

Path Integral Monte Carlo simulations of solid parahydrogen using many-body interaction potentials

by

Alexander Ibrahim

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2024

© Alexander Ibrahim 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Takamasa Momose
Professor, Dept. of Chemistry
University of British Columbia

Supervisor: Pierre-Nicholas Roy
Professor, Dept. of Chemistry
University of Waterloo

Committee Members: Jeff Chen
Professor, Dept. of Physics and Astronomy
University of Waterloo

Marcel Nooijen
Professor, Dept. of Chemistry
University of Waterloo

Internal-External Member: Scott Hopkins
Professor, Dept. of Chemistry
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Aside from chapter 1 and chapter 6, this thesis is composed of published manuscripts to which I have contributed substantially, under the supervision of Dr. Pierre-Nicholas Roy. In the following, we describe the publication from which the thesis chapter is sourced, as well as the contributions from my coauthors.

- Chapter 2: This chapter is reproduced from Alexander Ibrahim, Lecheng Wang, Tom Halverson, Robert James LeRoy, and Pierre-Nicholas Roy, “Equation of state and first principles prediction of the vibrational matrix shift of solid parahydrogen,” *Journal of Chemical Physics* **151**, 244501 (2019). Pierre-Nicholas Roy engaged in fruitful discussions and suggested changes to the study and the text of the manuscript. Pierre-Nicholas Roy, Lecheng Wang, Tom Halverson, and Robert James LeRoy contributed to the conceptualization and methodology of the work.
- Chapter 3: This chapter is reproduced from Alexander Ibrahim and Pierre-Nicholas Roy, “Three-body potential energy surface for parahydrogen” *Journal of Chemical Physics* **156**, 044301 (2022). Pierre-Nicholas Roy engaged in fruitful discussions and suggested changes to the study and the text of the manuscript.
- Chapter 4: This chapter is reproduced from Alexander Ibrahim and Pierre-Nicholas Roy, “Equation of state of solid parahydrogen using *ab initio* two-body and three-body interaction potentials” *Journal of Chemical Physics* **157**, 174503 (2022). Pierre-Nicholas Roy engaged in fruitful discussions and suggested changes to the study and the text of the manuscript.
- Chapter 5: This chapter is reproduced from Alexander Ibrahim and Pierre-Nicholas Roy, “A neural network-based four-body potential energy surface for parahydrogen” *Journal of Chemical Physics* **160**, 244308 (2024). Pierre-Nicholas Roy engaged in fruitful discussions and suggested changes to the study and the text of the manuscript.

Abstract

We construct *ab initio* many-body potential energy surfaces (PES) and use them to perform high-accuracy path integral Monte Carlo (PIMC) simulations of solid parahydrogen (*para*-H₂).

We first perform PIMC simulations of solid *para*-H₂ using the Faruk-Schmidt-Hinde (FSH) potential, an *ab initio* 1D two-body PES for *para*-H₂ constructed by the Roy group in 2015. The simulations are successful at reproducing experimental results for the equilibrium density and for the vibrational matrix shift for solid *para*-H₂. However, we find that the two-body PES on its own is too energetically repulsive at higher densities, and greatly overestimates the pressure as a function of density. To improve the accuracy of our simulations, we must include higher-order many-body interactions, such as the three-body and four-body interactions.

We then construct an isotropic *ab initio* three-body PES for *para*-H₂. The energies are calculated using the coupled cluster method with singles, doubles, and perturbative triples excitations (CCSD(T)). The calculations are performed using an AVTZ atom-centred basis set, with additional (*3s3p2d*) midbond functions. We use a machine learning method called the reproducing kernel Hilbert space (RKHS) method to construct a PES from the *ab initio* energies. The three-body PES is attractive at short distances. We perform PIMC simulations of solid *para*-H₂ using both the two-body FSH potential and the new three-body PES. The inclusion of the three-body PES improves the agreement with experiment at lower densities. However, at higher densities, the attractive interaction of the three-body PES overcorrects for the repulsive wall of the two-body PES, resulting in a severe underestimation of the pressure-density curve.

Next, we construct an isotropic *ab initio* four-body PES for *para*-H₂. The energies are calculated using the CCSD(T) method, using an AVDZ atom-centred basis set with additional (*3s3p2d*) midbond functions. We use a multilayer perceptron (MLP) to construct a PES from the *ab initio* energies. We find that the four-body PES is repulsive at short distances. We anticipate that the inclusion of the four-body PES alongside the aforementioned two-body and three-body PESs will improve the agreement of the PIMC simulations of solid *para*-H₂ with experiment to higher densities than previously found.

Acknowledgements

I would like to begin by thanking my supervisor, Dr. Pierre-Nicholas Roy. This thesis could not have been written without his support, guidance, and encouragement. His incredible energy, excitement, advice and constant motivation have been invaluable throughout my degree.

I would also like to thank the other members of my advisory committee, Dr. Marcel Nooijen, Dr. Jeff Chen, and Dr. Michel Gingras. Their constructive feedback and useful advice have greatly improved the quality of my research.

I am grateful for the constant support of the other members of the Theoretical Chemistry group at the University of Waterloo. Their constant help and encouragement have been vital to my work. In no particular order, I would like to thank Matthew Schmidt, Dmitri Iouchtchenko, Kevin Bishop, Adam Marr, Xichen Lou, Lindsay Orr, Neil Raymond, Spencer Yim, Jianying Sheng, Tom Halverson, Tapas Sahoo, Mike Lecours, Hui Li, Mark Zanon, Songhao Bao, Nithin Aaron, Shaeer Moeed, Jacob Chaussé, Megan Dawson, Tobias Serwatka, Estevao De Oliveira, and Wenxue Zhang.

Finally, I would like to thank my parents for their endless support and encouragement throughout the past six years.

Dedication

To mom and dad.

Table of Contents

Examining Committee	ii
Author's Declaration	iii
Statement of Contributions	iv
Abstract	v
Acknowledgements	vi
Dedication	vii
List of Abbreviations	xiii
1 Introduction	1
1.1 Parahydrogen	1
1.1.1 Molecular parahydrogen	1
1.1.2 Solid parahydrogen	2
1.2 Many-body interactions	3
1.2.1 What are many-body interactions?	3
1.2.2 Calculating many-body interaction energies	4
1.2.3 General properties of many-body interaction energies for parahydrogen	5

1.3	The time-independent Schrödinger equation	6
1.4	Electronic structure theory	7
1.4.1	Born-Oppenheimer Approximation	7
1.4.2	Electronic structure methods	8
1.4.3	Basis sets	9
1.4.4	The Counterpoise correction	9
1.5	Potential energy surfaces	10
1.6	Path integral Monte Carlo	12
1.7	Machine learning	15
1.7.1	Multilayer perceptrons	16
1.7.2	Benefits and drawbacks of a multilayer perceptron PES	17
1.8	Energy Unit Conversions	17
1.9	Overview of the thesis	18
2	Equation of state and first principles prediction of the vibrational matrix shift of solid parahydrogen	19
2.1	Introduction	19
2.2	Theory	21
2.2.1	Hamiltonian, pairwise approximation, and symmetry	21
2.2.2	Path Integral Monte Carlo method	24
2.3	Numerics	26
2.3.1	Extrapolating to an infinite number of time steps	26
2.3.2	Extrapolating to an infinite number of particles	28
2.4	Results and Discussion	30
2.4.1	The global EOS	30
2.4.2	EOS near equilibrium density	32
2.4.3	Pressure as a function of density	34
2.4.4	Compressibility of solid parahydrogen	36
2.4.5	Vibrational band origin shifts	37
2.5	Conclusion	42

3	Three-body potential energy surface for parahydrogen	43
3.1	Introduction	43
3.2	Creating the Potential Energy Surface	46
3.2.1	Coordinate System of the Potential Energy Surface	46
3.2.2	Electronic Structure Calculations	49
3.2.3	Constructing the Potential Energy Surface	50
3.3	Discussion and Results	59
3.3.1	Basis Set Comparison	59
3.3.2	Lebedev Quadrature Comparison	61
3.3.3	Effects of Data Mesh Spacing on Potential Energy Surface	62
3.3.4	HCP Lattice Analysis	65
3.3.5	Comparison Against Silvera-Goldman Potential	68
3.4	Conclusion	71
4	Equation of state of solid parahydrogen using <i>ab initio</i> two-body and three-body interaction potentials	72
4.1	Introduction	72
4.2	Model	75
4.2.1	Full Inclusion Method	76
4.2.2	Perturbative Method	77
4.3	Computation	78
4.3.1	Path Integral Monte Carlo	78
4.3.2	Trotter Error	78
4.3.3	Finite System Size Error	79
4.4	Results	83
4.4.1	Energy-Density Equation of State	83
4.4.2	Pressure-Density Equation of State	88
4.4.3	Centroid Distribution	90
4.4.4	Pair Distribution Functions	93
4.5	Conclusions	96

5	A neural network-based four-body potential energy surface for parahydrogen	98
5.1	Introduction	98
5.2	Creating the Potential Energy Surface	101
5.2.1	Electronic Structure Calculations	101
5.2.2	Managing Degrees of Freedom	102
5.2.3	Sample Generation	104
5.2.4	Feature Transformations	105
5.2.5	Rescaling the Energies	107
5.2.6	Neural Network and Training	109
5.3	Discussion and Analysis	113
5.3.1	Lebedev Quadrature Comparison	113
5.3.2	Basis Set and Method Comparison	114
5.3.3	Short Range: Exponential Decay	118
5.3.4	Long Range: Bade Potential	118
5.3.5	Interactions in Solid Parahydrogen	121
5.4	Conclusion	125
6	Conclusion	126
6.1	Future work	128
6.1.1	Using the four-body PES in PIMC simulations	128
6.1.2	Many-body force evaluations	128
6.1.3	Higher-order many-body PESs	128
6.1.4	PIMC simulations with more degrees of freedom	129
	References	130

A	Supplementary material for Chapter 5	146
A.1	Trilateration	146
A.2	Lexicographic Ordering	148
A.3	Basis Set Superposition Error Fit Constants	149
A.4	Estimates of Bade Coefficients	149
A.5	Training Error Curves for Smaller Models	151
A.6	Short Range: Exponential Decay Fit	153
A.7	Transitions between different parts of the PES	155
A.8	Contributions to the hcp lattice for each geometry	157

Abbreviations

AHR	Adiabatic Hindered Rotor
ATM	Axilrod-Teller-Muto
AVNZ	Augmented correlation-consistent polarized Valence N Zeta
AVDZ	Augmented correlation-consistent polarized Valence Double Zeta
AVTZ	Augmented correlation-consistent polarized Valence Triple Zeta
AVQZ	Augmented correlation-consistent polarized Valence Quadruple Zeta
BOA	Born-Oppenheimer Approximation
BOL	Born-Oppenheimer-Like
BSSE	Basis Set Superposition Error
CCSD	Coupled Cluster with Singles and Doubles
CCSD(T)	Coupled Cluster with Singles, Doubles, and perturbative Triples
CPU	Central Processing Unit
DMC	Diffusion Monte Carlo
EOS	Equation of State
fcc	face-centred cubic
FSH	Faruk-Schmidt-Hinde
GPU	Graphics Processing Unit
hcp	hexagonal close packed
HF	Hartree-Fock
MATM	Modified Axilrod-Teller-Muto
MLP	Multilayer Perceptron
MoRiBS	Molecular Rotors in Bosonic Solvents
MP2	Møller-Plesset 2
MSE	Mean Square Error
<i>ortho</i>-D₂	Orthodeuterium
<i>ortho</i>-H₂	Orthohydrogen
<i>para</i>-H₂	Parahydrogen
PES	Potential Energy Surface
PIGS	Path-Integral Ground State
PIMC	Path-Integral Monte Carlo
PIMD	Path-Integral Molecular Dynamics
ReLU	Rectified Linear Unit
RKHS	Reproducing Kernel Hilbert Spaces
RMSE	Root Mean Square Error
SCF	Self-Consistent Field
SG	Silvera-Goldman

SSP	Shifted Softplus
TISE	Time-Independent Schrödinger Equation
VMC	Variational Monte Carlo
ZTZP	Zero-Temperature Zero-Pressure

Chapter 1

Introduction

1.1 Parahydrogen

In the body of this thesis, we model interactions between collections of parahydrogen molecules and use those interactions to simulate solid parahydrogen. Here, we briefly discuss certain important properties of the molecule.

1.1.1 Molecular parahydrogen

Molecular hydrogen is the simplest molecule in existence, and has been the subject of theoretical research since the inception of quantum mechanics. It appears as one of two nuclear spin isomers. One of them is orthohydrogen (*ortho*-H₂), which occupies the triplet state, with a total nuclear spin of 1. The other one, of particular interest in this thesis, is parahydrogen (*para*-H₂), which occupies the singlet state, with a total nuclear spin of 0.

Due to symmetry restrictions, *para*-H₂ can only occupy the even rotational states $j = \{0, 2, 4, \dots\}$, while *ortho*-H₂ can only occupy the odd rotational states $j = \{1, 3, 5, \dots\}$.¹ At room temperature or higher, the Boltzmann factors for both spin isomers are saturated, and they appear in a 3 : 1 ratio of *ortho*-H₂ to *para*-H₂. However, because *para*-H₂ can occupy the lowest rotational state $j = 0$ it is more energetically favourable compared to *ortho*-H₂. In the presence of a strong external inhomogeneous magnetic field, *ortho*-H₂ and *para*-H₂ molecules can convert between one another. As a result, below 20 K, molecular hydrogen is about 99.9 % *para*-H₂.

Energetically, the translational, rotational, vibrational, and electronic degrees of freedom in *para*-H₂ all operate on different energy scales.¹ Let $E(\nu, j)$ be the energy of a *para*-H₂ molecule in a state with the vibrational quantum number ν , and the rotational quantum number j . We find that

$$\begin{aligned} E(\nu = 0, j = 1) - E(\nu = 0, j = 0) &\approx 120 \text{ cm}^{-1} \\ E(\nu = 0, j = 2) - E(\nu = 0, j = 0) &\approx 360 \text{ cm}^{-1} \\ E(\nu = 1, j = 0) - E(\nu = 0, j = 0) &\approx 4000 \text{ cm}^{-1}. \end{aligned}$$

This property of *para*-H₂ becomes relevant later in the thesis. We can take advantage of the difference in energy scales to separate the different degrees of freedom to a good approximation. This is incredibly helpful for mathematically modelling interactions between *para*-H₂ molecules.

1.1.2 Solid parahydrogen

Below 13.8 K, *para*-H₂ forms a quantum molecular solid with a variety of interesting non-classical features.²⁻⁵ One of the more striking of these features is the abnormally large zero-point kinetic energy of the solid, giving each *para*-H₂ molecule very large translational vibrations about its nominal lattice site.⁶⁻⁹ The weak intermolecular interactions between the *para*-H₂ molecules, as well as their very low mass, give the molecules a large thermal wavelength.^{7,9} At the equilibrium density of $\rho = 0.026 \text{ \AA}^{-3}$,⁹ the ratio of the root mean square displacement to the lattice constant is about 0.2.^{6,8}

Many of the other strange properties of solid *para*-H₂ arise as a result of this large zero-point translational motion. For example, solid *para*-H₂ has a much greater lattice constant than one would expect classically. At lower densities, we find that this large displacement about the lattice sites gives solid *para*-H₂ a radial distribution function similar to that of a liquid.¹⁰ It is only at higher densities that the underlying structure of the solid can be seen in the radial distribution function.

As a molecular solid, the individual *para*-H₂ molecules retain most of their free molecule character. This allows us to model bulk systems of *para*-H₂ as a sum of interactions of smaller clusters of *para*-H₂, such as pairs, triplets, quadruplets, and so on.¹

It has recently become apparent that pairwise interactions are not enough on their own to accurately model solid *para*-H₂. The shortcomings of the two-body interaction are especially noticeable at high densities; the short-range wall is too repulsive, overestimating

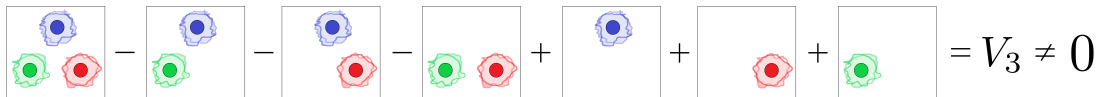
the pressure as the molecules move closer together. To accurately describe solid *para*-H₂, we need access to many-body interaction potentials.^{11–13} In most molecular systems, higher-order many-body interactions are typically considered perturbative or negligible compared to the two-body interaction. They typically only become relevant at very short ranges, or when very high accuracies are required. However, in solid *para*-H₂, many-body interactions are important at densities much lower than one would expect classically. The large translational motion of each *para*-H₂ molecule about its nominal lattice site cause the molecules to spend time at distances much shorter than the lattice constant. As a result, even if the many-body potential is negligibly weak at the nominal lattice sites, part of their probability distribution still bleeds into the short-range regions of coordinate space where the many-body potential is strong.

1.2 Many-body interactions

Higher-order many-body interactions, or those beyond the pair interaction, become important in simulations of solid *para*-H₂ at higher densities. They even modify the equation of state around the equilibrium density. In this work we create models for the three-body and four-body interaction potentials. Here we briefly discuss many-body interactions and how to calculate them.

1.2.1 What are many-body interactions?

A non-additive many-body interaction energy (often referred to as simply as a “many-body interaction energy”) is the interaction energy that remains when all the lower-order interaction energies have been subtracted out.^{14,15}



$$\boxed{\text{blue, red, green}} - \boxed{\text{blue, red}} - \boxed{\text{blue, green}} - \boxed{\text{red, green}} + \boxed{\text{blue}} + \boxed{\text{red}} + \boxed{\text{green}} = V_3 \neq 0$$

Figure 1.1: Depiction of the non-additive three-body interaction potential energy, and how subtracting the pair interaction energies generally results in a non-zero interaction potential. The solid red, blue, and green circles each represent a nucleus, and the semi-transparent shape around each nucleus represents the electron cloud. For simplicity, we show each particle as an atom, but the general idea extends to molecules as well. The energy of each monomer is added back to the total, to compensate for the fact that we indirectly subtract each monomer energy twice when subtracting all the pair energies.

In Fig. (1.1), we demonstrate this with the three-body interaction energy. Consider a collection of three interacting molecules. We can calculate the total energy of the three molecules together (depicted in the first box), and calculate the energies of the three individual pairs of molecules (depicted in the following three boxes). If the overall interaction potential were pairwise additive, then the total three-body interaction energy would be equal to the sum of the three pair interaction energies. In general, this is not the case. The presence of a third molecule perturbs the probability distribution of the electrons around the first two molecules, thus modifying the interaction. We can say that there is a remaining component of the total interaction energy between the three molecules that exists even after the interaction energies of the pairs have been removed. We refer to this remaining component as the three-body interaction energy. We can generalize this idea to higher-order many-body interactions, such as the four-body interaction, and so on.

1.2.2 Calculating many-body interaction energies

We can systematically derive the equations for calculating higher-order many-body interactions. Let $E_{ij\dots m}$ be the energy calculated by solving the time-independent Schrödinger equation (TISE) when molecules with labels $i, j, \dots, m$ are present. Next, we define the “monomer interaction energy” ΔE_i to simply be the energy itself: $\Delta E_i = E_i$. The n -body interaction energy $\Delta E_{12\dots n}$ is defined by taking the n -body energy $E_{12\dots n}$ and subtracting the interaction energies of all possible $(n-1)$ -body combinations, all possible $(n-2)$ -body combinations, and so on, until the interaction energies of all possible monomer combinations have been subtracted. The definition for each n -body interaction energy is a combination of sums and differences of lower-order many-body energies, and the aforementioned scheme ensures that the correct signs are used.

The two-body interaction energy ΔE_{12} is given by

$$\begin{aligned}\Delta E_{12} &= E_{12} - (\Delta E_1 + \Delta E_2) \\ &= E_{12} - (E_1 + E_2).\end{aligned}\tag{1.1}$$

Similarly, the three-body interaction energy ΔE_{123} is calculated by subtracting all the pair and monomer interaction energies,

$$\begin{aligned}\Delta E_{123} &= E_{123} - (\Delta E_{12} + \Delta E_{13} + \Delta E_{23}) - (\Delta E_1 + \Delta E_2 + \Delta E_3) \\ &= E_{123} - (E_{12} + E_{13} + E_{23}) + (E_1 + E_2 + E_3),\end{aligned}\tag{1.2}$$

and the four-body interaction energy ΔE_{1234} is calculated by subtracting all the three-body, two-body, and monomer interaction energies,

$$\begin{aligned}
\Delta E_{1234} &= E_{1234} \\
&\quad - (\Delta E_{123} + \Delta E_{124} + \Delta E_{134} + \Delta E_{234}) \\
&\quad - (\Delta E_{12} + \Delta E_{13} + \Delta E_{14} + \Delta E_{23} + \Delta E_{24} + \Delta E_{34}) \\
&\quad - (\Delta E_1 + \Delta E_2 + \Delta E_3 + \Delta E_4) \\
&= E_{1234} \\
&\quad - (E_{123} + E_{124} + E_{134} + E_{234}) \\
&\quad + (E_{12} + E_{13} + E_{14} + E_{23} + E_{24} + E_{34}) \\
&\quad - (E_1 + E_2 + E_3 + E_4).
\end{aligned} \tag{1.3}$$

1.2.3 General properties of many-body interaction energies for parahydrogen

The two-body, three-body, and four-body interaction energies for parahydrogen all share certain properties when described isotropically. At short intermolecular separations, each of them can be fit to an exponential decay function. At long intermolecular separations, each of them converges to an induced dispersion interaction. The dominant long-distance interaction for two-body potential is the induced dipole-dipole interaction, which decays as r^{-6} .¹⁶ For the three-body potential, it is the Axilrod-Teller-Muto potential, which decays as r^{-9} .^{17,18} For the four-body potential, it is the Bade potential, which decays as r^{-12} .¹⁹ The analytic dispersion interactions become unrealistic at short distances so we need to rely on *ab initio* calculations for cases where the molecules are close together.

At short distances, the two-body interaction is repulsive, the three-body interaction is net attractive, and the four-body interaction is repulsive. In fact, there seems to be a pattern in the many-body interactions for *para*-H₂, in which the sign of the interaction at short distances changes based on the parity of the interaction order. In a study on the many-body interactions in solid helium-4 by Tian *et al.*,²⁰ it was found that the two-body, four-body, and six-body interactions have net repulsive contributions, while the three-body and five-body interactions have net attractive contributions. Considering the similarities between solid helium-4 and solid *para*-H₂, it seems likely that this alternating trend might continue for the higher-order many-body interactions for *para*-H₂.

1.3 The time-independent Schrödinger equation

The electronic structure calculations and path-integral Monte Carlo simulations performed in this work are both ways of solving the TISE

$$\hat{H} |\phi\rangle = E |\phi\rangle \quad (1.4)$$

where $\hat{H}$ be the Hamiltonian that describes the system of particles, $|\phi\rangle$ is an eigenstate, and E is the energy of the eigenstate. However, before we discuss electronic structure theory or path integrals, we should start with the Hamiltonian at a lower level of approximation. In quantum chemistry, we are primarily concerned with the Hamiltonian describing a system of interacting electrons and nuclei

$$\hat{H} = \hat{K}_n + \hat{V}_n(\{\mathbf{R}_A\}) + \hat{H}_{ne}(\{\mathbf{r}_i, \mathbf{R}_A\}) \quad (1.5)$$

where we have the kinetic energy of the nuclei

$$\hat{K}_n = - \sum_A \frac{\hbar^2}{2M_A} \nabla_A^2, \quad (1.6)$$

the Coulombic potential energy between the nuclei

$$\hat{V}_n(\{\mathbf{R}_A\}) = \frac{1}{2} \sum_{A \neq B} \frac{Z_A Z_B e^2}{|\mathbf{R}_A - \mathbf{R}_B|}, \quad (1.7)$$

and the electronic Hamiltonian, representing the kinetic energy of the electrons and the Coulombic potential energy between the electrons and other electrons and nuclei

$$\hat{H}_{ne}(\{\mathbf{r}_i, \mathbf{R}_A\}) = - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,A} \frac{Z_A e^2}{|\mathbf{r}_i - \mathbf{R}_A|}. \quad (1.8)$$

In the above equations, lowercase indices label the electrons, uppercase indices label the nuclei, m_e is the electron mass, $\mathbf{r}_i$ represents an electron's position, M_A is a nucleon mass, $\mathbf{R}_A$ represents a nucleon's position, and Z_A represents a nucleon's atomic number.

The only systems described by Eq. (1.5) with an analytic solution are those with a single nucleon and a single electron, also known as hydrogen-like atoms. We must resort to approximations if we want to find solutions to systems that are any more complicated.

1.4 Electronic structure theory

In our first set of approximations, we account for the electron degrees of freedom. In this work, we make use of this theory when calculating both the three-body and four-body interaction energies. We first briefly describe the theory behind the approximation, and the key settings we use in the software that actually performs the calculations. Electronic structure theory is an entire subfield of research in and of itself. We limit our discussion to the general idea behind electronic structure calculations, a superficial overview of two particularly important methods, and the importance of basis sets.

1.4.1 Born-Oppenheimer Approximation

The Born-Oppenheimer approximation (BOA) is one of the most important approximations in quantum chemistry.¹ It allows us to separate the electron and nuclear degrees of freedom in the molecular Hamiltonian Eq. (1.5). The mass of a nucleon is about 1837 times greater than that of an electron. Atomic nuclei thus move on much longer time scales than electrons. An effective approximation is to assume that the nuclei are completely stationary, and solve the TISE assuming only the electrons are capable of motion. As the nuclei move, we assume that the electrons move quickly enough so as to immediately adapt to the positions of the nuclei. Using the BOA, Eq. (1.6) is neglected entirely, and Eq. (1.7) becomes a constant. We are left with only the electronic Hamiltonian Eq. (1.8), now parametrized by the atomic nuclei positions $\{\mathbf{R}_A\}$.

Suppose we solve Eq. (1.8) for the ground state electronic wave function $\mathcal{X}_0(\{\mathbf{r}_i\}; \{\mathbf{R}_A\})$ and the ground state electronic energy $E_0(\{\mathbf{R}_A\})$

$$\hat{H}_{ne}\mathcal{X}_0(\{\mathbf{r}_i\}; \{\mathbf{R}_A\}) = E_0(\{\mathbf{R}_A\})\mathcal{X}_0(\{\mathbf{r}_i\}; \{\mathbf{R}_A\}). \quad (1.9)$$

We can now express Eq. (1.5) as a function of only the atomic nuclei positions

$$\hat{H}(\{\mathbf{R}_A\}) \approx \frac{1}{2} \sum_{A \neq B} \frac{Z_A Z_B e^2}{|\mathbf{R}_A - \mathbf{R}_B|} + E_0(\{\mathbf{R}_A\}). \quad (1.10)$$

In practice, we solve for the electronic energy in Eq. (1.9) using software such as MRCC.^{21,22} For the input to these software packages, we provide the atomic positions and atoms themselves, a “method” for calculating the energy, a “basis set” with which to construct the wave function, and a variety of other settings such as convergence criteria.

1.4.2 Electronic structure methods

A common method for solving Eq. (1.9) is the Hartree-Fock (HF) method.²³ Among the features that characterize the HF method, two stand out. First, the HF method assumes that the wave function can be approximated by a single Slater determinant, which drastically simplifies the calculation. Second, the HF method is a type of mean field theory. Instead of accounting for all pairs of individual electron-electron Coulombic interactions, each electron interacts with an effective interaction field produced by all the other electrons. This approximation reduces the multi-electron problem into several one-electron problems. The HF method has a time and space complexity of $\mathcal{O}(n^4)$, where n is the number of basis functions (discussed later) which is relatively cheap in the context of electronic structure methods.

However, the HF method has many limitations. In general, molecular wave functions cannot be approximated by a single Slater determinant, and this decision in the HF method limits its possible solutions. Also, because it is a mean-field theory it neglects electron correlations entirely, which prevents it from predicting certain important effects, such as the dispersion interaction. The HF method typically only gives acceptable results in systems where electron correlations are very weak. We can relax some of the assumptions made in the HF method to come up with more powerful methods, known as post-Hartree-Fock *ab initio* methods. One of these methods is the coupled cluster method, which introduces electron correlations.²⁴

To perform the coupled cluster method, we must first perform the HF method to find the HF wave function $|\Psi_0\rangle$. We have an exponential operator $\exp\{\hat{T}\}$, which uses the total cluster operator $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots$, where $\hat{T}_1$ is the cluster operator for all one-electron excitations, $\hat{T}_2$ is the cluster operator for all two-electron excitations, and so on. These cluster operators are responsible for introducing electron correlations. The coupled cluster wave function is expressed as

$$|\Phi\rangle = \exp\{\hat{T}\} |\Psi_0\rangle . \quad (1.11)$$

To make the problem computationally tractable, we must truncate the total cluster operator to the first few operators. By solving the Schrödinger equation with Eq. (1.11), we can find the ground state coupled cluster electronic energy.

One especially useful version of the coupled cluster method is the version with full single and double excitations and perturbative triples excitations, referred to as CCSD(T). In quantum chemistry, CCSD(T) is referred to as the “gold standard,” because it is one of the most accurate methods that is still computationally feasible to use with relatively large molecules. The perturbative treatment of the triple excitations introduces a systemic error

that gives CCSD(T) a similar accuracy to coupled cluster methods of higher orders. The hydrogen molecule pair PES created by Hinde in 2008,²⁵ and the three-body and four-body PESs presented in this thesis, all use the CCSD(T) method when calculating their energies.

1.4.3 Basis sets

In electronic structure methods, each electron orbital in a Slater determinant can be expressed as a linear combination of atomic orbitals. The set of atomic orbitals used for this expansion is called the basis set.

The quantum chemistry literature contains many popular and standardized basis sets. They are designed in a way to generally maximize the quality of energies for specific systems while attempting to limit the computational expense. In particular, the augmented correlation-consistent polarized valence basis sets of cardinality N , abbreviated as AVNZ, are excellent for systems with weak dispersive energies.^{26,27} Using a larger basis set allows us to recover higher quality energies from our electronic structure methods, but also requires more computational effort. As the number of electrons in our system increases, we are forced to use smaller basis sets to make sure our calculations remain tractable.

For reference, the hydrogen molecule pair PES created by Hinde in 2008 was created using an AVQZ basis set,²⁵ the three-body *para*-H₂ PES presented in this work was created using an AVTZ basis set, and the four-body *para*-H₂ PES presented in this work was created using an AVDZ basis set.

1.4.4 The Counterpoise correction

By performing the calculations for the many-body interaction energies in Eqs. (1.1), (1.2) and (1.3), we introduce an error called the basis set superposition error (BSSE).

To explain the BSSE, we can take the calculation of the two-body interaction energy Eq. (1.1) as an example. When calculating the energy E_{12} , some of the basis functions centred on the atoms of molecule 1 overlap with the basis functions centred on the atoms of molecule 2, and vice versa. Thus it is as if molecule 1 has a slightly larger basis set during the calculation of E_{12} . Now suppose we calculate E_1 , with both molecule 2 and the basis sets of molecule 2 absent. Compared to molecule 1 during the calculation of E_{12} , molecule 1 in the calculation of E_1 has fewer basis set functions to work with. Thus the pair energy E_{12} , and the monomer energies E_1 and E_2 , are calculated at different levels of

quality. Taking their differences in Eq. (1.1) would introduce an additional error due to this inconsistency, which we call the BSSE.

We decide to account for the BSSE using the Counterpoise method.²⁸ In this method, the calculations for the energies E_{12} , E_1 and E_2 are all performed with the basis sets for both molecules present. For example, when we calculate E_1 , we remove the nuclei and electrons of molecule 2, but keep the basis sets of molecule 2 in the same locations we would find them in for the calculation of E_{12} . This strategy ensures that all calculations are performed at the same level of quality.

1.5 Potential energy surfaces

Thus far, we have described how to use the BOA to separate the motions of the atomic nuclei and the electrons, and how this allows us to find the energy of the system as a function of the nuclear positions. However, working with a potential as a function of nuclear positions is sometimes inconvenient, and we would like to make further simplifications. The general idea behind the BOA, where we adiabatically separate slow and fast degrees of freedom of a system, can be extended to other aspects of parahydrogen.

In 2008, Hinde published a full 6D PES for the *para*-H₂ pair interaction.²⁵ In 2014, Faruk *et al.* used the Adiabatic Hindered Rotor (AHR) method²⁹ to reduce Hinde’s 6D PES to a 1D PES of just the pair distance between the centres of mass.¹⁶ In the AHR method, we write the Hamiltonian in the space-fixed frame as

$$\hat{H}_{2B} = \hat{h}_1 + \hat{h}_2 + \hat{V}_{6D}(R, \rho_1, \rho_2, \theta_1, \theta_2, \phi) \quad (1.12)$$

where $\hat{h}_1$ and $\hat{h}_2$ are the monomer rovibrational Hamiltonians,³⁰ $\hat{V}_{6D}$ is Hinde’s 6D PES, R is the pair distance between the centres of mass, ρ_1 and ρ_2 are the two bond lengths, θ_1 and θ_2 are the two space-fixed azimuthal angles, and ϕ is the dihedral angle. We show the degrees of freedom for Hinde’s 6D PES in Fig. (1.2). In the AHR method, we consider the pair distance R to be the slow coordinate, and the remaining five rovibrational coordinates to be the fast coordinates. In a fashion similar to the BOA, we diagonalize over the five fast coordinates. The resulting energy eigenvalues are a function of the pair distance R . The Faruk-Schmidt-Hinde (FSH) potential is created by performing the AHR method and selecting the smallest energy eigenvalue for a range of values of R .³¹ We should emphasize that the diagonalization process uses not only the $j = 0$ rotational state, but also the $j = 2$ and $j = 4$ rotational states.

