitude to my supervisor, Professor Joel M. Bowman, without whom this dissertation would not be possible, for his guidance, teaching and continuous support. He has strengthened my critical understanding and creative thinking on modern quantum chemical research, and continues to serve as an important role-model for me. Through his instruction, I have become a better writer, presenter and independent thinker. His mentorship was paramount in providing a well rounded experience consistent with my long-term goals. I am extremely fortunate to have a supervisor like him.

I would like to thank Dr. Francesco Evangelista, and Dr. James T. Kindt for taking the time to be committee members of my graduate study and helpful suggestions. I am also extremely grateful to our past member Dr. Chen Qu for his patient guidance to develop my coding skill and our collaborator Prof. Paul L. Houston for insightful discussions.

I must also thank all the current and past members of the Bowman group: Dr. Qi Yu, and Qingfeng Kee Wang, especially Dr. Riccardo Conte for the valuable discussions, and Dr. Yimin Wang for her help and effort to solve problems on the computation facilities. I'm also grateful to Susan Browne for her coordination work. A special thanks to Dr. Simon Blakey, Kira Walsh, and Laura Hilado for their extreme support when I went through the difficult times and making my life easy at Emory.

I am also greatly thankful to my M.Sc. thesis supervisor Prof. Sridhar R. Gadre for

all his suggestions and unconditional support which makes my journey easier.

Last but not least, I want to deliver the deepest appreciation to my family, my father Late Ananda Nandi, and my mother Tapasi Nandi and my close relatives. They always gave me continuous encouragement and unconditional love. I also want to thank my new family member, Raima Dey Sahu, my fiance! She always gives me unconditional support and chooses to be accompany with me during any difficult times. I would also thank all my friends for their suggestions and encouragements. Because of them, my Ph.D study becomes smooth and enjoyable.

Table of Contents

List of Tables

List of Figures

Citations to Previously Published Work

Chapter 5 has been published as:

1. **Apurba Nandi, Chen Qu, and Joel M. Bowman*.** Using Gradients in Permutationally Invariant Polynomial Potential Fitting: A Demonstration for CH_4 Using as Few as 100 Configurations. ***J. Chem. Theory Comput.* 2019, *15*, 2826**.

Chapter 6 has been published as:

1. **Apurba Nandi, Chen Qu, and Joel M. Bowman*.** Full and Fragmented Permutationally Invariant Polynomial Potential Energy Surfaces for *trans* and *cis* *N*-methyl acetamide and Isomerization Saddle Points. ***J. Chem. Phys.* 2019, *151*, 084306**.

Chapter 7 has been published as:

1. **Apurba Nandi, Chen Qu, Paul L. Houston, Riccardo Conte, and Joel M. Bowman*.** Δ-machine Learning for Potential Energy Surfaces: A PIP Approach to Bring a DFT Based PES to CCSD(T) Level of Theory. ***J. Chem. Phys.* 2021, *154*, 051102**.

Chapter 8 has been published as:

1. **Subodh S. Khire, Nalini D. Gurav, Apurba Nandi,* and Shridhar R. Gadre*.** Enabling Rapid and Accurate Construction of CCSD(T)-Level Potential Energy Surface of Large Molecules Using Molecular Tailoring Approach. ***J. Phys. Chem. A.* 2022, *126*, 1458**.

Chapter 9 has been published as:

1. **Apurba Nandi,* Chen Qu,* Paul L. Houston,* Riccardo Conte,* Qi Yu,* and Joel M. Bowman*.** A CCSD(T)-Based 4-Body Potential for Water. ***J. Phys. Chem. Lett.* 2021, *12*, 10318**.

Chapter 10 has been published as:

1. **Chen Qu, Apurba Nandi, and Joel M. Bowman*.** ADIABATIC SWITCHING APPLIED TO THE VIBRATIONS OF *SYN*-CH_3CHOO AND IMPLICATIONS FOR "ZERO-POINT LEAK" AND ISOMERIZATION IN QUASICLASSICAL TRAJECTORY CALCULATIONS. ***Adv. Chem. Phys.* 2018, *163*, 151**.

Chapter 11 has been published as:

1. **Apurba Nandi, Chen Qu, and Joel M. Bowman *.** Diffusion Monte Carlo Calculations of Zero-Point Energies of Methanol and Deuterated Methanol. ***J. Comput. Chem.* 2019, *40*, 328**.

Chapter 12 has been published as:

1. **Apurba Nandi, Peng Zhang, Jun Chen, Hua Guo and Joel M. Bowman *.** Quasiclassical Simulations Based on Cluster Models Reveal Vibration-facilitated Roaming in the Isomerization of CO Adsorbed on NaCl. ***Nat. Chem.* 2021, *13*, 249**.

Chapter 1

Introduction

In chemistry, the systems we are dealing with consist of electrons and nuclei, and their motions are governed by the Schrödinger equation, except in extreme cases. In principle, the mysteries in any chemical process can be investigated accurately, if we are able to solve the exact Schrödinger equation. Unfortunately, however, even with the rapidly growing power of computational resources, solving the Schrödinger equation exactly for systems that only consist of a couple of atoms is still almost impossible. Therefore, various approximations have to be made to simplify the equation.

The famous Born-Oppenheimer (BO) approximation is the most important one based on the fact that the mass of an electron is much much smaller than the nucleus. Due to the massive difference in mass, the electrons move much much faster than the nuclei, and we can separate the motions of the electrons from nuclear motion. Therefore, instead of solving the Schrödinger equation simultaneously for all the electrons and nuclei, we can apply a two-step approach. In the first step, the Schrödinger equation for all the electrons at a set of instantaneous nuclear configurations can be solved. This allows us to construct the potential energy curve of a diatomic molecule, and in general, a potential

energy surface (PES) for any polyatomic species. In the second step, the Schrödinger equation of the nuclei can be solved using the PES obtained in the first step. In most cases, the BO approximation is exceptionally good, and the research presented in this dissertation is based on this approximation.

Even with the separation of the electronic and nuclear Schrödinger equation, solving either one of them is still very challenging. To solve the electronic part, a popular approach starts from the Hartree-Fock approximation, in which any one electron is assumed to move in an "average" potential due to all the other electrons. More accurate methods such as Møllet-Plesset perturbation (MP_n), coupled-cluster, and configuration interaction are all based on the Hartree-Fock theory. Another popular method worth mentioning is the density functional theory, in which functionals of electron density are used instead of the traditional wave function. These methods for the electronic structure problem are available in most quantum chemistry packages.

Next, we consider the Schrödinger equation of the nuclei. The dynamics of a molecule can be revealed when the molecular eigenstates and eigen-energies are obtained by solving time-dependent nuclear Schrödinger equation. The computational cost to solve these problems increases exponentially with increasing degrees of freedom, and this is the "curse of dimensionality" (Note that the electronic problem in fact also suffers from this "curse".). Many efforts have been taken into consideration to tackle the dimensionality problem. When the nuclear quantum effect is not significant, the nuclei are often treated as classical particles and those are governed by Newton's equation of motion. This greatly reduces the computational cost. However, when the nuclear quantum effect is big, several approximated quantum methods and semi-classical methods have been proposed in calculations of molecular eigenstates and eigen-energies.

The research presented in this dissertation basically follows the procedure described

above. It is seen that the PES plays an important role in the computation and prediction of the structure and dynamics of molecular species. Therefore, constructing an accurate PES is the first and the most significant part of each project. First, I performed electronic structure calculations to construct the PES, and then it was applied for the study of molecular vibrations and dynamics calculations of high dimensional molecular systems.

This dissertation is structured into three parts. The first part describes all the theories and computational methods I applied in all of my calculations. In Chapter 2, the Born-Oppenheimer approximation is explained in detail, and the method I used to construct the analytical representation of the PES is presented. The algorithms and methods employed in anharmonic vibrational calculations and dynamics simulations are described in Chapter 3 and Chapter 4, respectively.

The second part emphasizes the systematic developments in the PIP method for the construction of a robust analytical representation of PESs for large molecular systems. In Chapter 5, I describe the implementation of simultaneous energy-gradient fitting in the PIP method. Chapter 6 focuses on the implementation of fragmented basis approach for the development of accurate PESs for large molecular systems having more than 12 atoms. Using this approach I developed a full-dimensional PES of *N*-methyl acetamide and successfully applied this PES for quantum calculations. Chapter 7 emphasizes the Δ-Machine Learning approach for PES fitting, and how to correct a low-level DFT PES to a "Gold Standard" CCSD(T) level accuracy for large molecular systems. In Chapter 8, I present a rapid and accurate construction of full-dimensional CCSD(T)-level PESs for large molecules where fragmented-based electronic structure theory is employed to generate the CCSD(T) dataset in a cheaper way. Chapter 9 focuses on the construction of a very challenging PES of water 4-body interaction at the CCSD(T) level, which is one of the important parts of water many-body potential. This potential is successfully

applied for quantum calculations.

The final part of this dissertation covers my study on reaction dynamics and anharmonic vibration of gas-phase molecules using PIP PESs. Chapter 10 describes the application of the adiabatic switching method for "zero-point energy leak" and isomerization of *syn*-CH_3CHOO in quasiclassical trajectory calculations. Chapter 11 involves the study of quantum zero-point energies of methanol and its all deuterated isotopologs and isotopomers and the D/H exchange probability in their zero-point state which has a great interest in astrochemical research. In Chapter 12, I investigated a fascinating vibration-facilitated roaming mechanism in the isomerization of CO molecule on NaCl(100) surface by considering a $(CO\text{-}NaCl)_n$ finite cluster models. The novelty of this work is the isomerization was seen for highly excited CO vibrational states, in excellent agreement with the experiment. Then I present the nuclear quantum dynamics of Ethanol in Chapter 13, which reveals the "Leaky" nature of *trans* and *gauche* isomers when accurate vibrational zero-point energies and dynamical effects were taken into consideration. Chapter 13 focus on the details and results of semi-classical and fully-quantum anharmonic vibrational calculations of *trans* and *gauche* ethanol using recently developed CCSD(T) level PES. This is the first time fully-quantum anharmonic vibrational frequencies of ethanol are reported at the CCSD(T) level in excellent agreement with the reported experimental results.

Part I

Theories and Methods

Chapter 2

Potential Energy Surface

2.1 Born-Oppenheimer Approximation

The non-relativistic time-independent Hamiltonian (in atomic unit) for a molecule is:

$$\hat{H} = -\sum_{\alpha} \frac{1}{2M_\alpha} \nabla_\alpha^2 - \frac{1}{2} \sum_{i} \nabla_i^2 - \sum_{i,\alpha} \frac{Z_\alpha}{r_{i\alpha}} + \sum_{i<j} \frac{1}{r_{ij}} + \sum_{\alpha<\beta} \frac{Z_\alpha Z_\beta}{r_{\alpha\beta}}, \tag{2.1}$$

where α, β refer to the nuclei and i, j refer to the electrons; M_α and Z_α are the mass and charge of nucleus α. In a more compact notation, the Hamiltonian can be written as

$$\hat{H} = \hat{T}_N(\boldsymbol{R}) + \hat{T}_e(\boldsymbol{r}) + \hat{V}_{eN}(\boldsymbol{r}, \boldsymbol{R}) + \hat{V}_{ee}(\boldsymbol{r}) + \hat{V}_{NN}(\boldsymbol{R}), \tag{2.2}$$

where $\boldsymbol{R}$ denotes the set of nuclear coordinates and $\boldsymbol{r}$ is the set of electronic coordinates; $\hat{T}$ and $\hat{V}$ are the kinetic energy and potential operators, respectively.

The kinetic energy operator of the nuclei, $\hat{T}_N(\boldsymbol{r})$, can be neglected as it is more than three orders of magnitude smaller than $\hat{T}_e(\boldsymbol{r})$. Now according to Born-Oppenheimer approximation, the nuclei can be treated as fixed and we can separate the nuclear wave

function. Thus we can make the term $\hat{V}_{eN}(\boldsymbol{r}, \boldsymbol{R})$ parametrically depend on $\boldsymbol{R}$. That is, we can fix the $\boldsymbol{R}$ at a certain value $\boldsymbol{R_0}$, and solve for the electronic wavefunction $\Psi(\boldsymbol{r}; \boldsymbol{R_0})$:

$$\hat{H}_{el}\Psi(\boldsymbol{r}; \boldsymbol{R_0}) = E_{el}\Psi(\boldsymbol{r}; \boldsymbol{R_0}), \tag{2.3}$$

where

$$\hat{H}_{el} = \hat{T}_e(\boldsymbol{r}) + \hat{V}_{eN}(\boldsymbol{r}; \boldsymbol{R_0}) + \hat{V}_{ee}(\boldsymbol{r}) + \hat{V}_{NN}(\boldsymbol{R_0}). \tag{2.4}$$

When we repeat this for a set of $\boldsymbol{R_0}$, we obtain the potential energy surface (PES) of the molecule.

Suppose we have already solved equation 2.3 and obtained a set of eigenfunctions $\Psi_n(\boldsymbol{r}; \boldsymbol{R})$ and eigenvalues $E_n(\boldsymbol{R})$, i.e.

$$\hat{H}_{el}\Psi_n(\boldsymbol{r}; \boldsymbol{R}) = E_n(\boldsymbol{R})\Psi_n(\boldsymbol{r}; \boldsymbol{R}). \tag{2.5}$$

The exact solution of the original Hamiltonian in Eq. 2.2 can always be written in an infinite expansion.

$$\Phi(\boldsymbol{r}, \boldsymbol{R}) = \sum_k \Psi_k(\boldsymbol{r}; \boldsymbol{R}) * \chi_k(\boldsymbol{R}). \tag{2.6}$$

Insert this expansion into the original Schrödinger equation, multiply by $\Psi_n(\boldsymbol{r}; \boldsymbol{R})^*$, and integrate over the electronic coordinates $\boldsymbol{r}$, we obtain

$$\left[\hat{T}_N(\boldsymbol{R}) + E_n(\boldsymbol{R})\right]\chi_n(\boldsymbol{R}) - \sum_k \sum_\alpha \left[\frac{1}{M_\alpha} d^\alpha_{nk}(\boldsymbol{R}) \cdot \nabla_\alpha - \frac{1}{2M_\alpha} D^\alpha_{nk}(\boldsymbol{R})\right]\chi_k(\boldsymbol{R}) = \mathcal{E}\chi_n(\boldsymbol{R}), \tag{2.7}$$

where $d^\alpha_{nk}(\boldsymbol{R}) = \langle\Psi_n(\boldsymbol{r}; \boldsymbol{R})|\nabla_\alpha|\Psi_k(\boldsymbol{r}; \boldsymbol{R})\rangle$ and $D^\alpha_{nk}(\boldsymbol{R}) = \langle\Psi_n(\boldsymbol{r}; \boldsymbol{R})|\nabla^2_\alpha|\Psi_k(\boldsymbol{r}; \boldsymbol{R})\rangle$ are called non adiabatic couplings. In most cases, the couplings are very small and can be

safely neglected, and we obtain the following equation for the nuclear wave function:

$$\left[\hat{T}_N(\boldsymbol{R}) + E_n(\boldsymbol{R})\right] \chi_{nv}(\boldsymbol{R}) = \mathcal{E}_{nv}\chi_{nv}(\boldsymbol{R}). \tag{2.8}$$

The $E_n(\boldsymbol{R})$ is the PES and it is a hyper-dimensional function of the nuclear configuration for polyatomic molecules. With the PES, Eq. 2.8 could then be solved.

2.2 Permutationally Invariant Potential Energy Surface

In principle, we can always solve the electronic Schrödinger, Equation 2.3, "on the fly" at the desired nuclear configurations when necessary from electronic structure theory packages. However, the computational cost quickly becomes prohibitive as the level of electronic structure theory and the system size go up. Therefore, people have chosen an alternative approach, which is to develop an accurate analytical representation or a functional representation of potential energy in terms of the nuclear coordinates. As the exact analytical function form of the PES is unknown, developing an efficient and robust mathematical representation of a PES is very challenging. Initially, the PESs were constructed using the Shepard type interpolation method, i.e., the analytical potential agrees with *ab initio* energy exactly on any point in the database.[1–6] However, this becomes prohibitive when we move to systems with moderate size because the dimensionality of the problem increases so rapidly that the simple functions are unable to provide the accurate representation of PES. Therefore, the best strategy we can apply by selecting a well-behaved mathematical function $f(\boldsymbol{R};\boldsymbol{C})$ with a set of adjustable parameters $\boldsymbol{C}$ to represent the PES. The optimal parameters $\boldsymbol{C}$ can be determined via least-square process. First, a set

of electronic energies on different configurations $\{\boldsymbol{R}_i, V_i\}$ $(i = 1, \cdots, N)$ are computed, and then the parameters are determined by minimizing the residual $\sum_i (f(\boldsymbol{R}_i; \boldsymbol{C}) - V_i)^2$. There is extensive literature about obtaining the functional representation of the PES (e.g., see Ref. 7 and references therein).

Many efforts have been made in mathematical representations, functional form of $f(\boldsymbol{R}; \boldsymbol{C})$, of high-dimensional PESs in the last 30 years. Such as Neural Network (NN),[8–19] Gaussian Process Regression (GPR),[20–25] and Permutationally Invariant Polynomials (PIP) which is adopted in this work.[26–29] The variables used in the PIP method are "Morse variables" y_{ij}, which are the transformed internuclear distances r_{ij} via $y_{ij} = \exp(-r_{ij}/a)$, where a is in principle also an adjustable parameter but usually fixed at 2.0 Bohr.

The use of Morse variables guarantees that the potential does not diverge when the molecule dissociates. Therefore, the mathematical function can be written as

$$f(\boldsymbol{R}_i; \boldsymbol{C}) = \sum_{m=0}^{M} C_{\boldsymbol{b}} \left[\prod_{i<j}^{N} y_{ij}^{b_{ij}} \right] \quad \left(M = \sum b_{ij} \right), \tag{2.9}$$

where M is the maximum polynomial order, $\boldsymbol{b}$ stands for the ordered collection of the exponents b_{ij}, $C_{\boldsymbol{b}}$ are the linear coefficients that need to be determined, and N is the number of atoms.

The fundamental properties of a molecular potential are the invariant property with respect to permutations of like atoms as well as translational, and rotational invariant property. The use of Morse variables ensures that the potential is invariant with respect to translations and rotations of the molecule. So, the remaining permutational invariant property of a PES can be achieved by a straightforward way to replicate the electronic energies to all the equivalent configurations and fit the polynomial to this expanded

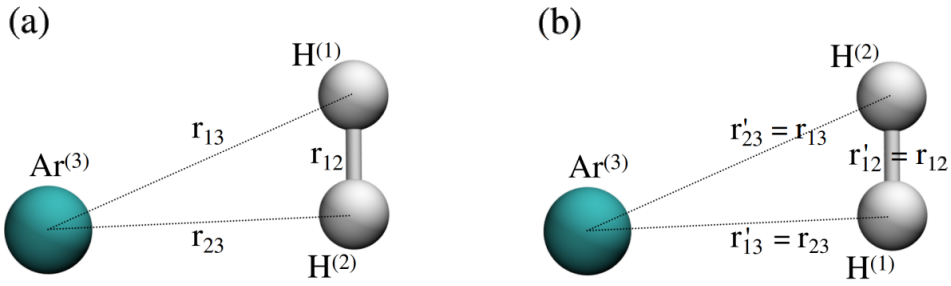


Figure 2.1: Permutation of the two identical H atoms in Ar−H_2.

dataset. However, a more elegant approach is to use a mathematical function in which the invariance property is built in. Two techniques have been developed by our group will be discussed here. In addition, several progresses have been made recently to incorporate the permutational invariance into NN,[30–34] and GPR[35] potentials.

2.2.1 Monomial symmetrization

The general form of a polynomial is shown in Eq. 2.9, which is not permutationally invariant, but this can be the starting point of monomial symmetrization. Here I just use a simple triatomic system as an example to illustrate the concept of this method, and then I will show how this method could be generalized to arbitrary molecules.

To begin with, consider the system Ar−H_2, and the molecule labeled as H(1), H(2), Ar(3) there are three internuclear distances, $\{r_{12}, r_{13}, r_{23}\}$. Then these internuclear distances are transformed to the Morse variables, $y_{ij} = \exp(r_{ij}/a)$, where we generally take the value of a = 2.0 Bohr. Now the potential of this triatomic system can be simply presented by the following expansion in monomials of Morse variables:

$$V(\boldsymbol{y}) = \sum_{m}^{M} C_{abc}[y_{12}^{a} y_{13}^{b} y_{23}^{c}]; (m = a + b + c). \tag{2.10}$$

This expression in general is not permutationally invariant, and this can be shown by first fixing the orders of one monomial, $y_{12}^a y_{13}^b y_{23}^c$. Now, consider the permutation of the H(1) and H(2). The monomial $y_{12}^a y_{13}^b y_{23}^c$ then maps onto $y_{12}^a y_{23}^b y_{13}^c$ which can be rewritten as the original monomial with permuted powers as $y_{12}^a y_{13}^c y_{23}^b$. Figure 2.1 shows the labels of each atom and all the internuclear distances (and thus the Morse variables) before and after the permutation. So the value of the sum of these monomials after the permutation becomes $C_{abc} y_{12}'^a y_{13}'^b y_{23}'^c + C_{acb} y_{12}'^a y_{13}'^c y_{23}'^b = C_{abc} y_{12}^a y_{23}^b y_{13}^c + C_{acb} y_{12}^a y_{23}^c y_{13}^b$. Clearly, there is no guarantee that the corresponding coefficients, C_{abc} and C_{acb} would be equal, with the exception when the permutation symmetry was included implicitly as mentioned, the numerical equality of the two coefficients might be achieved. Therefore, comparing the value of the sum of these two monomials before and after permutation, they are not equal unless the coefficients C_{abc} and C_{acb} have the same value. On the other hand, this illustration also suggests that the permutation symmetry can be easily ensured by using a single coefficient for both the terms, and thus the permutational invariance is built into the polynomial:

$$V(\boldsymbol{y}) = \sum_{m}^{M} D_{abc} \left[y_{12}^a y_{13}^b y_{23}^c + y_{12}^a y_{13}^c y_{23}^b \right]; (m = a + b + c). \tag{2.11}$$

For a general polyatomic molecule, we can find the mappings of all the possible permutations and use one coefficient for all the permutationally equivalent monomials. A general expression of the symmetrized monomials is

$$V(\boldsymbol{y}) = \sum_{m=0}^{M} C_{\boldsymbol{b}} \mathcal{S} \left[\prod_{i<j}^{N} y_{ij}^{b_{ij}} \right] \quad \left(m = \sum b_{ij} \right), \tag{2.12}$$

where $\mathcal{S}$ is the operator that symmetrizes the monomials. When the number of possible

permutations gets larger, finding the equivalent monomials becomes more difficult. Therefore, a more efficient approach has been implemented to representing these symmetrized monomials in a more compact way. In this approach, the symmetrized polynomials can be generated from the lower-order polynomials in a recursive way. This is called the monomial symmetrization approach (MSA). The rigorous derivation and proof of this theory are provided in Ref. 27. The details are beyond the scope of this work. Therefore, a brief description is given below.

Table 2.1: Monomials (m_i), and invariant polynomials (p_i) for A_2B type molecule up to degree 3.

Monomials	Polynomials
$m_0 = 1.0$	$p_0 = m_0$
$m_1 = y_{23}$	$p_1 = m_1 + m_2$
$m_2 = y_{13}$	$p_2 = m_3$
$m_3 = y_{12}$	$p_3 = m_4$
$m_4 = m_1 \times m_2$	$p_4 = p_1 \times p_2$
	$p_5 = p_1 \times p_1 - p_3 - p_3$
	$p_6 = p_2 \times p_2$
	$p_7 = p_2 \times p_3$
	$p_8 = p_3 \times p_1$
	$p_9 = p_2 \times p_5$
	$p_{10} = p_2 \times p_4$
	$p_{11} = p_1 \times p_5 - p_8$
	$p_{12} = p_2 \times p_6$

In MSA, all the possible monomials with a total degree less than some maximum value are enumerated, then symmetrized according to the permutation group of the molecule. Then, the polynomials are decomposed into the product of two polynomials of lower order, possibly subtracting a small number of same-order polynomials, when this is possible.

If this factorization is not possible, the polynomial is kept as the sum of monomials. Continuing with Ar−H_2 molecule (A_2B type system) as an example, the monomials and invariant polynomials up to the degree of 3 are listed in Table 5.1. As shown there, the first monomial is the constant term 1.0, and the following $n(n-1)/2$ monomials are just the Morse variables. The remaining monomials and then the polynomials can be obtained from previously computed ones. This approach was implemented in 2010 by Xie and Bowman in C++ and successfully applied to obtain a new PES of H_3O^+.[27]

2.2.2 Invariant polynomials

There is an alternative approach to obtain the invariant polynomials based on the invariant polynomial theory. This is mathematically more elegant and leads to a very compact expression that is more efficient to evaluate compared with the first one. The rigorous mathematical derivation and proof of this theory are provided in Ref. 36. The details are beyond the scope of this work, and are skipped here. On the basis of this theory, the invariant polynomial can be efficiently factorized into polynomials of primary invariants times secondary invariants. So, the potential can be expressed as

$$V(\boldsymbol{y}) = \sum_{d=0}^{M} C_{ab} h_a\left(\boldsymbol{p}(\boldsymbol{y})\right) q_b(\boldsymbol{y}), \tag{2.13}$$

where $\boldsymbol{p}(\boldsymbol{y})$ is the vector formed by the primary invariant polynomials and h_a is a polynomial of the primary invariants; $q_b(\boldsymbol{y})$ are secondary invariant polynomials, and d is the polynomial order whose maximum is set to M. For a molecule consisting of N atoms, there are $N(N-1)/2$ primary invariant polynomials. The computational algebra software MAGMA[37] is employed to obtain these primary and secondary invariant polynomials, and they have been implemented in a fitting library by Braams and Bowman.[26]

These two methods we mentioned to generate the permutationally invariant polynomials are numerically equivalent, however, the mathematical expressions are different. Compared to the MSA method, the invariant polynomials are generally more efficient to evaluate due to it's efficient factorization process as the products of primary and secondary invariant polynomials, but at the cost of less accessibility of the fitting basis functions. In addition to this, obtaining the analytical first derivatives from the unfactorized, MSA method is trivial, but the factorization in the invariant polynomial approach makes the derivation of the analytical first derivatives very difficult.

2.3 Dipole Moment Surface

The dipole moment is a three-dimensional vector, and its value is dependent on the choice of molecule coordinate. Thus the dipole moment surface (DMS) cannot be represented solely in terms of the internuclear and slightly different approach has been used for the dipole in terms of invariant polynomials. the expression of dipole moment, $\vec{\mu}$ can be written as a product of effective charges.

$$\vec{\mu}(\boldsymbol{R}) = \sum_i q_i(\boldsymbol{R})\vec{x}_i \tag{2.14}$$

Where $\boldsymbol{R}$ denotes the configuration, q_i is the effective charge on the ith nucleus, and $\vec{x}_i$ is the position vector of the ith nucleus. The effective charge is a scalar quantity, which can be expressed in the polynomials that are similar to those used for PESs.

Like PES the dipole moment should be invariant under the permutations of identical atoms, but there is a caveat to guarantee this invariance, which is the effective charges of like atoms should be "covariant" with respect to permutations. The term "covariant"

means, if the configuration transforms from $\boldsymbol{R}$ to $\boldsymbol{R}'$ by exchanging the identical nuclei i and j, the effective charges on these two nuclei have to satisfy the following relationship,

$$(q_j(\boldsymbol{R}'), q_i(\boldsymbol{R}')) = (q_i(\boldsymbol{R}), q_j(\boldsymbol{R}))\,, \tag{2.15}$$

and the charges on unique atoms are still invariant. So far, the covariant property has not yet been incorporated explicitly into the fitting basis, but rather we used a somewhat indirect and not quite optimal approach. Details can be found elsewhere.[26,38]

In addition to the covariant charges under permutations, the sum of the effective charges has to be equal to the net charge of the molecule. Ideally, this property should be built in the fitting basis, but in reality, we simply impose this as an additional constraint in the least-squares problem. Therefore, this condition is not strictly satisfied. This causes a drawback of the dipole fitting approach when we consider the dipole moment under translations. Let Z be the total charge of the molecule, and we apply a uniform displacement $\vec{r}$ to all the atoms. It must be $\vec{\mu}(\boldsymbol{R} + \vec{r}) = \vec{\mu}(\boldsymbol{R}) + Z\vec{r}$, but in the fitted dipole, the total charge is not exactly satisfied, and thus this property of the dipole under translations is not strictly satisfied as well.

2.4 Procedures

Here I describe the general procedures to construct a full-dimensional PES, and the procedures may vary slightly depending on the system and problem we are interested. The details of each individual PES will be discussed in the following chapters and can also be found in the corresponding references.

The very first step is the sampling of a configuration space and generate a dataset.

One simple way is the direct-product of grid. However, this could only apply for very small molecules. When we move to high-dimensional systems this method becomes prohibitive. Suppose, if we chose only 10 grid points, the total number of configurations would be 10^{3N-6} for a nonlinear polyatomic molecule, where N is the number of atoms. Even for $N = 4$, the number of configurations becomes too large. Another sophisticated way of sampling the configuration space is *ab initio* molecular dynamics (AIMD), at low-level electronic structure method such as DFT and MP2 with small basis sets. The AIMD simulations can be carried out at different stationary points, such as global or local minima, and saddle points, at several total energies to make sure the dataset has the coverage of extended region in the configuration space. The initial dataset is extracted from these AIMD trajectories.

Once the geometries are collected from low-level AIMD trajectory calculations, high-level *ab initio* calculations are performed on these configurations. Before starting the high level calculations, several test calculations are carried out to check the accuracy and computation time of a certain method/basis combination. Usually CCSD(T)-F12 method[39,40] with Dunning basis set (aug-)cc-pVXZ (X=D, T, ...)[41] provides good accuracy and efficiency. When bond-breaking process is involved, the single-reference coupled-cluster method fails, and multireference method is used. After all the *ab initio* energies are obtained, a linear regression fitting is performed to determine the coefficients of the initial PES.

The next step is to improve the initial fit if required. At first, the accuracy of this initial PES is examined by general fidelity tests such as, geometry optimization and normal mode frequency calculations of stationary points and then compared the results we obtained from this PES with the direct *ab initio* ones. If the agreement is not satisfactory, additional geometries are sampled by adding random displacements to those stationary

configurations. The energies of the additional geometries are calculated and added to the dataset. Furthermore, diffusion Monte Carlo simulations are performed to detect the "holes" in the PES, the regions having no data manifested by large negative values (lower than the global minimum), and then the additional configurations are added to the old data set. This iteration goes on until we are satisfied with the PES.

2.5 Summary

The approaches just mentioned have been used to developed a number of PESs that are discussed in detail in this thesis, and Table 2.2 shows a summary of PIP basis information about potential energy surfaces. These PESs are successfully applied to the dynamics simulations and anharmonic vibrational calculations. Further details can be found in the referred chapters as well as the cited references.

Table 2.2: A summary of a series of potential energy surfaces of which applications will be presented in this work.

Molecule	Max. Poly. Order	# Coeffecients	# Training Energies	# Training Gradients	Chapter
CH_4	5	208	9000	135000	5
N-methylacetamide	3	8040	6607	237852	6
Acetylacetone	2	86	430	....	8
Water 4-body	3	1649	2119	....	9
NaCl−CO	7	1716	8592	103104	12
Ethanol	4	14 752	11000	297000	13

Chapter 3

Molecular Vibrations

In this chapter I'll present the methods have been used to study the molecular vibrations. Vibrational properties of molecular system provide invaluable insight of its structure and dynamics. Advanced methods mainly focus on solving the nuclear Schrödinger equation which is a challenging task for polyatomic system. Many different models and techniques are reported. First, I will start with a simple harmonic model, where we assume that the potential is harmonic; and anharmonicity, and mode-dome couplings are negligible. Then the results could be systematically improved when vibrational self-consistent field (VSCF) and virtual-state configuration interaction (VCI) methods are employed. These methods have been implemented in the software called MULTIMODE. In addition to VSCF and VCI approach, diffusion Monte Carlo (DMC) is also described here which is another strategy to solve the vibrational ground state properties of molecule systems.

3.1 Normal Mode Analysis

The classical Hamiltonian of a polyatomic molecule can be written as

$$H = \sum_{i=1}^{3N} \frac{1}{2} m_i \dot{x_i}^2 + V(x_1, x_2, \cdots, x_{3N}), \tag{3.1}$$

where x_i, $\dot{x_i}$, and m_i are the positions, velocities, and masses of i^{th} degree of freedom. The subscript i runs from 1 to $3N$, where N is the number of atoms. Here $i = 1, 2, 3$ refers to the x, y, and z coordinates of the first atom, $i = 4, 5, 6$ refers to the x, y, and z of the second atom, etc. V is the potential and it is a function of the positions of all atoms.

Now we can apply famous Taylor series approximation and expand the potential, V about the minimum $\{x_i^{(0)}\}$ up to second-order terms:

$$V = V(\{x_i^{(0)}\}) + \sum_i V_i'(\{x_i^{(0)}\})(x_i - x_i^{(0)}) + \frac{1}{2} \sum_i \sum_j V_{ij}''(\{x_i^{(0)}\})(x_i - x_i^{(0)})(x_j - x_j^{(0)}), \tag{3.2}$$

where $V(\{x_i^{(0)}\})$ is the potential at the minimum, which could be an arbitrary value and for simplicity it is set to zero here. $V_i'(\{x_i^{(0)}\})$ are the first derivatives of the potential with respect to the coordinates. These first derivatives are zero as the potential is expanded about the minimum. The only terms that survive are the second derivatives:

$$V_{ij}''(\{x_i^{(0)}\}) = \left. \frac{\partial^2 V}{\partial x_i \partial x_j} \right|_{\{x_i^{(0)}\}}. \tag{3.3}$$

This is a non-separable Hamiltonian due to the coupling in second-order terms (off diagonal matrix elements). However, we can transform this coordinate system into a new coordinate system, Q, where this Hamiltonian matrix become diagonal and separable: $H = \sum_{i=1}^{3N} [T_i + V_i(Q_i)]$. This coordinate is the normal mode and it can be found using normal mode analysis.

First we define the mass-scaled Cartesian displacements q_i as $q_i \equiv \sqrt{m_i}(x_i - x_i^{(0)})$.

Thus, the classical Hamiltonian becomes

$$H = \frac{1}{2}\sum_{i=1}^{3N} \dot{q_i}^2 + \frac{1}{2}\sum_i \sum_j q_i F_{ij} q_j, \tag{3.4}$$

where F_{ij} are mass-scaled force constants:

$$F_{ij} \equiv \left.\frac{\partial^2 V}{\partial q_i \partial q_j}\right|_{\{0\}}. \tag{3.5}$$

In matrix notation, $\boldsymbol{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_{3N} \end{bmatrix}$, and the Hamiltonian is written as

$$H = \frac{1}{2}\dot{\boldsymbol{q}}^\top \dot{\boldsymbol{q}} + \frac{1}{2}\boldsymbol{q}^\top \boldsymbol{F} \boldsymbol{q}. \tag{3.6}$$

$\boldsymbol{F}$ is the mass-scaled Hessian matrix, which is a symmetric matrix. We know for a symmetric matrix, there exist an orthogonal matrix $\boldsymbol{L}$ that can diagonalize $\boldsymbol{F}$: $\boldsymbol{L}^\top \boldsymbol{F} \boldsymbol{L} = \boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_{3N})$. Let $\boldsymbol{q} = \boldsymbol{L}\boldsymbol{Q}$, and thus

$$H = \frac{1}{2}\dot{\boldsymbol{Q}}^\top \dot{\boldsymbol{Q}} + \frac{1}{2}\boldsymbol{Q}^\top \boldsymbol{\Lambda} \boldsymbol{Q} \tag{3.7}$$

$$= \sum_{i=1}^{3N} \left[\frac{1}{2}\dot{Q_i}^2 + \frac{1}{2}\lambda_i Q_i^2\right] \tag{3.8}$$

$$= \sum_{i=1}^{3N} \left[\frac{1}{2}\dot{Q_i}^2 + \frac{1}{2}\omega_i^2 Q_i^2\right], \tag{3.9}$$

where $\omega_i = \sqrt{\lambda_i}$.

Therefore, we've found the new coordinate $\boldsymbol{Q}$ in which the Hamiltonian can be written as the sum of individual Hamiltonians for each normal mode Q_i. Now the system looks like $3N$ uncoupled harmonic oscillators, with the harmonic frequencies $\omega_i = \sqrt{\lambda_i}$. Among these $3N$ normal modes, there are six zero-frequency modes, which corresponds to three translations and three rotations of a non-linear molecule.

Under this harmonic approximation, the quantum Hamiltonian of a molecule can be written as

$$\hat{H} = \sum_{i=1}^{3N-6} \left[-\frac{1}{2}\frac{\partial^2}{\partial Q_i^2} + \frac{1}{2}\omega_i^2 Q_i^2 \right]. \tag{3.10}$$

The eigenfunctions are just the direct products of $3N-6$ harmonic-oscillator wavefunctions, and the eigenvalues are the sums of harmonic-oscillator eigenenergies:

$$\psi_{n_1 \cdots n_{3N-6}} = \prod_{i=1}^{3N-6} \chi_{n_i}(Q_i); \tag{3.11}$$

$$E_{n_1 \cdots n_{3N-6}} = \sum_{i=1}^{3N-6} \hbar\omega_i (n_i + \frac{1}{2}), \tag{3.12}$$

where $\chi_{n_i}(Q_i)$ and n_i are the harmonic-oscillator eigenfunction and the quantum number of the i-th normal mode.

In this harmonic approximation, higher-order terms in the potential expansion are neglected. In many applications, those non-harmonic terms are also very important and should be taken into consideration. The non-harmonic terms can be treated as a perturbation to the Hamiltonian, and this leads to the vibrational second-order perturbation (VPT2) theory.[42,43]. In this work, I employ a more rigorous method, VSCF and VCI, which is very similar to the Hartree-Fock and post-Hartree methods in electronic structure theory.[44]

3.2 Vibrational Self-Consistent Field and Virtual-state Configuration Interaction

Here we write the molecular Hamiltonian in terms of normal mode coordinates:

$$\hat{H} = \sum_{i=1}^{3N-6} \hat{T}_i + V(\boldsymbol{Q}), \tag{3.13}$$

where $\boldsymbol{Q} = [Q_1 \ \cdots \ Q_{3N-6}]$ and $\hat{T}_i$ is the kinetic-energy operator of the i-th normal mode. Here the potential energy, $V(\boldsymbol{Q})$ is the full potential of the molecule.

In the VSCF method, the variational approach is used to solve this problem, and the trial wavefunction of a quantum state can be representated as a direct product of one-mode wavefunctions:

$$\psi(\boldsymbol{Q}) = \prod_{i=1}^{3N-6} \phi_i(Q_i). \tag{3.14}$$

The goal is to find a set of optimal one-mode functions $\{\phi_i(Q_i)\}$ $(i = 1, \cdots, 3N-6)$ that minimizes the energy functional

$$E[\{\phi_i\}] = \langle\psi|\hat{H}|\psi\rangle \tag{3.15}$$

under the condition $\langle\phi_i|\phi_i\rangle = 1$, i.e., ϕ_i are normalized. This problem could be solved using Lagrange's method of undetermined multiplier, and we construct the Lagrangian

$$\mathcal{L}[\{\phi_i\}, \varepsilon_1, \cdots, \varepsilon_{3N-6}] = \langle\psi|\hat{H}|\psi\rangle - \sum_{i=1}^{3N-6} \varepsilon_i \left(\langle\phi_i|\phi_i\rangle - 1\right). \tag{3.16}$$

By setting

$$\frac{\delta\mathcal{L}}{\delta\phi_i} = \frac{\delta\mathcal{L}}{\delta\varepsilon_i} = 0, \tag{3.17}$$

a set of $3N-6$ coupled equations ($i = 1, 2, \cdots, 3N-6$)

$$\left[\hat{T}_i + \left\langle \prod_{k\neq i}^{3N-6} \phi_k(Q_k) \middle| V(\boldsymbol{Q}) \middle| \prod_{k\neq i}^{3N-6} \phi_k(Q_k) \right\rangle\right] \phi_i(Q_i) = \varepsilon_i \phi_i(Q_i) \tag{3.18}$$

are obtained. These coupled equations are solved in an iterative way for each model wavefunction until self-consistency is reached. Each modal function $\phi_i(Q_i)$ is expressed as a linear combination of a finite set of basis functions multiplied by unknown coefficients.

$$\phi_i(Q_i) = \sum_{n_i}^{l_i} C_{n_i} \chi^{(n_i)}(Q_i). \tag{3.19}$$

where l_i is the number of basis functions for the i_{th} mode. The coefficients can be determined by diagonalizing the Hamiltonian matrix.

This VSCF method is similar to the Hartree-Fock method in electronic structure theory. It is a one-mode method, and the coupling between one mode and the others is taken into account in an "averaged" manner. The result of the VSCF calculation can be further improved by configuration interaction calculation, which is termed as virtual state configuration interaction method (VCI). Recall that for each mode, we solve the VSCF equations to obtain a set of eigenfunctions $\phi_i^{(v_i)}(Q_i)$ with different quantum number v_i. Therefore, a series of total wavefunctions can be constructed:

$$\psi_{\boldsymbol{v}}^{\mathrm{VSCF}}(\boldsymbol{Q}) = \prod_{i=1}^{3N-6} \phi_i^{(v_i)}(Q_i). \ (v_i = 0, 1, 2, \cdots; \text{ and } i = 1, \cdots, 3N-6) \tag{3.20}$$

One of them is the VSCF ground state, and the others are virtual states. In the VCI calculation, we expand the total wavefunction in terms of the VSCF ground and virtual

states:

$$\Psi(\boldsymbol{Q}) = \sum_{\boldsymbol{v}} C_{\boldsymbol{v}} \psi_{\boldsymbol{v}}^{\mathrm{VSCF}}(\boldsymbol{Q}), \tag{3.21}$$

and the coefficients $C_{\boldsymbol{v}}$ are obtained by diagonalizing the Hamiltonian matrix.

3.3 The Software "Multimode"

The VSCF and VCI methods are implemented in the software called "MULTIMODE".[45,46] This computational code is based on the rigorous Watson Hamiltonian for non-linear molecule. MULTIMODE has been widely applied to a variety of molecular systems using a n-mode representation of the full potential energy. However, there are few practical issues. In VSCF, integrals

$$\left\langle \prod_k \phi_k(Q_k) \middle| V(\boldsymbol{Q}) \middle| \prod_k \phi_k(Q_k) \right\rangle \tag{3.22}$$

have to be evaluated, and when the number of modes is large this multi-dimensional integration become computationally very expensive. Additionally, in the VCI calculation, the Hamiltonian matrix could be very large without any restriction in the expansion in Equation 3.21. The strategies have been taken into account to circumvent these issues in MULTIMODE are presented below.

3.3.1 Watson Hamiltonian

In the derivation of VSCF and VCI, the rotation-vibration coupling is neglected in the Hamiltonian. However, in MULTIMODE, the rigorous Watson Hamiltonian is used.[47] For

any non-linear molecule, the Watson Hamiltonian is represented in normal coordinates,

$$\begin{aligned}\hat{H} =& \frac{1}{2}\sum_{\alpha,\beta}\left(\hat{J}_\alpha - \hat{\pi}_\alpha\right)\mu_{\alpha\beta}\left(\hat{J}_\beta - \hat{\pi}_\beta\right) - \frac{1}{2}\sum_{k=1}^{3N-6}\frac{\partial^2}{\partial Q_k^2} \\ & - \frac{1}{8}\sum_{\alpha}\mu_{\alpha\alpha} + V(Q_1, \cdots, Q_{3N-6}),\end{aligned} \tag{3.23}$$

where $\alpha, \beta = x, y, z$; $\hat{J}_\alpha$ and $\hat{\pi}_\alpha$ are the total and vibrational angular momenta, respectively; $\mu_{\alpha\beta}$ is the inverse of the effective moment of inertia tensor and $V(Q_1, \cdots, Q_{3N-6})$ is the full potential in terms of the N normal coordinates. The vibrational angular momenta are given by

$$\hat{\pi}_\alpha = -i\sum_{k,l}\zeta_{k,l}^{\alpha} Q_k \frac{\partial}{\partial Q_l}, \tag{3.24}$$

where $\zeta_{k,l}^{\alpha}$ are Coriolis coupling constants, and these vibrational angular momentum terms usually cannot be neglected. In most cases, we solve the $J = 0$ Schrödinger equations.

This Hamiltonian works for semi-rigid molecules; however, for molecules with one large-amplitude motion, the reaction path version of Multimode should be used, which is based on the reaction path Hamiltonian.[48] This has been applied to molecules with internal rotation, such as CH_3OH and H_2O_2.[49,50] This reaction path Multimode is not used in this work, so it will not be described here. The theory and details are given in Ref. 48 and 51.

3.3.2 n-Mode representation of Potential

In MULTIMODE, the full dimensional potential of a molecule is written as a hierarchical n-mode representation (nMR):

$$V(Q_1, Q_2, \cdots, Q_m) = \sum_i V_i^{(1)}(Q_i) + \sum_{i<j} V_{ij}^{(2)}(Q_i, Q_j) + \sum_{i<j<k} V_{ijk}^{(3)}(Q_i, Q_j, Q_k) + \sum_{i<j<k<l} V_{ijkl}^{(4)}(Q_i, Q_j, Q_k, Q_l) + \cdots, \tag{3.25}$$

where one mode representation,

$$V_i^{(1)} = V(Q_i, Q_{l\neq i} = 0), \tag{3.26}$$

$$V_{ij}^{(2)} = V(Q_i, Q_j, Q_{l\neq i,j} = 0) - V_i^{(1)}(Q_i) - V_j^{(1)}(Q_j), \tag{3.27}$$

$$V_{ijk}^{(3)} = V(Q_i, Q_j, Q_k, Q_{l\neq i,j,k} = 0) - V_{ij}^{(2)} - V_{ik}^{(2)} - V_{jk}^{(2)} - V_i^{(1)} - V_j^{(1)} - V_k^{(1)}. \tag{3.28}$$

In MULTIMODE, this expansion is truncated, and the maximum number of modes allowed is six. Therefore, numerical quadratures with maximum dimensionality of six are needed, instead of $3N - 6$, and this could greatly reduce the computational cost of multi-dimensional integrals. In addition, efficient Gauss-Hermite quadrature is used in Multimode for numerical integration.

3.3.3 VCI excitation space and matrix pruning

We have seen above in VCI calculation the full-CI matrix size is the direct product of the number of basis for each mode, and it could be enormous when the molecule is large. Analogous to the truncated CI in electronic structure theory (for example, CISD or CISDT), constraints are imposed in MULTIMODE. We can use an "m-mode basis"

to restrict the excitation space to a maximum of m modes excited simultaneously. For each mode, we specify a maximum value of the quanta of excitation (called MAXBAS in Multimode), and in addition, the sum of quanta is limited by a user-specified value (called MAXSUM in Multimode). With all these restrictions, the size of the CI matrix is greatly reduced.

Even with this truncated CI approximation, the CI matrix becomes very large for moderate size molecules (10 to 15 atom systems) and this become very problematic for direct diagonalization. The following strategies are taken into account to overcome this issue. First, we impose the symmetry of the molecule which transform the full-CI matrix into a block diagonal matrix and we can easily diagonalized each block instead of the full matrix. In addition, we only need the lowest a few hundred or thousand eigenstates, and therefore, more efficient iterative process, block-Davidson method is implemented. Furthermore, rows and columns of the CI matrix can be eliminated, based on a perturbation test. This was initially implemented in Multimode by Handy and Carter,[52] and have been optimized recently in a vibrational calculation of CH_3NO_2.[53]

3.3.4 Infrared intensity

The intensities of vibrational transitions can be computed from MULTIMODE calculations using the "dump-restart" procedure[54] if the dipole moment surface (DMS) is available. During VSCF/VCI calculation, the VCI wave functions can be written to the disk, and they will be used for property calculation, such as, infrared intensity calculation. Now, the transition dipole matrix elements are calculated according to

$$R_{\alpha if} = \int \Psi_i(\boldsymbol{Q})\mu_\alpha(\boldsymbol{Q})\Psi_f(\boldsymbol{Q})d\boldsymbol{Q}, \tag{3.29}$$

where $\boldsymbol{Q}$ is the set of normal coordinates, and $\mu_\alpha(\boldsymbol{Q})$ is the α component ($\alpha = x, y, z$) of the dipole moment. Ψ_i and Ψ_f are the initial and final vibrational state of the transition.

The infrared intensity of the $i \rightarrow f$ transition is evaluated using the expression

$$I_{if} = \frac{8\pi^3 N_A}{3hc(4\pi\epsilon_0)} \nu \sum_\alpha |R_{\alpha if}|^2 (N_i - N_f), \tag{3.30}$$

where N_A is the Avogadro's number, ν is the wavenumber of the transition, and N_i is the number of the molecules in state i. If we only consider the transition originated from the vibrational ground state, $N_i - N_f$ is 1.

3.4 Diffusion Monte Carlo

The DMC method is the stochastic approach to solve the exact nuclear Schrödinger equation for the vibrational ground state energy and wavefunction. The concept behind the DMC method is to solve the time-dependent Schrödinger equation in imaginary time.[55–59] This is done by simulating a random walk of many replicas, also called "walkers", of the molecule, using a birth/death processes. This is a powerful method to extract the ground state wave function and energy of the molecule from the probability distribution of the random walkers. Here, the basic theory and a practical algorithm of DMC are described.

3.4.1 Theory

The procedure can be illustrated in one dimension (1D) for a particle of mass m that moves in a potential V(x). The time-dependent Schrödinger equation of this particle is written as

$$i\hbar \frac{\partial \psi(x,t)}{\partial t} = \hat{H}\psi(x,t) = -\frac{\hbar^2}{2m}\frac{\partial^2 \psi}{\partial x^2} + V(x)\psi, \tag{3.31}$$

where $\hat{H}$ is the Hamiltonian of the system. The wave function can be written as and the solution can be expressed as

$$\psi(x,t) = \sum_{n=0}^{\infty} c_n \phi(x)_n \exp\left[\frac{-iE_n t}{\hbar}\right], \tag{3.32}$$

where $\phi_n(x)$ are the eigenfunctions of the time-independent Schrödinger equation with the eigenvalues E_n,

$$\hat{H}\phi_n(x) = E_n \phi_n(x) \tag{3.33}$$

and the order of the eigenvalues are follows as

$$E_0 < E_1 \leq E_2 \leq E_3 \leq .... \tag{3.34}$$

Next, we shift the energy scale by E_R such that $V \rightarrow (V(x) - E_R)$ and $E_n \rightarrow (E_n - E_R)$, where E_R is a reference energy. Then we introduce a new variable $\tau = it$ to transform real time to imaginary time. So, the 1-D time dependent Schrödinger equation becomes

$$\hbar \frac{d}{d\tau}\Psi(x,\tau) = \frac{\hbar^2}{2m}\frac{d^2}{dx^2} - [V(x) - E_R]\Psi(x,\tau), \tag{3.35}$$

and the the wave functions become

$$\Psi(x,\tau) = \sum_{n=0}^{\infty} c_n \phi_n(x) \exp\left[\frac{-(E_n - E_R)\tau}{\hbar}\right]. \tag{3.36}$$

If $E_R > E_0$, $\psi(x,\tau)$ diverges when τ approaches ∞; if $E_R < E_0$, $\psi(x,\tau)$ decays to zero; only if $E_R = E_0$, $\psi(x,\tau)$ converges to ϕ_0. That means, if we choose the E_R to be the ground state energy E_0, and propagate the system to large imaginary time values, $\psi(x,\tau)$ converges to the ground state of the Hamiltonian $\hat{H}$. Of course E_0 is unknown in advance,

and its value could be determined by diffusion Monte Carlo method.

The Equation 3.35. of DMC method is similar to the diffusion equation with a first-order rate term

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - kC. \tag{3.37}$$

The "diffusion coefficient" of the imaginary-time Schrödinger equation is $D = \sqrt{\hbar/2m}$. The second derivative part can be modeled with a random walker process with large number of walkers, and the first-order term can be viewed as a source or sink of the walkers.

At sufficiently short times, the solution of Equation 3.37. can be approximated as

$$C(x,t) \approx U(x,t)C(x,0)\exp\left(-kt\right), \tag{3.38}$$

where

$$U(x,t) = \frac{1}{\sqrt{4\pi Dt}}\exp\left[-\frac{(x-x_0)^2}{4Dt}\right] \tag{3.39}$$

is the solution to the diffusion problem without the first-order term, using δ-function as the initial condition. Therefore, in the simulations, each time increment consists of two steps: in the first step the walkers are assigned a random displacement based on the distribution specified by Equation 3.39., and in the second step, the walkers are removed or replicated based on the probability density $\exp(-kt)$ from Equation 3.38. A practical algorithm is given next.

3.4.2 Algorithm

Initially, we generate $\mathcal{N}_0$ replicas from the same initial geometry, so that the initial condition can be viewed as a δ-function. At each step, a random displacement Δx is

assigned to each walker. This displacement is selected from a Gaussian distribution

$$P(\Delta x) = \frac{1}{\sqrt{4\pi D\Delta\tau}} \exp\left[-\frac{(\Delta x)^2}{4D\Delta\tau}\right], \tag{3.40}$$

with $D = \sqrt{\hbar/2m}$ and $\Delta\tau$ is the step size in the simulation. After the displacement, the potential of each walker, $V(x)$, is calculated and the corresponding weight function

$$W = \exp\left[-(V(x) - E_R)\Delta\tau\right] \tag{3.41}$$

is computed. For each walker, if $W < 1$, it will be removed with probability $1 - W$; if $W > 1$, a new walker with the same configuration will be added, and then an additional one may be added with probability $W - \text{int}(W)$, where $\text{int}(W)$ is the largest integer that does not exceed W. This is call birth/death process. When these are done for all the walkers, the reference energy is calculated as

$$E_R(\tau) = \langle V(\tau)\rangle - \alpha\frac{\mathcal{N}(\tau) - \mathcal{N}(0)}{\mathcal{N}(0)}, \tag{3.42}$$

where $\langle V(\tau)\rangle$ is the average potential over all walkers, and $\mathcal{N}(\tau)$ is the number of live walkers at imaginary time τ. Here α is a feed-back parameter that controls the fluctuations of the number of the walkers and the reference energy. After equilibration, the average of the reference energy over the imaginary time gives an estimate of the quantum zero-point energy, and the distribution of the walkers, when properly normalized, represents the ground-state wavefunction.

Chapter 4

Classical Trajectory Simulations

According to Born-Oppenheimer approximation, one can solve the time-dependent nuclear Schrödinger equation for the motion of nuclei. However, this is also very challenging, and rigorous quantum calculations become prohibitive for the system with more than four atoms.

Therefore, in many cases people neglect the nuclear quantum effects, and perform classical trajectory calculations where the motions of the nuclei can be described sufficiently well by classical mechanics. These classical trajectory simulations are widely used in the study of dynamical and chemical process. The motions of the nuclei follow the classical Hamilton's equations (also equivalent to Newton's and Lagrange's equations):

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \ \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \tag{4.1}$$

$$H = T(p,q) + V(q), \tag{4.2}$$

where q_i and p_i are the position and momentum of i-th atom, respectively. H is the classical Hamiltonian, which is the sum of the kinetic T(p,q) and potential energy V(q).

In molecular dynamics (MD) simulations, the classical equations of motion in Eq. 4.1 are integrated numerically if the potential is known. Note that our potential energy surfaces (PESs) allow very efficient integration, compared to "on-the-fly" *ab initio* molecular dynamics.

4.1 Initial Conditions

In the classical dynamic simulations, initial conditions of the sampling are very crucial, because these are usually chosen to directly compare with experimental results or other benchmarked calculations.[60] Several methods exist for the initial condition sampling, such as microcanonical sampling (NVE), canonical sampling (NVT), isobaric sampling (NPT), and normal mode sampling etc. Microcanonical sampling and normal mode sampling are the two most commonly used methods in our projects.

4.1.1 Microcanonical Sampling (NVE)

Microcanonical sampling is the simplest and straightforward sampling process which satisfy the total energy E constraint in the phase space. During its implementation, for a fixed molecular configuration the velocity of each atom is selected from a uniform distribution in $(-0.5, 0.5)$. Then the velocities are scaled to make the total kinetic constraint. However, this simple implementation fails to describe the zero-point motion of a molecule, and the total energy E could even be lower than the zero-point energy (ZPE) of a molecule. This could lead to poor estimation of reaction barriers for a polyatomic molecule having large amount of ZPE (tens of kcal/mol). On the other hand, this microcanonical sampling is unable to model mode-specific reactions where one or more normal modes are excited. Therefore, the normal mode sampling is introduced.

4.1.2 Normal Mode Sampling

Normal mode sampling is also a type of microcanonical sampling, where total energy E equals to the energy of a specific vibrational state. The implementation is very simple, and It is done in normal mode coordinates. First, we perform normal mode analysis by diagonalizing the mass-weighted force constant matrix, and we obtain the harmonic frequencies $\boldsymbol{\omega}$, and the corresponding normal mode vector $\boldsymbol{L}$. The Hamiltonian of an n-mode system can be approximately written by the sum of energies for each harmonic oscillators, given by:

$$H(P,Q) = E_{Total}, \tag{4.3}$$

$$E_{Total} = \sum_{i=1}^{n} E_i = \sum_{i=1}^{n} \frac{P_i^2 + \omega_i^2 Q_i^2}{2}, \tag{4.4}$$

where P_i and Q_i are the momentum and normal coordinate of i-th mode, respectively. Now, random values for Q_i and P_i are chosen by assigning a random phase to each mode for making an uniform distribution in the phase space.

$$Q_i = \frac{\sqrt{2E_i}}{\omega_i} \cos(2\pi r_i); \tag{4.5}$$

$$P_i = -\sqrt{2E_i} \sin(2\pi r_i), \tag{4.6}$$

where r_i is a random number from a uniformed distribution in $(0, 1)$, and E_i is the energy one would like to put in the i-th mode. Typically E_i is the harmonic zero-point energy (ZPE) of that mode, and in mode-specific dynamics, it could be the energy of a certain excited state. Then the Q and P are transformed back to Cartesian coordinates $\boldsymbol{q}$ and

momenta $\boldsymbol{p}$ using normal mode vector,$\boldsymbol{L}$ by:

$$\boldsymbol{q} = \boldsymbol{q}_0 + \boldsymbol{M}^{-1/2}\boldsymbol{L}\boldsymbol{Q} \tag{4.7}$$

$$\boldsymbol{p} = \boldsymbol{M}^{1/2}\boldsymbol{L}\boldsymbol{Q}, \tag{4.8}$$

where $\boldsymbol{q}_0$ is the coordinates of the equilibrium configuration, and $\boldsymbol{M}$ is a diagonal matrix whose elements are masses of the nuclei.

During this sampling process a spurious angular momentum $\boldsymbol{j}_s$ could be generated, and this is calculated by

$$\boldsymbol{j}_s = \sum_i \boldsymbol{r}_i \times m_i \dot{\boldsymbol{r}}_i, \tag{4.9}$$

where $\boldsymbol{r}_i$ is the position of the i-th nucleus. Assume that $\boldsymbol{j}_0$ be the desired angular momentum, so $\boldsymbol{j} = \boldsymbol{j}_0 - \boldsymbol{j}_s$ is the additional amount which should be added to the system. Therefore the velocity, $(\boldsymbol{I}^{-1}\boldsymbol{j}) \times \boldsymbol{r}_i$ is added to each atom, where $\boldsymbol{I}$ is the moment of inertia tensor.

Finally the internal energy E, after the two steps mentioned above, slightly deviates from the desired total internal energy E_0. Therefore, the Cartesian coordinates and momenta are scaled according to

$$q_i' = q_{i,0} + (q_i - q_{i,0})\sqrt{E_0/E} \tag{4.10}$$

$$p_i' = p_i\sqrt{E_0/E}. \tag{4.11}$$

We follow this process in loops until the actual internal energy agrees with the desired internal energy.

4.2 Final Conditions

In classical dynamics simulations, properties of the products such as the translational energy release, the rotational and vibrational energy distribution can be calculated from the coordinates and velocities of the nuclei at the termination of a classical trajectory. Here I briefly describe the procedures of final condition analysis.[61]

4.2.1 Translations and Rotations

During the classical dynamics simulation, the system should be translationally invariant, means the translational energy should be equal to zero. A reaction may have several products and but here we just focus on one fragment that is of interest. Upper case letters are used for the center-of-mass positions, and velocities of the product, and lower case letters for each individual atom. The center-of-mass velocity of a fragment can be calculated by

$$\boldsymbol{V} = \frac{\sum_i m_i \boldsymbol{v}_i}{M}, \tag{4.12}$$

where the sum is over all the atoms in that fragment, and M is the total mass of that fragment. Therefore the translational energy of this fragment is simply $E_{\text{trans}} = \frac{1}{2} M \left|\boldsymbol{V}\right|^2$. By setting $\boldsymbol{v}'_i = \boldsymbol{v}_i - \boldsymbol{V}$, we can remove this translational motion from the fragment and now the internal energy of this fragment becomes sum of the kinetic and potential energy.

$$E_{\text{int}} = T + V(q) = \sum_i \frac{1}{2} m_i \left|\boldsymbol{v}'_i\right|^2 + V(q), \tag{4.13}$$

where $V(q)$ is the potential energy in generalized coordinates relative to the equilibrium structure.

This internal energy consists of vibrational energy and rotational energy. The rota-

tional angular momentum can be calculated by $\boldsymbol{j} = \sum_i m_i \boldsymbol{r}'_i \times \boldsymbol{v}'_i$, where $\boldsymbol{r}'_i$ and $\boldsymbol{v}'_i$ are the position and velocity of atom i in center-of-mass frame. Then the rotational energy is given by

$$E_{\text{rot}} = \frac{1}{2}\boldsymbol{j}^\top \boldsymbol{I}^{-1}\boldsymbol{j}, \tag{4.14}$$

where $\boldsymbol{I}$ is the moment of inertial tensor of the fragment. For diatomic molecules, the rotational quantum number J can be determined using $|\boldsymbol{j}| = \sqrt{J(J+1)}\hbar$ and rounded to the nearest integer.

4.2.2 Vibration and Zero-point Energy Constraint

The vibrational energy can be written as $E_{\text{vib}} = E_{\text{int}} - E_{\text{rot}}$. For diatomic molecules, the vibrational quantum number n can be determined by

$$\left(n + \frac{1}{2}\right)\hbar\omega = E_{\text{vib}}, \tag{4.15}$$

where ω is the harmonic frequency.

ZPE violation is one of the common problem in classical simulation. The vibrational energy of the product could be lower than the ZPE, which is not allowed in quantum mechanics. To overcome this, a straightforward ZPE constraint could be applied as follow. If the total vibrational energy of the products is smaller than the sum of ZPEs of each individual product, the trajectory is discarded and will not be considered in the final condition analysis, or if the vibrational energy of any fragment is smaller than its ZPE, the trajectory is discarded. The former is the soft ZPE constraint and the latter is hard ZPE constraint.

Part II

Systematic Developments in PIP Method for Generating High-dimensional PESs

Chapter 5

Implementation of Simultaneous Energy-Gradients Fitting in Permutationally Invariant Polynomial Approach

5.1 Chapter Abstract

We describe software to incorporate electronic energies and gradients to develop high-dimensional potential energy surfaces, using a permutationally invariant polynomial basis. The energies and gradients are obtained using direct dynamics, using the efficient B3LYP/6-31+G(d) level of theory. The new software is described along with extensive testing and assessment of the benefits of using gradients as well as energies for CH_4. Starting with a dataset of 9,000 configurations, we examine training and testing on datasets of energies only and energies plus gradients with datasets as small as 50. In addition to

standard root-mean-square fitting errors of energies and gradients, normal-mode analyses and diffusion Monte Carlo calculations are performed to examine the fidelity of the fits using gradients. We show that a precisely fitted potential surface can be obtained using energies and gradients with only 100 or even just 50 widely-scattered configurations. Finally, several fits are done using all the data from direct-dynamics trajectories with 1,000 steps. These are more demanding fits compared to the one based on pruning datasets. The results of these fits are encouraging.

5.2 Overview

Developing potential energy surfaces (PESs) continues to be an active area of theoretical and computational research. One can get the PES "on the fly" by obtaining the electronic energies from electronic structure packages; however, the computational cost quickly becomes prohibitive as the level of electronic structure theory goes up. The alternative is the development of accurate analytical representations of PESs. The main challenge is to develop an efficient and robust mathematical representation of the PES, using datasets of electronic energies that span a high-dimensional space.

In the past 15 years, significant progress has been made in the non-parametric, mathematical representations of high-dimensional PESs for molecules and clusters that contain more than four atoms.[13–16,19,21–23,25,26,30,34,62–65] The three major methods currently in widespread use are Permutationally Invariant Polynomials (PIP)[26,65], Neural Networks (NN),[14–16,19,66–69] NN with PIP inputs[30,31,34,63,64,70], Gaussian Process regression (GPR),[20,21,25] and GPR with PIP inputs.[35] This field is developing rapidly and the references cited above are not meant to be comprehensive.

The applications using the various non-parametric methods can be divided into two

broad areas. One is isolated molecules, including chemical reactions[29,34,64,71] and hydrogen-bonded clusters.[65] In these applications, the PESs span large regions of configuration space but are limited to 5–10 atoms. A recent example to the 10-atom formic acid dimer has been reported,[72,73] which, however, is not a global PES. Most applications in this area make use of high-level electronic energies, e.g., CCSD(T). These are feasible to perform for this number of atoms and for tens of thousands of configurations. In general gradients are not used in the fits in this case, as these are expensive to calculate using CCSD(T) theory. The second application area deals with tens of atoms, and so an efficient electronic structure method, i.e., density functional theory, is used.[15,20,21,69] A high-level method such as CCSD(T) is not feasible for tens of heavy atoms such as Cu, Si, etc. or tens of water monomers.

Our group has developed the PIP approach over the past 15 years and with numerous applications.[26–29,65] This approach was first applied in 2003 to construct a global PES of CH_5^+ cation, a notoriously fluxional molecule, using a basis of polynomials that are invariant with respect to 120 permutations of the five equivalent H atoms.[74] The PESs developed in our group generally make use of datasets of 10^4–10^5 scattered electronic energies, typically obtained at the CCSD(T) level of theory. However, the scattered configurations are obtained using DFT direct dynamics, generally at several total energies. After pruning the DFT dataset to a relatively small size, but generally with energies that span a range of several hundred to roughly 20,000 cm^{-1}, new energies are obtained at the CCSD(T) level of theory. The fitted PESs are typically tested by making comparisons with direct *ab initio* calculations of harmonic normal-mode frequencies at stationary points and the results are generally highly precise.

For challenging cases with numerous minima and high-energy saddle points, global PESs are generally less precise and of course the accuracy of normal-mode frequencies de-

creases. Counter examples to this statement though are full-dimensional, permutationally invariant PESs for H_5^+,[75] the water dimer,[76,77] and the protonated water dimer[78] (the Zundel cation), using CCSD(T)/aug-cc-pVTZ energies. In these cases numerous low-lying saddle points and associated harmonic frequencies were very accurately described by the PESs. So, even without gradient data precise fits to electronic energies can produced highly accurate Hessians, at least at stationary points, where tests have been carried out.

Even though the datasets we have used for fits are very small relative to the number of energies in direct-product grids, 10^5 high-level electronic energies are clearly expensive to obtain. Such large datasets of energies can be required for very high-dimensional PESs, especially ones with numerous stationary points, e.g., PESs for CH_3CHO,[71,79] CH_3NO_2,[80] and CH_3CHOO.[81]

Recently, we examined the performance of PIP as well as GPR using PIP inputs for a variety of dataset sizes for four case studies,H_3O^+, $OCHCO^+$, H_2CO including *cis* and *trans*-HCOH, and formic acid dimer.[65] One major goal of that work was to see how small the datasets could be without sacrificing quantitative accuracy compared to direct *ab initio* results and results from fitting larger datasets for properties such as barrier heights and normal mode frequencies. We noted that both methods produce acceptable results on large test datasets using training dataset of 1,000 energies for H_3O^+ and 1,500 for $OCHCO^+$. While these are small datasets to train on, compared to the size of the datasets used in the original fits, a reasonable question to ask is can these be significantly further reduced if gradient data were used. We investigate that question in detail here by extending our software to perform PIP fits using energies to include gradient data as well. The extended software described here is available at https://github.com/szquchen/MSA-2.0.

We note that energies and gradients have been used in other non-parametric methods,

such as NN, reviewed in ref. 16, and GPR, reviewed in ref. 22. The applications have mainly been to atomic solids and atomic clusters, using DFT calculations of energies and gradients, which are analytical. Much less work using gradients has been done for isolated molecules. Neural network fits to electronic energies and gradients have been reported for the H + HBr reaction,[68] H_2O, O_3, and ClOOCl.[82,83] For semi-global PESs of H_2O and ClOOCl 282 and 1,693 configurations were used, respectively, and for the global PES of O_3 2,815 configurations were used in the fits to energies and gradients at the efficient MP2 level of theory where gradients are also analytical. Here "global" refers to a PESs that describe at least one dissociation channel.

5.3 Theory and software

5.3.1 Theory

The theory of permutationally invariant polynomials has been presented in several review articles,[26,28,29,65] and so we give a brief review of the essential details. In terms of a PIP basis, the potential energy, V, can be written as

$$V = \sum_{i=1}^{M} c_i p_i(\boldsymbol{y}), \tag{5.1}$$

where p_i are permutationally invariant polynomials of internuclear distances transformed to Morse variables, denoted $\boldsymbol{y}$, and c_i are linear coefficients which are determined using standard linear least-squares fitting, which up to now have been fits to electronic energies. This summation is up to a certain total polynomial order, M.

The invariant polynomials can be obtained by using monomial symmetrization.[26,65] The procedure for n atoms is to begin with a single monomial of the $n(n-1)/2$ Morse

variables, $y_{ij} = \exp(-r_{ij}/\lambda)$, where r_{ij} is the inter-nuclear distance between atoms i and j and λ is the range parameter. Then by applying the permutations of like atoms these variables map into another set of Morse variables and thus another monomial, but with the same total polynomial order. A permutationally invariant polynomial is generated by summing over the monomials.[26]

Consider the A_2B molecule as a simple example; atom A is labeled as 1 and 2, and B is labeled as 3. There is only one permutation and this permutation maps the monomial $y_{12}^a y_{13}^b y_{23}^c$ to the monomial $y_{12}^a y_{23}^b y_{13}^c$, and thus the invariant polynomial is the sum of the two monomials $y_{12}^a y_{13}^b y_{23}^c + y_{12}^a y_{23}^b y_{13}^c$. Other examples were given in detail in ref. 26. This approach is mainly pedagogical rather than practical except for very simple cases, e.g., PESs for H_2CO[84,85] and H_2CN.[86] Clearly, a more efficient approach would be based on representing these symmetrized monomials in a more compact way. This is done in two ways. One is based on monomial symmetrization but builds up the symmetrized polynomials from lower-order ones recursively.[27] This is called the monomial symmetrization approach (MSA). The second approach makes use of the invariant polynomial theory[36] and computational algebra software MAGMA[37]. Based on this theory, the invariant polynomial can be efficiently factorized into polynomials of primary invariants times secondary invariants. A library of such primary and secondary polynomials for many molecule types is available upon request.[26] For the present application, which is to incorporate gradient data into the fitting, it is easier to modify the MSA software.

We briefly describe the MSA algorithm and how the gradients can be computed analytically. In MSA, all the possible monomials with total degree less than some maximum value are enumerated then symmetrized according to the permutation group of the molecule, written for example as A_2B for a triatomic molecule with 2 like atoms, "A", and a single atom "B". Then the polynomials are decomposed into the product of

Table 5.1: Monomials (m_i), invariant polynomials (p_i), and their partial derivatives dm_j and dp_j for A_2B up to degree 3.

Monomials / Polynomials	Partial Derivative (with respect to x_i)
$m_0 = 1.0$	$dm_0 = 0.0$
$m_1 = y_{23}$	$dm_1 = \partial y_{23}/\partial x_i$
$m_2 = y_{13}$	$dm_2 = \partial y_{13}/\partial x_i$
$m_3 = y_{12}$	$dm_3 = \partial y_{12}/\partial x_i$
$m_4 = m_1 \times m_2$	$dm_4 = dm_1 \times m_2 + m_1 \times dm_2$
$p_0 = m_0$	$dp_0 = dm_0$
$p_1 = m_1 + m_2$	$dp_1 = dm_1 + dm_2$
$p_2 = m_3$	$dp_2 = dm_3$
$p_3 = m_4$	$dp_3 = dm_4$
$p_4 = p_1 \times p_2$	$dp_4 = dp_1 \times p_2 + p_1 \times dp_2$
$p_5 = p_1 \times p_1 - p_3 - p_3$	$dp_5 = dp_1 \times p_1 + p_1 \times dp_1 - dp_3 - dp_3$
$p_6 = p_2 \times p_2$	$dp_6 = dp_2 \times p_2 + p_2 \times dp_2$
$p_7 = p_2 \times p_3$	$dp_7 = dp_2 \times p_3 + p_2 \times dp_3$
$p_8 = p_3 \times p_1$	$dp_8 = dp_3 \times p_1 + p_3 \times dp_1$
$p_9 = p_2 \times p_5$	$dp_9 = dp_2 \times p_5 + p_2 \times dp_5$
$p_{10} = p_2 \times p_4$	$dp_{10} = dp_2 \times p_4 + p_2 \times dp_4$
$p_{11} = p_1 \times p_5 - p_8$	$dp_{11} = dp_1 \times p_5 + p_1 \times dp_5 - dp_8$
$p_{12} = p_2 \times p_6$	$dp_{12} = dp_2 \times p_6 + p_2 \times dp_6$

two polynomials of lower order, and possibly subtracting a small number of same-order polynomials, when this is possible. If this factorization is not possible, the polynomial is kept as the sum of monomials.[27]

Continuing with A_2B as an example, the monomials and invariant polynomials up to degree of 3 are listed in Table 5.1. As shown there, the first monomial is the constant term 1.0, and the following $n(n-1)/2$ monomials are just the Morse variables. The remaining monomials and then the polynomials can be obtained from previously computed ones. The partial derivatives of the first $n(n-1)/2+1$ monomials (in this example m_0 – m_3) with respect to the Cartesian coordinates can be obtained analytically, and the

derivatives of the following terms are computed based on previous ones, as shown in the second column of the table.

Software has been written to obtain the PIP basis and all components of the gradient and this is described below. However, first we describe how the standard least-squares optimization is done using energy and gradient data. Suppose there are M linear coefficients and the database contains N energies and $3N$ components of the gradient vector obtained at N configurations. For a molecule of n atoms, the linear least-squares equations are

$$\begin{bmatrix} p_1(1) & p_2(1) & p_3(1) & \cdots & p_M(1) \\ p_1(2) & p_2(2) & p_3(2) & \cdots & p_M(2) \\ \vdots & \vdots & & \vdots & \\ p_1(N) & p_2(N) & p_3(N) & \cdots & p_M(N) \\ \frac{\partial p_1(1)}{\partial x_1} & \frac{\partial p_2(1)}{\partial x_1} & \frac{\partial p_3(1)}{\partial x_1} & \cdots & \frac{\partial p_M(1)}{\partial x_1} \\ \vdots & \vdots & & \vdots & \\ \frac{\partial p_1(1)}{\partial x_{3n}} & \frac{\partial p_2(1)}{\partial x_{3n}} & \frac{\partial p_3(1)}{\partial x_{3n}} & \cdots & \frac{\partial p_M(1)}{\partial x_{3n}} \\ \vdots & \vdots & & \vdots & \\ \frac{\partial p_1(N)}{\partial x_1} & \frac{\partial p_2(N)}{\partial x_1} & \frac{\partial p_3(N)}{\partial x_1} & \cdots & \frac{\partial p_M(N)}{\partial x_1} \\ \vdots & \vdots & & \vdots & \\ \frac{\partial p_1(N)}{\partial x_{3n}} & \frac{\partial p_2(N)}{\partial x_{3n}} & \frac{\partial p_3(N)}{\partial x_{3n}} & \cdots & \frac{\partial p_M(N)}{\partial x_{3n}} \end{bmatrix} \begin{bmatrix} c_1 \\ \\ \\ c_2 \\ \\ \\ c_3 \\ \\ \vdots \\ \\ \\ c_M \end{bmatrix} = \begin{bmatrix} V(1) \\ V(2) \\ \vdots \\ V(N) \\ g_1(1) \\ \vdots \\ g_{3n}(1) \\ \vdots \\ g_1(N) \\ \vdots \\ g_{3n}(N) \end{bmatrix}, \tag{5.2}$$

where "(i)" denotes the i-th data point in the database, $[x_1, \cdots, x_{3n}]$ are the Cartesian coordinates, and $[g_1, \cdots, g_{3n}]$ are the components of gradient. The size of the matrix is $(N + 3nN) \times M$. The linear coefficients are obtained using standard least-squares methods.

5.3.2 MSA software

Several steps are involved in generating a PES using the PIPs and gradients with a given total polynomial order for a given molecule type and with input consisting of energies or energies plus gradients. These steps are illustrated in the flow chart shown in Figure 5.1. The chart indicates an iterative procedure based on "holes", which are unphysical, generally very negative, regions of a PES. This is not done here, as one goal is to assess the extent of such regions on PESs based on only energies and those based on energies plus gradients. It is perhaps worth noting that in general non-parametric methods to obtain PESs are not reliable in regions "far" from the training data. From our extensive experience with PIPs, these regions are where the true potential is very repulsive, i.e., of the order of 0.5 Hartree. This always occurs for small values of internuclear distances, and such regions are typically not extensively sampled. We will examine this in some detail below.

A Python script has been written that wraps a C++ code that generates the PIPs and gradients for a given molecule type, and a Perl script that creates a FORTRAN code for the PIPs and gradients and does the least-squares fitting. The user is queried to specify the maximum polynomial order, the permutational symmetry, and the file that contains *ab initio* energies (and gradients). The user also has the option to specify the range parameter λ in the Morse variables and add weights to each data point, or use the default values. The least-squares linear algebra is done with the freely available FORTRAN code DGELSS. As noted already the complete software is available at https://github.com/szquchen/MSA-2.0 and the necessary LAPACK routines are available at http://www.netlib.org/lapack/lapack-3.1.1/html/.