In the three-body and four-body PESs for *para*-H₂ we use a rigid rotor approximation, by fixing the bond length of each molecule to the rovibrational ground state value of 0.7668 Å.¹¹ We remove the angular orientation degrees of freedom using Lebedev quadratures.^{32,33} This is a technique used to approximate the integral of a function over a unit sphere using a sum of that function evaluated at specific angular orientations. After applying the Lebedev quadrature, the three-body and four-body PESs for *para*-H₂ become isotropic, and are spherically averaged to the $j = 0$ rotational state.

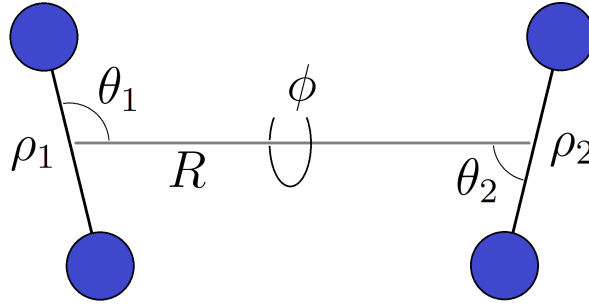


Figure 1.2: A graphical depiction of the six degrees of freedom that parametrize the pair potential between two hydrogen molecules.

The three-body and four-body PESs for *para*-H₂ use less sophisticated means of reducing the degrees of freedom than those used for the FSH potential. One reason is that the computational constraints increase with the number of particles, and it is too expensive to perform the required *ab initio* calculations to create the full three-body and four-body PESs to sufficient detail in the first place. Another reason is that, except at extremely high densities, the molecules in solid *para*-H₂ are already well modelled as spherically symmetric particles interacting through isotropic potentials. The anisotropic and vibrational components are typically considered small corrections to the total interaction potential. Also, the three-body and four-body PESs can be seen as corrections to the two-body PES. Thus any contributions from the anisotropy or the bond vibrations in the three-body and four-body PESs are themselves a correction to a correction, and we can more easily justify neglecting them.

The two-body, three-body, and four-body PESs described in this thesis all treat the *para*-H₂ molecules as pointwise particles. Each of them is a function of only the pair distances between the centres of mass of the *para*-H₂ molecules. We can express these

PESs, respectively, as

$$V_2(\mathbf{r}_1, \mathbf{r}_2) = V_2(r_{12}) \quad (1.13)$$

$$V_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = V_3(r_{12}, r_{13}, r_{23}) \quad (1.14)$$

$$V_4(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = V_4(r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}), \quad (1.15)$$

where $\mathbf{r}_i$ is the position of the centre of mass of the i^{th} *para*-H₂ molecule, and r_{ij} is the distance between the centres of mass of the i^{th} and j^{th} *para*-H₂ molecules.

The rotational and vibrational kinetic energies are accounted for in the monomer terms for the hydrogen molecules during the AHR method. The only remaining kinetic energy term is that of the translational motion of the centres of mass of the *para*-H₂ molecules. We can express the working Hamiltonian as

$$\hat{H}_0 = \hat{K} + \hat{V}(\{\mathbf{r}\}) \quad (1.16)$$

where $\hat{K}$ is the operator for the translational kinetic energy, and $\hat{V}(\{\mathbf{r}\})$ is the operator for the interaction potential after the reduction of the degrees of freedom. Depending on the context in this thesis, $\hat{V}$ can be the two-body PES on its own, the sum of the two-body and three-body PESs, or the sum of the two-body, three-body, and four-body PESs.

1.6 Path integral Monte Carlo

Even with the aforementioned simplifications, the remaining Hamiltonian is often still too complicated to explicitly diagonalize. Certain computational methods allow us to calculate the expectation value of operators without diagonalizing them. We use one of these methods, called the path-integral Monte Carlo (PIMC) method in this work.^{34,35} For simplicity, we describe the PIMC method using the primitive Trotter decomposition. However, there are other Trotter decompositions that lead to a lower systematic error.

We begin with the partition function Z , which can be recovered from the working Hamiltonian $\hat{H}_0$ using

$$Z = \int d\mathbf{q} \langle \mathbf{q} | e^{-\beta \hat{H}_0} | \mathbf{q} \rangle \quad (1.17)$$

where β is the thermodynamic beta, and for convenience of notation, we use $\mathbf{q}$ instead of $\{\mathbf{r}\}$ to represent the positions of all particles in the system. We refer to the exponential $e^{-\beta \hat{H}_0}$ as the density matrix.

The integral Eq. (1.17) would be much easier to represent if we could separate the kinetic and potential components. Unfortunately, performing this separation introduces a

systemic error. To demonstrate this, we can expand both $e^{-\beta\hat{K}}$ and $e^{-\beta\hat{V}}$ in a Taylor series and calculate the product $e^{-\beta\hat{K}}e^{-\beta\hat{V}}$.

$$e^{-\beta\hat{K}}e^{-\beta\hat{V}} = 1 - \beta(\hat{K} + \hat{V}) + \frac{\beta^2}{2}(\hat{K}^2 + 2\hat{K}\hat{V} + \hat{V}^2) + \mathcal{O}(\beta^3) \quad (1.18)$$

We then compare this with the expansion of $e^{-\beta\hat{H}_0}$ in a Taylor series.

$$\begin{aligned} e^{-\beta\hat{H}_0} &= 1 - \beta\hat{H}_0 + \frac{\beta^2}{2}\hat{H}_0^2 + \mathcal{O}(\beta^3) \\ &= 1 - \beta(\hat{K} + \hat{V}) + \frac{\beta^2}{2}(\hat{K}^2 + \hat{K}\hat{V} + \hat{V}\hat{K} + \hat{V}^2) + \mathcal{O}(\beta^3) \end{aligned} \quad (1.19)$$

The expansions of Eqs. (1.18) and (1.19) begin to differ in the terms proportional to β^2 . Because the kinetic and potential components do not commute, we find that separating the density matrix into factors of $e^{-\beta\hat{K}}$ and $e^{-\beta\hat{V}}$ introduces an error (called the Trotter error) that scales with β^2 .

$$e^{-\beta\hat{H}_0} = e^{-\beta\hat{K}}e^{-\beta\hat{V}} - \frac{\beta^2}{2}[\hat{K}, \hat{V}] + \mathcal{O}(\beta^3) \quad (1.20)$$

To work around this, we can perform Trotter factorization on the density matrix. Define $\tau = \beta/P$, where P is a positive integer. We can now express the partition function as

$$Z = \int d\mathbf{q} \langle \mathbf{q} | (e^{-\tau\hat{H}_0})^P | \mathbf{q} \rangle . \quad (1.21)$$

Separating the kinetic and potential components of each individual Trotterized density matrix $e^{-\tau\hat{H}_0}$ introduces a Trotter error that only scales with $\tau^2 = (\beta/P)^2$. By increasing the value of P , we can systematically reduce the Trotter error, and in the limit $P \rightarrow \infty$, the factorization becomes exact.

Next, we insert $P - 1$ instances of the resolution of the identity into the integral, each nested between two instances of the Trotterized density matrix. The partition function is now

$$Z = \int d\mathbf{q}_1 d\mathbf{q}_2 \dots d\mathbf{q}_P \langle \mathbf{q}_1 | e^{-\tau\hat{H}_0} | \mathbf{q}_2 \rangle \langle \mathbf{q}_2 | e^{-\tau\hat{H}_0} | \mathbf{q}_3 \rangle \dots \langle \mathbf{q}_P | e^{-\tau\hat{H}_0} | \mathbf{q}_1 \rangle , \quad (1.22)$$

where we relabel $\mathbf{q}$ as $\mathbf{q}_1$, and label the position coordinates for each resolution of the identity from 2 to P . The position coordinates for the matrix elements in Eq. (1.22) form a ring. There is a matrix element connecting $\mathbf{q}_1$ to $\mathbf{q}_2$, $\mathbf{q}_2$ to $\mathbf{q}_3$, and so on until we connect $\mathbf{q}_P$ back to $\mathbf{q}_1$. Acknowledging this periodicity and defining $\mathbf{q}_1 = \mathbf{q}_{P+1}$, we can simplify Eq. (1.22) into

$$Z = \int \prod_{i=1}^P d\mathbf{q}_i \langle \mathbf{q}_i | e^{-\tau\hat{H}_0} | \mathbf{q}_{i+1} \rangle . \quad (1.23)$$

Now we separate the kinetic and potential components of each matrix element.

$$\langle \mathbf{q}_i | e^{-\tau \hat{H}_0} | \mathbf{q}_{i+1} \rangle = \langle \mathbf{q}_i | e^{-\tau \hat{K}} e^{-\tau \hat{V}} | \mathbf{q}_{i+1} \rangle + \mathcal{O}(\tau^2) \quad (1.24)$$

$$\approx \int d\mathbf{q}' \langle \mathbf{q}_i | e^{-\tau \hat{K}} | \mathbf{q}' \rangle \langle \mathbf{q}' | e^{-\tau \hat{V}} | \mathbf{q}_{i+1} \rangle . \quad (1.25)$$

The potential is diagonal in the position representation, allowing us to write

$$\langle \mathbf{q}' | e^{-\tau \hat{V}} | \mathbf{q}_{i+1} \rangle = e^{-\tau V(\mathbf{q}')} \delta^{Nd}(\mathbf{q}' - \mathbf{q}_{i+1}) , \quad (1.26)$$

where $\delta^{Nd}(\mathbf{q})$ is the delta function for a system with a dimensionality of d that is composed of N particles. To evaluate the kinetic component, we insert two instances of the resolution of the identity in the momentum representation into the matrix element, express the kinetic energy operator in terms of the canonical momentum operator, and carry out the evaluation.

$$\langle \mathbf{q}_i | e^{-\tau \hat{K}} | \mathbf{q}' \rangle = \int d\mathbf{p}_0 d\mathbf{p}_1 \langle \mathbf{q}_i | \mathbf{p}_0 \rangle \langle \mathbf{p}_0 | e^{-\tau \hat{K}} | \mathbf{p}_1 \rangle \langle \mathbf{p}_1 | \mathbf{q}' \rangle \quad (1.27)$$

$$= \left(\frac{m}{2\pi\hbar^2\tau} \right)^{Nd/2} \exp \left[-\frac{m}{2\hbar^2\tau} (\mathbf{q}_i - \mathbf{q}')^2 \right] . \quad (1.28)$$

The matrix element for the Trotterized density matrix can now be expressed as

$$\langle \mathbf{q}_i | e^{-\tau \hat{H}_0} | \mathbf{q}_{i+1} \rangle \approx \left(\frac{m}{2\pi\hbar\tau} \right)^{Nd/2} \exp \left\{ -\tau \left[\frac{m}{2} \left(\frac{\mathbf{q}_{i+1} - \mathbf{q}_i}{\hbar\tau} \right)^2 + V(\mathbf{q}_i) \right] \right\} . \quad (1.29)$$

Expressing the path $\mathbf{q}_i$ as a list of positions of the N particles $\{\mathbf{r}_{1,i}, \mathbf{r}_{2,i}, \dots, \mathbf{r}_{N,i}\}$, we write

$$\langle \mathbf{q}_i | e^{-\tau \hat{H}_0} | \mathbf{q}_{i+1} \rangle \approx \left(\frac{m}{2\pi\hbar\tau} \right)^{Nd/2} \exp \left\{ -\tau \sum_{n=1}^N \left[\frac{m}{2} \left(\frac{\mathbf{r}_{n,i+1} - \mathbf{r}_{n,i}}{\hbar\tau} \right)^2 + V(\mathbf{r}_{n,i}) \right] \right\} . \quad (1.30)$$

Finally, we write our partition function as

$$Z \approx \left(\frac{m}{2\pi\hbar\tau} \right)^{Nd/2} \int \prod_{i=1}^P d\mathbf{q}_i \exp \left\{ -\tau \sum_{n=1}^N \left[\frac{m}{2} \left(\frac{\mathbf{r}_{n,i+1} - \mathbf{r}_{n,i}}{\hbar\tau} \right)^2 + V(\mathbf{r}_{n,i}) \right] \right\} . \quad (1.31)$$

We can sample Eq. (1.31) using the PIMC method. The set of all positions in the system $S = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N\} = \{\mathbf{r}_{1,1}, \dots, \mathbf{r}_{1,P}, \mathbf{r}_{2,1}, \dots, \mathbf{r}_{N,P}\}$ describes the state of the system (alongside other quantities such as the temperature, the simulation box size, and the number of particles, although these are typically fixed). Using a variety of Monte Carlo moves, we can alter these positions to explore different possible states.

Let $\mathcal{E}_{\hat{A}}(S^{(k)})$ be an unbiased estimator for the operator $\hat{A}$ applied to the k^{th} recorded state of the simulation $S^{(k)}$. If we sample N_K state snapshots from the simulation, we can estimate the expectation value of the operator $\hat{A}$ on this system as

$$\langle \hat{A} \rangle \approx \frac{1}{N_K} \sum_{k=1}^{N_K} \mathcal{E}_{\hat{A}}(S^{(k)}). \quad (1.32)$$

For example, the estimator for the total two-body potential energy is given by

$$\mathcal{E}_{\hat{V}_2}(S^{(k)}) = \frac{1}{P} \sum_{i=1}^P \sum_{n=1}^{N-1} \sum_{m=n+1}^N V_2(|\mathbf{r}_{n,i} - \mathbf{r}_{m,i}|). \quad (1.33)$$

We can derive estimators for other quantities of interest, such as the kinetic energy, the radial distribution function, the three-body potential energy, and so on. These estimators are invaluable for analyzing the PIMC simulations performed in this work.

1.7 Machine learning

The traditional method of creating a PES involves first selecting an analytic functional form with various parameters, and then modifying those parameters such that the PES best fits the data. This analytic form is often physically motivated, and thus we can more easily physically interpret the various parts of our PES. Even without a functional form, if the relationship between the inputs and outputs is simple enough, we can create a PES using simple fitting methods. For example, the FSH potential is a 1D pointwise pair interaction potential. Once given the set of inputs (the pair distances) and the corresponding outputs (the two-body interaction energies), we can trivially fit the potential using a 1D linear interpolation.

However, when the input is multidimensional, or the relationship between the inputs and outputs is too complicated, the traditional methods of creating a PES become too difficult to carry out. Many-body interactions are necessarily multidimensional, and it is far more challenging to find an analytic form or a simple computational method to fit them. To our benefit, advances in machine learning have made it much easier to create higher dimensional PESs for many-body interactions. For the present work, we use two machine learning tools, the reproducing kernel Hilbert space (RKHS) method, and the multilayer perceptron, to create our PESs.

The three-body PES is a function of only the three pair distances between the three molecules. We convert the three side lengths into a system of curvilinear rescaled Jacobi

coordinates, which are still simple enough to be easily interpretable. We then use a machine learning method called the reproducing kernel Hilbert space method, through the RKHS software toolkit by Unke *et al.*, to map the Jacobi coordinates to the energies.³⁶

The four-body PES is a function of the six relative pair distances between the four molecules. It is much more difficult to find a suitable coordinate transformation. For example, there are permutation issues associated with the six pair distances, and four different ways to violate the triangle inequality rule. Due to these difficulties, we decide to create the four-body PES using a multilayer perceptron.

1.7.1 Multilayer perceptrons

A multilayer perceptron (MLP) is a repeating sequence of inner products between matrices, with non-linear activation functions between them to break the linearity of the inner products. The MLP itself is referred to as the model, and the matrix elements of all the matrices that compose it are called the weights of the model. An input is passed to the model in the form of a column vector. We take the inner product of the input with the weights of the first matrix, and then pass the result to an activation function to introduce non-linearity, producing a new column vector. We then take the inner product of this new column vector with the second matrix, and then pass the result to another activation function. This process of taking inner products and using activation functions is repeated until we reach the final layer of the model. The value returned by the model is the “predicted output.”

Suppose we have a training set, composed of inputs and their corresponding “true outputs.” In the current context, the inputs are ordered sets of pair distances, and the corresponding outputs are the interaction energies. At first, the weights of the model are randomized, and thus if we were to pass any of the inputs into the model, the output would be completely random. At a superficial level, the process of creating the PES is really the process of modifying the weights of the multilayer perceptron in such a way that the model maps the inputs to their corresponding true outputs.

The process of modifying the model’s weights is an optimization problem. Suppose we pass the inputs from our training set into the MLP and recover the predicted outputs. We then use a function called the loss function to measure how similar the predicted outputs are to the corresponding true outputs. Using the chain rule, we can calculate the gradient of the loss function with respect to the model’s outputs, as well as all the weights in the model. We can then use one of several optimization algorithms, such as gradient descent to adjust the weights of the model. After each adjustment, the predicted outputs should

better match the corresponding true outputs. This process of modifying the weights is repeated until the difference between the predicted and true outputs becomes sufficiently small, at which point we have created our PES.

The above description of an MLP and how to train it is purposely oversimplified for brevity. For example, it is not only enough that the inputs in our training data set map to their corresponding outputs, but also that future unseen inputs map properly to their own corresponding outputs. Thus in addition to the training data set, there are also testing and validation data sets. The process of optimization and modifying the weights is also fairly complicated. We often use algorithms that are more advanced than gradient descent, such as Adam or stochastic gradient descent. There are also concerns such as overfitting and underfitting that we need to account for.

1.7.2 Benefits and drawbacks of a multilayer perceptron PES

Using an MLP to fit the training data makes the creation of the PES much more straightforward. We provide the model and the training data, and the training process produces a set of weights in the form of matrix elements capable of mapping the inputs to their corresponding outputs. The responsibility of learning the relationship between the inputs and outputs is taken over by the training process.

Compared to an analytic function, an MLP is more like a black box. The details of the MLP, such as the number of hidden layers or their sizes, are rarely physically motivated, and many different MLPs can produce very similar mappings. In addition, the resulting weights are often not unique. Repeating the training process with different initial random weights may result in creating a completely different set of trained weights which happen to map the inputs to the outputs. This makes a trained MLP more difficult to interpret.

1.8 Energy Unit Conversions

In this work, we show quantities of energy in both units of Kelvin (K) and reciprocal centimetres (cm^{-1}). The latter units are also called wavenumbers. Despite not being SI units, they are commonly used as units of energy in fields related to spectroscopy and low temperature physics.

The wavenumber (the quantity, not the unit) k of a photon in a vacuum is directly proportional to the photon's frequency $\nu = ck/2\pi$ and the photon's energy $E = \hbar ck$, where

c is the speed of light. This makes the wavenumber a useful quantity when working with photon energies. The wavenumber (the quantity) is expressed in dimensions of reciprocal length, and the most common unit is the reciprocal centimetre, also called a wavenumber. The choice of the reciprocal centimetre as opposed to the reciprocal of another unit of length, or even the standard SI unit of energy, is likely due to conveniences of scale. When we express the gaps between energy levels in molecules in units of wavenumbers, the numbers are typically neither extremely small nor extremely large. These same energy gaps expressed in units of joules are often on the order of 10^{-24} J to 10^{-18} J. These properties make the wavenumber convenient to use as a unit of energy for spectroscopists.

Expressing energies in units of Kelvin is convenient when discussing the occupancy of states in a canonical ensemble, because the ability for a system to occupy states of different energy levels depends on the temperature. For example, suppose we have a two-state system at room temperature whose excited state has an energy of 4.0×10^{-20} J. The extent to which this state is occupied is not immediately obvious to most readers. This same energy can be expressed as roughly 2897 K, which makes it clear that this state is rarely occupied at room temperature.

The conversions between SI units of energy and both wavenumbers and Kelvin are straightforward. To convert an energy in units of Kelvin to units of joules, multiply it by the Boltzmann constant $k_B = 1.380649 \times 10^{-23}$ J/K. To convert an energy in units of wavenumbers to units of joules, multiply it by $\hbar c \approx 1.98630 \times 10^{-23}$ J/cm⁻¹.

1.9 Overview of the thesis

The remainder of this thesis is organized as follows: In Chapter 2, we perform PIMC simulations of solid *para*-H₂ using the FSH pair potential to find the equation of state and the first vibrational matrix shift of the solid. In Chapter 3, we create an isotropic *ab initio* three-body PES for *para*-H₂, fit using the RKHS method. In Chapter 4, we again perform PIMC simulations of solid *para*-H₂, using both the FSH pair potential and the aforementioned three-body PES, to find the equation of state of the solid. In Chapter 5, we create an isotropic *ab initio* four-body PES for *para*-H₂, fit using a multilayer perceptron. Finally, in Chapter 6, we summarize what we have done throughout the thesis and describe avenues for future work.

Chapter 2

Equation of state and first principles prediction of the vibrational matrix shift of solid parahydrogen

This chapter is reproduced from Alexander Ibrahim, Lecheng Wang, Tom Halverson, Robert James LeRoy, and Pierre-Nicholas Roy, “Equation of state and first principles prediction of the vibrational matrix shift of solid parahydrogen,” *Journal of Chemical Physics* **151**, 244501 (2019), with the permission of AIP Publishing.³⁷

2.1 Introduction

For the past several decades, the properties of solid parahydrogen (*para*-H₂) have been studied extensively. Its weak intermolecular interactions and low mass make *para*-H₂ crystals very “soft” and give them many interesting quantum properties.^{7,9} Solid *para*-H₂ is one of the simplest examples of a quantum crystal.^{2–5,38–43} Also, as a spinless boson, *para*-H₂ has proven itself to be a viable candidate for superfluidity, both spectroscopically and theoretically.^{44–53} For example, computer simulations have shown that “glassy” *para*-H₂ possesses superfluid properties.⁵⁴ Experimentally, solid *para*-H₂ is an excellent accommodating host for impurity spectroscopy, owing to its relatively large lattice constant and the spherical symmetry of its $J = 0$ rotational level.^{55–58}

A large amount of research has been dedicated to characterizing the thermodynamics of *para*-H₂. For example, relative to its value in the gas phase, the frequency shift of the

Raman $Q_1(0)$ transition at $T = 4.2$ K was measured by Oka *et al.* to be -11.43 cm^{-1} in solid *para*- H_2 .⁴⁰ This is more than twice the vibrational frequency shift observed in large *para*- H_2 clusters.^{16,40,59} To date, no theoretical studies have been able to address this discrepancy. Another quantity attracting wide interest is the isotherm curve of *para*- H_2 , particularly at $T = 4.2$ K.^{60–63} Silvera’s review covers the subject far beyond the scope of the discussion here.⁹ The zero pressure molar volume and high compressibility of solid *para*- H_2 at 4.2 K both make a strong case that it is a quantum solid at that temperature.⁹ Further support for this idea comes from Schuch *et al.*, who have shown that solid *para*- H_2 has a hexagonal close-packing (*hcp*) structure in its equilibrium state.⁶⁴

Several computational studies have attempted to provide insights to the experiments outlined above. In 1996, Cheng and Whaley performed Diffusion Monte Carlo (DMC) calculations on solid *para*- H_2 at equilibrium density⁶⁵ using both the Buck and Silvera-Goldman (SG) empirical two-body pair potentials.^{62,66} The SG term includes a repulsive three-body correction term, while the Buck potential does not. In 2006, Operetto and Pederiva performed a similar DMC simulation over a wider density range. They predicted the equation of state (EOS) of solid *para*- H_2 and confirmed that its preferred crystal structure is *hcp*, matching the experimental consensus.⁴² Using path-integral Monte Carlo (PIMC) calculations, Osychenko *et al.* investigated the onset of superfluidity of glassy *para*- H_2 .⁵⁴ Their research highlighted the sensitivity of computational results to the choice of potential energy surface (PES), and emphasized the need for more accurate PESs for *para*- H_2 that better account for subtle quantum effects. For example, understanding the influence of finite size effects is paramount to accurately reproducing high-resolution spectroscopic measurements.

Recently, by solving a rovibrational Hamiltonian for *para*- H_2 monomers based on the 6D Hinde $\text{H}_2\text{--H}_2$ dimer PES,²⁵ Schmidt *et al.* reported first-principles adiabatic hindered rotor (AHR) reduced-dimension potentials for *para*- $\text{H}_2\text{--para-H}_2$ and *ortho*- $\text{D}_2\text{--ortho-D}_2$, for both the internal vibrational ground and excited states.³¹ These potentials were able to reproduce the experimentally observed infrared vibrational shifts for clusters of various sizes.¹⁶ The work presented in this paper uses these *para*- $\text{H}_2\text{--para-H}_2$ PES’s, which we refer to as the Faruk-Schmidt-Hinde (FSH) potential. We should note that, because information about the ro-vibrational states of the *para*- H_2 monomers has been averaged out of the aforementioned empirical Buck and SG potentials, they cannot predict vibrational band shifts.

Also presented in this paper are results from a PIMC program created to simulate a solid *para*- H_2 crystal using both face-centred cubic (*fcc*) and *hcp* structures.⁶⁷ For the EOS at $T = 4.2$ K, to help eliminate the systematic errors inherent to PIMC simulations, the number of time slices in the simulation and the number of *para*- H_2 molecules inside

the periodic boundary cell were extrapolated to infinity. Using this EOS we computed the pressure as a function of density and the compressibility as a function of pressure. We also computed the infrared vibrational frequency shift as a function of density. This was made possible due to the aforementioned work of Schmidt *et al.*, whose PES includes intramolecular stretching. To investigate the temperature dependence of the EOS and vibrational matrix shift, we also perform several simulations using a finite number of time slices and particles at several different isotherms.

The remainder of this paper is organized as follows: In Sec. 2.2 we describe our theoretical formalism of the solid *para*-H₂ system and the PIMC methodology. In Sec. 2.3 we provide an overview of our numerical approach to the elimination of systematic errors in our PIMC simulations. We present and discuss the results in Sec. 2.4, and end with concluding remarks in Sec. 2.5.

2.2 Theory

2.2.1 Hamiltonian, pairwise approximation, and symmetry

A lot of effort has gone into studying the quantum nature of solid *para*-H₂. However, because the bulk effects of solid *para*-H₂ only appear for large numbers of particles, exact calculations of these properties are intractable. We can make calculations more manageable by using a series of adiabatic approximations in a Born-Oppenheimer-like (BOL) manner, similar to common methods used in electronic structure. Coupled with a pairwise approximation, this simple methodology allows us to compute many thermodynamic properties while still using a first-principles PES.

The Hamiltonian for a system of N interacting *para*-H₂ molecules can be expressed as

$$\hat{H}(\mathbf{x}_1, \mathbf{p}_1, \dots, \mathbf{x}_N, \mathbf{p}_N) = \frac{1}{2\mu} \sum_{i=1}^N \mathbf{p}_i^2 + U(\mathbf{x}_1, \dots, \mathbf{x}_N), \quad (2.1)$$

where μ is the mass of each *para*-H₂ molecule, and $\mathbf{p}_i$ and $\mathbf{x}_i$ are the momentum and position of the i^{th} molecule, respectively. Here, we use a pairwise additive potential U as an approximation to the full, many body potential. Originally derived by Faruk *et al.*,¹⁶ this pairwise potential is computed by treating the *para*-H₂–*para*-H₂ separation as the “slow” coordinate in the BOL framework. The resulting Hamiltonian is

$$\hat{h}(R_{12}) = \hat{h}_1 + \hat{h}_2 + V^{6D}(r_1, r_2, \theta_1, \theta_2, \phi; R_{12}), \quad (2.2)$$

where V^{6D} is the *para*-H₂-*para*-H₂ rotor interaction potential produced by Hinde, and $\hat{h}_1$ and $\hat{h}_2$ are the rovibrational Hamiltonians for the individual *para*-H₂ molecules, respectively.²⁵ We then solve Eq. (2.2) at a series of intermolecular spacings $R_{12} = |\mathbf{x}_1 - \mathbf{x}_2|$ for the ground state eigenvalues, which in turn constitute the given pairwise potential. To compute the eigenvalues we expand Eq. (2.2) in a product basis of the molecular eigenfunctions. This yields a basis set of the form

$$|I\rangle = \bigotimes_{i=1}^2 |l_i m_i \nu_i\rangle = |l_1 m_1 \nu_1 l_2 m_2 \nu_2\rangle, \quad (2.3)$$

where $I = \{l_1, m_1, \nu_1, l_2, m_2, \nu_2\}$ is the composite index.

In general, because the individual *para*-H₂ molecules are indistinguishable, the (*para*-H₂)_N system must exhibit exchange symmetry. We know that the Raman modes have A_1 character - i.e. we are only interested in the eigenstates from the totally symmetric irreducible representation. However, because we are using a BOL approximation to separate the centre-of-mass motion from the internal rotation, we cannot explicitly apply the exchange symmetry to our basis. The full exchange operation permutes *all* coordinate indices, including $R_{12} \rightarrow R_{21}$, and because R_{12} is a parameter of Eq. (2.2), it is unaffected by the permutation operator, i.e. $\hat{P}_{12} |l_1 m_1 \nu_1 l_2 m_2 \nu_2\rangle = |l_2 m_2 \nu_2 l_1 m_1 \nu_1\rangle$. This presents a problem because our criterion for the BOL separation is simply energetic ordering, and we cannot guarantee the character of the ground state. To ensure that all the chosen states in the final pairwise potential have the proper A_1 symmetry, we define a projection operator for the totally symmetric irreducible representation,

$$\Delta_+ = \frac{1}{2} (\hat{I} + \hat{P}_{12}), \quad (2.4)$$

and apply it to Eq. (2.2),

$$\hat{h}_+(R_{12}) = \Delta_+^T \hat{h} \Delta_+, \quad (2.5)$$

where Δ_+^T is the transpose of Δ_+ . At first glance, this procedure turns out to be equivalent to starting with a symmetrized basis. However, this is only the case for the pair, since the only parametric exchange term is R_{12} , which is equivalent to R_{21} . In general, this does not hold for an arbitrary R_{ij} , but such an issue is beyond the scope of this paper.

Another quantity of interest is the “vibrational matrix shift”, the change in the ground state energy of the system after the introduction of an internal quantum of vibration. Experiments show that the energy spacing of this vibrational excitation is very large compared to the internal rotational energy level spacing,^{68–70} and thus the problem can be treated adiabatically. By taking the total vibrational quantum to be the sum of the individual

quanta, $\nu_t = \nu_1 + \nu_2$, we can create two surfaces by restricting $\nu_t = 0$ or $\nu_t = 1$. We then compute the ground state energy of Eq. (2.2) as a function of the pairwise distance, under the two vibrational restrictions,

$$\langle I | \hat{h}(R_{ij}) | J \rangle \rightarrow E_0^{\nu_t}(R_{ij}) \rightarrow V^{\nu_t}(R_{ij}). \quad (2.6)$$

We define $V^1(r)$ to be the potential energy function of a *para*-H₂ dimer where one molecule is in the $\nu = 1$ state and the other is in the $\nu = 0$ state, and $V^0(r)$ to be its counterpart where both *para*-H₂ molecule are in the $\nu = 0$ state.¹⁶

We can now express the many particle potential as a sum of the aforementioned pair potentials. For $\nu_t = 0$, this potential is

$$U^0(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{i < j} V^0(|\mathbf{x}_i - \mathbf{x}_j|). \quad (2.7)$$

The expression for the excited potential surface is more complicated. Our formalism only allows the placement of the vibrational quantum at a specific site, and to account for this limitation is a non-trivial combinatorics problem. Let ${}^N U_n^1$ be the excited potential energy for a system of N particles, where the vibrational excitation is located on particle n . For $N = 2$ the solution is trivial - we simply recover the excited pairwise curve

$${}^2 U_1^1(\mathbf{x}_1, \mathbf{x}_2) = {}^2 U_2^1(\mathbf{x}_1, \mathbf{x}_2) = V^1(R_{12}). \quad (2.8)$$

For $N = 3$, there are three possible locations for the excitation,

$$\begin{aligned} {}^3 U_1^1(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) &= V^1(R_{12}) + V^1(R_{13}) + V^0(R_{23}) \\ {}^3 U_2^1(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) &= V^1(R_{12}) + V^0(R_{13}) + V^1(R_{23}) \\ {}^3 U_3^1(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) &= V^0(R_{12}) + V^1(R_{13}) + V^1(R_{23}). \end{aligned} \quad (2.9)$$

Notice that the symmetric combination of the three expressions above, ${}^3 U^1 = \frac{1}{3}({}^3 U_1^1 + {}^3 U_2^1 + {}^3 U_3^1)$, has repeated terms that need to be dealt with. For a general N , the excited pairwise potential can be expressed as

$${}^N U^1(\mathbf{x}_1, \dots, \mathbf{x}_N) = c^0 \sum_{i < j} V^0(R_{ij}) + c^1 \sum_{i < j} V^1(R_{ij}), \quad (2.10)$$

where c^0 and c^1 are the fraction of terms in the ground and excited state sums respectively, and $c^0 + c^1 = 1$. We see by examining Eq. (2.9) that for $N = 3$, the symmetric combination yields doubles of each excited term. This is also true for $N \geq 4$, because the term $V^1(R_{ij})$ appears in both ${}^N U_i^1$ and ${}^N U_j^1$. Thus we conclude that $c^1 = 2/N$, and by extension,

$c^0 = (N-2)/N$. The final expressions for the ground state and excited state Hamiltonians for a (*para*-H₂)_N cluster can be written, respectively, as

$$\hat{H}^0(\mathbf{x}_1, \mathbf{p}_1, \dots, \mathbf{x}_N, \mathbf{p}_N) = \frac{1}{2\mu} \sum_{i=1}^N \mathbf{p}_i^2 + \sum_{i<j} V^0(R_{ij}), \quad (2.11)$$

and

$$\hat{H}^1(\mathbf{x}_1, \mathbf{p}_1, \dots, \mathbf{x}_N, \mathbf{p}_N) = \frac{1}{2\mu} \sum_{i=1}^N \mathbf{p}_i^2 + \frac{N-2}{N} \sum_{i<j} V^0(R_{ij}) + \frac{2}{N} \sum_{i<j} V^1(R_{ij}). \quad (2.12)$$

2.2.2 Path Integral Monte Carlo method

Given a temperature $\beta = 1/k_B T$ and the density operator $e^{-\beta\hat{H}}$, the canonical average of a quantum mechanical operator $\hat{O}$ is expressed using the PIMC method as

$$\begin{aligned} \langle \hat{O} \rangle_\beta &= \frac{1}{Z} \text{Tr} \{ \hat{O} e^{-\beta\hat{H}} \} \\ &= \frac{1}{Z} \int d\mathbf{q} d\mathbf{q}' \langle \mathbf{q} | \hat{O} | \mathbf{q}' \rangle \langle \mathbf{q}' | e^{-\beta\hat{H}} | \mathbf{q} \rangle, \end{aligned} \quad (2.13)$$

where $Z = \text{Tr}\{e^{-\beta\hat{H}}\}$ is the partition function. The translational coordinates of the *para*-H₂ molecules are denoted collectively as $\mathbf{q} = (\mathbf{r}_0, \{\mathbf{r}_{i>0}\})$. A modified version of the finite temperature MoRiBS-PIMC code developed by T. Zeng *et al.*⁷¹ was used to compute the density matrix and obtain the expectation value of the operator $\hat{O}$. The integral is discretized in imaginary time,

$$\langle \mathbf{q}' | e^{-\beta\hat{H}} | \mathbf{q} \rangle = \int \prod_{i=2}^P d\mathbf{q}_i \times \prod_{i=1}^P \langle \mathbf{q}_i | e^{-\tau\hat{H}} | \mathbf{q}_{i+1} \rangle, \quad (2.14)$$

where $\tau = \beta/P$ is the imaginary time step³⁴ and the individual time steps are labeled by $i = 1$ to P . In addition, the boundary conditions $\mathbf{q}_1 = \mathbf{q}'$ and $\mathbf{q}_{P+1} = \mathbf{q}$ are applied to Eq. (2.14). Bose exchange is practically absent in the density ranges explored and is not included in the simulations - its inclusion barely changes the calculated quantities while increasing the amount of computation time by roughly 40 %.