In the next section we present a demonstration of the new software to develop a PES

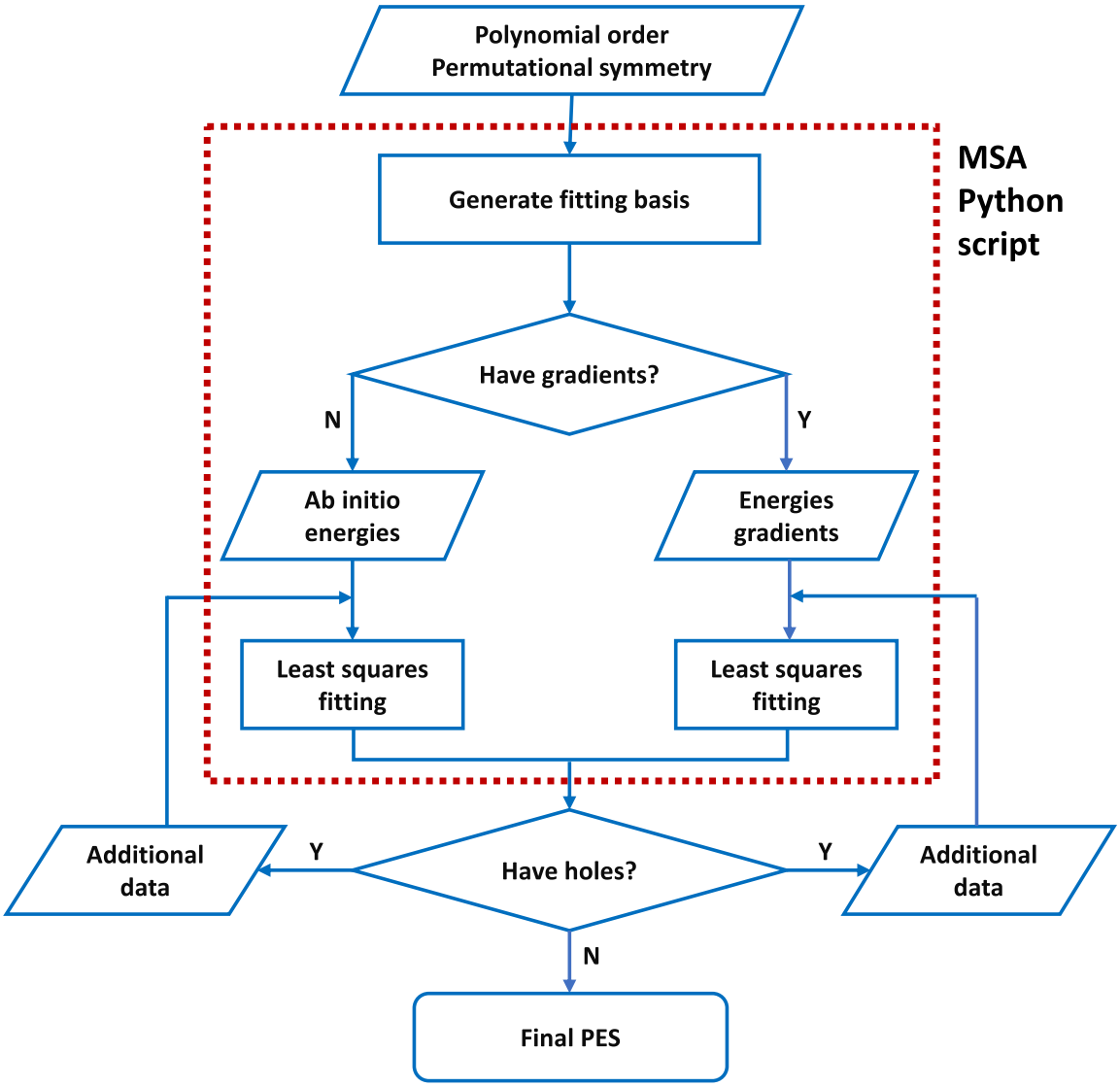


Figure 5.1: Flow chart of the PES fitting procedures. The procedures in the red rectangle are now integrated in one single Python script.

for methane with several goals in mind. Namely, we examine fits using just energies and fits using both energies and gradients. These are for a number of datasets of various sizes. In addition to reporting root-mean-square (RMS) fitting errors for the energies and gradients, we report normal mode analyses and also diffusion Monte Carlo calculations of the zero-point energy. First, we give a simple 1d fitting as a pedagogical example.

5.4 Applications

5.4.1 1d potential

As a simple example, consider fitting the the 1d Morse potential

$$V(r) = D[1 - \exp(-x/a)]^2,\ a = 2 \text{ bohr}, D = 4 \text{ eV} \tag{5.3}$$

using the Morse variable, $y = \exp(-x/3)$ with energies and derivative data. In this example the gradient data is equal in size to the energy data, but it does illustrate the advantage of using the gradient. We consider 8 training configurations, $x = -2$, -1, 1, 4, 5, 7, 10 and 15. These cover the region of the repulsive wall as well as the asymptote. Note, the minimum, $x = 0$, is not included in the training data set. We performed least-squares fits (no weighting of data) using energies and energies and derivatives with a maximum polynomial order of 4 (and so 5 linear coefficients). A higher polynomial order could be used for the fit with derivatives as there are more data in that fit, however, we do not do that. The RMS errors for the energies and the derivatives for a test set of 440 configurations are 0.0246 eV and 0.0292 eV/bohr, respectively for the energy-only fit (E-fit), and 0.0253 eV and 0.0142 eV/bohr, respectively for the energy and derivative fit (EG-fit). The RMS fitting error for the energy is about the same in the two fits and the

magnitude is in line with our expectations, based on the large energy range, 0 to roughly 12 eV, for the test data. Also, as expected the RMS error for the gradient is significantly smaller (about a factor of 2) in the EG-fit.

Figure 5.2 shows plots of the two fits as well as the Morse potential in the repulsive, minimum and long-range regions. As seen, and perhaps as expected, the fit including the derivatives is more precise than the fit with just energies, although both fits are quite precise. Indeed a plot of the two fits over the range -2.0 to 20.0 would appear to be virtually identical to the target Morse potential. The E- and EG-fit give a value of the

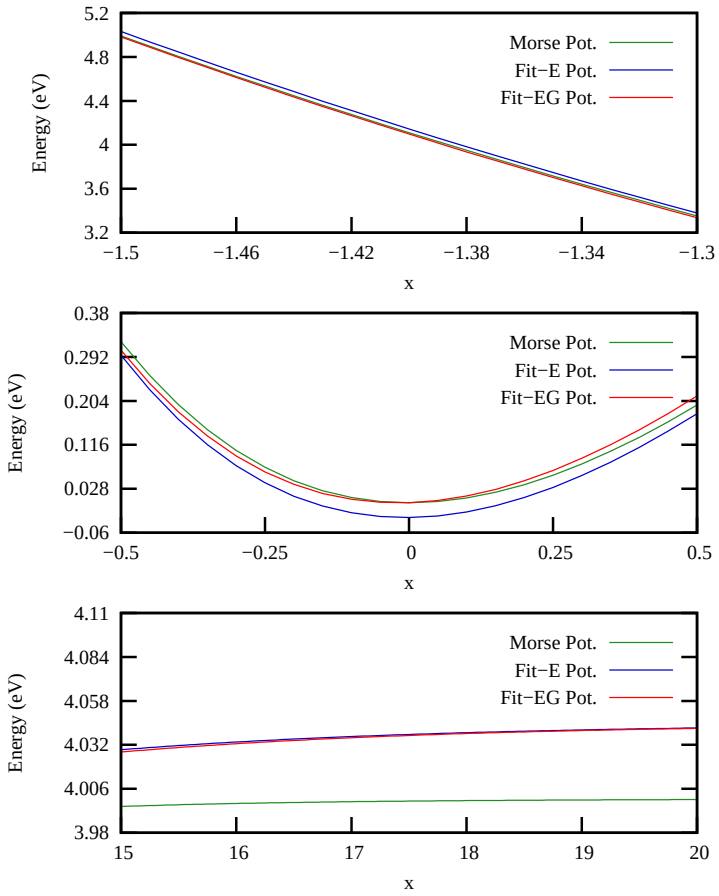


Figure 5.2: Comparison of Morse potential plot between energy-only fit (E-fit) and energy and gradient fit (EG-Fit) for the repulsive, minimum and long-range regions.

potential at the minimum of -0.0295 and 0.0002 eV, respectively. Both fits extrapolate (at 20 bohr) to 4.04 eV and so, considering the values at the minimum, the electronic dissociation energy from the E- and EG-fit are 4.06 and 4.03 eV, respectively. Clearly, and as expected, the EG-fit is more precise than the E-fit.

Next we present a detailed study of using energies and energies and gradients for a full-dimensional PES for methane.

5.4.2 CH_4

Methane is a five-atom, highly symmetric molecule and for each energy there are 15 gradient components and thus this is a realistic test case. The dataset of energies and gradients was generated from *ab initio* molecular dynamics (AIMD) simulations using B3LYP/6-31+g(d) level of theory. Three trajectories were calculated with different total energies to obtain a wide coverage of the configuration space. These trajectories were propagated for 30,000 time steps with the step size of 5.0 a.u. (about 0.12 fs) and the total energies in these are 1,000, 10,000, and 15,000 cm^{-1}. Both the energy and gradients were recorded every 10 steps so the full dataset consists of 9,000 configurations. These AIMD simulations were done using the MOLPRO quantum chemistry package.[87]

A number of training datasets which are subsets of the full dataset were used to examine the new fitting procedure. In all cases we selected an equal number of configurations from each trajectory. Thus, the energy range of each subset is roughly the same, i.e., 0 to 15,000 cm^{-1}. In the calculations reported in the next section nine datasets were used.

A histogram plot of the distribution of electronic energies from these three AIMD trajectories and three example subsets are shown in Figure 5.3. The sharp peak at low energies is just a result of the overlap of these energies from the three trajectories. Clearly,

there are other ways to obtain datasets, however, we find that the results shown below validate the distributions shown in this figure.

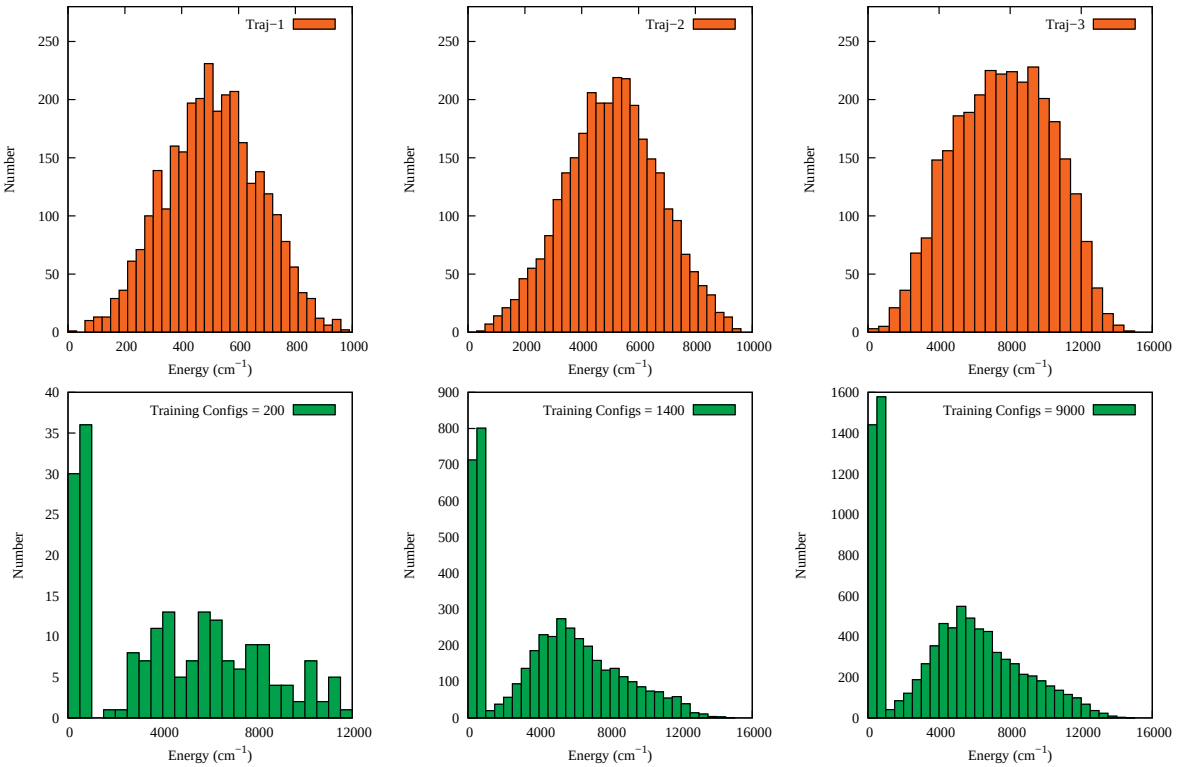


Figure 5.3: Distributions of electronic energies from three trajectories at indicated total energies and in green the actual datasets used for the PES fits.

For each dataset, two fits were done: one is a fit to energies only and the other is a fit to energies plus gradients. These two classes of PESs are denoted as PES-E and PES-EG. For all the PES-EG fits the maximum polynomial order is 5 and this results in 208 polynomials and thus 208 linear coefficients. This maximum polynomial order is also used for all PES-E fits except for the datasets of 300 and 100. In those cases the maximum polynomial order is 4, and there are 83 coefficients in this case. The largest

dataset we used consists of 9,000 energies and 135,000 components of gradients. Clearly, fits with energies and gradients could be done with a higher maximum polynomial order than with energies only. However, for the present purpose, which is a direct assessment of including gradient data for a given fitting basis, we kept the same polynomial order.

Diffusion Monte Carlo (DMC)[55,58,59] calculations and normal mode analyses were performed with these different fits to analyze the fidelity of the PESs and how the two fitting procedures differ with respect to the number of training configurations. The standard unbiased algorithm[58] is used for the DMC calculations. In each DMC "trajectory", 10,000 walkers were propagated for 25,000 time steps with the step size 5.0 a.u.; 10,000 steps were used to equilibrate the walkers, and the reference energies in the remaining 15,000 steps were used to compute the zero-point energy (ZPE).

5.5 Results and Discussion

5.5.1 Training and testing precision

First, consider the performance of the fits using training datasets of 9,000, 4,500, 3,000, 2,000, 1,500, 1,000, 600, 300 and 100 configurations. The RMS error of the energy and gradient magnitude for these datasets are given in Table 5.2.

As seen, the energy and gradient RMS errors are fairly similar and constant for datasets larger than 1,000 configurations. The EG PESs are more precise for the gradients than the E PESs and the converse holds for the energies. So, there seems to be no advantage in using gradient data for these larger datasets. The situation changes significantly for datasets of 1,000 and smaller number of configurations for the gradient precision, whereas the energy RMS error is only marginally different, i.e., changes of

Table 5.2: RMS fitting error of energies (RMS_E) (cm^{-1}) and gradient magnitude (RMS_G) (cm^{-1}/bohr) for PES-E and PES-EG with the indicated number of training configurations (N_{train}).

	PES-E [a]		PES-EG	
N_{train}	RMS_E	RMS_G	RMS_E	RMS_G
9000	2.84	17.36	3.88	14.20
4500	2.15	15.76	3.19	11.84
3000	2.17	17.84	3.62	12.69
2000	1.63	17.25	3.25	11.64
1500	1.58	18.22	3.25	11.73
1000	1.47	21.03	3.01	11.24
600	0.97	24.21	2.55	9.97
300	4.22	78.79	2.28	9.06
100	0.14	458.17	1.47	5.30

[a] Maximum polynomial order of 4 is used for $N_{train} = 100$ and 300; all the remaining fits including the ones with gradients use polynomial order 5.

roughly 2 cm^{-1}. For the datasets of 100 and 300 configurations, where the maximum polynomial order is 4 (and 83 coefficients), the gradient RMS error is 10 and 100 times larger for PES-E compared to PES-EG. Note with 100 energies the E-fit with 83 coefficients is an overfit and that causes the large error in the gradient magnitude. The EG-fit, however, gives excellent precision for the energies and the gradients.

Another important test is to examine the performance of PES-E and PES-EG on test datasets. For this test the datasets in Table 5.2 are used, except for the largest one of 9,000 and the test data consists of the remaining data from the 9,000 dataset. The results are given in Table 5.3. As seen in all cases, the testing RMS error of both energies and gradients is smaller in PES-EG than PES-E. The difference in these RMS errors increases significantly with a training set of 1,000 and reaches a very large difference for 100 and 300. Thus, it is clear that there is a major benefit to including gradients in small training

set of configurations.

Table 5.3: Test RMS error of energies (RMS_E) (cm^{-1}) and gradients (RMS_G) (cm^{-1}/bohr) for PES-E and PES-EG with the indicated number of training configurations (N_{train}) .

		PES-E [a]		PES-EG	
N_{train}	N_{test}	RMS_E	RMS_G	RMS_E	RMS_G
4500	4500	4.78	24.48	4.67	19.45
3000	6000	4.61	24.35	4.34	17.97
2000	7000	6.35	28.55	4.46	18.04
1500	7500	6.61	28.92	4.46	17.77
1000	8000	8.37	37.65	4.29	17.59
600	8400	16.35	72.35	4.68	19.67
300	8700	45.66	168.97	5.15	22.75
100	8900	208.00	913.72	8.88	39.84

[a] Maximum polynomial order of 4 is used for $N_{train} = 100$ and 300; all the remaining fits use polynomial order 5.

In order to make a more detailed examination of the errors in gradients, we calculated the cosine of the angle between *ab initio* gradient vector and PES-fitted gradient vector and also the ratio of the magnitudes of these two vectors. The $\cos\theta$ values and the ratios of magnitudes for the test dataset are mostly concentrated at 1.0 for the training dataset of 4,500 in both PES-E and PES-EG. However, for the training datasets of 100 and 300 the $\cos\theta$ and gradient vector magnitude ratio values are widely spread for PES-E relative to PES-EG. As a representative example, we plot $\cos\theta$ and the ratio of magnitudes for 8,700 test datasets using fits trained on 300 configurations in Figure 5.4. As seen, the PES-EG is far more precise than PES-E. This is not a surprising result, of course.

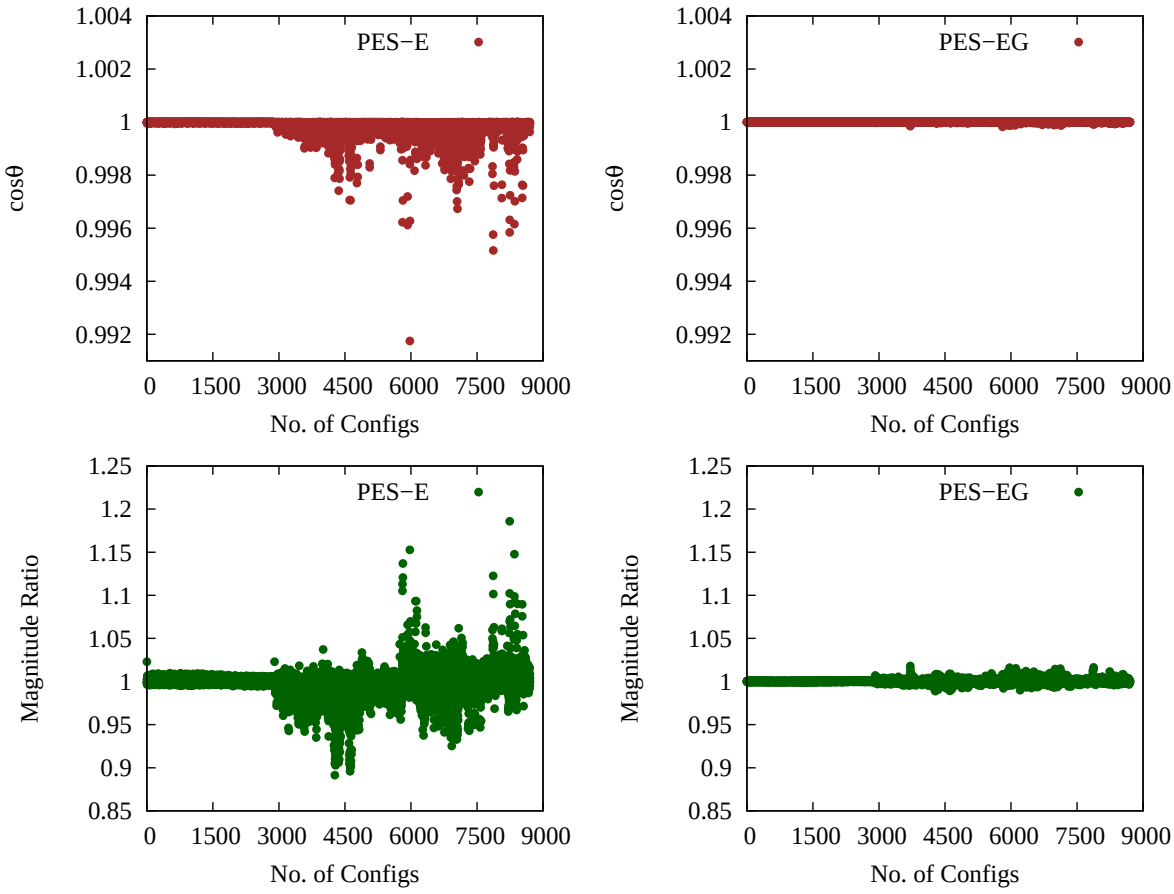


Figure 5.4: Plots of $\cos\theta$, where θ is the angle between *ab initio* gradients and the PES-fitted gradients and plot of vector magnitude ratio between *ab initio* gradients and the PES-fitted gradients for 300 training geometries to fit the PESs.

5.5.2 Normal mode analyses

To further examine the accuracy of the various fits, we performed normal mode analyses and compare the frequencies with the *ab initio* ones. Methane has two F, one E and one A vibrational modes. So, without permutational symmetry it would be very challenging to reproduce these degenerate modes from PESs. However, by using the full symmetry group of order 4!, every PES describes these degenerate modes exactly.

Comparisons are given for PES-E and PES-EG in Tables 5.4 and 5.5. First, consider relatively large training datasets, shown in Table 5.4. Both PES-E and PES-EG produce very precise results, with PES-EG slightly more precise. Thus, there is no significant advantage here in using gradient data in these fits. However, for smaller datasets the results in Table 5.5 show a significant advantage in using gradient data. Even for the case of just 50 configurations (with energies that still span the energy range shown in Figure 5.3) PES-EG produces results that are within 0–5 cm^{-1} of the *ab initio* ones. As noted in the table footnote 50 energies are insufficient to obtain a PES-E.

Table 5.4: Comparison of harmonic frequencies (cm^{-1}) of CH_4 between PES-E and PES-EG with the corresponding *ab initio* ones for indicated training sets.

		Training Geom=600		Training Geom=2000		Training Geom=4500	
Mode	*ab initio*	PES-E	PES-EG	PES-E	PES-EG	PES-E	PES-EG
1	1365	1367	1366	1367	1366	1367	1366
2	1365	1367	1366	1367	1366	1367	1366
3	1365	1367	1366	1367	1366	1367	1366
4	1581	1583	1580	1582	1580	1582	1580
5	1581	1583	1580	1582	1580	1582	1580
6	3042	3039	3042	3044	3042	3044	3042
7	3147	3145	3147	3147	3146	3147	3147
8	3147	3145	3147	3147	3146	3147	3147
9	3147	3145	3147	3147	3146	3147	3147

Table 5.5: Comparison of harmonic frequencies (cm^{-1}) of CH_4 between PES-E and PES-EG with the corresponding *ab initio* ones for indicated training sets

.

		Training Geom=50		Training Geom=100		Training Geom=300	
Mode	*ab initio*	PES-E [a]	PES-EG	PES-E [b]	PES-EG	PES-E [b]	PES-EG
1	1365	–	1367	1369	1366	1368	1365
2	1365	–	1367	1369	1366	1368	1365
3	1365	–	1367	1369	1366	1368	1365
4	1581	–	1582	1590	1581	1583	1581
5	1581	–	1582	1590	1581	1583	1581
6	3042	–	3037	2929	3042	3032	3042
7	3147	–	3147	3159	3146	3152	3147
8	3147	–	3147	3159	3146	3152	3147
9	3147	–	3147	3159	3146	3152	3147

[a] This fit was not done because the number of training data is too small;
[b] These use maximum polynomial order of 4, PES-EG uses maximum polynomial order 5.

These results on the normal mode frequencies continue to demonstrate the power of energy-based PESs in obtaining quantitatively accurate harmonic frequencies, provided the datasets are sufficiently large. The new results show the great advantage in also using gradient data for small datasets of configurations. In all cases, it should be noted that energies and gradient data at the equilibrium structure is not contained in the datasets.

To explore more regions of the configuration space we present results of diffusion Monte Carlo calculations of the zero point energy next.

5.5.3 Diffusion Monte Carlo Zero Point Energy

Diffusion Monte Carlo (DMC) calculations are a good probe of the quality of a PES in extended regions of the configuration space. One common issue with mathematical fitting of energies is unphysical behavior in regions without data. This can be dramati-

cally manifested by large negative values. These are referred to as holes. Generally, we have observed that these occur for highly repulsive configurations, i.e., short internuclear distances. Adding some data in these regions plus a re-fit generally eliminates these holes.

Here the objective is to examine the hole issue with respect to E and EG fitting. These holes are discovered in the course of a DMC trajectory and we have flagged their number for training sets of 100–4,500 configurations. In DMC calculations holes are benign, and can be easily dealt with provided they are in regions where the actual potential is very high. In unbiased DMC (which we employ) walkers are "killed" if they enter regions of large potential. This procedure is followed if a walker encounters a hole. Thus, for all PESs we do report a ZPE along with the number of holes encountered. The ZPE is an average over five independent DMC trajectories and the uncertainty in the ZPE is the standard deviation from the five independent calculations.

Results of these calculations for PES-E and PES-EG are given in Table 5.6. Consider PES-E first. As seen, the number of holes ranges from several thousand to almost half a million. These appear to be large values; however, they need to be considered relative to the number of configurations sampled in the five DMC trajectories. This number is of order 10^9 and so the number of holes is a very small fraction of the configurations sampled.

Considering now the result for the PES-EGs, we observe a dramatic reduction in the number of holes. Even for 300 and 100 training configurations, PES-EG generates only 307 and 6,369 holes and the ZPE values are good in accord with values for larger datasets. Thus, the rigorous ZPE of CH_4 can be obtained even with just 100 configuration datasets with energy plus gradient data. As an aside, it is worth noting that the ZPE of 9730 cm^{-1} for these DFT-based PESs is close to the exact quantum value of 9702 cm^{-1}[88] using an *ab initio*-based PES[89] of near spectroscopic accuracy.

Table 5.6: Diffusion Monte Carlo zero-point energy (cm^{-1}) of CH_4 and the number of holes (N_{hole}) for PES-E and PES-EG with the indicated number of training configurations (N_{train}).

	PES-E [a]		PES-EG	
N_{train}	N_{hole}	ZPE	N_{hole}	ZPE
4500	3529	9732 ± 4.5	259	9729 ± 2.5
3000	18069	9732 ± 2.3	1042	9730 ± 2.8
1000	17544	9727 ± 2.9	1147	9729 ± 3.5
600	2666	9728 ± 2.1	19	9732 ± 2.1
300	112481	9756 ± 2.0	307	9730 ± 1.9
100	467584	9793 ± 3.6	6369	9730 ± 2.4

[a] Maximum polynomial order of 4 is used for $N_{train} = 100$ and 300; all the remaining fits use polynomial order 5.

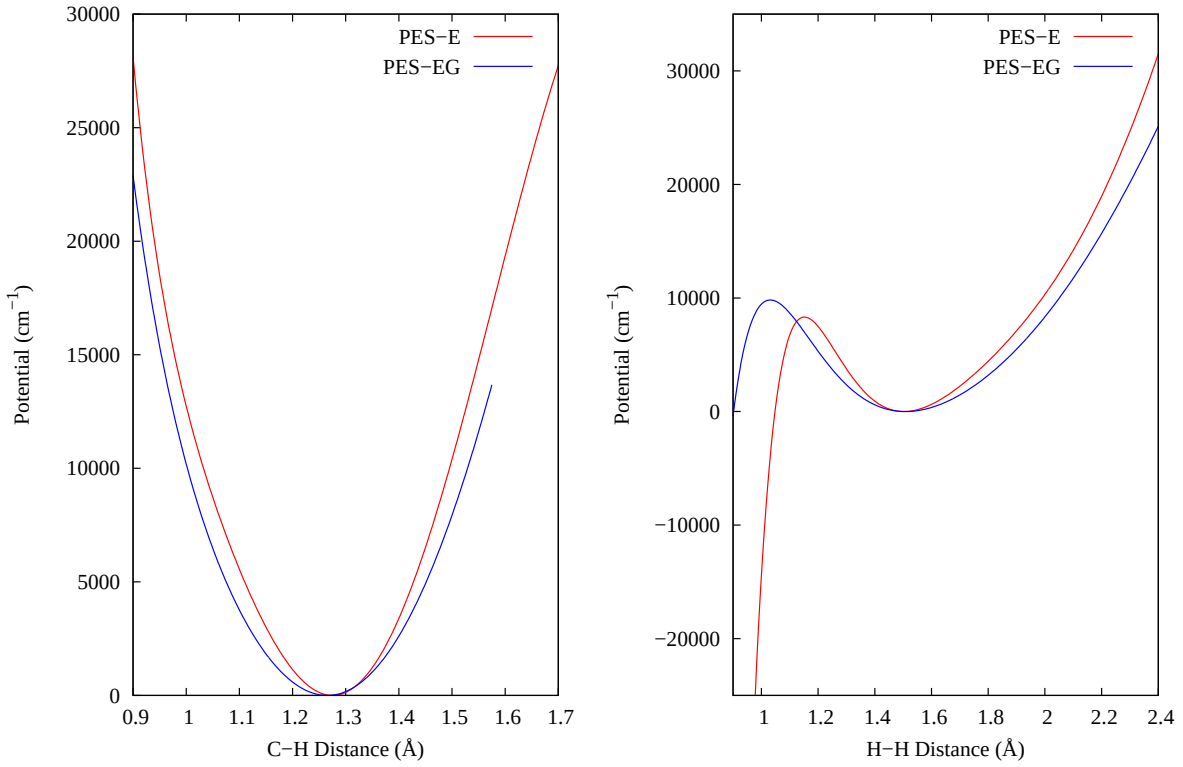


Figure 5.5: Plots of 1D cuts for distances indicated and for PES-EG and PES-E with 100 configurations.

To examine the effectiveness of using gradients in more detail, we show in Figure 5.5 1D cuts of the potential at the minimum in the C−H and the H−H internuclear distances for PES-E and PES-EG using 100 configurations. As seen, there is a "hole" in the latter cut at small H−H distances. However, the local maximum in the PES-EG cut occurs at smaller H−H distances and also at a significantly higher potential value than the one seen in the PES-E cut.

5.5.4 PES-EG from AIMD direct dynamics

Finally, we examined the fidelity of PES-EGs obtained using energies plus gradients from every 10th step of direct-dynamics AIMD trajectories. These were all initiated at the global minimum. Recall that this is not the procedure we followed to obtain the datasets used in the fits described and tested above. Specifically, we ran AIMD trajectories with the total energy of 10,000, 20,000, and 30,000 cm^{-1} for 1,000 time steps. Thus, energies and gradients at 100 configuration were used to obtain three corresponding PESs, denoted PES-EG-1, PES-EG-2, and PES-EG-3 corresponding to the three total energies. Each was used to perform geometry optimization, normal mode analysis and DMC calculations.

The RMS fitting errors are given in Table 5.7. These are all very small, with the expected growth for increasing total energy of the trajectory. Normal mode frequencies, given in Table 5.8, are in excellent agreement with the *ab initio* ones. As expected, the results from PES-EG-1, which uses data from the lowest energy trajectory is the most precise one.

The ZPEs and the number of holes from DMC calculations are given in Table 5.9. As seen, the ZPEs from PES-EG-2 and PES-EG-3 are the same and in excellent agreement with the ZPE shown in Table 5.6. As expected PES-EG-1 has more holes than PES-EG-2

Table 5.7: RMS fitting error of energies (RMS_E) (cm^{-1}) and gradients (RMS_G) (cm^{-1}/bohr) for PES-EG-1, PES-EG-2, and PES-EG-3.

PES	RMS_E	RMS_G
PES-EG-1	0.23	0.68
PES-EG-2	2.88	7.02
PES-EG-3	7.97	24.14

Table 5.8: Comparison of harmonic frequencies (cm^{-1}) of CH_4 from PES-EG-1, PES-EG-2, and PES-EG-3 with the corresponding *ab initio* ones

Mode	*ab initio*	PES-EG-1	PES-EG-2	PES-EG-3
1	1365	1365	1364	1363
2	1365	1365	1364	1363
3	1365	1365	1364	1363
4	1581	1582	1578	1582
5	1581	1582	1578	1582
6	3042	3042	3041	3040
7	3147	3147	3147	3145
8	3147	3147	3147	3145
9	3147	3147	3147	3145

and PES-EG-3. The reason for this is the small range of configurations sampled by the AIMD trajectory at 10,000 cm^{-1} total energy, which is just slightly higher than the ZPE.

Table 5.9: Diffusion Monte Carlo ZPE (cm^{-1}) and the number of holes (N_{hole}) for PES-EG-1, PES-EG-2, and PES-EG-3.

PES	N_{hole}	ZPE
PES-1	1452	9731 ± 3.4
PES-2	873	9728 ± 0.9
PES-3	364	9728 ± 3.3

As a final test of these three PES-EGs we ran a classical trajectory using each one at the total energy of each corresponding AIMD trajectory (propagated for 1,000 steps) and examined how long it took for the trajectory run on the PES-EG to find a hole. For a 10,000 cm^{-1} trajectory on PES-EG-1 the trajectory ran for millions of time steps without finding a hole. For a trajectory of total energy 20,000 cm^{-1} run on PES-EG-2 a hole was found at 4,000,000 time steps. Finally, the trajectory with total energy 30,000 cm^{-1} and run on PES-EG-3 found a hole at only 1,600 time steps. This general behavior of the trajectories with increasing total energy is understandable because the trajectory is of sampling larger regions of configuration space (actually phase space) with increasing energy whereas the datasets of the PESs are limited to a fixed small number (100) of configurations.

The fidelity of these potentials is excellent and overall we see that fitting with energies plus gradients produces high-fidelity results for this methane example even with configuration datasets as small as 50. Of course, the data being used in the fits including energies plus gradients is much large than using just energies. For methane the ratio of

gradient data to energy data is 15 to 1. In general it is $3n$ to 1, where n is the number of atoms. This large growth in data size is not a major concern for the MSA approach we have described here since in our experience the linear least-squares equations can be solved efficiently (in a few hours) for data sizes as large as several hundred of thousand. For the methane example here, 100 configurations results in a data size of only 1600. For a larger molecule, for example *N*-methyl-acetamide with 12 atoms, this number of configurations would result in a data size of only 3700. Of course more configurations are likely needed for larger molecules.

It is reasonable to ask if the cost of obtaining gradients is "worth it". The answer depends of course on the cost. From the point of view of AIMD simulations (which are ubiquitous), there is no additional cost as gradients are required in these simulations. So, one immediate use of the new software reported here is to simply take the output from an AIMD simulation and create a PES. This was already demonstrated here. At present, essentially all AIMD simulations use density functional theory and this is likely to continue at least for the near future. The absolute accuracy of this theory is in general below that of coupled cluster theory, at least at the level of CCSD(T). This level of theory is typically used for PESs based on energies using PIP, PIP-NN fitting approaches and the resultant PESs are highly accurate and predictive.[34,64,65] Unfortunately, coupled cluster energies and gradients are much expensive to calculate. However, as these calculations increase in efficiency it is likely that these will be used in the future.

5.6 Summary and Conclusions

We presented new software to incorporate electronic energy and gradient data in the representation of a potential energy surface using permutationally invariant polynomi-

als and linear least-squares optimization. The new software was described in detail and demonstrated extensively for methane. PESs were reported using just energies and energies plus gradients using a large range of training datasets spanning an energy range from 0 to 15,000 cm^{-1}. The data were obtained from long-time direct-dynamics calculations using efficient DFT theory (B3LYP). The performance of the PESs in terms of the RMS errors for energies and gradients, normal mode frequencies, and rigorous diffusion Monte Carlo zero point energy, obtained using diffusion Monte Carlo, was examined for both training and testing datasets. A major advantage of using energy plus gradient data was found for small configuration datasets (as small as 100 and 50) where these PESs, denoted PES-EG, were shown to be of essentially equivalent fidelity to those obtained using configuration datasets that are an order of magnitude larger. Unphysical regions of the PESs (holes) sampled by PESs including gradient data was found to be much smaller than with PESs fit to energies only.

Data from 100 configurations obtained from every tenth time step of direct-dynamics calculations run for 1,000 time steps were used to generate PES-EGs at several total energies. These PESs also performed very well, suggesting that the new software can be immediately used to obtain PESs from an direct-dynamics "AIMD" simulation.

Finally, the small configuration datasets that include gradient data for PIP-PES fitting appear well poised for obtaining high-dimensional *ab initio* PESs for molecules much larger than CH_4. Work along these lines is underway and the results will be reported shortly.

Chapter 6

Implementation of Fragmented basis in Permutationally Invariant Polynomial Approach for PES Fitting

6.1 Chapter Abstract

We report full and fragmented potential energy surfaces (PESs) for *N*-methyl acetamide that contain the *cis* and *trans* isomers and the saddle points separating them. The full PES uses Permutationally Invariant Polynomials (PIPs) in reduced symmetry which describe the three-fold symmetry of each methyl rotor. A more efficient PES is an extension of the fragmented PIP approach we reported recently. In this approach the set of Morse variables is partitioned and the fragmented PIP basis is the union of the PIP basis for each set of variables. This approach is general and can be used with Neural Network fits.

The fits are done using roughly 250,000 electronic energies and gradients obtained from direct dynamics, using B3LYP/cc-pVDZ level of theory. The full PIP basis in 66 Morse variables with maximum polynomial order of 3, contains 8,040 linear coefficients. The fragmented PIP basis, also with maximum polynomial of 3, contains 6,121 coefficients. The root-mean-square (RMS) errors of both PESs are roughly 100 cm^{-1} for energies and 15 cm^{-1}/bohr per atom for gradients, for energies up to roughly 45,000 cm^{-1}, relative to the *trans* minimum. Energies and normal mode frequencies of the *cis* and *trans* isomers for the full and fragmented PESs agree well with direct calculations. The energies of the two saddle points separating these minima are precisely given by both PESs. Diffusion Monte Carlo calculations of the zero-point energies of the two isomers are also reported.

6.2 Overview

Developing high-dimensional, *ab initio*-based potential energy surfaces (PESs) remains to be an active area of theoretical and computational research. In the past 15 years, significant progress has been made in the development of non-parametric, machine learning approaches to fit large data set of electronic energies for molecules and clusters that contains more than four atoms.[13–16,19,21–23,25,26,30,34,62–65] These include three popular methods which are Permutationally Invariant Polynomials (PIP)[26,65], Neural Networks (NN),[14–16,19,66–69] NN with PIP inputs,[30,31,34,63,64,70] Gaussian Process regression (GPR),[20,21,25] and GPR with PIP inputs.[35] There is a major motivation to extend these methods to large molecules of interest in chemistry and material science.

Our group has developed the PIP approach over the last 15 years to represent high dimensional PESs of molecules and molecular clusters with numerous applications.[26–28,65,90] This method makes use of Morse variables, which are transformed internuclear distances.

In 2003 this method was first applied to CH_5^+ cation to construct a global PES that is invariant with respect to 120 permutations of the five equivalent H atoms.[74] Generally the PESs make use of data sets of 10^4–10^5 scattered electronic energies, typically obtained at the CCSD(T) level of theory. This approach has been applied to obtain PES for more than 50 molecules, including reactive systems, and molecular clusters.[65] Of particular interest to this paper there are PESs for 7, 8, 9, and 10 atom systems, e.g., CH_3CHO, with many minima and saddle points,[91] CH_3CHOO,[92] malonaldehyde,[93] and formic acid dimer,[72] respectively.

There are bottlenecks for PIP methods as the molecular size increases. Consider first PIP-NN approaches. These use PIPs as the input to NN fits of *ab initio* electronic energies.[34,63] The minimum number needed to get a faithful representation of the molecular permutational symmetry has been discussed in the literature.[34,63,64,94] Zhang and co-workers have advocated the use of a set of PIPs known as fundamental invariants (FIs) to satisfy the above condition. The number of FIs is given by sum of the number of primary invariant polynomials plus the number of irreducible secondary invariant polynomials. The number of primary invariants is equal to the mathematical dimensionality of the PES. If all internuclear distances (or Morse variables) are used, this number is $N(N-1)/2$ and thus is fixed by N. The number of irreducible secondary invariant polynomials varies and depends on the number of atoms and the symmetry group; however, the number can rapidly exceed the number of primary invariants.[63] Zhang and co-workers have reported a number of high-level PIP(FI)-NN PESs for as many as 6 atoms. Guo and co-workers using PIP-NN have also reported a number of high-quality PESs. The largest system they reported is the 7-atom reaction $OH+CH_4$,[34] where 1,331 PIPs were used as input variables[94,95] and the NN fit contained 5,929 parameters. This is clearly a large number of PIPs for the input. Given the rapid increase in the number of PIPs or

FIs for input of NN fits with respect to the molecule size, this number is probably at or close to the practical limit of these approaches.

The PIP approaches we have developed,[26,27,65] and which we briefly review below, have their own practical limits; however, they are reached in a different manner from PIP-NN approaches. The size of the input of our PIP approach (i.e., the number of Morse variables) grows as order N^2, which is much milder than the scaling of the input size of PIP-NN PESs. The size of the PIP basis for the fit increases rapidly with the number of atoms and the total polynomial order and the size of the appropriate symmetry group for the molecule of interest. In general as the symmetry group increases the basis size decreases, as illustrated in detail elsewhere.[26,27] As noted above, the PIP approach has been applied for molecules with as many as 10 atoms. This number of atoms has been cited in the literature as the practical limit for PIP. This limit was extended using the PIP approach for 12-atom *trans*-*N*-methyl acetamide (*trans*-NMA).[96] The major point of that paper, which is preliminary to the present one, was to describe a fragmented PIP approach, with the goal of extending the PIP method to molecules even larger than *N*-methyl acetamide. As an aside, we note that the 10-atom limit was also just exceeded using PIPs in a calculations of of anharmonic rovibrational partition functions including torsional motion.[97]

Before presenting the details of the new work on NMA, we make some remarks on the the atom-centered NN approach due to Behler and Parinello,[15] and now denoted high-dimensional neural network (HD-NN). This is a powerful approach for a large number of atoms, i.e., of order 100. The PIP-NN and HD-NN approaches were recently discussed and compared in a very instructive way.[94] Of the several systems compared, the OH+CH_4 reaction was the largest one. As noted already, this 7-atom system is the largest one for which a PIP-NN PES has been reported. This is actually a small system for HD-NN;

however, the detailed comparison showed clearly the difference in computational effort in the PIP-NN and HD-NN PESs. In brief, the HD-NN method consists of NN fits to atomic energies with like atoms sharing the same NN fits. Thus for $OH+CH_4$ there are 7 NNs. The NN fits for H contains 6,781 parameters and 5,461 for C and O. For the PIP-NN PES there is a single NN fit, with 5,929 parameters. The input for the PIP-NN PES consists of 1,331 PIPs, nevertheless, that PES runs considerably faster than the HD-NN one.

At this point, it seems clear that for molecules with 10 or fewer atoms PIP approaches, including GP ones,[25,35] are probably the ones of choice, whereas for ca 100 or more atoms the HD-NN approaches (including the "message-passing" variant[98,99] developed very recently), as well as atom-centered GP approaches,[20,22] are currently the only general and feasible ones. (A caveat however, is that if the many-body(monomer) expansion of the PES converges rapidly, the PIP approach can and has been used with good success.[65,100])

There is, in our opinion, motivation to address the gap between these limits. We recently reported a new fragmented PIP approach to extend the PIP approach to more than 10 atoms, perhaps tens or more atoms and thus fill this gap.[96] We tested this approach on *trans* isomer of *N*-methyl acetamide, a 12-atom molecule with 66 Morse variables. A full PIP was used as well to provide the benchmark result and so the 10-atom limit was modestly exceeded.

Here we extend this work by reporting full PIP and fragmented PIP PESs for NMA,[96] that describe both the *trans* and *cis* isomers as well as the saddle points separating them. The fidelity of these new PESs is examined by considering energies and harmonic normal mode frequencies. The full PIP PES is used in diffusion Monte Carlo calculations of the zero-point energies of the *trans* and *cis* isomers.

The choice of NMA not only serves as an example of a moderately large molecule

to which the fragmented basis can be applied, it is a molecule that has been intensively studied since it is the smallest molecule in the peptide linkage in polypeptides and proteins. Thus, NMA has been studied both theoretically and experimentally. An important focus of these studies of the NMA molecule is the *cis-trans* isomerization. Recently, NMA have been focused on vibrational and electronic spectroscopy,[101–105] solvent shell structure,[106–108] and *cis-trans* equilibrium.[106–111] It was determined that the stability of *trans*-isomer is greater than the *cis*-isomer and the free energy difference is about 1.5–2.5 kcal/mol. A rough estimation of enthalpy change between these two conformers have been reported as 2.3 kcal/mol from an infrared spectroscopic measurement of NMA in a nitrogen matrix.[112]

Several theoretical investigations have been made of NMA using quantum mechanical methods mostly focused on thermal properties[106–108,110,113] and the vibrational frequencies in gas phase.[101,112,114] Many research groups have reported the *ab initio* calculations on NMA at the HF and/or MP2 level of theory with 4-31G(d), 6-31G(d), 6-31G(d,p), and 6-31++G(d,p) basis sets.[101,106–108,115–117] In addition, the study of conformation and hydrogen bonding effect of NMA and its complexes with water have reported at the DFT level, B3LYP and/or BLYP functionals with 6-311++G(d,p) basis set[110] and the results are consistence with the experimental data or *ab initio* results with larger basis set. In 2001, a detailed analysis of structures, dipole moments and thermodynamic properties of *cis*- and *trans*-NMA at HF, MP2 and DFT (B3LYP) level of theories with various Pople type basis sets was reported.[117] Thus, given this level of interest in NMA our motivation to select NMA to develop full-dimensional and fragmented PESs should be clear. Technically, it also is a challenge owing to the two distant methyl rotors.

6.3 Theory

The theory of permutationally invariant polynomials is well established and has been presented in several review articles,[26–29,65] and so we give a brief review of the essential details. In terms of a PIP basis, the potential energy, V, can be written in compact form as

$$V = \sum_{i=1}^{M} c_i p_i(\boldsymbol{y}), \tag{6.1}$$

where p_i are permutationally invariant polynomials of internuclear distances transformed to Morse variables, denoted $\boldsymbol{y}$, and c_i are linear coefficients which are determined using standard linear least-squares fitting. This summation is up to a certain total polynomial order, M.

The invariant polynomials can be obtained by symmetrizing the monomials.[26,65] The procedure for N atoms is to begin with a single monomial of the $N(N-1)/2$ Morse variables, $y_{ij} = \exp(-r_{ij}/\lambda)$, where r_{ij} is the inter-nuclear distance between atoms i and j and λ is the range parameter. For the trivial example of a triatomic molecule, the expansion in monomials is given by

$$V = \sum_{m=0}^{M} C_{abc} \left[y_{12}^a y_{13}^b y_{23}^c \right] ; \ (m = a + b + c), \tag{6.2}$$

where the summation is over all powers of the y_{ij} subject to the constraint that the total degree m is at most M. Then by applying the permutations of like atoms these variables map into another set of Morse variables and thus another monomial, but with the same total polynomial order. A permutationally invariant polynomial is generated by summing

over the monomials.[26] So, the eq. 13.2 becomes

$$V = \sum_{m=0}^{M} D_{abc} \hat{S} \left[y_{12}^{a} y_{13}^{b} y_{23}^{c} \right], \tag{6.3}$$

where $\hat{S}$ is the operator that symmetrizes the monomials.[26,27]

Consider the A_2B molecule as a simple example; atom A is labeled as 1 and 2, and B is labeled as 3. There is only one permutation and this permutation maps the monomial $y_{12}^{a} y_{13}^{b} y_{23}^{c}$ to the monomial $y_{12}^{a} y_{23}^{b} y_{13}^{c}$, and thus the invariant polynomial is the sum of the two monomials $y_{12}^{a} y_{13}^{b} y_{23}^{c} + y_{12}^{a} y_{23}^{b} y_{13}^{c}$. Other examples were given in detail in ref. 26.

There are two efficient approaches to generate these symmetrized monomials in a compact way. The first approach is based on invariant polynomial theory[36] and associated computational algebra software MAGMA.[37] Based on this theory, the invariant polynomials can be efficiently factorized into polynomials of primary invariants times secondary invariants. A library of such primary and secondary polynomials for many molecule types has been generated.[26] The second approach is based on monomial symmetrization and obtains the symmetrized polynomials from lower-order ones recursively.[27] This is called the monomial symmetrization approach (MSA). Analytical gradients can readily be obtained using the MSA approach and used for both fitting and subsequent evaluations. The latter was in fact done by Li and coworkers in 2016 in the context of trajectory calculations.[118] We recently implemented analytical gradients into the MSA algorithm to fit both energy and gradient simultaneously. This approach is used here and more details can be found elsewhere.[119]

To motivate the fragmented PIP approach, recall that the PIP basis can be generated

by starting with a monomial,[27] given by eq. 6.4

$$y_1^{n_1} y_2^{n_2} \cdots y_l^{n_l}, \quad l = N(N-1)/2, \tag{6.4}$$

where for simplicity the Morse variables are indexed by a single integer. These variables go to zero as the corresponding internuclear distance goes to infinity. This feature was a major advantage of these variables as this permits a realistic description of dissociation. Here this feature has a different advantage. Namely, for large molecules many internuclear distances are large and so the corresponding Morse variables are nearly zero. Thus, any monomial with these variable is nearly zero and can be discarded from the basis. This observation is the major motivation for the fragmented PIP approach we propose for large molecules.

One way to take advantage of this is to start with the full basis and then remove polynomials with very small Morse variables. However, this requires obtaining the full fitting basis first, and this could be prohibitive for large molecules.

Recently, we proposed a way to directly fragment the PIP basis.[96] This is done by partitioning the molecule into fragments, generating the PIP basis of each fragment and then making the the entire PIP basis as the union of these fragment bases. It is worth noting that this approach is not the same as the fragment approach to obtain the energy of large molecules, developed by several groups.[120–122] Here we are generating a single PIP basis for fitting the energies (and gradients) of a large molecule.

A general description of our approach is given below, followed by the details of the application to NMA. We note that a preliminary account of the fragmented approach was reported for the *trans* isomer of NMA.[96] Here the application is to the *cis* and *trans* isomers of NMA as well as the saddle points separating them.

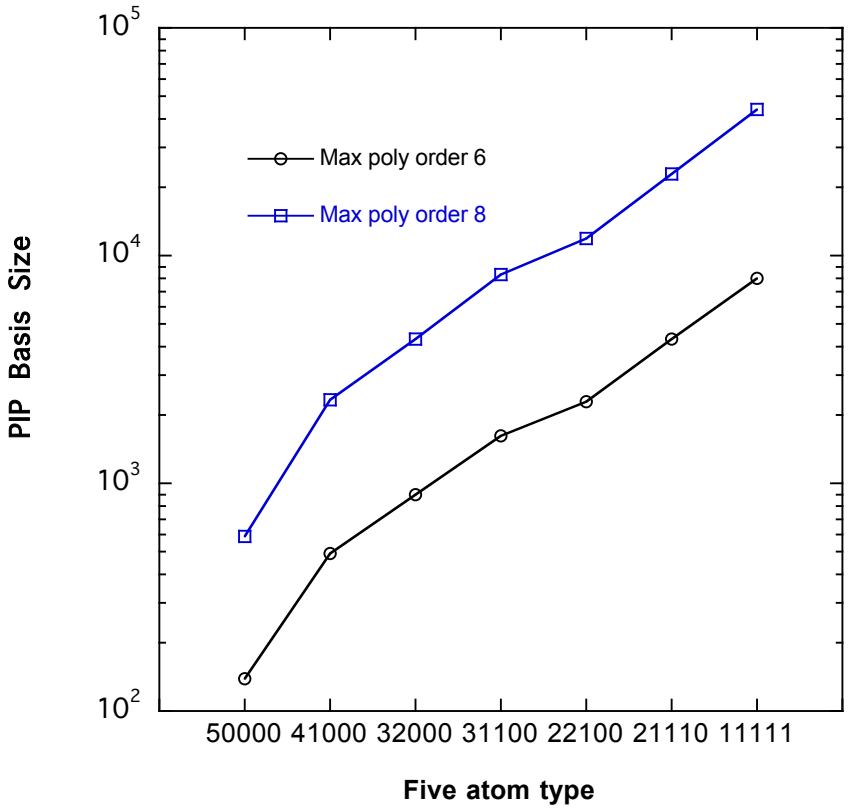


Figure 6.1: PIP basis size for a five-atom molecule for indicated maximum polynomial order spanning molecule type A_5 (50000) to ABDCE (11111).

To begin, consider a five-atom molecule, a small example. Figure 6.1 shows the size of the PIP basis for molecule types A_5, with maximum permutational symmetry, through ABCDE with no symmetry for maximum polynomial order of 6 and 8. Two points are clear: one is the large increase (roughly a factor of 5) in the basis for each molecule type as the maximum polynomial order increases from 6 to 8. The second one is the decrease in the basis size as the order of symmetric group increases, roughly by the size. So for A_5 the basis size is roughly 5! times smaller than the size for ABCDE, etc. However, the complexity of the PIPs increases as the symmetry increases. Without any attempt to factorize PIPs in A_5 consist of many more monomials (5! more in the limit of very high

polynomial order) than the for ABCDE. Thus considering the computational effort to evaluate V in terms of the PIP basis, there is a trade-off between the reduction in the basis size and the complexity of the polynomials. However, the effort to solve the linear least-squares equations to determine the unknown coefficients clearly benefits from the smallest basis. Thus there are several considerations in deciding on whether to use the maximum possible permutational symmetry or a reduced symmetry. Of course the feasibility of the permutation must also be considered. Clearly for CH_5^+ the full permutational symmetry must be used.

The number of variables in the PIP basis is another important parameter that impacts the size of basis. As an example, the basis size for A_2B, A_2BC, and A_2BCD with maximum polynomial order of 6 is 50, 502 and 4,264. The number of variables is 3, 6 and 10, respectively. clearly, there is highly non-linear growth in the size of the basis with respect to the number of variables, which grows like order N^2. This is of central importance when considering larger molecules.

This growth of the PIP basis with respect to the number of Morse variables and the polynomial order motivated us to develop the fragmented PIP approach.[96] The basic idea of the fragmented PIP approach is to partition atoms in a large molecule into fragments, generally with some atoms in common among adjacent fragments. Then PIP bases are determined for each fragment in terms of the corresponding set of Morse variables. Since the number of atoms in any fragment is less than N, the total number of atoms, the size of these PIP bases will be smaller, perhaps much smaller, than the full PIP basis, certainly for the same maximum total polynomial order.

Considering now NMA, there are 66 Morse variables and we already noted that the size of PIP basis using the minimal permutational symmetry using maximum polynomial orders of 3 (8,040) and 4 (95,965).[96] Clearly, polynomial orders of 6 or 8 would result in

completely unfeasible PIP basis sizes. It is perhaps worth noting that in some contexts NMA would not be considered a large molecule; however, it is large enough for our purpose which is to apply and test the fragmentation methodology which we describe next.

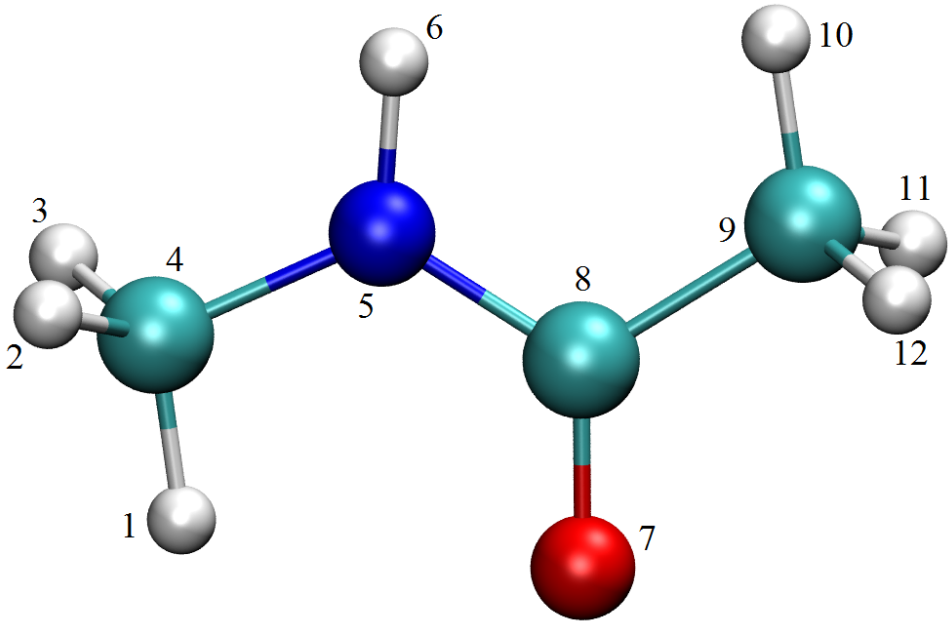


Figure 6.2: Structure of the *trans* *N*-methyl acetamide with atoms labeled as indicated.

To begin, we introduce an atom numbering scheme shown in Figure 13.4, which depicts the structure of *trans*-NMA. As seen, the H atoms on the two methyl rotors (atoms 1–3 and 10–12) are the most distant sets of atoms and indeed these are the ones with very small Morse variables. So, by just omitting these variables results in a simple two fragmentation scheme where one fragment consists of atoms 1–9, and the other consists of atoms 4–12. There are 36 Morse variables in each fragment and there are variables and basis functions in common. Clearly, the PIP basis for each fragment is smaller than the

full PIP basis which contains 66 Morse variables. This fragmentation scheme was used successfully previously with a maximum total polynomial order 3, where only *trans*-NMA was considered.[96]

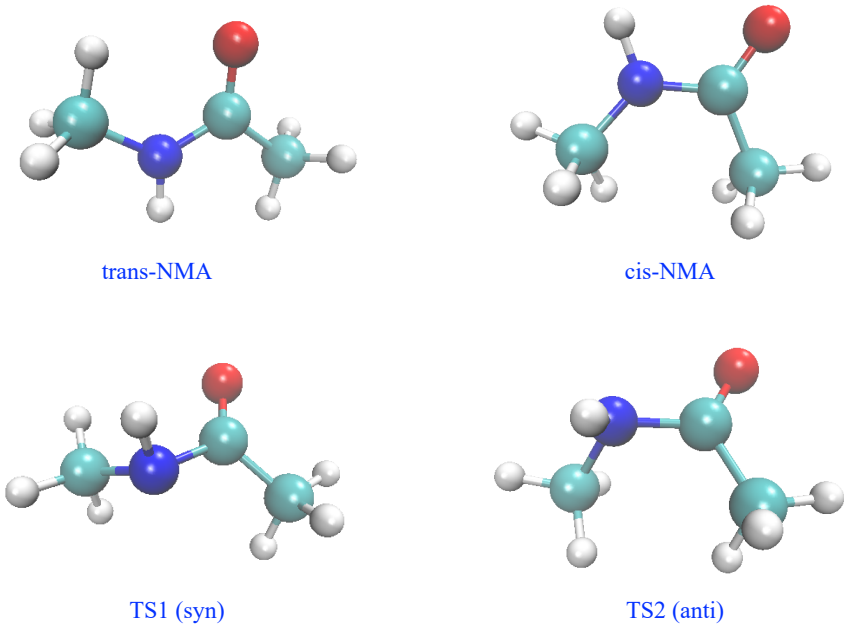


Figure 6.3: Structures of *trans* and *cis* *N*-methyl acetamide and their two isomerization transition states.

During our research to extend the this two-fragmented PES to include the *cis*-NMA isomer, as well as two saddle points (see Fig. 6.3), we found some issues, due to an increased interaction of the distal methyl H atoms 'H' (1–3 and 10–12) atoms, which are closer to each other, especially in the saddle points, than in *trans*-NMA. To account for this, still using the fragmentation method, we added a third 6-atom fragment consisting of the 6 'H' atoms. This fragment of 33 symmetry contains 15 Morse variables and the

corresponding PIP basis is much smaller compared to the 3111111 PIP basis. All results below use this 3-fragment scheme.

The PIP fitting bases of the fragments were generated using the latest MSA software which incorporates gradients.[123] For the first two fragments, 3111111 permutational symmetry is used, while for third fragment 33 permutational symmetry is used. We denote the PIP basis for each fragment as $\{p_i\}$, $\{q_j\}$, and $\{r_k\}$, and the Morse variables in each fragment as $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, and $\boldsymbol{y}_3$. Thus, in this 3-fragment scheme the potential is given by

$$V = \sum_i a_i p_i(\boldsymbol{y}_1) + \sum_j b_j q_j(\boldsymbol{y}_2) + \sum_k c_k r_k(\boldsymbol{y}_3), \tag{6.5}$$

i.e., the fitting basis of the molecule consists all the p, q and r. We note that, as written, this fitting basis has some redundancy. That is, there are some PIPs that are the same in $\{p_i\}$, $\{q_j\}$ and $\{r_k\}$. The number of redundant PIPs is a small fraction of the total PIP fragmented basis given above. They can in principle be removed although with some non-trivial programming effort. Practically, the redundancy is not an issue as the number of redundant PIPs is relatively small, as just noted, and also because the LAPACK software we use (DGELSS) to solve the least-squares problem is based on singular value decomposition, which can deal numerically with this rank-deficient least-squares problem.

Before presenting the computational details, we make some general remarks about the fragmentation approach. First, the above expression for a three-fragment PIP basis is a small (but adequate for the present purposes) example of a general fragmentation. Clearly, for a molecule with many more than 12 atoms more fragments would appear. As long as the number of atoms in each fragment is roughly 10 or less the software we have developed can efficiently generate the PIP basis. Fragments do not have to be adjacent, as illustrated here for the distal methyl groups. Also, it should be clear that the symmetry

and maximum polynomial order for each fragment can be independently chosen.

Finally, we make some remarks about the generality of the fragmented PIP approach. Recall this approach begins by partitioning the Morse variables and then forming PIP basis for each set of Morse variables, and finally forming the union of PIP basis. It appears that a similar strategy can be used with NNs and GPs. For example, for PIP-NNs, the analogue of eq. 6.5, would be to express the potential by

$$V = \mathrm{NN}_1(\boldsymbol{y}_1) + \mathrm{NN}_2(\boldsymbol{y}_2) + \cdots + \mathrm{NN}_i(\boldsymbol{y}_i), \tag{6.6}$$

where in this equation NN_i is a PIP neural network in the variables $\boldsymbol{y}_i$. The optimization of the NN parameters then proceeds in the usual way. Also, the input for each PIP-NN would be the appropriate PIP basis for the variable set, $\boldsymbol{y}_i$. Clearly, this input size could be much smaller than the one for a single PIP-NN fit in all Morse variables. For example, for NMA (recall, with 66 Morse variables) there are 3636 fundamental invariants, using 31111113 symmetry. This is a large number and clearly comes with significant computational overhead to evaluate the PIP-NN PES. Using a fragmented PIP-NN approach the size of the input for each term would be significantly smaller: there are only 134 FIs for the 9-atom fragment (3111111 symmetry) and 118 for the 6-atom fragment (33 symmetry).

6.4 Computational Details

The data sets of energies and gradients were generated from *ab initio* molecular dynamics (AIMD) simulations at several total energies by using microcanonical sampling (NVE). Initial conditions were chosen to obtain wide coverage of the configuration space

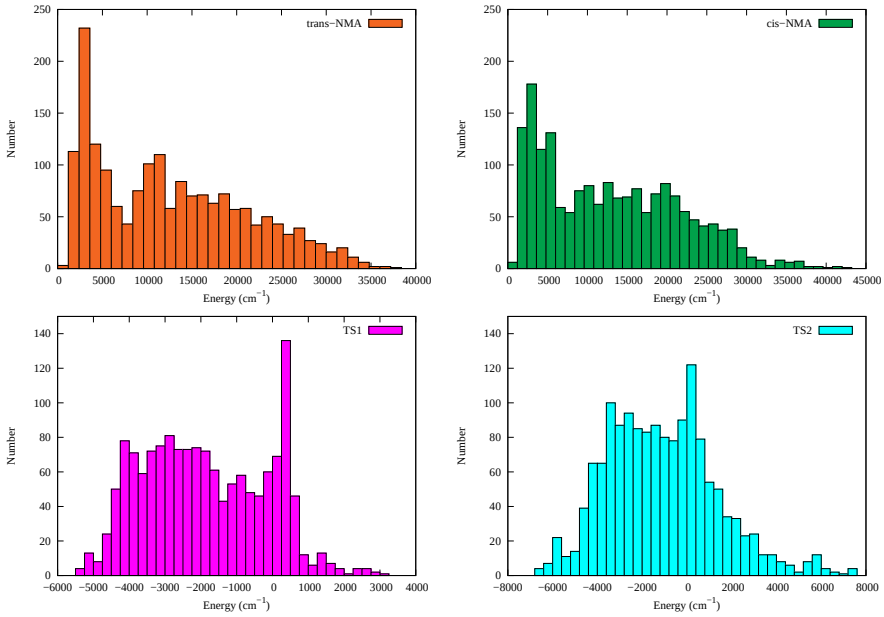


Figure 6.4: Distributions of electronic energies (cm^{-1}) of *trans*- and *cis*-isomer and two isomerization TSs relative to their saddle points energy from their corresponding AIMD trajectories.

for each isomer and the transition states. These AIMD trajectories were propagated for 3,000 time steps with the step size 5.0 a.u. (about 0.12 fs) and with total energies of 1,000, 5,000, 10,000, 20,000, 30,000, 40,000, and 50,000 cm^{-1}. The geometries and their corresponding 36 gradient components were recorded every 10 time steps from each trajectory to generate the fitting data set. These calculations were done at the efficient DFT(B3LYP)/cc-pVDZ level of theory, using Molpro quantum chemistry package.[87] The final data set consists of 6,607 energies and corresponding 237,852 gradients for a total size of 244,459. The distributions of the electronic energies of *trans* and *cis*-isomers and two isomerization TSs are shown in Figure 6.4. As seen, the distribution is very broad; both *trans* and *cis*-isomers cover the energy range up to 40,000–45,000 cm^{-1} and the two TSs cover the energy range up to 10,000–14,000 cm^{-1}, relative to the TS reference configuration energy.

The full PIP basis to fit the PES is generated using MSA software with a reduced permutational symmetry of 31111113 (this describes the identity of the hydrogen atoms within a methyl group which is essential to get the three fold torsional barrier) and a maximum polynomial order of 3. This leads to 8,040 linear coefficients (PIP basis), which are determined by solving a linear least-squares problem. The value of range parameter, $\lambda = 2.0$ bohr, is used during the fit. The RMS errors of this fitted PES are 126.4 cm^{-1} for all energies and 14 cm^{-1}/bohr per atom for gradients respectively.

The fragmented PIP basis is generated using our MSA software with a maximum polynomial order of 3. As noted already, this fragmented scheme contains two 9-atom fragments and a single 6-atom fragment. For the first two 9-atom fragments, the permutational symmetry 3111111 is used, while for the remaining 6-atom fragment, permutational symmetry 33 is used. This leads to 3,028 PIP bases for each of the 9-atom fragment and 65 PIP bases for the 6-atom fragment. This results in a total of 6,121 linear coefficients, which are determined by solving a linear least-squares problem. The RMS errors of this fitted PES are 148.4 cm^{-1} for energies and 15 cm^{-1}/bohr per atom for gradients respectively. These RMS errors are very similar to the full-PIP PES.

Geometry optimization, normal-mode analysis are performed to examine the fidelity of these fitted PESs. This full PES is also applied to calculate rigorous zero-point energies (ZPEs) of both *cis*- and *trans*-NMA using diffusion Monte Carlo (DMC) method. The concept behind DMC is to solve the time-dependent Schrödinger equation in imaginary time.[55,56,124] This is done by simulating a random walk of many replicas, also called "walkers", of the molecule, using a birth/death processes. Each DMC trajectory is propagated for 20,000 time steps using 10,000 random walkers with the step size 5.0 a.u.; 10,000 steps are used to equilibrate the walkers, and the reference energies in the remaining 10,000 steps are used to compute the ZPE. For each isomer, 15 DMC simulations were carried

out, and the final ZPE is the average of the 15 simulations. Statistical uncertainty of the zero-point energy is defined as the standard deviation of DMC energies over the total number of simulations. This is written as

$$\Delta E = \sqrt{\frac{1}{15}\sum_{i=1}^{15}(E_i - \bar{E})^2}, \tag{6.7}$$

where $\bar{E}$ is the average energy over the 15 simulations.

6.5 Results and Discussion

6.5.1 Full PIP PES

The PES is generated by using the full PIP basis is denoted full-PES. First, this full-PES is applied for the geometry optimization of both *trans* and *cis*-isomers and their two TSs. The structures of these isomers and saddle points are shown below, along with the energies relative to the *trans*-NMA.

As seen, *trans*-NMA is more stable than *cis*-NMA in accordance with previous reported results. Comparison of the *cis-trans* energy gap with the various theoretical results as well as the experimental value is shown in Table 6.1 and the agreement with our PES result is very good. We get this *cis-trans* energy gap as 818 cm^{-1} (2.32 kcal/mol) from our PES whereas the experimental value is 804 cm^{-1} (2.30 kcal/mol).[112]

Next to examine the vibrational frequency predictions of the PES, we performed normal mode analyses for both *trans* and *cis*-NMA and their two isomerization TSs. Comparisons of the harmonic frequencies for *trans* and *cis*-NMA and their two TSs with their corresponding *ab initio* ones are given in Tables 6.2 and 6.3. The agreement for

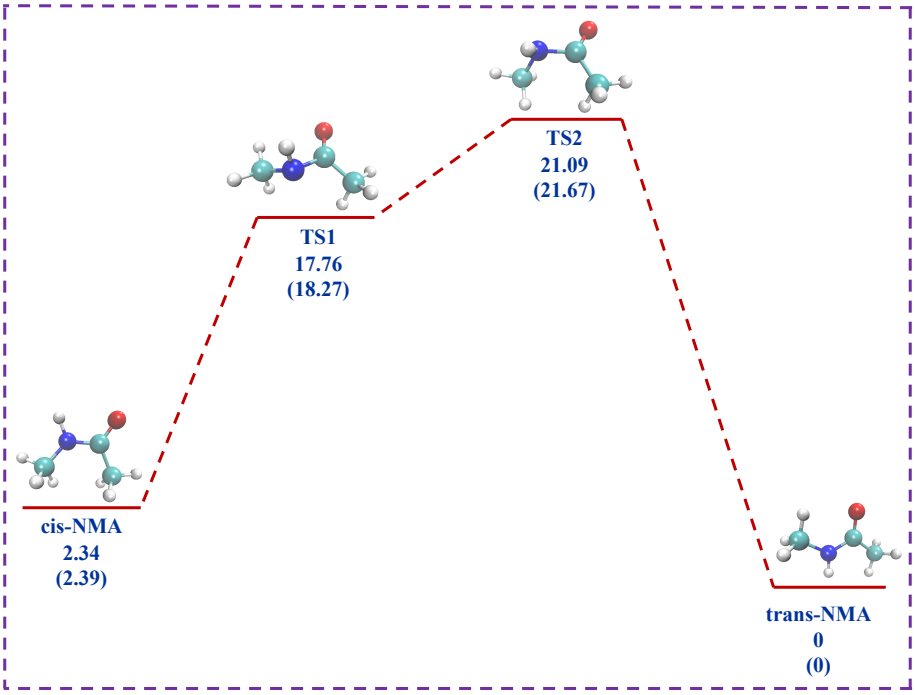


Figure 6.5: Schematic of the N-methyl acetamide potential energy surface showing stationary points and energies (kcal/mol), relative to the minimum, from the new fitted PES and *ab initio* energies (shown in parentheses).

Table 6.1: Comparison of *cis-trans* energy difference (ΔE) at different level of theory and the corresponding experimental value.

Method	ΔE (cm^{-1})	ΔE (kcal/mol)
HF/6-31G(d)	874	2.50
HF/6-311+G(d,p)	983	2.81
B3LYP/6-311+G(d,p)	867	2.48
MP2/6-311+G(d,P)	794	2.27
B3LYP/cc-pVDZ	818	2.32
Experiment	804	2.30

trans and *cis*-NMA with the direct B3LYP/cc-pVDZ frequencies is overall very good; the maximum error is 35 cm^{-1} but most of the frequencies are within several cm^{-1} of the *ab initio* ones. Somewhat larger differences are seen for the two isomerization TSs; however, the mean absolute deviations (MAD) are 28 and 24 cm $^{-1}$ respectively for TS1 and TS2. Recalling that only 6,607 geometries are used in the data sets, the good results for harmonic frequencies continue to demonstrate the power of simultaneous energy-gradient fitting method in obtaining such accurate harmonic frequencies.

Now we present the results of diffusion Monte Carlo calculations of the zero-point energy (ZPE) for both *cis* and *trans* isomers to explore more regions of the configuration space. Diffusion Monte Carlo calculation is a good probe of the quality of a PES in extended regions of the configuration space. We know one common issue in potential fitting is the unphysical behavior in regions without data and this can be dramatically manifested by large negative values. These are referred to as holes in PES. Generally, we have observed that these occur for highly repulsive configurations, i.e., short internuclear distances. Adding some data in these regions plus a refit generally eliminates these holes. The main purpose to present DMC calculations is to examine how our PES correctly describe the high energy regions for both isomers.

The ZPEs for *trans*-NMA and *cis*-NMA are 21905 ± 16 and 21934 ± 18 cm^{-1} respectively, whereas the harmonic ZPEs are 22206 and 22212 cm^{-1}.In these DMC calculations the holes detected in some trajectories is very small in number compared to the total number of configurations sampled in a trajectory. The number of holes is around $5 - 20$ where the total number of configurations sampled is on the order of 10^8. During the DMC propagation when a random walker encounters a hole (and thus in reality enter regions of large potential), we kill that walker and let the trajectory continue to propagate. This procedure follows our unbiased DMC algorithm. As an aside we believe this is the largest

Table 6.2: Comparison of harmonic frequencies (in cm^{-1}) between PES and the corresponding *ab initio* (B3LYP/cc-pVDZ) ones of *trans* and *cis*-NMA.

	trans-NMA		*cis*-NMA	
Mode	PES	*ab initio*	PES	*ab initio*
1	38	46	76	83
2	119	109	123	144
3	151	160	166	160
4	290	294	286	283
5	393	428	468	471
6	433	442	518	517
7	621	625	575	573
8	629	634	624	627
9	867	871	815	806
10	995	990	977	979
11	1038	1037	1043	1036
12	1113	1111	1087	1087
13	1132	1138	1137	1128
14	1167	1158	1187	1192
15	1260	1263	1330	1330
16	1391	1375	1393	1387
17	1415	1412	1439	1433
18	1434	1439	1447	1447
19	1474	1459	1462	1451
20	1485	1467	1467	1460
21	1491	1472	1478	1471
22	1552	1557	1508	1500
23	1772	1775	1798	1785
24	3019	3013	3001	2990
25	3040	3040	3040	3036
26	3072	3070	3069	3079
27	3125	3123	3107	3102
28	3127	3127	3116	3120
29	3137	3150	3155	3162
30	3631	3627	3592	3585

Table 6.3: Comparison of harmonic frequencies (in cm^{-1}) between PES and the corresponding *ab initio* (B3LYP/cc-pVDZ) ones of TS.

	TS1		TS2	
Mode	PES	*ab initio*	PES	*ab initio*
1	194i	157i	189i	139i
2	149	170	143	112
3	177	228	174	226
4	232	278	219	287
5	376	394	408	430
6	456	525	532	554
7	628	624	562	601
8	712	756	642	721
9	811	815	803	800
10	973	980	957	971
11	1036	1004	1024	1004
12	1083	1040	1080	1043
13	1088	1136	1125	1132
14	1159	1159	1158	1155
15	1234	1216	1200	1190
16	1367	1360	1355	1360
17	1400	1402	1363	1416
18	1436	1422	1428	1427
19	1444	1425	1440	1433
20	1479	1439	1465	1453
21	1503	1459	1492	1469
22	1527	1488	1501	1488
23	1802	1701	1845	1727
24	3020	2996	2968	2959
25	3050	3027	3032	3030
26	3064	3074	3071	3072
27	3104	3091	3086	3095
28	3111	3102	3106	3112
29	3158	3151	3151	3152
30	3461	3397	3439	3403

molecule for which a DMC calculation of the ZPE has been reported.

Next, we show the torsional potentials of the two methyl rotors of both *trans* and *cis*-isomers of NMA in Fig. 6.6 (a) and Fig. 6.7 (a), respectively, as a function of the torsion angles of the two methyl rotors. CH_3(CO) represents the CH_3 group attached to C=O group and CH_3(NH) represents the CH_3 group attached to N−H. In *trans*-NMA, the energy barriers of the CH_3(CO) rotor is 37.4 cm^{-1} and for the CH_3(NH) rotor it is 255.9 cm^{-1}. This result is close to the one obtained directly at B3LYP/cc-pVDZ calculations, namely 56.6 and 185.9 cm^{-1}, obtained by single-point calculations at the minimum and torsional barrier.

We see a flip of these two rotational barriers when we consider the *cis*-isomer of NMA. The energy barriers of CH_3(CO) rotor is 360.6 cm^{-1} and CH_3(NH) rotor is 61.3 cm^{-1} for the *cis*-NMA. This result also matches with the corresponding single point calculations at B3LYP/cc-pVDZ level, namely 301.4 and 67.6 cm^{-1}. Here, our purpose is to describe the 3-fold symmetry of the two methyl rotors qualitatively as the relevant permutational symmetry is incorporated in the PES and we are satisfied with these results.

6.5.2 Fragmented PIP PES

The fragmented basis PES using the 3-fragment scheme is denoted simply by frag-PIP PES. As a first test of this frag-PES, geometries of both *trans* and *cis*-isomers and the two TSs were optimized. Comparison of the relative energies obtaining from the two PESs (full- and frag-) is shown in the Table 6.4 along with the *ab initio* ones. It is seen that the relative energies from full-PES and frag-PIP ones are very close to each other and also match with the corresponding *ab initio* ones.

We also performed normal mode analyses using the frag-PIP PES at optimized ge-

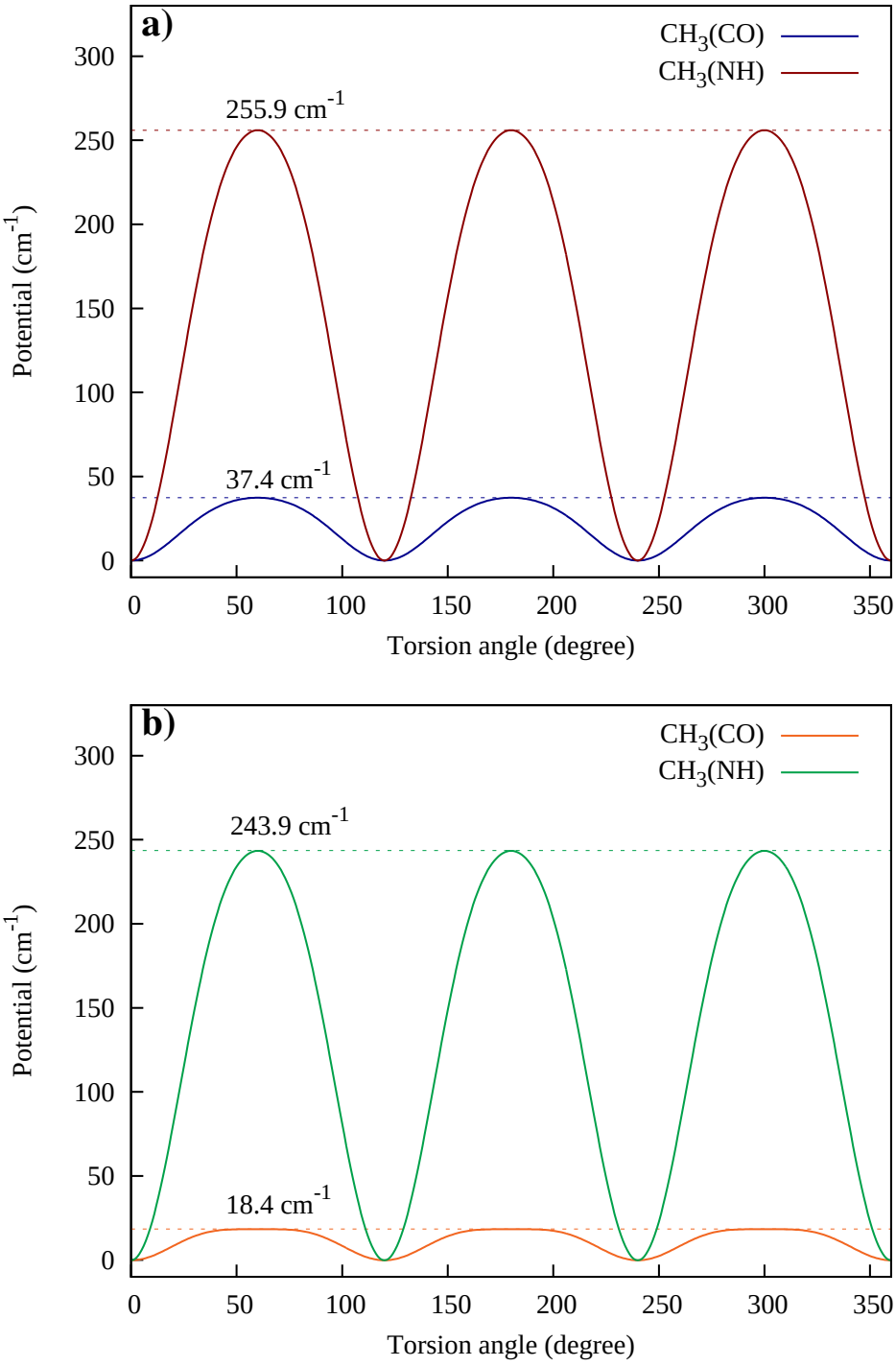


Figure 6.6: Torsional potential (not fully relaxed) of the two methyl rotors of *trans*-NMA from the full (a) and frag (b) PIP PESs .

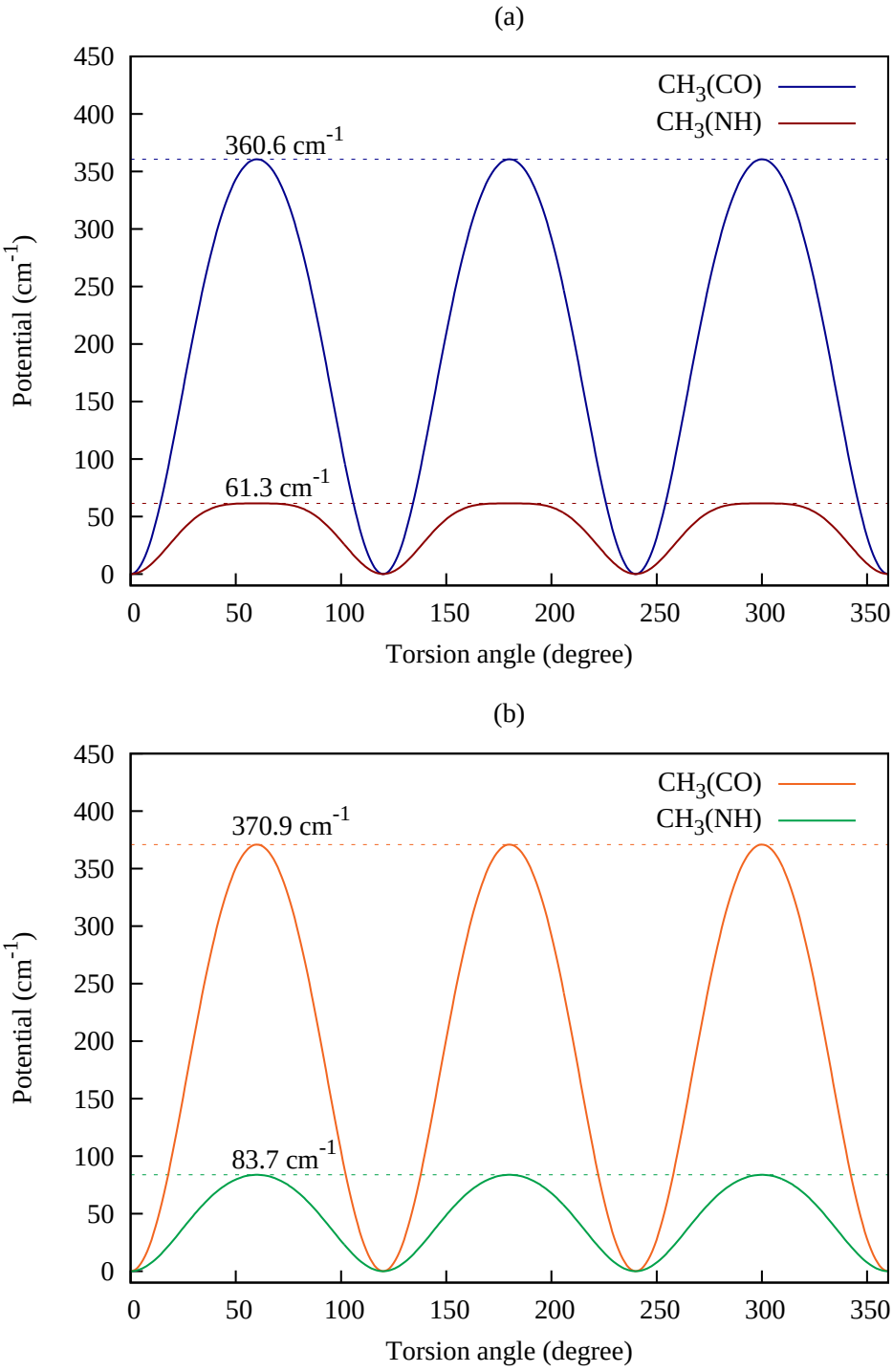


Figure 6.7: Torsional potential (not fully relaxed) of the two methyl rotors of *cis*-NMA from full (a) and frag-PIP (b) PESs.

Table 6.4: Comparison of relative energies (in cm^{-1}) of *N*-methyl acetamide stationary points from full-PES and frag-PIP PES with the *ab initio* ones.

Molecule	PIP PES	frag-PIP PES	*ab initio*
trans-NMA	0	0	0
cis-NMA	818	800	838
TS1	6211	6196	6391
TS2	7376	7353	7579

ometries of both *trans* and *cis*-NMA and their two TSs. A direct comparison of the harmonic frequencies for *trans* and *cis*-NMA and their two TSs from the full-PES and frag-PIP PES are given in Table 6.5 and 6.6. As seen, the frequencies from the frag-PIP PES are in good agreement with those from the full-PES, which as noted above are in good agreement with direct B3LP frequencies.

As we mentioned already, previously a 2-fragment scheme was employed to fit the PES of *trans*-NMA and the results were encouraging.[96] However, this 2-fragment scheme is unable to describe the *cis*-NMA and the two TSs. Specifically, using this two fragmented-PIP PES, we get very good normal mode frequencies for the *trans*-isomer in comparison to those from the full-PES, while for the *cis*-NMA and the two TSs we get energies that differ by 500–900 cm^{-1} from the full-PES results and large deviations for the normal mode frequencies. In fact we even could not locate first-order saddle points for TS1 and TS2. These failures are due to neglect of the inter-methyl interactions between the 3 'H' atoms of two methyl groups (1–3 and 10–12). These groups of atoms are closer in *cis*-NMA and TSs than in *trans*-NMA. Therefore, the addition of the third fragment (these six 'H' atoms) is necessary and we obtain very good results for geometry optimization and harmonic frequency calculations. As noted already, the third fragment is the smallest possible one that describes these six atoms with reduced 33 symmetry.

Table 6.5: Comparison of harmonic frequencies (in cm^{-1}) between full-PES and frag-PIP PES for *trans* and *cis*-NMA.

	trans-NMA		*cis*-NMA	
Mode	PIP PES	frag-PIP PES	PIP PES	frag-PIP PES
1	38	30	76	76
2	119	113	123	117
3	151	149	166	166
4	290	289	286	285
5	393	388	468	468
6	433	435	518	515
7	621	620	575	576
8	629	630	624	621
9	867	868	815	814
10	995	990	977	977
11	1038	1038	1043	1045
12	1113	1114	1087	1090
13	1132	1134	1137	1131
14	1167	1164	1187	1189
15	1260	1260	1330	1330
16	1391	1382	1393	1389
17	1415	1418	1439	1431
18	1434	1431	1447	1451
19	1474	1478	1462	1459
20	1485	1485	1467	1466
21	1491	1497	1478	1485
22	1552	1554	1508	1506
23	1772	1770	1798	1798
24	3019	3018	3001	3007
25	3040	3045	3040	3040
26	3072	3070	3069	3068
27	3125	3128	3107	3109
28	3127	3130	3116	3113
29	3137	3142	3155	3160
30	3631	3630	3592	3594

Table 6.6: Comparison of harmonic frequencies (in cm^{-1}) between full-PES and frag-PIP PES for the TSs.

	TS1		TS2	
Mode	PIP PES	frag-PIP PES	PIP-PES	frag-PIP PES
1	194i	197i	189i	191i
2	149	151	143	140
3	177	174	174	173
4	232	226	219	200
5	376	369	408	406
6	456	447	532	531
7	628	624	562	564
8	712	710	642	641
9	811	812	803	805
10	973	976	957	971
11	1036	1036	1024	1002
12	1083	1087	1080	1079
13	1088	1196	1125	1113
14	1159	1155	1158	1145
15	1234	1236	1200	1193
16	1367	1368	1355	1349
17	1400	1400	1363	1361
18	1436	1434	1428	1427
19	1444	1448	1440	1434
20	1479	1476	1465	1462
21	1503	1497	1492	1484
22	1527	1516	1501	1493
23	1802	1801	1845	1846
24	3020	3021	2968	2973
25	3050	3044	3032	3033
26	3064	3072	3071	3070
27	3104	3109	3086	3092
28	3111	3110	3106	3103
29	3158	3163	3151	3159
30	3461	3467	3439	3449

We also computed the torsional potential of two methyl rotors for both *trans* and *cis*-isomer using the frag-PIP PES. The energy barriers of CH_3(CO) rotor is 18.4 cm^{-1} and CH_3(NH) rotor is 243.6 cm^{-1} for the *trans*-NMA molecule and for *cis*-NMA these are 370.9 cm^{-1} and 83.7 cm^{-1} respectively. These results agree well with the full-PES results. The potential energy as a function of the torsional angles of the two methyl rotors of both *trans* and *cis*-isomers of NMA are shown in Fig. 6.6 (b) and Fig. 6.7 (b).

Previously, we reported the torsional potential and harmonic frequencies of *trans*-NMA using the PES having two 9-atom fragments and the result were generally very good.[96] If we apply this 2-fragment basis to fit the present data set and to compute the torsional barriers of *cis*-NMA, we get 346.6 cm^{-1} for CH_3(CO) rotor and 104.7 cm^{-1} for CH_3(NH) rotor respectively. So, the 2-fragment PES is also able to describe the torsional potentials well but not as well as the 3-fragment one. Another noticeable improvement in this work over the previous fragmented PES is the normal mode frequencies of *trans*-NMA. These frequencies were reported previously both for fragmented and full PESs and in both cases we got an imaginary frequency which is the lowest frequency mode. However, in the present work we get all real frequencies for *trans*-NMA for both fragmented PES and full PES. The reason behind this improvement is that the number of training geometries used to fit both PESs is larger than the previously reported PESs.

Recall that the PIP basis for each 9-atom fragment contains 3,028 terms with 36 Morse variables. The small 6-atom fragment of 33 symmetry contains 15 Morse variables and 65 terms. Clearly then this fragmented basis is more efficient to evaluate than the full PIP basis which contains 8,040 terms. A good test of the speed difference between the full and fragmented PESs is a standard DMC trajectory with the same number of random walkers and propagation time. The frag-PIP PES takes about 429 sec to finish a DMC trajectory calculation, while the full-PES takes about 793 sec which is almost

twice. This factor is larger than the ratio of the sizes of the basis. The reason is that some of PIPs in the full basis are more complex than those in the fragmented PIP basis and thus more costly to evaluate. The ZPEs obtained using the frag-PIP PES are within the statistical uncertainty of the DMC ZPEs reported above using the full PES.

6.6 Summary and Conclusions

In summary, we applied simultaneous energy-gradient fitting to obtain a full dimensional potential energy surface of 12-atom *N*-methyl acetamide molecule using both full and fragmented permutationally invariant polynomial basis. Both full- and frag-PIP PESs accurately describe *trans* and *cis* isomers and the two saddle points separating them. The frag-PIP PES uses a basis obtained from three fragments of NMA, namely two 9-atom fragments describing the 3H atoms on each methyl rotor and a third 6-atom one describing the 6 H atoms on two methyl groups. The quality of the frag-PIP PES is verified by tests against the full-PIP PES as well as direct DFT calculations. As expected, the frag-PIP PES is significantly faster than the full one. The approach of partitioning the Morse variables, which is a prerequisite for the fragmented PIP approach, is general and can be used with other fitting methods. Thus, the fragmented fitting approach can be applied to many molecules with more than 10 atoms.

Chapter 7

Implementation of Δ-Machine Learning in Permutationally Invariant Polynomial Fitting to Obtain CCSD(T) Level Accuracy

7.1 Chapter Abstract

"Δ-machine learning" refers to a machine learning approach to bring a property such as a potential energy surface (PES) based on low-level (LL) density functional theory (DFT) energies and gradients to close to a coupled cluster (CC) level of accuracy. Here we present such an approach that uses the permutationally invariant polynomial (PIP) method to fit high-dimensional PESs. The approach is represented by a simple equation, in obvious notation $V_{LL\rightarrow CC} = V_{LL} + \Delta V_{CC-LL}$, and demonstrated for CH_4, H_3O^+, and *trans* and *cis*-*N*-methyl acetamide (NMA), $CH_3CONHCH_3$. For these molecules the LL

PES, V_{LL}, is a PIP fit to DFT/B3LYP/6-31+G(d) energies and gradients and ΔV_{CC-LL} is a precise PIP fit obtained using a low-order PIP basis set and based on a relatively small number of CCSD(T) energies. For CH_4 these are new calculations adopting an aug-cc-pVDZ basis, for H_3O^+ previous CCSD(T)-F12/aug-cc-pVQZ energies are used, while for NMA new CCSD(T)-F12/aug-cc-pVDZ calculations are performed. With as few as 200 CCSD(T) energies, the new PESs are in excellent agreement with benchmark CCSD(T) results for the small molecules, and for 12-atom NMA training is done with 4696 CCSD(T) energies.

7.2 Overview

Correcting *ab initio*-based potential energy surfaces (PESs) has been a long-standing goal of computational chemistry. Several approaches dating from 30 years ago have been suggested. In one, a correction potential is added to an existing PES and parameters of the correction potential are optimized by matching ro-vibrational energies to experiment.[125–127] This approach relies on being able to calculate exact ro-vibrational energies to make the comparison with experiment robust. Thus, it has only been applied to triatomic molecules and it is limited to these and possibly tetratomics. Another approach is to modify an existing potential using scaling methods that go under the heading of "morphing".[128–130] An impressive example is a PES for HCN/HNC reported by Tennyson and co-workers[131] who morphed a CCSD(T)-based PES.[132]

More recent approaches using machine learning (ML) aim to bring a PES based on a low-level of electronic theory to a higher level. As the field moves to consideration of larger molecules and clusters, where high-level methods are prohibitively expensive, the motivation for doing this is obvious. There are two classes of such approaches, one is

"Δ-machine learning" (Δ-ML) and the other is "transfer learning".[133] Δ-ML, which is of direct relevance to the present paper, seeks to add a correction to a property obtained using an efficient and thus perforce low-level *ab initio* theory.[134–139] This approach includes an interesting, recent variant based on a "Pople" style composite approach.[135] In this sense the approach is related, in spirit at least, to the correction potential approach mentioned above, when the property is the PES. However, it is applicable to much larger molecules.

The transfer learning approach has been developed extensively in the context of neural networks[133] and so much of the work in that field has been brought into chemistry.[136–140] The idea of transfer learning comes from the fact that knowledge gained from solving one problem can often be used to solve another related problem. Therefore, a model learned for one task, e.g., a ML-PES fit to low-level electronic energies/gradients, can be reused as the starting point of the model for a different task, e.g., an ML-PES with the accuracy of a high-level electronic structure theory.

Most work using transfer learning or Δ-ML has been on developing general transferable force fields with application mainly in the area of thermochemistry and molecular dynamics simulations at room temperature and somewhat higher. Meuwly and co-workers have used transfer learning to improve neural network PESs for malonaldehyde, acetoacetaldehyde and acetylacetone.[139]

Here we report a Δ-ML approach for PESs, using the permutationally invariant polynomial (PIP) approach. The PIP approach has been applied to many PESs for molecules, including chemical reactions, dating back roughly 15 years. For reviews see Refs. 26,29,65. Recent extensions of the PIP software to incorporate electronic gradients[141,142] have extended the PIP approach to amino acids (glycine)[143] and molecules with 12 and 15 atoms, e.g., *N*-methyl acetamide,[142,144,145] tropolone,[146] and acetylacetone,[147] respectively. As is

widely appreciated in the field, incorporating gradients into fitting requires efficient, low-level electronic structure methods, such as density functional theory or MP2, as these provide analytical gradients.[148] These levels of theory were used for the PES fits of the three molecules mentioned above.

Our approach is given by the simple equation

$$V_{LL\rightarrow CC} = V_{LL} + \Delta V_{CC-LL}, \tag{7.1}$$

where $V_{LL\rightarrow CC}$ is the corrected PES, V_{LL} is a PES fit to low-level DFT electronic data, and ΔV_{CC-LL} is the correction PES based on high-level coupled cluster energies. The assumption underlying the hoped-for small number of high-level energies is that the difference ΔV_{CC-LL} is not as strongly varying as V_{LL} with respect to nuclear configuration.

We demonstrate the efficacy and high-fidelity of this approach for two small molecules, H_3O^+ and CH_4, and for 12-atom *N*-methyl acetamide (NMA). In all cases V_{LL} is a PIP fit to DFT energies and gradients and ΔV_{CC-LL} is a PIP fit to a much smaller data base of differences between CCSD(T) and DFT energies.

Unlike H_3O^+ and CH_4, for NMA there is no previous CCSD(T)-based PES and so the present CCSD(T)-corrected one is, we believe, the most accurate one available.

7.3 Computational Details

In order to develop a corrected PES we need to generate a data set of high and low-level energies for training and testing. In this study we need both DFT and CCSD(T) data sets. Training is done for the correction PES ΔV_{CC-LL} and testing is done for the corrected $V_{LL\rightarrow CC}$. Do note that this two-step "training and testing" is on different data

sets. Our objective is to see the impact of the training data set size on the fidelity of the corrected PES $V_{LL\rightarrow CC}$ for CH_4 and H_3O^+.

For CH_4 we take the DFT data sets from our recently reported work where the total of 9000 energies and their corresponding gradients were generated from *ab initio* molecular dynamics (AIMD) simulations, using the B3LYP/6-31+G(d) level of theory.[141] In that work we reported PES fits using a number of subsets of the DFT data which span the energy 0–15000 cm^{-1}. This PES was developed to demonstrate new PIP software for fitting energies and gradients.[123] This PES is based on precise fitting energies and gradients at 600 configurations. Here, this PES is brought to the CCSD(T) level of accuracy.

We generate a data set that contains CCSD(T)/aug-cc-pVDZ energies at 3000 configurations, taken from the previously reported DFT data.[141] A number of training data sets and one test data set, which are subsets of this 3000 data, are employed to examine the Δ-ML procedure. Training and testing is done on several partitions of this dataset. Figure 7.1 shows the distribution of CCSD(T) and DFT eneriges for the traing data set of 600 and test data set of 2000. As seen the range of these energies is 0 to roughly 15 000 cm^{-1}, relative to respective minima. The CCSD(T) and DFT distributions look very similar and this is because the difference in these energies (relative to their respective minima) are much smaller than the range of the energies.

We train ΔV_{CC-LL} using the difference of CCSD(T) and DFT absolute energies and then test on a different set. A variety of training data sets were used and details of one example are shown in Figure 7.2. There we plot ΔV_{CC-LL} versus the DFT energies for a training set of 600 configurations. Testing the fit is done on a larger dataset of 2000 energy differences, shown in the lower panel of this figure. For clarity we reference ΔV_{CC-LL} to the minimum of the difference of the CCSD(T) and DFT energies (which is roughly -19

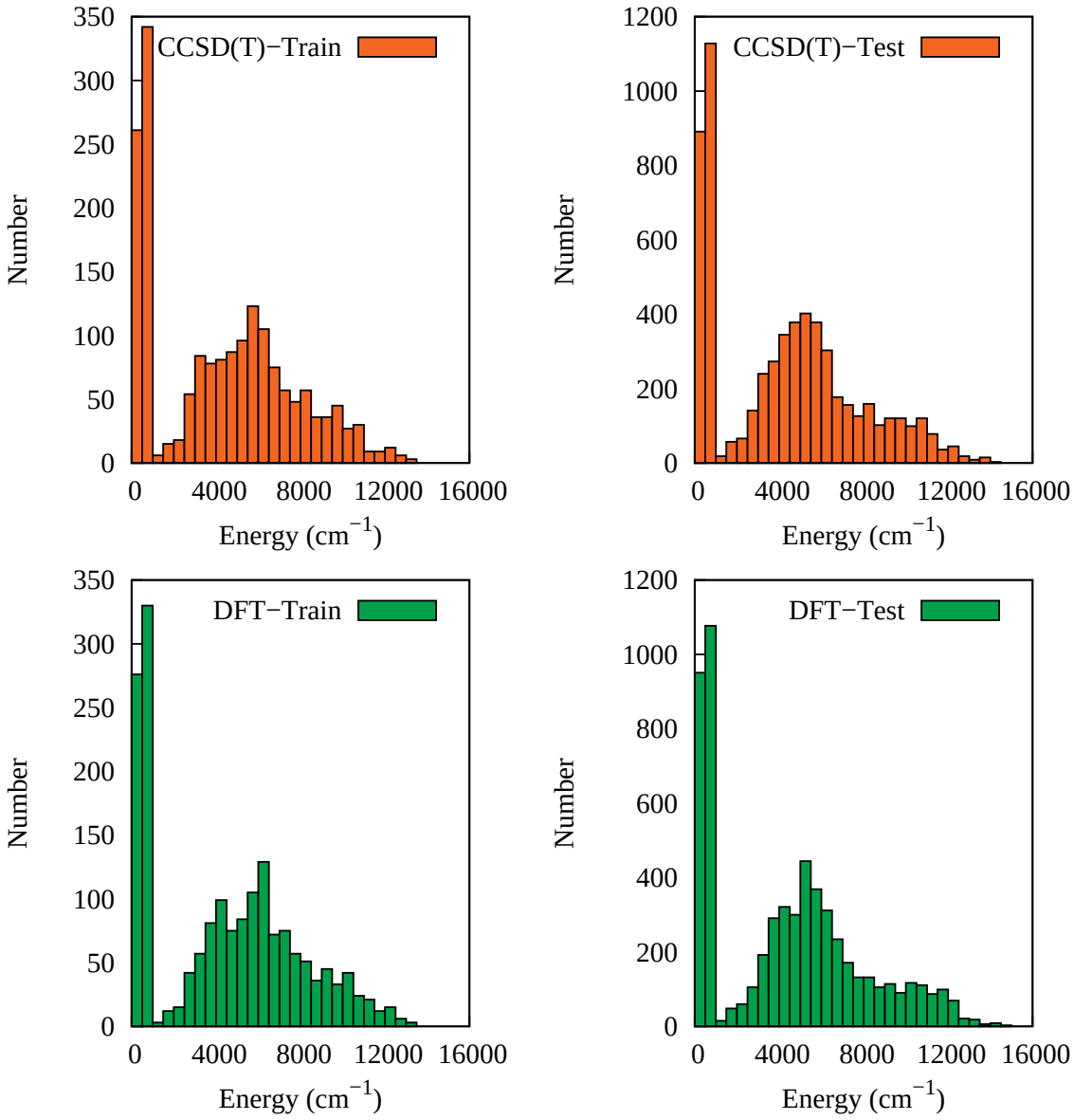


Figure 7.1: Distribution of energies (cm^{-1}) relative to the minimum value of CH_4 for both CCSD(T) and DFT training and test data sets.

820 cm^{-1}). As seen, the energy range of ΔV_{CC-LL} is about 1600 cm^{-1}, which is much smaller than the DFT energy range relative to the minimum value (which is roughly 15 000 cm^{-1}). Also note the training dataset looks like a "thinned" version of the larger test dataset.

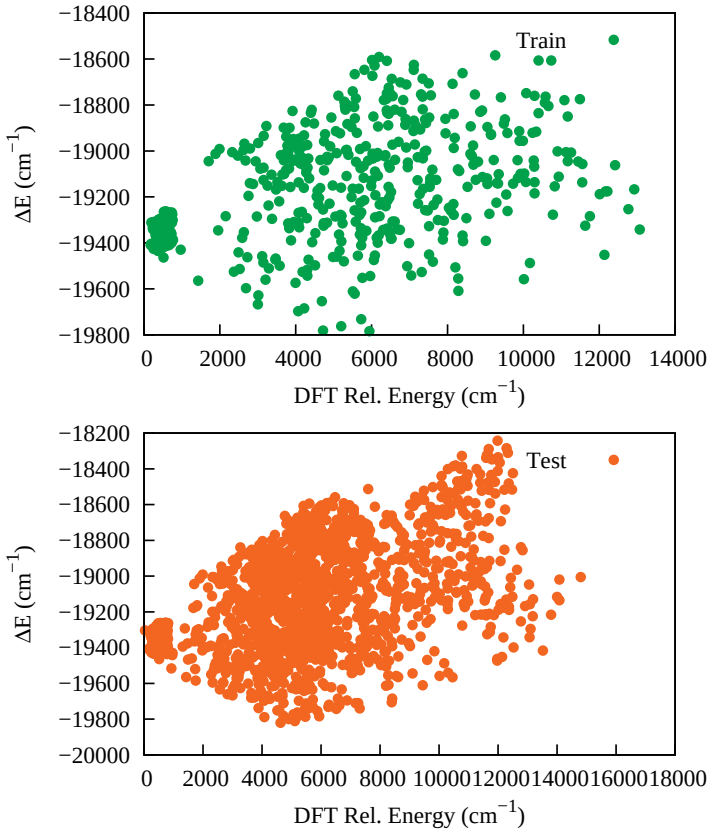


Figure 7.2: Plot of ΔV_{CC-LL} (relative to the reference value i.e. -19 820 cm^{-1}) vs DFT energy relative to the CH_4 minimum value with the indicated number of training and test data set.

ΔV_{CC-LL} is fit using a dataset of 600 energies using a PIP (describing the 4! permutations of the H atoms) basis with maximum polynomial order 5. This results in a basis of 208 terms. The linear coefficients were determined by solving the linear least-squares

problem; the fitting RMS error of this ΔV_{CC-LL} fit is 0.18 cm^{-1}. The fit was tested on the 2000 energy differences. The RMS test error in this case is 0.79 cm^{-1}.

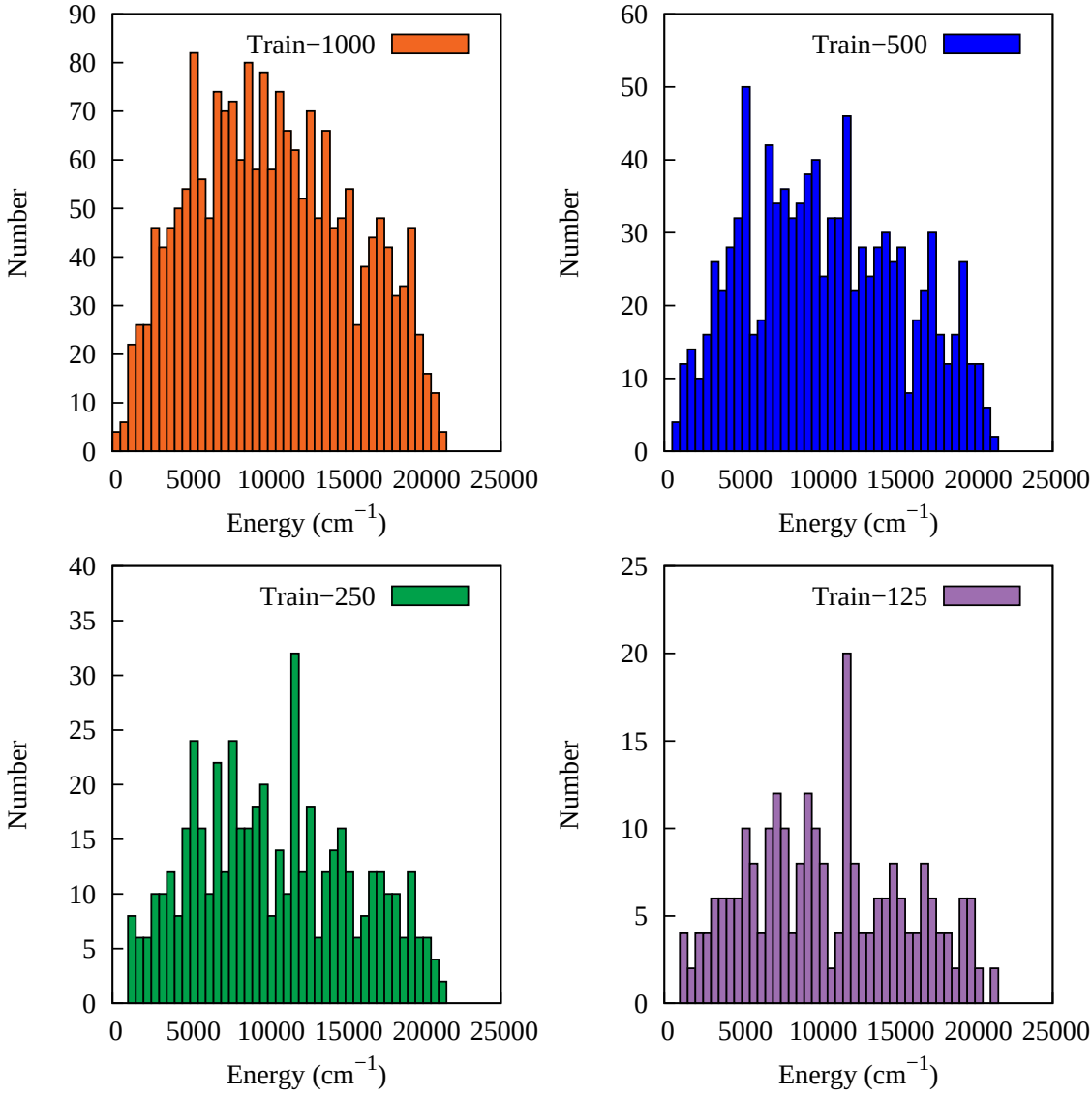


Figure 7.3: Distribution of DFT energies (cm^{-1}) relative to the minimum value of Hydronium ion with indicative training data set.

For H_3O^+ CCSD(T) energies are available from our previously reported PES, which

is a fit to 32 142 CCSD(T)/aug-cc-pVQZ energies.[149] From this large CCSD(T) data set, four different datasets are generated with energies in the range 0 to 24 000 cm^{-1} for new DFT calculations of energies and gradients, and the remaining geometries are considered as the corresponding test data set. These new DFT calculations are done at the efficient B3LYP/6-311+G(d,p) level of theory, using the Molpro quantum chemistry package.[87] Figure 7.3 shows the distribution of DFT energies for the all the four different sets of training data. As seen the range of these energies is 0 to roughly 23 000 cm^{-1}, relative to the minima for all the training sets. Note, these DFT configurations span the same large range of configurations as the much larger CCSD(T) ones, but have less dense sampling.

A data set of 1000 DFT energes and their corresponding gradients are used to fit the B3LYP/6-311+G(d,p) PES. We use a maximum polynomial of 7 that results total of 348 PIP basis functions. The corresponding linear coefficients are determined by solving linear least-square problem. The RMS errors of this fitted PES are 8.76 cm^{-1} for all energies and 2.13 cm^{-1}/bohr per atom for gradients.

For NMA we make use of previous DFT/B3LYP/cc-pVDZ energies and the corresponding PES that spans both the *trans* and *cis* isomers and barriers separating them.[145] New CCSD(T)-F12/aug-cc-pVDZ calculations are done at a sparse set (5430) of configurations that span the full range of configurations used in the previous work. This 5430 new CCSD(T) dataset is partitioned into a training data of 4696 geometries and the test data of 734 geometries. Figure. 7.4 shows the distribution of CCSD(T)-F12/AVDZ energies for the training and test data sets. Both the training and test data are distributed to a wide energy range that spans both the *trans* and *cis* isomers and barriers separating them. These datasets are used to obtain the ΔV_{CC-LL} PES.

The PIP fits of ΔV_{CC-LL} are done using our recent monomial symmeterization software.[123,141] Some details of the PIP bases are given in the next section. We note that

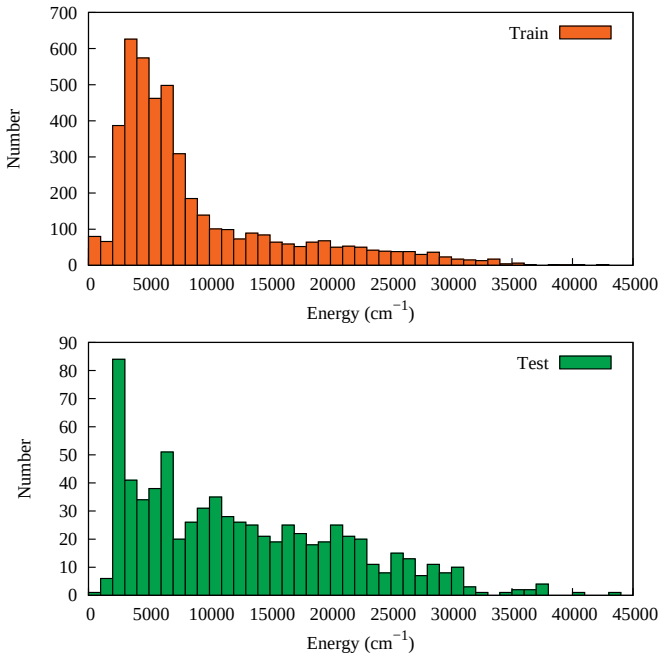


Figure 7.4: Distribution of CCSD(T)-F12/AVDZ energies (cm^{-1}) relative to the minimum value of *N*-methyl acetamide for both training and test data sets.

they are all small relative to typical PIP bases needed for precise fitting of the full PES for these molecules.

For all molecules the data sets are partitioned into several training and testing subsets to examine how few data are needed for training to get satisfactory results.

7.4 Results

We present root mean square (RMS) errors for $V_{LL\rightarrow CC}$ relative to direct CCSD(T) energies for a variety of ΔV_{CC-LL} fits. In addition, comparisons are made with direct CCSD(T) results for the geometry and harmonic frequencies of relevant stationary points. To assess the performance of the present approach these results are placed alongside the corresponding DFT ones.

7.4.1 CH_4

To obtain the corrected PES $V_{LL\rightarrow CC}$ we add ΔV_{CC-LL} to one of our previously reported DFT-based PESs for CH_4.[141] The fitting RMS error of that V_{LL} DFT PES is 2.55 cm^{-1}. To examine how well the corrected PES $V_{LL\rightarrow CC}$ reproduces the CCSD(T) energies we plot the energies of $V_{LL\rightarrow CC}$ vs the direct CCSD(T) energies for both training and test data sets in Figure. 7.5. As seen, the correspondence between the two is excellent. The overall RMS differences are 2.5 and 3.2 cm^{-1} for the training and testing set, respectively. This is already an indication of the success of the present approach to correct the DFT-based PES and bring the corrected PES very close to CCSD(T) energies.

To examine the robustness of these results, fits of ΔV_{CC-LL} were also done using datasets of 1000, 300, and 100 energies. The corresponding RMS differences between the $V_{LL\rightarrow CC}$ and CCSD(T) energies are given in Table 7.1. In case of the training set with $N_{Train} = 1000$, we use a maximum polynomial order of 5 to fit ΔV_{CC-LL}, and this leads to a basis of 208 terms. For the training data set of 300 and 100, we reduce the maximum polynomial order to avoid overfitting. As seen, the RMS errors are similar for all the training data sets. This is certainly notable for the training set of only 100 energy differences, where the RMS error is only 4.9 cm^{-1} for the 2000 test energies up to 15000 cm^{-1}. In this case the PIP basis for ΔV_{CC-LL} contains only 30 terms.

Next consider the performance of the Δ-ML approach for the equilibrium geometry and normal mode frequencies. The results are compared to the CCSD(T) and DFT ones in Table 9.1. As seen, the corrected PES $V_{LL\rightarrow CC}$ produces results in excellent agreement with direct CCSD(T) ones and also provides a large improvement compared to the DFT PES. Perhaps most impressive is the excellent accuracy achieved even with the smallest training dataset of 100 energies. For example, the rms deviation for the

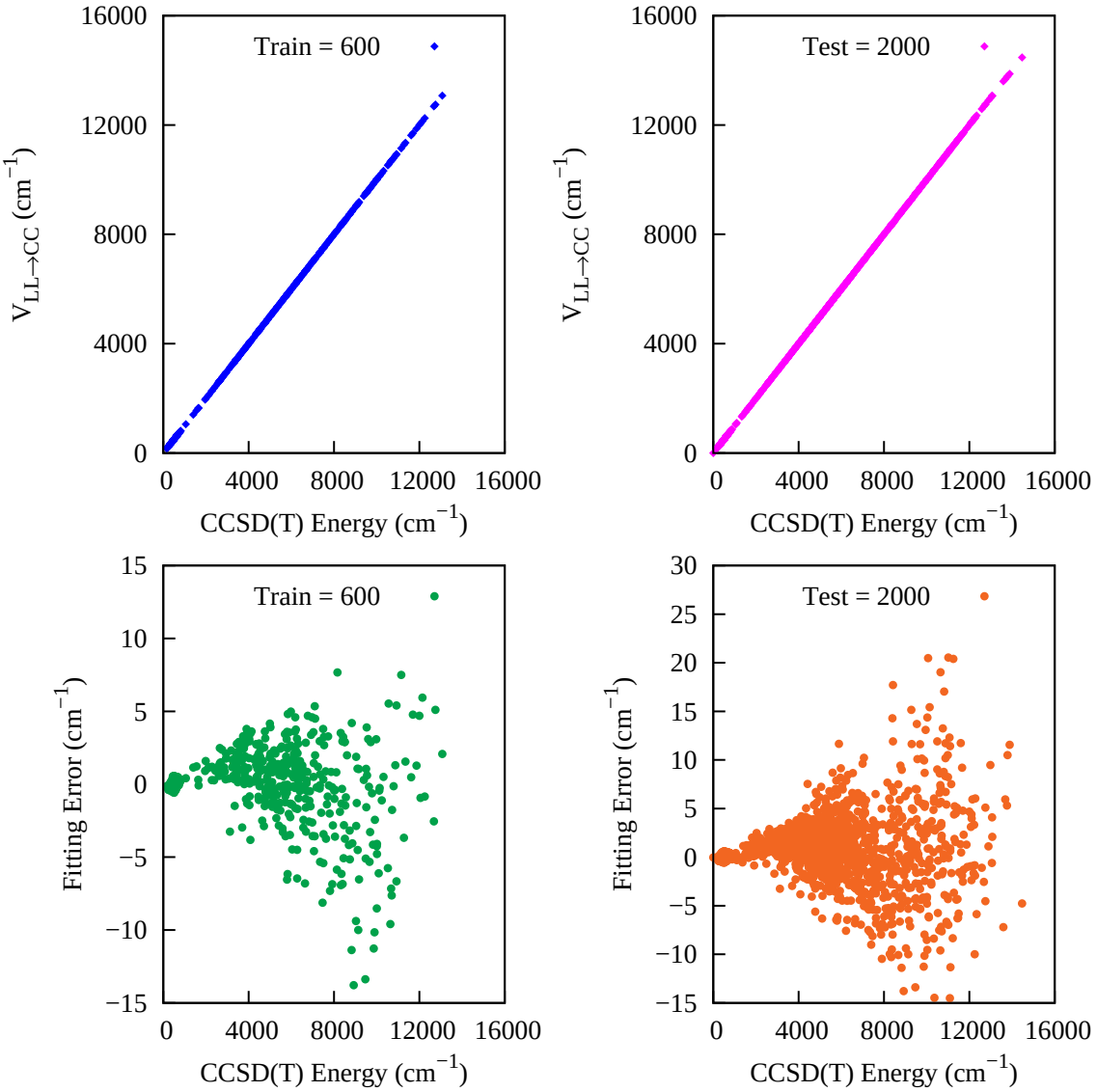


Figure 7.5: Two upper panels show energies of CH_4 from $V_{LL\rightarrow CC}$ vs direct CCSD(T) ones for the indicated data sets. The one labeled "Train" corresponds to the configurations used in the training of ΔV_{CC-LL} and the one labeled "Test" is just the remaining configurations. Corresponding fitting errors relative to the minimum energy are given in the lower panels.

Table 7.1: RMS error between direct CCSD(T) and $V_{LL\rightarrow CC}$ energies ($\mathrm{RMS_E}$) ($\mathrm{cm^{-1}}$) with the indicated number of test (N_{Test}) configurations for CH_4, where training on ΔV_{CC-LL} is done for various training data (N_{Train}).

N_{Train}	N_{Test}	$\mathrm{RMS_E}$
1000[a]	2000	3.14
600[a]	2000	3.22
300[b]	2000	4.41
100[c]	2000	4.88

[a] ΔV_{CC-LL} is trained with maximum polynomial order of 5, basis size of 208.
[b] ΔV_{CC-LL} is trained with maximum polynomial order of 4, basis size of 83.
[c] ΔV_{CC-LL} is trained with maximum polynomial order of 3, basis size of 30.

Table 7.2: Comparison of differences, δ, in bond lengths (angstroms) and harmonic frequencies ($\mathrm{cm^{-1}}$) relative to direct CCSD(T) benchmarks for the minimum of CH_4. Note 2.0(-5) means 2.0×10^{-5}, etc

	Geom. Param.		Harmonic Freq.			
N_{Train}	δ(C-H)	δ(H-H)	$\delta\nu_1$	$\delta\nu_2$	$\delta\nu_3$	$\delta\nu_4$
1000[a]	-2.0(-5)	-2.5(-5)	-0.9	1.8	-1.1	0.5
600[a]	-2.0(-5)	-2.5(-5)	-1.2	1.7	-1.3	0.6
300[b]	-3.0(-5)	-4.5(-5)	-0.2	1.7	-0.1	-0.1
100[c]	-2.0(-5)	-3.5(-5)	-0.2	1.5	-2.5	0.3
DFT	8.1(-3)	1.3(-2)	-46.1	-47.2	-25.32	-4.0

[a] Maximum polynomial order of 5, basis size of 208.
[b] Maximum polynomial order of 4, basis size of 83.
[c] Maximum polynomial order of 3, basis size of 30.

harmonic frequencies with respect to the CCSD(T) values is reduced from 31 cm^{-1} in the DFT PES to about 1 cm^{-1} for the corrected PES.

7.4.2 H_3O^+

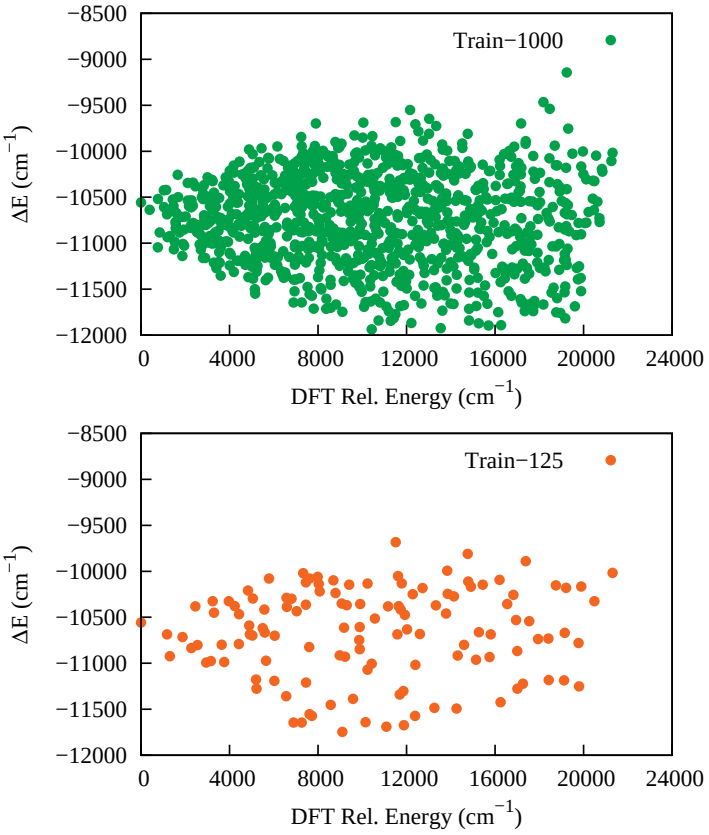


Figure 7.6: Plot of ΔV_{CC-LL} (relative to the reference value i.e. -12 110 cm^{-1}) vs DFT energy relative to the H_3O^+ minimum value with the indicated number of training data sets.

For H_3O^+ we trained ΔV_{CC-LL} on several sets of the difference of CCSD(T) and DFT absolute energies and then tested on the remaining data from the total of 32 142

configurations. In Figure. 7.6 we plot ΔV_{CC-LL} versus the DFT energies, relative to the DFT minimum for two training sets. We reference ΔV_{CC-LL} to the minimum of the difference between the CCSD(T) and DFT energies (which is roughly -12 110 cm^{-1}). As seen, the energy range of ΔV_{CC-LL} is about 3000 cm^{-1}, which is much smaller than the DFT energy range relative to the minimum value (which is roughly 23 000 cm^{-1}).

The performance of the ΔV_{CC-LL} fits is evaluated using the training data sets of 1000, 500, 250 and 125 configurations and the corresponding test data sets consist of the remaining data from the total of 32 142 configurations. The corresponding RMS differences between the $V_{LL\rightarrow CC}$ and CCSD(T) energies are given in Table 7.3. As seen, the RMS errors are similar for all the training data sets. Results for the training set of only 125 energy differences are particularly encouraging, where the RMS error is just 32 cm^{-1} for test energies up to 23 000 cm^{-1}. In this case the PIP basis for ΔV_{CC-LL} contains only 51 terms.

A plot of $V_{LL\rightarrow CC}$ vs direct CCSD(T) energies for the training set of 500 points and its corresponding test data is shown in Figure. 7.7. As seen, there is excellent precision; however, we see some large errors for the test data set. These come from high energy configurations which are irrelevant in this study. If needed, one can always improve these errors by adding the high energy data points into the training data set.

An examination of the fidelity of $V_{LL\rightarrow CC}$ for various properties is given in Tables 7.4 and 7.5, for the indicated training sets for ΔV_{CC-LL}. As seen, $V_{LL\rightarrow CC}$ produces results in excellent agreement with direct CCSD(T) ones and also a large improvement compared to the DFT PES. Most impressive is the high accuracy achieved even with the smallest training data set of 125 energies.

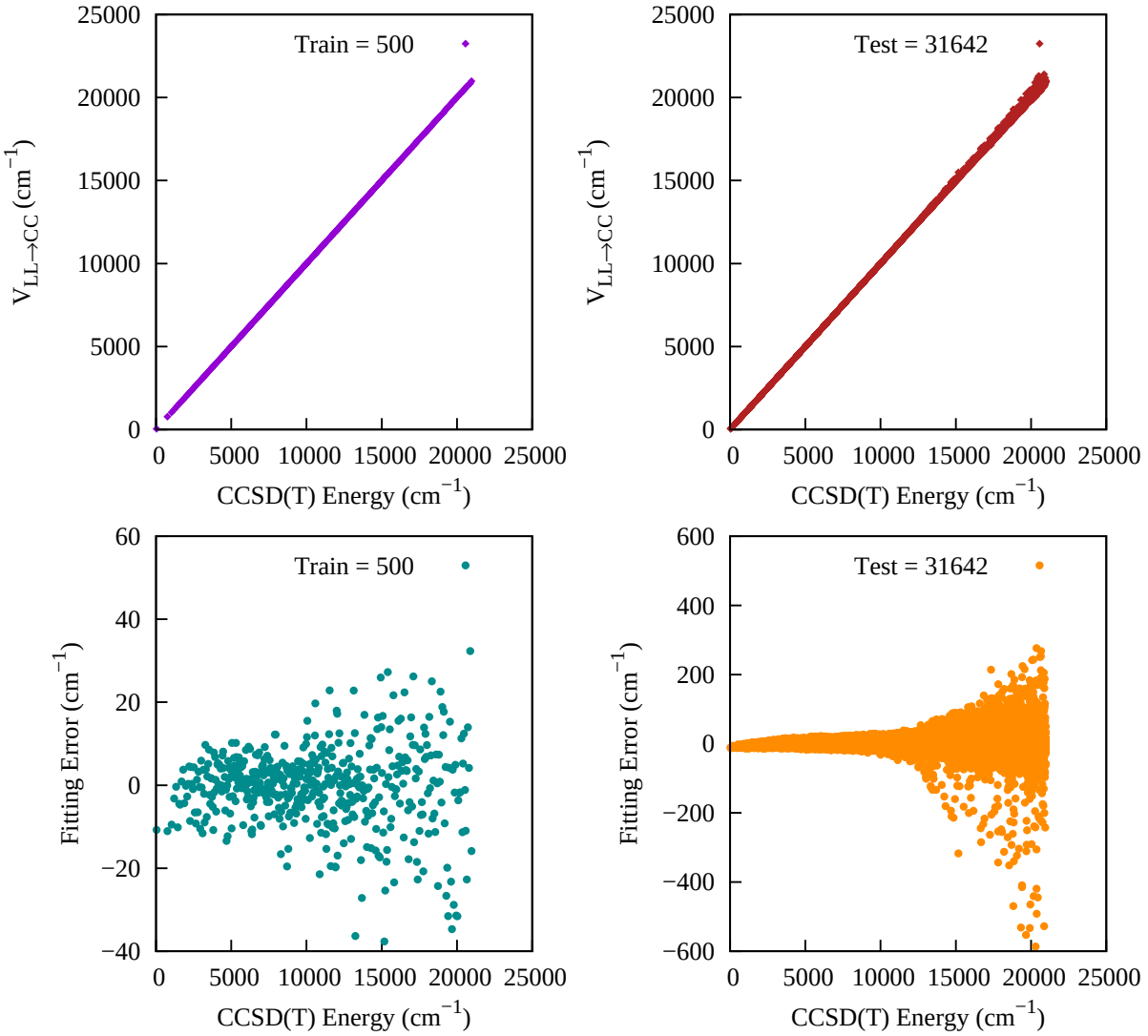


Figure 7.7: Two upper panels show energies of H_3O^+ from $V_{LL\rightarrow CC}$ vs direct CCSD(T) ones for the indicated data sets. The one labeled "Train" corresponds to the configurations used in the training of ΔV_{CC-LL} and the one labeled "Test" is just the remaining configurations. Corresponding fitting errors relative to the minimum energy are given in the lower panels.

Table 7.3: RMS error between direct CCSD(T) and $V_{LL\rightarrow CC}$ energies (RMS_E) (cm^{-1}) with the indicated number of test (N_{Test}) configurations for H_3O^+, where training on ΔV_{CC-LL} is done for various training data (N_{Train}).

N_{Train}	N_{Test}	RMS_E
1000[a]	31142	55.11
500[b]	31642	28.39
250[c]	31892	50.78
125[d]	32017	32.46

[a] ΔV_{CC-LL} is trained with maximum polynomial order of 7, basis size of 348.
[b] ΔV_{CC-LL} is trained with maximum polynomial order of 6, basis size of 196.
[c] ΔV_{CC-LL} is trained with maximum polynomial order of 5, basis size of 103.
[d] ΔV_{CC-LL} is trained with maximum polynomial order of 4, basis size of 51.

Table 7.4: Comparison of differences, δ, in bond lengths (angstroms) and harmonic frequencies (cm^{-1}) of the corrected PES, $V_{LL\rightarrow CC}$, relative to direct CCSD(T) benchmarks for the minimum of H_3O^+ for indicated training sets of ΔV_{CC-LL}. DFT PES results are also given. Note 3.0(-5) means 3.0 x 10^{-5}, etc.

	Geom. Param.		Harmonic Freq.			
N_{Train}	δ(O-H)	δ(H-H)	δv_1	δv_2	δv_3	δv_4
1000[a]	-3.0(-5)	-1.8(-4)	4.8	1.8	-4.4	3.3
500[b]	-5.0(-5)	-4.4(-4)	6.2	4.7	0.02	3.5
250[c]	-3.0(-5)	-7.8(-4)	2.6	4.8	6.3	0.02
125[d]	1.0(-5)	13.3(-4)	-9.1	-12.1	-8.6	3.02
DFT	-47.8(-4)	-24.1(-3)	125.9	26.5	26.5	33.7

[a] Maximum polynomial order of 7, basis size of 348.
[b] Maximum polynomial order of 6, basis size of 196.
[c] Maximum polynomial order of 5, basis size of 103.
[d] Maximum polynomial order of 4, basis size of 51.

Table 7.5: Comparison of differences, δ, in bond lengths (angstroms) and harmonic frequencies (cm^{-1}) of the corrected PES, $V_{LL\rightarrow CC}$, relative to direct CCSD(T) benchmarks for the saddle point of H_3O^+ for indicated training sets of ΔV_{CC-LL}. DFT PES results are also given. Note 3.0(-5) means 3.0 x 10^{-5}, etc.

	Geom. Param.		Harmonic Freq.				
N_{Train}	δ(O-H)	δ(H-H)	δv_1	δv_2	δv_3	δv_4	δ(Barrier)
1000[a]	-5.0(-5)	-9.0(-5)	-3.1i	3.3	-6.1	1.3	2
500[b]	-1.0(-5)	-2.0(-5)	-2.6i	2.0	-2.2	-0.7	10
250[c]	-2.2(-4)	-3.8(-4)	-1.2i	1.2	7.7	-4.3	7
125[d]	-1.0(-5)	-1.0(-5)	-0.7i	-3.7	-3.0	-4.8	-9
DFT	-70.6(-4)	-12.2(-3)	111.3i	17.6	45.5	58.7	297

[a] Maximum polynomial order of 7, basis size of 348.
[b] Maximum polynomial order of 6, basis size of 196.
[c] Maximum polynomial order of 5, basis size of 103.
[d] Maximum polynomial order of 4, basis size of 51.

7.4.3 *N*-methyl acetamide

We recently reported DFT-based PESs for 12-atom *N*-methyl acetamide (NMA) using full and fragmented PIP basis sets.[144,145] The idea of using a fragmented basis to extend the PIP approach to molecules with more than 10 atoms was illustrated for NMA. The data set for the more recent PES, which describes the *cis* and *trans* minima as well as saddle points separating them, consisted of energies and gradients. The full basis of maximum polynomial order of 3 has 8040 linear coefficients. The fragmented PIP basis, also with a maximum polynomial order of 3, contains 6121 coefficients.

The fits were done using 6607 energies and corresponding 237 852 gradient components for a total data size of 244 459. These data were obtained from direct dynamics, using the B3LYP/cc-pVDZ level. Clearly a data set of this size from CCSD(T) calculations is not feasible and so the present approach is needed in order to bring this

DFT-based PES close to CCSD(T) quality.

For the training and testing we calculated a total of 5430 CCSD(T)-F12/aug-cc-pVDZ energies. Training of ΔV_{CC-LL} was done on 4696 data points of the difference of direct CCSD(T) and DFT-PES absolute energies. Testing of $V_{LL\rightarrow CC}$ was done on 734 energies. The distribution of the electronic energies (shown in Figure. 7.4) for both the training and test data sets spans the large range of configurations used for the DFT-based PES, i.e., *trans* and *cis* isomers and their isomerization TSs.

In Figure. 7.8 we show the range of ΔV_{CC-LL} versus the DFT energies, relative to the DFT minimum for the training and test data sets. We reference ΔV_{CC-LL} to the minimum of the difference of the CCSD(T) and DFT energies (which is roughly -50 580 cm^{-1}). As seen, the energy range of ΔV_{CC-LL} is about 4500 cm^{-1}, which is much smaller than the DFT energy range relative to the minimum value (which is roughly 50 000 cm^{-1}). The PIP basis to fit the ΔV_{CC-LL} is generated using MSA software with the same reduced permutational symmetry of 31111113 (this describes the identity of the hydrogen atoms within a methyl group which is essential to get the three fold torsional barrier) used previously but and a maximum polynomial order of 2. This leads to 569 linear coefficients (PIP basis). The fitting RMS error of this ΔV_{CC-LL} is 57 cm^{-1}. A plot of $V_{LL\rightarrow CC}$ vs direct CCSD(T) energies for the training and test data is shown in Figure. 7.9. The RMS differences between the $V_{LL\rightarrow CC}$ and direct CCSD(T) energies for the training and test data sets are 57 and 147 cm^{-1}, respectively. A slight increment of the test RMS error is comparable with the DFT PES RMS error of 126 cm^{-1}.

We perform geometry optimization and normal mode analyses for both *trans* and *cis* isomers using this Δ-ML PES to examine the high fidelity of this PES, and we get significant improvement from the DFT PES, which predicts an incorrect minimum for the *trans*-isomer. Specifically, the torsion angle of one methyl rotor is shifted by 60 deg

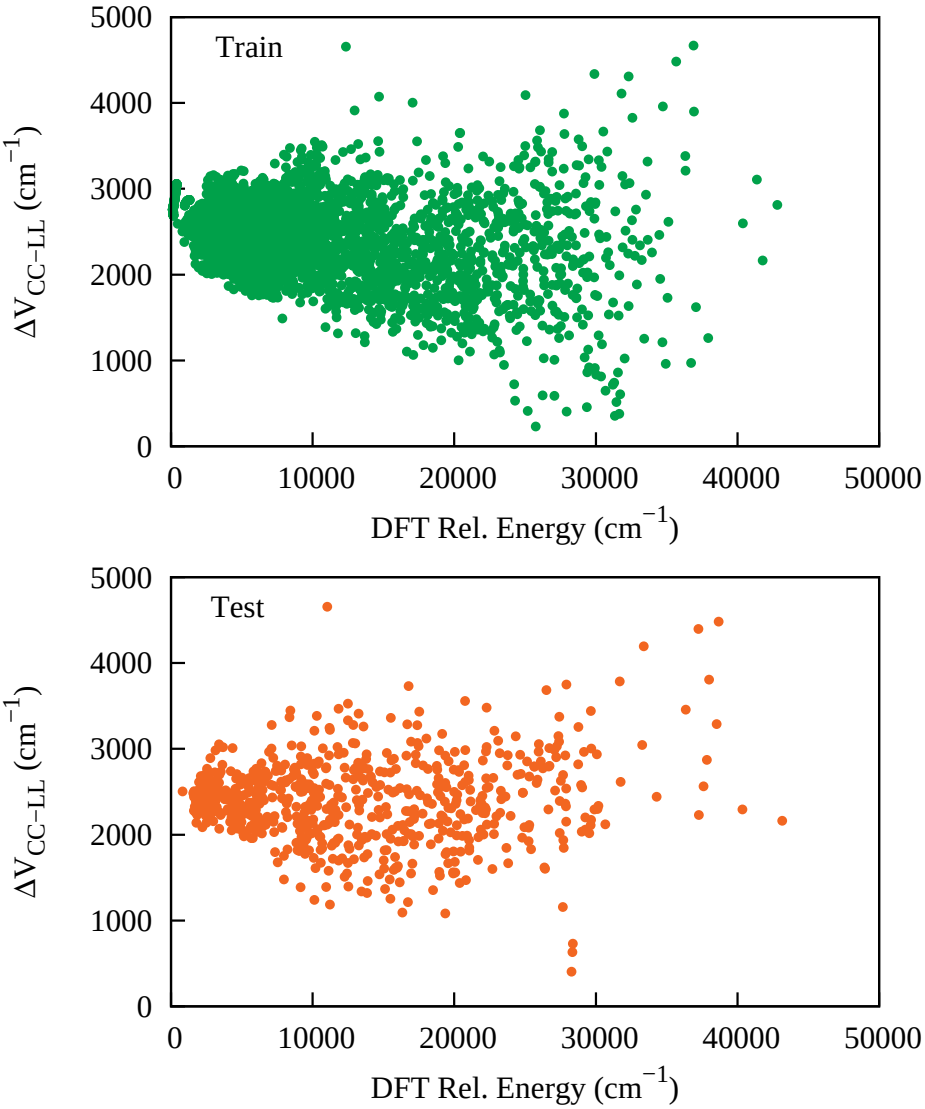


Figure 7.8: Plot of ΔV_{CC-LL} (relative to the reference value i.e. -50,200 cm^{-1}) vs DFT energy relative to the N-methyl acetamide minimum value for both training and test data set.

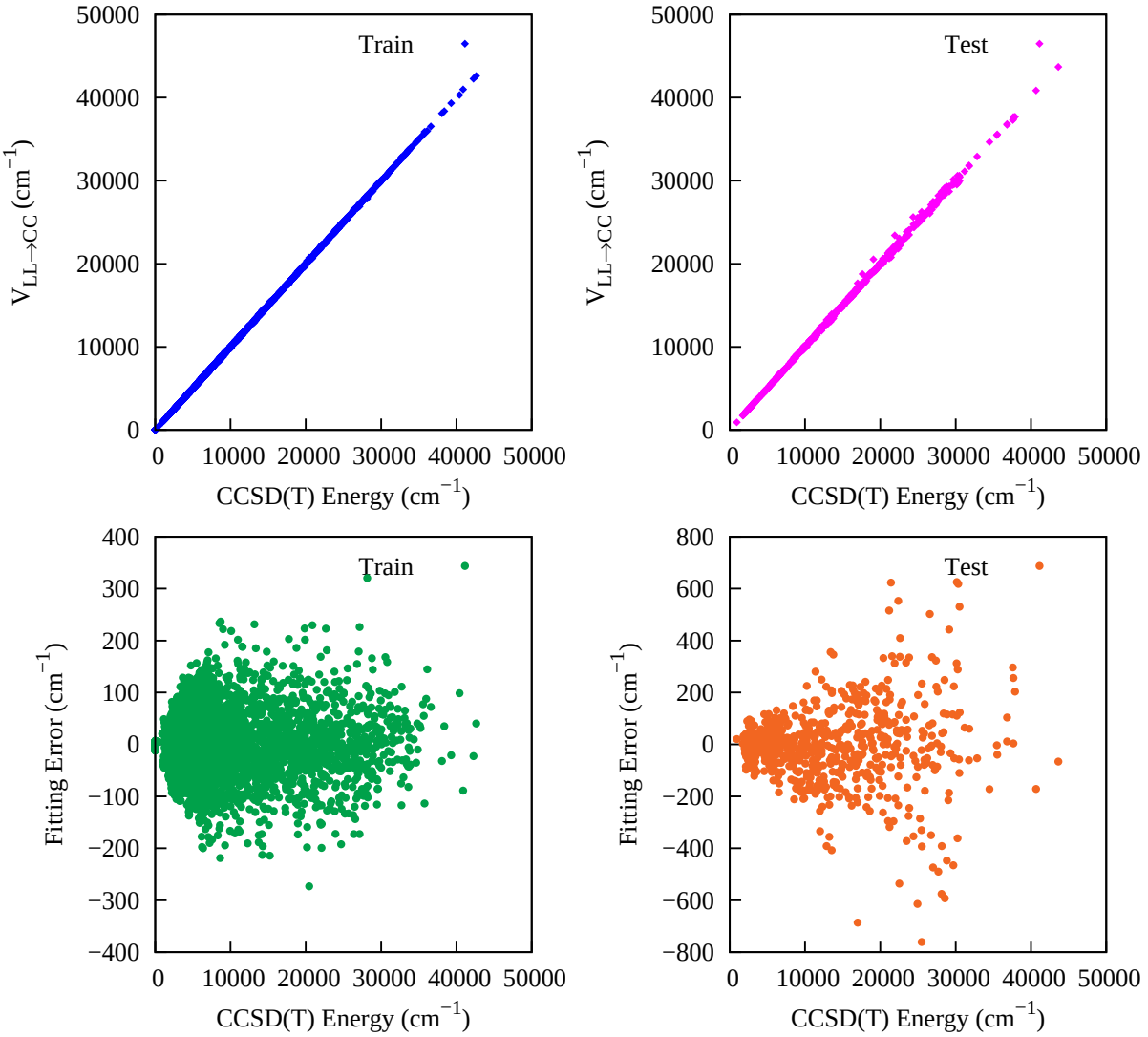


Figure 7.9: Two upper panels show energies of N-methyl acetamide from $V_{LL\rightarrow CC}$ vs direct CCSD(T) ones for the indicated data sets. The one labeled "Train" corresponds to the configurations used in the training of ΔV_{CC-LL} and the one labeled "Test" is just the remaining configurations. Corresponding fitting errors relative to the minimum energy are given in the lower panels.

relative to the CCSD(T) structure. These differences in structure are shown in the Figure. 7.10, while more discussion of the torsional barriers is given below. We also compare all the inter-nuclear distances for both *trans* and *cis* isomers with the direct CCSD(T)-F12 optimized geometry. The RMS error of differences, δ in bond lengths for Δ-ML PES relative to the direct CCSD(T)-F12 is 0.0037 and 0.0410 angstroms for *trans* and *cis* minimum, respectively. These RMS errors are 0.3349 and 0.0446 angstroms for DFT PES minimum structures relative to the direct CCSD(T)-F12 ones. It is expected that the RMS error for the *trans* isomer is quite big for DFT PES due to incorrect prediction of the minimum geometry.

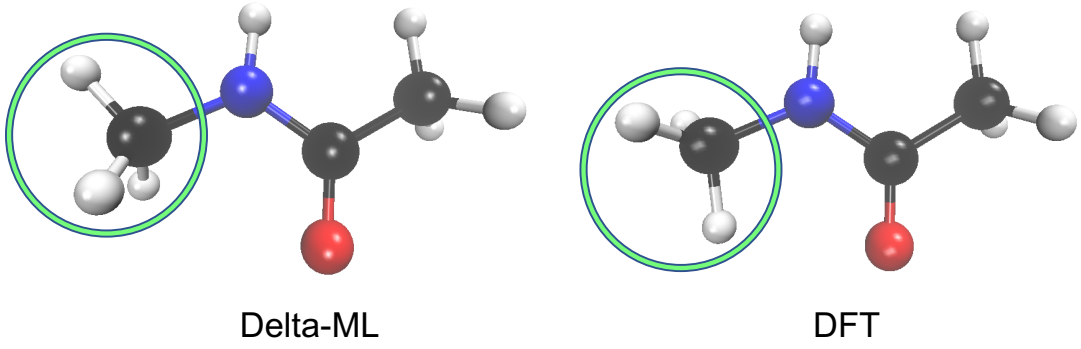


Figure 7.10: Comparison of *Trans*-NMA minimum geometry.

The *cis*-*trans* energy difference on the corrected PES is 782 cm^{-1}, which is 41 cm^{-1} below the direct CCSD(T) one. The RMS errors of harmonic frequencies between direct CCSD(T) one and the Δ-ML one are 15 and 13 cm^{-1}, respectively, for *trans* and *cis* isomers, whereas, these are 26 and 17 cm^{-1} for the DFT PES. The geometry differences are comparably small for the *cis*-isomer but large for the DFT PES for the *trans*-isomer, owing mainly to the error in the methyl rotor minimum on the DFT PES, noted already.

We also performed normal mode analyses for both *trans*- and *cis*-NMA. Comparisons of the harmonic frequencies for *trans*- and *cis*-NMA with their corresponding direct CCSD(T)-F12 ones and the DFT PES ones are given in Tables 7.6 and 7.7. It is seen that the agreement between Δ-ML PES and CCSD(T)-F12 is very good, specially for the high frequency modes. Note that we get two imaginary frequencies for the low frequency modes of *trans* isomer which are $124i$ and $144i$ from the direct CCSD(T)-F12 calculation. These are due to two methyl rotors which are almost free rotors for the *trans* isomer. Therefore, we also performed geometry optimization and normal mode analysis for the *trans* isomer at CCSD(T)/AVDZ level of theory. The frequencies we obtained at CCSD(T)/AVDZ level of theory are very similar to the CCSD(T)-F12 ones and also get one imaginary frequency which is $20i$.

Table 7.6: Comparison of normal mode frequencies (in cm^{-1}) between Δ-ML PES, DFT PES and the direct CCSD(T)-F12 ones for *trans*-NMA.

Mode	CCSD(T)-F12	Δ-ML	DFT-PES	Mode	CCSD(T)-F12	Δ-ML	DFT-PES
1	$124i$	34	38	16	1397	1415	1391
2	$144i$	52	119	17	1457	1465	1415
3	116	133	151	18	1468	1471	1434
4	251	265	290	19	1481	1498	1474
5	316	330	393	20	1481	1509	1485
6	416	426	433	21	1511	1554	1491
7	618	621	621	22	1563	1582	1552
8	624	633	629	23	1753	1741	1772
9	876	894	867	24	3045	3042	3019
10	996	1005	995	25	3050	3051	3040
11	1045	1056	1038	26	3110	3098	3072
12	1110	1107	1113	27	3133	3133	3125
13	1155	1153	1132	28	3137	3142	3127
14	1194	1180	1167	29	3140	3147	3137
15	1290	1283	1260	30	3681	3637	3631

Table 7.7: Comparison of normal mode frequencies (in cm^{-1}) between Δ-ML PES, DFT PES and the direct CCSD(T)-F12 ones for *cis*-NMA.

Mode	CCSD(T)-F12	Δ-ML	DFT-PES	Mode	CCSD(T)-F12	Δ-ML	DFT-PES
1	59	36	76	16	1413	1420	1393
2	120	131	123	17	1465	1468	1439
3	163	179	166	18	1472	1494	1447
4	285	287	286	19	1483	1497	1462
5	468	459	468	20	1492	1500	1467
6	504	501	518	21	1503	1522	1478
7	574	575	575	22	1530	1560	1508
8	605	609	624	23	1766	1777	1798
9	818	829	815	24	3025	3020	3001
10	996	992	977	25	3043	3066	3040
11	1051	1052	1043	26	3095	3075	3069
12	1098	1097	1087	27	3112	3114	3107
13	1155	1156	1137	28	3137	3131	3116
14	1211	1207	1187	29	3169	3171	3155
15	1349	1350	1330	30	3633	3602	3592

Finally, the most exciting results are the torsional potentials of the two methyl rotors of both *trans* and *cis*-isomers of NMA. We get significant improvement from our previously reported DFT PES results.[145] Detailed comparisons of the partially relaxed torsional barriers are given in Table 7.8. As seen, there are large differences between the DFT PES and CCSD(T) results for the CH_3−NH rotors for both *cis* and *trans* isomers. Overall, the Δ-ML PES barriers are significantly closer to the CCSD(T) ones than the DFT-PES ones.

Given the error in these DFT PES barriers, a detailed examination of the torsional potentials is warranted. These are shown in Fig. 7.11. These appear as would be expected, with the exception of panel a), where the Δ-ML potential has a small dip at 60 deg, instead of a barrier there. The barrier of 34 cm^{-1} given in Table 7.8 is thus at slightly

Table 7.8: Comparison of torsion barriers of methyl rotors, CH_3-NH and CH_3-CO (cm^{-1}) for *trans* and *cis* isomers of *N*-methyl acetamide.

trans-NMA	CH_3-NH	CH_3-CO
DFT PES	256	37
Δ-ML PES	34	74
CCSD(T)	42	103
cis-NMA	CH_3-NH	CH_3-CO
DFT PES	61	361
Δ-ML PES	153	366
CCSD(T)	148	303

the wrong location. The source of this offset is the large error in the DFT PES, which has a minimum 60 deg in error compared to the benchmark CCSD(T) result. The small artifact in the Δ-ML torsional potential is of minor consequence given that the CCSD(T) barrier is only 42 cm^{-1}.

To the best of our knowledge there is no experimental determination of these torsional barriers for either isomer of NMA. However, there is a report of the torsional barrier for acetamide of 24 cm^{-1}.[150] This barrier is consistent with the small barriers of 34 cm^{-1} (Δ-ML PES) and 74 cm^{-1} (Δ-ML PES) for *trans*-NMA. Also, it appears that the larger barriers for *cis*-NMA may be due to the closer proximity of these methyl rotors.

Next we make some comments about computation times on our cluster with Intel Xeon 2.40 GHz processors. First, to calculate the 5430 CCSD(T) energies required about 900 cpu-hours. (This was done using multiple nodes.) The time for 100 000 calculations of the corrected PES, $V_{LL\rightarrow CC}$, is the sum of 2.056 seconds for the DFT PES, V_{LL}, plus 0.126 seconds for the ΔV_{CC-LL} PES. Thus, the ΔV_{CC-LL} PES takes only 6% of the total cpu time.

To conclude this section, we note that preliminary work indicates that using about

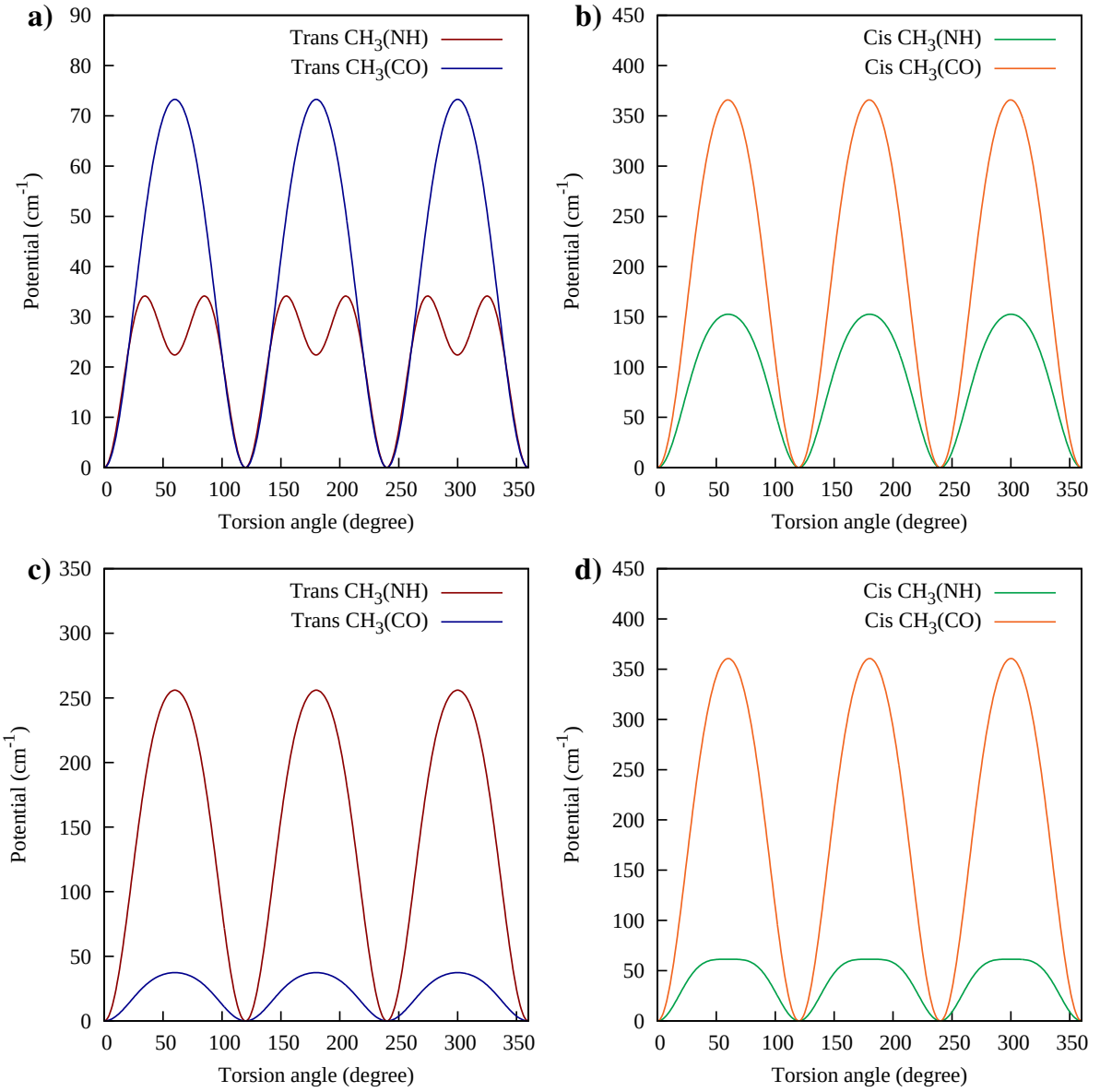


Figure 7.11: Torsional potentials (not fully relaxed) of the two methyl rotors of both *trans* and *cis*-NMA from Δ-ML PES a) and b), and DFT PES c) and d). Note, for the torsion indicated in red in c), the zero angle corresponds to a structure that is rotated by 60 deg relative to the corresponding and correct CCSD(T) torsional potential.

half the number of CCSD(T) energies, i.e., 2200 energies, produces a ΔV_{CC-LL} PES that is close to the quality of the one reported here. We plan to report the details of this along with even smaller data sets later.

7.4.4 Timings

Tables 7.9 and 7.10 show the computation time for evaluating the low-level DFT-based PES and the correction PES, ΔV_{CC-LL}. These calculations are carried out on a single core of Intel Xeon 2.40 GHz processor-based machines with 64 GB RAM. Clearly, computation of ΔV_{CC-LL} is much faster when the training data is 300 or 100 for CH_4 and 250 or 125 for H_3O^+.When the training data is large, the number of PIP basis functions is the same for both the V_{LL} and ΔV_{CC-LL} PES and, therefore, we do not see any time advantage. However, with decreasing size of the training data we reduce the maximum polynomial order and we get a great time advantage for the ΔV_{CC-LL} PES. Thus, the additional cost to bring the DFT-based PES to CCSD(T) level of accuracy is a small fraction of the cost of evaluating the DFT PES.

Table 7.9: Timings (sec) for 100 000 PES evaluations for CH_4.

N_{Train}	V_{LL}	ΔV_{CC-LL}	$V_{LL\rightarrow CC}$
1000[a]	0.07	0.07	0.14
600[a]	0.07	0.07	0.14
300[b]	0.07	0.04	0.11
100[c]	0.07	0.02	0.09

[a] Maximum polynomial order of 5, basis size of 208.
[b] Maximum polynomial order of 4, basis size of 83.
[c] Maximum polynomial order of 3, basis size of 30.

Table 7.10: Timings (sec) for 100 000 PES evaluations for H_3O^+.

N_{Train}	V_{LL}	ΔV_{CC-LL}	$V_{LL\rightarrow CC}$
1000[a]	0.04	0.04	0.08
500[b]	0.04	0.03	0.07
250[c]	0.04	0.02	0.06
125[d]	0.04	0.01	0.05

[a] Maximum polynomial order of 7, basis size of 348.
[b] Maximum polynomial order of 6, basis size of 196.
[c] Maximum polynomial order of 5, basis size of 103.
[d] Maximum polynomial order of 4, basis size of 51.