The vibrational transition frequency of the chromophore molecule changes when embedded in the *para*-H₂ solid. Define ΔE_{embed} and ΔE_{free} to be the transition frequencies

of an embedded and free molecule, respectively. The shift of the vibrational band origin of the dopant in the *para*-H₂ solid relative to that of the free chromophore molecule is

$$\Delta\nu = \Delta E_{\text{embed}} - \Delta E_{\text{free}}. \quad (2.15)$$

We can also evaluate Eq. (2.15) as the difference between $E_{\text{matrix}}^{\nu_t=1}$, the total energy of a *para*-H₂ lattice in which one *para*-H₂ molecule is in its $\nu = 1$ state while all the others are in the $\nu = 0$ state, and $E_{\text{matrix}}^{\nu_t=0}$, the total energy in which all *para*-H₂ molecules are in the $\nu = 0$ state,

$$\Delta\nu = E_{\text{matrix}}^{\nu_t=1} - E_{\text{matrix}}^{\nu_t=0}. \quad (2.16)$$

In other words, in the pure *para*-H₂ solid, the chromophore molecule is *para*-H₂ itself. Just as in pure *para*-H₂ clusters,¹⁶ the “chromophore” *para*-H₂ molecule in the solid cannot be singled out from the other *para*-H₂ molecules in realistic simulations, and the excitation $\nu_t = 1$ can diffuse by exchange.

The Hamiltonian of solid *para*-H₂ in the excited state $\nu_t = 1$ can be approximated¹⁶ as

$$\hat{H}^1 = \hat{H}^0 + \Delta\hat{V}, \quad (2.17)$$

where $\Delta\hat{V} = \hat{V}^1 - \hat{V}^0$ is the difference between the total potential energy of solid *para*-H₂ in the $\nu_t = 1$ state (where a single *para*-H₂ molecule is in the $\nu = 1$ state) and the $\nu_t = 0$ state (where all molecules are in the $\nu = 0$ state). Using first-order perturbation theory,^{16,72,73} we can write the vibrational frequency shift as

$$\Delta\nu = \langle \Delta\hat{V} \rangle. \quad (2.18)$$

We determine this quantity using the radial distribution function $g(r)$ of the *para*-H₂ molecules in the solid. However, it is easier to directly retrieve from the PIMC simulation the weight distribution function $h(r)$, where $h(r)dr = 4\pi r^2 \rho g(r)dr$, which is then normalized such that $\int dr h(r) = 1$. We can evaluate the vibrational frequency shift in a simulation of N *para*-H₂ molecules as

$$\begin{aligned} \Delta\nu &= 4\pi\rho(N-1) \int dr r^2 g(r) \Delta V(r), \\ &= (N-1) \int dr h(r) \Delta V(r), \end{aligned} \quad (2.19)$$

where $\Delta V(r) = V^1(r) - V^0(r)$.

The use of periodic boundary conditions allows us to simulate *para*-H₂ crystals of various densities for both the *fcc* and *hcp* structures. Each simulation is performed by constructing

a system of $D_1 \times D_2 \times D_3$ elementary cells for a given lattice type. The integers D_1 , D_2 , and D_3 are chosen such that the ratio of the three sides of the system to each other is as close as possible to unity for cells of manageable sizes. Cells of size $5 \times 3 \times 3$, $7 \times 4 \times 4$, $8 \times 5 \times 5$, and $10 \times 6 \times 6$ were constructed for the *hcp* structure, and cells of size $4 \times 4 \times 4$, $5 \times 5 \times 5$, $6 \times 6 \times 6$, and $7 \times 7 \times 7$ were used for the *fcc* structure. Each elementary cell contains 4 *para*-H₂ molecules, for both the *hcp* and *fcc* structures. The energy and vibrational band shift measurements from these systems of different finite sizes are then extrapolated to obtain their counterparts for a system of infinite size. This extrapolation is discussed in detail in Secs. 2.3.1 and 2.3.2.

2.3 Numerics

2.3.1 Extrapolating to an infinite number of time steps

The use of finitely many time slices in calculations is the primary source of systematic errors in any PIMC simulation. To compensate for this error, the calculation can be repeated several times for different numbers of time slices. We can then use the expression

$$\langle \hat{O} \rangle_{N,\tau} = \langle \hat{O} \rangle_{N,0} + B_2 \tau^2 + B_4 \tau^4 \quad (2.20)$$

to extrapolate the PIMC data to the limit $\tau \rightarrow 0$.⁷⁴ Here, $\langle \hat{O} \rangle_{N,\tau}$ can be either the energy E or the band origin shift $\Delta\nu$ for a system of N *para*-H₂ molecules simulated with the time step size τ . The parameters $\langle \hat{O} \rangle_{N,0}$, B_2 and B_4 are determined by applying Eq. (2.20) to several $(\langle \hat{O} \rangle_{N,\tau}, \tau)$ pairs, and $\langle \hat{O} \rangle_{N,0}$ is the extrapolated value. Most previously reported simulations of doped rare gas clusters^{73,75} started with extensive coverage tests to choose a value of τ that is “small enough” such that $\langle \hat{O} \rangle_{N,\tau}$ no longer changes appreciably by decreasing τ . The working calculations were then performed using only the chosen τ . To see why the extrapolation $\tau \rightarrow 0$ is important, we show the energy per particle ϵ for an *hcp* crystal of $N = 180$ *para*-H₂ molecules for different numbers of time slices in Fig. (2.1). One calculation of interest is to determine the equilibrium density of solid *para*-H₂, and we see in Fig. (2.1) that the equilibrium density of the EOS curve changes with τ . In Fig. (2.2) the energy per particle is plotted as a function of the imaginary time step for two densities, and fit using Eq. (2.20). For both examples, the extrapolated value is roughly 0.5 cm⁻¹ greater than the value found using $P = 192$ time slices. We will see in Sec. 2.4.2 that this correction is an order of magnitude greater than the difference between the ground state energy per particle of the *hcp* and *fcc* crystals. It is therefore necessary to extrapolate τ to the limit $\tau \rightarrow 0$ rather than to rely on a “small enough τ ” approximation.

Most observed and predicted equations of state were obtained at $T = 4.2$ K,^{9,60–63} and this is the temperature used for most of our simulations. We perform simulations with the numbers of time slices $P = \{64, 80, 96, 128, 160, 192\}$, and then use Eq. (2.20) to extrapolate the energy and band origin shift measurements to $\tau \rightarrow 0$. To study the temperature dependence of the EOS and vibrational matrix shift, an *hcp* periodic cell of $N = 448$ *para*-H₂ molecules is simulated at temperatures $T = \{2.10$ K, 4.20 K, 5.04 K, 6.30 K}, with the number of time slices adjusted such that τ is the same for each simulation.

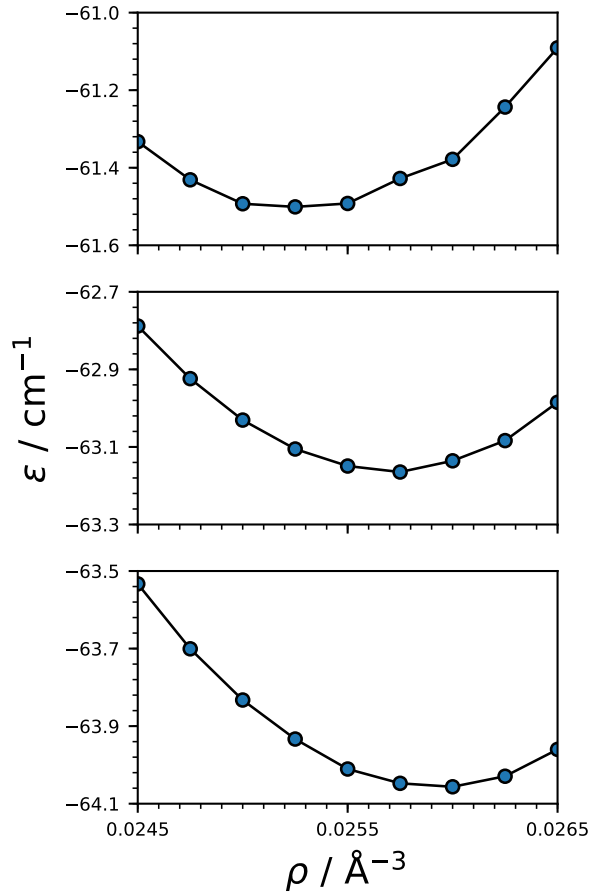


Figure 2.1: Energy per particle ϵ as a function of density ρ obtained for numbers of time steps $P = 192$ (top), $P = 96$ (centre), and $P = 80$ (bottom) in the PIMC simulation, where $\tau = \beta/P$. Each simulation was performed using $N = 180$ *para*-H₂ molecules. The density for which the EOS is at its minimum changes with P .

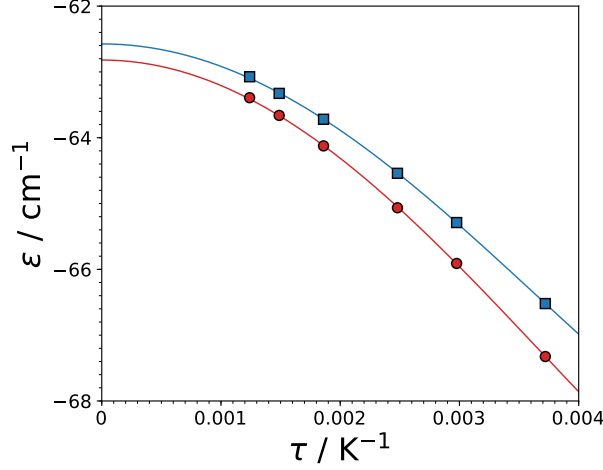


Figure 2.2: Energy per particle ϵ as a function of imaginary time step size τ of an *hcp* crystal with $N = 1440$ *para*-H₂ molecules and a density of $\rho = 0.0255 \text{ \AA}^{-3}$ (blue squares) and $\rho = 0.0245 \text{ \AA}^{-3}$ (red circles). The value of ϵ at $\tau = 0$ is found using Eq. (2.20).

2.3.2 Extrapolating to an infinite number of particles

The use of a finite number of *para*-H₂ molecules introduces an additional source of systematic error to the periodic boundary condition methodology, called the finite size effect.^{76,77} To account for this error we calculate physical observables for systems of different sizes and extrapolate these measurements to $N \rightarrow \infty$.

In Fig. (2.3) we see that, after having performed the imaginary time extrapolation $\tau \rightarrow \infty$, both the energy per particle $\epsilon = E/N$ and the band origin shift $\Delta\nu$ are linear functions of $1/N$. To explain why, consider a periodic cell with an edge length of $2L$. Due to its finite size, there are additional *para*-H₂ molecules located beyond a shell of radius $r > L$, where r is the distance between the centre of the periodic cell and the surface of the shell. At long ranges, the function $V^0(r)$ is dominated by the leading term of the dispersion interaction, which is proportional to r^{-6} .²⁹ Also, the weight distribution function $h(r)$ is determined by the density of *para*-H₂ on the surface area of the shell, so we can write $h(r) \sim r^2$. As given by the tail correction,⁷⁸ the contribution to ϵ from the additional *para*-H₂ molecules beyond the shell is

$$\Delta\epsilon = 2\pi\rho \int dr h(r)V^0(r) \sim \int_L^\infty dr r^2 \frac{1}{r^6} \sim \frac{1}{L^3}. \quad (2.21)$$

Because the density is uniform, the volume of the periodic cell is proportional to the number of *para*-H₂ molecules within, and we can write $\Delta\epsilon \sim L^{-3} \sim N^{-1}$. A similar calculation shows that the contribution from the additional *para*-H₂ molecules to the vibrational matrix shift is also inversely proportional to N . Thus to account for the finite size effect, simulations are performed for different numbers of particles and the PIMC data is extrapolated to the limit $N \rightarrow \infty$ using

$$\langle \hat{O} \rangle_{N,0} = \langle \hat{O} \rangle_{\infty,0} + \frac{C}{N}. \quad (2.22)$$

In Eq. (2.22), $\langle \hat{O} \rangle_{N,0}$ is the observable calculated using a system of N *para*-H₂ molecules with the extrapolation in Eq. (2.20) already applied, $\langle \hat{O} \rangle_{\infty,0}$ is its counterpart with the extrapolation $N \rightarrow \infty$ applied, and C is a constant. The values of ϵ and $\Delta\nu$ obtained with Eq. (2.22) are lower because more *para*-H₂-*para*-H₂ interaction pairs are included in a region where the potential is attractive.

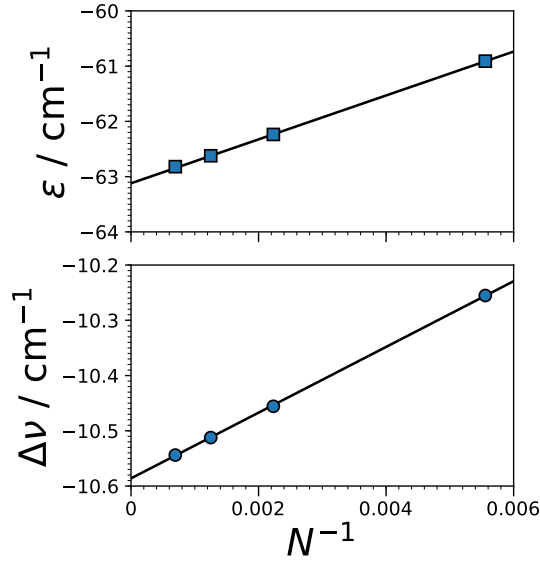


Figure 2.3: The energy per particle (ϵ , top) and vibrational frequency shift ($\Delta\nu$, bottom) at $\tau = 0$ as a function of the number of *para*-H₂ molecules N , for an *hcp* crystal at density of $\rho = 0.0255 \text{ \AA}^{-3}$. Using Eq. (2.21), we find both these quantities for an infinite lattice of *para*-H₂ molecules.

We see from Eq. (2.21) that the scaling of ϵ and $\Delta\nu$ with N is determined by the leading term of the long-range interaction - if the leading term is proportional to r^{-m} , then ϵ and $\Delta\nu$ change linearly with $N^{1-m/3}$.

2.4 Results and Discussion

2.4.1 The global EOS

To fit the calculated values of the energy per particle we use a Murnaghan-type curve⁷⁹

$$\epsilon(\rho) = a + b\rho + c\rho^\gamma, \quad (2.23)$$

where ϵ is the energy per particle, ρ is the density, and (a, b, c, γ) are the fit parameters. The EOS for an *hcp* crystal, after having performed the $\tau \rightarrow 0$ and $N \rightarrow \infty$ extrapolations, is presented in Fig. (2.4) on the density interval $0.02 \text{ \AA}^{-3} < \rho < 0.04 \text{ \AA}^{-3}$. Alongside it we also include several previously reported EOS curves.⁴² We can see from Fig. (2.4) that the resulting curve fits the measured data satisfyingly.

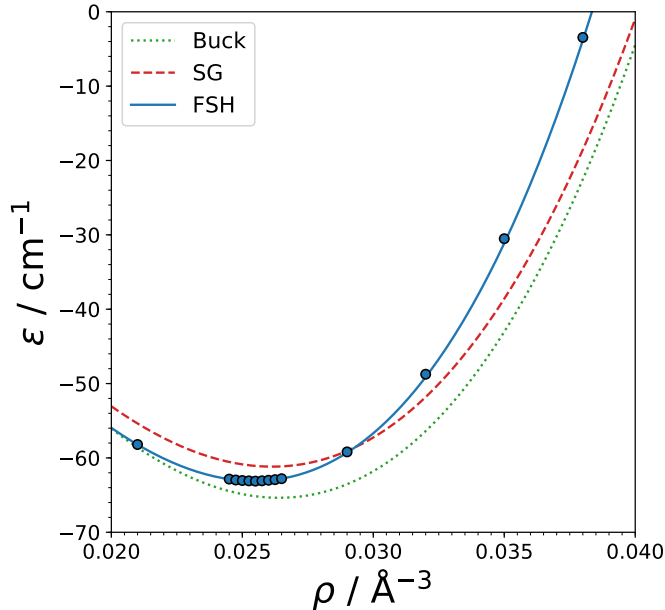


Figure 2.4: The EOS curves for an *hcp* lattice of *para*-H₂ molecules using the FSH potential at $T = 4.2$ K (solid blue, circles), the Buck potential at $T = 0$ K (green, dotted), and the SG potential at $T = 0$ K (red, dashed). The FSH potential provides the lowest equilibrium density of the three.

On this density interval, the EOS's for *fcc* and *hcp* crystals are nearly indistinguishable, and so the former is omitted.⁴² The fit parameters a , b , c , and γ are provided for the curves

in Fig. (2.4) in Table 2.1. From this fit we can determine the equilibrium density ρ_0 , the distance between two nearest neighbours R_0 , and the equilibrium energy per particle ϵ_0 , which we summarize in Table 2.2. The nature of these EOS curves is closely related to the position and depth of their corresponding potential energy curves, shown in Fig. (2.5). For example, we can see in Fig. (2.4) that the equilibrium density of the EOS curve of the FSH potential is lower than those of the SG and Buck potentials. This is because its minimum is deeper and lies at a greater distance than those of the other potentials. The FSH potential also has the greatest repulsive wall of the three, which is why, at high densities, its EOS curve is greater than those of the SG and Buck potentials. In the same fashion, the Buck potential has the lowest repulsive wall, which is consistent with the observation that its EOS lies below the others at high densities.

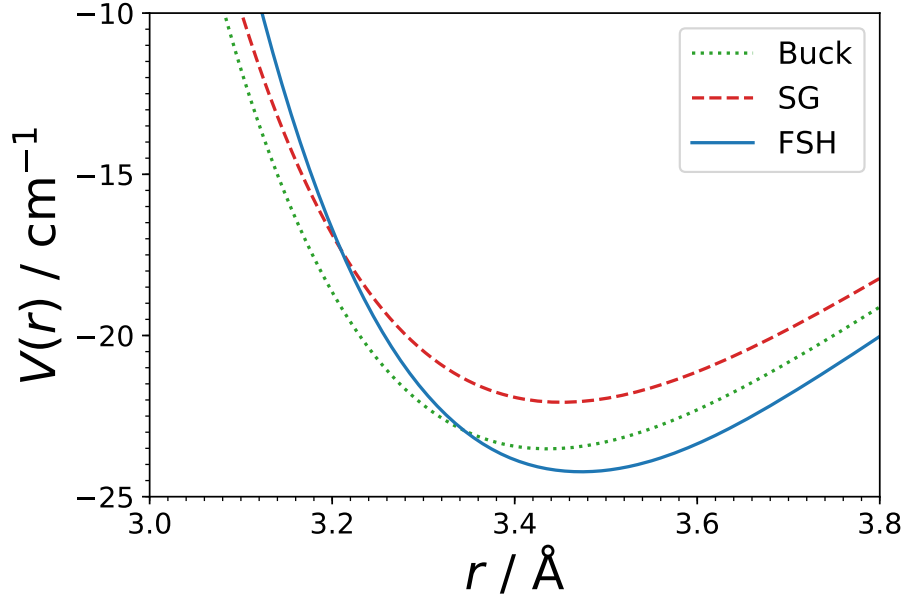


Figure 2.5: The *para*-H₂–*para*-H₂ potential energy curves for the FSH potential (blue, solid), the Buck potential (green, dotted) and the SG potential (red, dashed). The FSH potential has the greatest repulsive wall when $R < 3.2$ Å, and so it predicts a lower equilibrium density than the other two potentials.

We should also note that our simulations were performed at $T = 4.2$ K with the results extrapolated to $N \rightarrow \infty$. The DMC simulations for the Buck and SG potentials were performed at absolute zero and for a system of $N = 108$ *para*-H₂ molecules with a tail correction applied.⁴²

Table 2.1: Coefficients a , b , c , and γ obtained by fitting Eq. (2.23) to the (ρ, ϵ) pairs obtained from our PIMC simulation and from Ref. [42]. The density ρ is given in $\text{\AA}^{-3}$, and the energy per particle ϵ is given in cm^{-1} . Note that our results corrected for the finite size effect using Eq. (2.22) while in Ref. [42] the tail correction was performed.

	a	b	c	γ
FSH(fcc , $T = 4.2$ K)	28.1712	-4791.96	6.32311×10^7	3.96
FSH(hcp , $T = 4.2$ K)	27.5055	-4762.14	6.29562×10^7	3.96
Buck(hcp , $T = 0$ K)	36.4411	-5380.89	1.49110×10^7	3.52809
SG(hcp , $T = 0$ K)	33.4464	-5058.51	1.38902×10^7	3.51769

Table 2.2: The equilibrium density ρ_0 (in $\text{\AA}^{-3}$), the distance between nearest neighbours R_0 (in $\text{\AA}$), and the equilibrium energy per particle ϵ_0 (in cm^{-1}), of fcc and hcp *para*-H₂ crystals, obtained in our PIMC simulations and from Ref. [42]. Also included are experimental results from Refs. [9] and [42]. Note that our results corrected for the finite size effect using Eq. (2.22) while in Ref. [42] the tail correction was performed.

	fcc			hcp		
	ρ_0	R_0	ϵ_0	ρ_0	R_0	ϵ_0
FSH($T = 4.2$ K)	0.02547	3.815	-63.062	0.02546	3.816	-63.102
Buck($T = 0$ K)	0.02647	3.766	-64.99	0.02640	3.770	-65.36
SG($T = 0$ K)	0.02606	3.786	-60.82	0.02613	3.783	-61.18
exp($T = 0$ K)	-	-	-	0.02600 ⁹	3.789	-62.5 ⁴²

2.4.2 EOS near equilibrium density

The EOS curves of solid *para*-H₂ in both its fcc and hcp forms near the equilibrium density are shown in Fig. (2.6). All the (ρ, ϵ) points here were obtained by extrapolating both τ and N . We can see that the values of ϵ for the hcp crystal are always lower than those of the fcc crystal by about 0.04 cm^{-1} , making the hcp crystal the more stable of the two. This result agrees with experiments indicating that the hcp lattice is the preferred structure of a *para*-H₂ crystal around $T = 4.2$ K.^{9,64}

To study the temperature dependence of the EOS, we simulate the $7 \times 4 \times 4$ hcp lattice at temperatures $T = \{2.10 \text{ K}, 4.20 \text{ K}, 5.04 \text{ K}, 6.30 \text{ K}\}$, with the number of time slices adjusted

such that $\tau \approx 0.00248 \text{ K}^{-1}$ for each simulation. In other words, no N or τ extrapolation was performed for these simulations.

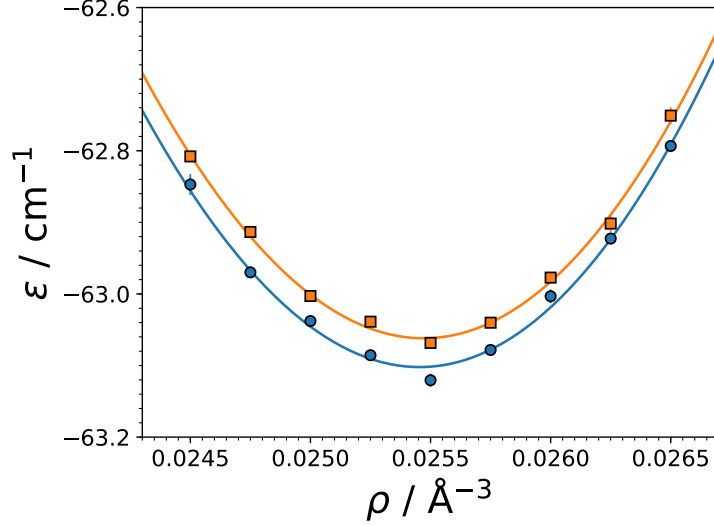


Figure 2.6: The EOS of solid *para*-H₂ near the equilibrium density for the *hcp* crystal (blue circles) and the *fcc* crystal (orange squares) after performing the imaginary time step extrapolation $\tau \rightarrow 0$ and the system size extrapolation $N \rightarrow \infty$. The *hcp* lattice structure is the more stable of the two.

Table 2.3: The equilibrium density ρ_0 (in $\text{\AA}^{-3}$), the distance between nearest neighbors R_0 (in $\text{\AA}$), and the equilibrium energy per particle ϵ_0 (in cm^{-1}), of *hcp para*-H₂ crystals at different temperatures. All simulations here are performed using the FSH potential, with $\tau \approx 0.00254 \text{ K}^{-1}$ and $N = 448$ *para*-H₂ molecules. The values are obtained using the fit Eq. (2.23), while fixing $\gamma = 3.96$.

	ρ_0	R_0	ϵ_0
$T = 2.10 \text{ K}$	0.02589	3.794	-64.513
$T = 4.20 \text{ K}$	0.02590	3.794	-64.506
$T = 5.04 \text{ K}$	0.02590	3.794	-64.496
$T = 6.30 \text{ K}$	0.02592	3.792	-64.468

The results in Table 2.3 and Fig. (2.7) indicate that as the temperature decreases, the EOS isotherm becomes deeper and the value of ϵ_0 decreases. However, the changes

in the EOS isotherm become less substantial at lower temperatures. For example, the change in the EOS curve from $T = 4.20$ K to $T = 2.10$ K is much smaller than the change from $T = 6.30$ K to $T = 5.04$ K, despite being a larger decrease in temperature. These observations agree with the general temperature invariance of the EOS curve of solid *para*-H₂ below $T = 4.2$ K, mentioned in Ref. [9].

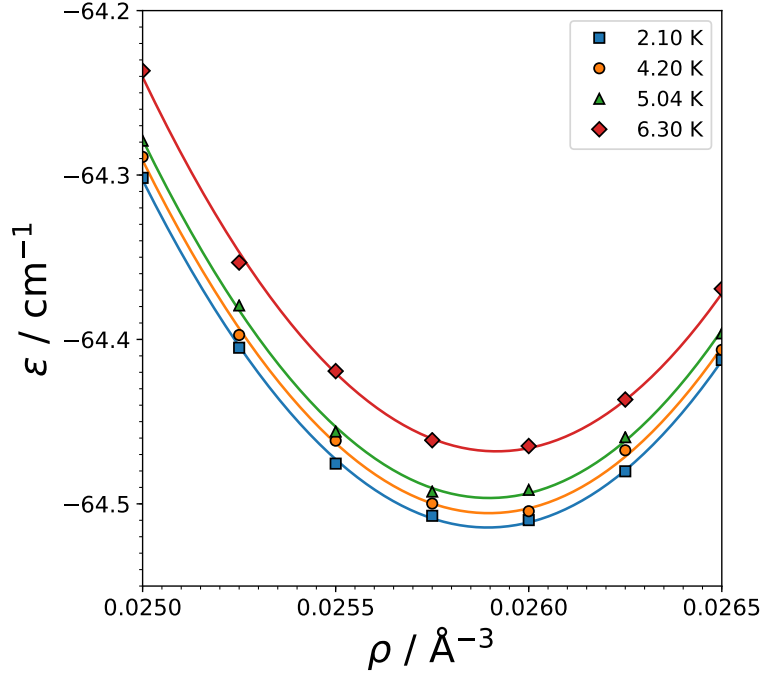


Figure 2.7: The EOS for a *para*-H₂ *hcp* crystal at temperatures of $T = 2.10$ K (blue squares), $T = 4.20$ K (orange circles), $T = 5.04$ K (green triangles), and $T = 6.30$ K (red diamonds), with $N = 448$ and P adjusted such that in each simulation, $\tau \approx 0.00248$ K⁻¹. The energy per particle at equilibrium decreases with temperature, but the equilibrium density itself remains unchanged.

2.4.3 Pressure as a function of density

The pressure as a function of density can be calculated from the EOS using⁴²

$$P = \rho^2 \left. \frac{\partial \epsilon}{\partial \rho} \right|_T. \quad (2.24)$$

In Fig. (2.8) we see the computed pressure curves on the density interval $0.02 \text{ \AA}^{-3} < \rho < 0.04 \text{ \AA}^{-3}$ for the FSH, Buck, and SG potentials⁴² and experimental data.^{9,63} The inset in Fig. (2.8) presents this data near the equilibrium density. The difference between the pressure curves for the *hcp* and *fcc* crystals is negligible on both density ranges, so only the results for the former are plotted. This is expected from Table 2.1, which shows that the values of b , c and γ for the *hcp* and *fcc* fits using Eq. (2.23) are nearly the same.

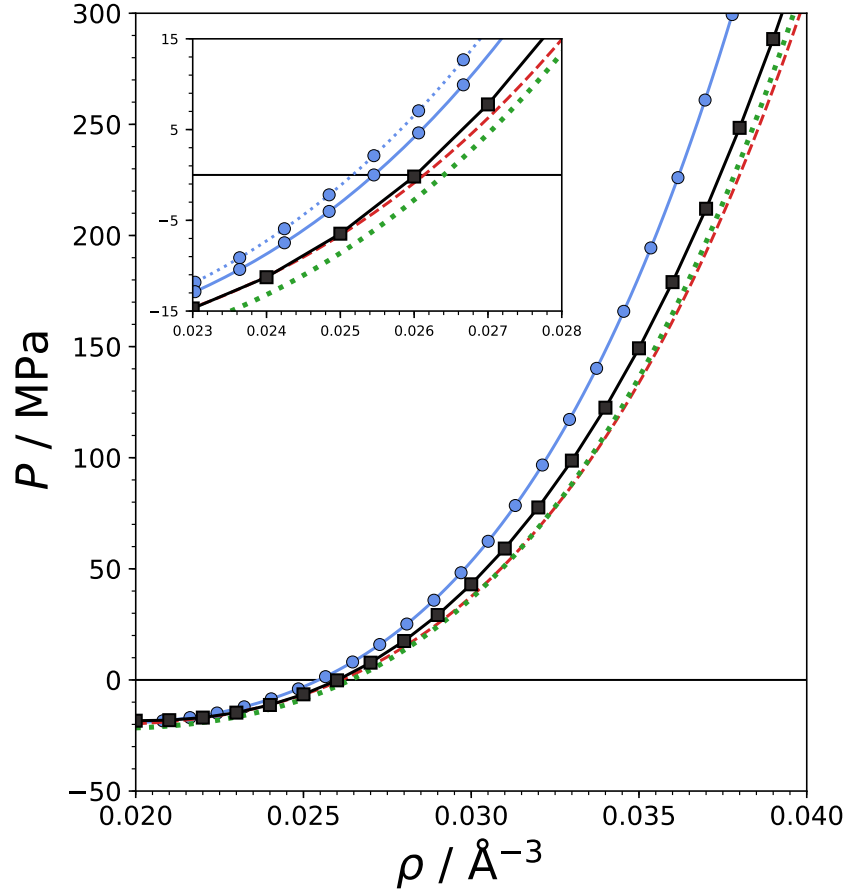


Figure 2.8: The pressure for an *hcp* crystal of *para*-H₂ molecules as a function of density. Plotted are the FSH potential at $T = 4.2 \text{ K}$ extrapolated to $N \rightarrow \infty$, (solid blue line with circles), the Buck potential at $T = 0 \text{ K}$ (dotted green line), the SG potential at $T = 0 \text{ K}$ (dashed red line) and the experimental results at $T = 4.2 \text{ K}$ (solid black line with squares). Also included within the inset is the FSH potential at $T = 4.2 \text{ K}$ and $N = 180$ (dotted blue line with circles).