7.5 Summary and Conclusions

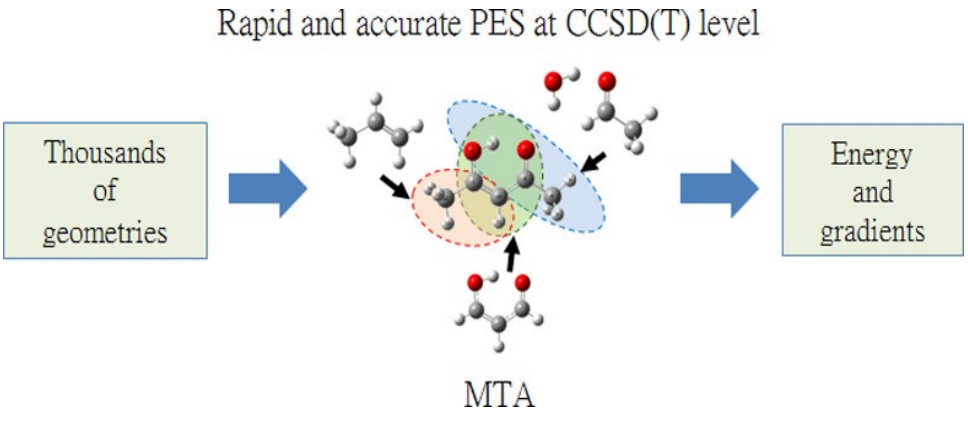
We reported an efficient and easy-to-implement correction to a low-level DFT PES based on a low-order PIP fit to the difference in a sparse set of high-level CCSD(T) and DFT energies. The correction was shown to produce a final PES with properties that are close to the corresponding CCSD(T) benchmark values for CH_4 and H_3O^+. Similar results were shown for *N*-methyl acetamide and this demonstrates that the approach should be widely applicable to large molecules. We plan to do this in the future for acetylacetone and tropolone, for which low-level PESs have recently been reported.[142,146,147] However, it would be difficult to present the rigorous tests against high-level coupled cluster results for say harmonic frequencies as these require a very large computational effort.

Finally, we note that the low-level PES can be based on any fitting method as can the correction PES. However, both should be consistent with respect to the same level of permutational invariance. We believe the PIP approach has advantages for the correction PES. One is that the fit is permutationally invariant and another, and perhaps more significant one, is that a low-order PIP fit can be both precise and efficient to evaluate.

Chapter 8

Rapid and Accurate Construction of CCSD(T)-level PES of large molecules using Fragmented Method

8.1 Chapter Abstract



The construction of a potential energy surface (PES) of even a medium-sized molecule employing correlated theory, such as CCSD(T), is arduous due to the high computational cost involved. The present study reports the possibility of efficiently constructing such a PES of molecules containing up to 15 atoms and 550 basis functions by employing the fragment-based molecular tailoring approach (MTA) on off-the-shelf hardware. The MTA energies at the CCSD(T)/aug-cc-pVTZ level for several geometries of three test molecules, viz., acetylacetone, *N*-methylacetamide, and tropolone, are reported. These energies are in excellent agreement with their full calculation counterparts with a time advantage factor of $3-5$. The energy barrier from the ground to transition state is also accurately captured. Further, we demonstrate the accuracy and efficiency of MTA for estimating the energy gradients at the CCSD(T) level. As a further application of our MTA methodology, the energies of acetylacetone at $\sim$430 geometries are computed at the CCSD(T)/aug-cc-pVTZ level and used for generating a Δ-machine learning (Δ-ML) PES. This leads to the H-transfer barrier of 3.02 kcal/mol, well in agreement with the benchmarked barrier of 3.19 kcal/mol. The fidelity of this Δ-ML PES is examined by geometry optimization and normal mode frequency calculations of global minima and saddle point geometries. We trust that the present work is a major development for the rapid and accurate construction of PES at the CCSD(T) level for molecules containing up to 20 atoms and 600 basis functions using off-the-shelf hardware.

8.2 Overview

Developing a high-dimensional potential energy surface (PES) of molecules based on advanced *ab initio* methods is an active theoretical and computational chemistry research area. Within the Born-Oppenheimer approximation, one can get the PES "on-the-fly"

by obtaining the molecular energies from electronic structure packages. However, the computational cost quickly becomes prohibitive as the level of electronic structure theory and the system size go up. An alternative approach is to develop an accurate analytical representation of PES using data sets of electronic energies that span a high-dimensional space. As the PES establishes a relationship between a nuclear configuration and the forces acting on nuclei, it has wide chemical applications. The PESs thus constructed have been employed in the literature for several types of reactions dynamics calculations, harmonic-anharmonic vibrational analysis, geometry optimization, etc.

In the past 15 years, significant progress has been made in developing non-parametric, machine-learning approaches for fitting large data sets of electronic energies for molecules and clusters that contain more than four atoms. Three major methods recently in widespread use for this purpose are: permutationally invariant polynomials (PIPs), Neural Networks (NN), and Gaussian Process Regression (GPR) as well as their combinations. Several reviews have summarized these approaches.[14,16,22,26,30,35,65,151,152] However, it is a major challenge to extend these methods to larger molecules of interest in chemistry and physics.

Increasing the molecular size creates a bottleneck for developing the PES at high-level theory. Due to the very high scaling of the "gold standard" coupled-cluster with perturbative triples [CCSD(T)] theory ($\sim$N7, N being the number of basis functions), its use is generally limited to the molecules containing $\sim$10 atoms. A large data set of energies is required for high-dimensional PES, and this data set requirement rapidly increases with the increasing number of atoms. Therefore, researchers have been focusing on the simultaneous energy plus gradient fitting method to reduce the number of data points substantially.[141] A significant advantage of using energy plus gradient data is found for small configuration data sets. The PES thus constructed exhibits equivalent fidelity to that obtained using only energy data sets, which are an order of magnitude larger.

Lower-level electronic structure methods such as density functional theory (DFT) and second-order Møller-Plesset perturbation theory (MP2) are typically used to deal with the PES of larger molecules. The PES of molecules having more than 10 atoms using the CCSD(T)/aug-cc-pVTZ (CCSD(T)/aVTZ) level of theory is generally conspicuous by their absence. In 2016, Bowman and co-workers computed the PES for the formic acid dimer $(HCOOH)_2$, a 10-atom system, using 13475 energies at CCSD(T)-F12/haTZ for H atom and aVTZ for C and O atoms' level of theory.[153] This PES was applied for zero-point energy (ZPE) computation using the diffusion Monte-Carlo (DMC) method and ground-state tunneling splitting for the H-transfer process. A nine-atom PES for the chemical reaction $Cl+C_2H_6$ was recently reported by Papp et al. using a composite MP2/CCSD(T) method.[154] Examples of potentials for six- and seven-atom chemical reactions, which are fit to tens of thousands or even hundred thousand CCSD(T) energies, have also been reported.[155,156] Recently, a 15-atom PES was calculated for acetylacetone, containing seven heavy atoms.[157] This was a major computational effort at the LCCSD(T)-F12/cc-pVTZ-F12 level of theory. This PES was obtained with 2151 LCCSD(T) energies using the Δ-machine-learning approach.

As pointed out earlier, developing the PES using CCSD(T) level of theory for molecules containing more than 10 atoms is prohibitively expensive. It also requires a larger dataset with the increasing dimensionality of the system, often leading to the use of lower-level methods such as DFT and MP2. However, the DFT and MP2 methods do not yield accurate results for reaction barriers, geometrical parameters, and vibrational frequencies compared to their CCSD- or higher counterparts. Thus, there is a need to generate a PES based on CCSD or CCSD(T) level calculations.

Quantum chemical calculations on molecular systems provide reliable insights into their structural parameters and reactivity. Though accurate, the methods capturing a

substantial part of electron correlation (such as MP2 or CCSD(T)) have a formidable scaling, leading to the requirement of high processor speed, memory, and secondary storage. This makes these methods unsuitable for systems comprising ∼100 atoms (MP2) or ∼20 atoms (CCSD(T)). In order to circumvent this high-scaling problem, several fragmentation-based (FB) methods have been proposed during the last two decades for treating large molecules. These methods have been extensively tested and benchmarked for obtaining the molecular energies, gradients, and Hessian matrix. An overview of such FB methods has been given in Ref. 158. One of such FB methods, viz. molecular tailoring approach (MTA), has been developed indigenously by us and appraised for various large molecules.

The MTA methodology was proposed, developed, and tested out on a variety of molecules/clusters by Gadre and co-workers.[159–164] The MTA framework is currently applicable to closed-shell systems for estimating the energies/energy gradients/Hessian matrix. Geometry optimization followed by IR/Raman spectral calculation is also available within the MTA. Here I briefly capture the essential features of MTA below.

Within MTA, a spatially extended, closed-shell parent molecule under investigation is divided into a series of primary and overlapping fragments. These fragments are subjected to computation instead of the entire parent molecule. Then, the MTA program 'patches' the results (e.g., Energy, E) of the fragments to get a good approximation to the property of the parent molecule (vide Eq. 8.1) employing set inclusion/exclusion principle.

$$E = \sum_{i} E^{F_i} - \sum_{i<j} E^{F_i \cap F_j} + \cdots + (-1)^k \sum_{i<j<\cdots<n} E^{F_i \cap F_j \cap \cdots \cap F_n} + \cdots. \tag{8.1}$$

Here, E^{F_i} represent the energy of i^{th} primary fragment; $E^{F_i \cap F_j}$ is the energy of the overlap fragment between the primary fragments i and j. The term n represents the number of

primary fragments, and k is the degree of overlap.

On account of this, the required computational expenses/resources for estimating a property of parent molecule are steeply reduced. The MTA procedure makes initial fragments by putting a sphere on all non-hydrogen atoms of the parent molecule. Then, these fragments are merged based on distance, with subsequent rounds of fragmentation. Thus, the near-neighbourhood of an atom is preserved in at least one of the main fragments. The quality of the results can be gauged a priory by the distance-based parameter called R-Goodness (RG). In general, the larger the RG, the better the preservation of the atoms' chemical environment. For more details about MTA's origin, fragmentation procedure, and capabilities, as well as the limitations of MTA, readers are directed to Ref. 159–163.

In general, the MTA-based energies carry an error of several milli-hartrees vis-á-vis the respective full calculation (FC) energies. In 2012, a grafting procedure was devised and thoroughly benchmarked[161,162] in order to reduce these error arising due to the fragmentation. Within the version of grafting procedure used here for calculating the energy,[161] the difference between the FC and MTA correlation energies computed with lower basis (LB) is added to the sum of MTA correlation energy at the higher basis (HB) and the Hartree-Fock (HF) FC energy at HB. All the MTA calculations are done using an identical fragmentation scheme, vide Eq. 8.2. I call such grafted energy (cf. Eq. 8.2) as the GMTA energy in the following discussion.

$$E_{GMTA} = E(HF)^{HB}_{FC} + E(CORR)^{HB}_{MTA} + (E(CORR)^{LB}_{FC} - E(CORR)^{LB}_{MTA}). \qquad (8.2)$$

Here, $\mathrm{E(HF)}^{HB}_{FC}$ refers to the FC HF energy computed at HB whereas $\mathrm{E(CORR)}^{HB}_{MTA}$ is MTA correlation energy computed at HB. The correlation energies in FC at LB and MTA at LB are denoted respectively as $\mathrm{E(CORR)}^{LB}_{FC}$ and $\mathrm{E(CORR)}^{LB}_{MTA}$. The grafting

correction is proposed based on our earlier observation that the difference between MTA and FC energies is almost independent of the basis set used.[164] As expected, this correction leads to a better estimation of the total energy and electronic properties. The grafting procedure is now built into the MTA code. The final energy printed out in the MTA code already incorporates the grafting correction. Because of this, in the discussion that follows, I denote the grafted MTA energy as E_{MTA}.

With this background, it is felt worthwhile to explore the MTA method for constructing a PES at the CCSD(T) level with augmented basis sets using off-the-shelf hardware. The computational details of these exploratory studies are given in the following Section.

8.3 Computational Details

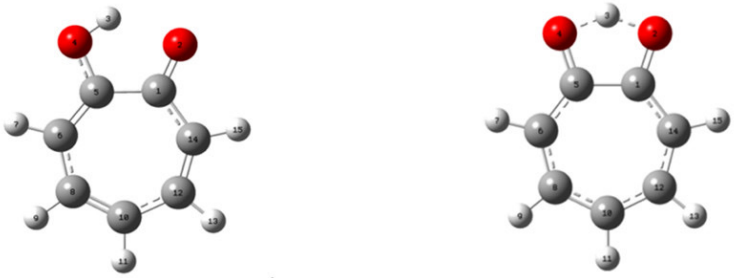
The attempt is made for the evaluation of energy of a few trial geometries on the PES for three test molecules, viz. acetylacetone (AA), *N*-methyl acetamide (NMA), and tropolone with the use of MTA. The AA, NMA, and tropolone molecules contain 15, 12, and 15 atoms, respectively. The starting ground- and transition state geometries and a few other geometries of these molecules are adopted from the articles by Bowman and coworkers.[145,165,166] Because of the enormous computational cost, they have used a modest level of theory and/or basis set. It would be worthwhile to carry out these calculations with a high level of theory and a large basis set for a meaningful comparison with the experimental findings. Therefore, these geometries (ground- and transition state) are further subjected to the geometry optimization at the MP2/aug-cc-pVTZ(MP2/aVTZ) level of theory. The number of basis functions for the aVTZ basis set associated with AA, NMA, and tropolone is 506, 391, and 552, respectively. The optimized structures of the test molecules are depicted in Figure 8.1. The single point (SP) FC energy eval-

(a) Acetylacetone (AA); Global minimum(GM) (left) and the Transition State (TS) (right)

trans-NMA cis-NMA

TS1-NMA TS2-NMA

(b) *trans* and *cis* structures of N-methylacetamide (NMA) and the two transition states



(c) Tropolone: Global Minimum (GM) (left); Transition State (TS) (right)

Figure 8.1: Geometrical structures with the corresponding transition states of (a) acetylacetone (AA), (b) *N*-methylacetamide (NMA), and (c) tropolone molecules optimized at the MP2/aVTZ level of theory. See text for details.

uations at MP2 and CCSD(T) levels using the aVTZ basis are carried out for these geometries. Moreover, the harmonic vibrational frequency computations are also done at the MP2/aVTZ level of theory to confirm their minimal nature. All the computations are done using the Gaussian suite of programs[167] using the default frozen core option, on a 16-core Hewlett-Packard (HP) server-grade machine unless otherwise indicated.

Being small in size, the molecules under study are subjected to manual fragmentation into the main (M) and overlapping (O) fragments, which are used for MTA computations. The fragmentation schemes for global minimum (GM) geometries of the test molecules are as shown in Figure 8.2.

In Figure 8.2, the main fragments are highlighted in light red, blue, and green ovals labeled as M1, M2, etc. The respective overlap fragments are labeled as O1, O2, etc. There are a maximum of three/four main fragments with an average size of half of the molecule and two/three overlapping ones. An identical fragmentation scheme is employed for the TS geometries as well.

The single point MTA energy evaluation for these geometries at MP2/aVTZ and CCSD(T)/aVTZ level are carried out using these fragments. The grafting correction is done employing cc-pVTZ basis set at the respective level of theories. Apart from GM and TS geometries, two other geometries of each test molecule are also explored to compare the MTA energies with their FC counterparts.

The SP energy gradient calculations at MP2/aVTZ and CCSD(T)/aVDZ levels are also carried out for one of the geometries (which is not a stationary point on the PES) of each of the three test molecules, with a view to benchmark the accuracy and efficiency of MTA. The comparison of the gradients at only one prototype geometry for each of the test molecules at FC and MTA levels is reported. A computation for estimating gradients at the CCSD(T)/aVTZ level for the prototype geometry of AA is also carried out. Since

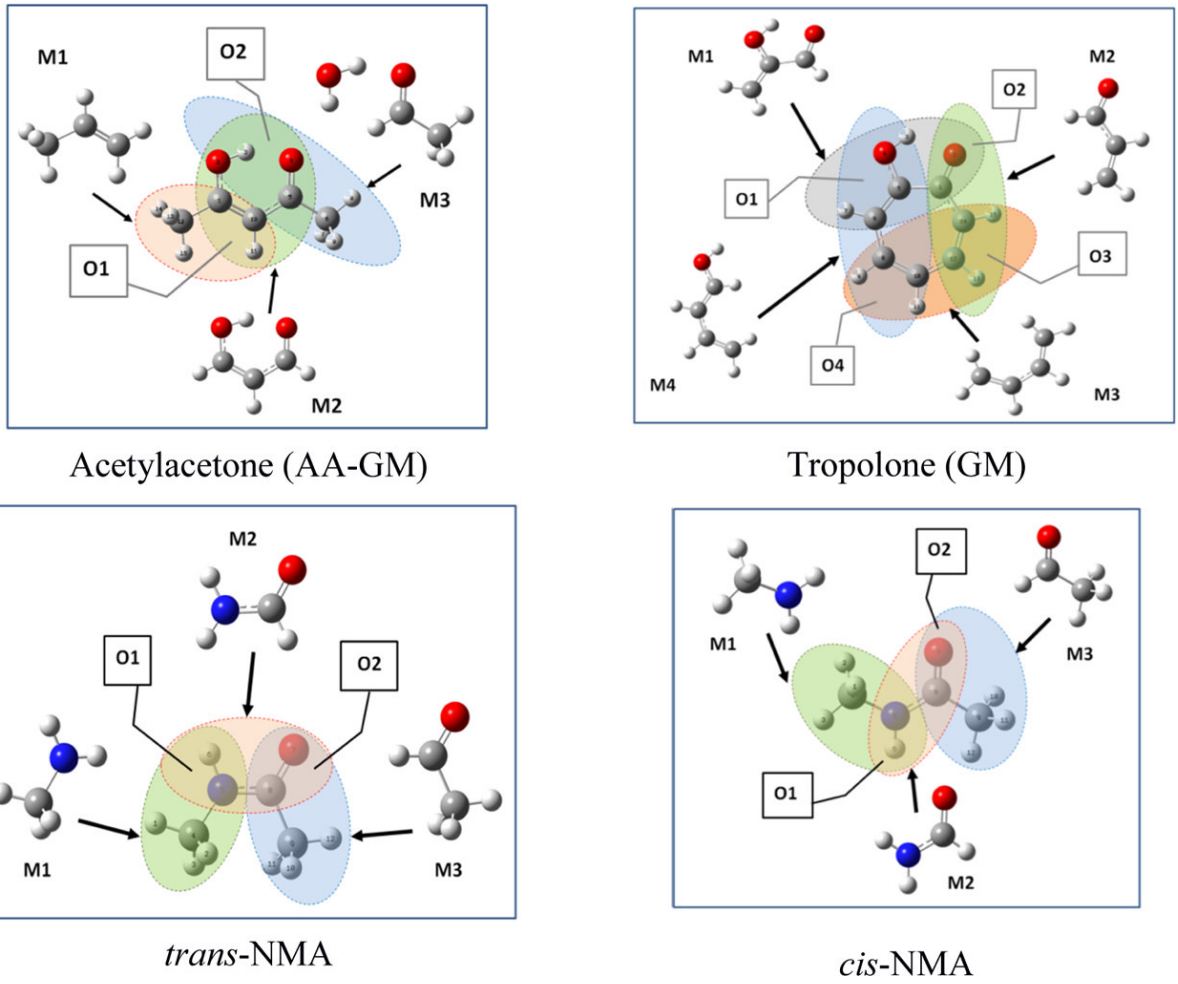


Figure 8.2: Fragmentation schemes implemented in MTA calculations for Acetylacetone (AA-GM), Tropolone (GM), trans-NMA, and cis-NMA. See text for details.

analytical gradients for CCSD(T) level theory is not available in the Gaussian package,[167] the numerical gradients (as implemented in the Gaussian) are used in the present study.

I report the appraisal of the accuracy and efficiency of MTA for reproducing the respective FC energies and energy gradients for three chosen test molecules. The results of these calculations are presented in the next Section.

8.4 Results and Discussion

Table 8.1: Single-Point Full Calculation(E_{FC}) and Grafted MTA (E_{MTA}) Energies[b] (in a.u.) Using aVTZ Basis for NMA Molecule Using MP2/aVTZ-Optimized Geometries, along with the (Signed) Error[a].

	MP2			CCSD(T)		
Geometry	E_{FC}	E_{MTA}	Δ	E_{FC}	E_{MTA}	Δ
Cis	-248.06688	-248.06625	-0.00063	-248.13930	-248.13875	-0.00054
Trans	-248.07080	-248.07024	-0.00055	-248.14308	-248.14263	-0.00045
TS1-NMA	-248.04284	-248.04218	-0.00066	-248.11685	-248.11630	-0.00054
TS2-NMA	-248.03753	-248.03695	-0.00058	-248.11169	-248.11123	-0.00046
N1	-248.06834	-248.06791	-0.00043	-248.14116	-248.14081	-0.00035
N2	-248.05584	-248.05547	-0.00037	-248.12912	-248.12856	-0.00055
	(1.5)	(2.5)		(93)	(29)	

[a] The respective wall clock timings (min) are given in parentheses. See text for details.
[b] The MTA energies are calculated employing eq 8.2.

The SP energies at MP2 and CCSD(T) for MP2/aVTZ optimized geometries of the NMA molecule (both *cis*- and *trans* isomers as well as two additional geometries N1, and N2) are displayed in Table 8.1 and also reports the signed error, $\Delta = E_{FC} - E_{MTA}$, at the corresponding level of theory, bringing out the excellent agreement of MTA energy

with its FC counterpart, the error being uniformly negative and numerically less than 0.7 mH. The wall-clock time shows a significant advantage (a factor of ∼3 or so) for MTA vis-á-vis the FC for computing the energies at CCSD(T) level. Note that for small test systems, no time advantage is seen in the case of MP2 computations. However, a small time advantage is noticed for the largest system, viz. tropolone (cf. Table 8.2), with the augmented basis sets employing MP2 method.

Table 8.2: Single-Point Full Calculation(E_{FC}) and Grafted MTA (E_{MTA}) Energies (in a.u.) for the Enol Forms of Acetylacetone (AA) and Tropolone Molecules Estimated Using aVTZ Basis Set[a].

	MP2			CCSD(T)		
Geometry	E_{FC}	E_{MTA}	Δ	E_{FC}	E_{MTA}	Δ
			Enol Form of Acetylacetone (AA)			
GM	-345.17691	-345.17641	-0.00050	-345.27104	-345.27056	-0.00048
TS	-345.17343	-345.17299	-0.00044	-345.26610	-345.26567	-0.00043
A1	-345.09946	-345.09910	-0.00036	-345.19219	-345.19181	-0.00038
A2	-345.09048	-345.09023	-0.00025	-345.18336	-345.18290	-0.00045
	(6)	(6)		(740)	(172)	
			Tropolone			
GM	-420.00707	-420.00667	-0.00040	-420.10670[b]	-420.10642	-0.00028
TS	-420.00049	-420.00067	0.00018	-420.09604[b]	-420.09600	-0.00004
T1	-419.92486	-419.92624	0.00138	-420.02997[b]	-420.02948	-0.00049
T2	-419.98716	-419.98826	0.00110	-420.08595[b]	-420.08663	0.00067
	(21)	(18)		(2040)	(421)	

[a] The wall clock timings (min) are given in parentheses and the (signed) error is given as Δ. See text for details.
[b] Calculation is done on 16 cores of a node.

Likewise, for the other two test molecules, viz. the enol form of AA and tropolone, the MTA- and FC energies at MP2 and CCSD(T) using aVTZ basis set are in shown in

Table 8.2. The energy differences are seen to (numerically) reduce with the elevation of the level of theory[164] for all the test cases testifying our earlier observation. In addition, a time advantage factor between 4 and 5 is noticed for both the molecules at CCSD(T) level theory.

The energy barrier between the ground- and transition-state is also seen to be well-estimated by MTA. For example, in the case of AA molecule at CCSD(T)/aVTZ level, the FC barrier of 4.96 mH is well-captured by MTA (4.89 mH). Similarly, for tropolone, the FC energy barrier of 10.66 mH is also estimated quite well by the MTA one (10.42 mH).

As mentioned earlier, it is helpful to have the energy gradients for different geometries of a molecule (which are not the GM or TS) to build its PES. However, the cost of FC gradient calculation with a high theory/basis set level is prohibitively difficult. Because of this, an appraisal of MTA for estimating the energy gradients (GR) is carried out.

Since FC gradient calculation is readily possible at the MP2/aVTZ and CCSD(T)/aug-cc-pVDZ (aVDZ) levels, I have used these levels of theory for benchmarking purposes. Finally, Table 8.3 compares the gradients (GR) calculated using FC and MTA at CCSD(T)/aVDZ level theory for the prototype geometry of NMA. For this purpose, I use a simpler version, viz. overall grafting correction, viz.

$$GR_{GMTA} = GR^{HB}_{MTA} + (GR^{LB}_{FC} - GR^{LB}_{MTA}). \tag{8.3}$$

The FC and MTA CCSD(T)/aVDZ level gradient results for acetylacetone and tropolone are also displayed in Table 8.4 and 8.5. These gradient results show a substantial time advantage factor for MTA gradient calculation, retaining a very good accuracy, especially for the gradients with a magnitude greater than 0.005, vis-á-vis their FC counterparts.

Table 8.3: Comparison between Full Calculation (GR_{FC}) and MTA (GR_{MTA}) gradients (in a.u.) calculated at CCSD(T)/aVDZ basis for the geometry of *N*-methyl acetamide molecule. See text for details.

	N-methyl acetamide					
Atom	GR_{FC}			GR_{MTA}		
H	-0.0013	-0.0012	-0.0030	-0.0003	-0.0015	-0.0027
H	-0.0031	-0.0021	0.0086	-0.0021	-0.0024	0.0083
H	0.0016	0.0143	0.0015	0.0029	0.0138	0.0015
C	0.0098	-0.0146	-0.0085	0.0046	-0.0123	-0.0084
N	-0.0095	0.0222	0.0020	-0.0088	0.0213	0.0019
H	0.0018	-0.0171	0.0001	0.0020	-0.0162	0.0001
O	-0.0011	-0.0070	0.0013	-0.0012	-0.0060	0.0013
C	-0.0045	0.0090	-0.0029	-0.0033	0.0096	-0.0029
C	-0.0061	-0.0183	-0.0004	-0.0066	-0.0193	-0.0003
H	0.0061	0.0041	-0.0072	0.0061	0.0043	-0.0071
H	0.0051	0.0000	0.0055	0.0051	0.0002	0.0055
H	0.0012	0.0107	0.0029	0.0013	0.0107	0.0029
	0.0222[a]	**0.0081**[b]	**355**[c]	**0.0213**[a]	**0.0078**[b]	**95**[c]

[a] Maximum gradient
[b] RMS gradient
[c] Wall clock time (min)

Table 8.4: Comparison between Full Calculation (GR_{FC}) and MTA (GR_{MTA}) gradients (in a.u.) calculated at CCSD(T)/aVDZ basis for the geometry of Acetylacetone molecule. See text for details.

Acetylacetone						
Atom	GR_{FC}			GR_{MTA}		
C	0.0302	0.0150	0.0230	0.0310	0.0151	0.0233
O	-0.0864	-0.0085	0.0353	-0.0866	-0.0088	0.0350
H	0.0166	0.0054	-0.0123	0.0165	0.0060	-0.0118
O	-0.0300	0.0009	0.0076	-0.0309	0.0004	0.0076
C	-0.0618	-0.0741	0.0020	-0.0618	-0.0736	0.0028
C	0.0467	0.0405	-0.1342	0.0476	0.0400	-0.1346
H	-0.0005	0.0247	0.1502	-0.0009	0.0247	0.1500
H	-0.0210	-0.0028	-0.0098	-0.0212	-0.0028	-0.0098
H	0.0184	-0.0151	-0.0098	0.0181	-0.0152	-0.0095
C	0.1029	0.0433	-0.0194	0.1020	0.0432	-0.0198
H	-0.0078	-0.0053	-0.0005	-0.0078	-0.0053	-0.0004
C	-0.0361	-0.0850	0.0023	-0.0356	-0.0856	0.0017
H	0.0340	-0.0010	0.0066	0.0340	-0.0009	0.0068
H	0.0172	0.0070	-0.0224	0.0170	0.0069	-0.0223
H	-0.0223	0.0550	-0.0185	-0.0225	0.0551	-0.0187
	0.1502[a]	**0.0459**[b]	**2520**[c]	**0.1500**[a]	**0.0459**[b]	**322**[c]

[a] Maximum gradient
[b] RMS gradient
[c] Wall clock time (min)

Table 8.5: Comparison between Full Calculation (GR_{FC}) and MTA (GR_{MTA}) gradients (in a.u.) calculated at CCSD(T)/aVDZ basis for the geometry of Tropolone molecule. See text for details.

Tropolone						
Atom	GR_{FC}			GR_{MTA}		
C	-0.0802	0.0277	0.0043	-0.0809	0.0276	0.0052
O	0.0206	0.0134	0.0175	0.0207	0.0139	0.0175
H	-0.0688	-0.0450	-0.0181	-0.0689	-0.0447	-0.0182
O	-0.0843	-0.0109	-0.0060	-0.0823	-0.0110	-0.0060
C	0.0988	-0.0693	0.0070	0.0969	-0.0666	0.0071
C	0.1608	0.0533	0.0033	0.1596	0.0551	0.0017
H	-0.0282	0.0081	-0.0025	-0.0283	0.0081	-0.0026
C	-0.0039	-0.0284	-0.0092	0.0024	-0.0320	-0.0075
H	-0.0107	0.0003	0.0049	-0.0108	0.0003	0.0048
C	0.0113	0.0051	0.0145	0.0101	0.0098	0.0133
H	0.0150	0.0065	-0.0021	0.0150	0.0066	-0.0019
C	0.0130	0.0001	-0.0335	0.0116	-0.0012	-0.0334
H	-0.0301	0.0046	0.0189	-0.0302	0.0046	0.0188
C	-0.0170	0.0565	-0.0045	-0.0139	0.0562	-0.0041
H	0.0038	-0.0221	0.0054	0.0036	-0.0217	0.0055
	0.1608[a]	**0.0408**[b]	**3480**[c*]	**0.1596**[a]	**0.0405**[b]	**1470**[c]

[a] Maximum gradient
[b] RMS gradient
[c] Wall clock time (min)
[*] Using Advanced computational facility

Moving ahead, the MTA gradient calculation at CCSD(T)/aVTZ level of theory for the smallest test case i.e., NMA is reported in Table 8.6. The elapsed time for this computation is somewhat large, viz. 1636 min. Although benchmarking with the respective FC results is not feasible, I trust that the results are accurate, with a significant time advantage employing off-the-shelf hardware. The CCSD(T)/VTZ level of theory is used for grafting correction. As this is a prototype computation, a single 16 core machine was employed for this calculation. However, the multiple nodes can be easily harnessed for production jobs because MTA code is highly parallel.

Table 8.6: MTA Gradients (GR_{MTA}, in a.u.) Calculated at CCSD(T)/aVTZ Level of Theory for the Geometry (N1) of the *N*-methyl acetamide Molecule. See text for details.

Atom	GR_{MTA}		
H	0.0033	0.0006	0.0037
H	0.0019	0.0000	0.0017
H	0.0014	0.0055	0.0016
C	0.0060	-0.0115	-0.0083
N	-0.0056	0.0235	0.0019
H	0.0017	-0.0098	0.0003
O	-0.0013	-0.0239	0.0014
C	-0.0076	0.0168	-0.0030
C	-0.0048	-0.0182	-0.0007
H	0.0016	0.0019	-0.0001
H	0.0009	-0.0020	-0.0015
H	0.0020	0.0180	0.0032
	0.0239[a]	**0.0085**[b]	**1636**[c]

[a] Maximum gradient
[b] RMS gradient
[c] Wall clock time (min)

Table 8.6 brings out the possibility of applying the MTA for obtaining the energy

gradients at such high level of theory viz. CCSD(T) using aVTZ basis set on an off-the-shelf hardware.

8.4.1 New PES for Acetylacetone at CCSD(T)/aVTZ Level

As an application of the present MTA-based methodology, we develop a new full-dimensional MTA-based PES of AA at CCSD(T)/aVTZ level of theory using "Δ-machine learning (Δ-ML)" approach. The dataset of ∼430 geometries used for generating this PES is taken from previously reported work by Qu et al.[157] The overall grafting procedure (similar to that used in Eq. 8.3) was used for evaluating the energies these geometries. The basic idea of Δ-ML approach is to construct a high-level, coupled cluster level PES starting from a lower level MP2 one by using a correction PES. The details of this Δ-ML approach can be found elsewhere.[168] This correction PES is a fit to a small number of high-level ab initio energies that span the same range of configurations used to obtain the lower-level (LL) PES. The corrected CCSD(T)-level PES is denoted as $V_{LL\rightarrow CC}$, viz.

$$V_{LL\rightarrow CC} = V_{LL} + \Delta V_{CC-LL}, \tag{8.4}$$

where V_{LL} is the lower-level PES and ΔV_{CC-LL} is the correction PES. Here V_{LL} was previously reported PES which is a fit to 5454 energies and their corresponding gradients computed at MP2/aVTZ level of theory.[165] In this work, I develop a correction PES, ΔV_{CC-LL}, using only ∼430 electronic energies computed at CCSD(T)/aVTZ level of theory using MTA approach. This is a permutationally invariant polynomials (PIPs) fit using maximum polynomial order of 2 with 1 2 5 7 permutational symmetry (This symmetry indicates that the 7 'H' atoms are permutable with each other, as the 2 'O' atoms and 5 'C' atoms. The transferring atom H, atom 1, is treated as distinguishable.) which

leads to 86 linear coefficients. These coefficients are determined by standard linear least-square regression method. These PIP bases are generated using the MSA software.[123] I perform both weighted average and unweighted fitting for ΔV_{CC-LL}. The root means square errors (RMSE) for both weighted and unweighted fitting are 25 and 106 cm^{-1}, respectively. In order to examine the performance of the $V_{LL\rightarrow CC}$ PES, we use weighted averaged ΔV_{CC-LL} fitting only. Geometry optimization and normal mode frequency calculations are performed for both the global minimum (GM) and H-transfer saddle point (SP) geometries to examine the fidelity of the $V_{LL\rightarrow CC}$ PES. All the normal mode frequencies for GM and SP geometries are listed in Table 8.7 and 8.8 along with previously reported MP2 and LCCSD(T) frequencies. I get the symmetric double well H-transfer barrier as 1057 cm^{-1} or 3.02 kcal/mol, in good agreement with the previously reported[169] value of 1119 cm^{-1} or 3.19 kcal/mol. Though this Δ-ML PES slightly underestimates the barrier height, it is a significant improvement over the MP2-based PES,[165] which leads to a barrier height of 745 cm^{-1} or 2.13 kcal/mol.

The systems under consideration, though apparently rather small, are computationally heavy at the CCSD(T)/aVTZ level. Thus, FC gradient calculations for these molecules may be possible only by the investment of very large hardware and wall clock time. However, for spatially extended large molecule, MTA can indeed be a possible solution.

8.5 Summary and conclusions

The present work has reported some exploratory studies toward an application of MTA for the construction of PES of the three chosen test molecules at MP2- and CCSD(T)/aVTZ level of theory. The energies of ground- and transition states, as well as other some other

Table 8.7: Harmonic frequencies (in cm^{-1}) of the global minimum geometry of acetylacetone from the indicated sources.

	Global Minimum			
Mode	[a]Direct [at LCCSD(T)]	[a]$V_{LL\rightarrow CC}$ [at LCCSD(T)]	$V_{LL\rightarrow CC}$ [at CCSD(T)/aVTZ]	[a]V_{LL} [at MP2]
1	113	97	95	97
2	133	120	117	119
3	169	157	158	153
4	197	191	187	189
5	236	227	205	229
6	372	359	361	364
7	392	389	378	390
8	505	502	507	507
9	554	570	560	567
10	643	652	648	650
11	654	657	653	656
12	793	801	799	803
13	919	881	886	921
14	942	912	910	936
15	951	934	930	942
16	1010	1005	984	1014
17	1040	1044	1039	1048
18	1050	1052	1046	1058
19	1072	1069	1063	1071
20	1192	1192	1182	1200
21	1276	1270	1249	1290
22	1393	1377	1379	1384
23	1405	1404	1407	1399
24	1424	1433	1418	1433
25	1462	1450	1449	1470
26	1480	1486	1482	1494
27	1483	1489	1482	1497
28	1488	1499	1491	1505
29	1502	1508	1505	1512
30	1670	1662	1635	1655
31	1709	1705	1696	1704
32	3047	3058	3048	2855
33	3052	3080	3079	3095
34	3118	3122	3107	3099
35	3122	3170	3154	3178
36	3157	3178	3161	3187
37	3165	3203	3187	3208
38	3220	3210	3194	3218
39	3257	3248	3249	3258

[a] Ref. 157

Table 8.8: Harmonic frequencies (in cm^{-1}) of the H-transfer saddle point geometry of acetylacetone from the indicated sources.

	Global Minimum			
Mode	[a]Direct [at LCCSD(T)]	[a]$V_{LL\rightarrow CC}$ [at LCCSD(T)]	$V_{LL\rightarrow CC}$ [at CCSD(T)/aVTZ]	[a]V_{LL} [at MP2]
1	1278i	1082i	999i	921i
2	100	57	64	53
3	121	62	69	57
4	165	159	156	156
5	198	197	198	198
6	289	286	280	285
7	412	416	410	417
8	537	530	534	531
9	540	540	545	539
10	578	589	577	580
11	661	654	640	645
12	767	739	730	740
13	781	751	741	756
14	949	947	939	953
15	992	981	971	979
16	1035	1032	1004	1039
17	1037	1039	1035	1043
18	1054	1059	1052	1060
19	1067	1060	1054	1062
20	1195	1182	1168	1189
21	1308	1250	1219	1223
22	1341	1341	1313	1347
23	1406	1401	1397	1409
24	1413	1418	1417	1422
25	1481	1490	1483	1496
26	1487	1491	1483	1496
27	1488	1494	1486	1500
28	1491	1496	1489	1502
29	1569	1560	1550	1567
30	1613	1617	1574	1629
31	1624	1648	1649	1670
32	1904	1744	1698	1685
33	3054	3058	3045	3098
34	3057	3079	3075	3099
35	3130	3179	3156	3190
36	3132	3180	3156	3190
37	3154	3196	3177	3207
38	3156	3197	3177	3208
39	3241	3273	3274	3282

[a] Ref. 157

(non-stationary) geometries of these molecules are probed using the MTA methodology. For all the test molecules, the energetics are accurately estimated, with the typical error (vis-á-vis the respective FC energy) being typically smaller than 1 milli-hartree. The energy barriers from the ground- to transition-state also turn out to be accurate for all the three test cases. It may be mentioned that the FC jobs at CCSD(T)/ aVTZ level theory in this work, albeit at a high computational cost, are done only for demonstration purpose. Furthermore, the accuracy of the MTA for computing the energy gradients is critically assessed. All the MTA-based energy gradients at CCSD(T)/aug-cc-pVDZ level are efficiently and accurately calculated using a single 16-core machine. With the inherently parallel nature of MTA, we expect a significant time saving with the use of multiple machines in parallel. Substantial time-saving can also be gained using other efficient parallel ab initio packages, such as ORCA and CFOUR etc. The customization of MTA to accommodate these software packages is underway.

As per our experience, grafting from CCSD(T)/cc-pVTZ level works the best for MTA calculations at CCSD(T)/aVTZ level calculations. This necessitates the full calculation of the single point energy at CCSD(T)/cc-pVTZ level, which may be feasible for molecules containing up to 20 atoms/400 basis functions. For going beyond this threshold, a further detailed study is required for choosing the level of theory and basis-set for effecting the grafting correction.

The actual application for exploring a few hundred energy values on the PES for one of the test molecules at CCSD(T)/aVTZ level theory is computationally rather expensive, but indeed possible with the use of MTA. A full dimensional MTA-based PES for the Acetylacetone molecule is developed using $\sim$430 CCSD(T)/aVTZ energies with Δ-ML approach. We obtain the H-transfer barrier height as 3.02 kcal/mol, in excellent agreement with the benchmarked value of 3.19 kcal/mol. Recently, Qu et al. published

a full dimensional AA PES at LCCSD(T)-F12/cc-pVTZ-F12 level of theory, showing a barrier height of 3.49 kcal/mol.[157] Note that present results are at CCSD(T)/aVTZ level with no local approximation.

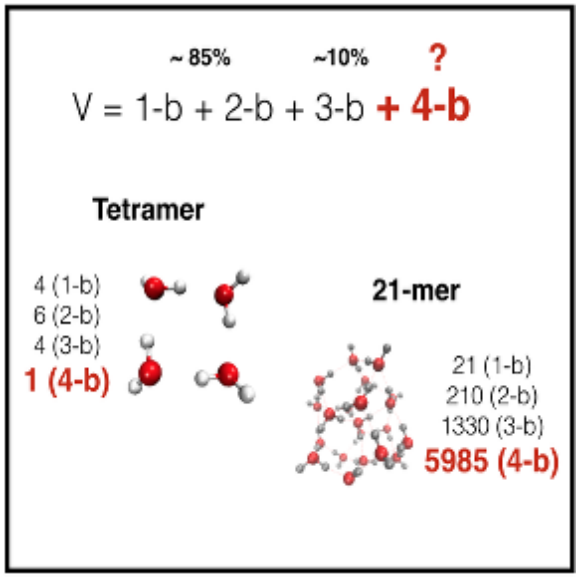
In summary, the present work has clearly demonstrated an efficient and economic method for constructing the PES, which can be readily extended to molecules containing up to 20 (first-row) atoms/600 basis functions using high level *ab initio* methods such as CCSD(T).

Chapter 9

Development of a CCSD(T)-based 4-body Potential for Water

9.1 Chapter Abstract

High-level, *ab initio* calculations find that the 4-body interaction is needed to account for near-100 percent of the total interaction energy for water clusters as large as the 21-mer. Motivated by this, we report a permutationally invariant polynomial, machine-learned, potential energy surface (PES) for the 4-body interaction. This PES is a fit to 2119, symmetry-unique, CCSD(T)-F12a/haTZ 4-b interaction energies. Configurations for these come from tetramer direct-dynamics calculations, fragments from an MD water simulation at 300 K, and from tetramer fragments in a variety of water clusters. The PIP basis is purified to ensure that the PES goes rigorously to zero in monomer+trimer and dimer + dimer dissociations. The 4-b energies of isomers of the hexamer calculated with the new PES are shown to be in better agreement with benchmark CCSD(T) results than those from the MB-pol potential. Tests on larger clusters further validate the high-



fidelity of the PES. The PES is shown to be fast to evaluate, taking 2.4 seconds for 10^5 evaluations on a single core of 2.4 GHz Intel Xeon processor, and significantly faster using a parallel version of the PES.

9.2 Overview

The many-body expansion (MBE) for the non-covalent interactions is pervasive in computational chemistry and has been applied in many guises for water. There are numerous studies of this expansion for water, and this large literature has been reviewed in a recent article.[170] There are many aspects of this expansion, but the one of particular interest here is the convergence of this expansion for water. This is not a trivial matter to examine since one has to ask, convergence with respect to what property? Certainly the most

common property that has been examined in this respect is the complete dissociation energy of water clusters to all monomers or, almost equivalently, the interaction energy of clusters. The former includes relaxation of the monomer geometry in the full dissociation limit whereas the latter does not.

The general conclusion from numerical studies on moderate sized clusters is that an accurate, i.e., to within a few percent, description can be obtained with the MBE truncated after 3-body interactions. However, it has been known for many years that the 4-b interaction, while small, may be not negligible. Theoretical work that came to this conclusion dates back more than 35 years with the pioneering work investigating many-body aspects of the classical induced interaction of rigid water.[171,172] Gregory and Clary investigated the effects of 4-b induction interactions on the zero-point energy of the water hexamer in 1996.[173] (These interactions are formally N-body, where N is the number of monomers and are referred to as N-body induction energies.) It was assumed in this early work that 3 and 4-b interactions were negligible at short range and thus only the long-range classical induced energies were focused on. In 2000, Xantheas reported calculations of the total interaction energies of isomers of the water hexamer at the MP2 level and concluded that the 4-b contribution to the total interaction energy varies from 1.4 to 4.4 percent, depending on the isomer.[174] In 2011, a many-body analysis of the lowest energy isomers for the hexamer, the prism and the cage, determined from CCSD(T)/aug-cc-pVTZ calculations that 4-b contributions to the interaction energy were -0.57 and -0.43 kcal/mol, respectively.[175] A more recent examination of the 4-b interaction for the hexamer isomers confirms these values.[176] This lowers the prism energy by about 49 cm^{-1} more than the cage. Recently, Heindel and Xantheas reported a careful numerical examination (focusing on elimination of basis set superposition errors) of the MBE for interaction energies of water clusters, consisting of 7, 10, 13, 16 and 21 monomers.[177]

Based on this and previous analyses, they concluded that the MBE can thus be safely (i.e., converged to less than 1 percent) truncated after the 4-body term and not after the 3-body level.

The above conclusion notwithstanding, it is the case that the MBE truncated at the 3-body term does capture roughly 95 percent of the interaction energy for water clusters. Thus, in the past 10 or so years several *ab initio* water potentials have been developed based on 1-, 2- and 3-body *ab initio* PESs or correction PESs. These are known by the acronyms WHBB,[178] HBB2-pol, MB-pol,[100] and CC-pol.[179] The WHBB, HBB2-pol, and MB-pol potentials use a spectroscopically accurate 1-b potential,[180] and for the 2-b numerical fits, they are calculated using permutationally invariant polynomials (PIPs)[181] and thousands of CCSD(T) energies.[182–184] The 3-b potential in WHBB is a PIP fit to thousands of MP2 energies, whereas the one in the HBB2-pol and MB-pol potentials is based on fits to thousands of CCSD(T) energies. To be more precise, MB-pol uses PIP fits to the difference in 2-b and 3-b energies obtained from the TTM4-F PES[185] and CCSD(T) energies. The CC-pol potential uses elaborate functional forms to represent the 2 and 3-b interactions and to fit them to many CCSD(T) energies.[179]

The WHBB potentials account for 4 and higher-body interactions by switching to TTM3-F potential[186] in the long-rage. This potential provides a sophisticated treatment of long-range electrostatics for an arbitrary number of monomers. The MB-pol potential uses a different strategy, since it is based on correction PESs at the 2 and 3-body level to the TTM4-F potential[185] plus additional electrostatic terms. A critical assessment of the accuracy of the TTM3-F and TTM4-F potentials for 4-b energies of isomers of the water hexamer against direct CCSD(T)-F12b/VTZ was reported in 2015.[176] While the TTM4-F 4-b potential is generally more accurate than the TTM3-F one, it has errors between around 0.1 and 0.35 kcal/mol. These are fairly large fractions of the 4-b energy itself.

Indeed these errors account substantially for the overall errors of the WHBB and MB-pol potentials for the interaction energies of hexamer isomers. In addition, these errors are potentially significant for rigorous studies of the relative energies of the prism and cage, which have been done with both the WHBB and MB-pol potentials and using rigorous treatments of the vibrational motion.[187] In the latest study using MB-pol, the cage was reported to be more stable at 0 K than the prism by roughly 0.1 kcal/mol, including a rigorous treatment of zero-point energy.[188] This difference is within the range of errors in the hexamer 4-b interaction in the MB-pol potential and so the conclusion probably should be viewed with some caution.

Based on this and previous work, there is strong motivation to develop a full-dimensional PES for the 4-b water interaction potential. This is a challenging 12-atom system that was beyond consideration during the time when WHBB, MB-pol, and CC-pol were developed. Attempts have been made by some of us to obtain analytical expressions of high dimensional 4-body interactions of molecular systems, i.e. hydronium-water-water-water 4-b interaction.[189] In that work, inspired by Skinner and coworkers' E3B water 3-body model,[190,191] we proposed empirical exponential functions for different types of 4-b geometries and fitted relevant parameters from limited *ab initio* data. However, these simple exponential functions can not meet the requirement of a robust and high dimensional 4-body potential.

9.3 Results and Discussion

Recently, we extended the PIP approach to describe PESs for 12-atom *N*-methyl acetamide[141,144], 15-atom tropolone[166], and acetylacetone.[165] Also noteworthy is a recent paper reporting a method to generate a very compact PIP fitting basis.[192] Both our

papers from 2019 and this one have clearly broken the oft-cited "10-atom" limit for PIP bases that appeared 12 years ago.[181] Because the long-range behavior is important for the 4-b interaction, here we use a PIP basis that is purified[193–195] so that the 4-b PES goes rigorously to zero as any monomer (or dimer) separates from the other group of monomers. The CCSD(T)-F12a/haTZ (aug-cc-pVTZ basis for O and cc-pVTZ for H) method was selected for the calculations of electronic energies as it provides accurate results at an acceptable computational cost.

Below we give details of the generation of the database of electronic energies and standard tests of the precision of the 4-b PES are given. Then the new PES is tested for the 4-b energies of 8 isomers of the water hexamer, and the total 4-b energies of a number of larger water clusters. The long-range behavior is examined for two dissociation 1d-profiles of the water tetramer to two dimers and to a monomer plus trimer.

It is instructive to investigate the contributions of the various n-body (n=2, 3, 4) interaction energies for the water tetramer. (Recall that these are obtained in a series of calculations where, for example, the 2-b energies correspond to the energies of all pairs of monomers, etc. Of course finally, the sum of these n-body interactions, n=2-4, must equal the total interaction energy of the tetramer.) This is shown in Figure 9.1 for the the movement of two dimers starting at the global minimum (GM) structure where just the distance between two equivalent dimers, denoted R, is varied. At large R we have two equivalent non-interacting water dimers. The total interaction is repulsive for R less than 2.5 A. But, as seen, the addition of the 3-b interaction to the 2-b ones moves the minimum to smaller R and likewise shifts the repulsive wall to smaller R. The same occurs, albeit to a smaller extent, by the addition of the 4-b interaction. Thus, over the range of the plot the 3-b and 4-b interactions are attractive. As seen, the 2-b interaction is the largest, followed by the 3-b, followed by the 4-b, as expected. The 4-b is very small

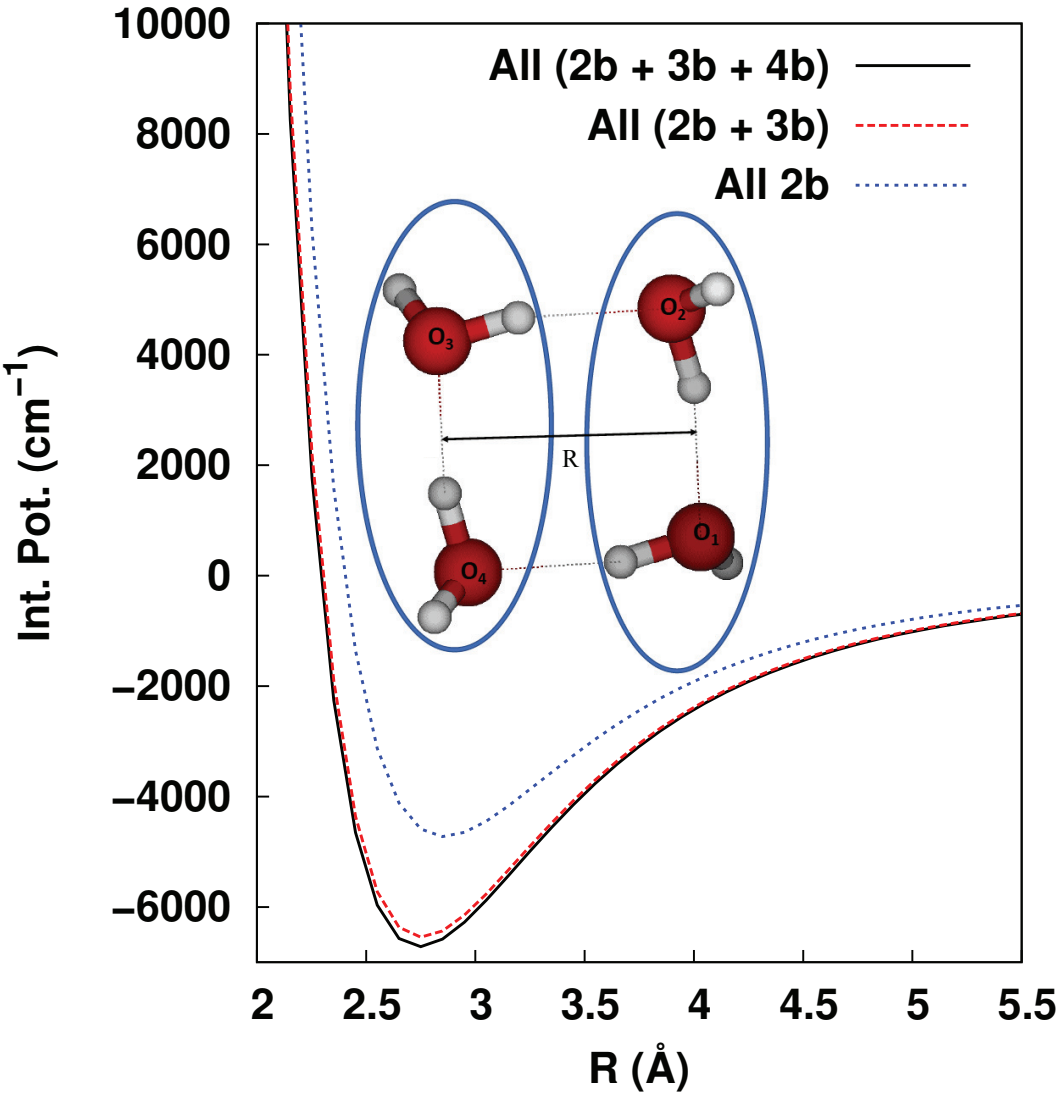


Figure 9.1: Interaction potential for the water tetramer including the n-body terms as indicated. See text for details of the 1d path defining this energy profile.

in this case; however, it should be noted that while there are 6 2-b interactions and 4 3-b interactions there is only a single 4-b interaction for the tetramer. Of course, as the number of monomers increases the number of 4-body interactions grows faster than the number of 2 and 3-b interactions.

The Figure 9.1 provides important evidence that the lower-limit of the range R for the 4-b PES should be roughly 2 Å. With that in mind a variety of approaches were used to generate a diverse database of 4-b configurations and energies. First, configurations were obtained from direct-dynamics calculations for the water tetramer. Second, tetramer fragments were selected from a classical dynamics simulation of water using the MB-pol potential. This source of data anticipates the eventual use of the 4-b potential for condensed phase simulations. Third, tetramer fragments were selected from the equilibrium configuration of isomers of the water hexamer, heptamer, decamer, and 13-mer. Lastly, configurations along the dimer+dimer dissociation profile shown in Figure 9.1. In total, 2119 configurations were obtained from this sampling, details of which are given next.

The NVE direct-dynamics calculations were done using the efficient B3LYP/6-31+G(d,p) level of theory. Trajectories were initiated from the two high-energy planar ring structures of the tetramer. These were generated by simply bringing the up-down-up-down ring structure of the global minimum to a planar structure by flipping the out-of-plane H atoms to the plane of the four O atoms. And for the first planar structure the distances between two equivalent dimers are 2.7515 and 2.7575 Å, respectively, and for the second one these are 2.3533 and 2.7575 Å, respectively. The energies of these structures are respectively 8,100 and 11,243 cm^{-1} above the GM. These planar structures along with the GM one are given in the Figure 9.2. Two trajectories were calculated from the first planar structure and three trajectories were calculated from the second planar structure. An initial kinetic energy of 1000 cm^{-1} was distributed randomly in these five

independent trajectories to all atoms at these initial configurations. The total energies of tetramers in the direct-dynamics trajectories are thus 9,100 and 12,243 cm^{-1}. These trajectories were propagated for 30,000 time steps with the step size of 5.0 a.u.(about 0.12 fs). A histogram distribution of tetramer potential energies, relative to the global minimum, is shown in Figure 9.3. As seen they span a range from roughly 2000 to 11 500 cm^{-1}. This indicates that the tetramer configurations from the AIMD trajectories span a large range of distorted geometries. The data generated from these trajectories was pruned by simply saving every 10th configuration (in time) to obtain 468 configurations for the final database. However, since dissociation of the tetramer did not occur in these direct-dynamics simulations, additional configurations are needed. These are describe next.

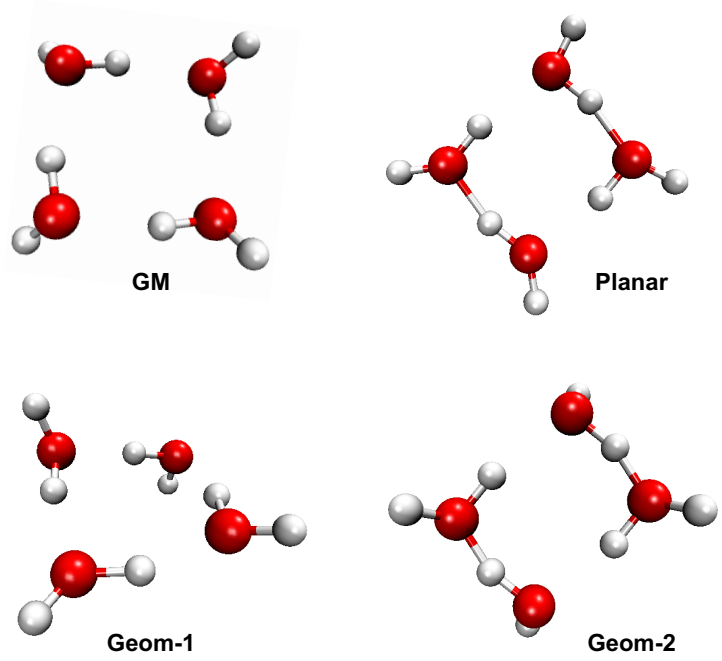


Figure 9.2: Water tetramer Geometries.

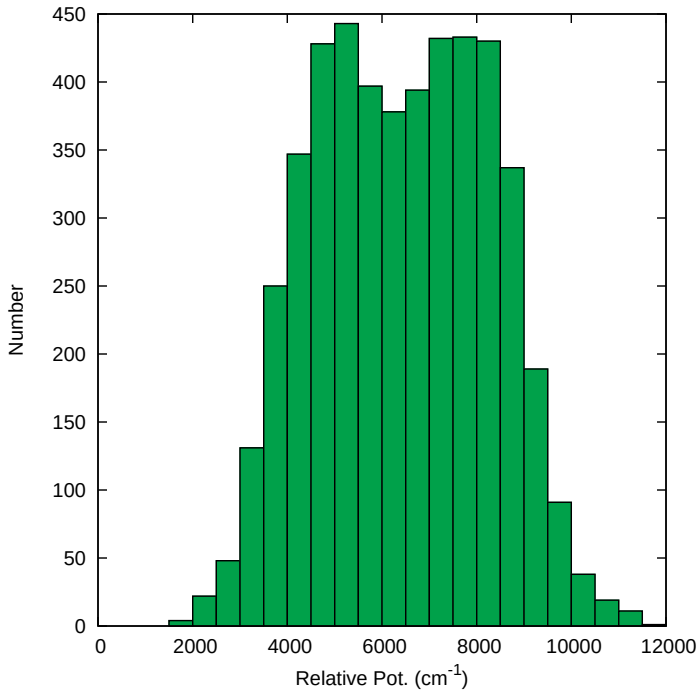


Figure 9.3: Tetramer potential energies from direct dynamics trajectories

Additional 4-b configurations were obtained from tetramer clusters selected from a 300 K MD simulation of water using 256 monomers and the MB-pol potential. These are defined as four monomers with the third largest O-O distance smaller than 4.5 Å. These we selected quasi-randomly from the large database of monomer configurations, and 540 configurations were selected.

Sixty one energies were obtained from the cut dissociating to the two dimers over the range 1.7 to 7.7 Å. Finally, tetramer configurations were obtained from 4 isomers of the water hexamer, namely the prism, cage, book-1, and cyclic-chair, the water heptamer, decamer, and 13-mer water clusters. All 4-body configurations were selected at the equilibrium structures of these clusters. For the hexamer isomers the structures are from ref. 196 and for the heptamer, decamer and 13-mer from ref. 177. In total 1050 configurations

from these clusters were calculated.

To construct the new PES, we calculated the 4-b energies at the CCSD(T)-F12a/haTZ level of theory for the above 2119 configurations. The choice of basis was made considering Basis Set Superposition Error (BSSE) and computational efficiency. Since BSSE is typically assessed for interaction energies, that is what we considered for three configurations, the minimum, Geom-1 and Geom-2. Table 9.1 shows this analysis for three bases. As seen all bases exhibit a relatively small BSSE and the results for the haTZ basis are very close to the much more expensive haQZ basis. So, based on this analysis we chose the haTZ basis for the CCSD(T) energies at the database configurations. The entire energy range of these 4-b energies is -1063.1 to 105.2 cm^{-1} and the distribution of most of these energies in a smaller range is shown in Figure 9.4. There are just a small number of 4-b energies outside this range.

Table 9.1: Comparison of BSSE corrected (corr.) and uncorrected (uncorr.) 4-b interaction energies (in cm^{-1}) for minimum geometry as well as two more geometries (Geom-1 and Geom-2) at various level of theory.

Method	Minimum		Geom-1		Geom-2	
	corr.	uncorr.	corr.	uncorr.	corr.	uncorr.
CCSD(T)-F12a/haDZ	-173.1	-169.4	-402.6	-395.3	-636.2	-636.0
CCSD(T)-F12a/haTZ	-174.4	-174.5	-405.6	-404.0	-647.2	-646.2
CCSD(T)-F12a/haQZ	-175.9	-175.3	-406.8	-404.8	-649.4	-648.7

Histograms of the OO distances in the tetramer configurations from the entire dataset are shown in 9.5. As seen they span a range from roughly 2 to 9 Å, indicating good coverage of this important distance. Histograms of the 4-b energies for the water heptamer and decamer are given in Fig. 9.6.

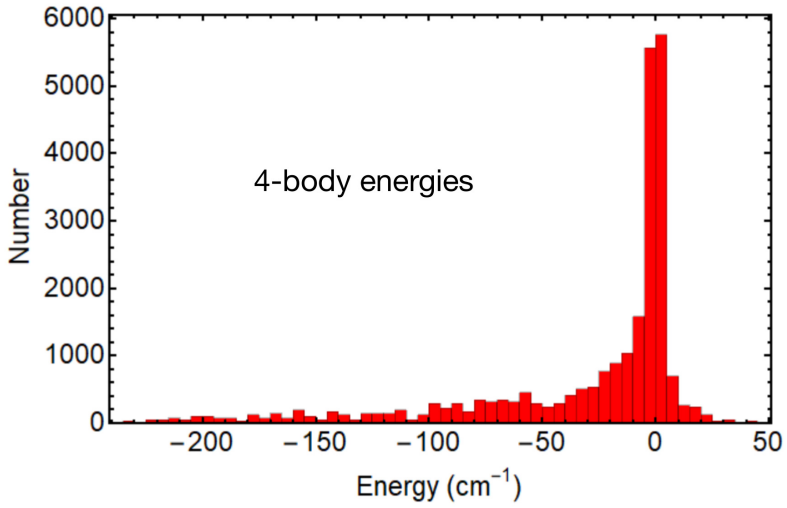


Figure 9.4: Histogram of the CCSD(T)-F12a/haTZ points used in the PES fit to the 4-body potential energy surface.

The PES is a linear least-squares fit to the data base of energies using a representation in permutationally invariant polynomials (PIPs).[27,142,181] The permutational symmetry that describes all the permutations of like atoms has the designation 84, meaning that the eight hydrogens permute with one another and the four oxygens permute with one another. However, it is also possible to use reduced symmetries if one replicates the data set with the proper number of permutations of the water molecules. Here we use the symmetry 22221111, meaning that the hydrogens within each water monomer permute with one another but that they do not permute between water molecules, and that the oxygens do not permute among themselves. In order to describe the invariance with respect to the 4! permutations of water monomers we replicated the symmetry-unique configuration 4! times. The reason for using this reduced symmetry is given below, where we discuss the total polynomial order for the fit.

The usual method for fitting the energy data set with the PIP polynomials is not

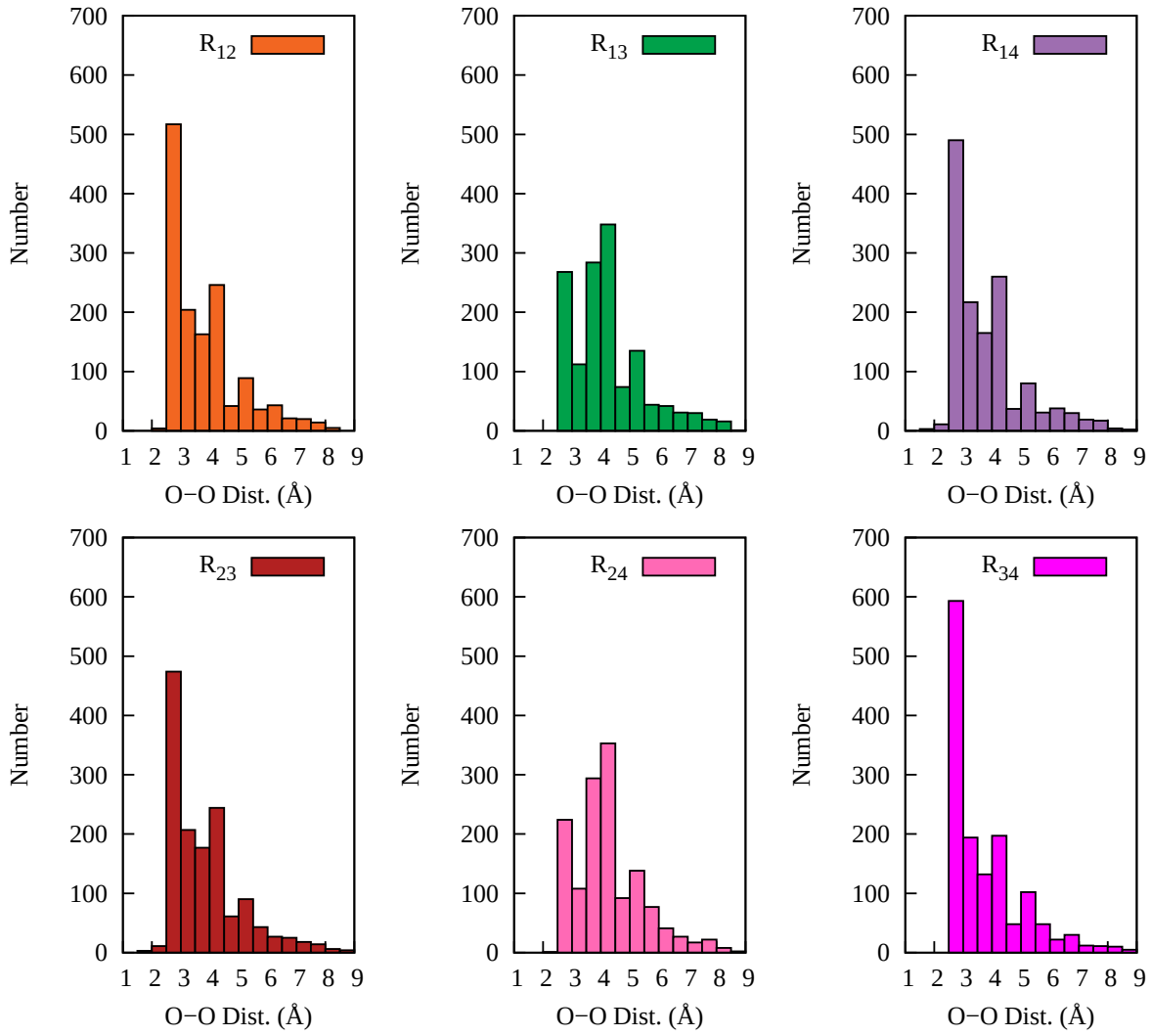


Figure 9.5: Distribution of OO distances for the entire dataset.

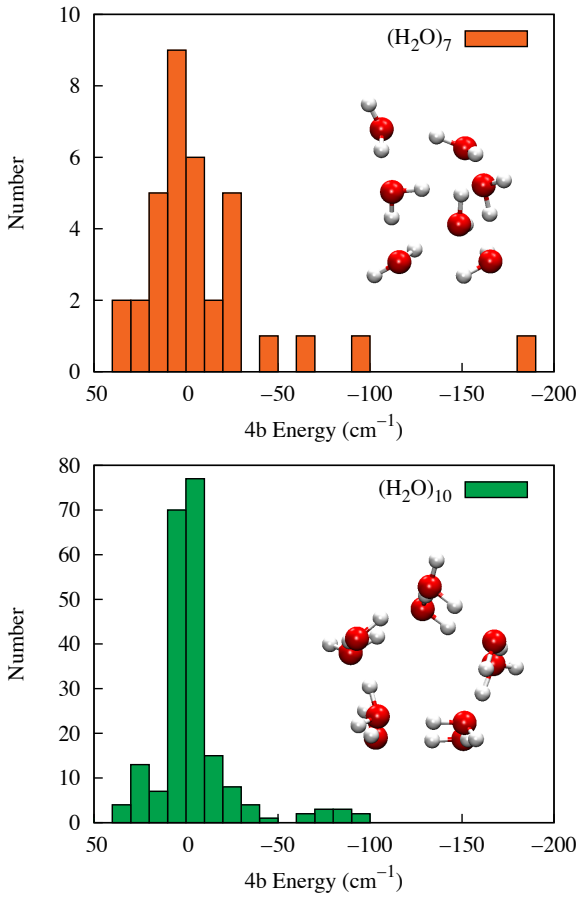


Figure 9.6: Histograms of 4-b energies for heptamer and decamer.

to use the internuclear distances directly in the PIPs but rather to transform them into functions such as Morse variables ($\exp(-r_{ij}/a)$) or $1/r_{ij}$ variables. For the current application, we used a mixture of these functions, namely Morse variables for the 12 intramolecular coordinates with the range parameter $a = 2$ Bohr, and $1/r_{ij}$ variables for the 54 intermolecular coordinates.

Another consideration for the basis set choice is the behavior of the PIPs as a monomer or dimer of the tetramer is removed to a great distance with respect to the remaining waters. By definition, the 4-body interaction energy must go to zero as any water monomer

or water dimer is distanced. Many of the PIPs do not have this property.[181] The process of eliminating these PIPs[193] is what we term "purification".[194,197,198] Our method of purification starts by assigning random numbers to all the internuclear distances for the four water molecules, 66 in this case, and evaluating the polynomials using Morse variables. One then adds a large number to the Cartesian coordinates of the water monomers, one at a time, and water dimers, two at a time, and then calculates the new values of the polynomials. If for any of these distancing operations the value of a polynomial does not become less than a small cutoff number (we used 10^{-6}), the polynomial is eliminated from the set.

For the maximum 84 symmetry and using a maximum polynomial order of 3, there are 86 PIPs, but only 2 remain after purification. Clearly this is not a usable basis. In the 22221111 symmetry with polynomial order 3, there are 10 737 PIPs, and 1649 remain after purification. Of course, we want the number of PIPs, equal to the number of unknown coefficients, to be less than the size of energy database but not to be so much less that we do not obtain good fit precision. Using 22221111 symmetry and replication the data set of 2119 $\times$ 24 we have 50 856 energies, and so 1649 coefficients should be sufficient and precise. Although we considered other possibilities, many were infeasible and, for some, the permutational replication of the data set was not effective because it mixed intra- and inter-monomer coordinates. In the end, we used the 22221111 symmetry with a data base size of 50 856 and 1649 coefficients. For this basis the fitting error for the energies, shown in Figure 9.7, is small with an RMS error for the entire dataset of 6.2 cm^{-1}.

We now present several tests of the new 4-b PES, ranging from 1d energy profiles to tests for variety of water clusters. We compare to CCSD(T)-F12a calculations as well as results from the MB-pol and TTM4-F potentials. The first test is a comparison of 4-b

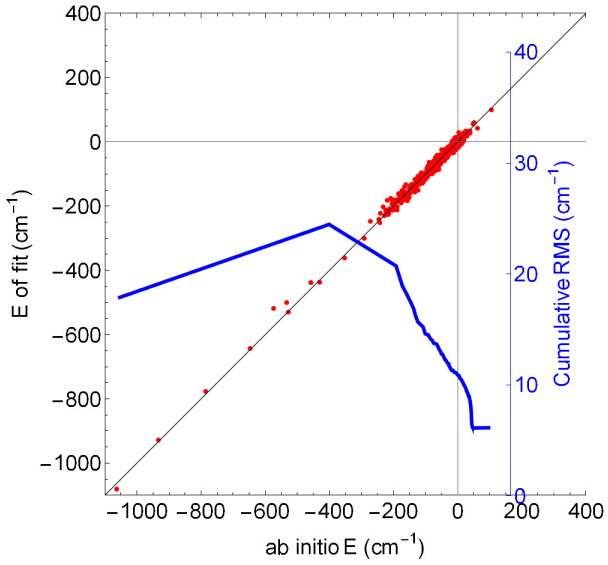


Figure 9.7: Correlation plot between *ab initio* energies and energies calculated by the PES fit. The correlation coefficient is $R^2 = 0.990$ and the RMS error is 6.2 cm^{-1}. (blue) The cumulative RMS error is in cm^{-1}.

energies from the PES, TTM4-F, and CCSD(T)-F12a for a 1-d profile showing the 4-b energy along the separation of the water tetramer to two rigid dimers passing through the cyclic minimum (OO equal to 2.75 Å). These are shown in Figure 9.8 from the current PES, direct CCSD(T)-F12a/haTZ, and TTM4-F calculations. Note, CCSD(T) energies for OO greater than 8 Åare not included in the training data set. As seen, over the range 1.7 to nearly 14 Å, the PES is in very good agreement with direct CCSD(T)-F12a/haTZ energies. The 4-b energies from the TTM4-F potential are also in good agreement with the benchmarks for OO distance greater than 2 Å. However, a major breakdown of the accuracy of that potential is seen for shorter distances. Evidently, this is where the classical electrostatic description of the two water dimer interaction fails, presumably due to significant chemical interaction between the electronic orbitals of the dimers. We focus

here on the TTM4-F potential as it plays a central role in the MB-pol water potential.[199] Namely, MB-pol uses high-level, *ab initio* correction PESs to TTM4-F at the 2 and 3-b level in the short range and relies on that potential for all 4-b and higher body water interactions and also for the interactions beyond the range of the corrections. Finally note the gradual approach to zero in this profile. At 11 Å the energy is less than -1.0 cm^{-1}

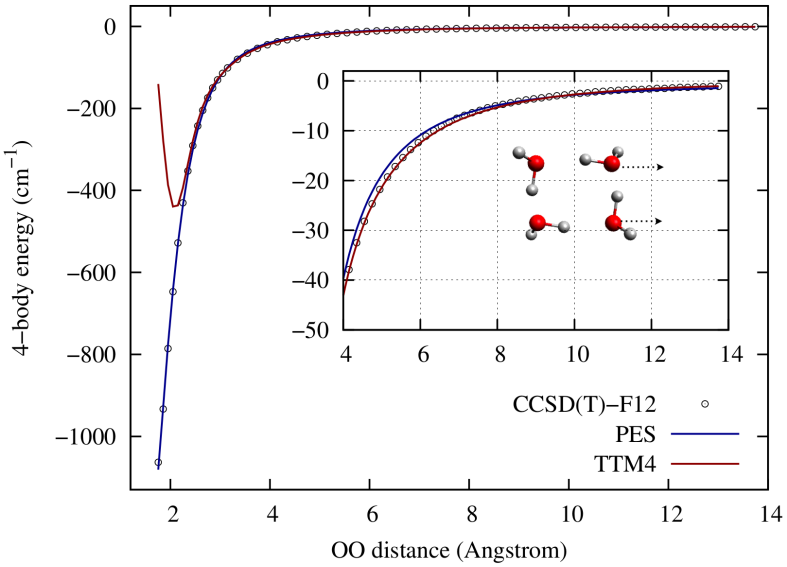


Figure 9.8: 4-b energies from indicated sources (see text for more details) as a function of the oxygen-oxygen distance between pairs of water dimers in the tetramer. The dashed arrows indicate the dimer pair that separates from the rigid tetramer. The equilibrium value of this distance is 2.7 Å.

A second profile showing PES and CCSD(T) energies is given in Figure 9.9. Here the tetramer dissociates to a monomer plus trimer over an OO range of 2.7 to 15 Å, and corresponding energy range of -175 cm^{-1} to 0. This is a significant test of the PES as no CCSD(T) data for this profile were included in the fitting database. As seen, the PES

agrees well with the benchmark results and the range of the interaction for this profile is roughly 7 Å.

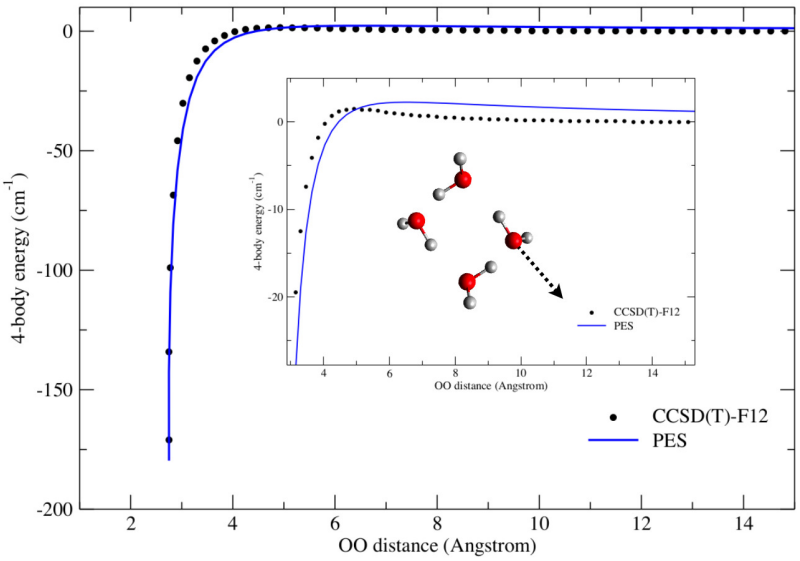


Figure 9.9: Test of the 4-b PES for a cut with a single monomer separating from the tetramer against CCSD(T)-F12a/haTZ energies. OO is the distance between the O atoms on the two monomers on the axis inferred from the arrow.

Both profiles show that the 4-b energy is rapidly becoming more negative at OO distances near and especially less than the equilibrium value of 2.7 Å. (This was already inferred in the discussion about Figure 9.1, where the many-body decomposition of the tetramer potential was shown.) In general, this indicates the possible importance of the 4-b in softening the repulsive interaction in the condensed phase, especially for high pressure conditions. At long range the two 1d-potentials show different behavior with trimer+monomer interaction at somewhat shorter range than the dimer+dimer interaction. Also, the 1-d potential for the former decays to zero slightly more slowly (a difference of 1–2 cm^{-1}) than the CCSD(T) reference energies. Below we comment on practical ap-

proaches to treat the long-range behavior of the 4-b interaction for uses in applications to the condensed phase with many monomers. Those comments will address in more detail this small difference between the PES and CCSD(T) energies for the trimer+monomer interaction.

Another test of the PES is the 4-b energies of the isomers of the water hexamer, prism, cage, book-1 (bk-1), book-2 (bk-2), bag, cyclic-chair (c-chair), cyclic-boat-1 (c-bt-1) and cyclic-boat-2 (c-bt-2). These isomers have been the focus of a number of papers, and the one by Medders et al. is of particular interest.[176] There the 4-body energy from several water potentials, including WHBB5, MB-pol, TTM3-F and TTM4-F were compared with CCSD(T)-F12/VTZ energies for eight isomers. Here we focus on the errors in the MB-pol and the new 4-b PES. These are shown in Figure 9.10. As seen, the 4-b PES errors are smaller than those from MB-pol. A numerical comparison of 4-b energies, the present CCSD(T) calculations and the previous ones from Medders et al. are given in Table 9.2. The two sets of CCSD(T) calculations are within 0.1 kcal/mol and generally less than that for the eight isomers. Given that MB-pol and TTM4-F 4-b energies differ by no more than 0.04 kcal/mol, and given the limited accuracy of TTM4-F shown in Figure 9.8, for OO distances relevant to the hexamer, the lower level of accuracy of MB-pol compared to the PES is not unexpected.

One aspect of an accurate 4-b PES is that it can be used for arbitrarily many monomers with virtually zero BSSE. Thus, it is possible to test the prediction of 4-b energies of larger water clusters using the 4-b PES and against benchmark calculations. A convenient set of benchmark calculations has been published by Heindel and Xantheas,[177] who provided geometries (in their Supporting Information) and reported many-body energies calculated with and without BSSE-correction at several levels of theory. Table 9.3 shows their results in columns 4 and 5 for MP2 and the basis set listed in col-

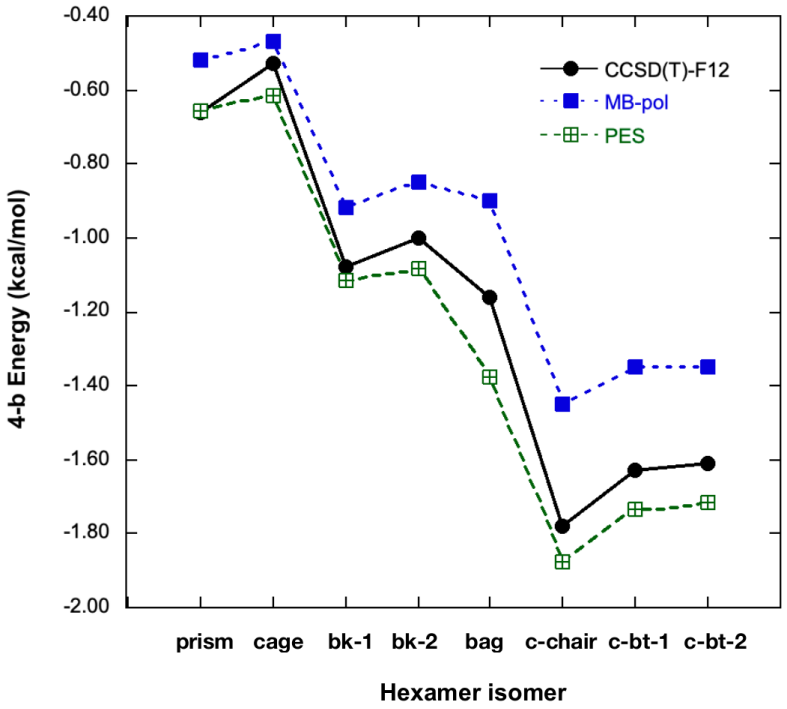


Figure 9.10: 4-b CCSD(T) energies for indicated isomers of the water hexamer from the 4-b PES, MB-pol, and previous CCSD(T)-F12/VTZ calculations.

Table 9.2: 4-b interaction energies (in kcal/mol) of eight isomers of the water hexamer from indicated sources.

Isomer	PES	Present CCSD(T)-F12a/haTZ	CCSD(T) -F12/VTZ[a]	MB-pol /TTM4-F
Prism	-0.66	-0.62	-0.66	-0.52
Cage	-0.61	-0.54	-0.53	-0.47
Book-1	-1.12	-1.16	-1.08	-0.92
Book-2	-1.08	-1.05	-1.00	-0.85
Bag	-1.38	-1.21	-1.16	-0.90
Chair	-1.88	-1.85	-1.78	-1.44
Boat-1	-1.73	-1.71	-1.63	-1.35
Boat-2	-1.71	-1.69	-1.61	-1.35

[a] From Medders et al J. Chem. Phys. 143, 104102 (2015).

umn 3. The results using the PES to calculate all 4-b energy in each n-mer are shown in the last column. The agreement is good through $n = 16$. Remarkably, even better agreement is obtained between the PES and benchmark CCSD(T)-F12a/haTZ calculations performed up to the 13-mer. For the two isomers of $n = 21$ the BSSE correction to MP2 calculations was not applied and coupled clusters energies were not obtained as both computations were evidently too intensive. As seen, the PES 4-b results are much less than the uncorrected MP2/aVDZ results. While this is almost certainly qualitatively correct, the quantitative accuracy of the 4-b results would require new benchmarks calculations, which for these larger clusters are currently not easily done.

Table 9.3: 4-b interaction energies (kcal/mol) for various water clusters, where the information in columns 3–5 is taken from the SI in ref. 177, numbers in column 6 are from our CCSD(T)-F12/haTZ calculations, and the numbers in column 7 are calculated using our 4-b PES.

$(H_2O)_n$ n=	No. of 4mers in nmer	H-X basis	MP2	MP2 BSSE-corrected	CCSD(T)-F12/haTZ	4-b PES
7[a]	35	aVTZ	-1.106	-0.874	-0.987	-1.012
10[a]	210	aVTZ	-3.028	-1.978	-2.576	-2.341
13[a]	715	aVDZ	-6.352	-1.499	-1.539	-1.571
16[a]	1820	aVDZ	-9.463	-2.179	-	-2.983
21[b]	5985	aVDZ	-20.976	-	-	-9.888
21[c]	5985	aVDZ	-19.892	-	-	-5.920

[a] These are global minimum structures;
[b] Fully solvated structure;
[c] All surface structure.

For most of the large clusters, many of the 4-mer configurations involve at least one water that is quite distant from the others, so that the 4-b energy is quite small. For example, Figure 9.11 shows a histogram of the PES-determined 4-b energies for

the "fully-solvated" 21-mer. There are large peaks near zero, although the distribution actually stretches from -95 to 54 cm^{-1}, with only a minor number of energies outside the range depicted in the figure. Since the total 4-b energy is the sum of all the 4-body contributions (5985 in the case of $n = 21$), there is evidently a large degree of cancellation. The issue of the many very small 4-b interactions is of course an important one and we comment on it after we discuss the computational effort to evaluate the 4-b PES, which we do next.

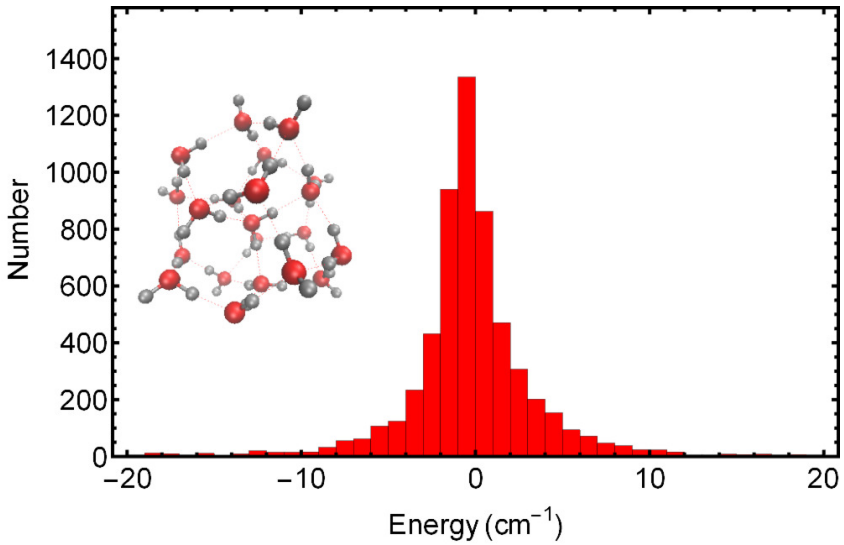


Figure 9.11: Distribution of PES 4-b energies for the fully solvated 21-mer

Finally, an important aspect of the new 4-b PES is the computational effort to use it. One hundred thousand evaluations of the 4-b PES takes 2.45 seconds on a single core of the 2.4 GHz Intel Xeon processor. This is significantly smaller than the computational effort to evaluate our WHBB PES for the water trimer.

While this computational effort is not large, it should be recalled that the number of 4-b interactions goes formally as N^4 for N water monomers, Thus, the cost would

be high if all the tetramers were to be evaluated for say a condensed phase simulations which require hundreds of monomers. A standard approach to deal with this is to use a distance-based cutoff via a standard switching function. This would greatly reduce the number of tetramers to be evaluated because in large clusters many tetramers have large inter-monomer distances and the 4-b energy for such tetramers is essentially zero and thus not necessary to evaluate. We demonstrate this below with results shown in Table 9.4, for clusters with 64, 128 and 256 monomers. In this case we evaluate the 4-b potential of tetramers whose largest O-O distance is smaller than 7 Å. (We don't necessarily recommend this for all applications of course, it is used here just for illustration.) With this choice more than 99% of the tetramers are not evaluated, and calculating the total 4-b energy takes seconds, as seen, rather than more than an hour. Parallelization is another possible strategy since the calculation of one tetramer is independent of the others. This has been done using OpenMP and as seen there is a substantial speed-up in the calculations.

Table 9.4: Computation cost of 4-b PES calculation with cut-off O-O distance of 7 Å

$(H_2O)_n$ n=	No. of total 4b	No. of calculated 4b	Time (s) single core	Time (s) 8 cores
64	635,376	5,078	0.18	0.03
128	10,668,000	24,460	0.79	0.12
256	174,792,640	58,804	1.80	0.25

Limiting the range of the 4-b interaction through the use of a multiplicative damping factor (switching function) that brings the 4-b to zero in a finite range is the well-established way to deal with the huge number of interactions. The parameters of such a function of course are varied, within practical limits, to establish systematic uncertainties

of any property with respect to this switching. An example from our group used damping of the 3-b interaction in $H_2-H_2O-H_2O$ interaction in a diffusion Monte Carlo quantum calculation of vibrational energies of H_2 in a hydrate clathrate.[200] In this calculation the WHBB water potential was used to describe the water interaction and a separate 1,2,3-body potential was used to describe the H_2-water interactions.[197] It is worth noting that this is a general many-body approach to describe the interaction of molecules and ions with water.

9.4 Summary and conclusions

In summary, we reported the first *ab initio*, full-dimensional potential energy surface for the 4-body interaction of water. The potential is a purified PIP fit to a diverse set of 2119 unique CCSD(T)-F12a/haTZ energies. This representation ensures that the PES rigorously goes to zero as a monomer or dimer is separated from a tetramer cluster. Tests of the 4-b PES demonstrate the high-fidelity compared to benchmark calculations. In addition, the 4-b PES was shown to significantly reduce errors in 4-b interaction energies for isomers of the water hexamer obtained from a high-level water potential that describes water interactions to the 3-b level. This new 4-b PES was shown to be fast to evaluate, and for applications to large numbers of water monomers a large fraction of 4-b configurations are beyond the range of the 4-b PES and so do not have to be evaluated. This was demonstrated for the 21-mer. The PES is also robust, at least for the present applications. In total roughly 100,000 4-b energies have been obtained using the new PES and none show huge negative values, i.e., "holes". The current 4-b PES could be an add-on to a water potential that either does not contain 4-b terms or has an approximate treatment of the 4-b, e.g., the TTMn family of potentials. Specifically, as

seen in Fig. 4, TTM4-F provides an accurate description of the 4-b interaction, except at short range where it fails. This suggests that a correction PES to the TTM4-F 4-b interaction might be a reasonable future project. However, the unphysical behavior of TTM4-F shown at short OO distances could be problematic for a correction PES. In any case, we plan to investigate this in the future. Finally, it should be clear that the present 4-b PES is "version 1.0". And while it appears to be both fast to evaluate and accurate, we anticipate being able to make significant speed-ups, the easiest one of which is to make use of multi-core architecture of all modern workstations and computer nodes.

Part III

Reaction Dynamics and Molecular Vibrations

Chapter 10

Implications of "Zero-Point Leak" and Isomerization of syn-CH_3CHOO in Quasiclassical Trajectory Calculations via Adiabatic Switching

10.1 Chapter Abstract

Quasiclassical trajectory calculations are distinguished from molecular dynamics calculations primarily by the difference in the phase-space used for sampling initial conditions. In the former, this phase space incorporates zero-point energy (ZPE), whereas the latter uses the classical phase space, typically at a given temperature. There are advantages and disadvantages to these approaches. Clearly, the quasiclassical one aims to mimic quantum mechanics, where zero-point energy is a major and essential difference with classical mechanics. But, "ZPE leak", a well-known issue with quasiclassical calculations, has

largely prevented the use of the quasiclassical method for larger systems and also is often mentioned as an issue even for smaller systems, where it is routinely used. By not incorporating ZPE, molecular dynamics calculations avoid this issue; however, at the cost of neglecting this important difference with quantum mechanics. In this article we examine ZPE leak from the point of view of rigorous semi-classical quantization; actually the lack thereof. We examine this issue here specifically for syn-CH_3CHOO, where the ZPE is roughly twice as large as the barrier for isomerization to the more stable isomer vinyl hydroperoxide. Semi-classical quantization is done for this 18 degree-of-freedom molecule using adiabatic switching in Cartesian coordinates, using the methodology recently described by us [C. Qu and J. M. Bowman, J. Phys. Chem. A., 120, 4988–4993 (2016)]. Good agreement for vibrational energies is found compared to accurate quantum calculations. Finally, some aspects of possible mode-specific isomerization to vinyl hydroperoxide is examined using the adiabatically switched trajectories.

10.2 Overview

Molecular dynamics, i.e., classical trajectory, calculations are by far the most commonly used simulation approach to investigate the nuclear motion in applications ranging from locating stationary points on potential energy surfaces (PESs) to IR and electronic spectroscopy to chemical reactions. There are of course known issues with this approach, especially when applied to H-atom motion. One is the lack of zero-point energy, which can be quite large. For H-atom motion in particular the issue is serious because even a single mode involving an H-atom, e.g., the OH stretch in water, can have a ZPE that is much larger than k_BT at room temperature. This means that a classical simulation of this stretch at room temperature produces essentially a localized harmonic description

of this motion. Among other implication, this means that the resulting IR band of this mode is essentially the harmonic one, which as is well known, can be upshifted from the correct one by roughly 100-200 cm^{-1}.

Adding zero-point motion to a classical simulation is done in so-called quasiclassical trajectory (QCT) calculations. It is well-known, and widely-used in gas-phase reactive scattering. Since its introduction by Karplus et al. for the $H+H_2$ reaction,[201] it has been used successfully in many atom+diatom reactions. For a diatomic molecule the addition of ZPE results in a perfectly periodic motion that is stationary in time and so there is no "ZPE leak" in this case. For chemical reactions, the addition of ZPE is known to produce much more accurate results than a simple classical simulation,[202] and it is also justified from a semi-classical point of view clearly. Similarly, a quasiclassical simulation of the IR spectrum of say OH is clearly superior to the classical one since the QCT one captures a significant effect of the anharmonicity, which is essentially neglected in a classical simulation except at very high temperature.

The problem with the QCT approach becomes apparent for applications with more than one mode, i.e., virtually all applications. To understand this, consider how ZPE is added for many modes. Typically, this is done by performing a normal mode analysis of the molecule, cluster, etc. and then applying straightforward semi-classical conditions on the harmonic motion. Details of this procedure and its use to obtain initial conditions for trajectory calculations can be found in the excellent 1998 article by Hase.[203] The problem is that the Hamiltonian that governs the nuclear motion is, in general, anharmonic and so the approximate harmonic semi-classical quantization breaks down, and the resulting mode-mode coupling causes energy transfer among the modes, which in the case of ZPE in the modes, results in the infamous "ZPE leak". This leak can be relatively benign or insidious. In the former case, the mode energies beat back and forth and on average

mode ZPE is sustained, at least for long-enough times for the simulation to finish, e.g., in X+CH_4 chemical reactions.[204,205] In the latter case, the leak can cause unphysical chemistry, e.g., bond-breaking. This is well known and is largely the reason given for not using QCT simulations in say water simulations, as even the water dimer shows this failure.[206] Just 15% ZPE leakage from the four OH-stretches is sufficient to break the H-bond. ZPE leak has also been noted recently in the context of a quantum thermostat for molecular dynamics simulations.[207]

Several "active" procedures to prevent this leakage among harmonic modes have been reported.[208–211] The earliest one was motivated by the model, 2-mode Henon-Heiles Hamiltonian

$$H = \frac{1}{2}(p_1^2 + p_2^2 + q_1^2 + q_2^2) + q_1^2 q_2 - \frac{1}{3}q_2^3. \tag{10.1}$$

Vibrational energy leak is particularly fast in this model as the two harmonic modes have the same frequency. The transfer of mode energy can be completely turned off at a certain threshold value by applying an impulse in the phase-space of each mode that changes the sign of the momentum, $p_1(t)$ or $p_2(t)$ at some time t, when the energy in mode 1 or mode 2 is equal to a threshold value. A generalization of this approach was used successfully to prevent ZPE leak in the water dimer[206] and trimer[212]. However, the procedure is cumbersome as it requires defining sets of normal modes as the trajectory evolves and so is not readily adaptable to larger clusters.

In any case, these methods are a patch to the fundamental problem, which is the lack of rigorous semi-classical quantization of the zero-point state. In principle rigorous semi-classical quantization determines the stationary classical action variables, which are the analogs of quantum numbers. For the unique zero-point state these quantum numbers are all zero and semi-classical quantization should be easiest for this state. So, assuming

that this can be done (and perhaps also assuming that there is a stationary quantum zero-point state), it seems clear that the classical motion of this state should be the most separable one and the one least likely to be plagued by islands of "chaos". Thus, correct semiclassical quantization of the zero-point state should lead to a stationary phase space with no ZPE leak, benign or insidious.

The theoretical and computational study of semiclassical quantization of molecular motion, in particular periodic or quasi-periodic vibrational motion, is large, with many contributions and advances made in the 1980s.[213–226] Nevertheless, the methodology is complex, as the forces, even at zero-point energies, are non-linear and the search for periodic orbits is challenging. Most of the applications were to model problems and only later to small molecules, e.g., triatomics and tetraatomics. For larger systems, some of the methods, especially those that search for periodic orbits and work for smaller molecules, become unwieldy.

We recently used adiabatic switching (AS)[219–226] in a calculation of quantization of the vibrational zero-point state and selected fundamentals of CH_4.[227] This was done using an accurate PES, for which exact quantum calculations of these states were available. Details of these calculations, which were done in Cartesian coordinates and which required some new methodology, are given below. As noted above, standard, approximate harmonic semi-classical quantization of these vibrational states of CH_4 was done (successfully) in studies of reaction dynamics. And so in this case ZPE and higher excited vibrational energy leak is benign. (For interesting further discussion of ZPE leak in reactive QCT calculations see ref. 228.)

Adiabatic switching appears to be straightforward to apply to high-dimensional systems, as the following brief description of it makes clear. The basic idea of AS is as follows: the Hamiltonian of the polyatomic molecule can be written as the sum of a

zeroth-order separable, and non-separable parts, $H = H_0 + \Delta H$. The separable part H_0 is assumed to be easily quantized semiclassically. Then a time-dependent Hamiltonian is introduced that evolves, via a simple switching function, from H_0 to H. If this is done sufficiently slowly, and according to the adiabatic theorem, the quantized state of H_0 will slowly evolve to the corresponding quantized state associated with the true Hamiltonian H. The application of AS in semiclassical quantization of mostly model systems has been investigated and reviewed.[219–226] The method, as will become clear below, is straightforward to apply and so it offers the prospect of being readily implemented in QCT calculations involving polyatomic reactants (and products). Recently, AS has been applied successfully to obtain the Wigner distribution for anharmonic coupled oscillators, starting with a zero-order separable distribution.[229]

The current application of AS is to the Criegee intermediate syn-CH_3CHOO. The Criegee intermediates are an important source of atmospheric OH radicals, and the dissociation of syn-CH_3CHOO has been studies extensively recently.[230–234] The mechanism of the dissociation of syn-CH_3CHOO to vinoxy+OH is based on the usual analysis of stationary points of the PES schematic. This indicates that syn-CH_3CHOO isomerizes to the vinyl hydroperoxide (VHP) through a five-membered ring-like transition state, and the VHP dissociates to OH+CH_2CHO products. While this mechanism is largely correct; there are added complexities in the region of the PES from VHP to the products,[233] and it is also incomplete at high internal energy of syn-CH_3CHOO, where a second, direct pathway to these products opens up.[234] In this paper, we focus on the issue of ZPE leak in the isomerization of syn-CH_3CHOO to VHP. The reason for choosing this isomerization, beyond the current interest in it, is that ZPE of syn-CH_3CHOO is almost twice the electronic barrier for isomerization. Thus, in a quasiclassical simulation, the isomerization might occur due to the ZPE leak. We apply AS to obtain the semi-classical

ZPE of syn-CH_3CHOO and investigate the time evolution of the molecule to see in fact if isomerization does occur. To push the method a bit, we also apply AS to the fundamental excitation of a mode that is predicted from a simple projection theory to promote isomerization.

This article is organized as follows: we briefly review the theory of adiabatic switching and provide the computational details in Section II. Results are presented in Section III and a summary and conclusions are given in Section IV.

10.3 Theory and Computational Details

10.3.1 Adiabatic switching

Adiabatic switching is founded in the adiabatic theorem of classical mechanics,[235] which states that certain constants of the motion may evolve invariantly in time. For semiclassical quantization, these are the good actions variables.[219–223] In AS a time-dependent Hamiltonian is given by

$$H(t) = H_0 + s(t)(H - H_0), \tag{10.2}$$

where H_0 is the zeroth-order Hamiltonian (which is the harmonic Hamiltonian in this work), and $s(t)$ is a switching function that varies monotonically from 0 to 1 over a finite time interval T and remains 1 when $t > T$. The familiar harmonic zeroth-order Hamiltonian for $3N$-6 vibrational modes is given by

$$H_0 = \sum_{i=1}^{3N-6} \left(\frac{P_i^2}{2} + \frac{\omega_i^2 Q_i^2}{2} \right), \tag{10.3}$$

where Q_i are the mass-scaled normal modes and ω_i are the corresponding harmonic frequencies. The switching function $s(t)$ in this work is chosen as[223,226]

$$s(t) = \frac{t}{T} - \frac{1}{2\pi}\sin\frac{2\pi t}{T}, \tag{10.4}$$

where T is the switching time.

The semiclassical quantization conditions of H_0 are the usual ones,

$$\oint P_i dQ_i = 2\pi(n_i + \frac{1}{2})\hbar, \tag{10.5}$$

where n_i are integers and these variables are classical actions. Since the H_0 is a separable harmonic-oscillator Hamiltonian, the semiclassical quantization can be achieved trivially, as shown next.

10.3.2 Quantization of H_0

To proceed, a standard normal-mode analysis is performed at the minimum structure, where the trajectories are initiated, in the present case the minimum of syn-CH_3CHOO. A set of harmonic frequencies ω_i and the corresponding normal mode eigenvectors L_i ($i = 1,\ 2,\ \cdots,\ 3N-6$) are obtained.

For each mode, the normal coordinate Q_i and the corresponding momentum P_i are in terms of action/angle variables by

$$Q_i = \left[(2n_i+1)\hbar/\omega_i\right]^{1/2}\cos\left(\omega_i t + \phi_i\right), \tag{10.6}$$

$$P_i = -\left[(2n_i+1)\hbar\omega_i\right]^{1/2}\sin\left(\omega_i t + \phi_i\right), \tag{10.7}$$

where n_i is the integral quantum number of mode i, and ϕ_i is the phase of the oscillator i, which varies uniformly from 0 and 2π.

For sampling purposes, e.g., in QCT calculations, at $t = 0$, the phase for each mode is randomly sampled from a uniform distribution on $(0, 2\pi)$. Once the random phases ϕ_i are determined, Q_i and P_i are obtained and then transformed rigorously to the Cartesian coordinates q_i and momenta p_i via

$$\boldsymbol{q} = \boldsymbol{q_0} + \boldsymbol{M}^{-1/2}\boldsymbol{L}\boldsymbol{Q}, \tag{10.8}$$

$$\boldsymbol{p} = \boldsymbol{M}^{1/2}\boldsymbol{L}\boldsymbol{P}, \tag{10.9}$$

where $\boldsymbol{q_0}$ is a vector of the Cartesian coordinates of the minimum; $\boldsymbol{L}$ is a matrix consists of the normal mode eigenvectors; $\boldsymbol{M}$ is a diagonal matrix whose elements are the masses of atoms.

These are then the initial conditions for the AS trajectory. AS can be done with a single trajectory, or as is more typical, for an ensemble of trajectories.

10.3.3 Propagation

To be completely general, we implemented AS in Cartesian coordinates and momenta, and so the method can be applied to any molecule, cluster, etc. The equations of motion

$$\dot{q}_i = \frac{\partial H(t)}{\partial p_i}, \ \dot{p}_i = -\frac{\partial H(t)}{\partial q_i} \tag{10.10}$$

are integrated numerically, where $H(t) = H_0 + s(t)(H - H_0)$. Using the rigorous H does lead to a complication, owing to the neglect of vibrational angular momentum terms in the above expression for H_0.[236] Without treating this coupling, evaluation of normal

modes from time-evolving Cartesian coordinates can become numerically very inaccurate as rotational normal modes develop non-zero values. To correct this the coordinate system is transformed to the original Eckart frame. The transformation to the Eckart frame is done by a rotation of the coordinate system $\boldsymbol{r}_i$ to a new one that satisfies the Eckart conditions. The rotation matrix, $\boldsymbol{C}$ is computed using the method of Dymarsky and Kudin[237]. The details of the procedure used in our group are given elsewhere,[205,238] where the context was the final state analysis of products. In brief, $\boldsymbol{C}$ is determined from the equation

$$\sum_{i=1}^{N} m_i \mathbf{r}_i^{eq} \times (\boldsymbol{C}\mathbf{r}_i - \mathbf{r}_i^{eq}) = 0, \tag{10.11}$$

where $\boldsymbol{r}_i^{eq}$ is the reference configuration where the normal mode analysis was performed.

Once the Cartesian coordinates in the Eckart frame are obtained, the reference set of normal coordinates is obtained as usual, and V_0 can be easily evaluated as

$$V_0 = \sum_{i=1}^{3N-6} \frac{\omega_i^2 Q_i^2}{2}. \tag{10.12}$$

and subtracted from the full potential. Note that the transformation to the Eckart frame is only performed to evaluate the harmonic potential, otherwise the classical propagation proceeds as usual in the original set of Cartesian coordinates and momenta. This transformation is easily implemented into any trajectory code.

10.3.4 Computational details

The PES of the CH_3CHOO we use is a full-dimensional one that is a permutationally invariant fit[26] to roughly 160 000 electronic energies (roughly 140 000 CCSD(T)-F12b/HaDZ and 20 000 CASPT2(12,10)/cc-pVDZ energies).[233,234,239,240] The first version

of the PES was restricted to the region of the syn-CH_3CHOO minimum and used in joint theory/experimental spectroscopic work.[239] The more recent one is far more extensive, describing region of the syn-CH_3CHOO minimum, the saddle point region to VHP, the VHP region, and dissociation to the products OH+vinoxy.[233,234] This recent version is employed in this work.

We performed AS calculations for syn-CH_3CHOO, with different switching times T for the zero-point state and one fundamental. The velocity Verlet algorithm was applied to numerically integrate the equations of motion, and the gradient at each step was calculated by numerical finite difference. The step size used in all the trajectory calculations was 0.06 fs. For each quantized state, three different switching times have been used, namely, T = 1209.5, 2419, and 4838 fs. After the full potential was completely turned on, the trajectories were further propagated for another T using the full PES.

10.4 Results and Discussion

Before presenting results from the AS, we show in Figure 10.1 a schematic of the potential energy of the syn-CH_3CHOO–VHP isomerization along a rectilinear reaction coordinate, namely the imaginary frequency normal coordinate of the saddle point separating these minima, denoted Q_{im}.[241] The potential shown is the one minimized with respect to the 3N-7 (17) normal modes of the saddle point for each value of Q_{im}. The Q_{im} can be viewed as the rectilinear reaction coordinate for the syn-CH_3CHOO–VHP isomerization. In the present context it is a very convenient way to monitor the approach to the isomerization barrier and indeed isomerization as well by determining the Q_{im} value at each step in a trajectory. This figure also shows the rigorous quantum ZPE of the syn-CH_3CHOO (12851 cm^{-1}), and it is almost twice as large as the barrier (6524 cm^{-1}) that separates

the syn-CH_3CHOO and the VHP.

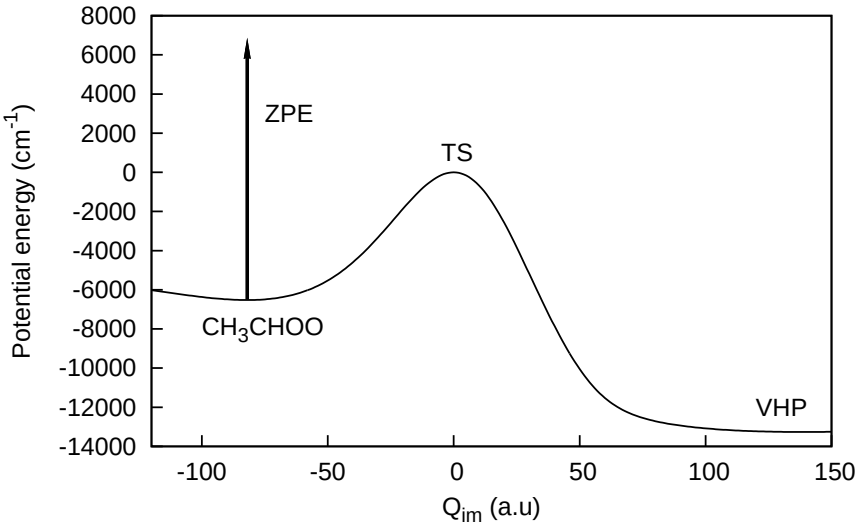


Figure 10.1: Schematic of the isomerization of syn-CH_3CHOO to VHP

Unlike methane, to which we successfully applied AS, syn-CH_3CHOO has a large-amplitude CH_3 torsion mode. When the excitation of this mode is large due to ZPE leak during the switching time, displacement of this torsion mode is large, and using the normal modes of the original reference geometry, as we do, can be problematic as there are three equivalent minima for the torsion. This is a fairly rare event and from previous diffusion Monte Carlo calculations of the ZPE,[240] we know that torsional mode is fairly localized (as expected given that the torsional barrier is 730 cm^{-1}). Thus, we discard trajectories which display this problematic behavior.

The vibrational energy of the syn-CH_3CHOO as a function of time for a single trajectory and three total switching times is shown in Figure 10.2. As seen, the initial harmonic ZPE (12981 cm^{-1}) changes as the potential is switched to the full potential and the vi-

brational energy of the syn-CH_3CHOO converges to a value that is close to the exact quantum ZPE of 12851 cm^{-1}. When averaging over all the trajectories (about 10 trajectories for each switching time), we obtained a ZPE of 12834 cm^{-1}, which is about 17 cm^{-1} below the quantum value. The decrease in energy with respect to time is generally monotonic with small oscillations. This result is in good accord with the one we observed earlier for the ZPE of CH_4, where the AS value is roughly 30 cm^{-1} below the exact one.[227]

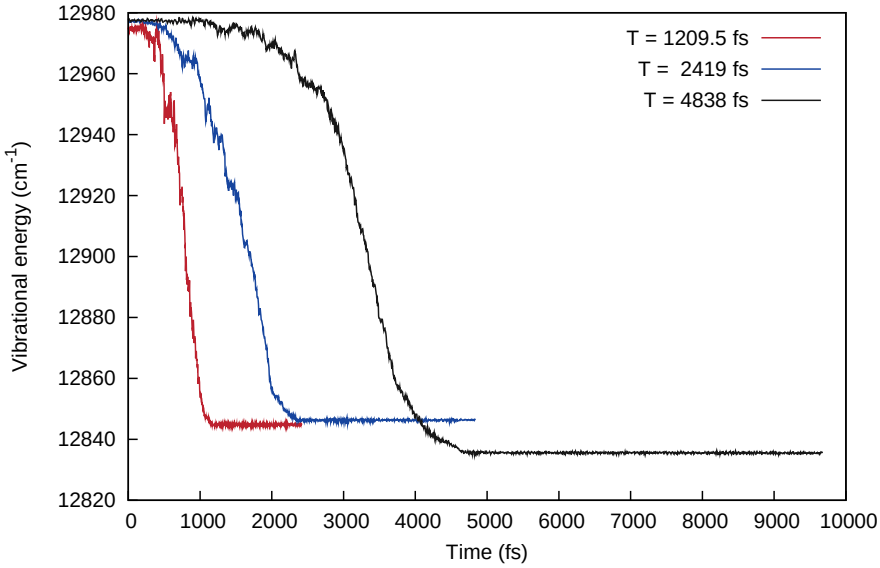


Figure 10.2: Zero-point vibrational energy of syn-CH_3CHOO using adiabatic switching for three switching times.

Figure 10.3 shows the time dependence of the Q_{im} value in an adiabatically switched trajectory. From the figure, we can see that the Q_{im} value oscillates around -80 (the syn-CH_3CHOO minimum), and it stays in the negative region, i.e., the syn-CH_3CHOO well. Note that the Q_{im} value has to reach 0.0 in order to reach the isomerization barrier to the VHP; therefore, even with a ZPE that is twice as large as the barrier, the syn-CH_3CHOO

→ VHP isomerization did not occur in these AS trajectories.

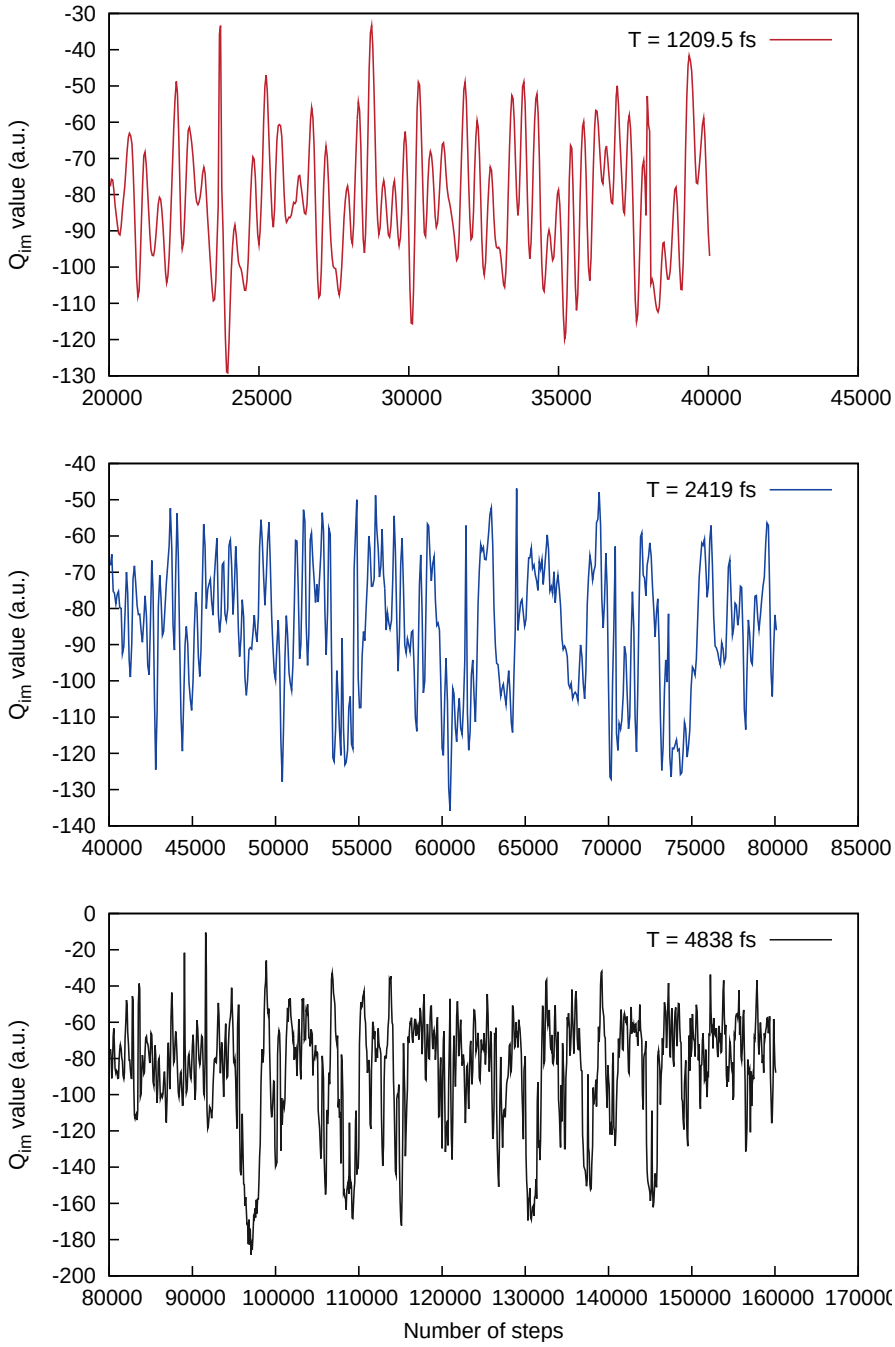


Figure 10.3: Time dependence of Q_{im} for zero-point adiabatically switched trajectory.

Table 10.1: The projection of the syn-CH_3CHOO normal modes on the imaginary-frequency mode of the saddle point.

Mode	frequency (cm^{-1})	Projection
1	211	-0.3207
2	301	-0.2729
3	445	0.0872
4	665	-0.1893
5	728	-0.1490
6	934	0.1002
7	981	0.1349
8	1044	-0.1916
9	1097	0.5280
10	1293	-0.0748
11	1428	0.0998
12	1456	0.3810
13	1474	-0.4346
14	1523	0.1992
15	2948	-0.0979
16	3094	0.0735
17	3110	0.0767
18	3189	0.0029

Next, we consider the isomerization when a promoting vibrational mode of the syn-CH_3CHOO is excited. The identification of this mode is based on a simple theory where the projections of the normal mode vectors onto the imaginary-frequency normal mode of the saddle point are used to determine, within a simple harmonic model, which modes promote isomerization when excited.[242] Note Jiang and Guo also proposed a similar model.[243] Therefore, we calculated the projection of the 18 modes of syn-CH_3CHOO onto the imaginary-frequency mode, and they are presented in Table 10.1. We find that mode 9 has the largest projection. Therefore we ran AS trajectories with this mode being excited. This can be done by simply setting $n_9 = 1$ in the initial quantization of the harmonic

Hamiltonian H_0. The vibrational energies of the syn-CH_3CHOO as a function of time for $T = 1209.5$ fs are shown in Figure 10.4. The average final energies of all these excited state trajectories (about 10 trajectories for each switching time) are 13923 cm^{-1}, leading to a semiclassical energy of 1089 cm^{-1} for the excitation of mode 9. In comparison, the excitation energy of this vibrational mode is 1097 cm^{-1} from quantum calculation.[239] Our semiclassical result is in very good (probably fortuitous) agreement with the quantum calculation. Figure 10.5 shows the time dependence of the Q_{im} value after the switching from trajectories with mode 9 excited. The Q_{im} values still stay negative, meaning that the isomerization did not occur in these trajectories.

The time dependence of $Q_{im}(t)$ for the AS trajectories is interesting, beyond serving as a diagnostic for unphysical isomerization. The time-dependence is clearly complex and in the future we plan to subject it to standard Fourier analysis. As well, the inner turning points are configurations (caustics) where the trajectory is reflected from the isomerization barrier. These can be used in simple estimates of the tunneling probability for isomerization. This would, in a sense, be the extension of our harmonic projection theory to obtain tunneling rates.[242] It would also be close in spirit to earlier work of Makri and Miller.[244] We plan to investigate this in the future.

We conclude this Discussion with some general comments. While AS is successful in obtaining vibrational energies in good agreement with quantum calculations, it is not assured that the switching has produced invariant tori in phase space with no traces of "chaotic" dynamics. Unfortunately, some of diagnostics such Poincaré surface-of-section analysis cannot be applied here owing the very large dimensionality of the phase space. However, the lack of isomerization does strongly suggest that the dynamics is more "regular" than "chaotic". Further, it would be interesting to see at what energy prompt isomerization occurs and indeed to continue the dynamics to the formation of the OH+vinoxy

products and to investigate the predictions of John Light's elegant phase space theory of chemical reactions,[245,246] which is directly applicable to this dissociation.

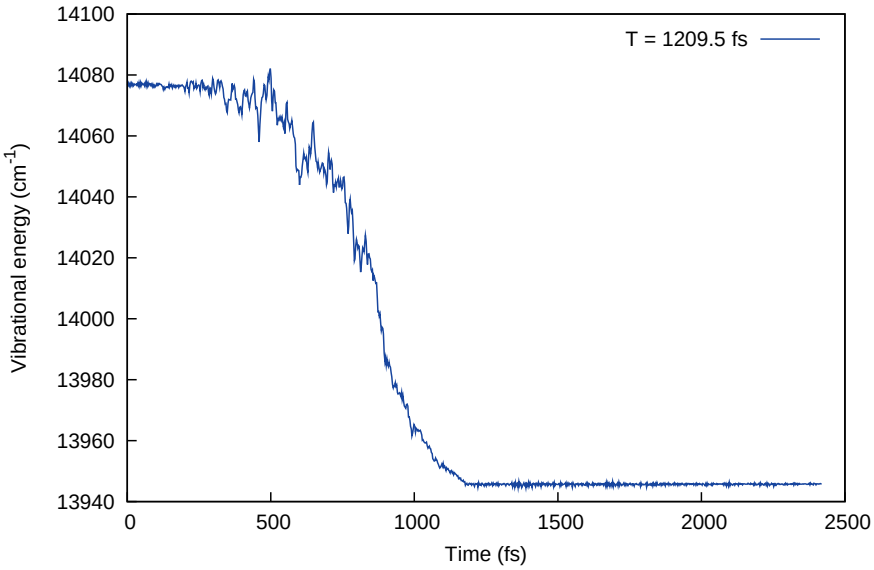


Figure 10.4: Vibrational energy of syn-CH_3CHOO with mode 9 excited, using adiabatic switching with one switching time.

10.5 Summary and conclusions

In this article, we presented a calculation of the zero-point energy of CH_3CHOO for the ground state and one fundamental using adiabatic switching with an existing *ab initio*-based potential energy surface. Adiabatic switching was done using a standard separable harmonic normal mode zeroth-order Hamiltonian to the full one in Cartesian coordinates. Details of the procedure were given, with special note taken of the transformation from Cartesian coordinates of the adiabatic switching trajectory to the reference normal mode coordinates. Comparison with available accurate quantum energies showed good

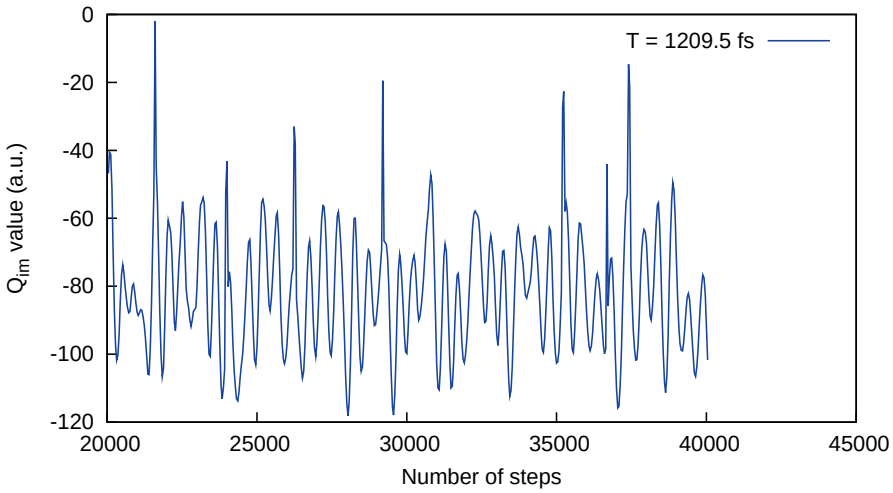


Figure 10.5: Time dependence of Q_{im} for excited adiabatically switched trajectory.

agreement. A major point of the paper was the demonstration of the lack of zero-point energy "leak" causing unphysical isomerization to vinyl hydroperoxide, even though the internal energy of the syn-CH_3CHOO is roughly twice the isomerization barrier. Our future work will combine adiabatic switching and our simple 1-D approach[242] to calculate the tunneling effect in the $CH_3CHOO \rightarrow$ VHP isomerization.

Chapter 11

Quantum Zero-point Energies of Methanol and Deuterated Methanol

11.1 Chapter Abstract

Diffusion Monte Carlo (DMC) simulations have been used to obtain quantum zero-point energies of methanol and all its isotopologs and isotopomers, using a new, accurate semi-global potential energy surface. This potential energy surface is a precise, permutationally invariant fit to 6,676 *ab initio* energies, obtained at the CCSD(T)-F12b/aug-cc-pVDZ level of theory. Quantum zero-point energies of deuterated methanol isotopomers are very close to each other and so a simple statistical argument can be used to estimate the populations of each isotopomer at very low temperatures. The DMC simulations also indicate that there is virtually zero probability for H/D exchange in the zero-point state.

11.2 Overview

Isotopic substitution plays a significant role in astrochemistry and biology. For example, in spite of the very low abundance of deuterium (about 10^5 times less than hydrogen) in the universe, D-containing isotopologues are commonly found in many interstellar and pre-stellar environments. Singly, doubly and triply deuterated molecules have been detected in various interstellar environments during the last couple of decades.[247] Methanol and ammonia are the two molecules which show the highest D-enhancements[248–250]. In 1988, singly deuterated methanol, CH_3OD, was detected,[251] and in 1993 CH_2DOH was observed in the Orion Compact Ridge.[252] The fractional H/D ratio for $[CH_3OD]/[CH_3OH]$ in the Orion Compact Ridge was found to be about 0.01–0.06, and about 1.1–1.5 for $[CH_2DOH]/[CH_3OD]$ respectively. Detection of the doubly deuterated methanol (CHD_2OH) was reported by Parise *et al.*[253] in 2002 in solar-type protostar IRAS 16293-2422. The abundance ratio of CHD_2OH/CH_3OH was found to be about 0.2 ± 0.1. In this work, two singly deuterated methanol were also observed and the abundance ratios are $[CH_2DOH]/[CH_3OH] = 0.9 \pm 0.3$ and $[CH_3OD]/[CH_3OH] = 0.04 \pm 0.02$. Later, triply deuterated methanol, CD_3OH, was found in solar-type protostar IRAS 16293-2422 in 2004.[250] The observed fractional ratio of $[CD_3OH]/[CH_3OH]$ is about 1.4%. However, other deuterated isotopomers (CH_2DOD and CHD_2OD) of methanol have not been observed yet in the interstellar environment.

In order to explain the abundance of the deuterated methanol and the absence of CH_2DOD and CHD_2OD in interstellar environment, the reaction mechanism of methanol formation is an active area in contemporary research. The experimental observations are consistent with the formation of methanol from successive hydrogenations of CO by reaction with atomic H/D on grain surfaces, if the atomic D/H ratio is 0.1–0.3.[254] The

study of Whittet *et al.* also supports this model.[255] Theoretically, in 1997, Charnley *et al.*[256] calculated a CH_2DOH/CH_3OD ratio of 3:1 in the Orion Compact Ridge, so the calculated ratio conflicts the observed ratio (1.1–1.5) and the authors show that when these mantles are evaporated, the $[CH_2DOH]/[CH_3OD]$ ratio can be altered by gas-phase reactions. We will discuss the CH_2DOH/CH_3OD ratio later in the paper.

In addition to its importance in astrochemistry, methanol is one of the molecules that has large-amplitude torsional motion with a small electronic barrier (roughly 360 cm^{-1})) to this internal rotation.[49,257,258] The advent of modern spectroscopic techniques and computational hardware have enabled exploration of methanol in great detail. Several experimental investigations have been performed in the last two decades to understand the interaction of this large amplitude motion with other normal modes.[259–262] High-level theoretical works on vibrational and torsional energies of CH_3OH have produced several *ab initio*-based force fields and potential energy surfaces.[49,258,263–267] However, The theoretical and experimental description of the deuterated methanol species is very limited. Firstly, In 1974, Serrallach *et al.* have reported experimental vibrational frequencies of methanol and its all deuterated isotopologs.[259] They used these experimental data for the determination of significant valence force field parameters by a simultaneous least-square adjustment. A theoretical study of anharmonic vibrational frequencies of CH_3OH, CH_3OD,CD_3OH, and CD_3OD have presented by Halonen and co-workers[263] in 2000. They employed CCSD(T)/aug-cc-pVTZ level of theory to calculate the harmonic part and MP4(SDQ)/aug-cc-pVTZ level of theory to calculate the anharmonic part of the potential energy surface and all the vibrational terms are calculated by using second order perturbation theory. In 2005, Sibert reported the anharmonic vibrational frequencies of methanol and its three deuterated analogs (CH_3OD,CD_3OH, and CD_3OD) using a combination of perturbative and variation method[258]. The effect of deuteration

on torsional tunneling splitting is clearly a very sensitive probe of the torsional barrier and dynamics. Mode coupling effects for this 6-atom molecule require a high-dimensional potential energy surface (PES). Huang *et al.* reported a full-dimensional PES in 2007[49]. This PES, which was a permutationally invariant fit to 19,315 *ab initio* energies at the CCSD(T)/aVTZ level of theory, is limited to the minima of methanol potential. It was used in MULTIMODE-Reaction Path calculations[49,268] having good agreement with experiment. Later in 2013, another PES was developed by Qu *et al.*[17] and roughly 46,000 *ab initio* energies were fitted using CCSD(T)-F12b level of theory and aug-cc-pVDZ basis set. This latest PES describes the region of the minimum of the potential as well as the ($CH_3 + OH$) dissociation channel[17] with excellent agreement of IUPAC evaluation of dissociation energy.[269]

One aspect of deuteration that has not been investigated theoretically is the zero-point energy of all isotopomers of partially deuteratated isotopologs of methanol. Specfically, this is energy ordering of D-substituted (mono, double and triple) methanol isomers with a focus on which position the D atom prefers to attach (either with C or with O) for partial deuteration. In this case, there are two possibilities to form deuterated-CH_3OH from CH_3OH, namely CH_3OD vs CH_2DOH, CH_2DOD vs CHD_2OH and CHD_2OD vs CD_3OH. To address this precisely for the zero-point state, we use the rigorous Diffusion Monte Carlo (DMC) approach with a new, accurate potential energy surface.

The paper is organized as follows. In section 2, we briefly present the details of the new PES, followed by the computational details of DMC calculations. In section 3, we present the results of our DMC simulations of all the D-containing isotopologues of methanol and the ZPE differences from each other. In section 4, we present the conclusions and remarks.

11.3 Theory and Computational Details

11.3.1 Potential Energy Surface of CH_3OH

We reported a permutationally invariant PES for methanol that describes the dissociation to CH_3+OH in 2013, based on a dataset of 45,199 configurations and corresponding CCST(T)-F12b/aVDZ energies/[17] The PES used in this work is a fit to a subset of 6,676 configurations in the global minimum and torsional saddle point regions of methanol. This more local PES uses the same permutationally invariant polynomial fitting basis and the same maximum polynomial order. This leads to 3,250 linear coefficients which were determined by solving a weighted linear least-squares problem. Each energy was weighted by $0.1/(0.1 + \delta E)$, where δE is the energy relative to the minimum in hartree. The number of *ab initio* data points and the fitting root-mean-square (rms) error as functions of the energy are shown in Figure 11.1. The rms error below 25000 cm^{-1} is only about 8 cm^{-1}, and the overall rms is 12.3 cm^{-1}. These are all smaller than the semi-global one presented in 2007,[49] and are significantly smaller than those of the dissociable PES,[17] and therefore it is much more precise for zero-point properties studied in this work.

The torsional barrier of this PES is 355.2 cm^{-1}, and it is in good agreement with the 359.5 cm^{-1} from CCSD(T)/VQZ calculations.[263] The harmonic frequencies of the minimum and the torsional saddle point are listed in Table 11.1, together with the *ab initio* ones calculated at CCSD(T)-F12b/aVDZ level of theory. The differences between the PES frequencies and the *ab initio* ones are generally within 5 cm^{-1}.

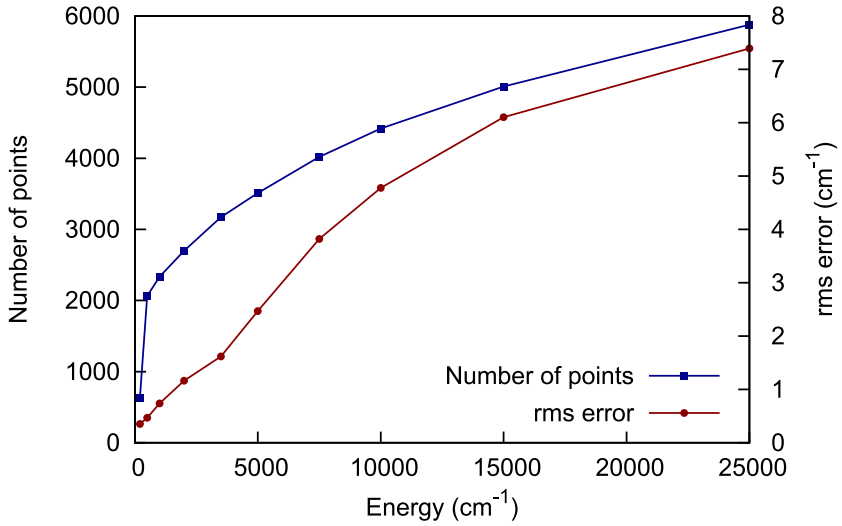


Figure 11.1: Number of *ab initio* data points and rms fitting error a function of the energy.

Table 11.1: Harmonic frequencies (in cm^{-1}) of the global minimum and saddle point of CH_3OH from the PES, and comparison with CCSD(T)-F12b/aVDZ calculations

	Minimum		Saddle point	
Mode	PES	*ab initio*	PES	*ab initio*
1	294	289	298*i*	284*i*
2	1068	1064	1068	1069
3	1091	1089	1094	1100
4	1179	1178	1188	1192
5	1384	1380	1361	1366
6	1483	1481	1486	1488
7	1505	1505	1498	1499
8	1517	1515	1524	1527
9	3014	3013	3028	3030
10	3070	3072	3097	3100
11	3133	3133	3103	3106
12	3865	3865	3899	3901

11.3.2 Diffusion Monte Carlo

In this work, we calculate rigorous quantum ZPEs of CH_3OH and its D-containing isotopologs and all isotopomers, using the Diffusion Monte Carlo procedure. The concept behind the DMC method is to solve time-dependent Schrödinger equation in imaginary time[55,56,124]. This imaginary time Schrödinger equation is solved by simulating a random walk of many replicas, "walkers", of the molecule, using a birth/death processes. This is a powerful method to extract the ground state wave function of the molecule from the probability distribution of the random walks, though it is computationally expensive as many walkers must be propagated for tens of thousands of steps in order to obtain an precise estimate of the ZPE.

The procedure can be illustrated in 1d for a mass m, subject to a potential V(x). The time dependent Schrödinger equation of this particle is written as

$$i\hbar\frac{d}{dt}\Psi(x,t) = \hat{H}\Psi(x,t), \tag{11.1}$$

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x), \tag{11.2}$$

where $\hat{H}$ is the Hamiltonian of the system. The wave function can written as

$$\Psi(x,t) = \sum_{n=0}^{\infty} c_n\phi_n(x)e^{-\frac{i}{\hbar}E_n t}, \tag{11.3}$$

where $\phi_n(x)$ are the eigenfunctions of the time-independent Schrödinger equation with the eigenvalues E_n

$$\hat{H}\phi_n(x) = E_n\phi_n(x) \tag{11.4}$$

and the order of the eigenvalues are follows as

$$E_0 < E_1 \leq E_2 \leq E_3 \leq .... \tag{11.5}$$

Next, we shift the energy scale by E_R such that $V \rightarrow (V(x) - E_R)$ and $E_n \rightarrow (E_n - E_R)$, where E_R is a reference energy. Then we introduce a new variable $\tau = it$ to transform real time to imaginary time. So, the 1-D time dependent Schrödinger equation becomes

$$\hbar \frac{d}{d\tau}\Psi(x,\tau) = \frac{\hbar^2}{2m}\frac{d^2}{dx^2} - [V(x) - E_R]\Psi(x,\tau), \tag{11.6}$$

and the the wave functions become

$$\Psi(x,\tau) = \sum_{n=0}^{\infty} c_n \phi_n(x) e^{-\frac{E_n - E_R}{\hbar}\tau}. \tag{11.7}$$

Clearly, as τ increases this sum converges to one term if $E_R = E_0$. Thus, $\Psi(x,\tau)$ converges to the ground state wave function $\phi_0(x)$ regardless of the choice of the initial wave function $\Psi(x,0)$, provided there is some overlap of the function with the ground state eigenstate. Here we use simple, unbiased DMC algorithm, proposed by Anderson[56].

This equation is solved by random walk propagation technique to simulate the evolution of Schrödinger equation in imaginary time. Initially, we generate $\mathcal{N}_0$ replicas from the same initial geometry of a molecule, as the distribution of $\Psi(x,0)$. This geometry is typically at the minimum of interest, which in this case is the global minimum. The replicas are described through a position $x_n^{(j)}$, where the suffix n represent the diffusive displacement and (j) represent the replicas. In the first step of the propagation, $\mathcal{N}_1$ replicas are generated by adding random displacement to each position of the molecule. These

position vectors x_1^j are determined by the following equation,

$$x_1^{(j)} = x_0^{(j)} + \sigma * \rho_1^{(j)} \tag{11.8}$$

where $\sigma = \sqrt{\frac{\hbar * \Delta\tau}{m}}$ and $\rho_1^{(j)}$ is a Gaussian random number with zero mean and variance equal to one and $\Delta\tau$ is the time step. This is the one step diffusive process of the system of replicas. Once the new replicas are determined, in the next step we calculate the weight, $W(x_1^{(j)})$.

$$W(x_1^{(j)}) = e^{-\frac{V(x_1^{(j)}) - E_R}{\hbar}\Delta\tau} \tag{11.9}$$

After that, another random number is choose to describe the birth/death process. If $[1 - W(x_1^{(j)})$]is grater than this random number then the replica (j) will remain and if $[1 - W(x_1^{(j)})]$ is less than this random number then the replica (j) will die; if $W(x_1^{(j)})$ grater than 1.0 then the replica (j) is duplicated. Thus total number of replicas $\mathcal{N}_1$ is determined after one diffusive process. Due to the birth/death process, the total number $\mathcal{N}_1$ of replicas will change from its original $\mathcal{N}_0$ value. According to the above equation (7), for reference energy E_R smaller then the ground state energy E_0, the replicas eventually all die; for E_R value larger than ground state energy E_0, the number of replicas $\mathcal{N}_n$ will increase. Only when $E_R = E_0$, the number of replicas fluctuate around an average value of $\mathcal{N}_0$. One can adjust the value of E_R by redefining the reference energy to keep the total number of replicas approximately constant, which follows as

$$E_R = \langle V \rangle + \alpha \left(1 - \frac{\mathcal{N}_1}{\mathcal{N}_0}\right). \tag{11.10}$$

Where, α is a positive parameter, is called the "feedback" parameter. α value is chosen to adjust the reference energy, E_R as to minimize the statistical fluctuations in replicas,

$\mathcal{N}_0$.[56,124]

11.3.3 Computational details

In this work, for methanol and all its isotopologs, we used 30,000 walkers, and they were propagated for 60,000 time steps with the step size of 5.0 a.u; 25,000 steps were used to equilibrate the walkers, and the reference energies in the remaining 35,000 steps were used to compute the ZPE. In all the simulations, the value of α was chosen to be 0.1.

For each isotopomer, 30 DMC simulations were carried out, and the final ZPE is the average of the 30 simulations. Statistical uncertainty of the DMC energy is defined as the standard deviation of DMC energies over the total number of simulations. This is written as

$$\Delta E = \sqrt{\frac{1}{30}\sum_{i=1}^{30}(E_i - \bar{E})^2}, \tag{11.11}$$

where $\bar{E}$ is the average energy over the 30 simulations.

11.4 Results and Discussion

To begin, we present the harmonic frequencies of CH_3OD, CH_2DOH, CH_2DOD, CHD_2OH, CHD_2OD, CD_3OH and CD_3OD in Table 11.2. From the first two columns we can determine the effect of locating the single D on the O atom versus on the methyl group. In particular, we see that the OD-stretch is 2814 cm^{-1} whereas the CD-stretch is 2233 cm^{-1}. Comparing these numbers to the corresponding ones for CH_3OH in Table 11.1, we see a roughly a significantly larger downshift for CH_3OD than for CH_2DOH. However, the effect on the ZPEs must take account of all the modes.

These effects are given at the harmonic level and from the DMC calculations in Table

Table 11.2: Harmonic frequencies (cm^{-1}) of isotopologs of methanol.

Mode	CH_3OD	CH_2DOH	CH_2DOD	CHD_2OH	CHD_2OD	CD_3OH	CD_3OD
1	234	288	226	284	218	279	214
2	881	929	843	911	800	866	785
3	1077	1056	979	933	961	910	910
4	1179	1093	1082	1047	1034	1010	1004
5	1255	1294	1187	1124	1094	1089	1070
6	1483	1378	1347	1293	1121	1093	1089
7	1505	1406	1385	1342	1348	1160	1102
8	1515	1499	1499	1415	1372	1326	1161
9	2814	2233	2233	2191	2210	2163	2163
10	3014	3042	2814	2278	2316	2278	2278
11	3070	3130	3042	3125	2815	2322	2322
12	3134	3865	3131	3865	3049	3865	2815

Table 11.3: Harmonic and DMC ZPE (cm^{-1}) of methanol and its all isotopologs. Values inside the parenthesis represent the statistical uncertainties in the DMC result.

Molecule	Harmonic ZPE	DMC ZPE
CH_3OH	11303	11107 (1.6)
CH_3OD	10581	10414 (1.7)
CH_2DOH	10607	10426 (1.5)
CH_2DOD	9884	9728 (1.7)
CHD_2OH	9903	9739 (1.5)
CHD_2OD	9168	9042 (2.0)
CD_3OH	9181	9044 (1.9)
CD_3OD	8455	8346 (1.4)

11.3. As expected the harmonic ZPEs are all higher than the rigorous DMC ones. The anharmonic ZPE of CH_3OH is 11107 cm^{-1} with a standard deviation of 1.6 cm^{-1} from 30 DMC simulations, whereas the harmonic one is 11303 cm^{-1}, which is about 200 cm^{-1} higher. This DMC ZPE perfectly matches with our previous work (11108 cm^{-1}).[17] The DMC ZPEs for mono-deuterated methanol, CH_3OD and CH_2DOH, are 10414 and 10426 cm^{-1} respectively, and their harmonic ones are 10581 and 10607 cm^{-1} respectively. Considering the differences in the ZPEs of the isotopomers, the harmonic result is 26 cm^{-1}, which is roughly twice the difference in the DMC anharmonic ZPEs. This indicates that these two isotopomers are very close to each other in energy. This statement is also true for doubly and triply deuterated methanol isotopomers. The difference of anharmonic ZPEs is only 11 cm^{-1} between CHD_2OH and CH_2DOD, and is only 2 cm^{-1} between CD_3OH and CHD_2OD, which is within the statistical uncertainty. For singly and doubly deuterated methanol, the $\sim$10 cm^{-1} difference in the ZPE is still larger than the statistical error of the DMC simulations, so we are able to determine that the lowest-energy isotopomers always have the deuterium in the OD position. On the other hand, for triply deuterated methanol, the 2 cm^{-1} difference is within the statistical uncertainty. Therefore, it is impossible to say which isotopomer is the most stable. Note that even though the harmonic approach overestimates the ZPE by roughly 200 cm^{-1}, it also predicts small difference in ZPEs between the two isotopomers for the same isotopolog of methanol.

Given the small differences in the DMC ZPEs, one can inquire about the robustness of the DMC results with respect to possible dynamical D/H exchange. Since this involves the breaking of two covalent bonds, one would expect that it is essentially a "zero-probability" process. We verified this in the DMC calculations for CH_2DOH and CH_3OD in the following way. We determined the distribution of CD and OD bond lengths in CH_2DOH and CH_3OD for all walkers along the DMC trajectories. If exchange happens

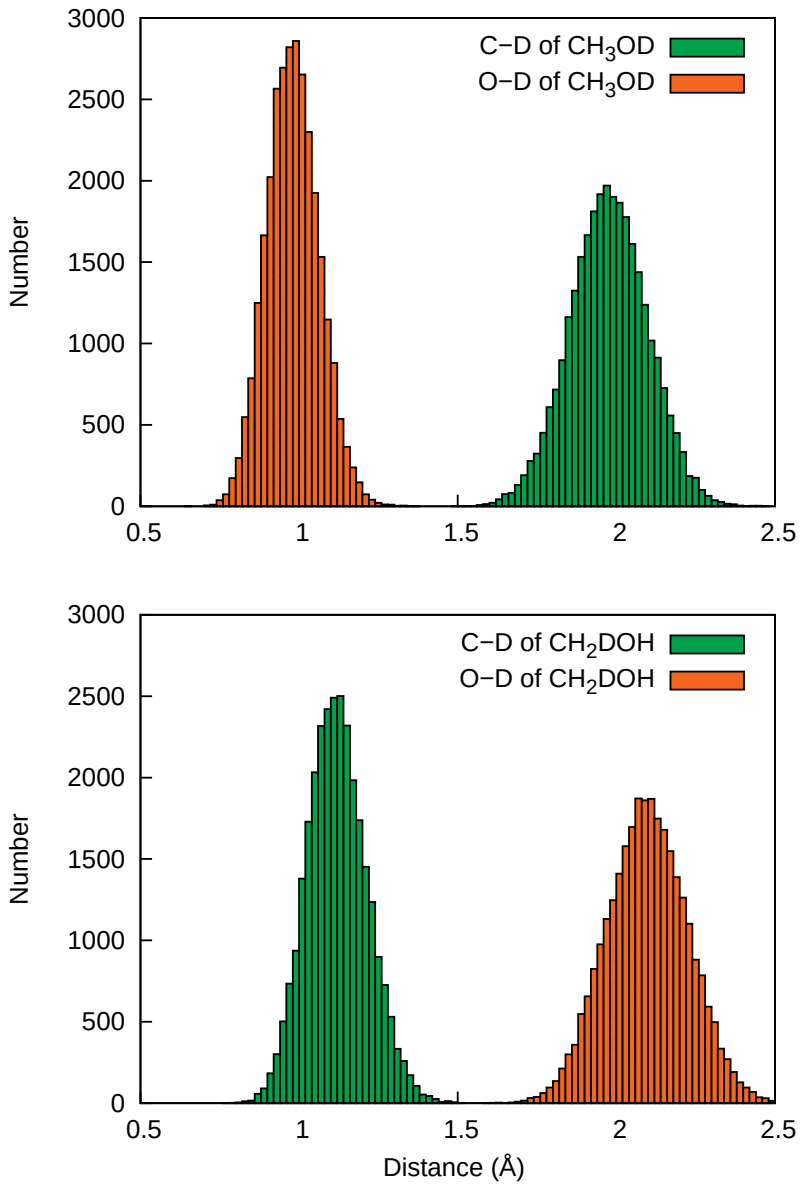


Figure 11.2: Distribution of C-D and O-D distance of CH_3OD and CH_2DOH. 29,619 and 29,696 configuration are selected from the DMC trajectory for CH_2DOH and CH_3OD, respectively.

with substantial probablity we would expect to see two peaks in the distribution of the CD/OD bond length. If exchange happens with small probabilty we would expect to see some overlap of the distributions; this would be necessary but not sufficient indicator of possible exchange. The distributions are shown in Figure 2, where only one peak is observed for each isotopomer. As expected the peaks are near the equilibrium CD/OD bond lengths. Further, there is no overlap of the green and red distributions. Thus, as expected the H/D exchange is dynamically a very low probability process and so each isotopomer would be expected to be stable in the ground vibrational state.

Next we discuss the possible implications of the DMC calculation on the abundance ratio of $[CH_2DOH]/[CH_3OD]$ in the interstellar medium discussed in the Introduction. As noted, there is some discrepancy between the observed ratio, which is roughly 1.5:1 and the one from a model, which is 3:1. We have shown that the ZPEs for these are essentially equal and so the population ratio at 0 K would be determined from a simple statistical argument which gives 3:1 for $[CH_2DOH]/[CH_3OD]$. That this agrees with the ratio obtained from an elaborate kinetic model with no consideration of ZPEs is certainly a coincidence. However, it seems like a reasonable "prior" expectation for this ratio.

11.5 Summary and Conclusions

Here, we presented the anharmonic ZPEs of methanol and its D-substituted isotopologs, using Diffusion monte carlo simulations and a new semi-global potential energy surface. The anharmonic ZPE of methanol decrease as the number of deuterium increases, which is expected, however the ZPEs of isotopmers are nearly equal and become essentially same for the triply D-substitued methanol. Our result supports the co-existence of two different isotopomers of singly, doubly and triply deuterated methanol in the cold interstellar

environments.

Chapter 12

Vibration-facilitated roaming in the isomerization of CO adsorbed on NaCl

12.1 Chapter Abstract

The desire to better understand the quantum nature of isomerization led to recent experimental observations of the vibrationally induced isomerization of OC-NaCl(100) ('C-Down') to CO-NaCl(100) ('C-Down'). To investigate the mechanism of this isomerization, we performed dynamics calculations using finite $(CO\text{-}NaCl)_n$ cluster models. We constructed new potential energy surfaces for CO-NaCl and CO-CO interactions using high-level ab initio data and report key properties of the bare CO-NaCl potential energy surface, which show much in common with the experiment. We investigated the isomerization dynamics using several cluster models and, in all cases, isomerization was seen for highly excited CO vibrational states, in agreement with experiments. A detailed

examination of the reaction trajectories indicates that isomerization occurs when the distance between CO and NaCl is larger than the distance at the conventional isomerization saddle point, which is a strong indicator of 'roaming'.

12.2 Overview

Isomerization is a pervasive phenomenon in chemistry and yet it has rarely been observed in real time experimentally. Theoretical work has been extensive on this topic; however, few quantum dynamical studies of isomerization have been reported. The isomerization of HCN to HNC is one notable exception, where isomerizing quantum states were reported using an accurate potential.[132,270] Another example is isomerization between acetylene and vinylidene, where theory was able to assign vibrational features in photodetachment of vinylidene anion.[271,272]

Recent experiments reported isomerization of CO adsorbed on NaCl(100)[273] (flipped to OC) for highly excited vibrational states of CO.[274] By resolving small differences in the emission spectra of CO and OC adsorbed on the surface the authors conclusively observed the isomerization from "C-down" to "O-down" on the surface. This experiment is a major step in observing isomerization dynamically, under state-controlled conditions in a condensed phase environment. The mechanism for this isomerization is complex as it involves excitation of the CO to high vibrational states via CO?CO energy pooling, followed by isomerization, subsequent emission and perhaps (partial) stabilization of the O-down isomer.

Several aspects of the experiment are important for theoretical modeling. First, it was determined that in the energetically favored C-down orientation[275] the CO bond axis makes an angle of approximately 25 deg relative to the surface normal.[273,275] For

the O-down isomer that angle is approximately 45 deg.[273] Second, the fundamental CO frequency, of the C-down isomer is blue-shifted by 7.6 cm^{-1}, whereas the O-down isomer fundamental is red-shifted by 9.3, cm^{-1}, relative to free CO.[273] Third, isomerization is only observed for the C-down isomer for CO for vibrational greater than roughly 20. And fourth the O-down isomer is also highly vibrationally excited, as evidenced by the overtone emission spectrum of that isomer.

In this Report we present a dynamics study of CO isomerization on the NaCl(100) surface by modeling the system by a finite CO-NaCl cluster. We develop a potential energy surface (PES) for the cluster that consists of a new *ab initio* CO-NaCl PES (for rigid NaCl) and a re-fit, using permutationally invariant polynomials (PIPs) to high-level *ab initio* energies for the CO-CO interaction reported recently in a precise PIP-Neural Network (PIP-NN) form.[276] Dynamics calculations are done for the bare CO-NaCl, a cluster where a central CO is surrounding by four neighboring COs and a cluster with 12 surrounding COs. Isomerization is found in all cases; however, there are quantitative differences. The results for the largest cluster model are closest to the experiment. In cases isomerization is found to occur at distances from the NaCl that strongly point to a roaming mechanism, which is vibrationally facilitate. Calculated emission spectra corresponding to experiment,[273] are also given.

In the next section we present essential details of the new PES for the cluster, key properties and how they relate to the experiment. Quasiclassical trajectory calculations of the isomerization are reported and interpreted in terms of a large amplitude roaming-type mechanism that opens up upon high vibrational excitation of the CO stretch.

We note previous theoretical work on CO adsorbed on NaCl(100). These include a study of the structure of CO monolayer adsorbed on NaCl(100) from molecular dynamics[277] using model potentials, a perturbation theory study of an adlayer CO on NaCl(100)

with consideration of CO-CO coupling,[278] a sophisticated kinetic model of CO-CO energy pooling,[279] a study of infrared fluorescence of a monolayer of ^{13}CO:NaCl(100) from highly vibrational excited states,[280] and a study of structures, energetics, and vibrational Davydov splittings.[281] None of these considered CO/OC isomerization. While the current paper was being revised, some of the current authors also reported PESs for CO interacting with NaCl(100) using density functional theory (DFT) and implications in isomerization of vibrationally excited CO.[282] The DFT PESs assume periodic boundary conditions, and thus are not suitable for dynamics calculations of CO isomerization.

12.3 Computational Details

12.3.1 Potential Energy Surface for CO-NaCl

In order to model the isomerization dynamics a potential energy surface was developed for CO-NaCl. The NaCl distance was held fixed at 2.82 angstrom, which is the distance in NaCl(100). In order to investigate the dynamics of highly vibrationally excited CO states, the PES must describe highly compressed and stretched CO. It is known that for such large displacements the ground electronic state of CO develops multi-reference character.[283] Rather than develop an expensive multi-reference CI-based PES, we addressed this challenge using a "plug and play" form for the PES.[284] Thus the total potential can be written as

$$V_{tot} = V_{int} + V_{\mathrm{CO}}(r_{CO}) + V_{\mathrm{NaCl}}, \tag{12.1}$$

where $V_{\mathrm{CO}}(r_{CO})$ is the potential for isolated CO as a function of r_{CO} (the CO inter nuclear distance) we developed based on multi-reference configuration interaction (MRCI) calculations. The interaction potential, V_{int} is given by the difference of the full potential

energy and the potential energy of just the CO at a general configuration of CO and NaCl. V_{NaCl} is the NaCl potential which is constant as the NaCl distance is held fixed at 2.82 angstrom and for convenience we have set this potential to be zero. As CO and NaCl separate V_{int} goes to zero and this is the origin of the term "interaction potential". Finally, we note in passing that a similar analysis of the CO potential was also used in Ref. 282.

We describe first the interaction potential and then the isolated CO potential.

Database and fit for the interaction potential

To begin, we performed constrained (fixed NaCl) geometry optimizations of the C and O-down minima at the CCSD/aug-cc-pVDZ level of theory using the MOLPRO quantum chemistry package.[87] These are used as the reference configurations for direct-dynamics (AIMD) calculations described below. These are used to establish a database of energies and gradients for fitting. Before we give the details of the AIMD calculations, which were done with MOLPRO, we note that the interaction potential is equal to the total potential minus the potential of the CO for a given CO-NaCl configuration. The same applies to the gradient.

The dataset of energies and gradients was obtained using direct dynamics, with the following sets of initial conditions. In one case, at each minimum the CO bond was stretched to 1.5 Angstrom with additional kinetic energies of 1000, and 5000 cm^{-1}. The six components of the momenta of C and O were determined by randomly subject to the constraint of the given kinetic energy. For NaCl the bond length was fixed at 2.82 angstrom and the components of Na and Cl momenta were set to zero at each time step. Trajectories were propagated for each minimum for 3000 time steps with the step size of 5.0 a.u. (about 0.12 fs).

Another batch of trajectories was run at each minimum but with the CO distance initially at the equilibrium value at each minimum. The trajectories were run with initial kinetic energies of 10000, 15000, and 30000 cm^{-1} relative to the C-down and O-down minimum and again with random sampling of the momenta. Seven trajectories were run at each minimum. These trajectories were also propagated for 3000 total time steps.

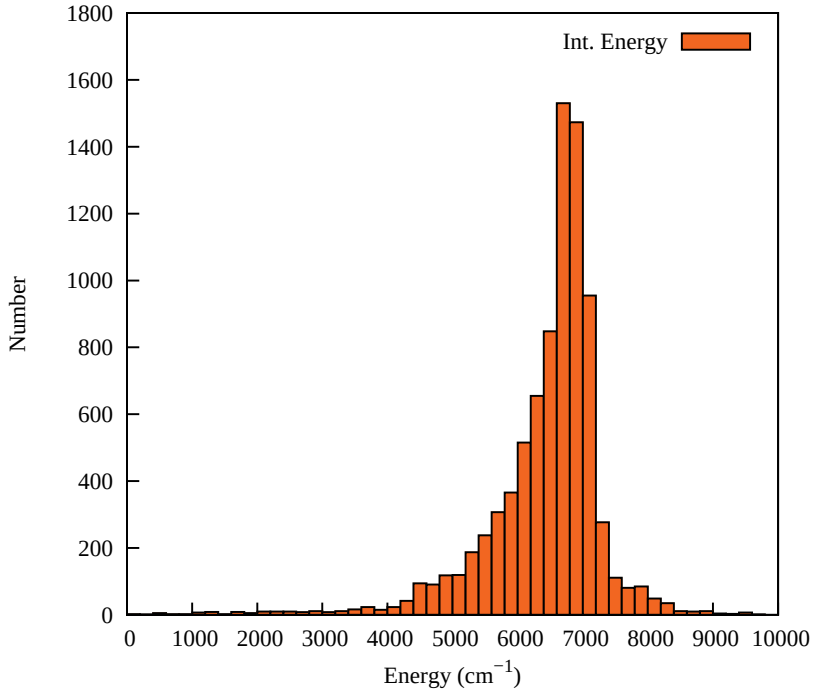


Figure 12.1: Distribution of interaction energies (cm^{-1}) relative to the minimum value.

These sets of trajectories provide large coverage of the configuration space for both C-down and O-down isomers and also for highly vibrationally excited CO. Sparse sampling of the complete data set was done so as to avoid nearly redundant data. The final data set consists of 8,592 energies and 103,104 gradients for a total data size of 111,696. The distributions of (interaction) energies used for the fitting database is shown in Fig. 12.1.

For reference histograms of the total potential energy (from CCSD calculations) are given in Fig. 12.2 for trajectories initiated at the indicate minimum.

These data were fit using our recent permutationally invariant polynomial (PIP) approach that incorporates gradients in the fitting.[285] These polynomials are functions of Morse variables, as explained in detail elsewhere.[26,65] The software to include gradients in the PIP basis fits is an extension of the monomial symmetrization MSA software[27] and is freely available.[123] The fit was done using a maximum polynomial order of 7. This leads to 1716 linear coefficients that were determined by solving the standard linear least-square problem. The root-mean-square (RMS) errors of this fitted PES are 27 cm^{-1} for energies and 22 cm^{-1}/bohr for all the gradient components.