We see from the tail correction Eq. (2.21) that the decrease in ϵ from the extrapolation of $N \rightarrow \infty$ becomes greater in magnitude at higher densities. This correction decreases the first derivative of the energy curve in Fig. (2.4) and decreases the pressure at higher densities. The inset of Fig. (2.8) shows that for the FSH potential, extrapolating the system size from $N = 180$ to $N \rightarrow \infty$ results in better agreement with experimental observations.

The P vs ρ curves are temperature-dependent.⁶³ This may explain why our curve in the inset of Fig. (2.8) lies above those obtained by Operetto *et al.*⁴² In addition, our results show greater discrepancies compared to experimental observations than do the curves obtained from Buck⁶⁶ and SG.⁶² This is because our simulation was performed using a first-principles pair potential based on the interaction in a *para*-H₂ dimer, with neither the empirical correction in the repulsive region nor the inclusion of three-body potential terms. At short ranges, the three-body *para*-H₂ potential is attractive.¹¹ Its inclusion to the FSH potential will lower the potential energy and decrease the pressure of the solid at higher densities.

2.4.4 Compressibility of solid parahydrogen

From the EOS of solid *para*-H₂ we can also recover its compressibility,⁴² given by

$$\kappa = \frac{1}{\rho} \frac{\partial \rho}{\partial P} \Big|_T. \quad (2.25)$$

Table 2.4: Coefficients a , b , c , and γ obtained by fitting Eq. (2.23) to the $(\rho, \Delta\nu)$ pairs obtained from our PIMC simulations. Also included is $\Delta\nu_0$ (in cm⁻¹), the vibrational matrix shift at equilibrium density ρ_0 . Both the simulation and experiment results are recorded for $T = 4.2$ K.

	a	b	c	γ	$\Delta\nu_0$
<i>fcc</i> , $N \rightarrow \infty$	8.0384	-1077.01	1.60252×10^6	3.30	-10.59
<i>hcp</i> , $N \rightarrow \infty$	8.0160	-1075.66	1.60036×10^6	3.30	-10.59
<i>hcp</i> , $N = 180$	7.7383	-1053.37	1.60765×10^6	3.30	-10.26
exp ⁴⁰	-	-	-	-	-11.43

In Fig. (2.9) we show the the compressibility as a function of pressure recovered from simulations using the FSH, Buck, and SG potentials, alongside experimental results provided in Ref. [80]. We see that the FSH potential is the most successful of the three at

predicting the compressibility at zero pressure, and that they each provide compressibilities that are too low as the pressure increases. In the case of the FSH potential, this deviation occurs because it is a first-principles pair interaction, and does not incorporate many-body effects. The three-body *para*-H₂ potential is attractive at short ranges,¹¹ and its inclusion to a simulation alongside the FSH potential will make solid *para*-H₂ more stable at higher densities and increase its compressibility.

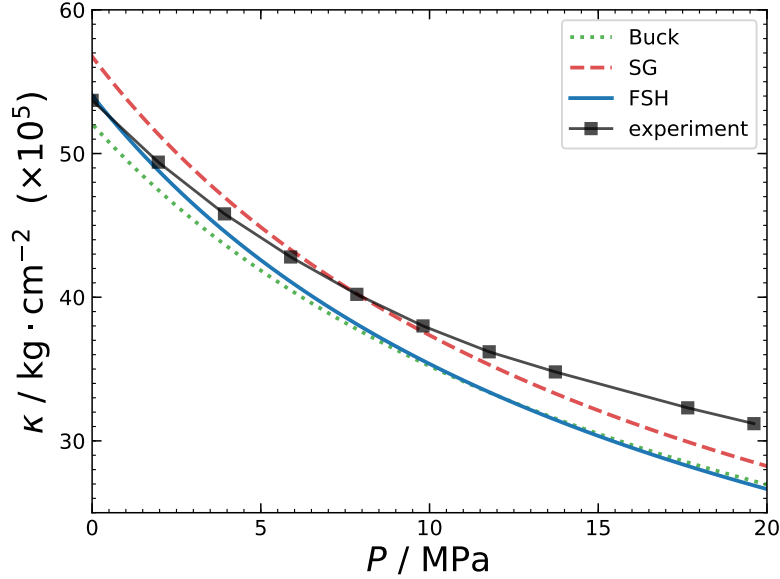


Figure 2.9: The compressibility κ for an *hcp* crystal of *para*-H₂ molecules as a function of pressure P , calculated using Eq. (2.25). Plotted are the FSH potential at $T = 4.2$ K (solid blue line), the Buck potential at $T = 0$ K (dotted green line), the SG potential at $T = 0$ K (dashed red line) and the experimental results for $T < 10$ K (solid black line with squares).

2.4.5 Vibrational band origin shifts

The calculated vibrational matrix shift $\Delta\nu$ as a function of density ρ for the *hcp* crystal in the density interval $0.02 \text{ \AA}^{-3} < \rho < 0.04 \text{ \AA}^{-3}$ is shown in Fig. (2.10). These $\Delta\nu$ values are calculated after performing both the N and τ extrapolation, and the resulting $(\rho, \Delta\nu)$ pairs are then fit using Eq. (2.23), the same equation used to fit the energy per particle. On the density range of Fig. (2.10), the results for the *hcp* and *fcc* crystals are nearly indistinguishable.

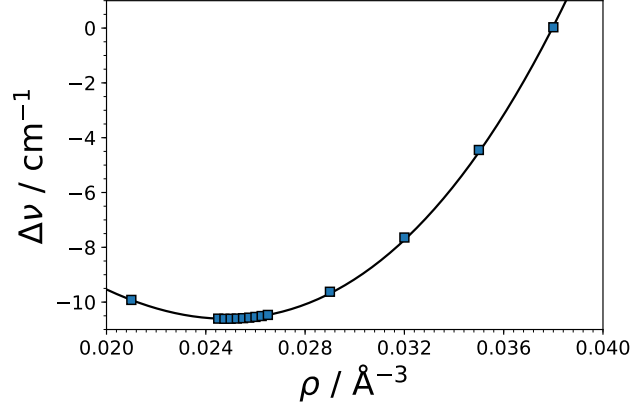


Figure 2.10: The vibrational matrix shift $\Delta\nu$ for an *hcp* crystal as a function of the density ρ , calculated using Eq. (2.19) after performing the extrapolations $\tau \rightarrow 0$ and $N \rightarrow \infty$. The corresponding curve for the *fcc* crystal is nearly indistinguishable from that of the *hcp* crystal on this density range.

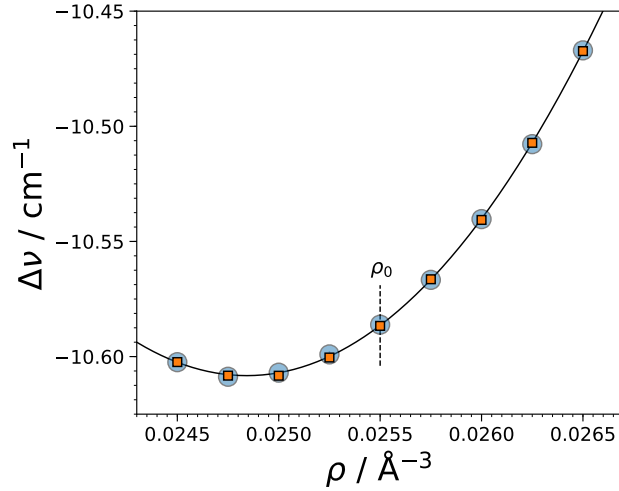


Figure 2.11: The vibrational matrix shift for the *hcp* (blue circles) and *fcc* (orange squares) crystals near the equilibrium density, calculated using Eq. (2.19) after performing the extrapolations $\tau \rightarrow 0$ and $N \rightarrow \infty$. Both lattices provide essentially the same vibrational shift. The minimum value of $\Delta\nu$ is found around $\rho = 0.0248 \text{ \AA}^{-3}$, lower than the equilibrium density $\rho_0 \approx 0.0255 \text{ \AA}^{-3}$ for the EOS indicated by the dashed line.

The minimum of the vibrational frequency shift is located near $\rho = 0.0248 \text{ \AA}^{-3}$, which is noticeably lower than the equilibrium density of $\rho_0 \approx 0.0255 \text{ \AA}^{-3}$ for the EOS curves. We show $\Delta\nu_0$, the vibrational band origin shift at the equilibrium density ρ_0 , alongside the fit parameters in Table 2.4. We see that the value of $\Delta\nu_0$ calculated in this report differs from the experimental value by only 8 %. In Fig. (2.11) we see that, once again, the vibrational matrix shifts for the *fcc* and *hcp* solids are essentially identical near the equilibrium density. We should thus expect that a measurement of $\Delta\nu$ is an inadequate way to determine if a sample of solid *para*-H₂ is in an *fcc* or *hcp* state.

The extrapolation of the system to infinite size is predicted by Eq. (2.19) to lower the vibrational frequency shift, because $V^1(r) < V^0(r)$ at large distances. In Fig. (2.12) we show $\Delta\nu$ vs ρ for an *hcp* lattice with $N = 180$ *para*-H₂ molecules and one having applied the extrapolation $N \rightarrow \infty$. The latter is about 0.3 cm^{-1} lower than the former, moving it closer to the experimental value of -11.43 cm^{-1} .⁴⁰

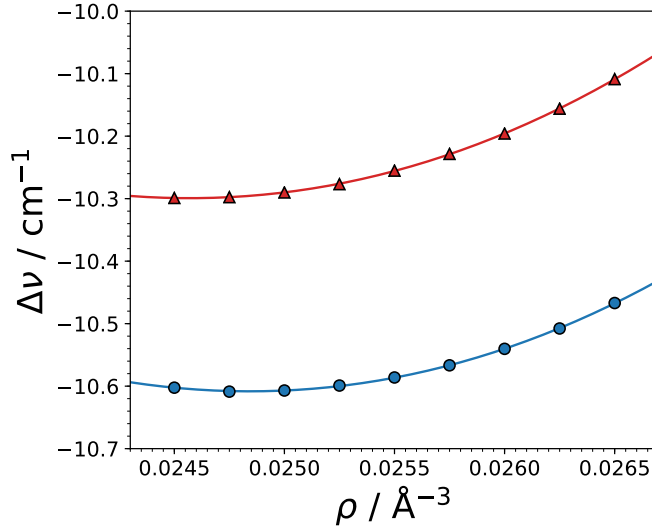


Figure 2.12: The vibrational matrix shift for an *hcp* crystal with $N = 180$ *para*-H₂ molecules (red triangles) and with the extrapolation $N \rightarrow \infty$ (blue circles). Performing the extrapolation decreases the value of $\Delta\nu_0$ from -10.26 cm^{-1} to -10.59 cm^{-1} , and brings it closer to the experimental value of -11.43 cm^{-1} .

The value of $\Delta\nu$ in large-sized ($N > 10$) *para*-H₂ clusters was found to be between -3 cm^{-1} and -4 cm^{-1} ,^{16,59} less than half of its value in solid *para*-H₂ (-11.43 cm^{-1}).⁴⁰ To explain this discrepancy, in Fig. (2.13) we can compare the radial distribution functions

for a cluster of $N = 33$ *para*-H₂ molecules¹⁶ and an *hcp* lattice of *para*-H₂ molecules with periodic boundary conditions applied. The $g(r)$ curves for solid *para*-H₂ for different values of N and τ are indistinguishable on the scale of Fig. (2.13), so the result for $N = 1440$ and $P = 192$ is chosen. Both radial distribution functions have been renormalized such that $4\pi \int dr r^2 \rho g(r) = (N - 1)$ based on the coefficient of Eq. (2.19). According to Eq. (2.19), the value of $\Delta\nu$ is determined by the integral of the product of $\Delta V(r) = V^1(r) - V^0(r)$ and the normalized radial distribution $g(r)$. We see in Fig. (2.13) that the radial distribution function of the *para*-H₂ lattice is greater than that of the *para*-H₂ cluster for all distances. The differing values of $\Delta\nu$ for the *para*-H₂ cluster and *para*-H₂ solid are simply a result of the different numbers of *para*-H₂-*para*-H₂ pairs. This is consistent with what we see in Ref. [16], where the value of $\Delta\nu$ for the cluster becomes more negative as the cluster size increases from 13 to 33.

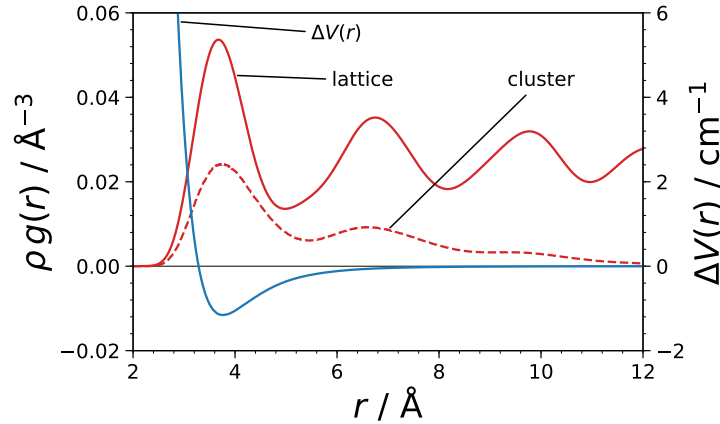


Figure 2.13: The normalized radial distribution $\rho g(r)$ for an *hcp* *para*-H₂ lattice (solid red line) and a cluster of $N = 33$ *para*-H₂ molecules (dashed red line), alongside the potential energy difference $\Delta V(r) = V^1(r) - V^0(r)$ (solid blue line). The magnitude of the first shell of the *para*-H₂ lattice is roughly twice that of the *para*-H₂ cluster in the region where $\Delta V(r)$ is the most negative. Note that while the radial distributions for the lattice and cluster were calculated the same way, the cluster does not have a well-defined density ρ .

In Fig. (2.14) we plot $\Delta\nu$ vs ρ for temperatures $T = \{2.10 \text{ K}, 4.20 \text{ K}, 5.04 \text{ K}, 6.30 \text{ K}\}$, with the number of time slices P adjusted such that $\tau = \beta/P \approx 0.00248 \text{ K}^{-1}$ for each simulation. The curves are fit using Eq. (2.23), and the fit parameters are provided in Table 2.5 next to their corresponding values of $\Delta\nu_0$. Similarly to the EOS isotherms of *para*-H₂ calculated with the FSH potential, the corresponding vibrational matrix shift

isotherms also decrease with temperature, and the changes are nearly negligible below $T = 4.2$ K.

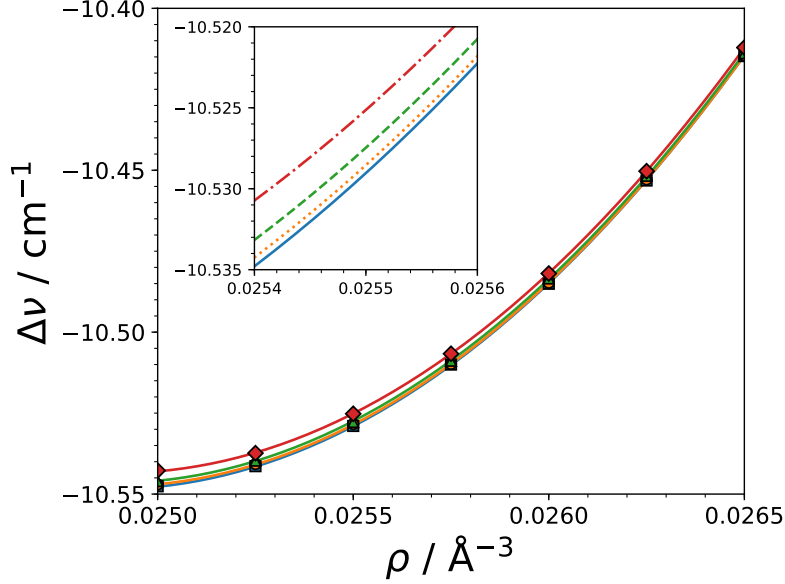


Figure 2.14: The vibrational matrix shift $\Delta\nu$ for a *para*-H₂ *hcp* crystal at temperatures of $T = 2.10$ K (blue squares, solid line in inset), $T = 4.20$ K (orange circles, dotted line in inset), $T = 5.04$ K (green triangles, dashed line in inset), and $T = 6.30$ K (red diamonds, dotted-dashed line in inset), with $N = 448$ and P adjusted such that in each simulation, $\tau = \beta/P \approx 0.00248$ K⁻¹.

Table 2.5: Coefficients a , b , c , and obtained by fitting Eq. (2.23) to the $(\rho, \Delta\nu)$ pairs, for *hcp para*-H₂ crystals at different temperatures. All simulations here are performed using $\tau \approx 0.00254$ K⁻¹ and $N = 448$ *para*-H₂ molecules. Each fit used a fixed exponent parameter $\gamma = 3.30$. Also included is $\Delta\nu_0$ (in cm⁻¹), the vibrational matrix shift at equilibrium density $\rho_0 = 0.02546$ Å⁻³.

	a	b	c	$\Delta\nu_0$
$T = 2.10$ K	7.952	-1067.07	1.58275×10^6	-10.5312
$T = 4.20$ K	7.970	-1067.87	1.58318×10^6	-10.5307
$T = 5.04$ K	7.908	-1064.28	1.57800×10^6	-10.5296
$T = 6.30$ K	8.035	-1070.83	1.58573×10^6	-10.5272

2.5 Conclusion

We have presented extensive PIMC studies of the EOS curves of both *fcc* and *hcp* structured solid *para*-H₂ based on a first-principles AHR effective pair potential for the *para*-H₂ dimer. To avoid the systematic errors associated with the finite size of the simulation cell and the finite value of imaginary time step, the results are extrapolated in the limits $N \rightarrow \infty$ and $\tau \rightarrow 0$. The calculated energy per particle as a function of density is then fit to a Murnaghan-type curve. The resulting EOS curves are compared with previously reported results, and the discrepancies are explained by differences in the two-body *para*-H₂–*para*-H₂ PESs and simulation methodologies. In addition, the equilibrium density obtained and the corresponding equilibrium energy per particle agree very well with experimental observations. The *hcp* structured solid *para*-H₂ crystal is shown to be more stable than the *fcc* structured one. The equilibrium density of the EOS curves was found to not vary with temperature. We also calculated the pressure as a function of density and the compressibility as a function of pressure, based on the obtained EOS curves, and provide a reasonable analysis of the results.

Furthermore, this work reports the first ever *a priori* prediction of a matrix vibrational band origin shift. This shift is predicted to be a marked function of density, and the value at the equilibrium density is quite close to the experimental result. The highly accurate EOS obtained should provide a reliable platform for further studies on the structural and spectroscopic properties of both pure and doped solid *para*-H₂.

Chapter 3

Three-body potential energy surface for parahydrogen

This chapter is reproduced from Alexander Ibrahim and Pierre-Nicholas Roy, “Three-body potential energy surface for parahydrogen” *Journal of Chemical Physics* **156**, 044301 (2022), with the permission of AIP Publishing.⁸¹

3.1 Introduction

Molecular hydrogen is the simplest of all molecular species, and is the target of a large body of fundamental research. In particular, many-body systems of one of its nuclear spin isomers, *para*-H₂ display a vast array of interesting quantum behaviours. For example, there exist both experimental^{44–48} and theoretical^{49,51,52,52,82–84} evidence of superfluid behaviour in pure and doped clusters of *para*-H₂, or doped clusters of *para*-H₂ embedded within helium droplets. Solid *para*-H₂ has been established as a quantum solid.^{1,2,4,7} In solid *para*-H₂, each molecule has a large zero-point motion about its lattice site (with a Lindemann ratio of about 0.2),^{6,8} owing to the low mass of, and weak intermolecular interactions between, *para*-H₂ molecules.^{7,9} These zero-point lattice vibrations are substantial enough to “inflate” the lattice, and give it many interesting quantum properties. In the zero-temperature, zero-pressure (ZTZP) limit, the lattice constant of solid *para*-H₂ is $R_0 = 3.79 \text{ \AA}$,⁹ considerably higher than what one would predict using a classical description. The combination of the spherical symmetry of its $j = 0$ rotational level, its weak intermolecular interactions, low chemical reactivity, and unsuspectingly large lattice constant, makes solid *para*-H₂ a

powerful host lattice for matrix isolation spectroscopy.^{85–91} Molecules small enough to occupy single substitution sites in solid *para*-H₂, such as H₂O,^{92,93} CH₄,^{94,95} and CO,⁹⁶ are capable of nearly free rotation.

Similarly to how *para*-H₂ is the simplest molecular species, solid *para*-H₂ is one of the simplest examples of a molecular solid. The intermolecular interactions are weak enough for the *para*-H₂ molecules within the crystal to retain many of their free molecule properties.^{1,9} In fact, it is possible to study many of the solid’s properties largely in terms of the free molecule properties. For example, except at very high pressures, the free vibrational and rotational quantum numbers ν and j remain good quantum numbers in the solid phase, and thus the internal vibrations and rotations of the *para*-H₂ molecules remain “free” within the solid.¹ In light of this, it is apparent that we can make large strides in our understanding of solid *para*-H₂ by studying the interactions between free molecular *para*-H₂.

We take our first steps into studying many-body collections of *para*-H₂ by looking at the pair potential. In the initial attempts at creating an interaction potential, scientists would assume a physical form informed by theoretical and phenomenological knowledge, with several adjustable parameters. These parameters would then be determined from experimental results, such as molecular scattering cross section data,^{66,97} virial and viscosity coefficients,⁹⁸ gas transport properties,⁹⁹ and solid state data.^{62,100} With advances in computing and electronic structure calculation research, many H₂–H₂ interaction potentials were created based on *ab initio* studies.^{16,25,101,102} One very commonly used and successful pair potential is the Silvera-Goldman (SG) potential,⁶² a semiempirical potential that uses a combination of solid state data and self-consistent field calculations.

A more complete picture of bulk systems of *para*-H₂ molecules requires the inclusion of many-body potentials. Advances in experimental and computational technology have made it possible to get increasingly accurate quantitative descriptions of condensed systems, and so research into many-body interactions has intensified.^{14,103} At long intermolecular separations, the triple-dipole interaction for spherical molecules is given by the Axilrod-Teller-Muto (ATM) potential^{17,18}

$$V_3(\{R_{ij}\}, \{\alpha_i\}) = C_9 \left[\frac{1 + 3 \cos \alpha_1 \cos \alpha_2 \cos \alpha_3}{R_{12}^3 R_{13}^3 R_{23}^3} \right] \quad (3.1)$$

where C_9 is a coefficient that determines the interaction strength, and in the triangle formed by molecules 1, 2, and 3, R_{ij} is the side length connecting molecules i and j , and α_i is the internal angle at molecule i . However, the $(\textit{para}\text{-H}_2)_3$ three-body contribution to the interaction energy deviates drastically from the ATM potential as the *para*-H₂ molecules move closer together. For most triangle configurations of the three molecules, including

those most relevant to the *hcp* lattice structure of solid *para*-H₂, the isotropic (*para*-H₂)₃ three-body contribution to the interaction energy at short intermolecular spacings is attractive,^{11,104} whereas the ATM potential predicts a strong net repulsive three-body contribution to the interaction energy. Some works attempt to mitigate this short-range repulsion by applying damping functions to the ATM potential.^{105–107} In addition, several *para*-H₂ pair potentials have been created that take many-body effects into account. For example, a potential by Moraldi¹⁰⁸ (and the modified version by Omiyinka and Boninsegni)¹³ softens the SG pair potential at short intermolecular distances, and is in excellent agreement with experiment. The SG potential has a two-body term that approximates the three-body interaction energy. Interestingly, this approximation term is repulsive. For brevity, from this point on we use the term “(*para*-H₂)₃ interaction energy” to refer to the (*para*-H₂)₃ three-body contribution to the total interaction energy, with no contribution from the pair interaction, unless otherwise specified.

In 2010, Manzhos *et al.* constructed an *ab initio* 9D (*para*-H₂)₃ potential energy surface (PES), with energies at the CCSD(T) level calculated using an aug-cc-pVTZ (AVTZ) atom-centred basis set.¹² They sampled the positions and angular orientations of H₂ molecules from a distribution, and fit the energies to a continuous functional form using a neural-network based algorithm. This PES gives the expected attractive behaviour at short intermolecular distances, and incorporates both intermolecular distance and angular orientation degrees of freedom. Garberoglio used a slightly modified version of this PES with much success to calculate the third virial coefficient of *para*-H₂.¹⁰⁹ However, it was noted that the PES by Manzhos *et al.* does not, without modifications, reproduce the correct behaviour at large intermolecular distances. Also, the black-box nature of the PES makes it more difficult to analyze; indeed, this was one of the factors that made it difficult to judge the accuracy of the third virial coefficient calculations.

Faruk *et al.* produced a 1D first-principles *para*-H₂–*para*-H₂ interaction potential,¹⁶ by applying the Adiabatic Hindered Rotor method²⁹ to a 6D H₂–H₂ PES published by Hinde.²⁵ We refer to this potential as the Faruk-Schmidt-Hinde (FSH) potential. The FSH potential was able to successfully reproduce experimentally observed vibrational energy shifts for *para*-H₂ clusters.^{16,31} In a recent paper, path-integral Monte Carlo (PIMC) simulations were performed using the FSH potential to generate an equation of state (EOS) of solid *para*-H₂.³⁷ The FSH potential was able to reasonably predict the ZTZP density and energy per particle, but greatly overestimated the pressure at higher densities compared to two other pair potentials, the SG and Buck⁶⁶ potentials. This is because the FSH potential does not account for the attractive three-body energies at short separations, and thus has a harder core than both the Buck and SG potentials.³¹ The inclusion of a three-body (*para*-H₂)₃ interaction potential alongside the FSH potential would thus decrease the

pressure of solid *para*-H₂ at higher densities, likely bringing it closer to experiment. Janssen and Avoird¹¹⁰ studied the dynamics and phase transitions of solid *para*-H₂ and *ortho*-D₂ using an *ab initio* pair potential, and concluded that three-body interactions are needed to correctly predict the solid’s behaviour at high pressures.

We construct a 3D *ab initio* isotropic PES for the (*para*-H₂)₃ trimer system. Our calculations are carried out at the CCSD(T) level, using the AVTZ atom-centred basis set, with an additional (3s3p2d) bond function.¹¹¹ We apply the rigid rotor approximation, and spherically average over the angular orientations of the H₂ molecules. The energies are fit to a smooth PES using the Reproducing-Kernel Hilbert Space (RKHS) toolkit by Unke *et al.*³⁶ The RKHS method is a machine learning method, primarily used to construct multidimensional PESs from *ab initio* data,^{112,113} and has successfully been used to create PESs for several other molecular systems.^{114–117} This PES is capable of calculating accurate energies for all (*para*-H₂)₃ systems that form triangle with all side lengths greater than 2.2 Å, and offers suitable extrapolations to calculate energies for smaller triangles. At large intermolecular separations, the PES smoothly transitions to the ATM potential. For the case where two molecules are less than 3.2 Å apart and the third is a large distance away, the PES smoothly transitions to a “modified” adaption of the ATM potential.

The remainder of this paper is structured as follows. In Sec. 3.2, we describe the coordinate system used to represent our (*para*-H₂)₃ interaction potential, the *ab initio* calculations used to find the energies, and the methods used to construct the PES. In Sec. 3.3, we assess the errors in the PES due to basis set size, spherical averaging, and sparsity in the *ab initio* input data mesh. We also look at the implications of using the presented PES in a study of solid *para*-H₂, and compare the presented PES against predictions made by the SG potential. In Sec. 3.4, we present our conclusions and plans for future work.

3.2 Creating the Potential Energy Surface

3.2.1 Coordinate System of the Potential Energy Surface

Label the molecules with numbers 1, 2, and 3. Let R_{ij} be the distance between the centres of mass of molecules i and j . The angular orientation of molecule i is given by the space-fixed polar and azimuthal angles (θ_i, ϕ_i) . We apply a rigid rotor approximation, fixing the bond length of each molecule to the vibrationally averaged H₂ ground state ($\nu = 0, j = 0$) interatomic distance of 1.449 a_0 . The rigid rotor approximation significantly reduces both

the number of electronic structure calculations required to create the PES and the cost of calculating the energy in a simulation. Also, because the three-body energy is a correction to a much larger two-body $(\text{H}_2)_2$ energy, the overall error due to the lack of bond length variations is expected to be small.

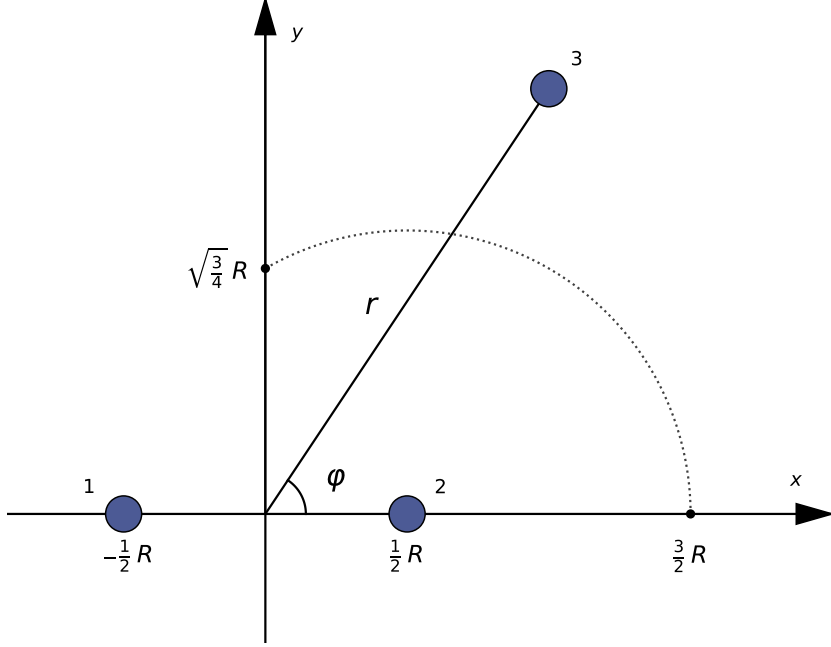


Figure 3.1: The centre-of-mass positions of the three *para*- H_2 molecules on the (x, y) -plane. After removing the bond length and angular degrees of freedom, each *para*- H_2 molecule can be represented as a point. Molecule 3 must lie outside of the region enclosed by the dotted line to satisfy the condition that $R_{12} \leq R_{23} \leq R_{13}$.

At fixed intermolecular distances $\{R_{ij}\}$, we apply the 6-point Lebedev quadrature to spherically average over the angular orientations of the three H_2 molecules.^{32, 118} More specifically, we align each H_2 molecule along either the space-fixed x -, y -, or z -axis. Done independently for all three molecules, there are 27 different combinations of space-fixed angles, and the $(\text{para-H}_2)_3$ interaction energy in our PES is the average of the interaction energies for these 27 different angular combinations. This averaging projects out the anisotropic components of the interaction energy, leaving only the isotropic term. An analysis of the error associated with using the 6-point Lebedev quadrature as opposed to a higher-order Lebedev quadrature is discussed in Sec. 3.3.2. Having removed both the

bond length and angle degrees of freedom, each H_2 molecule is treated as a point particle. The potential is now a function of the intermolecular distances, or equivalently, of the triangle formed by placing a *para*- H_2 molecule at each of its corners. From this point on, for brevity, when we refer to the “distance between two molecules”, we mean the distance between their centres of mass, and when we refer to the “interaction energy of a triangle configuration”, we refer to the interaction energy of the triangle where the centre of mass of a *para*- H_2 molecule lies at each corner.

In the present work, we use a “rescaled” Jacobi coordinate system to describe the relative centres of mass of the three *para*- H_2 molecules. First, label the molecules such that the pairwise distances satisfy $R_{12} \leq R_{23} \leq R_{13}$. Define $R = R_{12}$, and place molecules 1 and 2 on the x -axis, at $(-R/2, 0)$ and $(R/2, 0)$, respectively. Place molecule 3 in the positive quadrant of the (x, y) -plane. Draw a line connecting the origin to the position of molecule 3. Define r as the length of this line, and define $\varphi \in [0, \pi/2]$ as the angle between the positive x -axis and this line. This three-body coordinate system of *para*- H_2 molecules is depicted in Fig. (3.1). This coordinate system is analogous to representing the closest two molecules as a dimer of bond length R , while the third molecule is a monomer whose location in space is described relative to the centre of mass and angular orientation of the dimer.

The variables R , r , and φ can describe all triangles formed by the three *para*- H_2 molecules. However, due to the requirement that $R \leq R_{23} \leq R_{13}$, the allowed values of r are coupled to both R and φ . The minimum value of r is derived by setting $R_{23} = R$, and is given by

$$r_{\min} = \frac{R}{2} \left[\cos \varphi + \sqrt{3 + \cos^2 \varphi} \right]. \quad (3.2)$$

We decouple the coordinates R and φ from r using the dimensionless variable $s = r/r_{\min}$. The coordinate system used in the remainder of this paper is described by the independent variables (R, s, φ) , where $R \geq 0$, $s \geq 1$, and $\varphi \in [0, \pi/2]$. This “rescaled” Jacobi coordinate system is a convenient way of representing a PES for a three-body system of identical particles. Having three independent, non-repetitive coordinates makes it easy to set up a regular 3D grid. In addition, there is a simple, and unique, geometric interpretation associated with each coordinate from its interpretation as a dimer-monomer system. For example, the coordinate R represents both the bond length of the dimer and the overall size of the triangle, and changing it while keeping s and φ constant is akin to “rescaling” both the length of the dimer and the distance of the monomer from the dimer’s centre by the same factor. Changing s while keeping R and φ constant is akin to moving the monomer towards or away from the centre of the dimer. The angle φ represents the orientation of the monomer’s location relative to the long axis of the dimer.