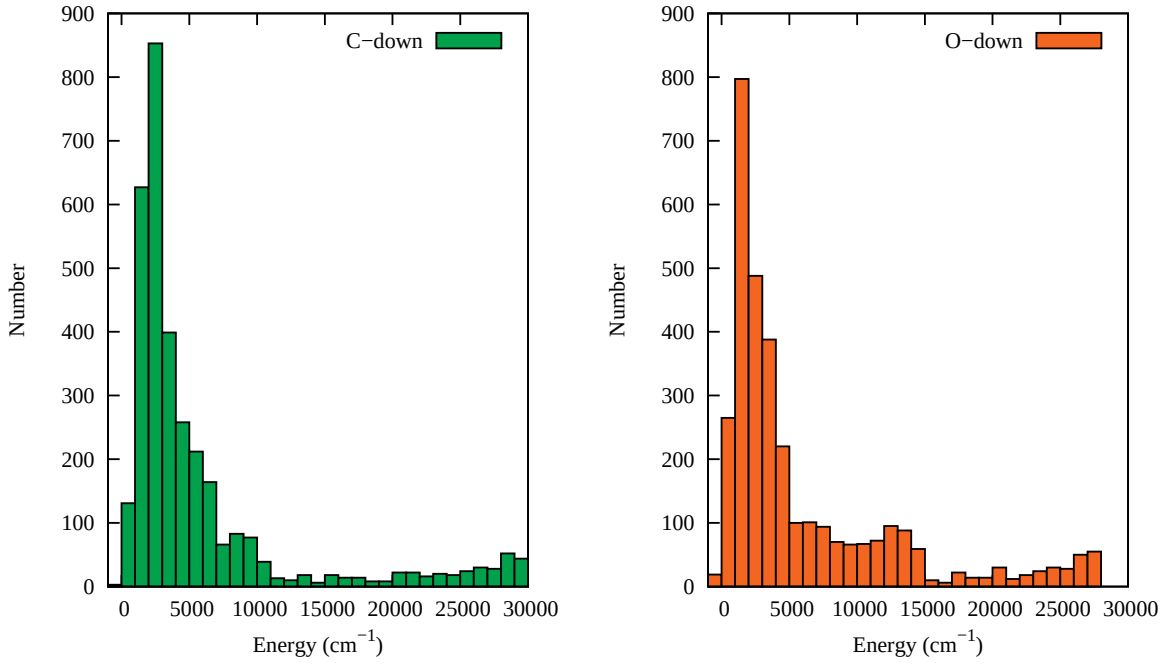


Figure 12.2: Distribution of electronic energies (cm^{-1}) of C-down and O-down isomers relative to their minimum energy.

MRCI CO Potential

It is known that the ground electronic state of CO is of multireference character at large internuclear distances, of interest to the present work. Tao et al. reported an accurate potential at the MRCI/aug-cc-pV5Z level of theory.[283] However, only a plot of the CO potential was available in that paper. Therefore, we performed MRCI/aug-cc-pVTZ calculations of this potential using MOLPRO.[87] These were done at 152 values of r_{CO} ranging from 0.88 to 2.4 angstrom. In this calculation, eight outermost orbitals and ten valence electrons of CO were selected as the active space. Spline interpolation was used to obtain the potential at arbitrary r_{CO}.

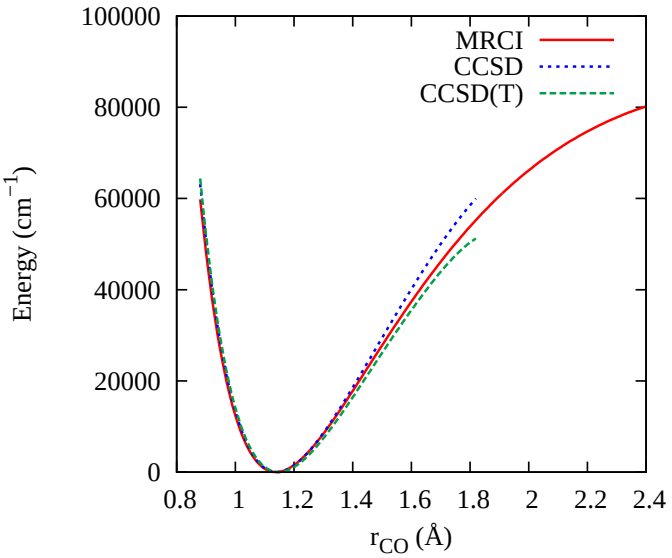


Figure 12.3: Comparison of ground state potential energy curve of isolated CO between CCSD, CCSD(T) and MRCI levels of theory.

This potential is shown in Fig. 12.3 along with ones obtained using the CCSD/aug-cc-pVDZ and CCSD(T)/aucg-cc-pVTZ levels of theory. As seen both CCSD and CCSD(T)

methods begin to fail when the CO internuclear distance is greater than 1.83 angstrom where the energy is roughly 50 000 cm^{-1}. In addition, the CC calculations showed convergence issues for r_{CO} greater than 1.85 angstrom.

We used this MRCI CO potential to calculate the numerically exact fundamental energy and obtained a value 43 cm^{-1} below the experimental value of 2143.3 cm^{-1} [273]. So, we applied a coordinate scaling trick[286] to adjust the potential so as to produce the experimental fundamental. In this simple case we simply multiplied the displacement variable, r_{CO}-r_e, where r_e is the equilibrium bond length, by 1.0215 before calling the potential routine.

It is perhaps useful to comment on using this scaled MRCI CO potential with the interaction potential based on CCSD/aug-cc-pVDZ(aVDZ) calculations described above. It appears that for energies up to around 50 000 cm^{-1} a CCSD CO potential would be adequate and for dynamics calculations up to that energy and indeed for the present work that is sufficient. However, for the calculation of the IR emission spectra, where we consider higher energies, we cannot rely on a CCSD or CCSD(T)-based CO potential. Thus, the present "plug and play" PES consisting of the MRCI CO potential plus the CCSD-based interaction potential serves for both the dynamics and emission spectra calculations.

12.3.2 (CO-NaCl) Cluster

The PES describing a monolayer of CO on NaCl consists of an average CO-NaCl PES plus a CO−CO interaction potential. First, we discuss the development of CO−CO interaction potential.

Table 12.1: Monomials, and symmetrized monomials for A_2B_2 molecule in reduced symmetry.

Atom Labels	Monomials	Normal order
1 2 3 4	$y_{12}^a y_{13}^b y_{14}^c y_{23}^d y_{24}^e y_{34}^f$	$y_{12}^a y_{13}^b y_{14}^c y_{23}^d y_{24}^e y_{34}^f$
2 1 4 3	$y_{12}^a y_{24}^b y_{23}^c y_{14}^d y_{13}^e y_{34}^f$	$y_{12}^a y_{13}^e y_{14}^d y_{23}^c y_{24}^b y_{34}^f$
Symm. term	$y_{12}^a y_{34}^f (y_{13}^b y_{14}^c y_{23}^d y_{24}^e + y_{13}^e y_{14}^d y_{23}^c y_{24}^b)$	

CO-CO Interaction PES

Recently, Guo and coworkers reported[276] an accurate full dimensional CO−CO interaction PES which is a PIP-NN fit of 16,000 geometries where electronic energies were obtained at all-electron explicitly correlated coupled cluster level with singles, doubles and perturbative triples (AE-CCSD(T)-F12b)[287] and extrapolated to the complete basis set (CBS) limit from the CVnZ-F12 (n = 2, 3) basis sets. For computational efficiency, we re-fit these data using PIPs, which provide the energy and analytical gradients.

In view of the fact that there is no atom exchange between the two COs, we use permutationl symmetry (22) for the PES. (This was also used for the PIP inputs of for the NN fit.) So, here we consider each CO as a monomer and we can permute these two CO molecules with each other. Recall that, in MSA invariant polynomials that are the Morse variables of inter nuclear distances, $y_{ij} = \exp(-r_{ij}/\lambda)$, are used to fit the data-set. However, we also require that this potential rigorously separate to two non-interaction CO molecules. To do that we "purified" the standard monomial symmetrized basis. The basic monomials and symmetrized monomials for this reduced permutational symmetry are given in Table 12.1.

Some of the symmetrized monomials (last line of Table 12.1), will have an incorrect dependence on the CO distance at large separation of two CO molecules. So, we need

to remove these polynomials at large CO−CO distance. (We term this "purifying the basis".) The polynomials which contain powers of y_{13} or y_{24} or combine power of $y_{13}y_{24}$ will not be zero at large separation of two CO. This means y_{13}^m or y_{24}^n or $y_{13}^m y_{24}^n \neq 0$, when m or n = 1,2,..., $\leq$ M, M is the maximum polynomial order. Here we use maximum polynomial order as 8 and the number of basis functions before purification is 3004. After removing the terms y_{13}^m and y_{24}^n and $y_{13}^m y_{24}^n$ when m or n = 1,2,..., $\leq$ 8, the basis size is reduced to 2959 (with some duplications). After removing duplicates, the total number total of polynomials (symmetrized monomials) is 1565. The dataset of roughly 16,000 energies[276] was fit with a final total precision of 7 cm^{-1}. Analytical gradients were coded up as well using some recursive routines. A sample of these gradients is shown up to degree 3 in Table 12.2.

Average Potential

The average CO-NaCl potential described in the text makes use of the CO-NaCl PES described above in the obvious way. Fig. 12.4 of shows the geometric arrangement of the Na and two Cl atoms. As seen, the CO-axis of the C-down and O-down minima lie along a 45 deg bisector of the X-Y plane defined by NaCl and NaCl′. The C-down minimum has a polar (tilt) angle of 34 degrees with respect to the normal to the plane of this triangle (the Z-axis in the figure). The O-down minimum has a tilt angle of 51 deg. These agree well with the experimentally reported tilt angles of about 25 and 45 deg, respectively.[273]

The goal is produce a potential where the CO is oriented between the two Cl atoms. To see that this is indeed the case for the average potential we show 1-D potentials for C-down and O-down isomers as a function of the azimuthal angle from the X-axis (depicted in Fig. 12.4) in Fig. 12.5. As seen, these are single well potentials with minima at 45 deg. Further properties of these potentials are given below.

Table 12.2: Invariant purified polynomials (p_i), and their partial derivatives dp_j for A_2B_2 up to degree 2.

Polynomials	Partial Derivative (with respect to x_i)
$p_0 = 0.0$	$dp_0 = 0.0$
$p_1 = y_{12}$	$dp_1 = \partial y_{12}/\partial x_i$
$p_2 = y_{14} + y_{23}$	$dp_2 = \partial y_{14}/\partial x_i + \partial y_{23}/\partial x_i$
$p_3 = y_{34}$	$dp_3 = \partial y_{34}/\partial x_i$
$p_4 = p_1 \times (y_{13} + y_{24})$	$dp_4 = dp_1 \times (y_{13} + y_{24}) + p_1 \times (\partial y_{13}/\partial x_i + \partial y_{24}/\partial x_i)$
$p_5 = p_1 \times (y_{14} + y_{23})$	$dp_5 = dp_1 \times (y_{14} + y_{23}) + p_1 \times (\partial y_{14}/\partial x_i + \partial y_{23}/\partial x_i)$
$p_6 = p_1 \times p_3$	$dp_6 = dp_1 \times p_3 + p_1 \times dp_3$
$p_7 = y_{13}y_{14} + y_{23}y_{24}$	$dp_7 = y_{13}y_{14} \times \partial y_{13}/\partial x_i + y_{13}y_{14} \times \partial y_{14}/\partial x_i$ $+ y_{23}y_{24} \times \partial y_{23}/\partial x_i + y_{23}y_{24} \times \partial y_{24}/\partial x_i$
$p_8 = y_{13}y_{23} + y_{14}y_{24}$	$dp_8 = y_{13}y_{23} \times \partial y_{13}/\partial x_i + y_{13}y_{23} \times \partial y_{23}/\partial x_i$ $+ y_{14}y_{24} \times \partial y_{14}/\partial x_i + y_{23}y_{24} \times \partial y_{24}/\partial x_i$
$p_9 = p_3 \times (y_{13} + y_{24})$	$dp_9 = dp_3 \times (y_{13} + y_{24}) + p_3 \times (\partial y_{13}/\partial x_i + \partial y_{24}/\partial x_i)$
$p_{10} = y_{14}y_{23}$	$dp_{10} = y_{14}y_{23} \times \partial y_{14}/\partial x_i + y_{14}y_{23} \times \partial y_{23}/\partial x_i$
$p_{11} = p_2 \times p_3$	$dp_{11} = dp_2 \times p_3 + p_2 \times dp_3$
$p_{12} = p_1 \times p_1$	$dp_{12} = dp_1 \times p_1 + p_1 \times dp_1$
$p_{13} = y_{14}^2 + y_{23}^2$	$dp_{13} = 2 \times y_{14} \times \partial y_{14}/\partial x_i + 2 \times y_{23} \times \partial y_{23}/\partial x_i$
$p_{15} = p_3 \times p_3$	$dp_{15} = dp_3 \times p_3 + p_3 \times dp_3$

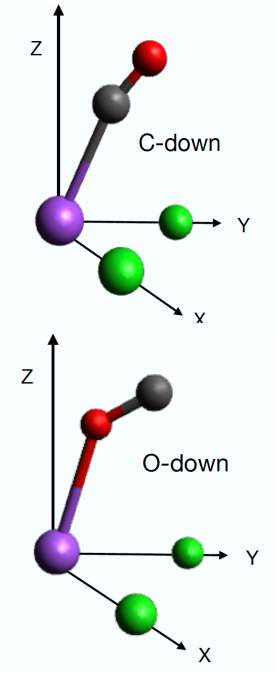


Figure 12.4: Structures of the C-down and O-down minima from the average CO-NaCl potential. C is indicated by dark gray, O by red, Na by purple and Cl by green.

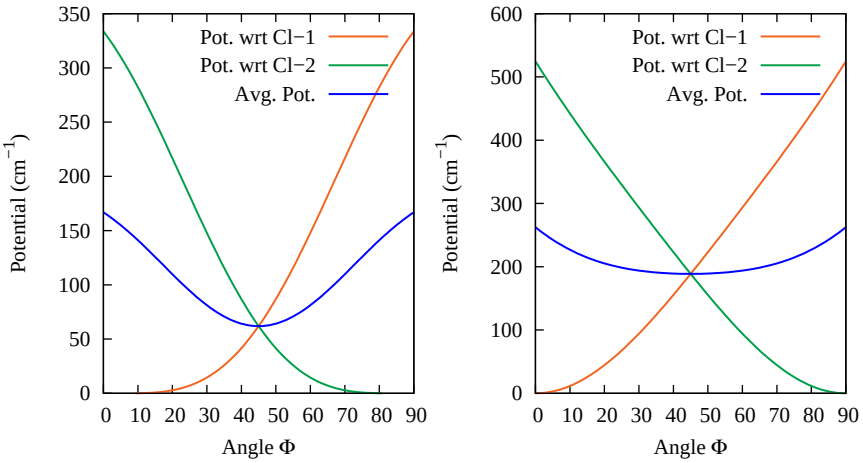


Figure 12.5: Average potential for both C-down (left one) and O-down (right one) isomers. When angle (ϕ) is zero CO or OC is co-planar with Na and Cl-1 and angle (ϕ) is 90 degree CO or OC is co-planar with Na and Cl-2.

Cluster Potential

The NaCl-CO cluster model consists of a number of NaCl units arranged according to 100 face and then adsorbed COs, with each C associated with one Na and shared with two Cl atoms. The average CO-NaClCl′ potential describes the bonded interaction of each CO to the given Na. The arrangement of the CO is such that a central CO is surrounded by COs. For the largest cluster there are 12 surrounding COs and the total potential is the sum of 13 average potentials and all 78 CO−CO potentials. Note that these 78 CO−CO interaction potentials, most of them are very weak due to their large distance between many pairs. Therefor, we considered only 32 CO−CO interaction potentials which are the strongest and moderately strong (the nearest and next nearest neighbors of any CO).

12.4 Results

In order to model the isomerization dynamics of a monolayer of CO on NaCl(100) we make note of the following properties (and assumptions) of such a monolayer. (Note, that while it is reasonable to refer to Na as Na^+ and Cl as Cl^- we do not do that below.) First, each CO is associated with a given Na on the surface and there is no migration of COs to other Na sites. However, since it is not possible to assign a single Cl to a given Na some consideration of the surrounding Cls must be given. Of the four nearest neighbor Cls only two are available for each CO in a monolayer and so each CO is shared by two Cls. This template for each CO and Na and two Cls is used to create a cluster model for the monolayer. There is a central CO molecule in this cluster that is surrounded by other CO molecules. Isomerization dynamics are modeled by vibrationally exciting the central CO. That CO interacts with other COs in a pairwise fashion and we consider the CO-CO interactions among the surrounding COs within a certain distance, which we

describe below.

The PES that corresponds to this model consists of a new *ab inito* PES for the CO-NaCl interaction with the Na-Cl distance fixed at 2.82 angstrom (the distance in NaCl(100)) plus an *ab inito* PES describing CO-CO interactions. Both PESs describe highly vibrationally excited CO. The CO-NaCl interaction PES is used in the Cl-sharing scenario by using an average of the CO-NaCl and CO-NaCl′ potentials where Cl and Cl′ denote the relevant shared Cls, as is shown in detail below.

Details of CO-NaCl PES, which is a permutationally invariant polynomial (PIP) fit to 8592 energies and gradients at CCSD/aug-cc-pVDZ level of theory and which describes the C-down and O-down minima, are given in the Computational section. The PIP PES describing the CO-CO interaction is also described in the Computational section. This PES is a refit the PIP-NN PES.[276] The PIP re-fit is done to have fast evaluation of energy and analytical gradients. A final aspect of this model is the treatment of the interaction of CO with Cl. For isolated CO-NaCl the CO and NaCl are coplanar with the O end pointed towards the Cl (as expected) as discussed in detail in the Computational section. For the O-down isomer all atoms are also coplanar with C end pointed to the Cl. This is not qualitatively correct for a monolayer of CO on the NaCl surface where the CO axis is not coplanar with Na and Cl, as discussed in the recent DFT study.[282]

The approach to address this is to use the average potential for CO-NaCl and CO-NaCl′ where the Cl and Cl′ are nearest neighbors to a given Na. As seen in the Computational section the average CO-NaClCl′ potential gives a minimum with the CO axis bisecting the right triangle made by Na and two Cl.

We performed constrained (fixed NaCl distance) geometry optimizations of the C-down and O-down minima using the single CO-NaCl PES and directly from CCCD/AVDZ, CCSD(T)(with an aug-cc-pVTZ basis for C and O atoms, and aug-cc-pwCVTZ basis

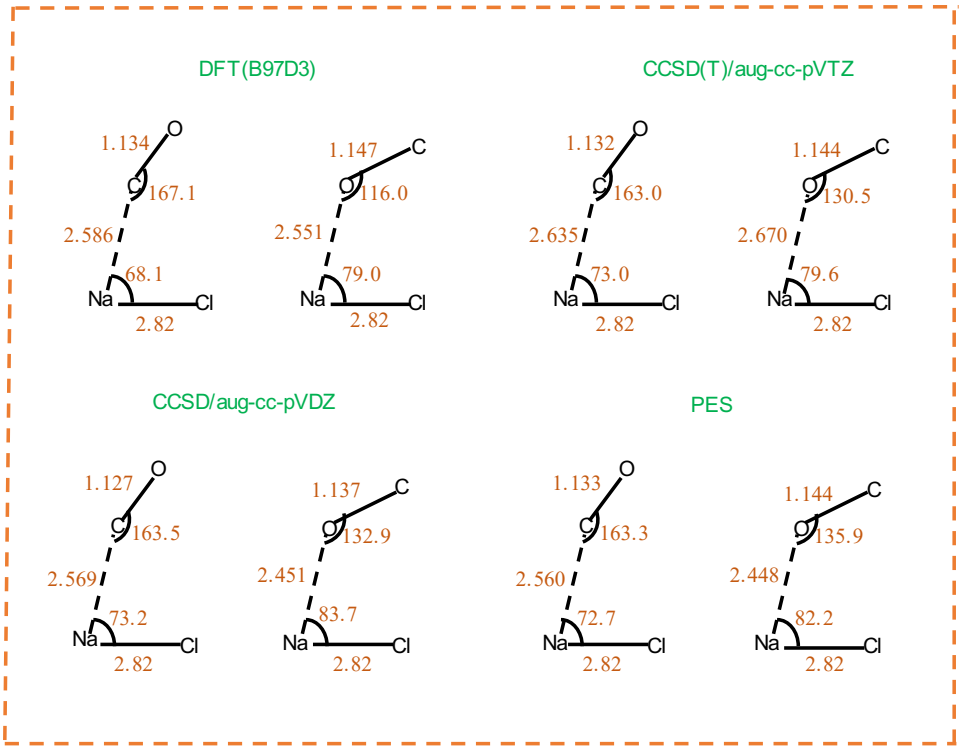


Figure 12.6: Geometrical parameters (angstroms and degrees) of C-down and O-down isomers from the bare CO-NaCl PES and at indicated level of theory. See text for details.

set for Cl and Na atoms[288]), and DFT/B97D3 calculations using aug-pc-3 basis for C, O, Cl atoms and aug-pc-4 basis for Na atom.[289,290] The energy difference between C-down and O-down minimum is 640 cm^{-1} from the PES is close to the one obtained directly from CCSD/AVDZ calculations, 645 cm^{-1} but is less than the one obtained from CCSD(T)/aug-cc-pVTZ calculations, 1045 cm^{-1}. Comparison of their geometrical parameters at different level of theories are shown graphically in Fig. 12.6. As seen these minima are well described by the PES. We also performed the geometry optimization for the central CO using both the average CO-NaClCl′ PES and also the 13-CO cluster PES for both C-down and O-down isomers. The energy differences between C-down and O-down minimum are 766 cm^{-1} and 729 cm^{-1} for the CO-NaClCl′ and for the cluster

PESs, respectively.

We also located a constrained isomerization saddle point on the CO-NaClCl′ averge PES. The barrier height is about 2568 cm^{-1}. This SP is shown in Fig. 12.7. We did not locate an exact SP from the cluster PES as this SP is very flat. The flatness of the PES is also consistent with the recent DFT study.[282]

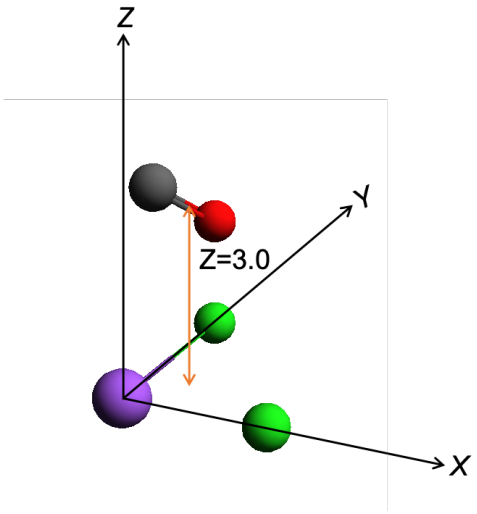


Figure 12.7: Saddle point geometry from CO-NaClCl′ average PES

We performed normal mode analyses for both C-down and O-down isomers using the single CO-NaCl PES with the constraint of a fixed NaCl distance. The results are given in Table 12.3 for which is of relevance to experiment.[273] As seen, with the exception of the CO-stretch, the remaining four vibrational modes are very low frequency. The mode of interest is the CO-stretch, which has frequencies of 2170 and 2079 cm^{-1} for the C-down and O-down isomers, respectively. These are respectively blue (+43 cm^{-1}) and red-shifted (-46 cm^{-1}) relative to the harmonic frequency of the isolated CO (from the PES). The normal mode frequencies from both CO-NaClCl′ average PES and $(CO\text{-}NaCl)_{13}$ PES are

given in Tables 12.4 and 12.5. The CO stretch frequency for both C-down and O-down mimina is very close to each other and they are also blue shifted and red shifted relative to the isolated CO. The SP of the CO-NaClCl′ PES was confirmed by the harmonic frequency calculations, where we obtained one imaginary frequency which is $56i$, which indicates a flat isomerization barrier and likely roaming dynamics. The other frequencies are given in Table 12.4.

Table 12.3: Harmonic frequencies (in cm^{-1}) between C-down, O-down isomers from single CO-NaCl PES.

Mode	C-down	O-down
1	39	31
2	121	90
3	153	108
4	218	114
5	2170	2079

Table 12.4: Harmonic frequencies (in cm^{-1}) between C-down, O-down isomers and their isomerization SP from CO-NaClCl′ average PES.

Mode	C-down	O-down	SP
1	14	35	$56i$
2	119	100	27
3	182	109	59
4	182	126	67
5	2071	2080	2125

We now turn to the cluster model consisting of a central CO with 12 surrounding COs

Table 12.5: Harmonic frequencies (in cm^{-1}) between C-down, O-down from cluster PES.

Mode	C-down	O-down
1	32	56
2	119	103
3	177	119
4	185	120
5	2171	2080

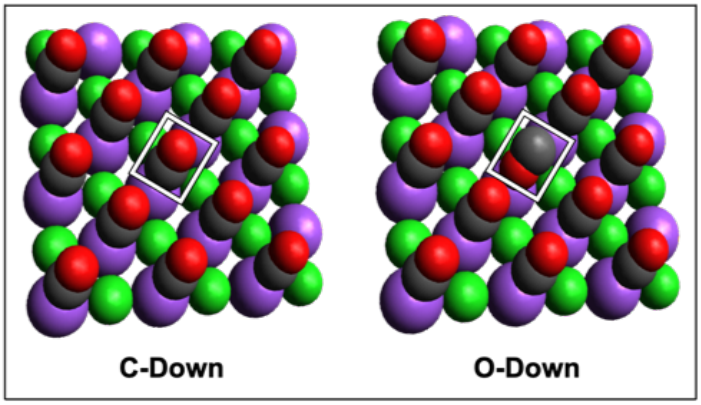


Figure 12.8: Depictions of the central CO for C-down and O-down isomers for the cluster with 13COs. C is indicated by dark gray, O by red, Na by purple, and Cl by green. The energy of the O-down isomer is 729 $cm^{?1}$ relative to the C-down isomer. In this model only the central CO flips from the C-down to the O-down minimum. Surrounding COs remain in the C-down configuration.

and 18 NaCls. This cluster is depicted in Fig. 12.8 for the C-down and O-down minima. At the former all COs are C-down and oriented in the same way; for the O-down minimum we indicate this isomer only for the central CO. The total potential is expressed as sum of all 13 CO-NaClCl′ average potentials and CO-CO 2-body interactions. Note, it is not necessary to consider all (78) CO-CO interactions, because most are weak owing to the large distances between many pairs. At equilibrium, the nearest-neighbor CO-CO distance between centers of mass is 3.99 angstrom. Next nearest neighbors are 5.64 angstrom apart. Restricting to these, there are 32 CO-CO interactions and these range

from -123 to -15 cm^{-1}.

Because the CO-CO interactions are weak compared to the interaction between the COs and the associated NaClCl′ units the CO equilibrium configurations are expected to be close to those for the averaged CO-NaClCl′ potential. Thus, full geometry optimzation of this 62-atom cluster was not done. Geometry optimization and normal mode analyses were performed for the central CO using this cluster PES for both C-down and O-down isomers. The energy of the O-down isomer is 729 cm^{-1} relative to the C-down isomer. This is close to the 766 cm^{-1} energy difference for the average CO-NaClCl′ potential. (This energy difference is similar to the energy difference of around 750 cm^{-1} based on an electrostatic model without short-range interactions and at a fixed distance of the CO above the surface.[274])

Normal mode frequencies of the central CO at the stationary points are already given above in Table 12.5. The CO-stretch has frequencies of 2171 and 2080 cm^{-1} for the C-down and O-down isomers, respectively. These are respectively blue (+44 cm^{-1}) and red-shifted (-47 cm^{-1}) relative to the harmonic frequency of the isolated CO (from the PES). These are in qualitative agreement with experiment; however, the values of the shifts are larger than seen in the experiment. The normal mode frequencies of the frustrated CO rotations and translations are in the range 10 to 220 cm^{-1} for the minima.

In anticipation of the dynamics calculations, which focus on vibrationally excited CO, we show several plots that indicate the change in energetics for highly stretched CO. The first are one-dimensional plots of unrelaxed potentials as a function of γ for two values of r_{CO}. The angle γ is defined precisely in the Fig. 12.9 and all other coordinates are fixed at the C-down minimum. These potential cuts are shown in Fig. 12.10. In these plots the CO is simply rotated about an axis perpendicular to the plane of the figure showing the C-down equilibrium configuration in Fig. 12.4. For CO at equilibrium the (unrelaxed)

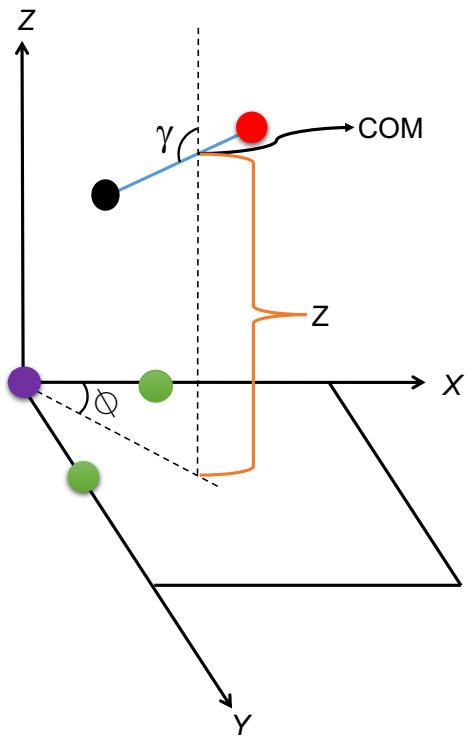


Figure 12.9: Z is the normal distance from the NaCl surface to center of mass of the CO and γ is the spherical polar angle of the CO such that when γ is 146 and 326 degrees the cluster becomes C-down and O-down, respectively.

isomeriation barrier is roughly 2700 cm^{-1} but in the stretched case it is roughly half that value, relative to the C-down minimum. Note the large change in the O-down minimum relative to the C-down one. We also tried to estimate the dissociation energy from both average and cluster PESs. These dissociation energies are 2137 and 2727 cm^{-1} for average and cluster PESs when the CO (central one) is at ground vibrational state and these are 570 and 1014 cm^{-1} when the CO (central one) is $v = 20$ vibrational state. We already mentioned earlier about the large change in the PES upon CO vibrational excitation. O-down minimum becomes much lower in energy compare to C-down minimum when CO is highly vibrationally excited. Therefor, we also computed the dissociation energy

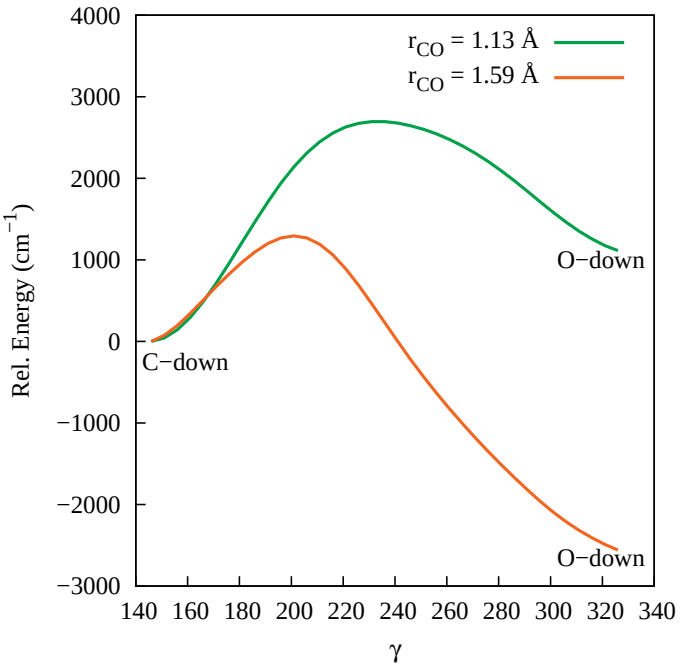


Figure 12.10: One-dimensional unrelaxed potential cuts for isomerization. These are as a function of γ(deg) relative to the C-down isomer for the indicated values of r_{CO} corresponding to v = 0 (1.13 angstrom) and the outer turning point for v = 20 (1.59 angstrom) vibrational states. These plots show a dramatic effect of stretching the CO on the energetics of the isomerization.

with respect to O-down minimum when CO is at $v = 20$ vibrational state and these are 2451 and 3049 cm^{-1} for average and cluster PES, respectively. We have seen a significant increment of this dissociation energy after adding CO-CO interactions.

Next are contour plots of the interaction potential of the central CO, shown in Fig. 12.11. Fig. 12.9 shows the coordinates of the contour plots. In these the Z-component of the position vector of the center of mass of the central CO is the additional variable. In the upper plot with r$_{CO}$ at r$_e$ the deeper minimum at 146 deg is C-down and the unrelaxed barrier is at around 230 deg and at Z of about 5 angstrom. For r$_{CO}$ equals 1.59 angstrom the unrelaxed O-down minimum is much deeper than the C-down minimum

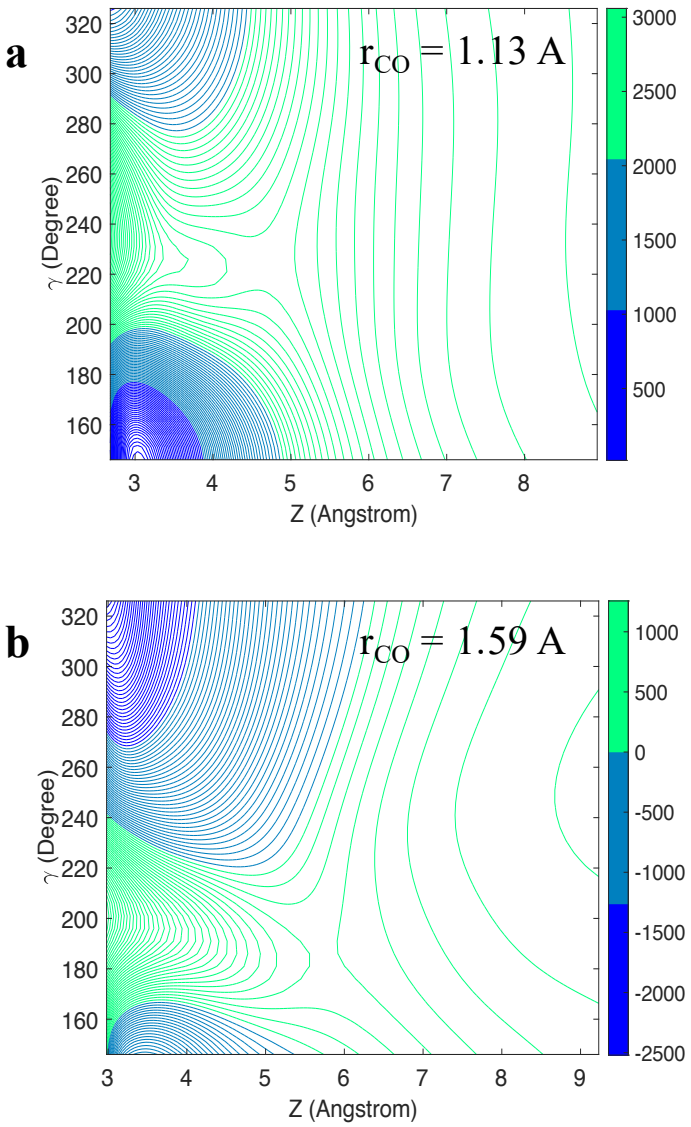


Figure 12.11: Contour plots for the large cluster PES in the variables Z and γ, defined in the text, for the indicated values of r_{CO}. Contour values are in 31 cm^{-1} increments for the top one and 38 cm^{-1} increments for the below one.

and also spans a much large space than the C-down minimum and the unrelaxed barrier moves to nearly 6 angstrom. Another difference is the extent of anisotropy in γ for the stretched-CO plot. As seen this region extends to Z-values greater than 8.0 angstrom.

These changes are relevant the dynamics presented next (and also to experiment where isomerization is seen when the CO (C-down) is highly vibrationally excited[273]).

From these plots we expect different dynamics for CO in the ground and this highly excited vibrational state. For the latter where the interaction extends to large distances above the surface, roaming[291–293] is likely. This is confirmed from analysis of the dynamics calculations given next.

12.4.1 Normal mode CO-stretch Potentials, and Vibrational Energies for C and O-down isomers

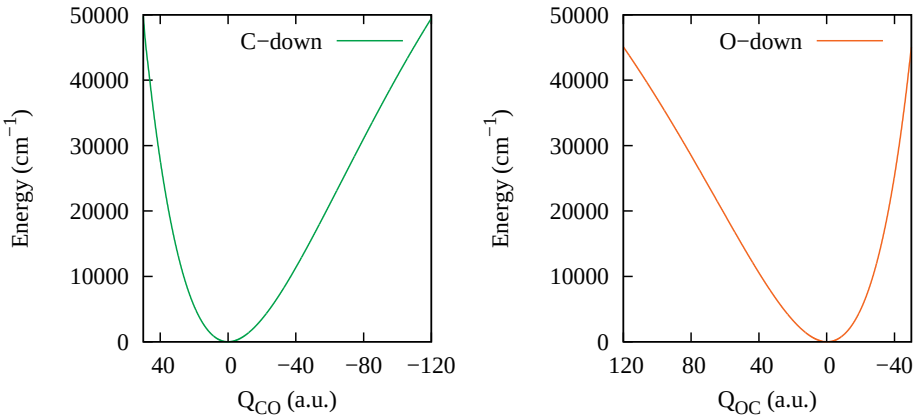


Figure 12.12: Potential energy (cm^{-1}) a functions of normal mode of the CO-stretch for the C-down and O-down isomers from cluster PES.

This 13-COs cluster PES was used in calculations of exact wavefunctions and energies of the CO-stretch for stretching mode on NaCl for both C-down and O-down isomers. The one-dimensional CO/OC potential cuts as a function of the corresponding normal mode at the PES minima are generated to solve the one-dimensional Schrödinger equation us-

ing the Colbert-Miller discrete variable representation of the Hamiltonian.[294] These two potential cuts are shown in Fig. 12.12. Discrete variable representation (DVR) technique was applied to compute the eigenvalues of these two adiabatic potentials. A list of these eigenvalues are given in Table 12.6. It has been clearly seen that vibrational energy difference between C-down and O-down is increasing with increasing the vibrational quantum number, v. When v=8 this vibrational energy difference exceeds the energy difference between C-down and O-down isomers (ΔE = 735 cm^{-1}). Another interesting aspect we have notices that when we stretch the CO bond the O-down isomer become lower in energy than the C-down isomer. C-down isomer becomes lower in energy when r_{CO} is less than 1.2 angstrom and after that O-down becomes lower in energy. These details can be found in Fig. 12.13. We also note that CO stretching of C-down isomer is blue shifted and O-down isomer is red shifted from isolated CO molecule which qualitatively agrees with the experimental result.

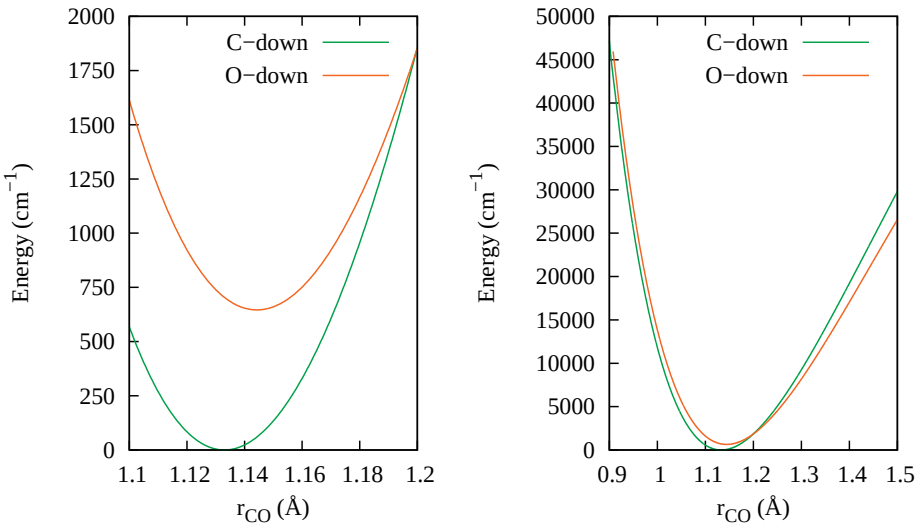


Figure 12.13: Potential energy (cm^{-1}) of the CO-stretch for both C-down and O-down isomers shown together with the zero at the C-down minima from cluster PES.

Table 12.6: Eigenvalues (cm^{-1}) of the central "CO" stretch for the C-down and O-down isomers from cluster PES.

Quantum State	C-down (cm^{-1})	O-down (cm^{-1})	Diff. (cm^{-1})
$v = 0$	1085.79	1044.03	41.76
$v = 1$	3238.00	3113.39	124.61
$v = 2$	5365.05	5156.55	208.50
$v = 3$	7467.08	7173.59	293.49
$v = 4$	9544.11	9164.49	379.62
$v = 5$	11596.19	11129.32	466.87
$v = 6$	13623.39	13068.13	555.26
$v = 7$	15625.76	14980.98	644.78
$v = 8$	17603.39	16867.93	735.46
$v = 9$	19556.31	18729.06	827.25
$v = 10$	21484.61	20564.47	920.14
$v = 11$	23388.38	22374.31	1014.07
$v = 12$	25267.73	24158.75	1108.98
$v = 13$	27122.83	25918.09	1204.74
$v = 14$	28953.94	27652.76	1301.18
$v = 15$	30761.45	29363.36	1398.09
$v = 16$	32545.91	31050.66	1495.25
$v = 17$	34308.04	32715.57	1592.47
$v = 18$	36048.70	34359.08	1689.62
$v = 19$	37768.81	35982.13	1786.68
$v = 20$	39469.24	37585.50	1883.74
$v = 21$	41150.68	39169.73	1980.95
$v = 22$	42813.60	40735.05	2078.55
$v = 23$	44458.11	42281.40	2176.71
$v = 24$	46084.02	43808.58	2275.44
$v = 25$	47690.89	45316.39	2374.50
$v = 26$	49278.20	46805.65	2472.55
$v = 27$	50845.84	48281.63	2564.21

12.4.2 Quasiclassical Trajectory Calculations

To study the C-down to O-down isomerization process, quasiclassical trajectory calculations were performed using CO-NaClCl′ and cluster PESs. In all cases the initial condi-

tions for the central C-down CO vibrational excited state were obtained by stretching the CO to an outer turning point and keeping all the 12 COs at their equilibrium value. The normal coordinate Q_{CO} was extended to a value such that the potential (shown in Fig. 12.12) equals the energy of a given vibrational state from Table 12.6. The corresponding Cartesian coordinates of the C and O were obtained from the usual transformation from normal coordinates to Cartesians and the corresponding velocity vector was set to zero. Thus, the energy in this mode is the correct quantum vibrational energy. An additional amount of kinetic energy of 300 cm^{-1} is added randomly and microcanonically to other degrees of freedom of the central CO to account for the thermal energy corresponding to the experimental temperature of 7 K.[273] NaCl bond distance was fixed at 2.82 angstrom during the propagation of each trajectory. 500 trajectories were run for each vibrational state v and each trajectory was run for a maximum of 200,000 time steps (roughly 20 ps). As a trajectory propagates we monitor the Na-C and Na-O distances. If they become nearly equal we examine the trajectory in detail to determine if and for how long the Na-O distance is shorter than the Na-C one. This is our criterion for isomerization.

Such trajectory calculations were also done for the bare CO-NaCl and the smaller cluster with 4 surrounding COs. In all cases, we observed isomerization based on the above criterion, desorption, or failure of the trajectory (after 200 000 time steps) due to unphysical edge effects of the finite cluster where some of the outermost COs sample configurations that are far from those described by the average CO-NaClCl′ PES.

For the bare NaCl-CO, we observed isomerization starting at $v = 12$. For $v = 15$ and above desorption is the major outcome. While this is not unexpected (since the total energy of the CO is more than ten times the dissociation energy), it is at odds with experiment, which has little or no desorption. For the small, 5 CO-cluster isomerization starts at $v = 16$ and desorption does not become a major process until $v = 20$. For the

large 13-CO cluster we performed QCT calculations for v = 16,18,20,22 and 24. In this case isomerization starts when $v = 22$ which is in much closer accord with experiment. So adding the CO-CO interactions suppresses desorption and also moves isomerization to higher vibrational states. We discuss this further below.

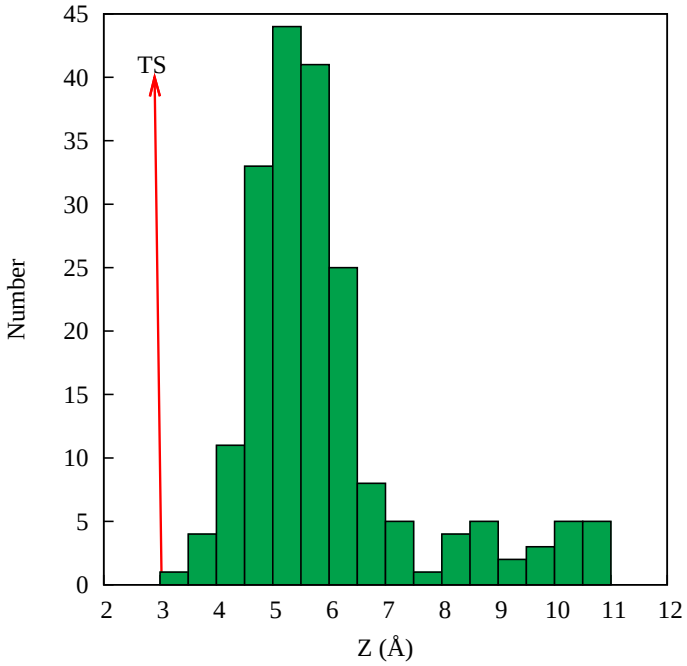


Figure 12.14: Histogram of the distribution of Z, the distance above the surface, for isomerizing trajectories for $v = 22$ and 24. The red arrow indicates the value of Z at the saddle point transition state for the average CO-NaCl potential.

For the large cluster for $v = 22$, 63 trajectories isomerized, only 9 trajectories directly desorbed, 32 trajectories completely propagated for 200 000 time steps without any isomerization or desorption. The remaining 396 eventually failed due to the edge effect mentioned above. For $v = 24$ we observed 126 isomerizing trajectories, 185 desorbed and only 2 trajectories completely propagated for 200 000 times steps. The remainder

failed due to the edge effect. In total, there were 189 isomerizing trajectories for these two vibrational states. For these, we calculated the value of Z where the Na-C and Na-O distances are first about equal. A histogram of these Z-values is shown in Fig. 12.14. As seen, the major peak is at around 5-6 angstrom; this is much larger than the conventional SP distance of 3.0 angstrom.

Plots of the time dependence of Z and the interaction potential of an isomerizing trajectory for $v = 22$ are shown in Fig. 12.15. Initially the CO is at the C-down minimum with a stretched CO bond length and Z is at the equilibrium value. Z increases with time and reaches a plateau value of around 4.7 angstrom before declining again to less than 2.5 angstrom. The corresponding interaction potential oscillates strongly except in the plateau region. This behavior can be understood as follows. At early times the C-down CO stretch is interacting with NaCl surface largely based on ion-dipole forces. At times after isomerization the O-down CO-stretch is similarly interacting (albeit more strongly) with the surface. In both cases the dipole moment of the highly excited CO changes sign as the molecule compresses and stretches (see Fig. 12.19 in the next section for a plot of the dipole moment vs r_{CO}.) Thus, if the total interaction is mostly due to dipole-ion interactions, the oscillation in the dipole moment explains the oscillations of the interaction potential. By contrast, at large Z the potential is nearly flat. This is because of longer range and hence a smaller interaction and also because the CO dipole is nearly parallel to the surface as so the dipole interaction is much reduced.

This plateau region both in Z and the potential at long range where isomerization occurs satisfies the general conditions of roaming.[291,295] A plot of the C and O distances to the Na atom is shown in Fig. 12.16 for this trajectory. This plot shows the early history of the trajectory where C is closer to Na than O, the plateau region where the two distances are nearly equal (and also the point where they are equal) and the later part of the

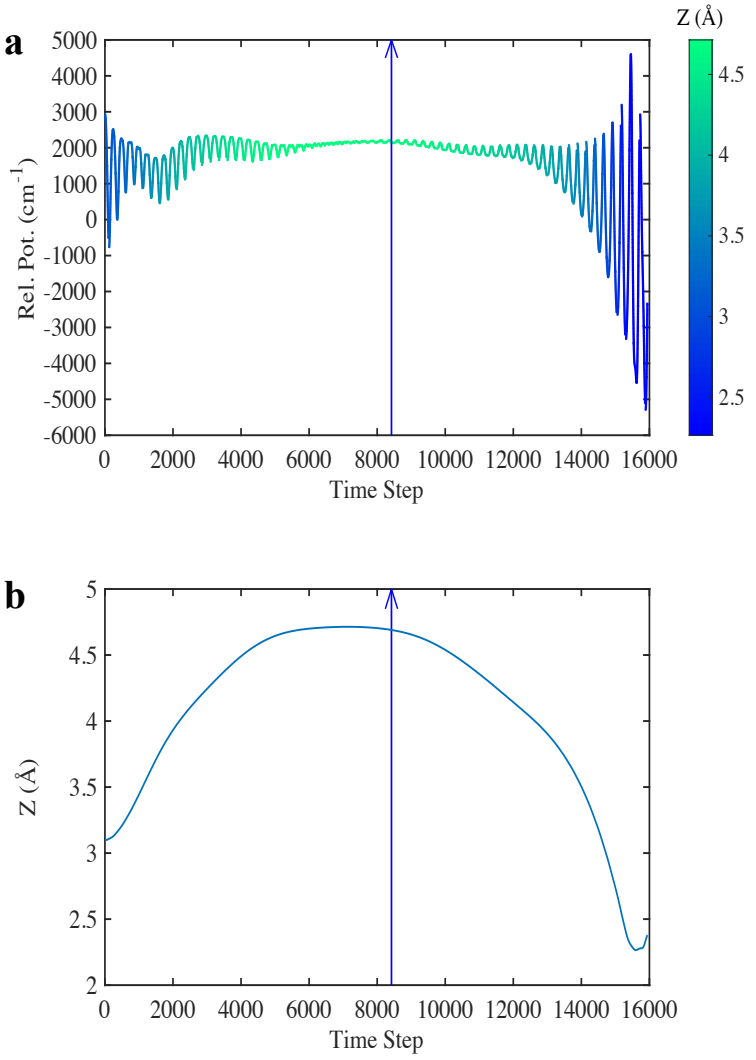


Figure 12.15: Signature of roaming dynamics. Pannel **'a'** and **'b'** shown the time dependence of the interaction potential (a) and Z, the distance above the surface (b) for an isomerizing trajectory for v = 22 for the large cluster. The vertical arrow indicates the time at which the isomerization occurs. The plateau region is where roaming and isomerization occur. Note the decrease in the oscillation of the interaction potential. This is due to both the relatively large distance from the surface and also due to the orientation of the CO dipole moment which has a reduced interaction with the surface ions.

trajectory where the O atom is closer to the Na. Snapshots of a representative isomerizing

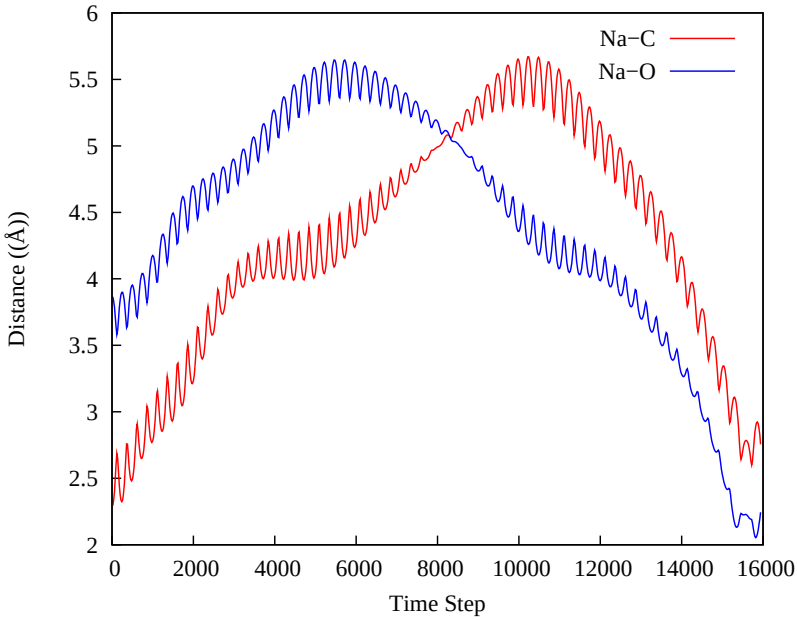


Figure 12.16: A plot of C and O distances to the Na atom during the propagation of an isomerizing trajectory.

trajectory for $v = 22$ are shown in Fig. 12.17. As seen the CO initially separates from the NaCl and isomerization begins at where thus, isomerization predominatly occurs at values of R much larger than the SP value.

Another important result of the present calculations is that CO is vibrationally excited after isomerization. This is indicated in Fig. 12.18, which shows r_{CO} as a function of time prior to and then after isomerization, for CO at $v = 22$. As seen isomerization occurs rapidly on the time scale of this figure; however, the high vibrational motion indicated in this figure is maintained during the brief period of isomerization. As seen there is virtually no change in the vibrational state. Thus the O-down isomer is highly vibrationally excited and so emission from this state for the O-down isomer can occur.

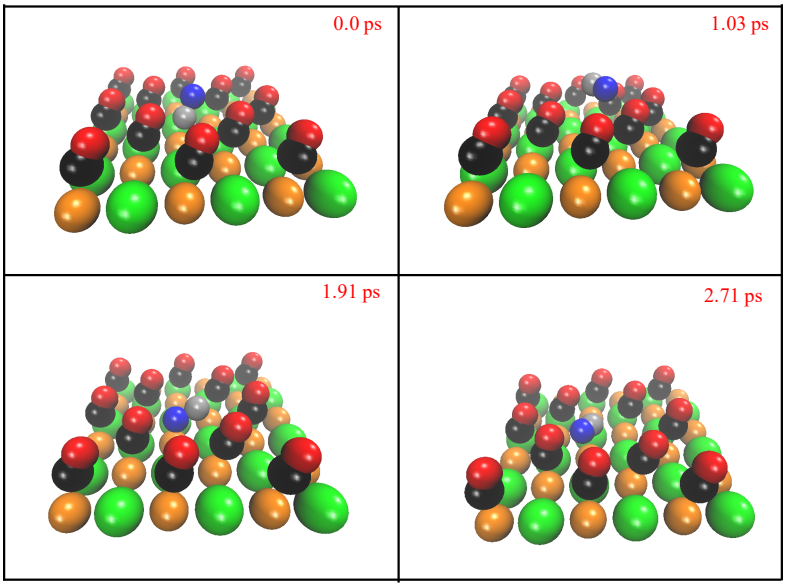


Figure 12.17: Snapshots of a representative trajectory starting from the C-down configuration and exhibiting isomerization.

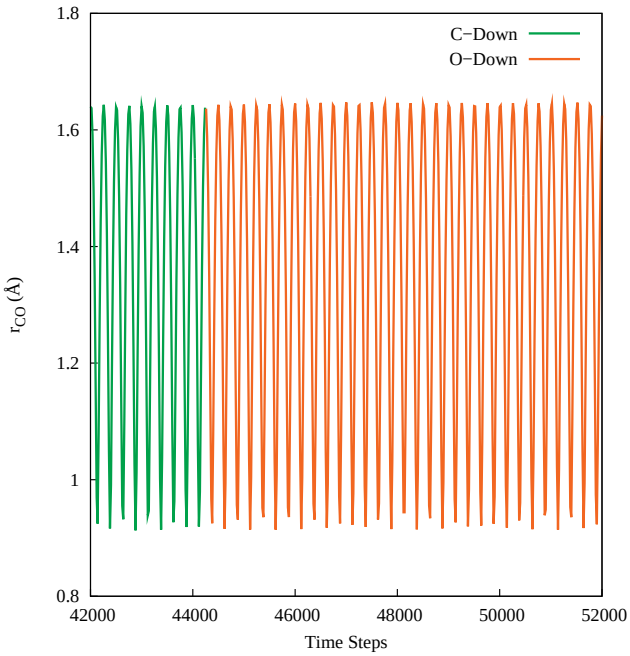


Figure 12.18: CO-stretch internuclear distance versus time for $v = 22$ just prior to isomerization and after isomerization

12.4.3 IR emission Spectra for C and O-down isomers

The 1D vibrational wavefunctions for the C-down and O-down CO-stretch were used in standard calculations of the IR emission spectrum by doing numerical integrals with the isolated CO-dipole moment.[296] The dipole moment of isolated CO is shown in Fig. 12.19. It is clearly seen the dipole moment vector changes sign when the CO is highly stretched and compressed. Using the isolated CO dipole is of course not exact; however, in the absence of a full-dimensional dipole moment surface of CO-NaCl we had to resort to this approximation.

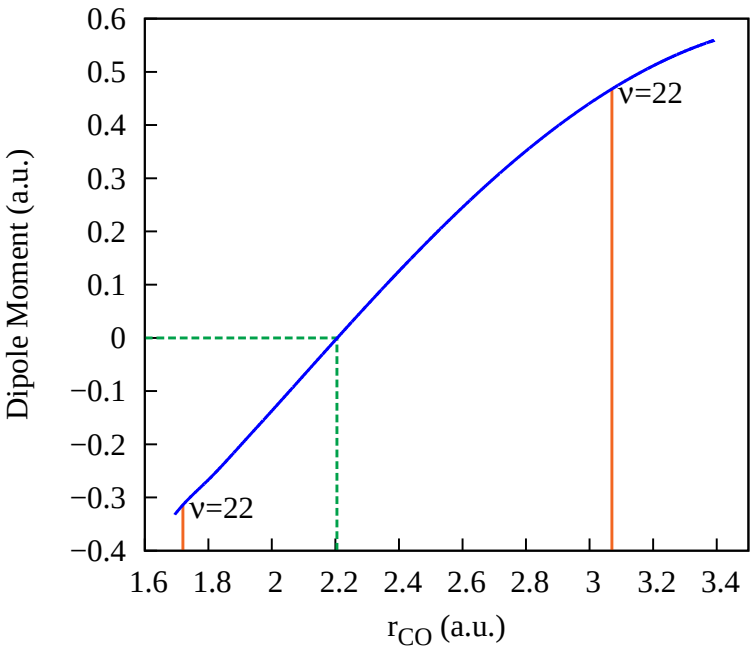


Figure 12.19: Dipole moment of isolated CO with the indication of both inner and outer turning point (orange line) at $v = 22$ state.

To make further contact with experiment, we calculated the emission spectra for Δ

v = -2 transitions of the C-down and O-down isomers. This was done using numerically exact wavefunctions and energies by solving the one-dimensional Schrödinger equation using the CO/OC cut potentials as a function of the corresponding normal mode at the PES minima. These two cut potentials are shown in Fig. 12.12 along with details of the wavefunction calculations and numerical matrix elements using an accurate *ab initio* dipole moment.[296] The simulated spectra are shown in Fig. 12.20 as sticks. As seen, the two spectra are interleaved with both showing increasing intensity and decreasing overtone gaps as v increases, as expected owing to the increasing non-linearity of the dipole moment at large values of r_{CO}, in agreement with experiment.

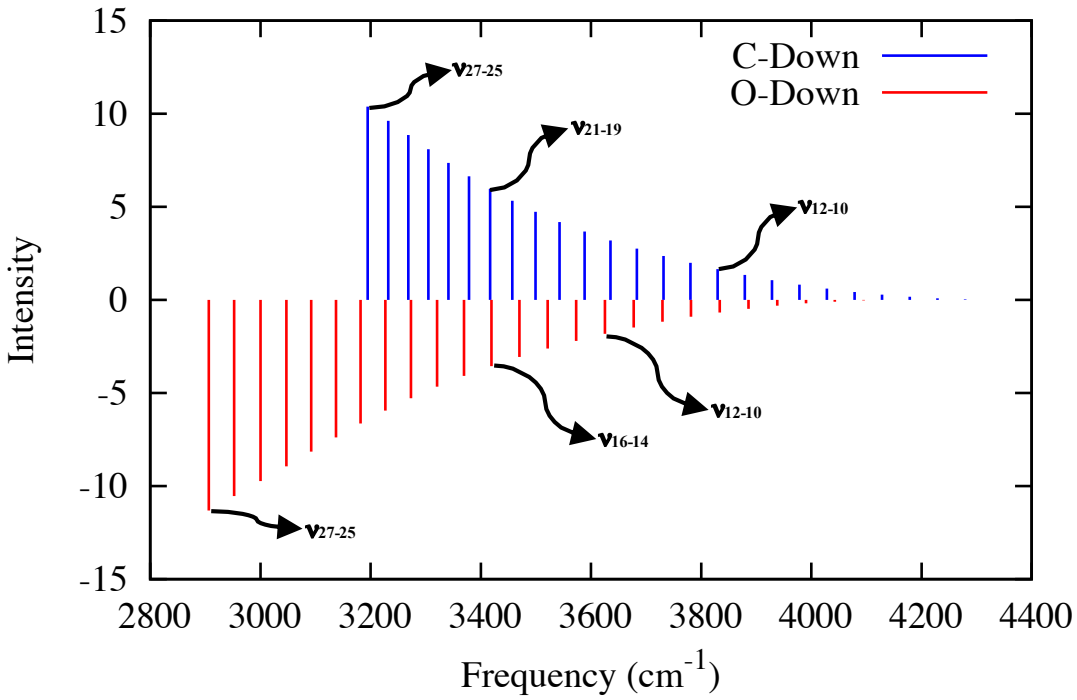


Figure 12.20: Calculated Δv=-2 emission spectra of both C-down and O-down isomers with the indication of their vibrational transition.

12.5 Summary and Conclusions

The present trajectory calculations find that isomerization occurs only for highly excited vibrational states of the C-down isomer. The threshold for isomerization increases as surrounding COs are added to the bare CO-NaCl moiety, with the threshold at $v = 22$ for the 13-CO cluster mode. This threshold is consistent with the experiment, which shows emission from the C and O-down isomers for vibrational states up to v of 27. Second, the present calculations show that CO continues to be highly vibrationally excited after isomerization. This is also clearly in agreement with experiment. Also, from the calculations, it appears that emission from much lower-lying O-down vibrational states seen in experiment occurs not from prompt isomerization but from relaxation of higher vibrationally excited O-down CO.

The driving force for this vibrationally induced isomerization is the major change in the potential energy surface from energetically favoring the C-down isomer to greatly favoring the O-down isomer upon high vibrational excitation of the CO-stretch. Also, the region of interaction between the CO and NaCl extends to longer range upon high vibrational excitation. The isomerization dynamics thus occurs over large range of distances of the CO from the NaCl surface, very reminiscent of roaming.[291–293] For the specific case of isomerization, recall that roaming was reported in unimolecular dissociation of energized CH_3NO_2.[295,297–299] In this case internally excited CH_3NO_2 can undergo a frustrated dissociation to fragments, CH_3 and NO_2, where the incipient fragments roam at large distance in a flat part of the potential and go on to form CH_3O+NO products.

Finally, we make some comments of the effects of CO-CO interactions on the central active CO. These interactions provide a channel for energy transfer from the vibrationally excited CO. This qualitatively at least accounts for the increase in the threshold for

isomerization in going from NaCl-CO to the small 5-CO cluster to the larger 13-CO cluster. Second, the large amount of desorption observed for bare NaCl-CO is reduced for the small CO-cluster and further reduced for the 13-CO cluster. Some of this effect can again be accounted for by the CO-CO energy transfer. Also, because the CO-CO interactions are attractive, the effect of them is to increase the dissociation energy (which we estimate to be roughly 600 cm^{-1}), which also results in less desorption.

The calculated IR overtone emission spectra are in accord with the experimental ones in showing significant intensity for high v-states.[273] The present calculations find a threshold for isomerization at $v = 22$, whereas experiment cannot determine such a threshold from the emission spectra. There are indeed strong experimental emission lines for the O-down isomer in the region of $v = 20$-28, but also a long progression at lower v all the way down to around $v = 4$. The present calculations interpret these lines below around 20 as due to vibrational relaxation of the highly excited O-down isomer and not prompt emission following isomerization of the C-down isomer in these low v states.

In summary, the present calculations verifies the vibrationally induced isomerization reported experimentally dominated by a roaming pathway, where the isomerization occurs at large separations, roughly 5 angstroms larger than the separation at the conventional isomerization saddle point. There are aspects of the experiment that will need to be addressed in future work. One is the details of the energy pooling that is needed to prepare CO in highly excited vibrational states. Another is the relaxation of highly excited O-down isomers and the subsequent reformation of the C-down isomer.

Chapter 13

Nuclear quantum dynamics reveal the leaky nature of gas-phase *trans* and *gauche* Ethanol conformers

13.1 Chapter Abstract

Ethanol is a molecule of fundamental interest in combustion, astrochemistry, and condensed phase as a solvent. It is characterized by two methyl rotors and *trans (anti)* and *gauche* conformers, which are known to be very close in energy. Here we show that based on rigorous quantum calculations of the vibrational zero-point state, using a new ab initio potential energy surface (PES), the ground state resembles the *trans* conformer but substantial delocalization to the *gauche* conformer is present. This explains experimental issues about identification and isolation of the two conformers. This "leak" effect is partially quenched when deuterating the OH group, which further demonstrates the need for a quantum mechanical approach. Diffusion Monte Carlo (DMC) and full-dimensional

semiclassical dynamics calculations are employed. The new PES is obtained by means of a Δ-Machine learning approach starting from a pre-existing low level (LL) density functional theory (DFT) surface. This surface is brought to the CCSD(T) level of theory using a relatively small number of *ab initio* CCSD(T) energies. Agreement between the corrected PES and direct *ab initio* results for standard tests is excellent. One- and two-dimensional discrete variable representation calculations focusing on the *trans-gauche* torsional motion are also reported, in reasonable agreement with experiment.

13.2 Overview

Ethanol is one of the most important organic molecules with many applications in industrial products, chemicals, and solvents. It is also the leading biofuel in the transportation sector, where it is mainly used in a form of reformulated gasoline[300,301] and studied from scientific, industrial, and environmental perspectives for its role in internal combustion engines.

Ethanol exists as a mixture of *trans* (or *anti*) and *gauche* (+/−) conformers in both solid, liquid, and gaseous state.[302–304] Therefore, the energy difference between the *trans* and *gauche* conformers is expected to be very small. This is corroborated by the data extracted upon fitting models to spectroscopic experiments in the microwave and far-infrared portion of the electromagnetic spectrum, which estimate the energy gap at 0.12 kcal/mol or 41 cm^{-1} in favor of the more stable *trans* conformer.[303,305] Therefore, there is an anticipated preponderance of the *gauche* form at room temperature (62%) because of its two-fold degeneracy (+/−). Furthermore, ethanol has two isomerization saddle points and a three-fold methyl torsional potential, which makes its potential surface much more complex.

Reports on ethanol in the literature have been often accompanied by several experimental studies of its isomers. In 1980, Quade and co-workers reported microwave torsional-rotational spectra of *gauche* ethanol[303] and later Durig and Larsen presented a detailed examination of the torsional modes.[305] Rotational isomerization of ethanol in nitrogen and argon matrices has been recorded under various conditions of temperature and irradiation in the OH and CO stretches by Coussan *et al.*[306] In 2013, comparative analysis of low-temperature FTIR absorption spectra were reported for ethanol isolated in an argon matrix by Balevicius and co-workers.[307] It was observed that in an argon matrix ethanol is predominantly in the *trans* configuration, although the most intense absorption lines of the *gauche* conformer were still observed in the spectra of the samples. Recently, the *trans-gauche* conformational distribution of ethanol has been investigated using the O-H and symmetric C-C-O stretching infrared spectra in argon and nitrogen matrix.[308] It was found that the *trans* conformer is more populated in nitrogen mixture whereas the *gauche* conformer is more populated in the argon mixture. After thermal cyclisation in the matrix, the *trans* conformer isomerises to the *gauche* conformer in a nitrogen matrix but the reverse happens in an argon matrix. Finally, Pearson *et al.* (PBD) also reported a comprehensive analysis of the threefold asymmetric rotational?torsional spectrum of ethanol in the torsional ground state of the OH internal rotation.[309] Zheng et al. considered the partition functions of rotors in ethanol and performed helpful calculations on the energy levels.[310]

Ethanol has also been investigated extensively using electronic structure calculations to understand its energetics and complex potential energy surface (PES). In 2004, calculations have been performed at the MP2/aug-cc-pVTZ and CCSD(T)/aug-cc-pVTZ levels of theory by Dyczmons.[311] It is reported that the *trans* isomer is 0.52 kJ mol^{-1} or 44 cm^{-1} more stable than the *gauche* isomer and the energy barrier for the torsional

motion of the OH group for *trans* to *gauche* isomerisation is 3.9 kJ mol^{-1} or 326 cm^{-1}. Recently, a high level calculation has been performed at the CCSD(T)/aug-cc-pVQZ level of theory by Kirschner and co-workers.[312] It was found that the *trans* isomer is more stable by 0.53 kJ mol^{-1} or 44 cm^{-1} compared to *gauche* isomer. Thus, it is concluded that the *trans* conformer is more stable in the gas phase compared to the *gauche* conformer. Remarkably, a thorough investigation on conformational analysis by systematically improving the basis set and the level of electron correlation of ethanol has been reported by Kahn and Bruice in 2005.[313] Their best estimate of the *trans-gauche* energy gap is 0.134 kcal mol^{-1} or 47 cm^{-1} and the energies of the two isomerization TSs (*eclipsed* and *syn*) are 1.08 kcal mol^{-1} or 378 cm^{-1} and 1.20 kcal mol^{-1} or 420 cm^{-1}, respectively, relative to the *trans* minimum. They came to the common conclusion that the *trans* conformer is more stable in the gas phase compared to the *gauche* conformer. Very recently, Grimme and co-workers reported combined implicit and explicit solvation protocols for the quantum simulation of ethanol conformers in the gas phase, liquid phase and in CCl_4 solutions. The implicit treatment of solvation effects suggested that the ratio of the *trans* and *gauche* conformers of ethanol increases only slightly when going from gas phase to a CCl_4 solution, and to neat liquid.[314]

However, we note that both experiments and theoretical calculations may not have been conclusive in describing the *trans-gauche* dichotomy of gas-phase ethanol. On the one hand, the experiments referenced above appear to deal with a mixture of the two conformers and to be even affected by experimental conditions. For instance, in Ref. 306 it is shown that infrared experiments performed in the 8-30 K temperature range point at temperature-dependent *trans-gauche* isomerism when a nitrogen matrix is employed, while the temperature dependence vanishes and evidence of *trans* isomer only is found when an argon matrix is used. Other experiments found a mixture of the two conformers

also in argon matrix, but with abundance conclusions at odds and an inter-conversion rate dependent on temperature and matrix type. On the other hand, accurate but static theoretical calculations have been performed only at the level of electronic structure, while quantum nuclear effects have not been taken into consideration or have been estimated just with basic and inaccurate harmonic approaches.

The main goal of this study is to investigate the energetics of ethanol and its challenging conformational properties including quantum nuclear effects. This is obtained by means of rigorous diffusion Monte Carlo (DMC) and semiclassical calculations able to describe nuclear quantum effects performed on a new "gold standard" *ab initio* CCSD(T) PES, which we have constructed for this investigation.

Developing high-dimensional, *ab initio*-based PESs remains an active area of theoretical and computational research. Significant progress has been made in the development of machine learning (ML) approaches to generate PESs for systems with more than five atoms, based on fitting thousands of CCSD(T) energies.[64,65,90,315] Examples of potentials for 6 and 7-atom chemical reactions which are fits to tens of thousands or even hundred thousand CCSD(T) energies have been reported.[316,317] However, there is a bottleneck for developing the PES at high level theory with the increase of molecular size. Due to the steep scaling of the "gold standard" CCSD(T) theory ($\sim N^7$, N being the number of basis functions), it is computationally demanding to fit PESs for systems with a larger and larger number of atoms.

The increasing dimensionality of the PES with the increase of number of atoms requires a large number of training datasets to fit the PES. Thus, the use of lower-level methods such as density functional theory (DFT) and second-order Møller-Plesset perturbation (MP2) theory is understandable, but probably not accurate enough for precise investigations like the one here targeted. To circumvent this bottleneck, researchers are

applying ML approaches to bring a PES based on a low-level of electronic structure theory (DFT or MP2) to a higher level (CCSD(T)) one. One way to achieve this is by means of the Δ-machine learning (Δ-ML) approach, in which a correction is made to a property dataset obtained using an efficient, low-level *ab initio* theory such as DFT or MP2.[134,136–138,318,319]

We apply a Δ-machine learning approach that we recently reported[319,320] to take a DFT-level PES of ethanol that we recently reported[321] (details of the dataset of energies and gradients are given in that paper) to the CCSD(T) level using a manageable subset of *ab initio* CCSD(T) energy points. The new PES is tested against the usual fidelity tests and then employed for the challenging DMC and SC simulations and for an investigation of the wavefunctions of the $-CH_3$ and OH motions, for which Quade et al. have suggested a geared motion by analyzing microwave spectra.[322,323]

13.3 Theory and Computational Details

13.3.1 Δ Machine Learning for PES construction

The theory underneath our Δ-ML approach is very simple[319,320] and can be presented in a simple equation

$$V_{\mathrm{LL}\rightarrow\mathrm{CC}} = V_{\mathrm{LL}} + \Delta V_{\mathrm{CC\text{-}LL}}, \tag{13.1}$$

where $V_{\mathrm{LL}\rightarrow\mathrm{CC}}$ is the corrected PES, V_{LL} is a PES fit to low-level DFT electronic data, and $\Delta V_{\mathrm{CC\text{-}LL}}$ is the correction PES based on high-level coupled cluster energies. It is noted that the difference between CCSD(T) and DFT energies, $\Delta V_{\mathrm{CC\text{-}LL}}$, is not as strongly varying as V_{LL} with respect to the nuclear configurations and therefore just a small number of high-level electronic energies are adequate to fit the correction PES. In the

present application to ethanol, we computed a total of 2319 CCSD(T)-F12a/aug-cc-pVDZ electronic energies and performed training on a subset of these data in size of 2069 energies. This choice of basis was made to balance between accuracy and computational efficiency. We do compare several key energies using this basis with the published results using aVQZ basis.

Here we employ PIP approach to fit both the V_{LL} and $\Delta V_{\mathrm{CC\text{-}LL}}$ PESs. The theory of permutationally invariant polynomial is well established and has been presented in several review articles.[26–28,65,90] In terms of a PIP basis, the potential energy, V, can be written in compact form as

$$V(\boldsymbol{x}) = \sum_{i=1}^{n_p} c_i p_i(\boldsymbol{x}), \tag{13.2}$$

where c_{i} are linear coefficients, p_{i} are PIPs, n_{p} is the total number of polynomials for a given maximum polynomial order and $\boldsymbol{x}$ are Morse variables. For example, $x_{\alpha\beta}$ is given by $\exp(-r_{\alpha\beta}/\lambda)$, where $r_{\alpha\beta}$ is the internuclear distance between atoms α and β. The range (hyper)parameter, λ, was chosen to be 2 bohr. The linear coefficients are obtained using standard least squares methods for a large data sets of electronic energies (and for large molecules' gradients as well) at scattered geometries.

In order to develop a corrected PES, we need to generate a dataset of high and low-level energies for training and testing. In this study, we need both DFT and CCSD(T) datasets. Training is done for the correction PES $\Delta V_{\mathrm{CC\text{-}LL}}$, and testing is done for the corrected $V_{\mathrm{LL}\rightarrow\mathrm{CC}}$. Do note that this two-step "training and testing" is on different datasets.

Here we take the DFT dataset from our recently reported "MDQM21" dataset[321] where a total of 11000 energies and their corresponding gradients were generated from *ab initio* molecular dynamics (AIMD) simulations at B3LYP/6-311+G(d,p) level of theory. The DFT PES ($\mathrm{V_{LL}}$) was a fit using 8 500 DFT data, which span the energy range of

0?35 000 cm^{-1}. Here, we generate a sparse dataset that contains CCSD(T)-F12a/aVDZ energies at 2319 configurations, taken from the "MDQM21" dataset. The following procedure is employed to generate the dataset of 2319 configurations. First, we took every 8th geometry from the DFT training dataset of 8500 configurations, which gives a set of 1063 geometries. Then we took half of the DFT test dataset of 2500 geometries. From the another half of the DFT test dataset, we took just 6 geometries having energy greater than 30 000 cm^{-1} relative to the minima. These lead to a total of 2319 configurations subject to CCSD(T) single point energy computation. This 2319-geometry dataset is partitioned into a training dataset of 2069 geometries and a test dataset of 250 geometries, respectively. Histogram plots of the distribution of DFT and CCSD(T) electronic energies are shown in Figure 13.1, where it can be seen that both the DFT and CCSD(T) datasets span a similar energy range. Geometry optimization and normal-mode analysis are performed to examine the fidelity of the $V_{LL\rightarrow CC}$ PES.