Triangles of specific forms have convenient representations in the rescaled Jacobi coordinate system. As examples, the coordinates $(R, 1, \pi/2)$, $(R, 1, \tan^{-1}(1/2))$, and $(R, 1, 0)$ represent the equilateral triangle of side length R , the right-angled triangle of leg length R , and the case where the molecules lie on a straight line with a spacing of R . The coordinates $(R_0, s, \pi/2)$ represent the set of isosceles triangles of base length R_0 and height $r \geq \sqrt{3/4}R_0$. All triangles with the same value of s and φ are similar triangles.

In practice, we calculate the rescaled Jacobi coordinates from the relative pair distances using the following procedure: Consider a triangle with side lengths $R_{12} \leq R_{23} \leq R_{13}$. Trivially, we set $R = R_{12}$. Next, we calculate the distance from the origin to molecule 3 using

$$r = \sqrt{\frac{1}{2} \left[R_{13}^2 + R_{23}^2 - \frac{1}{2} R_{12}^2 \right]}. \quad (3.3)$$

The coordinates φ and s are then found using

$$\cos \varphi = \frac{1}{2} \frac{R_{13}^2 - R_{23}^2}{R_{12}r} \quad (3.4)$$

and

$$s = \frac{2r}{R} \left[\cos \varphi + \sqrt{3 + \cos^2 \varphi} \right]^{-1}. \quad (3.5)$$

3.2.2 Electronic Structure Calculations

We perform electronic structure calculations for the $(\text{H}_2)_3$ interaction energy using the MRCC (version 2019) program.¹¹⁹ The calculations are done with the coupled-cluster method with single, double, and perturbative triple excitations, which is abbreviated as CCSD(T).¹²⁰ We use an AVTZ atom-centred basis set for each of the six hydrogen atoms, and supplement this with an additional set of $(3s3p2d)$ midbond functions¹¹¹ positioned at the center of mass of the system, following the example by Hinde.²⁵ The $(para\text{-H}_2)_3$ calculations by Wind *et al.* and Manzhos *et al.* did not make use of midbond functions.^{12,104} It has been shown that the AVTZ basis set provides similar energies to the aug-cc-pVQZ (AVQZ) basis set for the $(\text{H}_2)_3$ trimer.^{11,104}

For a fixed coordinate (R, s, φ) , let n label each of the 27 angular configurations of the 6-point Lebedev quadrature. The counterpoise-corrected²⁸ $(\text{H}_2)_3$ trimer interaction energy of the n^{th} angular orientation (with (R, s, φ) omitted on the RHS for brevity) is calculated using

$$E^n(R, s, \varphi) = E_{123}^n - (E_{12[3]}^n + E_{13[2]}^n + E_{23[1]}^n) + (E_{1[23]}^n + E_{2[13]}^n + E_{3[12]}^n). \quad (3.6)$$

In the above equation, E_{123}^n is the total CCSD(T) energy of the trimer in the n^{th} angular configuration. We define $E_{ij[k]}^n$ to be the CCSD(T) energy of the i^{th} and j^{th} molecules in the complete one-electron basis set of the trimer. In other words, $E_{ij[k]}^n$ is the CCSD(T) energy of the $(\text{H}_2)_3$ trimer where the nuclei and electrons of the k^{th} molecule are removed but its associated basis functions remain. Similarly, $E_{i[jk]}^n$ is the CCSD(T) energy of the i^{th} molecule in the complete one-electron basis set of the trimer. The spherically-averaged *ab initio* (*para*- H_2)₃ interaction energy V_{CC} is calculated using

$$V_{\text{CC}}(R, s, \varphi) = \frac{1}{27} \sum_{n=1}^{27} E^n(R, s, \varphi) . \quad (3.7)$$

To create the input data mesh for the PES, we calculate V_{CC} for 21 values of $R/\text{\AA} \in [2.20, 6.25]$, 17 values of $s \in [1.0, 3.85]$, and 19 values of $\varphi \in [0, \pi/2]$. The spacing of φ values is uniform, while the spacings for R and s are non-uniform, with a denser selection of points at smaller values and a sparser selection at larger values. The error analysis associated with the spacing of points in the input data mesh is done in Sec 3.3.3.

The precision of the CCSD(T) energies calculated by the MRCC program are determined by several convergence criteria. The calculations for these energies were performed with all default settings (for the 2019 version), except for `cctol`, which is set to 9. We perform sample calculations using the AVTZ basis set using `cctol` = 10, for two triangular configurations. First, we do so for the equilateral triangle configuration $(R, 1, \pi/2)$, ranging from $R/\text{\AA}$ = 2.20 to 6.25. All the sample interaction energies change by less than 0.0006 cm^{-1} . At shorter intermolecular spacings, this is much smaller than the difference between (*para*- H_2)₃ energies calculated using the AVTZ and AVQZ atom-centred basis sets. Next, we do so for the $(2.2 \text{\AA}, s, 0)$ configuration, where two molecules are a distance 2.2 $\text{\AA}$ apart on the x -axis, and the third molecule moves away from them along the x -axis. We see the same behaviour. At intermolecular separations great enough where average (*para*- H_2)₃ interaction energy becomes comparable to the error, the PES will have already been adjusted to an empirical form, as we shall see in Secs. 3.2.3 and 3.2.3.

3.2.3 Constructing the Potential Energy Surface

We construct a smooth PES using a combination of several methods. First, we use the RKHS toolkit provided by Unke *et al.* to construct the “initial” PES using the CCSD(T) energies.³⁶ The RKHS method has several convenient properties. For example, the resultant PES is smooth everywhere, and reproduces all the input CCSD(T) energies exactly.

While the RKHS method is used for interpolating interaction energies within the (R, s, φ) input data mesh, we use a number of other strategies to extrapolate the PES outside of the grid.

Large R Extrapolation

For cases where all three molecules are far apart ($R_{ij} > 5.5 \text{ \AA}$), the *ab initio* (*para*-H₂)₃ interaction energies approach the ATM potential. Expressed using (R, s, φ) coordinates, the ATM potential is

$$V_{\text{ATM}}(R, s, \varphi) = C_9 \frac{64}{R^9} \frac{1 + f(s, \varphi)}{[T(s, \varphi)]^{3/2}} \quad (3.8)$$

where, using

$$W(s, \varphi) = s^2 \left[\cos \varphi + \sqrt{3 + \cos^2 \varphi} \right]^2 = 4 \frac{r^2}{R^2}, \quad (3.9)$$

we have

$$T(s, \varphi) = 1 - 2W(s, \varphi) \cos 2\varphi + W^2(s, \varphi) \quad (3.10)$$

and

$$f(s, \varphi) = -3 \frac{(1 - W(s, \varphi) \cos^2 \varphi)(1 - W(s, \varphi))}{T(s, \varphi)}. \quad (3.11)$$

We use the coefficient $C_9 = 34336.220 \text{ cm}^{-1} \text{ \AA}^9$.^{11,121}

The *ab initio* (*para*-H₂)₃ interaction energies calculated using the AVDZ, AVTZ, and AVQZ basis sets do not precisely converge to the ATM potential at long distances. We ensure a smooth transition from the *ab initio* AVTZ CCSD(T) energies to the ATM potential using the following procedure. First, we fix a pair of values $s = \bar{s}$ and $\varphi = \bar{\varphi}$ from the input data mesh. Then we plot both the ATM potential and the set of *ab initio* (*para*-H₂)₃ interaction energies $\{V_{\text{CC}}(R; \bar{s}, \bar{\varphi})\}$ against R , and rescale all the energies by $C_9^{-1} R^9$ (see Fig. (3.2)). With this rescaling, the ATM potential appears as a horizontal line. Next we pick a value of R from the input data mesh, call it R_A , beyond which the rescaled *ab initio* energies appear to flatten, indicating that the PES has begun its R^{-9} decay. We intend to adjust the PES such that $V_3(R_A; \bar{s}, \bar{\varphi}) = V_{\text{ATM}}(R_A; \bar{s}, \bar{\varphi})$. We also pick a value $R = R_C$, not too far below R_A , at which the ATM and *ab initio* energies have very clearly diverged when looking at the rescaled plot. The resultant PES energies $\{V_3(R; \bar{s}, \bar{\varphi})\}$ are calculated (omitting $\bar{s}$ and $\bar{\varphi}$ on the RHS) using

$$V_3(R; \bar{s}, \bar{\varphi}) = \begin{cases} V_{\text{ATM}}(R) & R \geq R_A, \\ V_{\text{CC}}(R) [1 + \eta(R, R_A)] & R_C \leq R < R_A, \\ V_{\text{CC}}(R) & R < R_C \end{cases} \quad (3.12)$$

where

$$\eta(R, R_A) = \left[\frac{V_{\text{ATM}}(R_A) - V_{\text{CC}}(R_A)}{V_{\text{CC}}(R_A)} \right] e^{-\lambda(R-R_A)^2}, \quad (3.13)$$

and $\lambda = \lambda(\bar{s}, \bar{\varphi})$, often set between 2 and 5, determines the rate of decay. In other words, we use the relative difference between the CCSD(T) and ATM energies at the chosen distance R_A to smoothly taper the PES towards the ATM energies, for $R_C < R < R_A$. To ensure the PES remains smooth, we apply these changes to the data level and reapply the RKHS method. As an example of the aforementioned procedure, in Fig. (3.2) for the equilateral triangle configuration, we show the AVDZ-, AVTZ-, and AVQZ-level *ab initio* (*para*-H₂)₃ energy curves, the ATM potential, and the final adjusted PES. We see that the unadjusted *ab initio* energies do not properly converge to the ATM potential at long distances; each underestimates the effective value of C_9 by about 3%. The current PES, constructed by applying Eq. (3.12) to the AVTZ curve, smoothly transitions to the ATM potential.

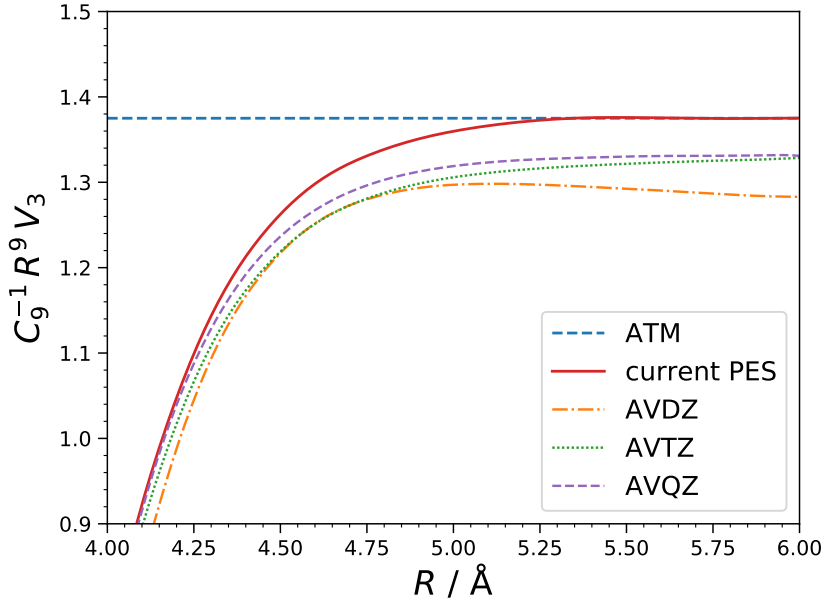


Figure 3.2: The (*para*-H₂)₃ interaction energy for three *para*-H₂ molecules at the corners of an equilateral triangle of side length R , multiplied by $C_9^{-1} R^9$. Shown are the ATM potential (horizontal, blue, dashed), and the RKHS-constructed *ab initio* CCSD(T) energies calculated using the AVDZ (orange, dash-dotted), AVTZ (green, dotted), and AVQZ (purple, dashed) basis sets. Also shown is the current PES, constructed by applying the adjustments of Eq. (3.12) to the AVTZ *ab initio* energies, with $R_A = 5.35$ Å and $R_C = 3.6$ Å.

Large s Extrapolation For $R < 3.2 \text{ \AA}$

For cases where two molecules are close to one another ($R < 3.2 \text{ \AA}$), but the third molecule is farther away (large s), the electronic structure calculations do not approach the ATM potential. Instead, *ab initio* energies in the “small R , large s ” regime converge to a modified ATM potential (the “MATM potential”), which takes the form

$$V_{\text{MATM}}(R, s, \varphi) = C_9 \frac{64}{R^9} \frac{a(R, \varphi) + b(R, \varphi) f(s, \varphi)}{[T(s, \varphi)]^{3/2}}. \quad (3.14)$$

Compared to Eq. (3.8), the only difference is the appearance of coefficients $a(R, \varphi)$ and $b(R, \varphi)$ in the numerator.

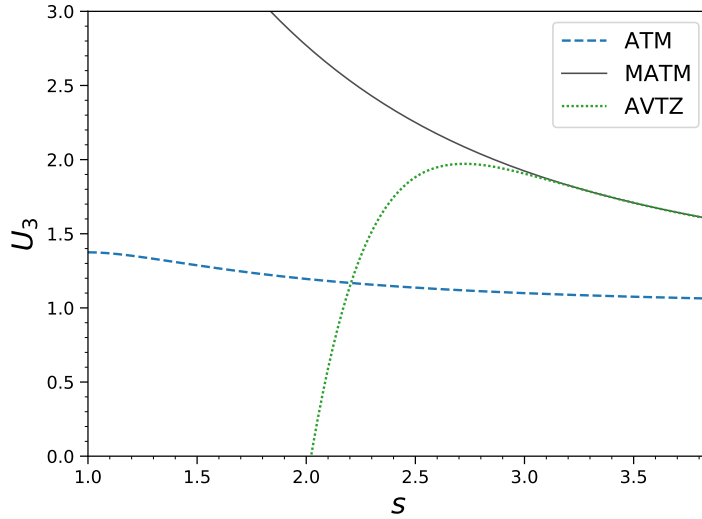


Figure 3.3: The numerator term $U_3(s; R, \varphi) = a + b f(s, \varphi)$, for $(R, \varphi) = (2.2 \text{ \AA}, \pi/2)$. We show the curves for the ATM potential (blue, dashed), the RKHS-constructed AVTZ-level *ab initio* energies (green, dotted), and the MATM potential constructed to fit these *ab initio* energies. In this case, $a = 1.06$ and $b = 8.64$.

We calculate $a(R, \varphi)$ and $b(R, \varphi)$ as follows. Define $U_3(s; R, \varphi) = a + b f(s, \varphi)$; for the *ab initio* energies, we find $U_3(s; R, \varphi)$ by multiplying the energies by $64^{-1} C_9^{-1} R^9 T^{3/2}(s, \varphi)$. First, we fix the values of $R = \bar{R}$ and $\varphi = \bar{\varphi}$ from the input data mesh. Then, we pick some s' for which $U_3(s; \bar{R}, \bar{\varphi})$ has transitioned to its long-range MATM behaviour for $s \geq s'$, and

calculate $a(\bar{R}, \bar{\varphi})$ and $b(\bar{R}, \bar{\varphi})$ using

$$b = \left. \frac{\partial U_3 / \partial s}{\partial f / \partial s} \right|_{s'} \quad \text{and} \quad a = U_3(s') - b f(s') . \quad (3.15)$$

For $R > 3.2 \text{ \AA}$, the coefficients $a(R, \varphi)$ and $b(R, \varphi)$ both converge to unity and take the form of the nominal ATM potential. To give the constructed PES the correct long-range behaviour along the s -coordinate, we construct a new input mesh using

$$V_3(s; \bar{R}, \bar{\varphi}) = \begin{cases} V_{\text{CC}}(s; \bar{R}, \bar{\varphi}) & s \leq s' \\ a(\bar{R}, \bar{\varphi}) + b(\bar{R}, \bar{\varphi}) f(s; \bar{\varphi}) & s > s' \end{cases} \quad (3.16)$$

where s' is the value of s used to define the MATM potential in Eq. (3.15). To ensure the transition occurs smoothly, we reapply the RKHS method on this new input mesh. In Fig. (3.3), we plot $U_3(s)$ for $(R, \varphi) = (2.2 \text{ \AA}, \pi/2)$. We see that the *ab initio* (*para*-H₂)₃ interaction energies do not converge to the ATM potential when two *para*-H₂ molecules are close together. Instead, in this regime we fit the MATM potential to the long-range behaviour of the *ab initio* energies along the s -coordinate.

An unfortunate effect of using Eq. (3.15) to calculate $a(R, \varphi)$ and $b(R, \varphi)$ is that both coefficients blow up as φ approaches zero. To see why, note that the partial derivative of Eq. (3.11),

$$\frac{\partial f}{\partial s} = \sin^2 \varphi \frac{6W}{s} \left[\frac{3 + 2W - (1 + 4 \cos^2 \varphi) W^2}{T^2} \right], \quad (3.17)$$

is present in the denominator of Eq. (3.15), and that, near $\varphi = 0$, b blows up as $\sin^{-2} \varphi$. To work around this singularity, we perform the MATM fit of *ab initio* energies along ($\varphi = 0$) using $f(s, \varphi_{\text{cf}})$, with the cutoff $\varphi_{\text{cf}} = 10^{-5}$. This approximation has no effect whatsoever on the unextrapolated *ab initio* energies. Performing the fit using $\varphi_{\text{cf}} = 10^{-3}$ or $\varphi_{\text{cf}} = 10^{-4}$ shows that this has little effect on the extrapolated energies, because as long as $\varphi_{\text{cf}} \neq 0$, the coefficients a and b simply adjust themselves to maintain the MATM fit.

When $R < 3.2 \text{ \AA}$, we can use Eq. (3.14) to calculate (*para*-H₂)₃ interaction energies when $s > 3.85$. For most values of φ , we calculate coefficients a and b using

$$a = \frac{U_3(s_0)f(s_1) - U_3(s_1)f(s_0)}{f(s_1) - f(s_0)} \quad (3.18)$$

and

$$b = \frac{U_3(s_1) - U_3(s_0)}{f(s_1) - f(s_0)} \quad (3.19)$$

where s_0 and $s_1 > s_0$ are selected values of s for which the PES has already transitioned into the MATM behaviour (we use $(s_0, s_1) = (3.55, 3.85)$). However, for $\varphi < \varphi_{\text{cf}} = 10^{-5}$, Eqs. (3.18) and (3.19) become numerically unstable, and the extrapolation is instead performed using the built-in long-range behaviour of the RKHS method.

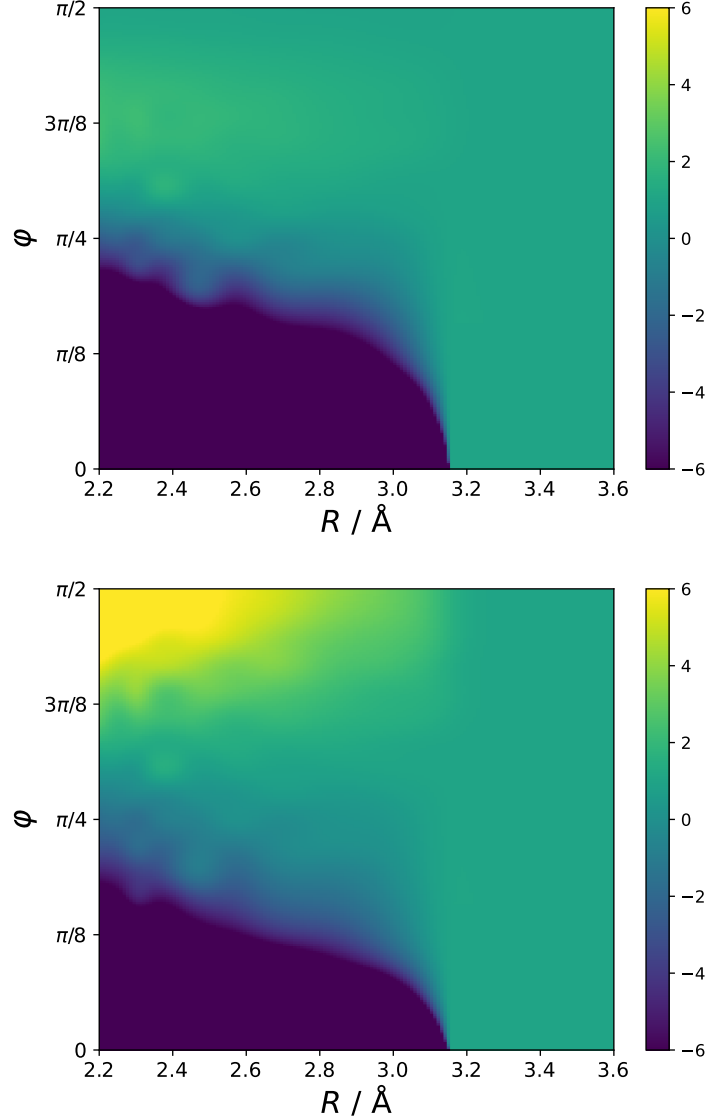


Figure 3.4: Colormap of $a(R, \varphi)$ (top panel) and $b(R, \varphi)$ (bottom panel) for the final constructed PES, calculated using Eq. (3.15).

When $R > 3.2 \text{ \AA}$, the coefficients $a(R, \varphi)$ and $b(R, \varphi)$ both converge to about unity, with numerical noise. We remove this noise from $a(R, \varphi)$ by transitioning from the numerically calculated coefficients of Eq. (3.15) to identically unity using

$$a(R, \varphi) := \bar{\omega}(R; R_m, \delta R) a(R, \varphi) + \omega(R; R_m, \delta R) \quad (3.20)$$

with $R_m = 3.2 \text{ \AA}$ and $\delta R = 0.1 \text{ \AA}$. The same filter with the same parameters is applied to the coefficient $b(R, \varphi)$. In Eq. (3.20), ω is given by

$$\omega(x; x_m, \delta x) = \begin{cases} 0 & x \leq x_0, \\ \frac{1}{2} [1 - \cos(\pi k)] & x_0 < x < x_1, \\ 1 & x \geq x_1, \end{cases} \quad (3.21)$$

and $\bar{\omega}$ is given by

$$\bar{\omega}(x; x_m, \delta x) = 1 - \omega(x; x_m, \delta x) \quad (3.22)$$

where we use $x_0 = x_m - \frac{1}{2}\delta x$, $x_1 = x_m + \frac{1}{2}\delta x$, and $k = (x - x_0)/\delta x$.

We show colormaps of $a(R, \varphi)$ and $b(R, \varphi)$ respectively in Fig. (3.4). For most of the region $\varphi > 0.9$, a is somewhat larger than unity, and peaks with $a = 2.28$ near $\varphi \approx 3\pi/8$. For $R < 3.2 \text{ \AA}$, as φ shrinks, the coefficient a blows up. As expected, a and b both become unity past $R = 3.2 \text{ \AA}$. The colormap of $b(R, \varphi)$ behaves similarly to its counterpart, except when $R < 3.0 \text{ \AA}$ and $\varphi > 1.0$. This region of coordinate space corresponds to small equilateral and “almost-equilateral” triangles. The coefficient b grows fairly large in this region, increasing to a maximum of about 8.64 when $(R, \varphi) = (2.2 \text{ \AA}, \pi/2)$.

Small R Extrapolation

For most values of s and φ , we can extrapolate the $(para\text{-H}_2)_3$ interaction energy below $R = 2.2 \text{ \AA}$ using an exponential fit

$$V_{\text{EX}}(R; s, \varphi) = V_3(R_{\min}) \exp(-c(R - R_{\min})), \quad (3.23)$$

where $c = c(s, \varphi)$ and (omitting s and φ on the RHS)

$$c(s, \varphi) = \frac{1}{\Delta R} \ln \left| \frac{V_3(R_{\min} + \Delta R)}{V_3(R_{\min})} \right| \quad (3.24)$$

with $R_{\min} = 2.2 \text{ \AA}$ and $\Delta R = 0.05 \text{ \AA}$. However, there are regions in the (s, φ) plane for which c is very large, and the extrapolation blows up very quickly below $R_{\min}$. These

are regions where, in Eq. (3.24), the magnitude of the numerator is a factor of about 2 or greater than that of the denominator. This problem is often worsened when $c(s, \varphi)$ is defined using the exact derivative of $V_3(R)$ instead of a finite approximation. On closer inspection, this behaviour only happens when $|V_3(R_{\min}; s, \varphi)|$ is already very close to zero (typically $< 0.05 \text{ cm}^{-1}$), and so a small absolute change in energy results in a large relative change in energy. Thus, pre-empting instances where the exponential fit will blow up due to a very large $c(s, \varphi)$, we transition from the exponential fit to a less accurate but numerically stable linear fit, using

$$V_3(R) = \bar{\omega}(c; c_m, \delta c)V_{\text{LI}}(R) + \omega(c; c_m, \delta c)V_{\text{EX}}(R) \quad (3.25)$$

where $c_m = 7$, $\delta c = 2$, ω and $\bar{\omega}$ are given by Eqs. (3.21) and (3.22), respectively, and $V_{\text{LI}}(R)$ is the linear fit below $R = R_{\min}$ whose slope is determined by $V_3(R_{\min})$ and $V_3(R_{\min} + \Delta R)$. Because the $(para\text{-H}_2)_3$ interaction energies in regions where the linear fit is used are already very small, as long as the $(para\text{-H}_2)_3$ interaction potential is not extrapolated too far below $R_{\min}$, the absolute error introduced by using the linear fit will also be very small. When used alongside a $(para\text{-H}_2)_2$ pair interaction potential, the error would be negligible. It should be emphasized that this short-range extrapolation along the R coordinate is not meant to give accurate $(para\text{-H}_2)_3$ interaction energies below $R_{\min}$. Rather its primary purpose is to make it possible to sample the PES at short distances without it blowing up. Such short range configurations are very unfavourable, and are expected to be rare except at very high pressures and densities. To ensure a smooth transition of the PES across $R = R_{\min}$, four phantom values of R with a spacing of $\Delta R = 0.05 \text{ \AA}$ are created down to $R = 2.0 \text{ \AA}$, and the RKHS method is reapplied.

In the case where both $R < 2.2 \text{ \AA}$ and $s > 3.85$, we first perform the MATM extrapolation along the s -coordinate described in Sec. 3.2.3 to calculate $V_3(2.2 \text{ \AA}, s, \varphi)$ and $V_3(2.25 \text{ \AA}, s, \varphi)$. We then use these two energies to extrapolate to values below $R = 2.2 \text{ \AA}$ with Eqs. (3.23), (3.24), and (3.25). Again, we should stress that this extrapolation purposely sacrifices accuracy for stability, and it should not be expected to give highly accurate $(para\text{-H}_2)_3$ interaction energies.

Summary

To summarize the construction of the current PES described in this subsection, we show in Fig. (3.5) the method used to calculate the energies in each region of the (R, s) -plane.

With the exception of the MATM potential fit for the case $\varphi < 10^{-5}$, the φ coordinate is treated the same way in each region depicted in Fig. (3.5). In the “ATM” region, the

current PES is identical to the ATM potential. In the “MATM” region, the energies are calculated using Eqs. (3.18) and (3.19) in Sec. 3.2.3. In the “Exponential*” and “MATM + Exponential*” regions, we use the extrapolation procedures described in Sec. 3.2.3. The “RKHS” region is a combination of purely AVTZ-level *ab initio* (*para*-H₂)₃ interaction energies and the adjustments to the ATM and MATM potential forms described in Secs. 3.2.3 and 3.2.3. At the border between the ATM and RKHS regions, the RKHS energies are already converged to the functional form of the ATM potential, and the transition is smooth. The same is true for the transition between the MATM and RKHS regions.

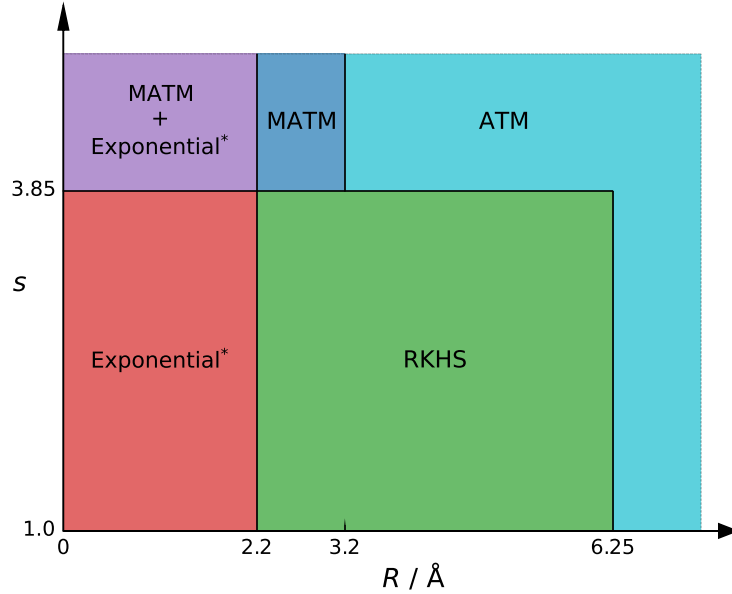


Figure 3.5: The regions in the (R, s) -plane, labeled with the method used to calculate (*para*-H₂)₃ interaction energies within them. Details describing each region are given in the text.

In principle, it does not matter in which order the adjustments in Secs. 3.2.3 and 3.2.3 are made. In practice, the two adjustments overlap when both R and s are large enough, and performing the MATM adjustments in Sec. 3.2.3 first will automatically perform the required adjustments of Sec. 3.2.3 for many values of s and φ . The extrapolation to energies below $R = 2.2$ Å can be done independently of either of the aforementioned two adjustments.

3.3 Discussion and Results

3.3.1 Basis Set Comparison

We estimate the error due to the truncation of the one-electron basis set by comparing the $(para\text{-H}_2)_3$ interaction energies calculated using the AVDZ, AVTZ, and AVQZ atom-centred basis sets at several coordinates. In each case, we supplement the atom-centred basis sets with the $(3s3p2d)$ bond function located at the centre of mass of the trimer. The energies are listed in Table 3.1.

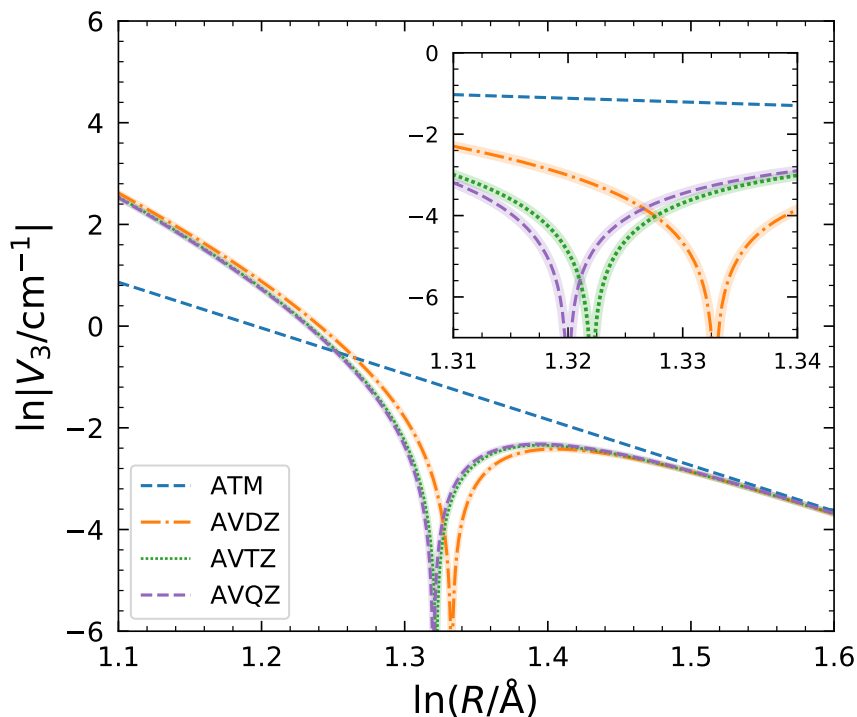


Figure 3.6: The $(para\text{-H}_2)_3$ interaction energy for three $para\text{-H}_2$ molecules at the corners of an equilateral triangle of side length R . Shown are the ATM potential (straight, blue, dashed), and the RKHS-constructed *ab initio* CCSD(T) energies calculated using the AVDZ (orange, dash-dotted), AVTZ (green, dotted), and AVQZ (purple, dashed) basis sets. The cusp in each *ab initio* curve is located where the curve changes sign from negative to positive. The inset zooms into the region where the cusps are grouped together, and the differences are more pronounced.

Table 3.1: Interaction energies for the (*para*-H₂)₃ trimer (cm⁻¹) at select (R, s, φ) coordinates (with R in Å), calculated at the CCSD(T) level using the AVDZ, AVTZ, and AVQZ atom-centred basis sets, supplemented with the ($3s3p2d$) bond function at the centre of mass.

R	s	φ	AVDZ	AVTZ	AVQZ
2.20	1.00	0	39.3985	35.1058	34.4951
		$\pi/6$	21.3524	18.6636	18.1592
		$\pi/3$	-94.2664	-90.8722	-90.2855
		$\pi/2$	-594.3297	-578.3776	-575.2212
	1.25	0	2.6753	2.0036	1.9247
		$\pi/6$	2.8044	2.2194	2.1194
		$\pi/3$	-10.7030	-10.0275	-9.9676
		$\pi/2$	-166.8678	-160.3304	-159.1005
	1.75	0	-0.2085	-0.2181	-0.2134
		$\pi/6$	-0.1402	-0.1441	-0.1445
		$\pi/3$	0.2278	0.2379	0.2343
		$\pi/2$	-7.0983	-6.2079	-6.1053
	2.95	0	0.1483	0.0160	0.0041
		$\pi/6$	0.1973	0.1050	0.0896
		$\pi/3$	-0.9241	-0.7669	-0.7571
		$\pi/2$	-18.1219	-16.8327	-16.6135
	1.25	0	-0.1009	-0.1071	-0.1082
		$\pi/6$	-0.0568	-0.0635	-0.0656
		$\pi/3$	0.0815	0.0950	0.0938
		$\pi/2$	-2.1565	-1.8514	-1.8101
	1.75	0	-0.0122	-0.0126	-0.0135
		$\pi/6$	-0.0095	-0.0096	-0.0103
		$\pi/3$	0.0187	0.0181	0.0171
		$\pi/2$	0.1580	0.1692	0.1685
4.15	1.00	0	-0.0240	-0.0268	-0.0255
		$\pi/6$	-0.0144	-0.0150	-0.0154
		$\pi/3$	0.0219	0.0222	0.0221
		$\pi/2$	0.0867	0.0903	0.0917

In essentially each case, we see a much larger jump in energy when we change from the AVDZ to the AVTZ basis set, than from the AVTZ to the AVQZ basis set. We also

see that, except for certain cases where the energies are near the point of changing signs the difference between energies calculated using the AVTZ and AVQZ basis sets are either between 1 – 3% of each other, or within 0.1 cm^{-1} of each other. At larger intermolecular separations, the differences are even smaller. These observations are consistent with the previous findings that the $(para\text{-H}_2)_3$ interaction energies converge very quickly with respect to the choice of one-electron basis set.^{11,104}

In Fig. (3.6), we compare the interaction energies for three $para\text{-H}_2$ molecules arranged in an equilateral triangle configuration, calculated using the AVDZ, AVTZ, and AVQZ atom-centred basis sets (the “CCSD(T) energy curves”), alongside the ATM potential. The CCSD(T) energy curves were each created using the RKHS method. All three of the CCSD(T) energy curves change from being strongly attractive to weakly repulsive as the intermolecular distance increases, and the cusp indicates the distance at which the sign of the energy changes.

The ATM potential has the expected R^{-9} decay. The CCSD(T) interaction energies decay exponentially at small R , but trend with and converge to the ATM potential at large R . In the inset, we see that the change in the potential energy curve is much smaller going from using the AVQZ basis set to the AVTZ basis set, than from the AVTZ basis set to the AVDZ basis set. Calculations using the AVQZ basis set are roughly 20 times slower than those using the AVTZ basis set. Thus, performing our calculations using the AVTZ basis set should provide sufficiently accurate energies while also being computationally tractable.

3.3.2 Lebedev Quadrature Comparison

To assess the error introduced by the use of the 6-point Lebedev quadrature scheme, we compare *ab initio* $(para\text{-H}_2)_3$ interaction energies that are spherically averaged using the 6-point and 14-point Lebedev quadrature schemes. We do so for two triangle geometries, the $(R, 1, \pi/2)$ geometry (equilateral triangles of side length R) and the $(R, 1, 0)$ geometry (three points on a straight line, each a distance R from its immediate neighbour). The energies are calculated using the AVDZ atom-centred basis, supplemented with the $(3s3p2d)$ midbond functions. We show in Fig. (3.7) the difference between the 6-point and 14-point integrated energies for these two geometries. Both curves display an exponential decay, although in the case of the collinear configuration this exponential decay is briefly interrupted by a sign change near $R = 2.7 \text{ \AA}$. Below $R = 2.95 \text{ \AA}$, the error from the 6-point Lebedev quadrature is about 0.2% and 2% for the equilateral and collinear configurations, respectively. For example, at $R = 2.2 \text{ \AA}$, the 6-point equilateral (collinear) configuration energy is -594.329 cm^{-1} (39.399 cm^{-1}), while the difference between the 6-point and

14-point energies is only -1.235 cm^{-1} (0.845 cm^{-1}). Aside from instances near where the interaction energy changes sign, the error decreases even more beyond $R = 2.95 \text{ \AA}$.

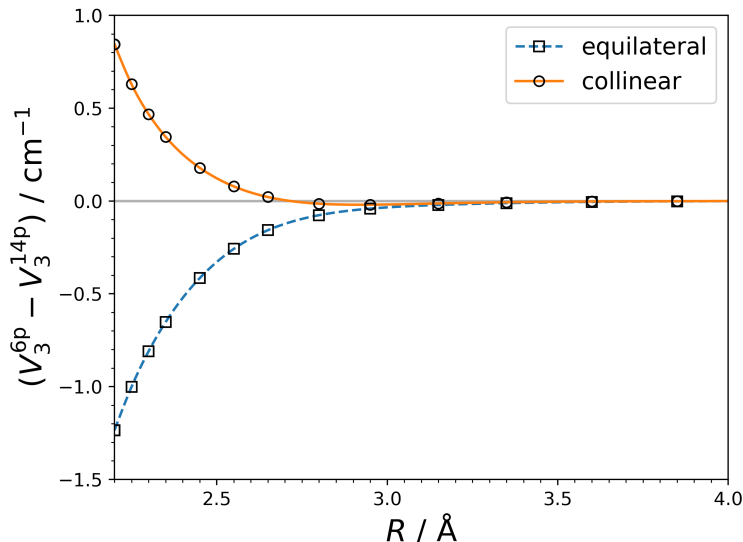


Figure 3.7: The difference between the 6-point (V_3^{6p}) and 14-point (V_3^{14p}) Lebedev quadrature averaged *ab initio* (*para*-H₂)₃ interaction energies. The energies are calculated using the AVDZ basis set. The curves shown are those for the equilateral triangle configuration of side length R (orange, solid, circles) and the collinear triangle configuration with spacing R (blue, dashed, squares).

For at least the isotropic component of the (*para*-H₂)₃ interaction potential, the 6-point Lebedev scheme is acceptable. Whereas the 6-point Lebedev quadrature takes the average of only 27 energies, its 14-point counterpart requires the energies of 343 angular orientations. To switch to the 14-point Lebedev quadrature requires an order of magnitude more electronic structure calculations, while providing relatively little improvement.

3.3.3 Effects of Data Mesh Spacing on Potential Energy Surface

As mentioned previously, the input data mesh used to create the PES is dense at short intermolecular spacings, and gradually increases as the molecules move farther apart. We do this because the AVTZ calculations used in the final constructed PES are fairly expensive. It is both unnecessary and intractable to use the same grid spacing at short intermolecular

separations ($\Delta R = 0.05 \text{ \AA}$ and $\Delta s = 0.05$) throughout the entire data mesh. We must find out where, and by what amount, the data mesh can be made sparser while maintaining the accuracy of the interpolation.

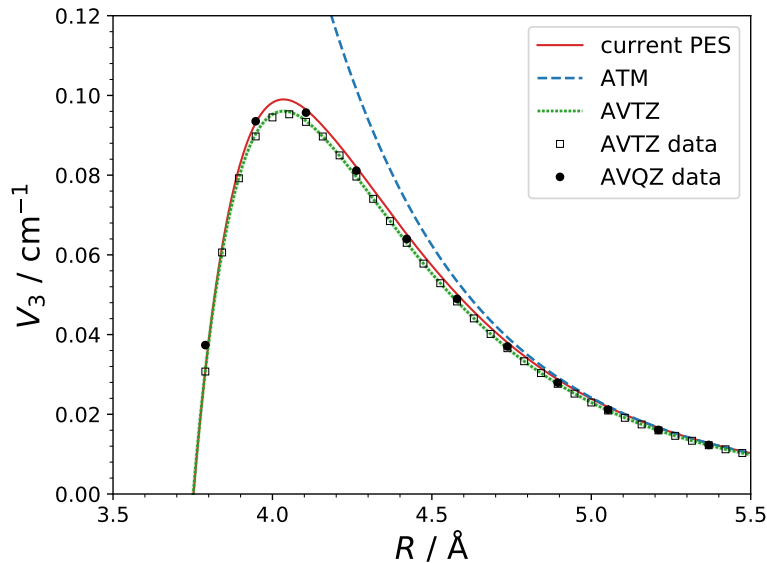


Figure 3.8: The $(para\text{-H}_2)_3$ interaction energy for three $para\text{-H}_2$ molecules at the corners of an equilateral triangle of side length R . Shown are the ATM potential (blue, dashed), the AVTZ-level RKHS-constructed PES (green, dotted), and the final PES (red, solid) presented in this work. At this distance, the AVTZ-level PES is constructed from training data with a mesh spacing of $\Delta R = 0.3 \text{ \AA}$. Also shown are energies calculated using the AVTZ basis (empty squares, spacing of $\Delta R = 0.053 \text{ \AA}$) and the AVQZ basis (solid circles, spacing of $\Delta R = 0.159 \text{ \AA}$); these energies were not used to construct any shown curve.

To do so, we calculated $(para\text{-H}_2)_3$ interaction energies for a uniform data mesh in the range of $(R/\text{\AA}, s, \varphi) \in [2.20, 4.75] \times [1.0, 5.0] \times [0, \pi/2]$, with grid spacings of $\Delta R = 0.05 \text{ \AA}$, $\Delta s = 0.046$, and $\Delta \varphi = \pi/20$. Overall, the data set was made up of $52 \times 100 \times 19 = 98800$ points. Calculations were performed with the cheaper AVDZ basis set + $(3s3p2d)$ midbond function. We construct a PES from this uniform data mesh using the RKHS method.

In a process of trial and error, we remove rows of R or s data at different sections of the input mesh, reconstruct the PES with the RKHS method, and compare its accuracy against the PES constructed with the full grid. For $R < 2.35 \text{ \AA}$, we find that the data mesh must remain dense to match the initial accuracy, at $\Delta R = 0.05 \text{ \AA}$. However, as R increases,

we can gradually increase the spacing by $0.05 \text{ \AA}$ at different intervals up to $\Delta R = 0.30 \text{ \AA}$ when $R > 3.85 \text{ \AA}$. A similar trend is seen with the s coordinate.

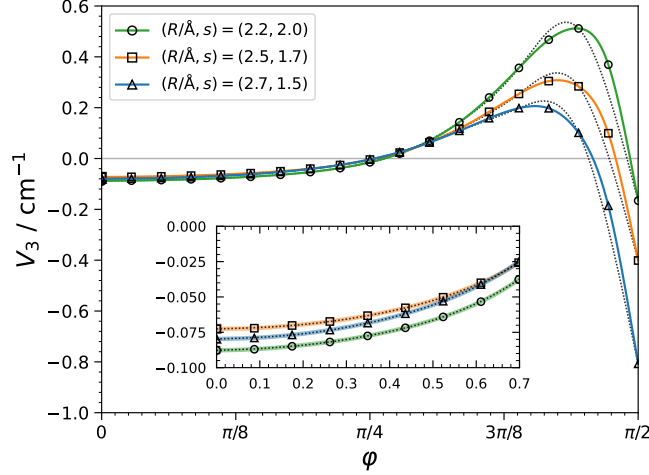


Figure 3.9: The $(para\text{-H}_2)_3$ interaction energy curves along the φ coordinate. The RKHS method is applied to the AVTZ-level *ab initio* $(para\text{-H}_2)_3$ interaction energies, using two different spacings along the φ coordinate. We show the curves for a spacing of $\Delta\varphi = 5^\circ$ [coloured lines, green for $(R/\text{\AA}, s) = (2.2, 2.0)$, orange for $(R/\text{\AA}, s) = (2.5, 1.7)$, blue for $(R/\text{\AA}, s) = (2.7, 1.5)$] and $\Delta\varphi = 10^\circ$ [corresponding black dashed lines]. The markers represent the *ab initio* data at $\Delta\varphi = 5^\circ$ spacing. The inset zooms into the plot at smaller φ .

To compare against the RKHS method, we use the tricubic interpolation method to construct another PES from the dense uniform input data mesh. The implementation of the tricubic interpolation we use is given by Lekien *et al.*¹²² The two methods produce very similar energies (typically less than 10^{-6} cm^{-1} difference at large R and s , and less than 10^{-2} cm^{-1} difference at small R and s). The only notable exception was between $R = 2.20 \text{ \AA}$ and $R = 2.25 \text{ \AA}$, where the tricubic PES deviated from the expected exponential behaviour, while the RKHS PES reproduced the expected trend successfully. Ultimately, the RKHS method was chosen over the tricubic interpolation method, because the accuracy of the latter worsened more quickly as the input data mesh spacing increased.

The spacing in the R -coordinate for $R > 3.85 \text{ \AA}$ is fairly large, at $\Delta R = 0.3 \text{ \AA}$, and we want to see how accurately the PES interpolates the $(para\text{-H}_2)_3$ interaction energies at this distance. In Fig. (3.8), for the equilateral triangle configuration, we plot curves for the current PES, the AVTZ-level RKHS-constructed PES, and the ATM potential. Also

shown are AVTZ-level energies with a spacing of $\Delta R = 0.053 \text{ \AA}$ that act as a “test set”, unused in the creation of the AVTZ-level PES. Despite being constructed with a mesh spacing about 6 times greater than the test energies, the AVTZ-level PES provides an excellent interpolation for the independent energy points. Looking at the current PES, we see that as a consequence of correcting the long range behaviour in the AVTZ-level PES, the empirical adjustments also shift the curve closer to the AVQZ energies. We see similar behaviour for the s coordinate at large s , where the spacing is $\Delta s = 0.3$. This is expected, as the RKHS method interpolates the R and s coordinates in a similar manner.

The φ coordinate has a finite range, and is interpolated differently by the RKHS toolkit.³⁶ To check how the density of points along the φ coordinate affects the accuracy of the fit, we reconstruct the current PES, but instead of using 19 values of φ with a uniform spacing of $\Delta\varphi = 5^\circ$, we remove every other inner point, leaving only 10 values with a uniform spacing of $\Delta\varphi = 10^\circ$. In Fig. (3.9), we plot both the original empirically-adjusted PES and the “sparser” version of the PES, for three different (R, s) pairs. For $\varphi < 3\pi/8$, both PESs provide similar interpolations, but at larger φ , the interpolation becomes more sensitive to the density of the φ coordinate data. The sparser PES performs poorly at predicting the removed energies. We should note that we specifically chose values of R and s for which the relationship between the data spacing and the interaction accuracy is especially pronounced. For most values of R and s , such as those where the $(para\text{-H}_2)_3$ interaction energy is on the order of tens or hundreds of cm^{-1} , the decrease in accuracy from doubling the spacing $\Delta\varphi$ at large φ is far less drastic, usually on the order of $< 1\%$. This is a good sign, as it indicates that the interpolation quality is already well converged with respect to the density of data along the φ coordinate.

3.3.4 HCP Lattice Analysis

Consider all the triangles formed by a central “reference” $para\text{-H}_2$ molecule and any two of its twelve nearest neighbours inside an hcp lattice, and let α be the interior angle at the reference $para\text{-H}_2$ molecule of this triangle. We find a total of 66 triangles. There are, in order of increasing interior angle, 24 instances where $\alpha = 60^\circ$ (i.e. equilateral triangles), 12 instances where $\alpha = 90^\circ$ (i.e. right-angled triangles), 3 instances where $\alpha = 109.47^\circ$, 18 instances where $\alpha = 120^\circ$, 6 instances where $\alpha = 146.44^\circ$, and 3 instances where $\alpha = 180^\circ$ (i.e. three molecules along a straight line).

In Fig. (3.10), we show the $(para\text{-H}_2)_3$ interaction energy for each triangle as a function of the lattice constant, for the current PES and the ATM potential. In each instance, the *ab initio* interaction energy diverges from the ATM potential as the lattice constant

decreases. Interestingly, for each triplet where the ATM potential becomes more repulsive as the lattice spacing decreases, the *ab initio* (*para*-H₂)₃ interaction energy switches sign and becomes progressively more attractive, and vice-versa. This growing deviation shows the inadequacy of the ATM potential to study three-body interactions in solid *para*-H₂ at high densities.

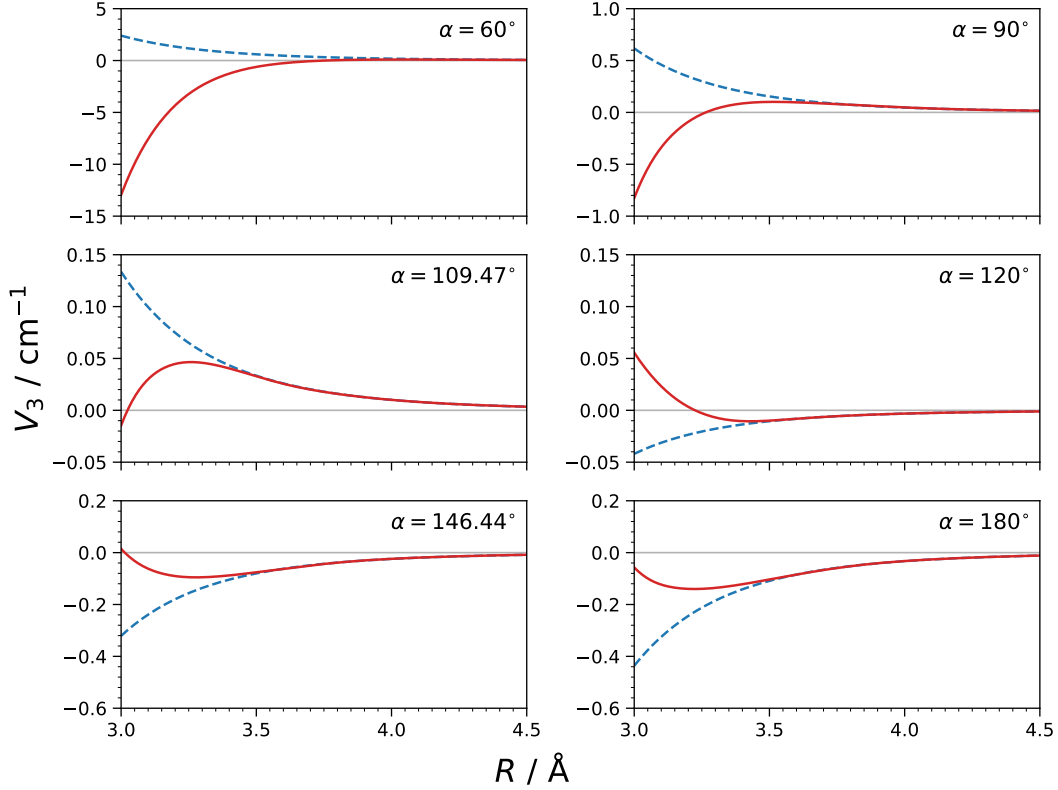


Figure 3.10: The (*para*-H₂)₃ interaction energy using the current PES (solid red line) and the ATM potential (dashed blue line) as a function of the lattice constant R , for the six types of triangles one can make using a central “reference” *para*-H₂ molecule and 2 of its 12 neighbours in a frozen *hcp* lattice. We ask readers to pay attention to the vertical scales of each subplot, as they generally differ for each triangle. Each subplot is labeled using its respective interior angle α of the triangle at the reference molecule. Note that each subplot corresponds to the (*para*-H₂)₃ interaction energy of a single triangle, not the sum of the contributions from all triangles with the same interior angle. For example, the subplot with the label $\alpha = 60^\circ$ is the (*para*-H₂)₃ interaction energy for a single equilateral triangle, not all 24 equilateral triangles of the 66 triangles considered.

At $R = 3 \text{ \AA}$, the strength of the $(para\text{-H}_2)_3$ interaction energy from the equilateral triangle configuration is the greatest of the nearest-neighbour triangles – it is an order of magnitude greater than that of the right-angled triangle, and at least two orders of magnitude greater than those of the other triangles. Together with its plurality among the aforementioned 66 triangles, we should expect the equilateral triangle to contribute the majority of the $(para\text{-H}_2)_3$ interaction energy in an *hcp* lattice.

Table 3.2: The average $(para\text{-H}_2)_3$ interaction energy (cm^{-1}) per particle between a central reference *para*-H₂ molecule and 2 of its 12 nearest neighbours in a frozen *hcp* lattice, for different lattice constants R ($\text{\AA}$). Shown are the contributions from all 66 triangles (ϵ^{nn}) and the 42 non-equilateral triangles ($\epsilon_{\text{neq}}^{nn}$), for the current PES and the ATM potential.

R	current PES		ATM	
	ϵ^{nn}	$\epsilon_{\text{neq}}^{nn}$	ϵ^{nn}	$\epsilon_{\text{neq}}^{nn}$
2.20	-5055.32	-428.30	333.57	22.72
2.40	-2051.03	-137.21	152.44	9.47
2.60	-801.71	-41.96	74.17	4.61
2.80	-300.24	-11.93	38.07	2.36
3.00	-106.64	-3.01	20.46	1.27
3.79	0.40	0.15	2.50	0.16

To see if this is the case, in Fig. (3.11), we compare the contribution to the average interaction energy per particle from all 66 triangles (ϵ^{nn}), and from only the 42 non-equilateral triangles ($\epsilon_{\text{neq}}^{nn}$), using the current PES and ATM potential. These calculations assume a frozen lattice. Select example energies are shown in Table 3.2. At the ZTZP lattice constant ($R = 3.79 \text{ \AA}$), the difference between ϵ^{nn} and $\epsilon_{\text{neq}}^{nn}$ is relatively small (0.40 cm^{-1} and 0.15 cm^{-1} , respectively). However, as the lattice constant decreases, the gap widens considerably, and the equilateral triangles dominate the contribution to the $(para\text{-H}_2)_3$ interaction energy per particle. This behaviour is caused, not only by the large outlying interaction energy of the equilateral triangle, but also by a partial cancellation of the interaction energies of the other 42 triangles. In contrast with the current *ab initio* PES, the ATM potential increases the energy per particle.

We can come up with a simple physically intuitive reason for why the interaction energy is the strongest for the equilateral triangle, and weakens as we increase the internal angle α to form other triangles. In an isotropic *para*-H₂–*para*-H₂ pair potential, the interaction energy strengthens as the intermolecular spacing decreases. Analogously, the isotropic

$(para\text{-H}_2)_3$ interaction energy strengthens as the total perimeter of the triangle decreases. This trend is apparent when looking at the ATM potential, given by Eq. (3.1). If we fix the pair distances $R_{12} = R_{23} = R$, and increase the internal angle α_2 from 60° to 180° , then the distance R_{13} increases from R to $2R$ and the denominator becomes 8 times larger. Similarly, we can see from Eq. (3.8) that if we fix s and φ (and thus constrain ourselves to a set of similar triangles), then the interaction potential increases as the perimeter (which is proportional to R) decreases.

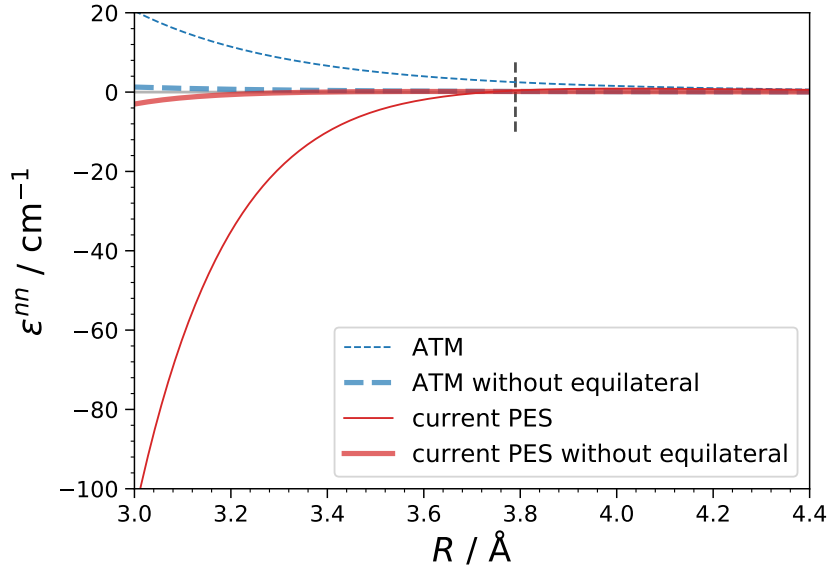


Figure 3.11: The average $(para\text{-H}_2)_3$ interaction energy per particle as a function of lattice constant R , between a central reference $para\text{-H}_2$ molecule and 2 of its 12 nearest neighbours in a frozen hcp lattice. The curves correspond to the contributions from all 66 triangles for the current PES (thin solid red line) and the ATM potential (thin dashed blue line), and the contributions from the 42 non-equilateral triangles for the current PES (thick solid red line) and the ATM potential (thick dashed blue line).

3.3.5 Comparison Against Silvera-Goldman Potential

How will the introduction of an *ab initio* $(para\text{-H}_2)_3$ PES influence simulations of condensed $para\text{-H}_2$ systems? To estimate its influence, we calculate the energy per molecule ϵ as a function of the density ρ of a solid $para\text{-H}_2$ hcp lattice at $T = 0$ K with no zero-point

motion. From this equation of state (EOS), we calculate the pressure using

$$P = \rho^2 \left. \frac{\partial \epsilon}{\partial \rho} \right|_T. \quad (3.26)$$

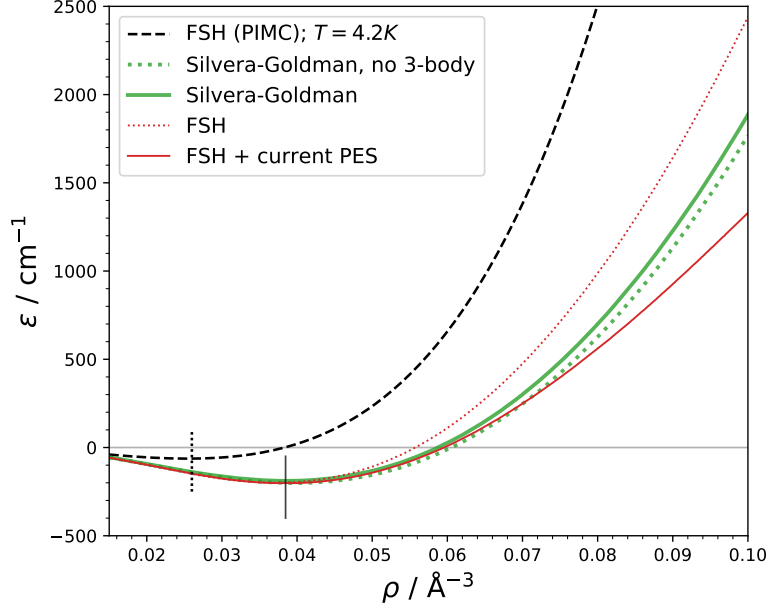


Figure 3.12: The energy per particle ϵ as a function of density ρ for a frozen solid *para*-H₂ *hcp* lattice. The curves correspond to the interaction combinations of the FSH potential (red, dotted, thin), the (FSH + current PES) combination (red, solid, thin), the SG potential (green, solid, thick), and the SG potential without its three-body approximation term (green, dotted, thick). Also shown is the curve for the FSH potential calculated from a PIMC simulation (black, dashed) at $T = 4.2\text{ K}$.³⁷ The shift in the energy if the simulation had been done at $T = 0\text{ K}$ would not be visible on the scale of this figure. The dotted and solid vertical lines indicate the equilibrium density of the PIMC-simulated FSH curve ($\rho = 0.0255\text{ \AA}^{-3}$), the classical FSH curve ($\rho = 0.0370\text{ \AA}^{-3}$) and (FSH + current PES) curve ($\rho = 0.0385\text{ \AA}^{-3}$).

We perform this frozen lattice calculation using four different combinations of two-body and three-body *para*-H₂ intermolecular potentials. These are the *para*-H₂–*para*-H₂ FSH potential on its own, the FSH potential with the current (*para*-H₂)₃ interaction PES, the

SG potential, and the SG potential without its three-body approximation term. The EOS and pressure curves are shown in Figs. (3.12) and (3.13), respectively.

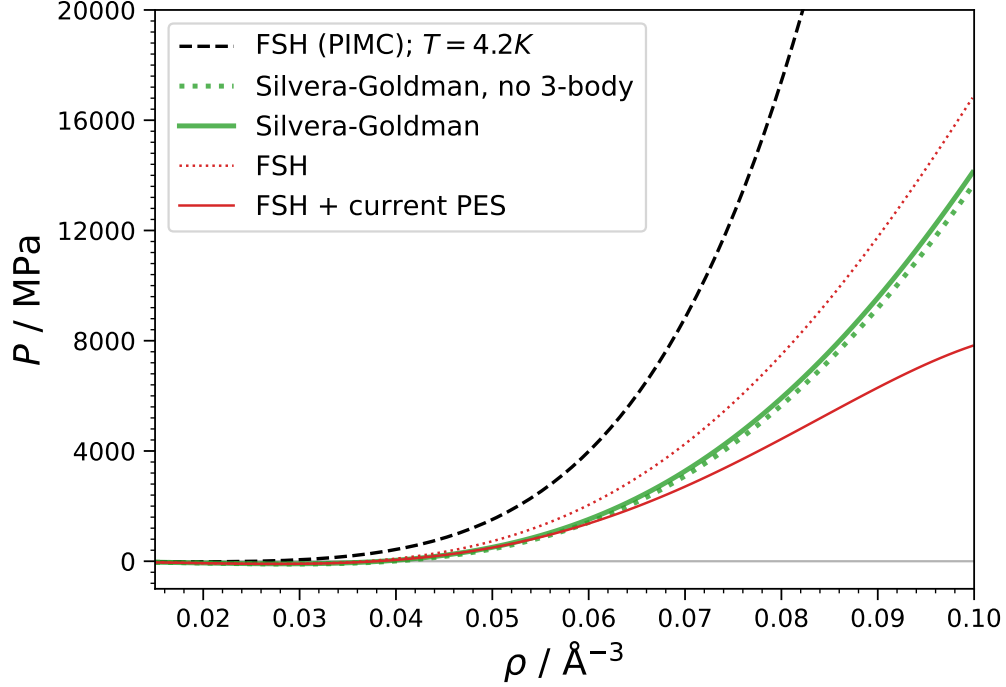


Figure 3.13: The pressure P as a function of density ρ for a frozen solid $para\text{-H}_2$ hcp lattice. The curves follow the same labelling in Fig. (3.12).

Each of the frozen lattice EOS curves predict a much higher ZTZP equilibrium density and a much lower energy per particle than seen experimentally ($\rho = 0.026 \text{ \AA}^{-3}$ and $\epsilon = -62.5 \text{ cm}^{-1}$, respectively).^{9,42} This discrepancy arises because the frozen lattice calculation does not account for the lattice inflation caused by the zero-point motion of the $para\text{-H}_2$ molecules. We see in Fig. (3.13) that at high densities, compared to the SG potential, the FSH potential on its own predicts a much steeper pressure curve, while the (FSH + current PES) combination decreases the pressure. The former observation is in line with previous findings,³⁷ where, due to its hard core, the FSH potential greatly overestimates the pressure when $0.02 \text{ \AA}^{-3} < \rho < 0.04 \text{ \AA}^{-3}$. The latter observation is a good sign, as we know that the SG potential itself begins to overestimate the pressure when extended past $\rho = 0.1 \text{ \AA}^{-3}$.^{13,108}

3.4 Conclusion

We have presented a 3D isotropic *ab initio* (*para*-H₂)₃ interaction energy PES, calculated at the CCSD(T) level of theory using an AVTZ atom-centred basis set with an additional (*3s3p2d*) midbond function. We have presented in detail the procedure used to construct the PES. The RKHS method interpolates between the data mesh. At long intermolecular separations, the (*para*-H₂)₃ interaction energy converges to the ATM potential. In situations where two *para*-H₂ molecules are close together and the third is far away, the (*para*-H₂)₃ interaction takes the form of a modified version of the ATM potential. At short intermolecular spacings, the (*para*-H₂)₃ interaction energy trends exponentially with R .

The (R, s, φ) coordinate system is a convenient and physically-insightful way to represent the (*para*-H₂)₃ system, and we recommend its use for other three-body systems of identical particles. The accuracy of the interpolation was found to be more sensitive to the input data mesh spacings at small values of s and R , and large values of φ . Switching from the 6-point to the 14-point Lebedev quadrature provides only minor improvements to the energy.

Among a *para*-H₂ molecule and its nearest neighbours in an *hcp* lattice, the equilateral triangle configuration contributes the majority of the (*para*-H₂)₃ interaction energy. For many applications, such as computer simulations of solids, one might sufficiently approximate the (*para*-H₂)₃ interaction energy using only equilateral and “almost-equilateral” triangles. Using the (R, s, φ) coordinate system, this corresponds to creating a PES using an input data mesh with a full range of R values, but only a small range of s values past $s = 1$, and a small range of φ values below $\varphi = \pi/2$. This approximation would greatly reduce the coordinate space one would need to cover to create an adequate three-body PES.