13.3.2 Diffusion Monte Carlo

This PES is also applied to compute rigorous quantum zero-point energies (ZPEs) of ethanol and its single deuterated isotopologues using unbiased DMC calculations. The concept behind DMC is to solve the time-dependent Schrödinger equation in imaginary time.[55,56,124] This is done by simulating a random walk of many replicas, also called "walkers", of the molecule, using a birth/death processes. At each step, a random displacement in each degree of freedom is assigned to each walker, and this walker may remain alive (and may give birth to a new walker) or be killed by comparing its potential energy, E_i, with a reference energy, E_r. For the ground state, the probability of birth or

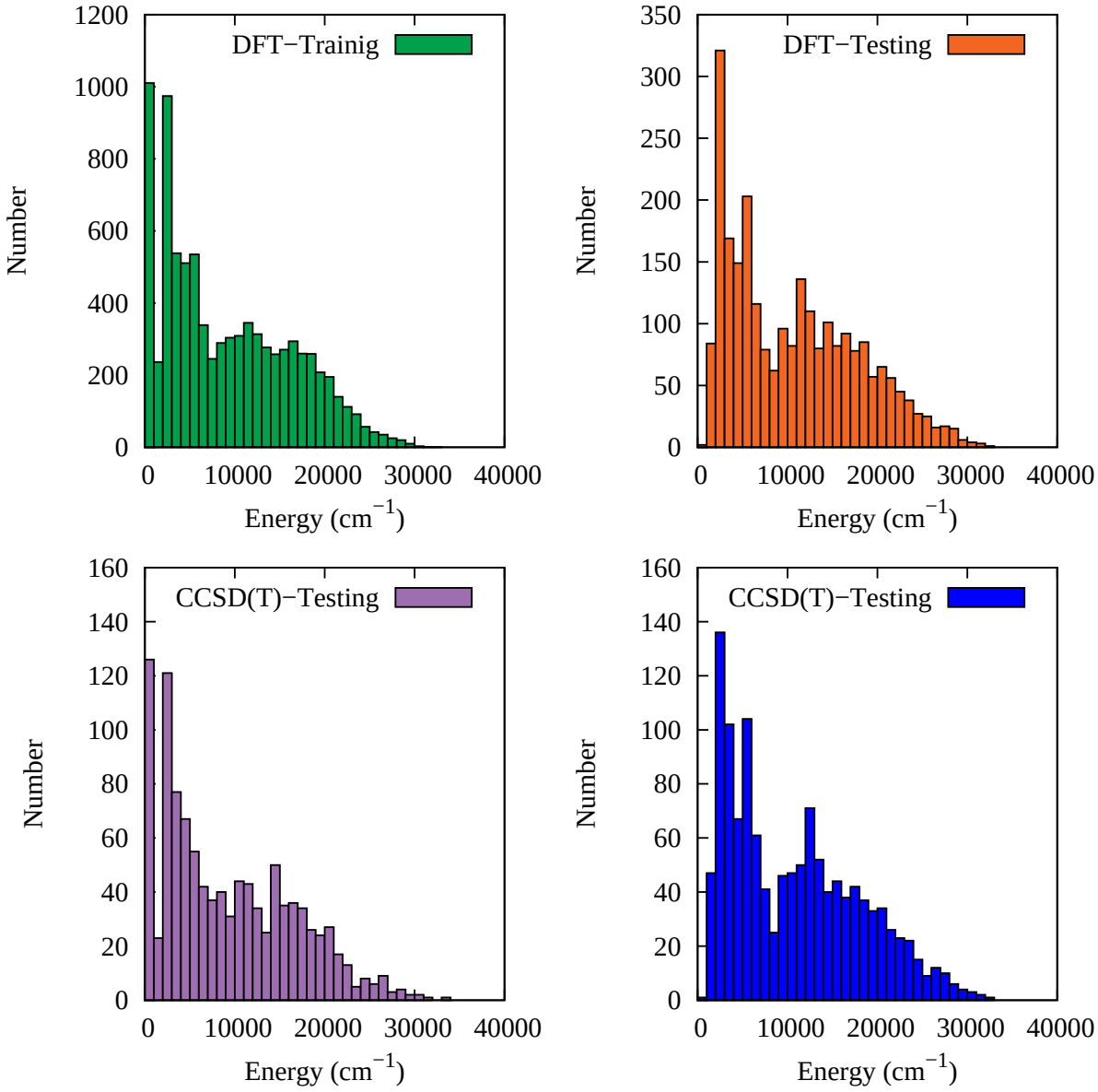


Figure 13.1: Distributions of DFT and CCSD(T) electronic energies (cm^{-1}) of both training and test datasets relative to their respective minimum value.

death is given as:

$$P_{\text{birth}} = \exp\left[-(E_i - E_r)\Delta\tau\right] - 1 \ (E_i < E_r) \tag{13.3}$$

$$P_{\text{death}} = 1 - \exp\left[-(E_i - E_r)\Delta\tau\right] \ (E_i > E_r), \tag{13.4}$$

where $\Delta\tau$ is the step size in imaginary time. After removing all dead walkers, the reference energy is updated using the equation

$$E_r(\tau) = \langle V(\tau)\rangle - \alpha\frac{N(\tau) - N(0)}{N(0)}, \tag{13.5}$$

where τ is the imaginary time; $\langle V(\tau)\rangle$ is the average potential over all the walkers that are alive; $N(\tau)$ is the number of live walkers at time τ; α is a parameter that can control the fluctuations in the number of walkers and the reference energy. Finally, the average of the reference energy over the imaginary time gives an estimate of ZPE.

In this study, each DMC trajectory is propagated for 30,000 time steps with step size of 5.0 a.u.; 20,000 steps are used to equilibrate the walkers, and the reference energies in the remaining 10,000 steps are used to compute the ZPE. For each isomer, 15 DMC simulations (or trajectories) were carried out, and the final ZPE is the average of the 15 simulations. Statistical uncertainty of the zero-point energy is defined as the standard deviation of DMC energies over the total number of simulations. This is written as

$$\Delta E = \sqrt{\frac{1}{15}\sum_{i=1}^{15}(E_i - \bar{E})^2}, \tag{13.6}$$

where $\bar{E}$ is the average energy over the 15 simulations. We also perform DMC calculations on three single deuterated isotopologues employing 15 DMC trajectories. For *trans*- and

gauche-CH_3CH_2OH and *trans*- and *gauche*-CH_3CH_2OD 40,000 random walkers are used, while for CH_2DCH_2OH and CH_3CHDOH only 20,000 random walkers are employed.

We note that we have used DMC calculations of zero-point energies in numerous similar applications using ML potential energy surfaces. Some recent examples and additional details of our implementation can be found in refs. 145,166,320.

13.3.3 Adiabatically Switched Semiclassical Initial Value Representation

Calculation of ethanol (*trans* and *gauche*) ZPEs, and those of its deuterated isotopologues can be performed from a dynamical point of view by means of the adiabatically switched semiclassical initial value representation (AS-SCIVR) technique. The goal is to corroborate DMC findings employing a completely different, but still full-dimensional, technique.[324–326] AS SCIVR is a recently developed two-step semiclassical approach able to regain quantum effects starting from classical trajectories. In this it is quite similar to standard semiclassical techniques[327–330], but it differs in the way the starting conditions of the semiclassical dynamics run are selected. In AS-SCIVR a preliminary adiabatic switching dynamics is performed. On the basis of the adiabatic theorem, this allows to start from harmonic quantization and approximately preserve quantization after switching on the true system Hamiltonian. The exit molecular geometry and momenta of the adiabatic switching run serve as starting conditions for the subsequent semiclassical dynamics trajectory. This entire procedure is applied to a distribution of harmonically quantized starting conditions.

In practice the adiabatic switching Hamiltonian is[331–333]

$$H_{\text{as}} = [1 - \lambda(t)]\, H_{\text{harm}} + \lambda(t) H_{\text{anh}}, \tag{13.7}$$

where $\lambda(t)$ is the following switching function

$$\lambda(t) = \frac{t}{T_{\text{AS}}} - \frac{1}{2\pi} \sin\left(\frac{2\pi t}{T_{\text{AS}}}\right), \tag{13.8}$$

H_{harm} is the harmonic Hamiltonian built from the harmonic frequency of vibration calculated by Hessian matrix diagonalization at the equilibrium geometry $\mathbf{q}_{\text{eq}}$, and H_{anh} is the actual molecular vibrational Hamiltonian. In our simulations T_{AS} has been chosen equal to 25000 a.u. (about 0.6 ps) and time steps of 10 a.u. have been employed. 5400 trajectories are evolved according to the Hamiltonian in Eq.(14.3) by means of a 4th order symplectic algorithm[334] starting from harmonic ZPE quantization.

Once the adiabatic switching run is over, the trajectories are evolved according to H_{anh} for another 25000 a.u. with same step size to collect the dynamical data needed for the semiclassical calculation. This relies on Kaledin and Miller's time-averaged version of semiclassical spectroscopy.[335,336] Therefore, the working formula is

$$I_{as}(E) = \left(\frac{1}{2\pi\hbar}\right)^{N_v} \sum_{i=1}^{N_{traj}} \frac{1}{2\pi\hbar T} \left| \int_0^T dt e^{\frac{i}{\hbar}[S_t(\mathbf{p}_{as},\mathbf{q}_{as}) + Et + \phi_t(\mathbf{p}_{as},\mathbf{q}_{as})]} \langle \Psi(\mathbf{p}_{eq},\mathbf{q}_{eq}) | g(\mathbf{p}'_t,\mathbf{q}'_t) \rangle \right|^2, \tag{13.9}$$

where $I_{\text{as}}(E)$ indicates that a vibrational spectral density is calculated as a function of the vibrational energy E. I_{as} is peaked at the eigenvalues of the vibrational Hamiltonian, the lowest one being the ZPE. Eq.(14.5) is made of several terms. N_{v} is the number of vibrational degrees of freedom of the system, i.e. 21 in the case of ethanol. T is the total evolution time of the dynamics for the semiclassical part of the simulation. As anticipated, we chose T equal to 25000 a.u. with a time step size of 10 a.u. $(\mathbf{p}'_{\text{t}}, \mathbf{q}'_{\text{t}})$ is

the instantaneous full-dimensional phase space trajectory. The semiclassical trajectory is started at time 0 from the final phase space condition $(\mathbf{p}_{\mathrm{as}}, \mathbf{q}_{\mathrm{as}})$ of the adiabatic switching part of the simulation. S_{t} is the classical action along the semiclassical trajectory, and ϕ_{t} is the phase of the Herman-Kluk pre-exponential factor based on the elements of the stability matrix and defined as

$$\phi_t = \mathrm{phase}\left[\sqrt{\left|\frac{1}{2}\left(\frac{\partial \mathbf{q}'_t}{\partial \mathbf{q}_{as}} + \Gamma^{-1}\frac{\partial \mathbf{p}'_t}{\partial \mathbf{p}_{as}}\Gamma - i\hbar\frac{\partial \mathbf{q}'_t}{\partial \mathbf{p}_{as}}\Gamma + \frac{i\Gamma^{-1}}{\hbar}\frac{\partial \mathbf{p}'_t}{\partial \mathbf{q}_{as}}\right)\right|}\right], \tag{13.10}$$

where Γ is an $N_{\mathrm{v}} \times N_{\mathrm{v}}$ matrix usually chosen to be diagonal with elements numerically equal to the harmonic frequencies. We note that evolution in time of ϕ_{t} requires calculation of the Hessian matrix, which represents the bottleneck of the AS-SCIVR approach and semiclassical methods broadly speaking. Based on Liouville's theorem, the stability (or monodromy) matrix has the property to have its determinant equal to 1 along the entire trajectory. However, classical chaotic dynamics can lead to numerical inaccuracies in the propagation, so, following a common procedure in semiclassical calculations, we have rejected the trajectories based on a 1% tolerance threshold on the monodromy matrix determinant value. Finally, the working formula is completed by a quantum mechanical overlap between a quantum reference state $|\Psi\rangle$ and a coherent state $|g\rangle$ characterized by the following representation in configuration space

$$\langle \mathbf{q}|g(\mathbf{p}'_t, \mathbf{q}'_t)\rangle = \left(\frac{\det(\Gamma)}{\pi^{N_\nu}}\right)\exp\left\{-(\mathbf{q}-\mathbf{q}'_t)^T\frac{\Gamma}{2}(\mathbf{q}-\mathbf{q}'_t) + \frac{i}{\hbar}\mathbf{p}_t'^T(\mathbf{q}-\mathbf{q}'_t)\right\}. \tag{13.11}$$

The reference state $|\Psi\rangle$ is usually chosen to be itself a coherent state. In Eq. (14.5) $|\Psi\rangle$ is written as $|\Psi(\mathbf{p}_{\mathrm{eq}}, \mathbf{q}_{\mathrm{eq}})\rangle$, where $\mathbf{p}_{\mathrm{eq}}$ stands for the linear momenta obtained in harmonic approximation setting the geometry at the equilibrium one $(\mathbf{q}_{\mathrm{eq}})$.

AS SCIVR allows for a full-dimensional investigation of zero-point energies of ethanol isomers. It is based on classical molecular dynamics and it is able to regain quantum effects by means of a stationary-phase approximation to Feynman's quantum propagator. Therefore, AS SCIVR is a very different approach from the stochastic DMC one, and we employ it to corroborate the outcomes of DMC calculations. There is no straightforward way to provide an uncertainty for these kinds of semiclassical approaches, but numerous previous studies (the interested reader can have a look, for instance, at Refs. 337–339) indicate that the method is able to approximate quantum results with an error ranging from very few wavenumbers to 20-30 cm^{-1}. We expect to find the same results of the benchmark DMC calculation within this range of uncertainty. AS SCIVR can also provide more information than DMC about excited states and quantum vibrational frequencies (including anharmonic overtones and combination bands). Calculation of ethanol fundamental frequencies of vibration including Fermi resonaces is left for a future work.

13.4 Results and Discussion

13.4.1 The starting low level PES (V_{LL})

The low level PES, V_{LL} was developed using the efficient B3LYP/6-311+G(d,p) level of theory. For the fit, we used maximum polynomial order of 4 with permutationally symmetry 321111, which leads to a total of 14752 PIPs in the fitting basis set. These were used to fit a dataset of 8500 energies and their corresponding gradients. The fitting RMS errors for energies and gradients are 40 cm^{-1} and 73 cm^{-1} $bohr^{-1}$, respectively. Testing was done on 2500 geometries. The testing RMS errors for energies and gradients

are 51 cm^{-1} and 106 cm^{-1} $bohr^{-1}$, respectively.

13.4.2 The correction PES ($\Delta V_{CC\text{-}LL}$)

A dataset of 2319 geometries are sparsely selected from the "MDQM21" DFT dataset and CCSD(T)-F12a/aug-cc-pVDZ energy computations are performed at those geometries. To develop the correction PES, we train $\Delta V_{CC\text{-}LL}$ on the difference between the CCSD(T) and DFT absolute energies of 2069 geometries and test the obtained surface on the remaining 250 geometries. A plot of $\Delta V_{CC\text{-}LL}$ versus the DFT energies for both training and test datasets is shown in Figure 13.2. Note that we reference $\Delta V_{CC\text{-}LL}$ to the minimum of the difference between the CCSD(T) and DFT energies (roughly 35 732 cm^{-1}). As seen, the energy range of $\Delta V_{CC\text{-}LL}$ is about 1800 cm^{-1}, which is much smaller than the DFT energy range relative to the minimum value (roughly 35 000 cm^{-1}).

The difference $\Delta V_{CC\text{-}LL}$ is not as strongly varying as V_{LL} with respect to the nuclear configuration. Therefore, low-order polynomials will be adequate to fit the correction PES. We use maximum polynomial order of 2 with permutational symmetry 321111 to fit the training dataset which leads to a total of 208 unknown linear coefficients (equivalent to the number of terms in the PIP fitting basis set). These coefficients are determined by solving a linear least-squares problem. The PIP basis to fit this PES is generated using our "in-house" MSA software.[123,141] The fitting RMS error of this $\Delta V_{CC\text{-}LL}$ fit is 25 cm^{-1}. The fit is tested on the 250 energy differences and the RMS test error in this case is 41 cm^{-1}.

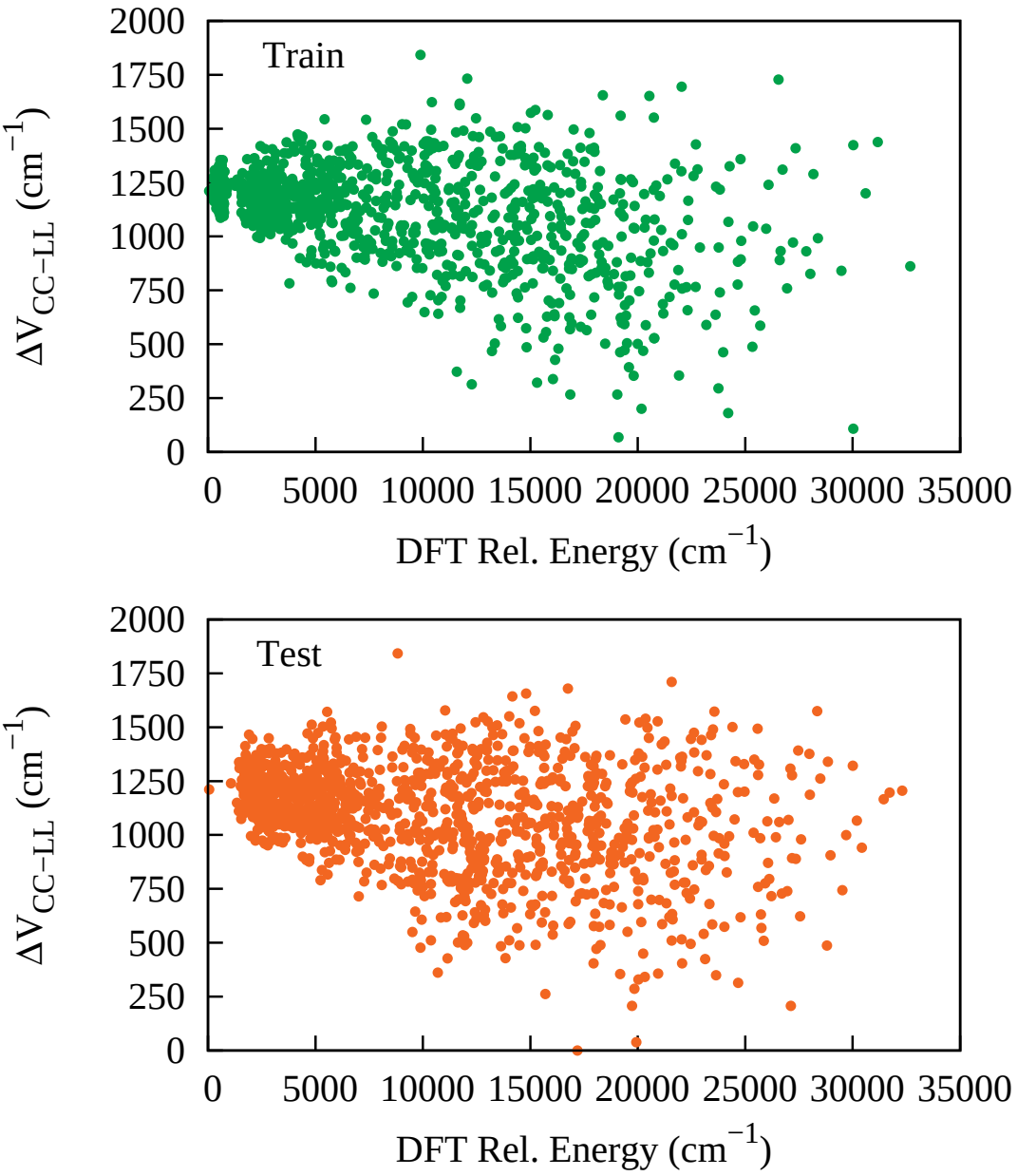


Figure 13.2: Plot of ΔV_{CC-LL} (relative to the reference value i.e. -35 732 cm^{-1}) vs DFT energy relative to the CH_3CH_2OH minimum value with the indicated number of training data sets.

13.4.3 The New CCSD(T) Ethanol PES ($V_{LL\rightarrow CC}$)

To obtain the CCSD(T) energies we add the correction $\Delta V_{CC\text{-}LL}$ to the low-level DFT PES, V_{LL}. A plot of $V_{LL\rightarrow CC}$ vs corresponding direct CCSD(T) energies for the training set of 2069 points and the test set of 250 points is shown in Figure 13.3. As seen, there is overall excellent precision; however, we see a few larger errors. The RMS differences between the $V_{LL\rightarrow CC}$ and direct CCSD(T) energies for the training and test datasets are 49 cm^{-1} and 63 cm^{-1}, respectively.

To examine this fidelity of the new $V_{LL\rightarrow CC}$ PES, we perform geometry optimization and normal mode frequency calculation of both *trans* and *gauche* isomers and their two isomerization saddle point geometries. They are the *eclipsed* one, in which the hydroxylic hydrogen eclipses with the hydrogen of the adjacent CH_2 group, and the *syn* one, in which the hydroxylic hydrogen is above the methyl group. The structures of these isomers and saddle points are shown in Figure 13.4. We get the PES optimized energies within 5 cm^{-1} of the direct CCSD(T)-F12a calculation and find that the *trans* isomer is lower in energy by 38 cm^{-1}. Next, to examine the vibrational frequency predictions of the PES, we perform normal mode analyses for both *trans* and *gauche* isomers and their isomerization saddle points. The comparison of harmonic mode frequencies of *trans* and *gauche* ethanol with their corresponding *ab initio* ones are shown in Table 13.1. The agreement with the direct CCSD(T)-F12a/aug-cc-pVDZ frequencies is overall very good; the maximum error is 21 cm^{-1} for the lowest frequency mode of *trans* conformer, but most of the frequencies are within a few cm^{-1} of the *ab initio* ones and the mean absolute error (MAE) is only 4 cm^{-1}. The *gauche* isomer shows even better agreement with the *ab initio* data. The two *trans* - *gauche* isomerization saddle point geometries such as *eclipse* and *syn* ones are confirmed by obtaining one imaginary frequency. The normal mode frequencies of

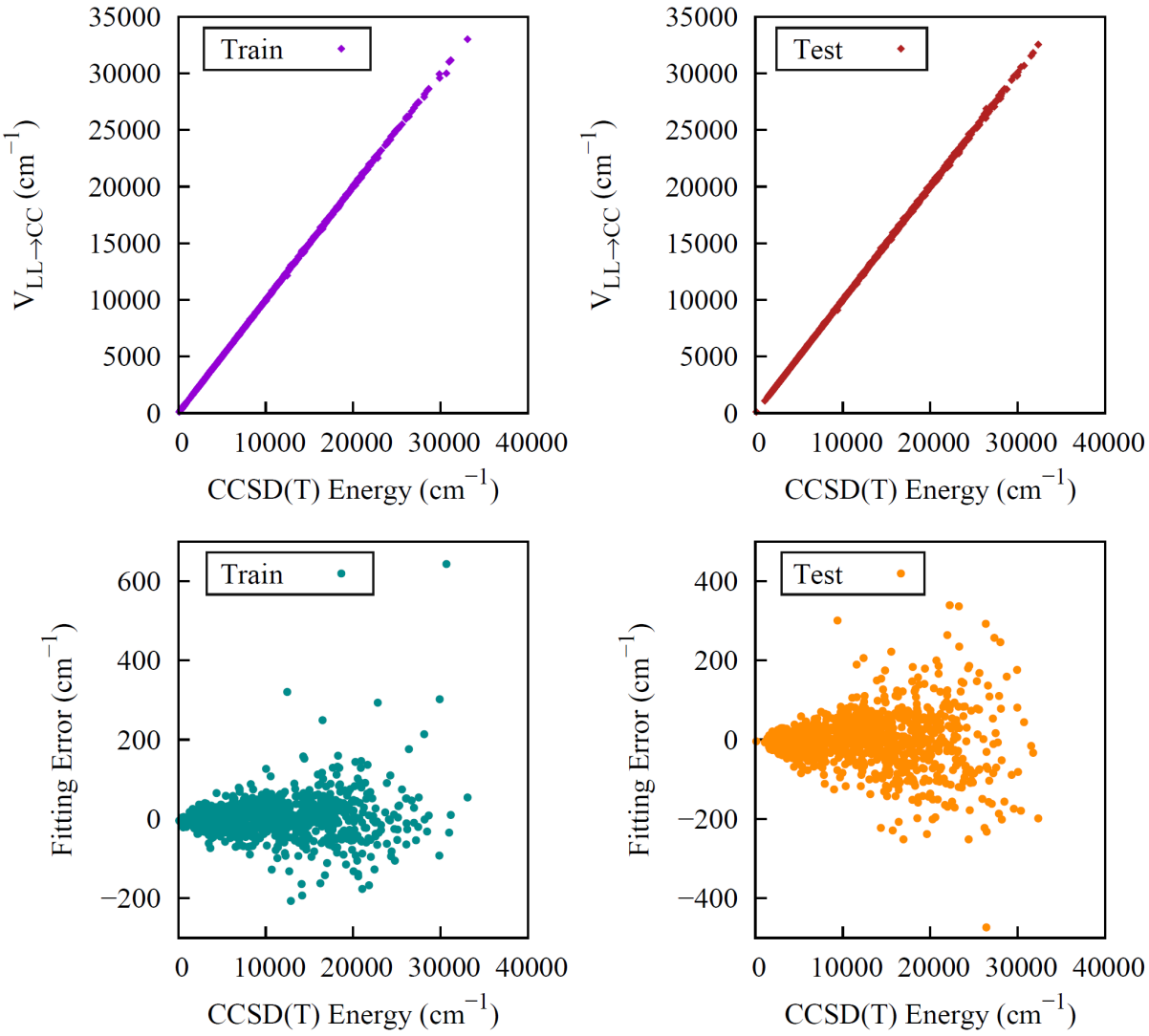


Figure 13.3: The two upper panels show energies of CH_3CH_2OH from $V_{LL\rightarrow CC}$ vs direct CCSD(T) ones for the indicated data sets. The one labeled "Train" corresponds to the configurations used in the training of $\Delta V_{CC\text{-}LL}$ and the one labeled "Test" is just the set of remaining configurations. Corresponding fitting errors relative to the minimum energy are given in the lower panels.

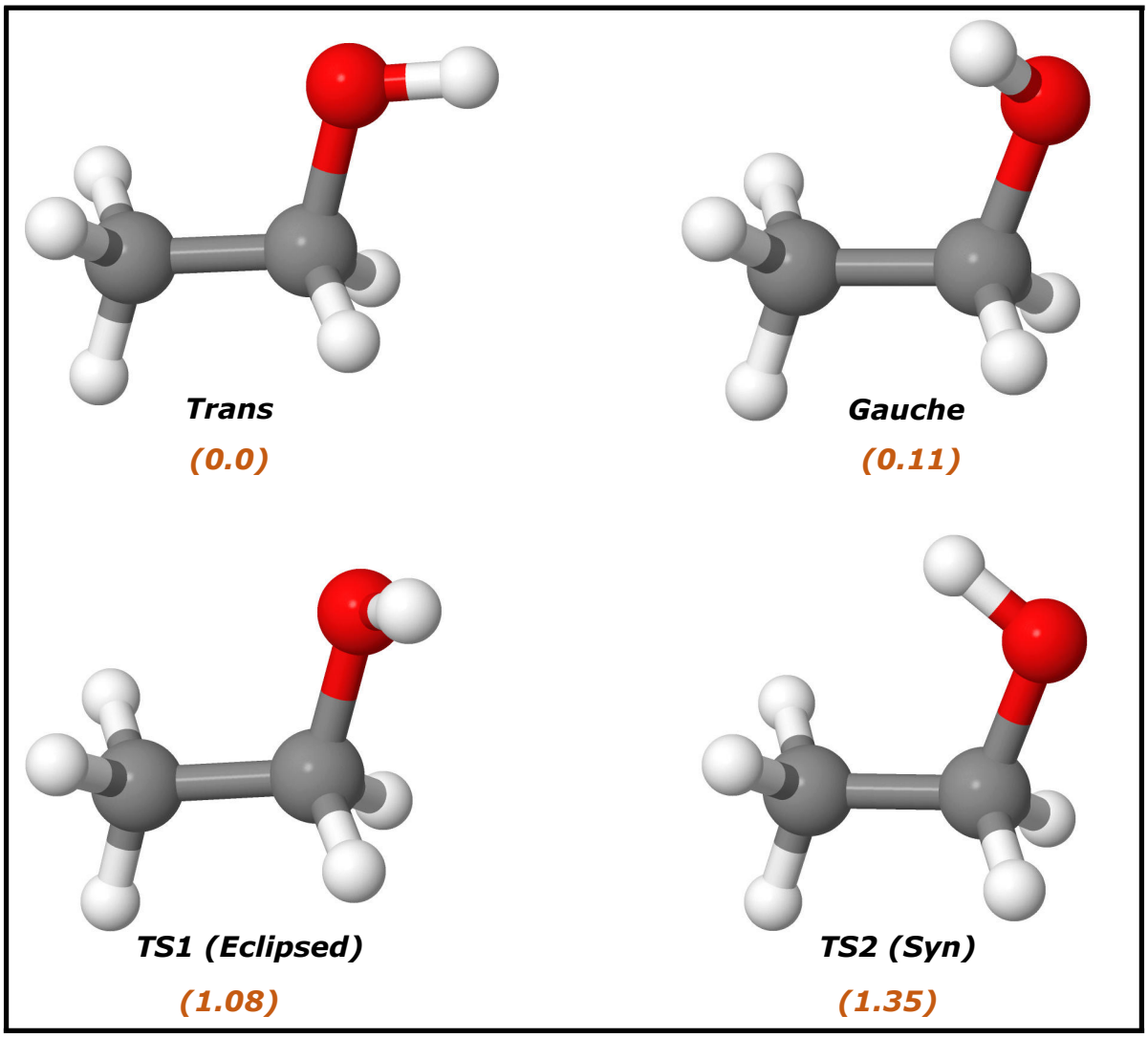


Figure 13.4: Geometry of *trans* and *gauche* conformers of ethanol and their two isomerization TSs and their electronic energies (kcal/mol) relative to the *trans* minimum from Δ-ML PES.

this saddle point geometry are given in Table 13.2. The barrier height of *trans* - *gauche* isomerization with respect to eclipse and syn TSs are found to be 377 cm^{-1} and 472 cm^{-1}, respectively, and the corresponding direct *ab initio* values are 389 cm^{-1} and 438 cm^{-1}. These are in excellent agreement with the experimental barrier heights of 402 cm^{-1} and 444 cm^{-1}.[305]

Table 13.1: Comparison of harmonic frequencies (in cm^{-1}) between $V_{LL\rightarrow CC}$ PES and the corresponding *ab initio* (CCSD(T)-F12a/aug-cc-pVDZ) ones of both *trans* and *gauche* isomers of ethanol.

	trans-ethanol			*gauche*-ethanol		
Mode	Δ-ML PES	*ab initio*	Diff.	Δ-ML PES	*ab initio*	Diff.
1	243	222	-21	268	258	-10
2	273	274	1	278	271	-7
3	417	413	-4	424	420	-4
4	818	813	-5	804	803	-1
5	909	907	-2	894	895	1
6	1055	1049	-6	1075	1069	-6
7	1115	1115	0	1094	1096	2
8	1181	1180	-1	1144	1141	-3
9	1284	1274	-10	1290	1284	-6
10	1302	1300	-2	1375	1374	-1
11	1403	1402	-1	1406	1402	-4
12	1454	1456	2	1424	1426	2
13	1488	1484	-4	1490	1491	1
14	1500	1501	1	1496	1497	1
15	1530	1531	1	1519	1522	3
16	2995	3001	6	3007	3014	7
17	3028	3036	8	3020	3028	8
18	3036	3042	6	3088	3089	1
19	3120	3122	2	3108	3108	0
20	3126	3127	1	3121	3123	2
21	3862	3853	-9	3845	3837	-8

Table 13.2: Comparison of harmonic frequencies (in cm^{-1}) between $V_{LL\rightarrow CC}$ PES and the corresponding *ab initio* (CCSD(T)-F12a/aug-cc-pVDZ) ones of both *eclipsed* and *syn* TSs of Ethanol.

	eclipsed			syn		
Mode	Δ-ML PES	*ab initio*	Diff.	Δ-ML PES	*ab initio*	Diff.
1	267i	287i	20i	332i	300i	-32i
2	261	256	-5	270	271	1
3	420	416	-4	411	414	3
4	800	797	-3	812	807	-5
5	899	899	0	892	892	0
6	1058	1064	6	1057	1061	4
7	1106	1106	0	1105	1109	4
8	1133	1132	-1	1186	1187	1
9	1285	1285	0	1307	1298	-9
10	1370	1358	-12	1308	1306	-2
11	1399	1397	-2	1406	1402	-4
12	1428	1427	-1	1446	1440	-6
13	1485	1486	1	1493	1493	0
14	1500	1598	-2	1507	1502	-5
15	1522	1520	-2	1534	1539	5
16	3020	3028	8	3015	3027	12
17	3028	3034	6	3027	3030	3
18	3059	3069	10	3054	3061	7
19	3112	3123	1	3103	3106	3
20	3123	3124	1	3109	3113	-6
21	3896	3890	-6	3872	3865	-7

Another comparison to the experiment we are able to perform thanks to the new PES concerns the torsional barrier for the methyl rotor. The methyl rotor torsional potentials (not fully relaxed) for both *trans* and *gauche* isomers as a function of the torsional angle are shown in Figure 13.5. It is seen that results from the PES are very close to the ones obtained from direct *ab initio* calculations at CCSD(T) level by means of a set of single point calculations. We obtain that the methyl torsional barriers for *trans* and *gauche* isomers are 1208 cm^{-1} and 1324 cm^{-1}, respectively. The methyl torsional barrier heights extrapolated from microwave spectroscopy for the *trans* and *gauche* isomers are 1174 cm^{-1} and 1331 cm^{-1}.[303,340,341] A different experimental analysis of the infrared and Raman spectra determined the methyl torsional barriers to be 1185 cm^{-1} and 1251 cm^{-1} for *trans* and *gauche*, respectively.[305] To complete our investigation of torsional barriers, we also report the methyl rotor torsional potential (not fully relaxed) for TS1 and TS2 geometries as a function of the CH_3 torsional in Figure 13.6. We get perfect three-fold symmetry with barrier heights of 1283 and 1404 cm^{-1}, respectively.

This is another proof of the accuracy of the new PES and another evidence of experimental results obtained from ethanol vibrational spectroscopy being not conclusive. So far only electronic energies have been investigated, but we now move to consider nuclear quantum effects.

As a remarkable quantum nuclear application of the PES, we present the results of diffusion Monte Carlo (DMC) calculations of the zero-point energy (ZPE) for both *trans* and *gauche* isomers and singly deuterated isotopologues. In addition to that, it is well known that a DMC calculation is a very challenging test to examine the quality of a PES in extended regions of the configuration space. A common issue in PES fitting is the unphysical behavior in the extrapolated regions where the fitting dataset is lacking data, and this is dramatically manifested by large negative values. These are referred

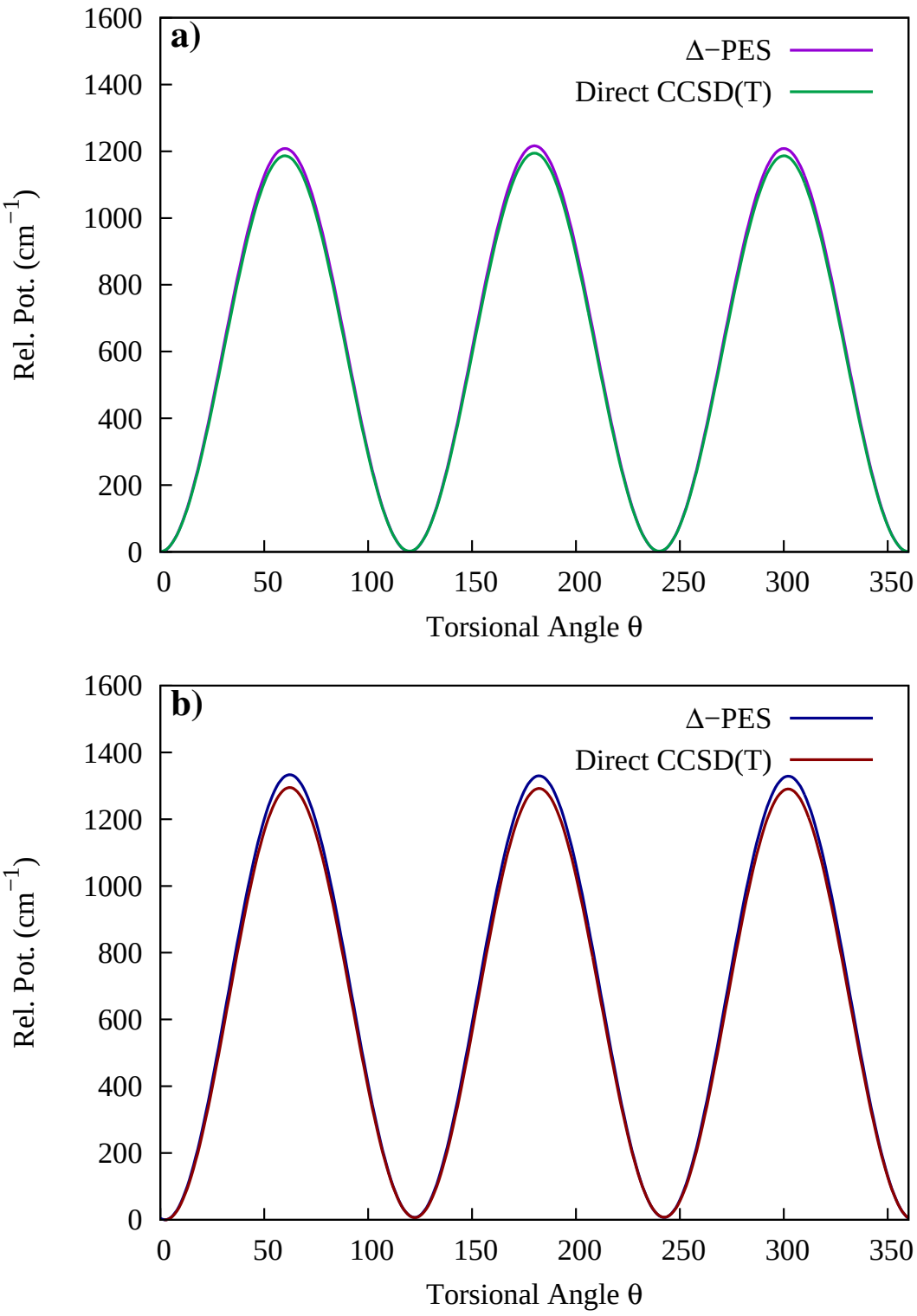


Figure 13.5: Comparison of torsional potential (not fully relaxed) of the methyl rotor of *trans* (a) and *gauche* (b) Ethanol between direct CCSD(T) and Δ-ML PES.

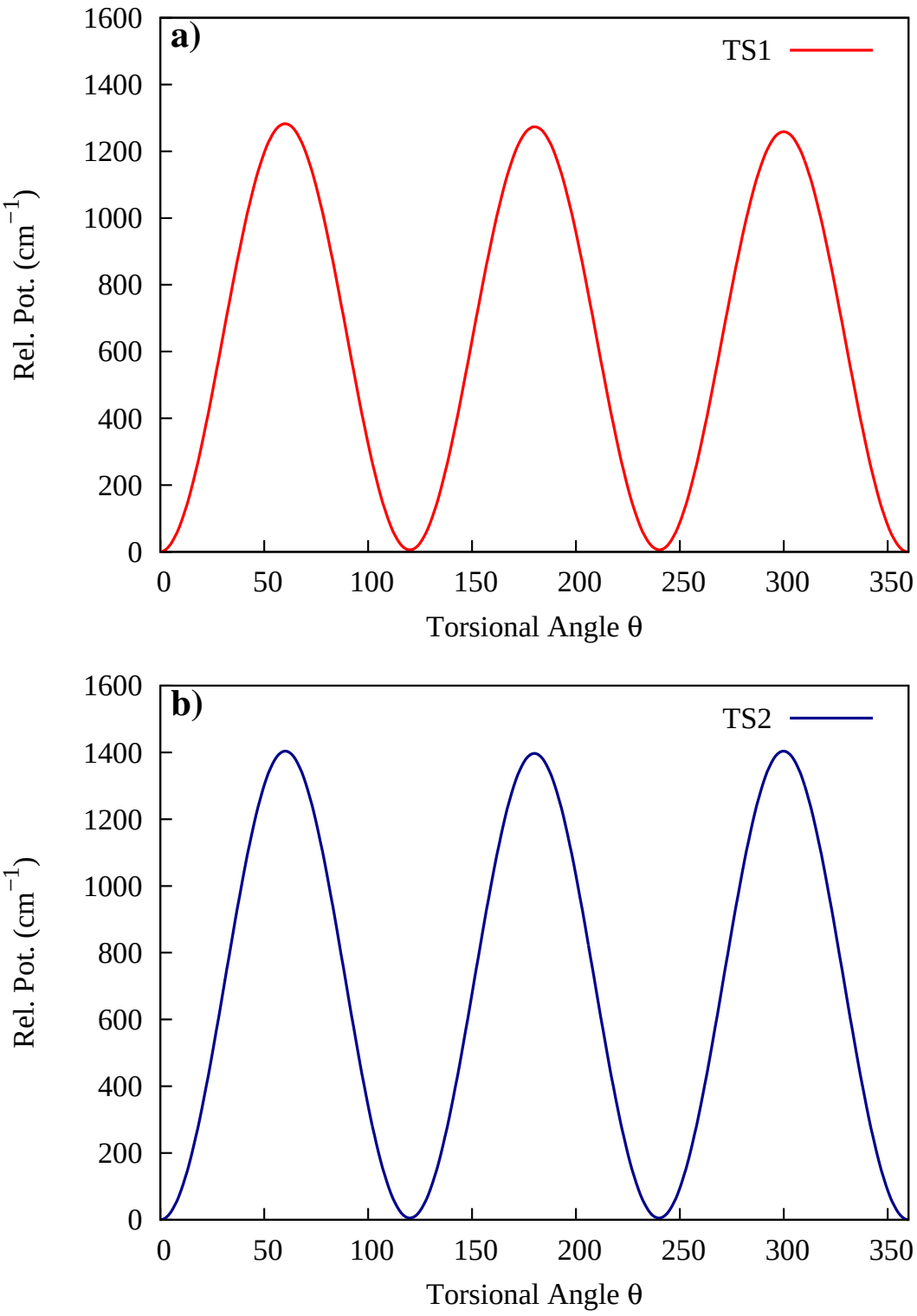


Figure 13.6: Torsional potential (not fully relaxed) of the methyl rotor of TS1 (a) and TS2 (b) geometry of Ethanol.

to as "holes" in the PES. Generally, we have observed that "holes" occur for highly repulsive configurations, i.e., short internuclear distances. Adding some more data in these regions and perform a refit generally eliminates the issue. So, one goal of presenting DMC calculations is also to demonstrate that our PES correctly describes the high energy regions of ethanol and it is therefore suitable for quantum approaches that need to sample these regions.

Table 13.3: Harmonic, DMC, and SC ZPEs (cm^{-1}) of *trans* and *gauche* ethanol and singly deuterated isotopologues. The zero of energy is set at the electronic global minimum. Values inside the parentheses represent statistical uncertainties in the DMC results.

Molecule	Harmonic ZPE	DMC ZPE	SC ZPE
CH_3CH_2OH(*trans*)	17568	17321 (9)	17298
CH_3CH_2OH(*gauche*)	17621	17321 (6)	17317
CH_3CH_2OD(*trans*)	16842	16619 (6)	16598
CH_3CH_2OD(*gauche*)	16894	16619 (8)	16611
CH_2DCH_2OH(*trans*)	16874	16649 (7)	16622
CH_3CDHOH(*trans*)	16836	16613 (9)	16586

Table 13.3 shows the DMC ZPEs of ethanol (both isomers) and singly deuterated isotopologues of the *trans* conformer along with semiclassical and harmonic ZPEs. It is seen that the agreement between AS-SCIVR and DMC ZPEs is very good and within method uncertainties (for SC methods uncertainty is typically within 20-30 cm^{-1}). Relative to the electronic global minimum, i.e. the bottom of the *trans* conformer well, the DMC ZPEs of *trans* and *gauche* isomers are 17321 $\pm$ 9 cm^{-1} and 17321 $\pm$ 6 cm^{-1}, respectively, whereas the corresponding SC ones are 17298 cm^{-1} and 17317 cm^{-1}, and the harmonic ZPEs are 17568 cm^{-1} and 17621 cm^{-1}. The harmonic ZPEs of the *trans* and *gauche* overestimate the true ZPE values by about 250-300 cm^{-1} revealing a substantial level

of anharmonicity. We note that in the DMC calculations very few "holes" are detected and in just a couple of trajectories. The total number of "holes" detected is 44, which is negligible compared to the total number of configurations (of the order of 10^{11}) sampled during the DMC trajectory calculations. This demonstrates that our PES can be in practice considered "hole-free". A further certification of this is given by the AS-SCIVR simulations, which are successfully run at energies close to the ZPE one. During DMC propagation, when a random walker encounters a "hole" (and thus it enters a region of large potential energy), we kill that walker and let the trajectory continue to propagate. This procedure follows our unbiased DMC algorithm.

We believe this is the first time the quantum anharmonic ZPE of ethanol is reported at CCSD(T) level of theory. At this point a comparison of our values to the experimentally-derived ones is very insightful. Since the experiment has the ZPE in it, we compare our DMC and SC results with 41 cm^{-1}, which is the experimental energy difference value we already anticipated in the Introduction. The bare electronic energy difference on the PES is 38 cm^{-1} with the *trans* conformer being the lower energy one. SC calculations estimate an energy difference of 19 cm^{-1} still in favor of the *trans* conformer, while DMC results have the two conformers basically degenerate. These values suggest an energy gap narrower than the experimentally-derived one, with SC and DMC results in agreement within uncertainty.

The DMC vibrational ground-state wavefunctions for hydrogens for both *trans* and *gauche* conformers are shown in Figure 13.7. The DMC results clearly show that the ground-state wavefunction has a *trans* fingerprint even when starting from the *gauche* conformer. On the other hand, the ground-state wavefunction is partly delocalized at the *gauche* geometry. This conclusion is corroborated by the top panel of Figure 13.8, which shows the distribution of walkers at the end of DMC trajectories (15 DMC trajectories

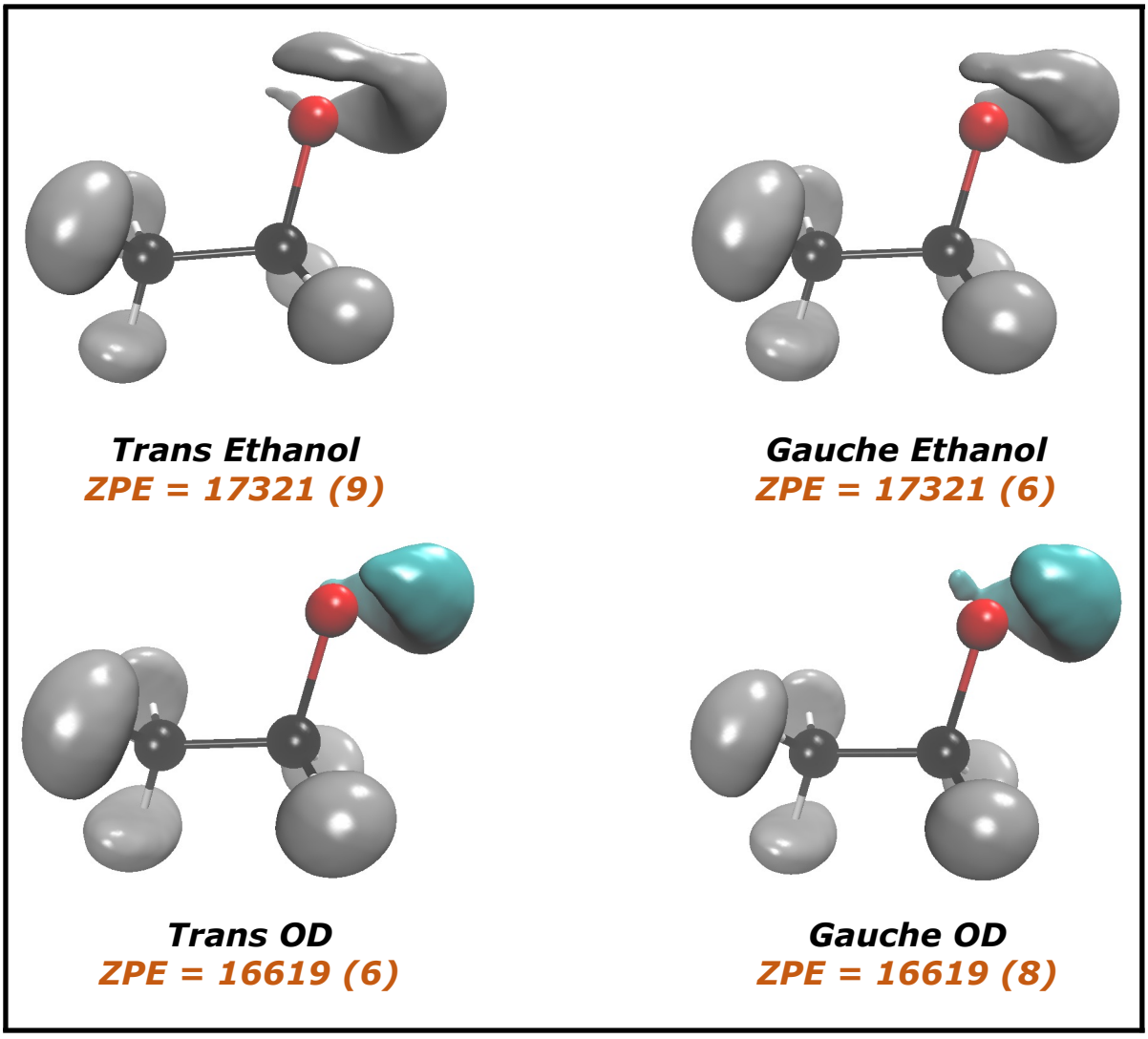


Figure 13.7: Vibrational ground-state wavefunction. The two upper panels represent the *trans* and *gauche*-ethanol and the two lower panels represent the *trans*-CH_3CH_2OD and *gauche*-CH_3CH_2OD. The hydrogen atom attached to the oxygen atom has been removed to help the eye. ZPEs values are reported with uncertainties in parentheses.

are computed, so total number of walkers are roughly $15 \times 40\,000 = 600\,000$.) started from the *trans* configuration relative to the C1-C2-O-H torsional angle. The *gauche* geometry is found at the torsional angle of ± 120 degrees.

We also present the vibrational ground-state wavefunction from DMC calculations for the OD motion in *trans*-CH_3CH_2OD, i.e. one of the singly deuterated isotopologues of *trans* ethanol in Figure 13.7. The ZPEs for the deuterated isotopologue is still equivalent with very similar wavefunctions. In the case of deuteration, the bottom panel of Figure 13.8 shows that the torsional angle distribution is more centered at the *trans* geometry and only very few walkers are found at *gauche* geometry. This shows that, on the one hand, quantum delocalization is somewhat quenched by the deuteration, while, on the other hand, starting from the deuterated *gauche* conformer still leads to the deuterated *trans* one.

The wave function of the OD motion still looks delocalized, but an interesting effect of deuteration on the dynamics of ethanol can be pointed out by examining AS-SCIVR calculations. In fact, as anticipated, a certain rate of AS-SCIVR trajectories are numerically unstable and discarded according to a threshold parameter, as defined in the Theory and Computational Details section. The rejection rate we find is about 55% for both the *trans* and *gauche* conformers and also for the methyl-deuterated isotopologues. Conversely, for CH_3CH_2OD the rejection rate decreases to about 20% and 38% for the *trans* and *gauche* conformer, respectively. This somehow strengthens DMC calculations by providing evidence of a more vibrationally-localized motion for OD with respect to OH and a clue of a reduced influence of the "leak" effect.

Then, it is interesting to compare the 1-D O-H torsional potential determined from our full-dimensional PES with the model used by Pearson, Brauer, and Drouin (PBD).[309] As shown in Figure 13.9 the two are very similar. The relative potential energies of the

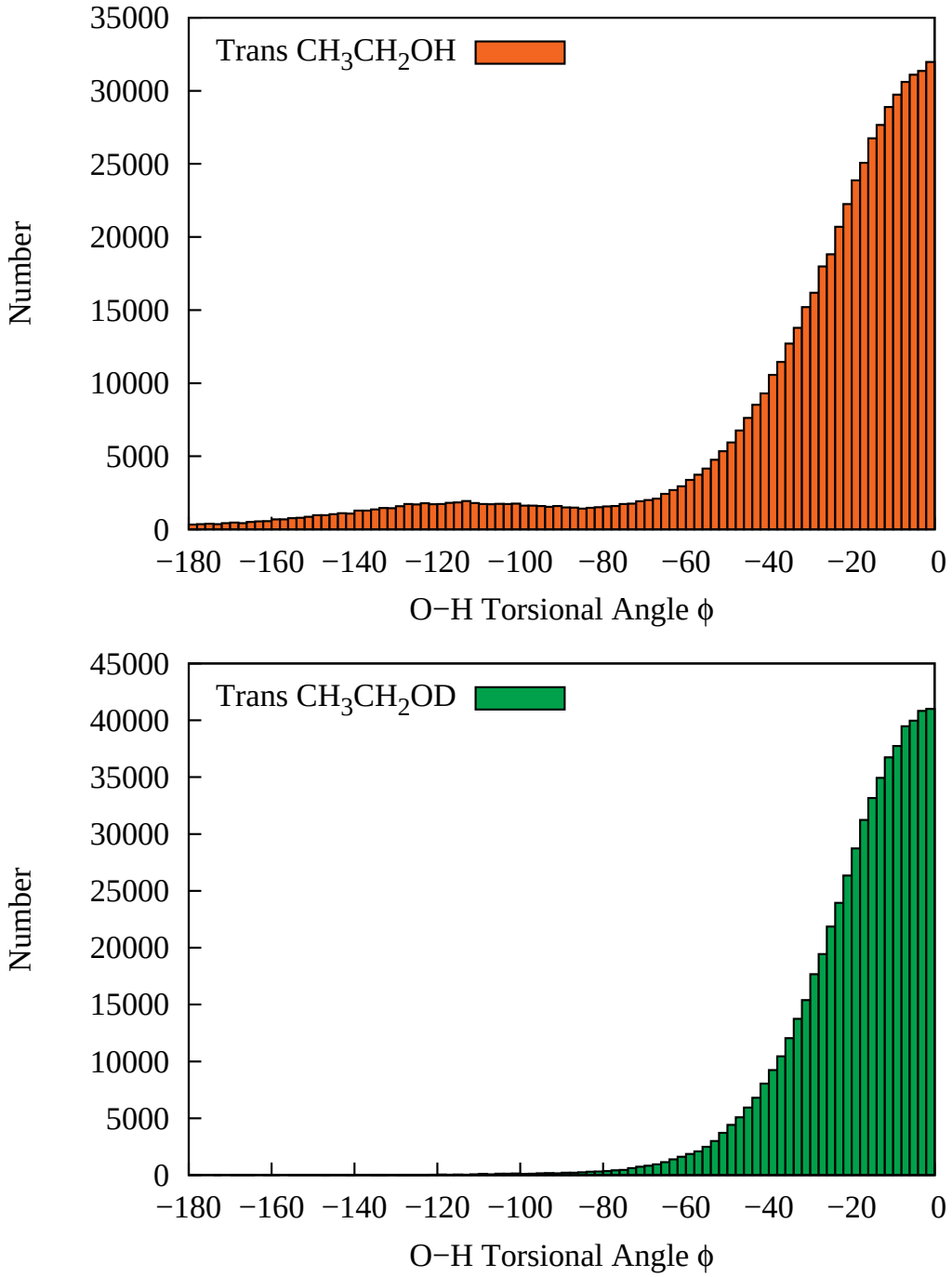


Figure 13.8: Distribution of C1-C2-O-H torsional angle (ϕ) from the DMC walkers. The upper panel represents the *trans*-CH_3CH_2OH and the lower panel represents the *trans*-CH_3CH_2OD.

gauche state and TS1 with respect to the *trans* state are nearly the same, while TS2 is somewhat higher in energy for the 1-D potential from our PES as compared to PBD. Recall that the 1-D OH torsional is not fully relaxed, so some minor adjustments to it might be anticipated.

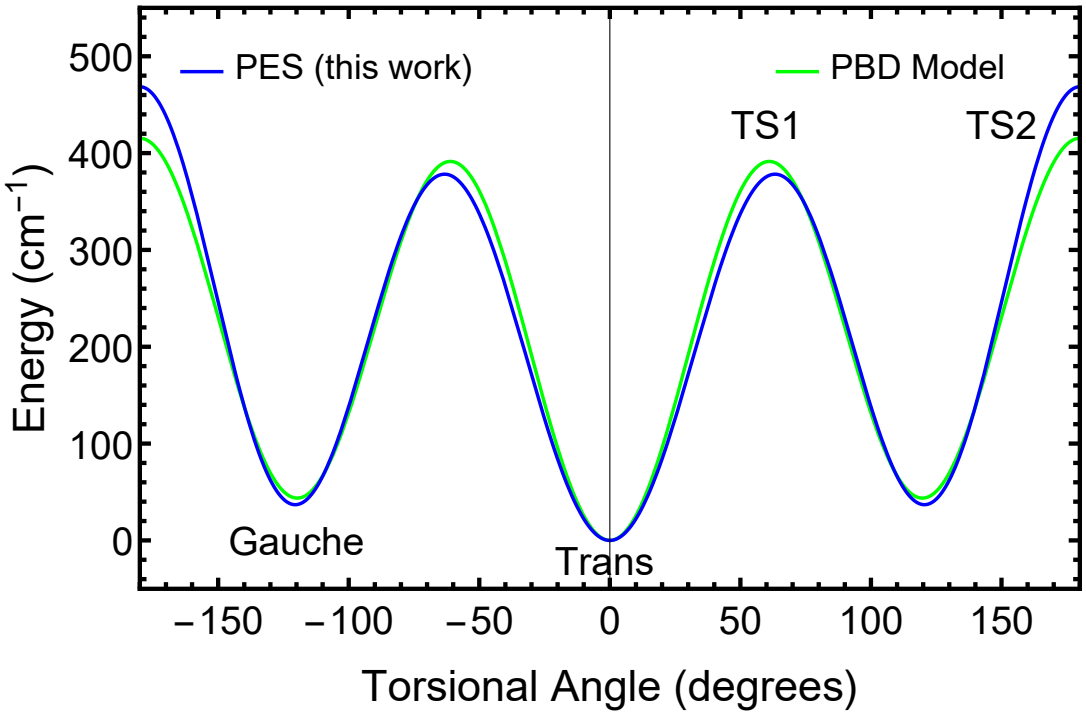


Figure 13.9: Comparison of C1-C2-O-H torsional potentials from this work (blue) and from PBD (green).

Having an 1-D model available is always advantageous because one can easily compute the energy levels and the corresponding wavefunctions. Our preferred method for doing so is by using the Discrete Variable Representation (DVR) techniques described in ref. 342. For the problem at hand, we use the azimuthal (0 to 2 π interval, periodic) variant. There is really only one adjustable parameter, the moment of inertia of the rotor. An

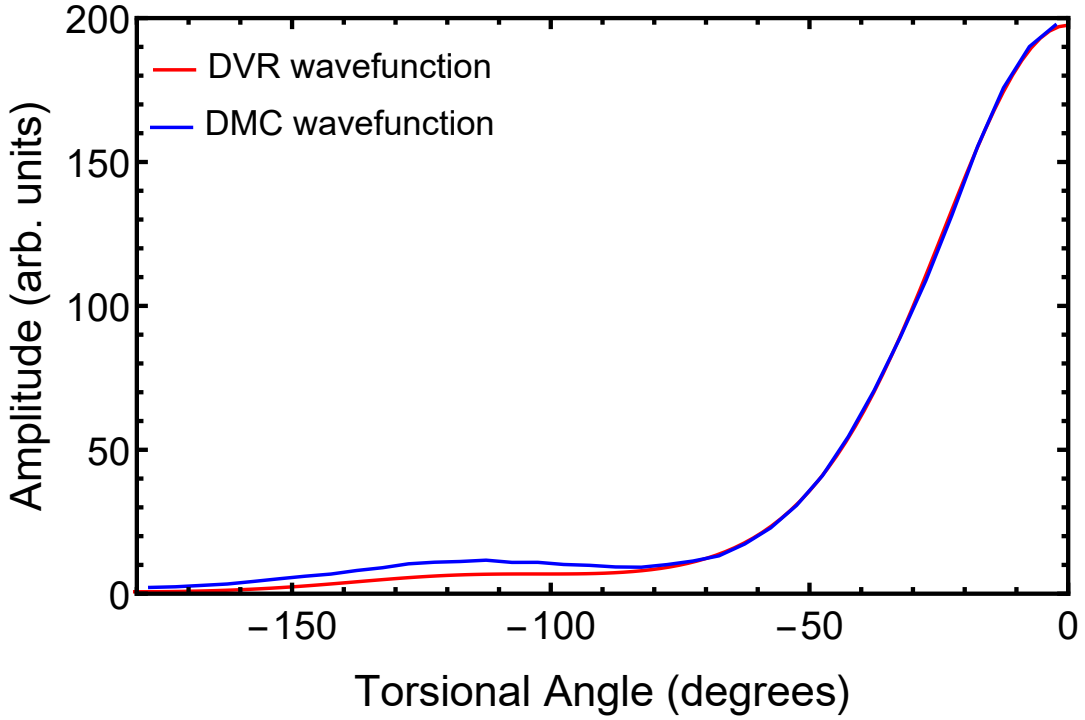


Figure 13.10: Comparison of the ground state OH torsional wavefunctions as determined from Discrete Variable Representation calculation on a 1-D cut (red) and from Diffusion Monte Carlo calculations on the full-dimensional PES (blue). Note that both wavefunctions have substantial amplitude near 120°, the geometry of the *gauche* state.

estimate for this might be $\mu_{O-H} \times r_{OH}^2$, where $\mu_{\text{O-H}}$ is the reduced mass of the OH in atomic units, and r_{OH} is the equilibrium distance of the O-H bond in bohr. For ethanol, this is about $3.2/(N_{\text{AV}} * m_{\text{e}})$, where N_{AV} is Avogadro's number and m_{e} is the mass of the electron. We reduced this numerical value from 3.2 to 2.7 so that, when applied to the PBD model torsional potential, we obtained agreement with their energy differences. With this moment of inertia then applied to our own PES, we obtained the energy levels and wavefunctions shown in Fig. 13.11, where the wavefunctions for only the first two levels are shown. It is interesting to note that there is substantial wave function amplitude for the *gauche* state at the geometry of the *trans* state and for the *trans* state at the geometry of the *gauche* state, an observation that was shown for the *trans* state also in the DMC results of Figure 13.8 based on the full-dimensional PES. In fact, the DMC *trans* wavefunction from Figure 13.8 and the wavefunction from Figure 13.11 are nearly identical, as shown in Figure 13.10.

Of course, a 1-D potential tells only a small part of the story. Two cuts of the 1-D CH_3 torsional potential have previously been shown in Figure 13.5. When we combine these cuts with two others (taken at the OH torsional angles corresponding to TS1 and TS2, see Figure 13.6) as well as with the OH torsional potential of Figures 13.9 and 13.11, we can obtain a reasonable fit for a 2-D potential of the combined motions of the OH and the CH_3, as shown in Figure 13.12. As described in the caption, when both OH and CH_3 are rotating, the minimum energy path for moving, for example, from the well at $\{\theta, \phi\} = \{0°, -120°\}$ to $\{360°, 120°\}$ is not at all straight, but rather follows the dashed black saw-tooth path reflecting the geared motion of the two rotors. As anticipated in the Introduction, this geared motion in ethanol has been suggested previously by Quade and colleagues from analysis of microwave spectra, but, to our knowledge, it has not previously been shown via a full-dimensional PES.

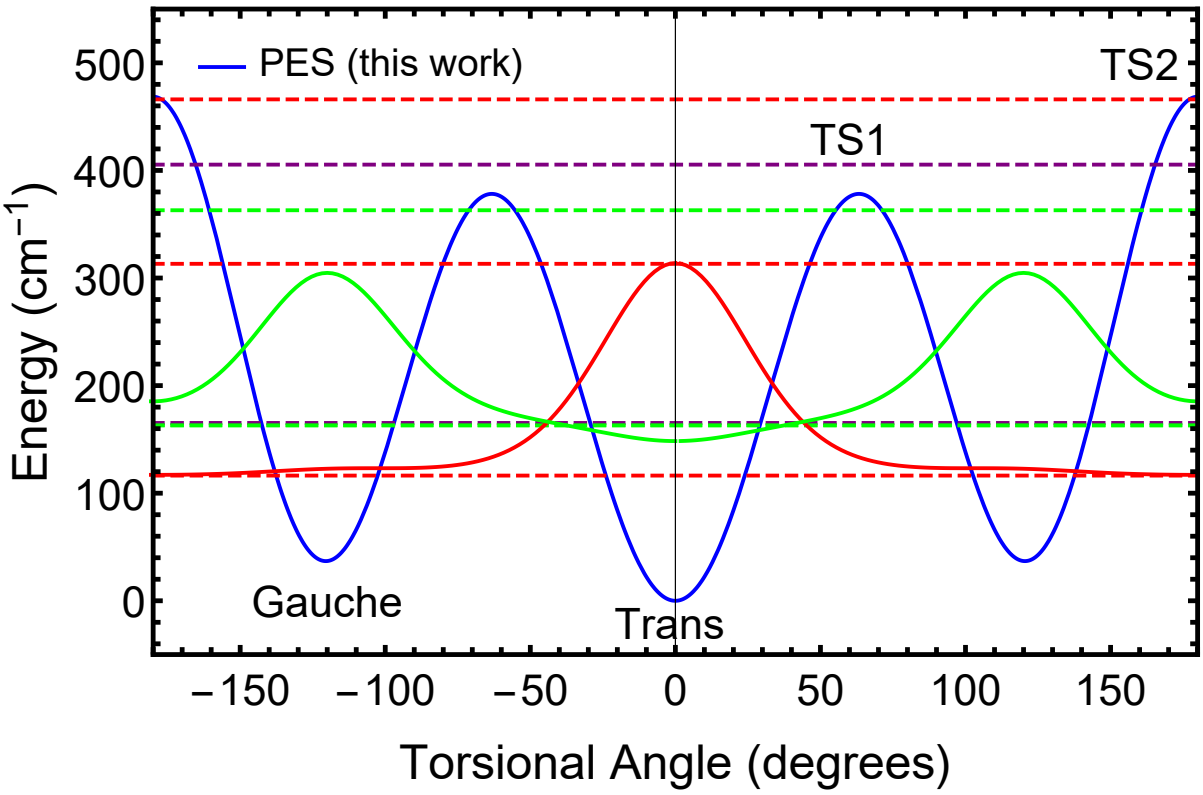


Figure 13.11: DVR results for energies and wavefunctions based on the 1-D C1-C2-O-H torsional potential from this work. The solid blue curve gives the potential, while the dotted lines give the first seven energy levels (there are two levels at 163.1 and 165.3 cm^{-1}). The solid red and green lines give the wavefunctions corresponding to the two lowest torsional energy levels.

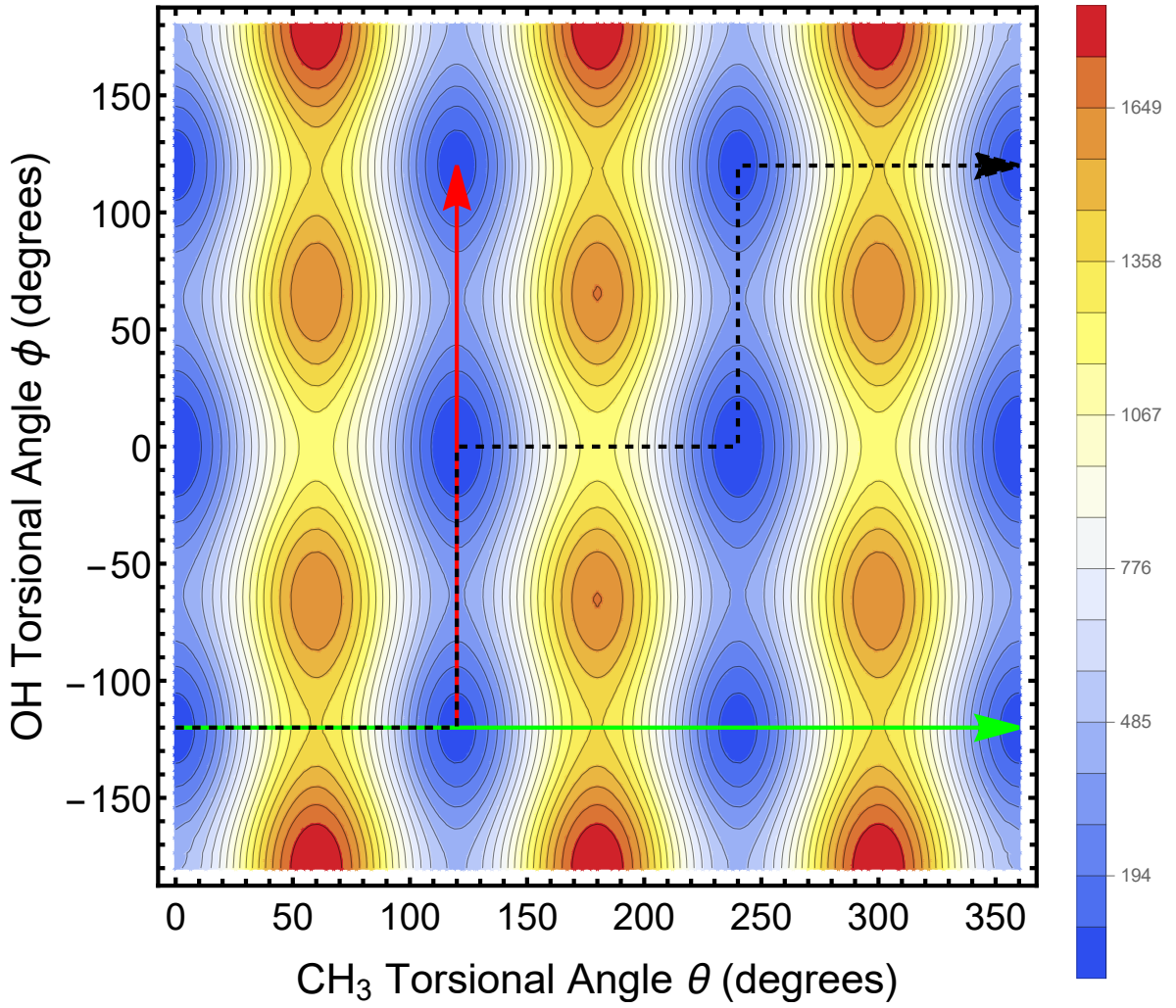


Figure 13.12: Two-dimensional contour plot of the potential energy as a function of OH torsional angle and CH_3 torsional angle. The color scale gives the potential in cm^{-1}. If the CH_3 torsional angle is constant, for example at 0°, 120°, 240°, or 360°, the lowest energy path for the OH rotational motion is in the vertical direction, for example along the red arrow. If the OH torsional angle is constant, for example at -120°, 0°, or 120°, then the lowest energy path for CH_3 rotational motion is in the horizontal direction, for example along the green arrow. If both the OH and CH_3 are rotating, instead of moving in a straight line, say from $\{\theta, \phi\}$={0°, -120°} to {360°, 120°}, the lowest energy path is to move along the saw-tooth arrow, which describes a geared motion in which the horizontal and vertical displacements take place along the minimum energy paths.

The functional form of the 2-D fit to the methyl and OH torsional motions shown in the 2-D contour plot (Figure 13.12) is presented here. The best values of the variables in Table 13.4 were obtained by simultaneously fitting five cuts of the OH and CH_3 torsion calculated from the full-dimensional PES. There were two unknown parameters. These cuts are shown in Figures. 13.5, 13.11, and Figure 13.6. The fits are virtually indistinguishable from the data and the values of the constants are listed in the Table 13.4.

$$\begin{aligned} V_{OH}(\phi) &= 0.5 \sum_{n=1}^{4} V_{nOH}(1 - \mathrm{Cos}(n\phi)), \\ V_{CH3}(\theta) &= V_{CH3}^{\phi=0}(0.5)(1 - \mathrm{Cos}(3\theta)), \\ \mathrm{Correction}(\phi) &= 1 + (\sum_{n=1}^{3} V_{nx}(1 - \mathrm{Cos}(n\phi)) \\ V(\theta, \phi) &= V_{CH3}(\mathrm{Correction}(\phi)) \times (0.5)(1 - \mathrm{Cos}(3\theta)) + V_{OH}(\phi) \end{aligned} \tag{13.12}$$

Table 13.4: Constants for the two-dimensional potential for the OH and CH_3 torsion in ethanol.

Constant in Eq. (1)	Value (cm^{-1})
V_{1x}	0.0653
V_{2x}	0.000147
V_{3x}	0.00827
V_{CH3}	1208.4
V_{1OH}	86.3
V_{2OH}	-4.37
V_{3OH}	381.9
V_{4OH}	-32.7

1-D DVR result for the CH_3 potential is shown in Figure 13.13. The moment of inertia for the methyl rotor was taken here to be $10.5/(N_{AV}m_e)$.

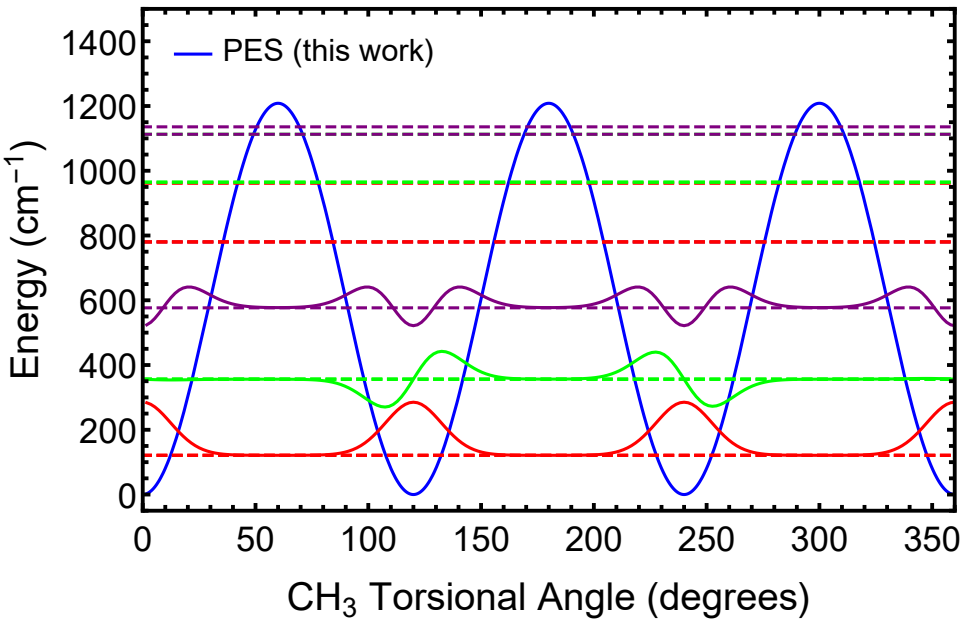


Figure 13.13: 1-D DVR results for the CH_3 torsional potential, whose potential is shown in the blue curve. The energy levels are shown as dotted lines, while the wavefunctions for the lowest three levels are shown as solid red, green, and purple lines.

Given the 2D potential in Eq. (13.12) and the parameters in Table 13.4, we can predict how the OH torsion will vary as a function of the CH_3 torsional angle θ, as shown in Figure 13.14. Not surprisingly, the barriers and the gauche conformation increase in energy as the methyl rotates so that one CH bond eclipses the OH bond. The figure demonstrates substantial interaction between the methyl and OH torsional motions.

The functional form of the 2-D torsional motions just mentioned also made it possible to perform a 2D DVR calculation of the combined energy levels and wavefunctions. Moments of inertia of $2.7/(M_{\mathrm{AV}}m_{\mathrm{e}})$ for the OH rotor and $10.5/(M_{\mathrm{AV}}m_{\mathrm{e}})$ for the CH_3

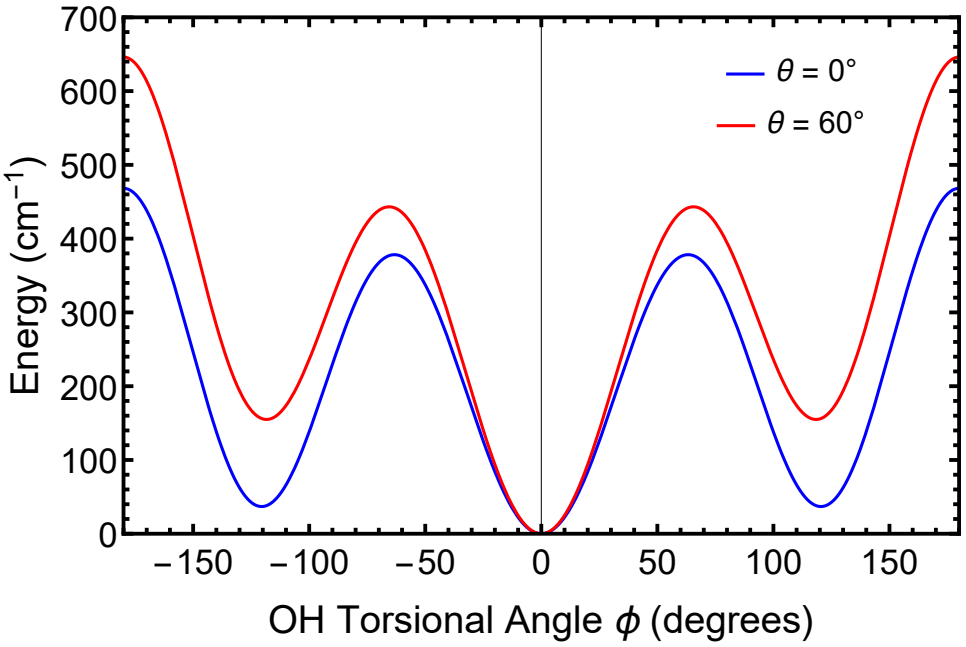


Figure 13.14: OH torsional potential for $\theta = 0$ and $\theta = 60$ degrees, normalized to have the same minimum.

rotor gave the best agreement with the experimental data summarized in PBD.[309] The results are shown in Table 13.5, where the first column gives the observed transitions, the second column gives our transition estimates based on the 2-D model (which was fit to five cuts through the full dimensional PES) and the third column gives the DVR results if instead of the full model potential, we use a separable potential having no cross terms between functions of θ and ϕ. The agreement is good, though certainly not perfect. It should be noted, however, that the 2-D potential is based on unrelaxed cuts and on a fit to 5 cuts of the potential; other cuts could modify the 2-D fit to the full-dimensional surface. There could be adjustments due to either effect. Nonetheless, it is remarkable that the *ab initio* surface is in such reasonable agreement with experiment. Following our calculations, we found that Zheng et al. had recommended moments of inertia for the two rotors based on their electronic structure calculations and the resulting low-lying

energy levels. Their results, converted to atomic units, are $2.66/(M_{\mathrm{AV}}m_{\mathrm{e}})$ for the OH rotor and $9.32/(M_{\mathrm{AV}}m_{\mathrm{e}})$ for the methyl rotor, very close to the values we found to be in best agreement with the experimental results of PBD.

Table 13.5: Comparison of experimental energy levels relative to the lowest level, our 2-D DVR calculations, and our 2-D DVR calculations omitting cross terms in the 2D torsional potential. All energies are in cm^{-1}.

Level	v_{OH}	v_{CH3}	Experiment	full 2-D potential	Omitting cross terms
e_1	0	0	0	0	0
e_1	0	0	39.5	52.3	46.7
o_1	0	0	42.8	54.4	48.9
o_2	1	0	202.6	198.2	196.9
e_2	1	0	238.6	236.	235.1
o_3	1	0	285.9	293.6	289.
e_0	0	1	244.4	251.8	246.5
e_1	0	1	?	299.4(?)	281.9(?)
o_1	0	1	?	301.4(?)	284.9(?)
e_0	0	2	475.5	472.1 477.7	468. 473.
e_1	0	2	529.49	532.	504.4
o_1	0	2	532.8	533.8	524.4

13.5 Summary and conclusions

We presented a new potential energy surface for ethanol at the CCSD(T) level of theory. This was achieved by a Δ-ML method applied to a recent B3LYP-based PES that we previously reported. The new PES was validated for torsional barriers and harmonic frequencies against direct CCSD(T) calculations for the *trans* and *gauche* conformers

and their isomerization TSs. Diffusion Monte Carlo and semiclassical calculations were reported for the zero-point energies of CH_3CH_2OH and several singly deuterated isotopologues. DMC wavefunctions have also been presented.

Our main goal was to investigate the energetics of ethanol which was known to be characterized by two conformers very close in energy. To achieve this goal we needed a way to perform high-level quantum stochastic and dynamical simulations. Therefore our first effort was to construct a "gold-standard" PES of ethanol suitable for quantum calculations that require sampling of the high energy region of the phase space. This is a real need for accurate quantum simulations and not just an exotic requirement. The DMC and SC applications reported demonstrate that not only we achieved our goal, but that the PES is robust for application of methods spanning a large portion of the configuration space.