The use of the current (*para*-H₂)₃ interaction PES alongside the FSH potential decreases its EOS and pressure for a frozen *hcp* lattice at high densities to below that of the SG potential. This indicates, albeit only qualitatively, that the (FSH + current PES) combination is a more powerful model than the SG potential for predicting the properties of many-body *para*-H₂ systems. It has yet to be seen whether the improvements actually occur when the zero-point motion of the *para*-H₂ molecules is taken into account, as in a path integral simulation. Such a project is currently being carried out in our group, for both solid *para*-H₂ and clusters of *para*-H₂ molecules. It is also possible that the influence of the current three-body PES on the structural properties of many-body *para*-H₂ systems can affect these systems’ vibrational excitation shifts.^{16,37}

Chapter 4

Equation of state of solid parahydrogen using *ab initio* two-body and three-body interaction potentials

This chapter is reproduced from Alexander Ibrahim and Pierre-Nicholas Roy, “Equation of state of solid parahydrogen using *ab initio* two-body and three-body interaction potentials” Journal of Chemical Physics **157**, 174503 (2022), with the permission of AIP Publishing.¹²³

4.1 Introduction

Solid *para*-H₂ is the simplest quantum molecular solid that exists in nature. It is the target of a sizeable body of active research.^{8,124–128} As a quantum solid, solid *para*-H₂ possesses many interesting and exotic physical properties. It has a triple point of $(T, P) = (13.8 \text{ K}, 7.042 \text{ kPa})$.¹²⁹ Upon freezing, solid *para*-H₂ exists in its stable equilibrium form as a hexagonal close-packed (*hcp*) lattice. Due to its low mass, each *para*-H₂ molecule has a very large zero-point energy, and is very delocalized about its nominal lattice site. At the experimental equilibrium density of $\rho_0 = 0.0260 \text{ \AA}^{-3}$, the Lindemann ratio of solid *para*-H₂ is about 0.2.^{6,8} These zero-point motions expand the lattice to the extent that, at the equilibrium density, the *para*-H₂ molecules are pushed to a mean distance beyond the pair potential’s minimum. As a result, solid *para*-H₂ is highly compressible.⁹

As a molecular solid, a key property of solid *para*-H₂ is that the individual *para*-H₂ molecules within retain most of their free molecule properties. The interactions of an individual *para*-H₂ molecule with its surroundings in the solid only slightly perturb its behaviour from that of a free molecule. For example, except at very high pressures, the free molecule rotational and vibrational quantum numbers j and ν remain good quantum numbers in the solid. We can readily model solid *para*-H₂ using interactions between collections of free *para*-H₂ molecules, such as pairs and triplets.

Because of its simplicity and relative ease of modelling compared to more complicated systems, solid *para*-H₂ is a convenient benchmark against which we can test our current theoretical knowledge of quantum many-body physics. Constructing the equation of state (EOS) is an excellent way to test the accuracy of a potential energy surface (PES) for *para*-H₂ under different density and state conditions.

Most simulations of bulk condensed *para*-H₂ are done using static point-wise pair interaction potentials. Such potentials have many advantages. For example, they are computationally fast and easy to implement; a dense lookup table and linear interpolation are often enough. The Silvera-Goldman (SG) and Buck potentials are two of the most popular two-body potentials.^{62,66} Both provide very good predictions at low and moderate densities, but overestimate pressures at higher densities.¹³ The Faruk-Schmidt-Hinde (FSH) potential is a more recent *ab initio* first principles pair potential,^{16,31} and due to its construction, it contains no implicit many-body contributions. Path-integral Monte Carlo (PIMC) simulations show that, even at relatively low densities (below 0.04 Å⁻³) the FSH potential overestimates the pressure of solid *para*-H₂ even more drastically than do the SG and Buck potentials.³⁷

Each of the aforementioned SG, Buck, and FSH potentials has a core wall that is too rigid, causing them to overestimate the pressure of solid *para*-H₂ at high densities. In 2013, Omiyinka and Boninsegni¹³ modified the Moraldi potential¹⁰⁸ to create a new pair potential, whose highlighting feature is its drastically softened core wall when compared to the SG potential. Simulations using this pair potential reproduce the experimental pressure-density curve for *para*-H₂ remarkably well between the densities of 0.076 Å⁻³ and 0.273 Å⁻³.

Several computational studies of condensed systems of *para*-H₂ emphasize the importance of three-body interactions at high densities.^{11-13,37} These three-body interactions are attractive at short intermolecular separations, and their inclusion effectively softens the core wall of the pair potential. The Axilrod-Teller-Muto (ATM) potential is a commonly-referenced potential for triple-dipole interactions.¹⁷ Unfortunately, despite having the correct long-range behaviour, it fails to reproduce the desired short-range behaviour even

qualitatively. The ATM potential predicts a net repulsive interaction at high densities, the opposite of what is required.

This research group recently published a static *ab initio* three-body interaction potential energy surface (PES) for *para*-H₂.⁸¹ This new PES has the expected net attractive interaction at short ranges, as well as the expected long-range behaviour of the ATM potential.

We can also look at studies of a closely related system, solid helium-4, for additional insights on solid *para*-H₂. Solid helium-4 is a quantum molecular solid above 25 bar, its atoms have large zero-point motions and interact weakly through dispersion interactions, and, like solid *para*-H₂, it is commonly simulated using quantum Monte Carlo techniques. In 2017, Barnes and Hinde performed variational Monte Carlo (VMC) and path-integral Ground State (PIGS) simulations of solid helium-4.¹³⁰ They used the Aziz-87 pair potential,¹³¹ and the three-body potential of Cencek *et al.*¹³² Simulations using both the pair and three-body potentials reproduce the experimental pressure-volume data down to about 4.0 cm³/mol, whereas those with only the pair potential were successful only down to about 8.0 cm³/mol. Given the similarities between solid helium-4 and solid *para*-H₂, the results of Barnes and Hinde are an indication that three-body interactions may greatly affect the EOS of solid *para*-H₂.

In this work, we compute the EOS of solid *para*-H₂ at $T = 4.2$ K using PIMC simulations. The system is constructed as an *hcp* lattice between the densities of 0.024 Å⁻³ and 0.1 Å⁻³. We use a potential tail calculation, for both the pair and three-body potentials, to correct the finite system size error. We also explore how these tail corrections vary as a function of density. Depending on the requirements, as described in Sec. 4.3, we correct the Trotter factorization error either via extrapolation, or by choosing a suitably small imaginary time step.

Following the work by Barnes and Hinde¹³⁰ we explore two different ways of including the effects of the three-body interaction into our simulations. The first is the full inclusion method, in which both the pair and three-body potentials are used for the Monte Carlo sampling as well as in the energy estimators. The second is the perturbative method, in which only the pair potential is used for the Monte Carlo sampling, but both potentials are used in the energy estimators. The key draw of the perturbative method is that it drastically reduces the computational costs of sampling. The perturbative case also gives us, with no additional effort, the EOS for which only the pair potential is used in the estimator. We also look at how each of these two methods of incorporating the three-body interaction affects the structure of solid *para*-H₂. For brevity, in this paper we will refer to “simulations performed with the perturbative treatment of the three-body potential” as

“perturbative treatment simulations”, and “simulations performed with the full inclusion treatment of the three-body potential” as “full inclusion treatment simulations.”

The remainder of this paper is structured as follows. In Sec. 4.2 we present the model of the system, including the Hamiltonian used, the properties and construction of the two-body and three-body PESs, and a more detailed description of the perturbative and full inclusion treatments. In Sec. 4.3, we provide a review of the computational techniques used in the simulations. In Sec. 4.4, we present the results and a discussion thereof. Finally, we present our conclusions and suggestions for future work.

4.2 Model

We model our system as a collection of N point-wise *para*-H₂ molecules in three dimensions. They are contained in an *hcp* lattice within a rectangular prism, with periodic boundary conditions applied in all directions. The Hamiltonian for our system is

$$\hat{H} = -\frac{\hbar^2}{2m} \sum_{i=1}^N \nabla_i^2 + \sum_{i<j}^N V_2(r_{ij}) + \sum_{i<j<k}^N V_3(r_{ij}, r_{ik}, r_{jk}), \quad (4.1)$$

where V_2 is the FSH pair interaction, V_3 is the *ab initio* three-body interaction,⁸¹ and r_{ij} is the relative distance between the centres of mass of molecules with labels i and j .

The FSH potential is a static *ab initio* pair interaction potential for *para*-H₂. In 2014, Faruk *et al.* applied the adiabatic hindered rotor (AHR) method^{29,73,133–136} to a 6-dimensional H₂–H₂ dimer potential created by Hinde.^{16,25,31} This process produced several pointwise 1D pair potentials, each of whose only parameter is the relative distance between each molecule’s centre of mass. These potentials have been used to successfully simulate and predict the experimentally observed first vibrational shifts of small clusters of *para*-H₂ molecules.^{16,31,137} This AHR approximation has also recently been successfully applied to help construct neural network interaction potentials.¹³⁸

We refer to one of these aforementioned potentials, that which gives the ground state interaction energy for two *para*-H₂ molecules, as the FSH potential. As a purely *ab initio* pair potential, the FSH potential does not implicitly incorporate higher many-body effects. It has both a deeper attractive well and a much more repulsive core than the SG and Buck potentials. When used on its own, it greatly overestimates the pressure of solid *para*-H₂, even at relatively low densities (below 0.04 Å^{−3}).³⁷

This research group recently published a static *ab initio* three-body PES for *para*-H₂.⁸¹ Its energies are calculated at the correlated coupled-cluster theory level, with single, double,

and perturbative triple excitations. The calculations were performed using an augmented correlation-consistent triple zeta atom-centred basis set, supplemented by a $(3s3p2d)$ mid-bond basis. The PES is isotropic. Each hydrogen molecule is fixed to the vibrationally averaged ground state bond length of 1.449 Å. The angular degrees of freedom are projected out using the $L = 3$ Lebedev quadrature,³² in which each effective interaction energy is the average interaction energy of 27 space-fixed angular orientations. The PES is constructed using the reproducing kernel Hilbert Space (RKHS) toolkit by Unke *et al.*³⁶ The RKHS method is a machine learning method that has seen much success constructing multi-dimensional PESs.^{139–142} In addition, we apply a number of phenomenological modifications and extreme-range extensions to the PES. These modifications ensure that the potential converges smoothly to the ATM potential at large intermolecular separations, and that its energies increase in magnitude exponentially at short intermolecular separations. The training data for the three-body potential only goes down to triangles with at least one side length greater than 2.2 Å. This is very similar to the lower limit for the FSH potential. Luckily, even at the highest densities relevant to the findings of this paper, the *para*-H₂ molecules spend very little time at such short distances.

In principle, because *para*-H₂ is a boson, we should treat it as an indistinguishable particle that obeys Bose statistics. In the context of a PIMC simulation, we account for Bose statistics using the worm algorithm,^{143,144} which expresses the permutation of identical bosons via the connection and exchange of beads between the ring polymers. We find, however, that even at the highest densities explored here, that bosonic exchanges between the molecules are almost entirely absent. Thus in practice, we can treat the *para*-H₂ molecules as distinguishable particles.

Inspired by previous work by Barnes and Hinde,¹³⁰ we will compare results from two different methods, each of which incorporates the three-body interactions in a different way: the perturbative method and the full inclusion method. With no additional effort, the perturbative treatment simulations also give us results for the case where the three-body potential is used in neither sampling nor the estimators, i.e. pure two-body potential simulations. These methods are described in more detail below.

4.2.1 Full Inclusion Method

In the full incorporation method, we use the pair and three-body potentials in both the configuration sampling stage and the energy estimators. During sampling, only three-body interactions between a given molecule and its 12 nearest neighbours are taken into account. This is an appropriate approximation; the three-body interaction falls off quickly enough

that, even at the highest densities considered in this paper, the contribution from the pair potential beyond the nearest neighbours is typically 2 to 3 orders of magnitude greater than that from the three-body potential. When using the energy estimator, we follow the Attard minimum image convention¹⁴⁵ and include all three-body interactions in the lattice up to the cutoff distance of $L/2$, where L is the shortest length of the simulation box. The explicit inclusion of the three-body potential in the sampling stage greatly increases the computational costs. The evaluation of the three-body potential itself is more expensive than for pair potentials, and even when limited to the nearest-neighbours, there are 66 interactions to account for per molecule. Compared to using pair interactions alone, the inclusion of the three-body interactions increases the computational cost of the PIMC simulations by a factor of around 40.

4.2.2 Perturbative Method

In the perturbative method, we use only the pair potential during configuration sampling, but use both the pair and three-body potentials for the energy estimators. Like in the full inclusion method, the energy estimators include all three-body interactions up to the cutoff distance $L/2$. Including the three-body potential perturbatively incurs a near negligible computational cost to the simulation. This is because, relative to the sampling stage, the estimation stage takes up a very small portion of the simulation time. To understand the accuracy trade-off of the perturbative method, we should remember that the three-body potential is attractive at short ranges. Its full inclusion in the sampling stage makes high density configurations less energetically unfavourable, and thus we should expect a change the positional distributions of the molecules in the solid. The accuracy of the perturbative method relies on the assumption that this change in positional distributions is very small. Our findings suggest that this assumption is sound (see Secs. 4.4.3 and 4.4.4). This assumption also holds in solid helium-4. Barnes and Hinde¹³⁰ found that, when compared to the perturbative method, the full inclusion of the three-body interaction only slightly shifted the pair distribution of solid helium-4 towards shorter distances. This holds even up to the relatively high density of $0.15 \text{ \AA}^{-3}$.

In the remainder of this paper, to describe the category of simulation that the presented results come from, we will use the following labels. Results from PIMC simulations with pair interactions only, for both the sampling and the energy estimator, are labelled as FSH-[2B]. Results from simulations with pair interactions and perturbatively incorporated three-body interactions are labelled as FSH-[2B(3B)]. Results from simulations with pair interactions and fully incorporated three-body interactions are labelled as FSH-[2B+3B].

4.3 Computation

4.3.1 Path Integral Monte Carlo

Consider a system of particles described by a Hamiltonian $\hat{H}$ at a finite temperature $\beta = 1/k_B T$. The expectation value of the operator $\hat{A}$ in this system can be described using

$$\begin{aligned}\langle \hat{A} \rangle &= \frac{1}{Z} \text{Tr} \{ \hat{A} e^{-\beta \hat{H}} \} \\ &= \frac{1}{Z} \int d\mathbf{q} d\mathbf{q}' \langle \mathbf{q} | \hat{A} | \mathbf{q}' \rangle \langle \mathbf{q}' | e^{-\beta \hat{H}} | \mathbf{q} \rangle,\end{aligned}\quad (4.2)$$

where $e^{-\beta \hat{H}}$ is the density matrix, $Z = \text{Tr} \{ e^{-\beta \hat{H}} \}$ is the partition function, and $\mathbf{q}$ represents the positions of all particles in the system.

Using the path integral formulation we take the identity operator resolved onto the position space,

$$\hat{I} = \sum |\mathbf{q}\rangle \langle \mathbf{q}| \quad (4.3)$$

and apply it P times to discretize the integral Eq. (4.2) in imaginary time

$$\langle \mathbf{q}' | e^{-\beta \hat{H}} | \mathbf{q} \rangle = \int \prod_{i=2}^P d\mathbf{q}_i \times \prod_{i=1}^P \langle \mathbf{q}_i | e^{-\tau \hat{H}} | \mathbf{q}_{i+1} \rangle, \quad (4.4)$$

where $\tau = \beta/P$ is the imaginary time step, and the index i labels the individual time steps. We apply the boundary conditions $\mathbf{q}_1 = \mathbf{q}'$ and $\mathbf{q}_{P+1} = \mathbf{q}$, thus forming a ring of coordinates $\{\mathbf{q}_i\}$. We can sample the integral Eq. (4.2) with the PIMC method.³⁴

4.3.2 Trotter Error

Note that when we separate the Hamiltonian in Eq. (4.4) into potential and kinetic components, we introduce a systematic Trotter factorization error into our expectation values.^{34,74} This error is a function of the imaginary time step size τ , and vanishes in the limit $\tau \rightarrow 0$. A previous study of the EOS of solid *para*-H₂ shows that this Trotter factorization error is significant.³⁷

One method to eliminate the Trotter error is to perform several simulations, each with a different value of τ . Each simulation produces a different expectation value $\langle \hat{A} \rangle_\tau$. We then fit these expectation values to the equation⁷⁴

$$\langle \hat{A} \rangle_\tau = \langle \hat{A} \rangle_0 + b_2 \tau^2 + b_4 \tau^4, \quad (4.5)$$

where $\langle \hat{A} \rangle_\tau$ is the expectation value of $\hat{A}$ in the limit $\tau \rightarrow 0$, and b_2 and b_4 are fit parameters. When both the kinetic and potential energies are small in magnitude, the systematic Trotter errors are small, and we can successfully use Eq. (4.5) using relatively large values of τ for fit values. When the kinetic or potential energies are large, we need smaller values of τ to use Eq. (4.5), and this extrapolation method becomes less advantageous.

Another approach is to select a small enough value of τ such that the Trotter error becomes tolerably small. This is much simpler, but the resulting kinetic energy energies have a much larger variance.

We use both the “small τ ” and “ τ extrapolation” methods, each on a different range of densities, to calculate different features of the EOS.

We use the “small τ ” method, using $P = 960$ beads ($\tau \approx 2.48 \times 10^{-4} \text{ K}^{-1}$), to calculate the energies in a sparse grid between $\rho = 0.024 \text{ \AA}^{-3}$ and $\rho = 0.1 \text{ \AA}^{-3}$. We find that, compared to the extrapolated value, this choice of time step leads to a Trotter error in the total energy of about 0.2 % at $0.026 \text{ \AA}^{-3}$ and $0.04 \text{ \AA}^{-3}$, and about 2.5 % at $0.1 \text{ \AA}^{-3}$. As we will see in later sections, the simulation results at high densities are incredibly divorced from the experimental results, and thus the increased systematic Trotter error at high densities should not be too concerning. The standard error of the mean of the energies from the “small τ ” method is on the order of 0.1 cm^{-1} , whereas the energies in concern span a range of nearly 4000 cm^{-1} (see Sec. 4.4.1).

We use the “ τ extrapolation” method to calculate the energies in a fine grid near the equilibrium density (around $\rho_0 = 0.026 \text{ \AA}^{-3}$). We perform multiple simulations at each density, using $P = \{64, 80, 96, 128, 192\}$ beads, and extrapolate the energies to $\tau \rightarrow 0$ using Eq. (4.5). The energies around the equilibrium density are low enough for the extrapolation to work well, and we need low variance results to accurately find the equilibrium density. The standard error of the mean of the energies from the “ τ extrapolation” method is on the order of 0.01 cm^{-1} , whereas the energies around the equilibrium density span a range of only about 2 cm^{-1} .

4.3.3 Finite System Size Error

In systems with periodic boundary conditions, we must apply minimum image rules to ensure the correct distances between molecules are chosen. For example, due to the periodicity, two molecules near opposite edges of the simulation box are actually very close together, and our simulations must account for this. For the pair potential, we apply the standard minimum image convention with a cutoff distance of $L/2$, where L is the shortest

length of the simulation box. Interactions between pairs of molecules greater than $L/2$ apart are ignored. For three-body interactions, this minimum image convention does not work – it has a tendency to underestimate the size of triangles. There are two common ways to overcome this. Some authors use a cutoff distance of $L/4$. However, at high densities, this distance becomes too short to give accurate results.¹⁴⁶ We instead apply the modifications to the minimum image convention developed by Attard,¹⁴⁵ which lets us keep all triangles for which each of its side lengths is less than the $L/2$ cutoff.

Our simulations are performed using a finite number of molecules N inside a box of side length L . Thus the energy estimators cannot account for interactions with molecules outside of the simulation box, and our potential energy expectation values suffer from systematic finite system size errors. We correct these errors by applying potential tail corrections for both the two-body and three-body potentials.⁷⁸

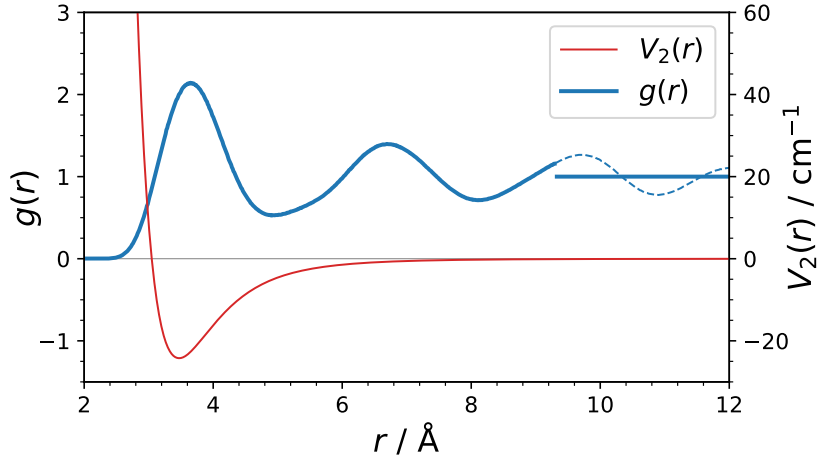


Figure 4.1: The radial pair distribution function $g(r)$ (blue) from a simulation of $N = 448$ *para*-H₂ molecules using $P = 64$ beads, at a density of $\rho = 0.026 \text{ \AA}^{-3}$. Also shown is the FSH pair potential $V_2(r)$ (red). In the region $r > L/2 \approx 9.3 \text{ \AA}$, the thin, dashed blue line is the $g(r)$ of the $N = 448$ simulation, and the thick, horizontal blue line represents the tail correction made when we instead use a $g(r)$ from a simulation with $N = 180$ molecules.

To calculate the tail correction energy $\epsilon_N^{(t)}$ for the pair potential, we assume that the radial pair distribution function $g(r)$ outside our simulation box is uniform (equal to unity),

and then integrate the pair potential outside this box.

$$\begin{aligned}\epsilon_N^{(t)} &= \frac{N\rho}{2} \int_{L/2}^{\infty} d^3r V_2(r)g(r) & (\text{pair}) \\ &\approx \frac{N\rho}{2} \int_{L/2}^{\infty} d^3r V_2(r) & (4.6)\end{aligned}$$

In Fig. 4.1, we show the radial pair distribution function $g(r)$ at the equilibrium density of $\rho = 0.026 \text{ \AA}^{-3}$, from a simulation of $N = 448$ *para*-H₂ molecules, using $P = 64$ beads. We also show the FSH pair potential $V_2(r)$. For $N = 448$ at this density, the cutoff distance is roughly $12.3 \text{ \AA}$. The radial pair distribution function for a simulation with $N = 180$ molecules is effectively identical to its shown counterpart with $N = 448$ molecules, but only until its own shorter cutoff distance of $L/2 \approx 9.3 \text{ \AA}$. The thin, oscillating, dashed blue line beyond the cutoff distance $L/2 \approx 9.3 \text{ \AA}$ in Fig. 4.1 is the “tail” that we want to approximate. When we set $g(r) = 1$ for $r > L/2$ in Eq. (4.6), we can represent this visually in Fig. 4.1, by replacing the “true tail,” (the thin, oscillating blue line), with the thick, horizontal blue line that overlaps it.

The tail correction is the most effective when performed at distances for which the oscillations in the radial pair distribution function, and the magnitude of the pair potential, are both relatively small. For example, at the equilibrium density, for a simulation with $N = 180$ molecules, we see in Fig. 4.1 that the tail correction would begin well into the r^{-6} decay of the pair interaction. However, we will soon see that the tail correction is still reasonable even at the highest densities explored in this paper, especially when compared with the total energy scales of the calculations.

For the three-body interaction tail correction we follow Eq. (2.150) in Ref. [78], and account for interactions from all triangles with at least one side length greater than $L/2$. This calculation requires the three-body distribution function, which, following the aforementioned reference, we approximate using a product of three pair distribution functions $g(r)$ from the simulation.

To test how well the tail corrections work, we perform perturbative treatment simulations of solid *para*-H₂ for two system sizes, $N = 180$ and $N = 448$. These simulations take place at temperature $T = 4.2 \text{ K}$, using $P = 64$ time slices, and for 31 density values between $0.024 \text{ \AA}^{-3}$ and $0.1 \text{ \AA}^{-3}$. In Fig. 4.2, we show the pair and three-body energy corrections for both lattice sizes. As expected, the tail corrections are smaller in magnitude for the larger $N = 448$ case than for the smaller $N = 180$ case. This is expected, because a larger system should have a smaller system size error to correct for. At all considered densities, the pair tail correction is attractive, while the three-body tail correction is repulsive. In addition, the pair tail correction is always greater in magnitude than that of the

three-body tail correction. This reflects the fact that, at large distances, the pair potential decays with an attractive $-C_6 r^{-6}$ interaction energy, while the three-body potential decays with a weaker, repulsive $C_9 r^{-9}$ interaction energy.

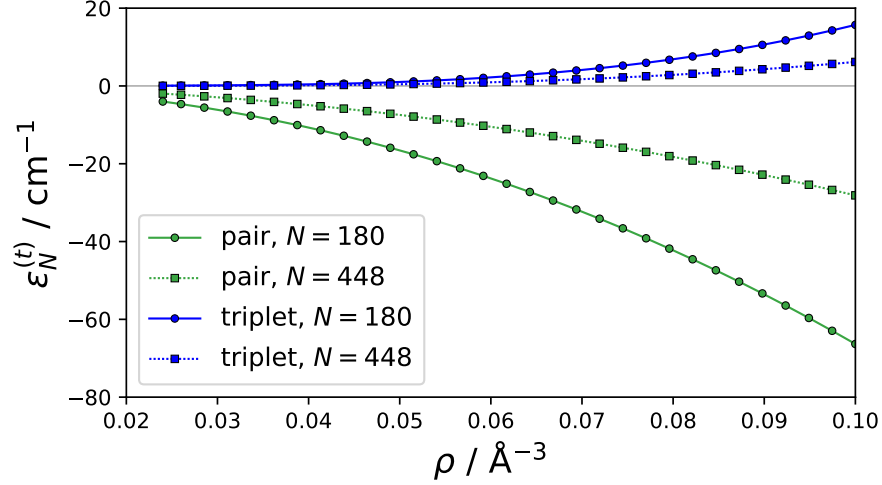


Figure 4.2: The potential tail correction energy per molecule $\epsilon_N^{(t)}$ as a function of density ρ . These energies are added to the simulation energies to correct for the finite system size error. Shown are the correction energies for simulations of $N = 180$ particles (solid, circles) and $N = 448$ particles (dotted, squares), separated into the pair interactions (green) and the three-body interactions (blue).

From the simulations we recover the average energy per particle for a system of N particles, $\epsilon_N^{(s)}$, for which the tail correction has not been applied. We now consider two cases. We can calculate the total energy per particle either with the tail correction, *i.e.* $\epsilon_N \approx \epsilon_N^{(s)} + \epsilon_N^{(t)}$, or without the tail correction, *i.e.* $\epsilon_N \approx \epsilon_N^{(s)}$. We then calculate the difference in the energy per particle of both system sizes, $\epsilon_{180} - \epsilon_{448}$, for both the corrected and uncorrected cases. If the tail correction is effective, we should expect that it makes up for the difference in the simulation sizes, and thus $\epsilon_{180} - \epsilon_{448}$ should be close to zero. We can verify this in Fig. 4.3. Without the tail correction, the energy gap per particle $\epsilon_{180} - \epsilon_{448}$ grows with the density. The application of the tail correction suppresses the energy difference almost completely at lower densities, and still relatively well at the highest densities considered. Thus with the tail corrections we can justify performing simulations using only $N = 180$ particles.

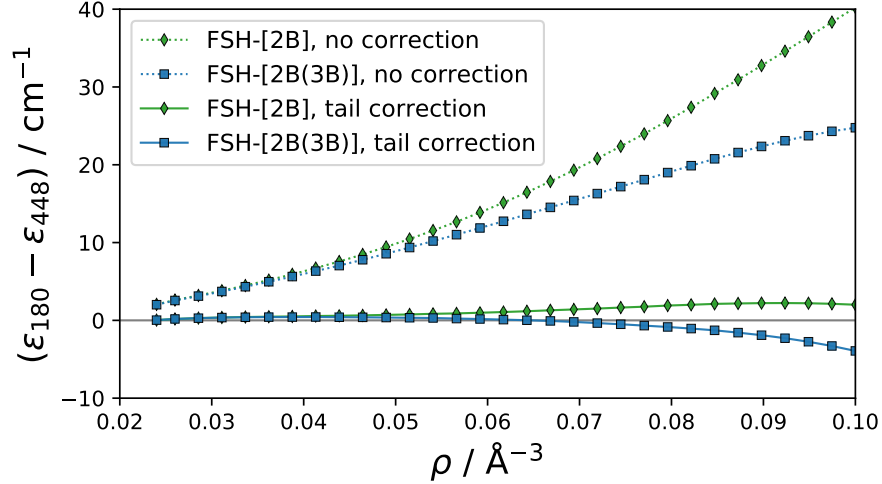


Figure 4.3: The difference in the energy per molecule calculated from simulations using $N = 180$ particles (ϵ_{180}) and $N = 448$ particles (ϵ_{448}). The differences are calculated for both the FSH-[2B] energies (green, diamonds) and the FSH-[2B(3B)] energies (blue, squares). The solid curves represent the cases where simulations of both sizes have had their respective potential tail corrections applied before the difference is taken (*i.e.* $\epsilon_N \approx \epsilon_N^{(s)} + \epsilon_N^{(t)}$). The dotted curves represent the cases where the tail corrections have been ignored (*i.e.* $\epsilon_N \approx \epsilon_N^{(s)}$). For the FSH-[2B] case, the tail correction $\epsilon_N^{(t)}$ is made up of only the pair tail contribution in Fig. 4.2, whereas for the FSH-[2B(3B)] case, $\epsilon_N^{(t)}$ is made up of both the pair and three-body tail contributions. Having an energy difference close to zero represents a successful finite system size correction.

4.4 Results

4.4.1 Energy-Density Equation of State

We perform PIMC simulations of solid *para*-H₂ at $T = 4.2$ K, between the densities of $0.024 \text{ \AA}^{-3}$ and $0.1 \text{ \AA}^{-3}$ for the perturbative simulations, and between the densities of $0.0617 \text{ \AA}^{-3}$ and $0.1 \text{ \AA}^{-3}$ for the full inclusion simulations. Due to the expense of the full inclusion calculations, they are only carried out at the higher densities, where the influence of the three-body potential is the greatest. We use the standard potential energy estimators for both the pair and three-body potential energies, taking into account periodic

boundary conditions and potential tail corrections, as discussed in Sec. 4.3. We use the primitive estimator for the kinetic energy.

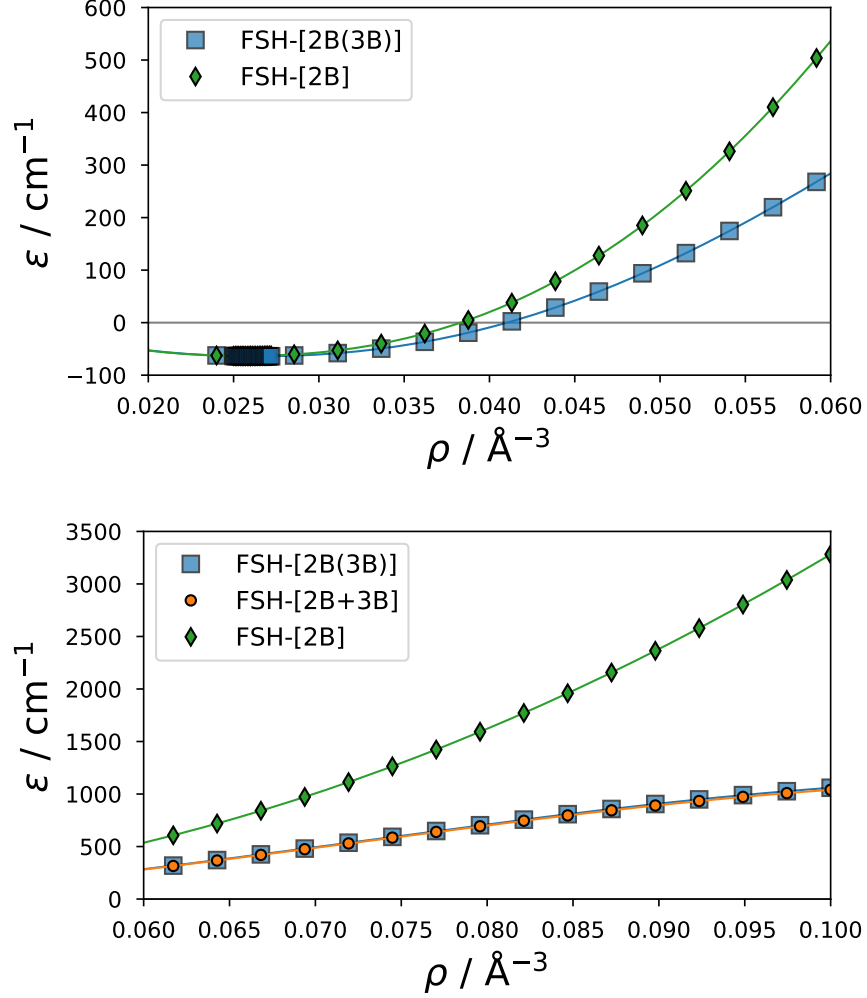


Figure 4.4: The energy per molecule ϵ as a function of the density ρ of solid *para*-H₂. The data points are shown for simulations performed under the FSH-[2B] conditions (green, diamonds), the FSH-[2B(3B)] conditions (blue, squares), and the FSH-[2B+3B] conditions (orange, circles). The curves represent the fitting of the data to a modified Birch equation, Eq. (4.7). All energies were calculated using the “small τ ” method to account for the Trotter error, except for those in the densely packed region around $\rho_0 = 0.026 \text{ \AA}^{-3}$, which were calculated using the “ τ extrapolation” method.

We fit the FSH-[2B] and FSH-[2B(3B)] total energies per particle to a modified Birch EOS¹⁴⁷

$$\epsilon(\rho) = \epsilon_0 - P_0/\rho + \sum_{n=1}^5 \tilde{\kappa}_n \rho^{2n/3} . \quad (4.7)$$

The fit parameters ϵ_0 , P_0 , and $\{\tilde{\kappa}_n\}$ are shown in Table 4.1. For the purposes of being able to create a curve for the pressure-density figures in Sec. 4.4.2, we also fit the FSH-[2B+3B] total energies to Eq. (4.7). However, because the energies for the FSH-[2B+3B] simulations were collected for a more limited range of densities, the fit parameters are very highly correlated to one another and are not very meaningful.