Our quantum results provided us with a breakthrough in the chemistry of ethanol since we found that the ground state is of *trans* type with a leak to the *gauche* conformer. Indeed, DMC ZPE evaluations return the same value starting from both *trans* and *gauche* geometries. A semiclassical estimate of the first excited state starting from the *gauche* conformer provides a reduced energy difference with respect to the energy gap between conformers found by electronic structure calculations. This is also at odds with harmonic estimates, which anticipate an increased gap. In our view, the "leak" effect and the reduced energy difference eventually explain experimental discrepancies in ethanol investigations and the difficulty to isolate the two conformers even at low temperatures.

We also notice that this result points to a striking resemblance with glycine as discussed in one of our previous works.[143] We found that the 8 identified isomers of glycine reduced to 4 couples of conformers once zero-point energy and nuclear dynamics effects were taken into account. However, the impact of this finding was minor compared to the

one for ethanol because the three main and experimentally investigated conformers of glycine were still energetically well separated. We think these results, and especially those presented for ethanol, provide a new insight on the chemistry of small organic molecules demonstrating the need to take nuclear quantum effects into account.

We employed the new potential to study the motions of the $-CH_3$ and $-OH$ rotors at the quantum mechanical level. DMC and DVR results are in very good agreement and the computed DVR wavefunctions confirm the presence of the "leak" effect. Furthermore, the previously suggested geared motion of the rotors is confirmed by our calculations, and the 2-D model of the torsions based on cuts through the full-dimensional potential provides reasonable energy levels when compared to experiment.

Finally, as a perspective and as anticipated, we notice that semiclassical calculations are able to evaluate the energy of vibrationally excited states and therefore, given the high fidelity of the PES, they will be employed, together with MULTIMODE calculations, in a future work for determining ethanol fundamental frequencies of vibration.

Chapter 14

Semiclassical and Multimode Calculations for the Vibrational Energies of *Trans* and *Gauche* Ethanol

14.1 Chapter Abstract

A full-dimensional, permutationally invariant polynomial potential energy surface based on the Δ-Machine learning approach, recently constructed for ethanol, is here employed for semiclassical and MULTIMODE calculations to determine the anharmonic vibrational spectra for both *trans* and *gauche* conformers of ethanol. Both semiclassical and MULTIMODE estimates agree well the with experimental data. We show that significant mixing between the quantum states and Fermi resonances occur when low frequency bending modes are included during vibrational self-consistent field (VSCF) and virtual-

state configuration interaction (VCI) calculations via MULTIMODE. The same effects are also accurately described by the full-dimensional semiclassical calculations. This is the first time such high-level anharmonic calculations at "gold-standard" CCSD(T) level are reported.

14.2 Overview

From the very early days of quantum mechanics the holy grail of theoretical chemists has been to put quantum mechanics to use as a computational tool that would some day rival the precision of experiment. The foundational, specific goal has been to develop first-principles, i.e., from quantum mechanics, potentials that govern nuclear motion. Progress in doing this for ever-larger molecules has been dramatic in the past 15 or so years.

Developing high-dimensional, *ab initio*-based potential energy surfaces (PESs) remains an active area of theoretical and computational research. Significant progress has been made in the development of machine learning (ML) approaches to generate potential energy surfaces (PESs) for systems with more than five atoms, based on fitting thousands of CCSD(T) energies.[64,65,90,315] The quantum chemical methods capturing a substantial part of electron correlation such as coupled-cluster with singles and doubles (CCSD), coupled-cluster with perturbative triples [CCSD(T)] etc. have a formidable scaling, leading to the requirement of high speed processor, memory, and secondary storage. Thus, there is a bottleneck for developing the PES at high level theory with the increase of molecular size. Due to the steep scaling of the "gold standard" CCSD(T) theory ($\sim N^7$, N being the number of basis functions), it is computationally demanding to build PES of systems with more than 9-10 atoms. (Many researchers do not consider this number of atoms as a "large molecule", however, it is used here as a computational boundary

for the CCSD(T) method.) Therefore, people are bound to use low level electronic structure methods such as density functional theory (DFT) and Møller-Plesset second order perturbation theory (MP2) to generate PESs for large molecules.

The PESs of molecules having more than 10 atoms using CCSD(T) level of theory are generally conspicuous by their absence. One 10-atom PES using the method we are aware of is the formic acid dimer $(HCOOH)_2$,[343] which contains 6 heavy atoms. It was developed by Bowman and co-workers in 2016. This was a major computational effort at the CCSD(T)-F12a/haTZ (VTZ for H and aVTZ for C and O) level of theory, which was a fit to 13475 electronic energies. A 9-atom PES for the chemical reaction $Cl + C_2H_6$ was recently developed by Papp et al. using a composite MP2/CCSD(T) method.[344] Examples of potentials for 6 and 7-atom chemical reactions which are fits to tens of thousands or even hundred thousand CCSD(T) energies have also been reported.[64,90,315–317]

The increasing dimensionality of the PES with the increase in number of atoms requires large training datasets to fit the PES. Thus, given the intense interest, and progress, in moving to larger molecules and clusters, where high-level methods are prohibitively expensive, the use of lower-level methods such as DFT and MP2 is understandable. These methods also provide analytical gradients, and this is an important source of data needed for larger systems. Our group has made use of the permutationally invariant polynomial (PIP) approach for developing PESs of *N*-methyl acetamide,[144,145] glycine[143] and tropolone.[166]

To circumvent this bottleneck, people are applying machine learning (ML) approaches to bring a PES based on a low-level of electronic structure theory (DFT or MP2) to a higher level (CCSD(T)) one. There are two popular methods currently being investigated to achieve this goal. One is the "transfer learning" (TL), and the other is the "Δ-machine learning" (Δ-ML).

TL has been developed extensively in the context of artificial neural networks,[133] and much of the work in that field has been brought into chemistry.[136–138,140,318] The basic idea of TL is that a fit obtained from one source of data (perhaps a large one) can be corrected for a related problem by using limited data and by making hopefully small training alterations to the parameters obtained in the first fit. Therefore, in the present context of PES fitting, an ML-PES fit to low-level electronic energies/gradients can be reused as the starting point of the model for an ML-PES with the accuracy of a high-level electronic structure theory. As noted, this is typically done with artificial neural networks, where weights and biases trained on lower-level data hopefully require minor changes in response to additional training using high-level data. Recently, Meuwly and co-workers applied TL to improve the MP2-based neural network PESs for malonaldehyde, acetoacetaldehyde and acetylacetone using thousands of local CCSD(T) energies .[318]

The other approach is Δ-machine learning. In this approach a correction is made to a property dataset obtained using an efficient, low-level *ab initio* theory such as DFT or MP2.[134,136–138,318] We applied this Δ-machine learning approach to correct a PES based on DFT electronic energies and gradients.[319] Initially, this was successfully done for CH_4 and H_3O^+, and for 12-atom *N*-methyl acetamide.[319] In all cases, the coupled cluster energies were obtained over the same large span of configurations used to get the lower-level DFT PES. For *N*-methyl acetamide these included both the *cis* and *trans* isomers and the saddle points separating them. Later, Δ-machine learning was extensively applied to the 15-atom acetylacetone (AcAc, $CH_3COCH_2COCH_3$) molecule. Not only was a full-dimensional PES at CCSD(T) level developed, but it was successfully applied to compute the quantum zero point energy and ground state wavefunction using diffusion Monte-Carlo (DMC) algorithm as well as to determine the tunneling splitting of H-transfer process.[320]

The focus of most work on TL or Δ-ML learning has been on developing transferable force fields, with applications mainly in the thermochemistry and molecular dynamics simulations at room temperature or somewhat higher. This motivates us to develop a full-dimensional PES of 9-atom ethanol (CH_3CH_2OH) molecule at CCSD(T) level applying the Δ-ML approach.[345]

Ethanol is widely used as a solvent in chemical reactions, and it has great importance in combustion chemistry. Ethanol is the leading biofuel in the transportation sector, where it is mainly used in a form of reformulated gasoline.[300,301] Thus, the study of ethanol chemistry in internal combustion engines is of high interest from scientific, industrial, and environmental perspectives. Ethanol exists as a mixture of *trans* or *anti* and *gauche* (+/-) conformers in both solid, liquid, and gaseous state.[302–304] It is well known that the energy gap between the two conformers is quite small; experimentally it is observed that ΔG is 0.12 (0.02) kcal/mol in favour of the *trans* conformer.[303] Ethanol also has a three-fold methyl torsional potential which makes its potential surface much more complex. These aspects have been investigated when presenting the new PES. Diffusion Monte Carlo (DMC) calculations performed on the new PES have shown that the global minimum is of the *trans* configuration even when starting from the *gauche* geometry. In this work we complete our study of ethanol by examining the fundamental frequencies of vibration of both conformers, which are expected to be influenced by quantum state mixing and Fermi resonances. To accomplish this task we employ full-dimensional semiclassical (SC) calculations and vibrational self-consistent field (VSCF) and virtual-state configuration interaction (VCI) calculations via MULTIMODE.

14.3 Theory and Computational Details

14.3.1 CCSD(T) PES of Ethanol

The full-dimensional CCSD(T) PES of ethanol used here has been recently reported,[345] so we give only a brief summary here. The development of this PES can be divided into two parts – low-level DFT PES (V_{LL}) and a correction PES (ΔV_{CC-LL}). Initially, a low-level DFT PES is developed using the efficient B3LYP/6-311+G(d,p) level of theory, then a correction is made using a sparse set of a relatively small number of *ab initio* CCSD(T) energies to determine the Δ-ML surface using our recently developed approach.[319]

This V_{LL} PES is a permutationally invariant polynomial fit to 8500 energies and their corresponding gradients at B3LYP/6-311+G(d,p) level of theory spanning the energy range of 0 - 35 000 cm^{-1}. For this fit, we used a maximum polynomial order of 4 with permutationally symmetry 321111, leading to a total of 14752 PIP basis functions and linear coefficients whose values were determined by linear least-squares regression. More details of this PES can be found elsewhere.[321]

To develop the correction PES, we train ΔV_{CC-LL} on the difference between the CCSD(T)-F12a/aug-cc-pVDZ and DFT absolute energies for 2069 geometries. A low-order PIP fit was employed because the difference ΔV_{CC-LL} is not as strongly varying as V_{LL} with respect to the nuclear configuration. We used maximum polynomial order of 2 with permutational symmetry 321111 to fit the training dataset which leads to a total of 208 PIP basis. The PIP basis to fit these V_{LL} and ΔV_{CC-LL} PESs were generated using our ?in-house? MSA software.[123,141]

To obtain the CCSD(T) PES we add the correction ΔV_{CC-LL} to the low-level DFT PES, V_{LL}. Investigation of this potential surface located two conformers of ethanol as well as two isomerization saddle points connecting them. Rigorous diffusion Monte Carlo

(DMC) and semiclassical calculations are also performed to compute anharmonic zero-point energies (ZPEs) of these conformers. Details of this PES for the interested reader can be found in ref. 345.

14.3.2 MULTIMODE Calculations

The harmonic vibrational frequencies could be systematically improved by use of vibrational self-consistent field (VSCF) and virtual-state configuration interaction (VCI) methods. These methods have been implemented in our "in-house" software called MULTIMODE. First, we present a brief recap of the VSCF[346,347] and VSCF/VCI scheme[348] in MULTIMODE.[45,46,349] The computational code is based on the rigorous Watson Hamiltonian[47] in mass-scaled normal coordinates, $\boldsymbol{Q}$, for non-linear molecule. This Hamiltonian is represented as

$$\hat{H} = \frac{1}{2}\sum_{\alpha\beta}(\hat{J}_\alpha - \hat{\pi}_\alpha)\mu_{\alpha\beta}(\hat{J}_\beta - \hat{\pi}_\beta) - \frac{1}{2}\sum_{k}^{F}\frac{\partial^2}{\partial Q_k^2} - \frac{1}{8}\sum_{\alpha}\mu_{\alpha\alpha} + V(\boldsymbol{Q}), \tag{14.1}$$

where $\alpha(\beta)$ represent the x, y, z coordinates, $\hat{J}_\alpha$ and $\hat{\pi_\alpha}$ are the components of the total and vibrational angular momenta respectively, $\mu_{\alpha\beta}$ is the inverse of effective moment of inertia tensor, and $V(\boldsymbol{Q})$ is the full potential in terms of normal coordinates. The number of normal modes is denoted by F, and for non-linear molecules F equals $3N-6$. In many applications of this Hamiltonian in the literature, the vibrational angular momentum terms are neglected and this approximation leads to an inaccurate result. Therefore, We include these terms in the MULTIMODE software.

In general there are two major bottlenecks in applications to the VSCF/VCI scheme. One is the numerical evaluation of matrix elements (multi-dimensional integrals) and the second is the size of the H-matrix. Both naively have exponential dependence on the

number of normal coordinates.

As an effective approach to deal with exponential scaling of matrix elements we represent the full potential in a hierarchical n-mode representation (nMR).[349] In normal coordinates, this representation is given by

$$V(Q_1, Q_2, \cdots, Q_F) = \sum_i V_i^{(1)}(Q_i) + \sum_{i,j} V_{ij}^{(2)}(Q_i, Q_j) + \sum_{i,j,k} V_{ijk}^{(3)}(Q_i, Q_j, Q_k) + \sum_{i,j,k,l} V_{ijkl}^{(4)}(Q_i, Q_j, Q_k, Q_l) + \cdots, \tag{14.2}$$

where $V_i^{(1)}(Q_i)$ is the one-mode potential, i.e., the 1D cut through the full-dimensional PES in each mode, one-by-one, $V_{ij}^{(2)}(Q_i, Q_j)$ is the intrinsic 2-mode potential among all pairs of modes, etc. Here, intrinsic means that the any n-mode term is zero if any of the arguments is zero. Also, each term in the representation is in principle of infinite order in the sense of a Taylor series expansion. So for example, $V^{(1)}(Q)$ might look like a full Morse potential.

This representation has been used for nearly twenty years by a number of research groups; a sample of these are refs. 45,46,349–353. It continues to be actively used in a variety of applications and theoretical developments.[354–359] In MULTIMODE the maximum value of n is 6. However, from numerous tests it appears that a 4MR typically gives energies that are converged to within roughly 1–5 cm^{-1} [360–362]. Thus we generally use 4MR with an existing full-dimensional PES and this is also done here.

The second major bottleneck to all VCI calculations is the size of the H-matrix, which as noted already can scale exponentially with the number of vibrational modes. There are many strategies to deal with this. Basically, they all limit the size of the excitation space, with many schemes taken from electronic structure theory. For example,

the excitation space can be limited by using the hierarchical scheme of single, double, triple, etc. excitations. MULTIMODE uses this among other schemes and can consider up to quintuple excitations. A major difference with electronic structure theory is that the nuclear interactions go beyond 2-body. This is immediately clear from the n-mode representation. Thus, MULTIMODE tailors the excitation scheme for each term in this representation. Other schemes to prune the CI basis have been suggested and the reader is referred to reviews[53,352,354,360,363–367] for more details.

We note that the above basic VSCF/VCI scheme with the n-mode representation has been implemented in Molpro by Rauhut and co-workers with the option to obtain the electronic energies directly on n-mode grids, with n up to 4 or from an existing potential.[368] Of course numerous enhancements and modifications to the basic scheme can be found there.

Finally, some comments on the limitations of rectilinear normal modes and thus the Watson Hamiltonian are in order for ethanol, which has low energy torsional modes. These are not expected to be accurately described, especially for excited states which will exhibit large amplitude curvilinear motion. We typically either include these modes, albeit with just a small number of basis functions, or drop then from the calculation. Both of these strategies are examined here. With respect to large amplitude curvilinear motion or torsional modes, we do note that the reaction path version of MULTIMODE[49] is able to describe these. However, because such motion is not the focus of the present work, we do not use this version, as it is also more computationally demanding than the version we adopt here. Thus, the spectra we present are more reliable quantitatively in the high-frequency region than in the low-frequency region.

14.3.3 Semiclassical Theory

An alternative approach we employ to calculate the fundamental frequencies of vibration for the two conformers is represented by the adiabatically switched (AS) semiclassical initial value representation (SCIVR) technique.[324,326] AS SCIVR is a recently developed two-level procedure able to regain quantum effects starting from classical trajectories. Therefore, AS SCIVR is a member of the family of semiclassical methods[327,328] and it features a characteristic way to determine the starting conditions of the dynamics. In AS SCIVR one starts from harmonic quantization and slowly switches on the actual system Hamiltonian. The final molecular geometry and momenta of the adiabatic switching run serve as starting conditions for the subsequent semiclassical dynamics trajectory. This procedure is applied to a distribution of harmonically quantized starting conditions.

The adiabatic switching Hamiltonian is[331–333]

$$H_{\mathrm{as}} = [1 - \lambda(t)]\, H_{\mathrm{harm}} + \lambda(t) H_{\mathrm{anh}}, \tag{14.3}$$

where $\lambda(t)$ is a switching function

$$\lambda(t) = \frac{t}{T_{\mathrm{AS}}} - \frac{1}{2\pi} \sin\left(\frac{2\pi t}{T_{\mathrm{AS}}}\right), \tag{14.4}$$

H_{harm} is the harmonic Hamiltonian built from the harmonic frequencies of vibration, and H_{anh} is the molecular vibrational Hamiltonian. We chose T_{AS} equal to 25000 a.u. (about 0.6 ps), and we employed time steps of 10 a.u. for a total of 4400 trajectories.

Once the adiabatic switching run is over, the trajectories are evolved according to H_{anh} for another 25000 a.u. with same step size to collect the dynamical data needed for the semiclassical calculation. For this purpose we use Kaledin and Miller's time-average

formula

$$I_{as}(E) = \left(\frac{1}{2\pi\hbar}\right)^{N_v} \sum_{i=1}^{N_{traj}} \frac{1}{2\pi\hbar T} \left| \int_0^T dt e^{\frac{i}{\hbar}[S_t(\mathbf{p}_{as},\mathbf{q}_{as})+Et+\phi_t(\mathbf{p}_{as},\mathbf{q}_{as})]} \langle \Psi(\mathbf{p}_{eq},\mathbf{q}_{eq}) | g(\mathbf{p}'_t,\mathbf{q}'_t) \rangle \right|^2 , \tag{14.5}$$

where $I_{as}(E)$ indicates that a vibrational spectral density is calculated as a function of the vibrational energy E. I_{as} is peaked at the eigenvalues of the vibrational Hamiltonian, the lowest one being the ZPE. Frequencies of vibration are obtained by difference between the relevant eigenvalues and the ZPE. In Eq.(14.5) N_v is the number of vibrational degrees of freedom of the system, i.e. 21 in the case of ethanol. T is the total evolution time of the dynamics for the semiclassical part of the simulation. As anticipated, we chose T equal to 25000 a.u. with a time step size of 10 a.u. $(\mathbf{p}'_t, \mathbf{q}'_t)$ is the instantaneous full-dimensional phase space trajectory started at time 0 from the final adiabatic-switching phase space condition $(\mathbf{p}_{as}, \mathbf{q}_{as})$. S_t is the classical action along the semiclassical trajectory, and ϕ_t is the phase of the Herman-Kluk pre-exponential factor based on the elements of the stability matrix and defined as

$$\phi_t = \text{phase}\left[\sqrt{\left|\frac{1}{2}\left(\frac{\partial \mathbf{q}'_t}{\partial \mathbf{q}_{as}} + \Gamma^{-1}\frac{\partial \mathbf{p}'_t}{\partial \mathbf{p}_{as}}\Gamma - i\hbar\frac{\partial \mathbf{q}'_t}{\partial \mathbf{p}_{as}}\Gamma + \frac{i\Gamma^{-1}}{\hbar}\frac{\partial \mathbf{p}'_t}{\partial \mathbf{q}_{as}}\right)\right|}\right] , \tag{14.6}$$

where Γ is an $N_v \times N_v$ matrix usually chosen to be diagonal with elements numerically equal to the harmonic frequencies. Based on Liouville's theorem, the stability (or monodromy) matrix has the property to have its determinant equal to 1 along the entire trajectory. However, classical chaotic dynamics can lead to numerical inaccuracies in the propagation, so, following a common procedure in semiclassical calculations, we have rejected the trajectories based on a 1% tolerance threshold on the monodromy matrix

determinant value. Finally, the working formula is completed by the quantum mechanical overlap between a quantum reference state $|\Psi\rangle$ and a coherent state $|g\rangle$ with the following representation in configuration space

$$\langle \mathbf{q}|g(\mathbf{p}'_t,\mathbf{q}'_t)\rangle = \left(\frac{\det(\Gamma)}{\pi^{N_\nu}}\right)\exp\left\{-(\mathbf{q}-\mathbf{q}'_t)^T\frac{\Gamma}{2}(\mathbf{q}-\mathbf{q}'_t)+\frac{i}{\hbar}\mathbf{p}'^T_t(\mathbf{q}-\mathbf{q}'_t)\right\}. \tag{14.7}$$

The reference state $|\Psi\rangle$ is usually chosen to be itself a coherent state. In Eq. (14.5) $|\Psi\rangle$ is written as $|\Psi(\mathbf{p}_{eq},\mathbf{q}_{eq})\rangle$, where $\mathbf{p}_{eq}$ stands for the linear momenta obtained in harmonic approximation setting the geometry at the equilibrium one ($\mathbf{q}_{eq}$).

14.4 Results and Discussion

Calculations are performed using Version 5.1.4 of MULTIMODE.[46,346,349] During all the calculations, a four-mode representation of the potential in mass-scaled normal coordinates and a three-mode representation of the effective inverse moment of inertia for the vibrational angular momentum terms in the exact Watson Hamiltonian are used.[47]. The formalism is based on CI from the virtual space of the ground vibrational state VSCF Hamiltonian. Here we explore reduced-mode models, i.e 11-mode and 15-mode models, with maximum mode excitation of 10 10 10 8, which means that the first three modes can be excited up to a maximum of 10 quanta and the fourth mode can be excited up to a maximum of 8 quanta. This leads to CI matrix sizes of 45486 and 155026 for 11-mode, and 15-mode calculations, respectively. We compute 200 CI vibrational states up to the energy of 4000 cm^{-1}.

Tables 14.1 and 14.2 show the comparison of Multimode and semiclassical AS-SCIVR anharmonic frequencies with the corresponding experimental and harmonic ones. It is

Table 14.1: Vibrational frequencies for trans-ethanol. Harmonic, Multimode (4MR), AS-SCIVR, and experimental IR gas values from Ref. 314. Experimental values in parenthesis refer to the Raman gas experiment of Ref. 369.

Mode	Harmonic	Multimode (4MR, 11 Modes)	Multimode (4MR, 15 Modes)	AS SCIVR	Experiment
7	1115	****	1101	1088	1090
8	1181	****	1160	1148	1166
9	1284	****	1251	1242	1241
10	1302	****	1276	1271	1275
11	1403	1367	1370	1363	1367
12	1454	1427	1426	1420	1450 (1430)
13	1488	1444	1443	1440	1455
14	1500	1459	1459	1456	1480 (1460)
15	1530	1491	1489	1481	1500 (1460)
16	2995	2810[a]	2811[a]	2881	2888
17	3029	2864[a]	2884[a]	2888	2902
18	3037	2939[a]	2937[a]	2933	2922
19	3120	2976[a]	2977[a]	2983	2987
20	3126	2978[a]	2981[a]	2986	2992
21	3862	3679	3672	3676	3676

[a] These states are mixed and not pure ones. See text for details.

Table 14.2: Vibrational frequencies for gauche-ethanol. Harmonic, Multimode (4MR), AS-SCIVR, and experimental IR gas values from Ref. 314. Experimental values in parenthesis refer to the Raman gas experiment of Ref. 369.

Mode	Harmonic	Multimode (4MR, 11 Modes)	Multimode (4MR, 15 Modes)	AS SCIVR	Experiment
7	1094	****	1085	1060	1066
8	1144	****	1124	1115	1117
9	1290	****	1259	1247	1249
10	1375	****	1337	1332	1342
11	1406	1370	1370	1369	1373
12	1424	1393	1393	1386	1394 (1430)
13	1490	1448	1447	1448	1460
14	1496	1454	1454	1451	1465 (1460)
15	1519	1480	1480	1475	1493 (1460)
16	3007	2867[a]	2880[a]	2885	2912
17	3020	2874[a]	2923[a]	2909	2936
18	3089	2927[a]	2931[a] (2950)[a]	2957	2972
19	3108	2958[a]	2959[a] (2979)[a]	2977	2987
20	3121	2976[a]	2978[a]	2985	2994
21	3845	3659	3653	3655	3662

[a] These states are mixed ones. See text and Table 4 for details.

seen the excellent agreement between Multimode and AS-SCIVR frequencies as well as experimental ones. We see substantial improvement moving from 11-mode to 15-mode Multimode calculations specially for the high frequency modes which are highly coupled. However, we see a big difference for mode 16 of *trans*-Ethanol in which the MULTIMODE calculation is differed by 70 and 77 cm^{-1} with respect to the AS-SCIVR frequency and the experimental frequency, respectively.

Table 14.3: Three largest VSCF/VCI expansion coefficients of indicated energies (cm^{-1}) frequencies for *trans*-Ethanol.

trans-**Ethanol**			
Energy	Mode	State	Coeff.
	bend-12	$2\nu_{12}$	0.6051
2811	stretch-16	ν_{16}	-0.5347
	bend-11+bend-12	$\nu_{11} + \nu_{12}$	-0.4571
	stretch-17	ν_{17}	-0.8634
2884	twist-10 + wag-12	$\nu_{10} + \nu_{12}$	0.2818
	rock-8 + bend-15	$\nu_{8} + \nu_{15}$	-0.2263
	stretch-18	ν_{18}	0.6277
2937	bend-14	$2\nu_{14}$	-0.5442
	bend-13	$2\nu_{13}$	0.3019
	stretch-19	ν_{19}	0.8309
2977	bend-14 + bend-15	$\nu_{14} + \nu_{15}$	-0.3612
	bend-14	$2\nu_{14}$	-0.2379
	stretch-20	ν_{20}	0.9237
2981	bend-13 + bend-14	$\nu_{13} + \nu_{14}$	0.1871
	wag-12 + bend-13	$\nu_{12} + \nu_{13}$	-0.1319

The most exciting finding during these MULTIMODE calculations are the presence of

Table 14.4: Three largest absolute magnitude coupling coefficients of Multimode frequencies for *gauche*-Ethanol.

gauche-**Ethanol**			
Multimode Frequency	Coupling Modes	Quanta	Coeff.
	stretch-16	ν_{16}	0.5852
2880	bend-12 + bend-14	$\nu_{12} + \nu_{14}$	-0.4138
	bend-13	$2\nu_{13}$	0.2968
	bend-14	$2\nu_{14}$	-0.5771
2923	stretch-17	ν_{17}	-0.4987
	bend-13	$2\nu_{13}$	-0.3752
	stretch-18	ν_{18}	-0.5134
2931	bend-13 + bend-15	$\nu_{13} + \nu_{15}$	0.4367
	bend-14 + bend-15	$\nu_{14} + \nu_{15}$	0.4249
	stretch-18	ν_{18}	0.5068
2950	bend-15	$2\nu_{15}$	-0.4341
	bend-14 + bend-15	$\nu_{14} + \nu_{15}$	0.4097
	bend-15	$2\nu_{15}$	-0.6244
2959	stretch-19	ν_{19}	-0.4643
	stretch-18	ν_{18}	-0.4046
	stretch-20	ν_{20}	0.8087
2978	bend-14 + bend-15	$\nu_{14} + \nu_{15}$	0.3754
	stretch-19	ν_{19}	-0.2759
	bend-15	$2\nu_{15}$	0.5336
2979	stretch-19	ν_{19}	-0.5294
	bend-14 + bend-15	$\nu_{14} + \nu_{15}$	0.3885

mixed states and Fermi resonances. From the Tables. 14.3 and 14.4, it is clearly seen that the high frequency modes (mode 16-20) are highly coupled for both the isomers. Most of the coupling are mainly occurred between C-H stretching modes and low frequency bending, twisting, and rocking modes and their coefficients are big which reveal that the mixing are significant. Due to this strong mixing, the anharmonic frequencies are lowered by 100-150 cm^{-1}.

The presence of mixed states and Fermi resonances found with Multimode are confirmed by the AS-SCIVR simulations. There are two aspects that make AS SCIVR efficient in this task. One is that trajectories are fully dimensional and this allows one to take into account couplings between all modes. Secondly, the presence of coherent states in Eq. 14.5, which have a Gaussian shape and therefore a tail in phase space, allows one to correctly collect quantum eigenenergies even if the energy of run trajectories is not perfectly tailored for the state under investigation. A good example of this is given by the AS-SCIVR estimate of mode 16 of the *trans* conformer, which is very close to the experimental value even if the mode is involved in a Fermi resonance.

14.5 Summary and Conclusions

Recently developed a full-dimensional CCSD(T) PES of ethanol has been employed to compute the anharmonic vibrational frequencies of both *trans* and *gauche*-ethanol. We found good agreement between MULTIMODE and semi-classical AS SCIVR calculations as well as the previously reported experimental results as shown in Tables 14.1 and 14.2. In addition, we also observed significant mixing between the vibrational states and Fermi resonances when low frequency bending modes are include during VCI calculations.

References

[1] Schatz, G. C. *Rev. Mod. Phys.* **1989**, *61*, 669.

[2] Ischtwan, J.; Collins, M. A. *J. Chem. Phys.* **1994**, *100*, 8080.

[3] Thompson, K. C.; Jordan, M. J. T.; Collins, M. A. *J. Chem. Phys.* **1998**, *108*, 8302.

[4] Hollebeek, T.; Ho, T.-S.; Rabitz, H. *Annu. Rev. Phys. Chem.* **1999**, *50*, 537.

[5] Bettens, R. P. A.; Collins, M. A. *J. Chem. Phys.* **1999**, *111*, 816.

[6] Majumder, M.; Ndengue, S. A.; Dawes, R. *Mol. Phys.* **2016**, *114*, 1.

[7] Marquardt, R.; Quack, M. In *Handbook of High-resolution Spectroscopy*; Quack, M., Merkt, F., Eds.; John Wiley & Sons, Ltd., 2011; p 511.

[8] Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. *J. Chem. Phys.* **1995**, *103*, 4129.

[9] Brown, D. F. R.; Gibbs, M. N.; Clary, D. C. *J. Chem. Phys.* **1996**, *10*, 7597.

[10] Raff, L. M.; Malshe, M.; Hagan, M.; Doughan, D. I.; Rockley, M. G.; Komanduri, R. *J. Chem. Phys.* **2005**, *122*, 084104.

[11] Manzhos, S.; Wang, X.; Dawes, R.; Carrington, T. *J. Phys. Chem. A* **2006**, *110*, 5295.

[12] Manzhos, S.; Carrington, T. *J. Chem. Phys.* **2006**, *125*, 084109.

[13] Handley, C. M.; Popelier, P. L. A. *J. Phys. Chem. A* **2010**, *114*, 3371–3383.

[14] Behler, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930–17955.

[15] Behler, J. *J. Chem. Phys.* **2011**, *134*, 074106.

[16] Behler, J. *Int. J. Quantum Chem.* **2015**, *115*, 1032–1050.

[17] Chen, J.; Xu, X.; Xu, X.; Zhang, D. H. *J. Chem. Phys.* **2013**, *138*, 154301.

[18] Gastegger, M.; Marquetand, P. *J. Chem. Theory Comput.* **2015**, *11*, 2187.

[19] Manzhos, S.; Dawes, R.; Carrington, T. *Int. J. Quantum Chem.* **2015**, *115*, 1012–1020.

[20] Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. *Phys. Rev. Lett.* **2010**, *104*, 136403.

[21] Bartók, A. P.; Kondor, R.; Csányi, G. *Phys. Rev. B* **2013**, *87*, 184115.

[22] Bartók, A. P.; Csányi, G. *Int. J. Quantum Chem.* **2015**, *115*, 1051–1057.

[23] Uteva, E.; Graham, R. S.; Wilkinson, R. D.; Wheatley, R. J. *J. Chem. Phys.* **2017**, *147*, 161706.

[24] Kolb, B.; Marshall, P.; Zhao, B.; Jiang, B.; Guo, H. *J. Phys. Chem. A* **2017**, *121*, 2552–2557.

[25] Kamath, A.; Vargas-Hernández, R. A.; Krems, R.; Carrington, T.; Manzhos, S. *J. Chem. Phys.* **2018**, *148*, 241702.

[26] Braams, B. J.; Bowman, J. M. *Int. Rev. Phys. Chem.* **2009**, *28*, 577.

[27] Xie, Z.; Bowman, J. M. *J. Chem. Theory Comput.* **2010**, *6*, 26–34.

[28] Bowman, J. M.; Braams, B. J.; Carter, S.; C., C.; Czakó, G.; Fu, B.; Huang, X.; Kamarchik, E.; Sharma, A. R.; C., S. B.; Wang, Y.; Xie, Z. *J. Phys. Chem. Lett.* **2010**, *1*, 1866.

[29] Bowman, J. M.; Czakó, G.; Fu, B. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8094.

[30] Jiang, B.; Guo, H. *J. Chem. Phys.* **2013**, *139*, 054112.

[31] Li, J.; Jiang, B.; Guo, H. *J. Chem. Phys.* **2013**, *139*, 204103.

[32] Jiang, B.; Guo, H. *J. Chem. Phys.* **2014**, *141*, 034109.

[33] Shao, K.; Chen, J.; Zhao, Z.; Zhang, D. H. *J. Chem. Phys.* **2016**, *145*, 071101.

[34] Jiang, B.; Li, J.; Guo, H. *Int. Rev. Phys. Chem.* **2016**, *35*, 479–506.

[35] Qu, C.; Yu, Q.; Van Hoozen, B. L.; Bowman, J. M.; Vargas-Hernández, R. A. *J. Chem. Theory Comput.* **2018**, *14*, 3381–3396.

[36] Derksen, H.; Kemper, G. *Computational Invariant Theory*; Springer Verlag: Berlin, Heidelberg, New York, 2002.

[37] Bosma, W.; Cannon, J.; Playoust, C. *J. Symb. Comp.* **1997**, *24*, 235.

[38] Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2011**, *134*, 094509.

[39] Adler, T. B.; Knizia, G.; Werner, H.-J. *J. Chem. Phys.* **2007**, *127*, 221106.

[40] Knizia, G.; Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2009**, *130*, 054104.

[41] Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

[42] Nielson, H. H. *Rev. Mod. Phys.* **1951**, *23*, 90.

[43] Barone, V. *J. Chem. Phys.* **2005**, *122*, 014108.

[44] Christoffel, K. M.; Bowman, J. M. *Chem. Phys. Lett.* **1982**, *85*, 220.

[45] Carter, S.; Bowman, J. M.; Hancy, N. C. *Theor. Chem. Acc.* **1998**, *100*, 191.

[46] Bowman, J. M.; Carter, S.; Huang, X. *Int. Rev. Phys. Chem.* **2003**, *22*, 533.

[47] Watson, J. K. G. *Mol. Phys.* **1968**, *15*, 479.

[48] Miller, H., William; Handy, N. C.; Adams, J. E. *J. Chem. Phys.* **1980**, *72*, 99.

[49] Bowman, J. M.; Huang, X.; Handy, N. C.; Carter, S. *J. Phys. Chem. A* **2007**, *111*, 7317.

[50] Carter, S.; Handy, N. C. *J. Chem. Phys.* **2000**, *113*, 987.

[51] Tew, D. P.; Handy, N. C.; Carter, S.; Irle, S.; Bowman, J. M. *Mol. Phys.* **2003**, *101*, 3513.

[52] Handy, N. C.; Carter, S. *Mol. Phys.* **2004**, *102*, 2201.

[53] Wang, X.; Carter, S.; Bowman, J. M. *J. Phys. Chem. A* **2015**, 11632.

[54] Burcl, R.; Carter, S.; Handy, N. C. *Chem. Phys. Lett.* **2003**, *380*, 237.

[55] Anderson, J. B. *J. Chem. Phys.* **1975**, *63*, 1499.

[56] Anderson, J. B. *J. Chem. Phys.* **1976**, *65*, 4121.

[57] Suhm, M. A.; Watts, R. O. *Phys. Rep.* **1991**, *204*, 293–329.

[58] Kosztin, I.; Faber, B.; Schulten, K. *Am. J. Phys.* **1996**, *64*, 633.

[59] McCoy, A. B. *Int. Rev. Phys. Chem.* **2006**, *25*, 77–107.

[60] Hase, W. L. In *Encyclopeida of Computational Chemistry*; Allinger, N. L., Ed.; Wiley: New York, 1998; p 402.

[61] Hase, W. L. In *Encyclopeida of Computational Chemistry*; Allinger, N. L., Ed.; Wiley: New York, 1998; p 399.

[62] Collins, M. A. *Theor. Chem. Acc.* **2002**, *108*, 313–324.

[63] Shao, K.; Chen, J.; Zhao, Z.; Zhang, D. H. *J. Chem. Phys.* **2016**, *145*, 071101.

[64] Fu, B.; Zhang, D. H. *J. Chem. Theory Comput.* **2018**, *14*, 2289–2303.

[65] Qu, C.; Yu, Q.; Bowman, J. M. *Annu. Rev. Phys. Chem.* **2018**, *69*, 6.1–6.25.

[66] Doughan, D. I.; Raff, L. M.; Rockley, M. G.; Hagan, M.; Agrawal, P. M.; Komanduri, R. *J. Chem. Phys.* **2006**, *124*, 054321.

[67] Manzhos, S.; Carrington, T. *J. Chem. Phys.* **2006**, *125*, 084109.

[68] Pukrittayakamee, A.; Malshe, M.; Hagan, M.; Raff, L. M.; Narulkar, R.; Bukkapatnum, S.; Komanduri, R. *J. Chem. Phys.* **2009**, *130*, 134101.

[69] Nongnuch, A.; Behler, J. *Phys. Rev. B* **2012**, *85*, 045439.

[70] Guan, Y.; Yang, S.; Zhang, D. H. *J. Phys. Chem. A* **2018**, *122*, 3140–3147.

[71] Fu, B.; Han, Y.-C.; Bowman, J. M.; Angelucci, L.; Balucani, N.; Leonori, F.; Casavecchia, P. *Proc. Natl. Acad. Sci. U.S.A* **2012**, *109*, 9733–9738.

[72] Qu, C.; Bowman, J. M. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24835.

[73] Qu, C.; Bowman, J. M. *J. Chem. Phys.* **2018**, *148*, 241713.

[74] Brown, A.; Braams, B. J.; Christoffel, K.; Jin, Z.; Bowman, J. M. *J. Chem. Phys.* **2003**, *119*, 8790–8793.

[75] Xie, Z.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2005**, *122*, 224307.

[76] Huang, X.; Braams, B. J.; Bowman, J. M.; Kelly, R. E. A.; Tennyson, J.; Groenenboom, G. C.; van der Avoird, A. *J. Chem. Phys.* **2008**, *128*, 034312.

[77] Shank, A.; Wang, Y. M.; Kaledin, A.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2009**, *130*, 144314.

[78] Huang, X.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2005**, *122*, 044308.

[79] Heazlewood, B. R.; Jordan, M. J. T.; Kable, S. H.; Selby, T. M.; Osborn, D. L.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *Proc. Natl. Acad. Sci. U.S.A* **2008**, *105*, 12719–12724.

[80] Dey, A.; Fernando, R.; Abeysekera, C.; Homayoon, Z.; Bowman, J.; G Suits, A. *J. Chem. Phys.* **2014**, *140*, 054305.

[81] Kidwell, N. M.; Li, H.; Wang, X.; Bowman, J. M.; Lester, M. I. *Nat. Chem.* **2016**, *8*, 509–514.

[82] Nguyen-Truong, H. T.; Le, H. M. *J. Phys. Chem. A* **2012**, *116*, 4629–4638.

[83] Nguyen-Truong, H. T.; Le, H. M. *Chem. Phys. Lett.* **2015**, *629*, 40 – 45.

[84] Wang, X.; Houston, P. L.; Bowman, J. M. *Philos. Trans. Royal Soc. A* **2017**, *375*, 2092–2100.

[85] Yang, B.; Balakrishnan, N.; Zhang, P.; Wang, X.; Bowman, J. M.; Forrey, R. C.; Stancil, P. C. *J. Chem. Phys.* **2016**, *145*, 034308.

[86] Wang, X.; Bowman, J. M. *J. Chem. Theory Comput.* **2013**, *9*, 901–908.

[87] Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. MOLPRO, version 2015.1, a package of ab initio programs. 2015; see http://www.molpro.net.

[88] Qu, C.; Bowman, J. M. *J. Phys. Chem. A* **2016**, *120*, 4988–4993.

[89] Yurchenko, S. N.; Tennyson, J.; Barber, R. J.; Thiel, W. *J. Mol. Spectrosc.* **2013**, *291*, 69 – 76.

[90] Bowman, J. M.; Czakó, G.; Fu, B. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8094–8111.

[91] Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Phys. Chem. A* **2008**, *112*, 9344–9351.

[92] Kidwell, N. M.; Li, H.; Wang, X.; Bowman, J. M.; Lester, M. I. *Nat. Chem.* **2016**, *8*, 509–514.

[93] Wang, Y.; Braams, B. J.; Bowman, J. M.; Carter, S.; Tew, D. P. *J. Chem. Phys.* **2008**, *128*, 224314.

[94] Li, J.; Song, K.; Behler, J. *Phys. Chem. Chem. Phys.* **2019**, *21*, 9672–9682.

[95] Li, J.; Guo, H. *J. Chem. Phys.* **2015**, *143*, 221103.

[96] Qu, C.; Bowman, J. M. *J. Chem. Phys.* **2019**, *150*, 141101.

[97] Jasper, A. W.; Harding, L. B.; Knight, C.; Georgievskii, Y. *J. Phys. Chem. A* **2019**, *123*, 6210–6228.

[98] Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. *J. Chem. Phys.* **2018**, *148*, 241722.

[99] Unke, O. T.; Meuwly, M. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693.

[100] Paesani, F. *Acc. Chem. Res.* **2016**, *49*, 1844–1851.

[101] Mirkin, N. G.; Krimm, S. *J. Am. Chem. Soc.* **1991**, *113*, 9742–9747.

[102] Chen, X. G.; Schweitzer-Stenner, R.; Asher, S. A.; Mirkin, N. G.; Krimm, S. *J. Phys. Chem.* **1995**, *99*, 3074–3083.

[103] Gaigeot, M.-P.; Martinez, M.; Vuilleumier, R. *Mol. Phys.* **2007**, *105*, 2857–2878.

[104] šebek, J.; Kejík, Z.; Bouř, P. *J. Phys. Chem. A* **2006**, *110*, 4702–4711.

[105] Hirst, J. D.; Hirst, D. M.; Brooks, C. L. *J. Phys. Chem. A* **1997**, *101*, 4821–4827.

[106] Jorgensen, W. L.; Gao, J. *J. Am. Chem. Soc.* **1988**, *110*, 4212–4216.

[107] Radzicka, A.; Pedersen, L.; Wolfenden, R. *Biochemistry* **1988**, *27*, 4538–4541.

[108] Luque, F. J.; Orozco, M. *J. Org. Chem.* **1993**, *58*, 6397–6405.

[109] Song, S.; Asher, S. A.; Krimm, S.; Shaw, K. D. *J. Am. Chem. Soc.* **1991**, *113*, 1155–1163.

[110] Han, W.-G.; Suhai, S. *J. Phys. Chem.* **1996**, *100*, 3942–3949.

[111] Mantz, Y. A.; Branduardi, D.; Bussi, G.; Parrinello, M. *J. Phys. Chem. B* **2009**, *113*, 12521–12529.

[112] Ataka, S.; Takeuchi, H.; Tasumi, M. *J. Mol. Struct.* **1984**, *113*, 147–160.

[113] Mirkin, N. G.; Krimm, S. *J. Mol. Struct.* **1995**, *334*, 1–6.

[114] Markham, L. M.; Hudson, B. S. *J. Phys. Chem.* **1996**, *100*, 2731–2737.

[115] Kang, Y. K. *J. Phys. Chem. B* **2000**, *104*, 8321–8326.

[116] Torii, H.; Tatsumi, T.; Kanazawa, T.; Tasumi, M. *J. Phys. Chem. B* **1998**, *102*, 309–314.

[117] Kang, Y. K. *J. Mol. Struct.* **2001**, *546*, 183–193.

[118] Lu, D.; Li, J. *J. Chem. Phys* **2016**, *145*, 014303.

[119] Nandi, A.; Qu, C.; Bowman, J. M. *J. Chem. Theory Comput.* **2019**, *15*, 2826–2835.

[120] Li, S.; Li, W.; Ma, J. *Acc. Chem. Res.* **2014**, *47*, 2712–2720.

[121] He, X.; Zhu, T.; Wang, X.; Liu, J.; Zhang, J. Z. H. *Acc. Chem. Res.* **2014**, *47*, 2748–2757.

[122] Collins, M. A.; Cvitkovic, M. W.; Bettens, R. P. A. *Acc. Chem. Res.* **2014**, *47*, 2776–2785.

[123] MSA Software with Gradients. `https://github.com/szquchen/MSA-2.0`, Accessed: 2019-01-20.

[124] Kosztin, I.; Faber, B.; Schulten, K. *Am. J. Phys.* **1996**, *64*, 633–644.

[125] Hamilton, I. P.; Light, J. C.; Whaley, K. B. *J. Chem. Phys.* **1986**, *85*, 5151–5157.

[126] Wu, Q.; Zhang, J. Z. *Chem. Phys. Letts* **1996**, *252*, 195 – 200.

[127] Skokov, S.; Peterson, K. A.; Bowman, J. M. *Chem. Phys. Letts* **1999**, *312*, 494 – 502.

[128] Gazdy, B.; Bowman, J. M. *J. Chem. Phys.* **1991**, *95*, 6309–6316.

[129] Bowman, J. M.; Gazdy, B. *J. Chem. Phys.* **1991**, *94*, 816–817.

[130] Meuwly, M.; Hutson, J. M. *J. Chem. Phys.* **1999**, *110*, 8338–8347.

[131] van Mourik, T.; Harris, G. J.; Polyansky, O. L.; Tennyson, J.; Csszr, A. G.; Knowles, P. J. *J. Chem. Phys.* **2001**, *115*, 3706–3718.

[132] Bowman, J. M.; Gazdy, B.; Bentley, J. A.; Lee, T. J.; Dateo, C. E. *J. Chem. Phys.* **1993**, *99*, 308–323.

[133] Pan, S. J.; Yang, Q. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359.

[134] Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.

[135] Zaspel, P.; Huang, B.; Harbrecht, H.; von Lilienfeld, O. A. *J. Chem. Theory and Comput.* **2019**, *15*, 1546–1559.

[136] Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Mller, K.-R.; Tkatchenko, A. *J. Chem. Phys.* **2019**, *150*, 114102.

[137] Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. *Nat. Commun.* **2018**, *9*, 3887.

[138] Stöhr, M.; Medrano Sandonas, L.; Tkatchenko, A. *J. Phys. Chem. Letts.* **2020**, *11*, 6835–6843.

[139] Käser, S.; Unke, O.; Meuwly, M. *New Journal of Physics* **2020**, *22*, 055002.

[140] Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. *Nat. Commun.* **2019**, *10*, 2903–2906.

[141] Nandi, A.; Qu, C.; Bowman, J. M. *J. Chem. Theor. Comp.* **2019**, *15*.

[142] Conte, R.; Qu, C.; Houston, P. L.; Bowman, J. M. *J. Chem. Theory Comput.* **2020**, *16*, 3264–3272.

[143] Conte, R.; Houston, P. L.; Qu, C.; Li, J.; Bowman, J. M. *J. Chem. Phys.* **2020**, *153*, 244301.

[144] Qu, C.; Bowman, J. M. *J. Chem. Phys.* **2019**, *150*, 141101.

[145] Nandi, A.; Qu, C.; Bowman, J. M. *J. Chem. Phys.* **2019**, *151*, 084306.

[146] Houston, P. L.; Conte, R.; Qu, C.; Bowman, J. M. *J. Chem. Phys.* **2020**, *153*, 024107:1–11.

[147] Qu, C.; Conte, R.; Houston, P. L.; Bowman, J. M. *Phys. Chem. Chem. Phys.* **2020**, –.

[148] Dral, P. O.; Owens, A.; Dral, A.; Csnyi, G. *J. Chem. Phys.* **2020**, *152*, 204110.

[149] Yu, Q.; Bowman, J. M. *J. Chem. Theory Comput.* **2016**, *12*, 1549.

[150] Suenram, R.; Golubiatnikov, G.; Leonov, I.; Hougen, J.; Ortigoso, J.; Kleiner, I.; Fraser, G. *J. Molec. Spec.* **2001**, *208*, 188 – 193.

[151] Cui, J.; Krems, R. V. *J. Phys. B* **2016**, *49*, 224001.

[152] Jiang, B.; Li, J.; Guo, H. *Int. Rev. Phys. Chem.* **2016**, *35*, 479–506.

[153] Qu, C.; Bowman, J. M. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24835–24840.

[154] Papp, D.; Tajti, V.; Győri, T.; Czakó, G. *J. Phys. Chem. Lett.* **2020**, *11*, 4762–4767.

[155] Fu, Y.-L.; Lu, X.; Han, Y.-C.; Fu, B.; Zhang, D. H.; Bowman, J. M. *Chem. Sci.* **2020**, *11*, 2148–2154.

[156] Lu, D.; Behler, J.; Li, J. *J. Phys. Chem. A* **2020**, *124*, 5737–5745.

[157] Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. *J. Phys. Chem. Lett.* **2021**, *12*, 4902–4909.

[158] Collins, M. A.; Bettens, R. P. A. *Chem. Rev.* **2015**, *115*, 5607–5642, PMID: 25843427.

[159] Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. *J. Phys. Chem.* **1994**, *98*, 9165–9169.

[160] Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109.

[161] Furtado, J. P.; Rahalkar, A. P.; Shanker, S.; Bandyopadhyay, P.; Gadre, S. R. *J. Phys. Chem. Lett.* **2012**, *3*, 2253–2258.

[162] Sahu, N.; Gadre, S. R. *Acc. Chem. Res.* **2014**, *47*, 2739–2747, PMID: 24798296.

[163] Khire, S. S.; Bartolotti, L. J.; Gadre, S. R. *J. Chem. Phys.* **2018**, *149*, 064112.

[164] Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. *J. Comput. Chem.* **2010**, *31*, 2405–2418.

[165] Qu, C.; Conte, R.; Houston, P. L.; Bowman, J. M. *Phys. Chem. Chem. Phys.* **2021**, *23*, 7758–7767.

[166] Houston, P. L.; Conte, R.; Qu, C.; Bowman, J. M. *J. Chem. Phys.* **2020**, *153*, 024107.

[167] Frisch, M. J. et al. Gaussian 09, revision C.01. 2010.

[168] Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. *J. Chem. Phys.* **2021**, *154*, 051102.

[169] Howard, D. L.; Kjaergaard, H. G.; Huang, J.; Meuwly, M. *J. Phys. Chem. A* **2015**, *119*, 7980–7990.

[170] Cisneros, G. A.; Wikfeldt, K. T.; Ojame, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartk, A. P.; Csnyi, G.; Molinero, V.; Paesani, F. *Chem. Rev.* **2016**, *116*, 7501–7528.

[171] Detrich, J.; Corongiu, G.; Clementi, E. *Chem. Phys. Lett.* **1984**, *112*, 426–430.

[172] Lybrand, T. P.; Kollman, P. A. *J. Chem. Phys.* **1985**, *83*, 2923–2933.

[173] Gregory, J. K.; Clary, D. C. *J. Phys. Chem.* **1996**, *100*, 18014–18022.

[174] Xantheas, S. S. *Chem. Phys.* **2000**, *258*, 225–231.

[175] Gra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. *J. Chem. Phys.* **2011**, *135*, 224102.

[176] Medders, G. R.; Götz, A. W.; Morales, M. A.; Bajaj, P.; Paesani, F. *J. Chem. Phys.* **2015**, *143*, 104102.

[177] Heindel, J. P.; Xantheas, S. S. *J. Chem. Theory Comput.* **2020**, *16*, 6843–6855.

[178] Wang, Y. M.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2009**, *131*, 054511.

[179] Gora, U.; Cencek, W.; Podeszwa, R.; van der Avoird, A.; Szalewicz, K. *J. Chem. Phys.* **2014**, *140*, 194101.

[180] Partridge, H.; Schwenke, D. W. *J. Chem. Phys.* **1997**, *106*, 4618.

[181] Braams, B. J.; Bowman, J. M. *Int. Rev. Phys. Chem.* **2009**, *28*, 577–606.

[182] Huang, X. C.; Braams, B. J.; Bowman, J. M. *J. Phys. Chem. A* **2006**, *110*, 445–451.

[183] Shank, A.; Wang, Y.; Kaledin, A.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2009**, *130*, 144314.

[184] Babin, V.; Leforestier, C.; Paesani, F. *J. Chem. Theory Comput.* **2013**, *9*, 5395–5403.

[185] Burnham, C. J.; Anick, D. J.; Mankoo, P. K.; Reiter, G. F. *J. Chem. Phys.* **2008**, *128*, 154519.

[186] Fanourgakis, G. S.; Xantheas, S. S. *J. Chem. Phys.* **2008**, *128*, 074506.

[187] Wang, Y. M.; Babin, V.; Bowman, J. M.; Paesani, F. *J. Am. Chem. Soc.* **2012**, *134*, 11116–11119.

[188] Mallory, J. D.; Mandelshtam, V. A. *J. Chem. Phys.* **2016**, *145*, 064308.

[189] Heindel, J. P.; Yu, Q.; Bowman, J. M.; Xantheas, S. S. *J. Chem. Theory. Comput.* **2018**, *14*, 4553–4566.

[190] Kumar, R.; Skinner, J. L. *J. Phys. Chem. B* **2008**, *112*, 8311–8318.

[191] Tainter, C. J.; Pieniazek, P. A.; Lin, Y. S.; Skinner, J. L. *J. Chem. Phys.* **2011**, *134*, 184501.

[192] Moberg, D. R.; Jasper, A. W.; Davis, M. J. *J. Phys. Chem. Lett.* **2021**, *12*, 9169–9174.

[193] Paukku, Y.; Yang, K. R.; Varga, Z.; Truhlar, D. G. *J. Chem. Phys.* **2013**, *139*, 044309.

[194] Conte, R.; Houston, P. L.; Bowman, J. M. *J. Chem. Phys.* **2014**, *140*, 151101.

[195] Conte, R.; Qu, C.; Bowman, J. M. *J. Chem. Theory Comput.* **2015**, *11*, 1631–1638.

[196] Howard, J. C.; Tschumper, G. S. *J. Chem. Theory Comput.* **2015**, *11*, 2126–2136.

[197] Homayoon, Z.; Conte, R.; Qu, C.; Bowman, J. M. *J. Chem. Phys.* **2015**, *143*, 084302.

[198] Qu, C.; Conte, R.; Houston, P. L.; Bowman, J. M. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8172–8181.

[199] Babin, V.; Medders, G. R.; Paesani, F. *J. Chem. Theory Comput.* **2014**, *10*, 1599–1607.

[200] Qu, C.; Bowman, J. M. *J. Phys. Chem. A* **2019**, *123*, 329–335.

[201] Karplus, M.; Porter, R. N.; Sharma, R. D. *J. Chem. Phys.* **1965**, *43*, 3259.

[202] Wang, X.; Bowman, J. M. *J. Chem. Theory Comput.* **2013**, *9*, 901–908.

[203] Hase, W. L. In *Encyclopeida of Computational Chemistry*; Allinger, N. L., Ed.; Wiley: New York, 1998; Vol. 1; pp 399–407.

[204] Czakó, G.; Bowman, J. M. *Science* **2011**, *334*, 343–346.

[205] Czakó, G.; Bowman, J. M. *J. Phys. Chem. A* **2014**, *118*, 2839.

[206] Czakó, G.; Kaledin, A. L.; Bowman, J. M. *J. Chem. Phys.* **2010**, *132*, 164103.

[207] Brieuc, F.; Bronstein, Y.; Dammak, H.; Depondt, P.; Finocchi, F.; Hayoun, M. *J. Chem. Theory Comput.* **2016**, *12*, 5688–5697.

[208] Bowman, J. M.; Gazdy, B.; Sun, Q. Y. *J. Chem. Phys.* **1989**, *91*, 2859–2862.

[209] Miller, W. H.; Hase, W. L.; Darling, C. L. *J. Chem. Phys.* **1989**, *91*, 2863–2868.

[210] Xie, Z.; Bowman, J. M. *J. Phys. Chem. A* **2006**, *110*, 5446–5449.

[211] Bonhommeau, D.; Truhlar, D. G. *J. Chem. Phys.* **2008**, *129*, 014302.

[212] Czakó, G.; Kaledin, A. L.; Bowman, J. M. *Chem. Phys. Lett.* **2010**, *500*, 217–222.

[213] Noid, D. W.; Koszykowski, M. L.; Marcus, R. A. *J. Chem. Phys.* **1980**, *73*, 391–395.

[214] De Leon, N.; Heller, E. J. *J. Chem. Phys.* **1984**, *81*, 5957.

[215] Miller, W. H. *J. Chem. Phys.* **1984**, *81*, 3573.

[216] Eaker, C. W.; Schatz, G. C.; De Leon, N.; Heller, E. J. *J. Chem. Phys.* **1984**, *81*, 5913.

[217] Martens, C. C.; Ezra, G. S. *J. Chem. Phys.* **1985**, *83*, 2990.

[218] Ezra, G. S.; Martens, C. C.; Fried, L. E. *J. Phys. Chem.* **1987**, *91*, 3721.

[219] Solov'ev, E. A. *Sov. Phys. JETP* **1978**, *48*, 635.

[220] Johnson, B. R. *J. Chem. Phys.* **1985**, *83*, 1204.

[221] Skodje, R. T.; Borondo, F.; Reinhardt, W. P. *J. Chem. Phys.* **1985**, *82*, 4611.

[222] Skodje, R. T.; Borondo, F. *J. Chem. Phys.* **1986**, *84*, 1533.

[223] Johnson, B. R. *J. Chem. Phys.* **1987**, *86*, 1445.

[224] Zakrzewski, J.; Saini, S.; Taylor, H. S. *Phys. Rev. A* **1988**, *38*, 3877.

[225] Saini, S.; Zakrzewski, J.; Taylor, H. S. *Phys. Rev. A* **1988**, *38*, 3900.

[226] Sun, Q.; Bowman, J. M.; Gazdy, B. *J. Chem. Phys.* **1988**, *89*, 3124.

[227] Qu, C.; Bowman, J. M. *J. Phys. Chem. A* **2016**, *120*, 4988–4993.

[228] Nagy, T.; Vikár, A.; Lendvay, G. *J. Chem. Phys.* **2016**, *144*, 014104.

[229] Bose, A.; Makri, N. *J. Chem. Phys.* **2015**, *143*, 114114.

[230] Taatjes, C. A.; Welz, O.; Eskola, A. J.; Savee, J. D.; Scheer, A. M.; Shallcross, D. E.; Rotavera, B.; Lee, E. P.; Dyke, J. M.; Mok, D. K.; Osborn, D. L.; Percival, C. J. *Science* **2013**, *340*, 177–180.

[231] Beames, J. M.; Liu, F.; Lu, L.; Lester, M. I. *J. Chem. Phys.* **2013**, *138*, 244307.

[232] Liu, F.; Beames, J. M.; Petit, A. S.; McCoy, A. B.; Lester, M. I. *Science* **2014**, *345*, 1596–1598.

[233] Kidwell, N. M.; Li, H.; Wang, X.; Bowman, J. M.; Lester, M. I. *Nat. Chem.* **2016**, *8*, 509–514.

[234] Wang, X.; Bowman, J. M. *J. Phys. Chem. Letts.* **2016**, *7*, 3359–3364.

[235] Landau, L.; Lifshitz, E. M. *Mechanics*, 3rd ed.; Pergamon: Oxford, 1976.

[236] Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; McGraw-Hill: New York, 1955.

[237] Dymarsky, A. Y.; Kudin, K. N. *J. Chem. Phys.* **2005**, *122*, 124103.

[238] Czakó, G. *J. Phys. Chem. A* **2012**, *116*, 7467.

[239] Lin, H.-Y.; Huang, Y.-H.; Wang, X.; Bowman, J. M.; Nishimura, Y.; Witek, H. A.; Lee, Y.-P. *Nat. Commun.* **2015**, *6*, 7012.

[240] Bowman, J. M.; Wang, X.; Homayoon, Z. *J. Molec. Spect.* **2015**, *311*, 2–11.

[241] Wang, Y.; Bowman, J. M. *J. Chem. Phys.* **2008**, *129*, 121103.

[242] Wang, Y.; Bowman, J. M. *J. Chem. Phys.* **2013**, *139*, 154303.

[243] Jiang, B.; Guo, H. *J. Chem. Phys.* **2013**, *138*, 234104.

[244] Makri, N.; Miller, W. H. *J. Chem. Phys.* **1989**, *91*, 4026.

[245] Light, J. C. *J. Chem. Phys.* **1964**, *40*, 3221–3229.

[246] Pechukas, P.; Light, J. C.; Rankin, C. *J. Chem. Phys.* **1966**, *44*, 794–805.

[247] Roueff, E.; Gerin, M. *Space Sci. Rev.* **2003**, *106*, 61–72.

[248] Lis, D. C.; Roueff, E.; Gerin, M.; Phillips, T. G.; Coudert, L. H.; van der Tak, F. F. S.; Schilke, P. *Astrophys. J. Lett.* **2002**, *571*, L55.

[249] van der Tak, F. F. S.; Schilke, P.; Müller, H. S. P.; Lis, D. C.; Phillips, T. G.; Gerin, M.; Roueff, E. *Astron. Astrophys.* **2002**, *388*, L53–L56.

[250] Parise, B.; Castets, A.; Herbst, E.; Caux, E.; Ceccarelli, C.; Mukhopadhyay, I.; Tielens, A. G. G. M. *Astron. Astrophys.* **2004**, *416*, 159–163.

[251] Mauersberger, R.; Henkel, C.; Jacq, T.; Walmsley, C. M. *Astron. Astrophys.* **1988**, *194*, L1–L4.

[252] Jacq, T.; Walmsley, C. M.; Mauersberger, R.; Anderson, T.; Herbst, E.; De Lucia, F. C. *Astron. Astrophys.* **1993**, *271*, 276.

[253] Parise, B.; Ceccarelli, C.; Tielens, A. G. G. M.; Herbst, E.; Lefloch, B.; Caux, E.; Castets, A.; Mukhopadhyay, I.; Pagani, L.; Loinard, L. *Astron. Astrophys.* **2002**, *393*, L49–L53.

[254] Parise, B.; Ceccarelli, C.; Tielens, A. G. G. M.; Castets, A.; Caux, E.; Lefloch, B.; Maret, S. *Astron. Astrophys.* **2006**, *453*, 949–958.

[255] Whittet, D. C. B.; Cook, A. M.; Herbst, E.; Chiar, J. E.; Shenoy, S. S. *Astrophys. J.* **2011**, *742*, 28.

[256] Charnley, S. B.; Tielens, A. G. G. M.; Rodgers, S. D. *Astrophys. J.* **1997**, *482*, L203.

[257] Xu, L. H.; Hougen, J. T. *J. Mol. Spectrosc.* **1995**, *169*, 396–409.

[258] Sibert, E. L.; Castillo-Chará, J. *J. Chem. Phys.* **2005**, *122*, 194306.

[259] Serrallach, A.; Meyer, R.; Günthard, H. *J. Mol. Spectrosc.* **1974**, *52*, 94.

[260] Boyarkin, O. V.; Lubich, L.; Settle, R. D. F.; Perry, D. S.; Rizzo, T. R. *J. Chem. Phys.* **1997**, *107*, 8409–8422.

[261] Xu, L.-H.; Wang, X.; Cronin, T. J.; Perry, D. S.; Fraser, G. T.; Pine, A. S. *J. Mol. Spectrosc.* **1997**, *185*, 158–172.

[262] Rueda, D.; Boyarkin, O. V.; Rizzo, T. R.; Chirokolava, A.; Perry, D. S. *J. Chem. Phys.* **2005**, *122*, 044314.

[263] Miani, A.; Hänninen, V.; Horn, M.; Halonen, L. *Mol. Phys.* **2000**, *98*, 1737–1748.

[264] Xu, L.-H. *J. Chem. Phys.* **2000**, *113*, 3980.

[265] Hänninen, V.; Halonen, L. *Mol. Phys.* **2003**, *101*, 2907–2916.

[266] Fehrensen, B.; Luckhaus, D.; Quack, M.; Willeke, M.; Rizzo, T. R. *J. Chem. Phys.* **2003**, *119*, 5534–5544.

[267] Yu, H.-G.; Muckerman, J. T. *J. Phys. Chem. A* **2004**, *108*, 8615–8623.

[268] Carter, S.; Handy, N. C.; Bowman, J. M. *Mol. Phys.* **2009**, *107*, 727–737.

[269] Ruscic, B.; Boggs, J. E.; Burcat, A.; Császár, A. G.; Demaison, J.; Janoschek, R.; Martin, J. M. L.; Morton, M. L.; Rossi, M. J.; Stanton, J. F.; Szalay, P. G.; Westmoreland, P. R.; Zabel, F.; Bérces, T. *J. Phys. Chem. Ref. Data* **2005**, *34*, 573.

[270] Bowman, J. M. *Science* **2000**, *290*, 724–725.

[271] Zou, S.; Bowman, J. M.; Brown, A. *J. Chem. Phys.* **2003**, *118*, 10012–10023.

[272] DeVine, J. A.; Weichman, M. L.; Laws, B.; Chang, J.; Babin, M. C.; Balerdi, G.; Xie, C.; Malbon, C. L.; Lineberger, W. C.; Yarkony, D. R.; Field, R. W.; Gibson, S. T.; Ma, J.; Guo, H.; Neumark, D. M. *Science* **2017**, *358*, 336–339.

[273] Lau, J. A.; Choudhury, A.; Chen, L.; Schwarzer, D.; Verma, V. B.; Wodtke, A. M. *Science* **2020**, *367*, 175–178.

[274] Chen, L.; Lau, J. A.; Schwarzer, D.; Meyer, J.; Verma, V. B.; Wodtke, A. M. *Science* **2019**, *363*, 158–161.

[275] Vogt, J.; Vogt, B. *J. Chem. Phys.* **2014**, *141*, 214708.

[276] Chen, J.; Li, J.; Bowman, J. M.; Guo, H. *J. Chem. Phys.* **2020**, *153*, 054310.

[277] Hoang, P. N. M.; Picaud, S.; Girardet, C.; Meredith, A. W. *J. Chem. Phys.* **1996**, *105*, 8453–8462.

[278] Meredith, A. W.; Stone, A. J. *J. Chem. Phys.* **1996**, *104*, 3058–3070.

[279] Corcelli, S. A.; Tully, J. C. *J. Chem. Phys.* **2002**, *116*, 8079–8092.

[280] Boney, E. T. D.; Marcus, R. A. *J. Chem. Phys.* **2013**, *139*, 184712.

[281] Boese, A. D.; Saalfrank, P. *J. Phys. Chem. C* **2016**, *120*, 12637–12653.

[282] Chen, J.; Hariharan, S.; Meyer, J.; Guo, H. *J. Phys. Chem. C* **2020**, *124*, 19146–19156.

[283] Lu, P.-F.; Yan, L.; Yu, Z.-Y.; Gao, Y.-F.; Gao, T. *Communications in Theoretical Physics* **2013**, *59*, 193–198.

[284] Qu, C.; Conte, R.; Houston, P. L.; Bowman, J. M. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8172–8181.

[285] Nandi, A.; Qu, C.; Bowman, J. M. *J. Chem. Theory Comput.* **2019**, *15*, 2826–2835.

[286] Bowman, J. M.; Gazdy, B. *Journal of Chemical Physics* **1991**, *94*, 816–817.

[287] Knizia, G.; Adler, T. B.; Werner, H.-J. *Journal of Chemical Physics* **2009**, *130*, 054104.

[288] Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, T. H.; Wilson, A. K. *Theoretical Chemistry Accounts* **2011**, *128*, 69.

[289] Jensen, F. *Journal of Chemical Physics* **2002**, *116*, 7372–7379.

[290] Jensen, F. *Journal of Physical Chemistry A* **2007**, *111*, 11198–11204.

[291] Townsend, D.; Lahankar, S. A.; Lee, S. K.; Chambreau, S. D.; Suits, A. G.; Zhang, X.; Rheinecker, J.; Harding, L. B.; Bowman, J. M. *Science* **2004**, *306*, 1158–1161.

[292] Suits, A. G. *Accounts of Chemical Research* **2008**, *41*, 873–881.

[293] Bowman, J. M.; Shepler, B. C. *Annual Review of Physical Chemistry* **2011**, *62*, 531–553.

[294] Colbert, D. T.; Miller, W. H. *The Journal of Chemical Physics* **1992**, *96*, 1982–1991.

[295] Homayoon, Z.; Bowman, J. M. *Journal of Physical Chemistry A* **2013**, *117*, 11665–11672.

[296] Langhoff, S. R.; Bauschlicher, C. W. *Journal of Chemical Physics* **1995**, *102*, 5220–5225.

[297] Zhu, R.; Lin, M. *Chemical Physics Letters* **2009**, *478*, 11 – 16.

[298] Dey, A.; Fernando, R.; Abeysekera, C.; Homayoon, Z.; Bowman, J. M.; Suits, A. G. *Journal of Chemical Physics* **2014**, *140*, 054305.

[299] Wodtke, A. M.; Hintsa, E. J.; Lee, Y. T. *Journal of Physical Chemistry* **1986**, *90*, 3549–3558.

[300] Sarathy, S. M.; Owald, P.; Hansen, N.; Kohse-Hinghaus, K. *Prog. Energy Combust. Sci.* **2014**, *44*, 40–102.

[301] Barraza-Botet, C. L.; Wagnon, S. W.; Wooldridge, M. S. *J. Phys. Chem. A.* **2016**, *120*, 7408–7418.

[302] Jönsson, P. G. *Acta Cryst. B.* **1976**, *32*, 232–235.

[303] Kakar, R. K.; Quade, C. R. *J. Chem. Phys.* **1980**, *72*, 4300–4307.

[304] Chen, L.; Zhu, W.; Lin, K.; Hu, N.; Yu, Y.; Zhou, X.; Yuan, L.-F.; Hu, S.-M.; Luo, Y. *J. Phys. Chem. A.* **2015**, *119*, 3209–3217.

[305] Durig, J.; Larsen, R. *J. Mol. Struct.* **1990**, *238*, 195–222.

[306] Coussan, S.; Bouteiller, Y.; Perchard, J. P.; Zheng, W. Q. *J. Phys. Chem. A.* **1998**, *102*, 5789–5793.

[307] Pitsevich, G. A.; Doroshenko, I. Y.; Pogorelov, V. Y.; Sablinskas, V.; Balevicius, V. *Low Temp. Phys.* **2013**, *39*, 389–400.

[308] Agrawal, S. K.; Pal, D.; Chakraborty, A.; Chakraborty, S. *Chem. Phys.* **2020**, *537*, 110851.

[309] Pearson, J. C.; Brauer, C. S.; Drouin, B. J. *J. Mol. Spectrosc.* **2008**, *251*, 394–409.

[310] Zheng, J.; Yu, T.; Papajak, E.; Alecu, I. M.; Mielke, S. L.; Truhlar, D. G. *Physical Chemistry Chemical Physics* **2011**, *13*, 10885–10907.

[311] Dyczmons, V. *J. Phys. Chem. A.* **2004**, *108*, 2080–2086.

[312] Kirschner, K. N.; Heiden, W.; Reith, D. *ACS Omega* **2018**, *3*, 419–432.

[313] Kahn, K.; Bruice, T. C. *ChemPhysChem* **2005**, *6*, 487–495.

[314] Katsyuba, S. A.; Gerasimova, T. P.; Spicher, S.; Bohle, F.; Grimme, S. *J. Comp. Chem.* **2022**, *43*, 279–288.

[315] Jiang, B.; Li, J.; Guo, H. *J. Phys. Chem. Lett.* **2020**, *11*, 5120–5131.

[316] Fu, Y.-L.; Lu, X.; Han, Y.-C.; Fu, B.; Zhang, D. H.; Bowman, J. M. *Chem. Sci.* **2020**, *11*, 2148–2154.

[317] Lu, D.; Behler, J.; Li, J. *J. Phys. Chem. A* **2020**, *124*, 5737–5745.

[318] Käser, S.; Unke, O.; Meuwly, M. *New J. Phys.* **2020**, *22*, 055002.

[319] Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. *J. Chem. Phys.* **2021**, *154*, 051102.

[320] Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. *J. Phys. Chem. Lett.* **2021**, *12*, 4902–4909.

[321] Houston, P. L.; Qu, C.; Nandi, A.; Conte, R.; Yu, Q.; Bowman, J. M. *J. Chem. Phys.* **2022**, *156*, 044120.

[322] Su, C. F.; Quade, C. R. *J. Mol. Spectrosc.* **1998**, *188*, 1–8.

[323] Suenram, R. D.; Lovas, F. J.; Quade, C. R.; Su, C. F. *J. Mol. Spectrosc.* **1998**, *188*, 9–13.

[324] Conte, R.; Parma, L.; Aieta, C.; Rognoni, A.; Ceotto, M. *J. Chem. Phys.* **2019**, *151*, 214107.

[325] Botti, G.; Ceotto, M.; Conte, R. *J. Chem. Phys.* **2021**, *155*, 234102.

[326] Botti, G.; Aieta, C.; Conte, R. *J. Chem. Phys.* **2022**, *156*, 164303.

[327] Miller, W. H. *J. Phys. Chem. A* **2001**, *105*, 2942–2955.

[328] Huber, D.; Heller, E. J. *J. Chem. Phys.* **1987**, *87*, 5302.

[329] Aieta, C.; Micciarelli, M.; Bertaina, G.; Ceotto, M. *Nat. Comm.* **2020**, *11*, 4384.

[330] Aieta, C.; Bertaina, G.; Micciarelli, M.; Ceotto, M. *J. Chem. Phys.* **2020**, *153*, 214117.

[331] Sun, Q.; Bowman, J. M.; Gazdy, B. *J. Chem. Phys.* **1988**, *89*, 3124–3130.

[332] Saini, S.; Zakrzewski, J.; Taylor, H. S. *Phys. Rev. A* **1988**, *38*, 3900–3908.

[333] Nagy, T.; Lendvay, G. *J. Phys. Chem. Lett.* **2017**, *8*, 4621–4626.

[334] Brewer, M. L.; Hulme, J. S.; Manolopoulos, D. E. *J. Chem. Phys.* **1997**, *106*, 4832–4839.

[335] Kaledin, A. L.; Miller, W. H. *J. Chem. Phys.* **2003**, *118*, 7174–7182.

[336] Kaledin, A. L.; Miller, W. H. *J. Chem. Phys.* **2003**, *119*, 3078–3084.

[337] Conte, R.; Gabas, F.; Botti, G.; Zhuang, Y.; Ceotto, M. *J. Chem. Phys.* **2019**, *150*, 244118.

[338] Gandolfi, M.; Rognoni, A.; Aieta, C.; Conte, R.; Ceotto, M. *J. Chem. Phys.* **2020**, *153*, 204104.

[339] Rognoni, A.; Conte, R.; Ceotto, M. *Chem. Sci.* **2021**, *12*, 2060–2064.

[340] Quade, C. *J. Mol. Spectrosc.* **2000**, *203*, 200–202.

[341] Pearson, J. C.; Sastry, K. V. L. N.; Winnewisser, M.; Herbst, E.; De Lucia, F. C. *J. Phys. Chem. Ref. Data* **1995**, *24*, 1–32.

[342] Colbert, D. T.; Miller, W. H. *J. Chem. Phys.* **1992**, *96*, 1982–1991.

[343] Qu, C.; Bowman, J. M. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24835–24840.

[344] Papp, D.; Tajti, V.; Gyri, T.; Czak, G. *J. Phys. Chem. Lett.* **2020**, *11*, 4762–4767.

[345] Nandi, A.; Conte, R.; Qu, C.; Houston, P. L.; Yu, Q.; Bowman, J. M. Quantum calculations on a CCSD(T) machine-learned PES reveal the leaky nature of gas-phase trans and gauche ethanol conformers. 2022; `https://arxiv.org/abs/2206.02297`.

[346] Bowman, J. M. *J. Chem. Phys.* **1978**, *68*, 608–610.

[347] Bowman, J. M. *Acc. Chem. Res.* **1986**, *19*, 202–208.

[348] Christoffel, K.; Bowman, J. **1982**, *85*, 220–224.

[349] Carter, S.; Culik, S. J.; Bowman, J. M. *J. Chem. Phys.* **1997**, *107*, 10458–10469.

[350] Ostrowski, L.; Ziegler, B.; Rauhut, G. *J. Chem. Phys.* **2016**, *145*, 104103:1–9.

[351] Ziegler, B.; Rauhut, G. *J. Chem. Phys.* **2016**, *144*, 114114:1–11.

[352] Christiansen, O. *Phys. Chem. Chem. Phys.* **2012**, *14*, 6672–6687.

[353] König, C.; Christiansen, O. **2015**, *142*, 144115:1–19.

[354] Schrder, B.; Rauhut, G. *J. Chem. Phys.* **2021**, *154*, 124114:1–12.

[355] Dinu, D. F.; Ziegler, B.; Podewitz, M.; Liedl, K. R.; Loerting, T.; Grothe, H.; Rauhut, G. *J. Mol. Spectrosc.* **2020**, *367*, 111224:1–11.

[356] Erfort, S.; Tschöpe, M.; Rauhut, G. *J. Chem. Phys.* **2020**, *152*, 244104:1–14.

[357] Schmitz, G.; Artiukhin, D. G.; Christiansen, O. *J. Chem. Phys.* **2019**, *150*, 131102:1–7.

[358] Madsen, N. K.; Jensen, R. B.; Christiansen, O. *J. Chem. Phys.* **2021**, *154*, 054113:1–14.

[359] Moitra, T.; Madsen, D.; Christiansen, O.; Coriani, S. *J. Chem. Phys.* **2020**, *153*, 234111:1–18.

[360] Bowman, J. M.; Carrington, T.; Meyer, H.-D. *Mol. Phys.* **2008**, *106*, 2145–2182.

[361] Carter, S.; Bowman, J. M.; Handy, N. C. *Mol. Phys.* **2012**, *110*, 775–781.

[362] Carter, S.; Sharma, A. R.; Bowman, J. M. **2012**, *137*, 154301:1–19.

[363] Roy, T. K.; Gerber, R. B. *Phys. Chem. Chem. Phys.* **2013**, *15*, 9468–9492.

[364] Oschetzki, D.; Rauhut, G. *Phys. Chem. Chem. Phys.* **2014**, *16*, 16426–16435.

[365] Császár, A. G.; Fabri, C.; Szidarovszky, T.; Matyus, E.; Furtenbacher, T.; Czakó, G. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1085–1106.

[366] Tennyson, J. *J. Chem. Phys.* **2016**, *145*, 120901:1–8.

[367] Carrington, T. *J. Chem. Phys.* **2017**, *146*, 120902:1–10.

[368] Werner, H.-J. et al. *J. Chem. Phys.* **2020**, *152*, 144107:1–24.

[369] Durig, J. R.; Deeb, H.; Darkhalil, I. D.; Klaassen, J. J.; Gounev, T. K.; Ganguly, A. *J. Mol. Spectrosc.* **2011**, *985*, 202–210.