Table 4.1: Fit parameters for the energy-density EOS, Eq. (4.7), for the FSH-[2B] and FSH-[2B(3B)] simulation sets.

	FSH-[2B]	FSH-[2B(3B)]
$\epsilon_0(\text{cm}^{-1})$	16.78 ± 0.87	18.67 ± 0.78
$P_0(\text{cm}^{-1} \text{ \AA}^{-3})$	-4842.4 ± 222.7	-5763.3 ± 200.8
$\kappa_1(\text{cm}^{-1} \text{ \AA}^2)$	118465.1 ± 5070.8	149565.7 ± 4572.1
$\kappa_2(\text{cm}^{-1} \text{ \AA}^4)$	-1312002.4 ± 53696.9	-1789644.0 ± 48415.6
$\kappa_3(\text{cm}^{-1} \text{ \AA}^6)$	6880148.8 ± 302556.8	10629544.1 ± 272798.6
$\kappa_4(\text{cm}^{-1} \text{ \AA}^8)$	-14867506.4 ± 877135.3	-28398452.1 ± 790863.2
$\kappa_5(\text{cm}^{-1} \text{ \AA}^{10})$	14135049.1 ± 1030210.7	26646661.8 ± 928881.8

In Fig. 4.4, we plot the energy-density results for the FSH-[2B], FSH-[2B(3B)], and FSH-[2B+3B] simulation cases. The FSH-[2B] energies increase with density far more strongly than the other categories. The FSH-[2B(3B)] and FSH-[2B+3B] energies are very similar to one another, with the latter only slightly lower than the former. At $0.0617 \text{ \AA}^{-3}$ their difference is 3.6 cm^{-1} (about 1 % error), and even up to $0.1 \text{ \AA}^{-3}$, the difference is only 18.1 cm^{-1} (1.8 %).

The equilibrium density of the EOS is indicated by the minimum of the energy-density curve. In Fig. 4.5, we show the FSH-[2B] and FSH-[2B(3B)] energy curves near the experimental equilibrium density $\rho_0 = 0.0260 \text{ \AA}^{-3}$.⁹ Our comparisons of the FSH-[2B(3B)] and FSH-[2B+3B], as well as later results shown in Fig. 4.7, show that the difference between the perturbative and full inclusion energies rapidly decreases as we go to lower densities. Thus we expect the equilibrium density to be low enough such that the perturbative and full inclusion results have similar minima. The FSH-[2B] curve underpredicts

the minimum, at $\rho = 0.02551 \text{ \AA}^{-3}$, reproducing our findings in Ref. [37]. Meanwhile, the FSH-[2B(3B)] curve overpredicts the minimum, at $\rho = 0.02646 \text{ \AA}^{-3}$. This indicates that, when the three-body interaction is used as the sole many-body correction term, its attractiveness overcorrects the repulsiveness of the pair interaction even as low as the equilibrium density.

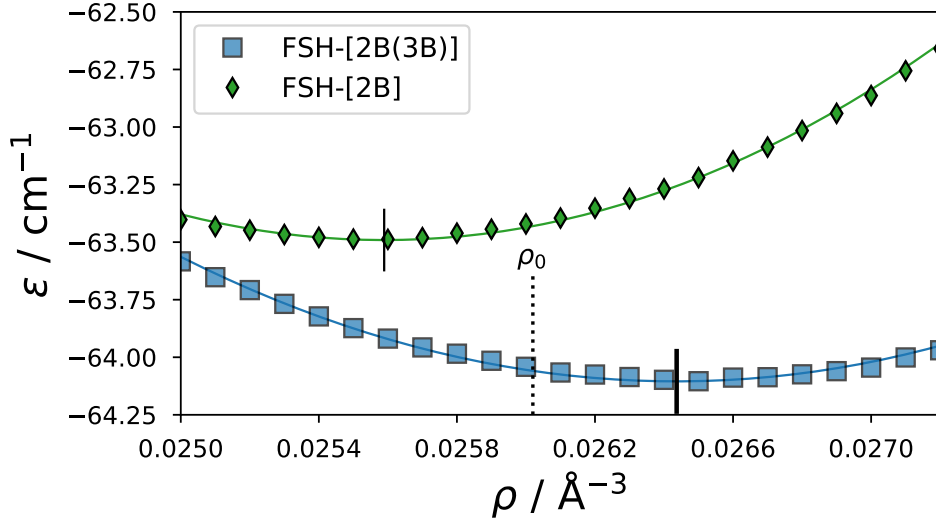


Figure 4.5: The energy per molecule ϵ as a function of density ρ , for simulations performed under the FSH-[2B] conditions (green, diamonds), and the FSH-[2B(3B)] conditions (blue, squares). The minima of the FSH-[2B] and FSH-[2B(3B)] curves are indicated by the thin and thick solid vertical black lines, respectively. The experimental equilibrium density ρ_0 is indicated by the vertical dotted line. The energies in this figure were calculated using the “ τ extrapolation” method to account for the Trotter error.

In Fig. 4.6, we show the individual energies that comprise the FSH-[2B(3B)] total energy curve, i.e. the kinetic, the pair potential, and the three-body potential energies. At low densities, all three types of energies have comparable magnitudes. As the density increases, both the pair and three-body potential energies grow exponentially in magnitude, each at roughly the same rate. The kinetic energy is much smaller in magnitude, and increases linearly. However, the findings of Omiyinka and Boninsegni¹³ suggest that this linear trend only holds for lower densities. It appears clear that the potential energies play the most significant role in the EOS of solid *para*-H₂.

Additionally, we can see how the inclusion or omission of the three-body interaction during sampling affects each of the aforementioned three energies. In Fig. 4.7, we take the kinetic, pair potential, and three-body potential energies calculated using both the full inclusion and perturbative simulations, and plot their respective differences as a function of density between $\rho = 0.06 \text{ \AA}^{-3}$ and $\rho = 0.1 \text{ \AA}^{-3}$. The full inclusion of the three-body interaction causes a shift in energy for all three energy types, and the magnitude of this shift increases with density. These results give some insight as to why the total energy from the perturbative and full inclusion simulations are so similar. The increase in the pair potential energy nearly counteracts the combined decrease in the kinetic and three-body potential energies. Our findings are in line with those of Barnes and Hinde in their simulations of solid helium,¹³⁰ wherein the perturbative and full inclusion treatments of the helium three-body potential of Cencek *et al.*¹³² give very similar total energies.

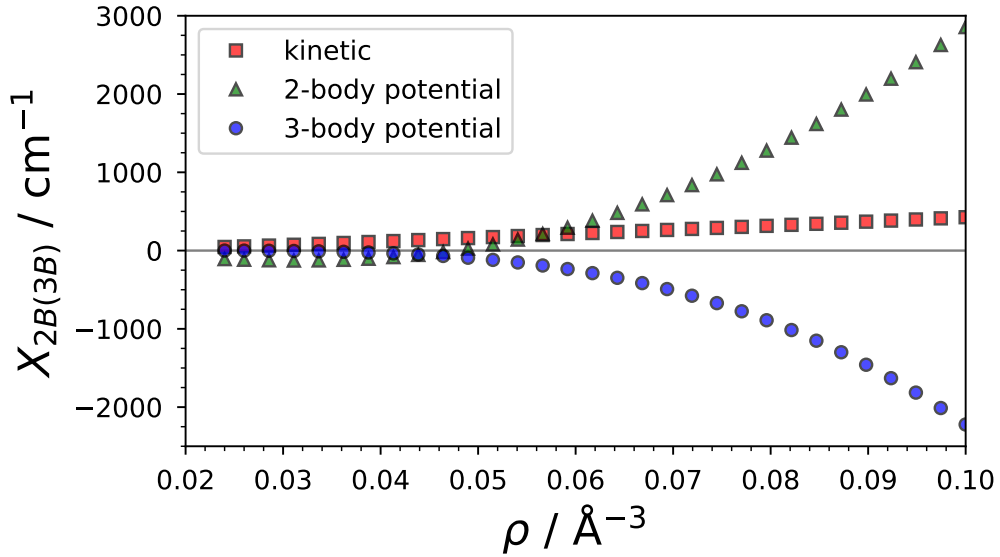


Figure 4.6: The energy per particle as a function of density ρ , separated into its kinetic (red, squares), pair potential (green, triangles), and three-body potential (blue, circles) components. The simulations were carried out under the FSH-[2B(3B)] conditions. The variable X on the vertical axis label is a placeholder for the various energy components.

To a large degree of accuracy, when studying properties that depend only on the total energy, it is enough for us to perform our simulations with only a perturbative treatment of the three-body potential. However, if we need the individual kinetic and potential energies

of the solid, then the three-body potential must be included in the sampling.

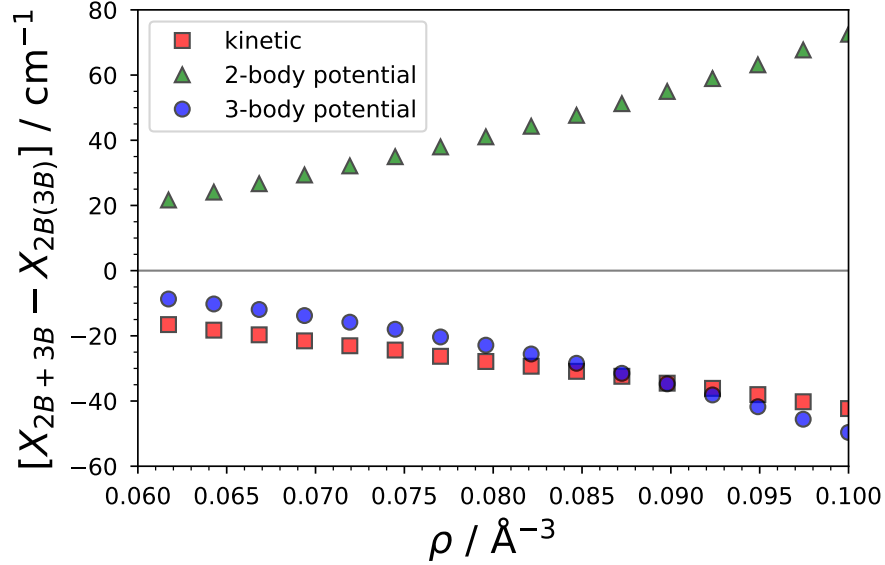


Figure 4.7: The difference in energy components as a function of density ρ , between energies calculated from FSH-[2B+3B] and FSH-[2B(3B)] simulations. The component labels match those in Fig. 4.6. The variable X on the vertical axis label is a placeholder for the various energy components.

4.4.2 Pressure-Density Equation of State

We calculate pressures from the energy curves in Sec. 4.4.1 using

$$P(\rho) = \rho^2 \frac{\partial \epsilon}{\partial \rho} \Big|_T = P_0 + \frac{2}{3} \sum_{n=1}^5 n \tilde{\kappa}_n \rho^{(2n+3)/3} \quad (4.8)$$

In Fig. 4.8, we show the pressure-density curves for the FSH-[2B], FSH-[2B(3B)], and FSH-[2B+3B] cases alongside experimental measurements.

As expected from Ref. [37], the FSH-[2B] curve greatly overshoots experiment, even near the saturation density. Meanwhile, the FSH-[2B(3B)] curve significantly underestimates the experimental data, even at moderately low densities (below $0.04 \text{ \AA}^{-3}$). As the density increases further, the FSH-[2B(3B)] provides an increasingly worsening prediction of the

experiment than the FSH-[2B] curve. Beyond $0.08 \text{ \AA}^{-3}$, the three-body potential even leads to an unphysical decrease in the pressure-density curve. The combination of the pair and three-body interaction potentials is far too attractive at high densities.

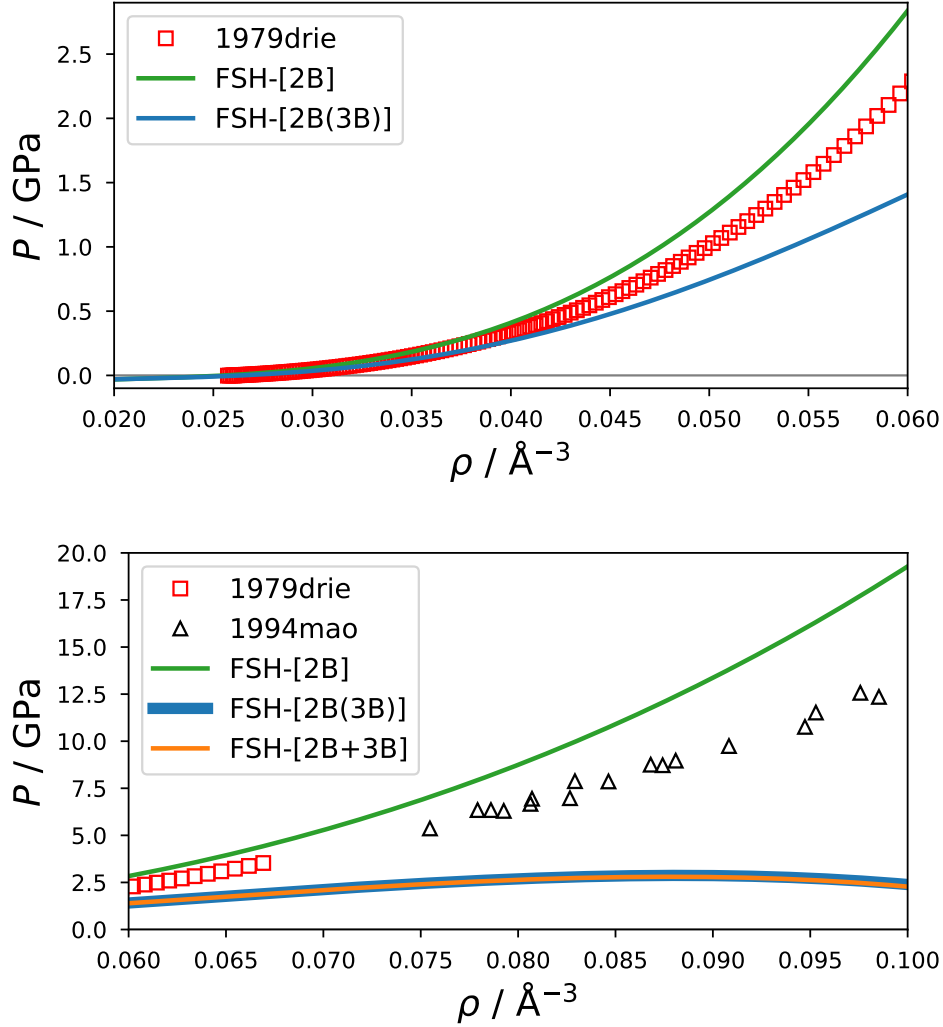


Figure 4.8: The pressure P of solid *para*-H₂ as a function of density ρ . The curves are calculated using Eq. (4.8), and correspond to simulations performed under the FSH-[2B] conditions (green), the FSH-[2B(3B)] conditions (blue), and the FSH-[2B+3B] conditions (orange). Also shown are experimental measurements from Ref. [63] (red squares) and Ref. [148] (black triangles).

Similar results have been seen in prior simulation studies of solid helium. Barnes and Hinde¹³⁰ carried out PIGS simulations of solid helium-4 using the Aziz pair potential¹³¹ and the three-body Cencek potential.¹³² They found that this combination of interactions leads to an underestimation of the pressure when the helium is compressed to sufficiently low molar volumes. Chang and Boninsegni¹⁴⁹ performed PIMC simulations of liquid and solid helium-4 using two separate combinations of pair and three-body interactions.^{150–153} For both combinations of potentials, their simulations underestimated the pressure curve of solid helium-4 at low molar volumes.

We should note that many assumptions in our model of solid *para*-H₂, for example, that it is an *hcp* lattice, and that it is composed of freely rotating molecules, depend on solid *para*-H₂ being in the low-pressure phase I form.¹⁵⁴ At low temperatures, solid *para*-H₂ remains in its phase I form below about 60 GPa,¹⁵⁵ which is much greater than what any of the pressure curves in Fig. 4.8 reach. Thus, we can be confident that in all simulations done for this paper, the solid is in phase I.

4.4.3 Centroid Distribution

Solid *para*-H₂ is a quantum solid, and its molecules have a large zero-point motion about their nominal lattice sites. This behaviour is largely a consequence of the low mass of, and weak intermolecular interactions between, the *para*-H₂ molecules. The PIMC method accounts for this motion by representing the quantum “spreading” of a particle in space using the position distribution of beads in the PIMC ring polymer. We can observe the zero-point motion of the molecules by sampling the distance of the beads from the centroid of their corresponding molecule

$$\hat{c}(r) = \left\langle \sum_{n=1}^N \sum_{i=1}^P \delta(r - |\mathbf{r}_{i,n} - \mathbf{r}_{c,n}|) \right\rangle \quad (4.9)$$

where $\mathbf{r}_{i,n}$ is the i^{th} bead on the n^{th} particle, and $\mathbf{r}_{c,n}$ is the centroid of the n^{th} particle, given by

$$\mathbf{r}_{c,n} = \frac{1}{P} \sum_{i=1}^P \mathbf{r}_{i,n} . \quad (4.10)$$

The above estimator $\hat{c}(r)$ produces a histogram in the shape of a Gaussian multiplied by r^2 . We rescale $\hat{c}(r)$, and normalize it from $r = [0, \infty)$ to get

$$c(r) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp \left[-\frac{1}{2} \frac{r^2}{\sigma^2} \right] . \quad (4.11)$$

Equation. (4.11) allows us to calculate the standard deviation σ , which represents the spread of the position distribution of each *para*-H₂ molecule. We show examples of Gaussian distributions representing this zero-point motion in Fig. 4.9, for two different densities.

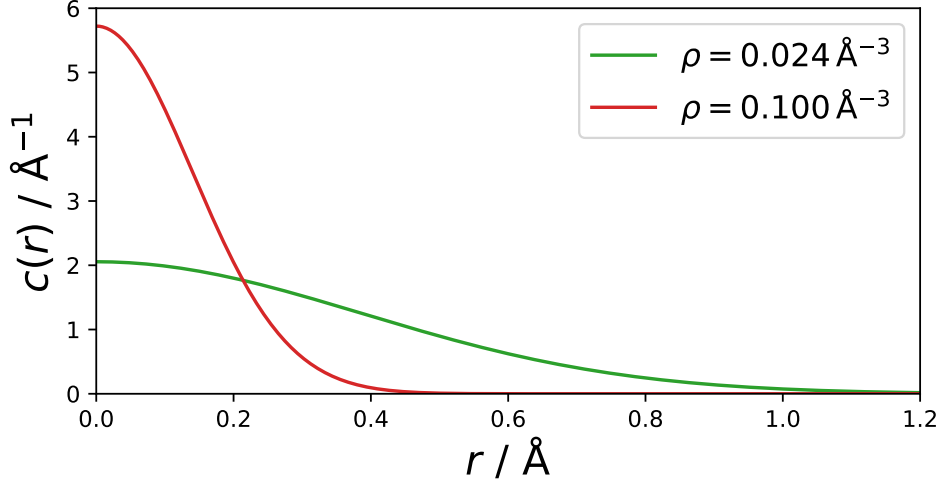


Figure 4.9: Normalized distribution of bead positions from the centroid of a *para*-H₂ molecule in solid *para*-H₂, fit to Eq. (4.11). The results are collected from an FSH-[2B(3B)] simulation at a relatively low density ($\rho = 0.024 \text{ \AA}^{-3}$, green curve) and a relatively high density ($\rho = 0.1 \text{ \AA}^{-3}$, burgandy curve). The distribution is sharper and more classical at the higher density.

In Fig. 4.10, we show the ratio σ/a as a function of density ρ , collected from perturbative and full inclusion treatment simulations. Here, a is the lattice constant. We see that the value of σ decreases with density. As mentioned earlier, one of the contributors to the presence of zero-point motion is the weak intermolecular interaction between *para*-H₂ molecules. As the density increases, the repulsive interactions become stronger, and reduce the spread of the bead positions about their nominal lattice sites. The structure becomes sharper and more classical at higher densities.

Even at the highest densities considered, where the effect of the three-body potential is the strongest, there is negligible difference in the centroid distributions between the two cases. Thus it appears that, despite the fact that the addition of the three-body potential increases the average distance between neighbouring beads (as indicated by the decrease in the kinetic energy seen in Sec. 4.4.1), the overall quantum mechanical “spreading” of

each individual particle remains roughly unchanged.

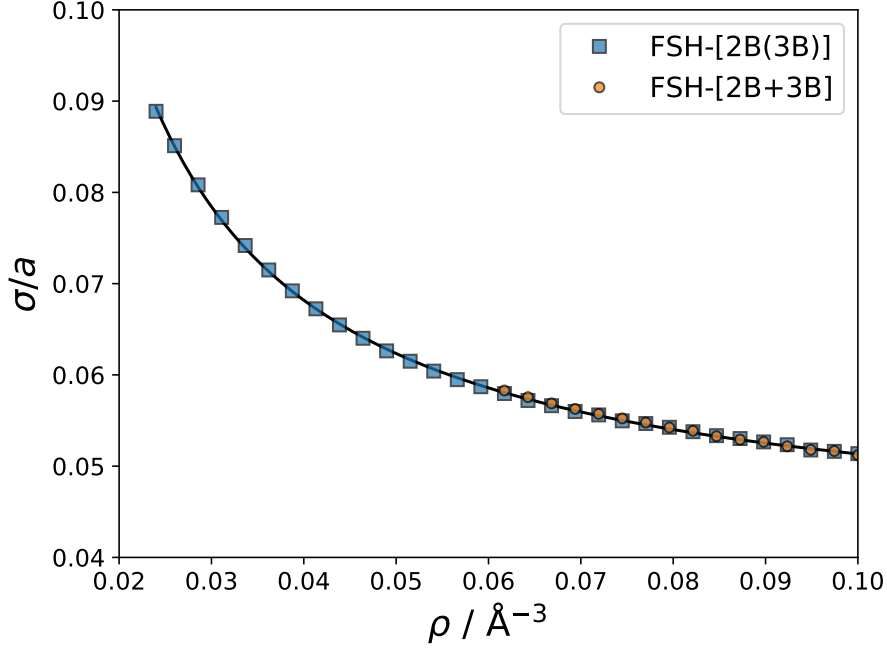


Figure 4.10: The ratio σ/a as a function of density ρ , collected from simulations performed under the FSH-[2B(3B)] conditions (blue, squares), and the FSH-[2B+3B] conditions (orange, circles). Here, σ is the standard deviation of the centroid bead distribution in Eq. (4.11), and a is the lattice constant of the solid at that density. The curve represents the fit to Eq. (4.12). The perturbative treatment of the three-body interaction has no effect on the structure of the solid, and so the FSH-[2B] results would be identical to the FSH-[2B(3B)] results.

We can fit the perturbative treatment data shown in Fig. 4.10 to the fit curve

$$\frac{\sigma}{a} = s_0 + \frac{s_1}{\rho} + \frac{s_2}{\rho^2}, \quad (4.12)$$

where $s_0 = 4.63(2) \times 10^{-2}$, $s_1 = 1.08(2) \times 10^{-3} \text{\AA}^{-3}$, and $s_2 = 5.1(5) \times 10^{-6} \text{\AA}^{-6}$.

4.4.4 Pair Distribution Functions

The inclusion of the three-body potential softens the overall interaction energy at high densities, leading to changes in the structural properties of the lattice. Two quantities we can look at to see this change are the radial pair distribution function $g(r)$, whose (un-normalized) estimator is given by

$$\hat{g}(r) = \left\langle \sum_{n < m}^N \sum_i^P \delta(r - |\mathbf{r}_{i,n} - \mathbf{r}_{i,m}|) \right\rangle \quad (4.13)$$

and the centroid radial pair distribution function $g_c(r_c)$, whose (un-normalized) estimator is

$$\hat{g}_c(r_c) = \left\langle \sum_{n < m}^N \delta(r_c - |\mathbf{r}_{c,n} - \mathbf{r}_{c,m}|) \right\rangle \quad (4.14)$$

where $\mathbf{r}_{c,n}$ is given by Eq. (4.10).

In Fig. 4.11, we show an example of the normalized $g(r)$ curves collected from perturbative and full inclusion treatment simulations, superimposed on each other. Both sets of data come from simulations performed at the density $\rho = 0.1 \text{ \AA}^{-3}$. The two curves are nearly identical in form. There are very small deviations between the two, and these are the most pronounced around the first shell. Define $\Delta g(r)$ to be the radial pair distribution function of the perturbative treatment simulation subtracted from that of the full inclusion treatment simulation. We plot $\Delta g(r)$ around the first shell in the inset of Fig. 4.11. The inclusion of the three-body interaction during sampling softens the pair distribution function, and we can see this especially in the first shell, which becomes noticeably flatter and wider. Note that we show this example for the highest density explored in this work, $\rho = 0.1 \text{ \AA}^{-3}$, because it is the case for which the change in the structure is the most pronounced. Similar behaviours occur at lower densities, but with a smaller change in the distribution.

In Fig. 4.12, we show the corresponding $g_c(r_c)$ curves under the same conditions and context of Fig. 4.11. The centroid radial pair distribution is much more localized than its corresponding radial pair distribution function. As was the case for the $g(r)$ functions, the perturbative and full inclusion treatment simulations produce very similar $g_c(r_c)$ functions. The differences are most easily seen around the first peak. We define $\Delta g_c(r_c)$ similarly to how we defined $g(r)$ earlier, and plot it around the first peak in the inset of Fig. 4.12. Once again, we see that the three-body interaction softens the peak in the first shell of the distribution.

We can use our findings thus far to describe the change in the lattice structure when the three-body interaction is introduced. Recall from Sec. 4.4.3 that the three-body interaction has a negligible effect on the centroid bead distribution, even up to $\rho = 0.1 \text{ \AA}^{-3}$. Also, we see from Figs. 4.11 and 4.12 that both the radial pair distribution and the centroid radial pair distribution soften around the first shell after including the three-body interaction in the sampling. These findings indicate that, at least up to the densities explored in this paper, the three-body interaction does not increase the overall quantum mechanical “spreading” of each individual particle, but it does give each particle more freedom to move around its nominal lattice site.

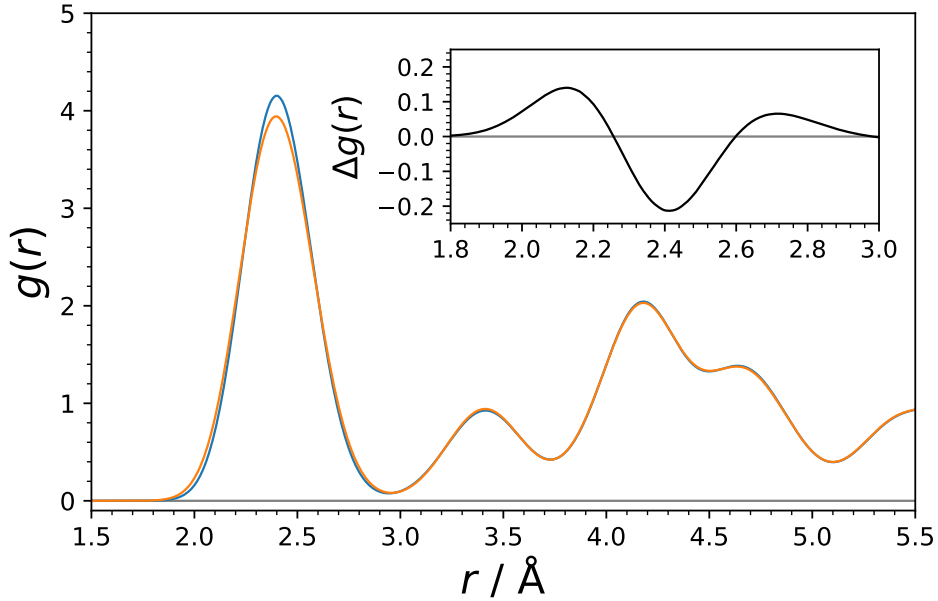


Figure 4.11: The radial pair distribution function $g(r)$ of solid *para*-H₂ at a density $\rho = 0.1 \text{ \AA}^{-3}$, given by Eq. (4.13). The results are collected from simulations performed under the FSH-[2B(3B)] conditions (blue, underlying curve), and the FSH-[2B+3B] conditions (orange, overlaying curve). Readers should focus on regions where the underlying blue curve is visible. In the inset we show $\Delta g(r)$, which is the difference between the orange and blue curves in the region around the first shell of the two $g(r)$ distributions.

We should admit, of course, that the arguments presented here about the effect of the three-body interaction on the solid’s structure come from a model that fails to produce the experimental EOS. However, these structural changes are consequences of softening the

overall interaction energies of the solid at higher densities. Thus, even if we had used a model that produced the correct EOS (*i.e.* one that shifts the FSH-[2B] pressure-density curve down to the experimental data instead of shooting past it) we would likely have seen similar changes in the structure, qualitatively. The radial and centroid radial pair distributions would both still have softened, but the changes would be less pronounced.

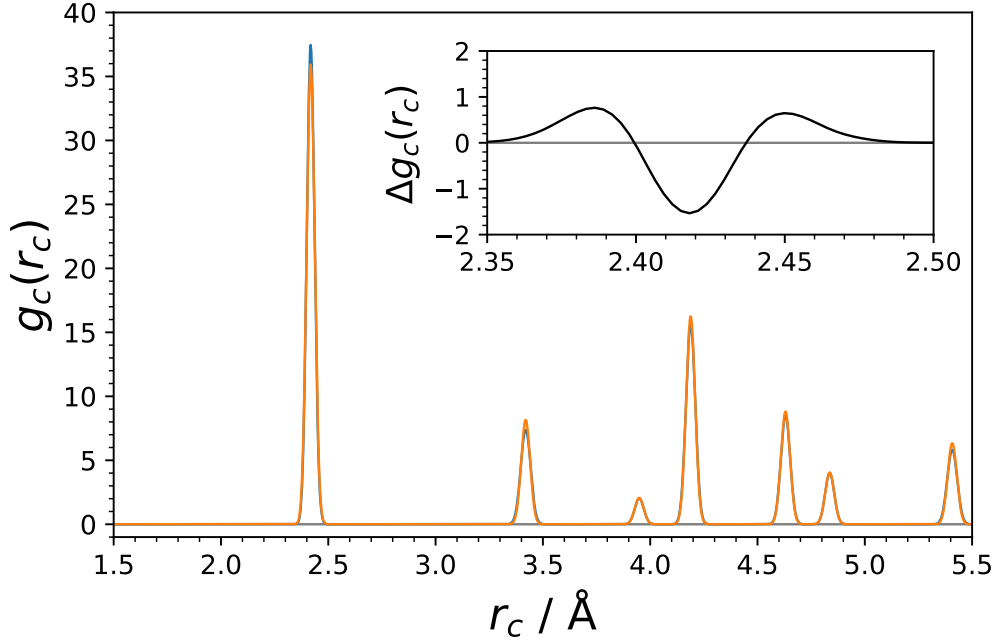


Figure 4.12: The radial centroid pair distribution function $g_c(r_c)$ of solid *para*-H₂ at a density $\rho = 0.1 \text{ \AA}^{-3}$, given by Eq. (4.14). The simulation labels are identical to those in the caption of Fig. 4.11. In the inset we show $\Delta g_c(r_c)$, which is the difference between the orange and blue curves in the region around the first shell of the two $g_c(r_c)$ distributions.

We can explore how the structure of the lattice changes as a function of density. In Fig. 4.13, we show the pair distributions $g(r)$ for three different densities, collected from perturbative treatment simulations. A change in the density has a far greater effect on the radial pair distribution than does the presence of the three-body interaction potential. As the density increases, the first shell becomes narrower and more peaked. Features of the pair distribution function that are hidden at low densities by the quantum delocalization of the *para*-H₂ molecules become visible at higher densities. For example, focus on the regions indicated by the dashed lines in Fig. 4.13. At a density of $\rho = 0.024 \text{ \AA}^{-3}$, this

region appears as a single, uniform shell. As the density increases, the shell develops sharper, more distinguishable features.

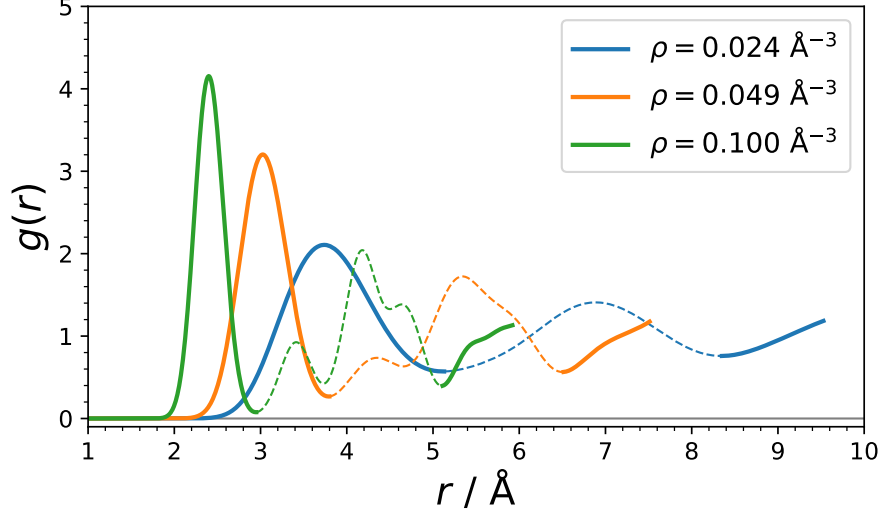


Figure 4.13: The radial pair distribution function $g(r)$ of solid *para*-H₂, collected from FSH-[2B] simulations at densities of $\rho = 0.024 \text{ \AA}^{-3}$ (blue curve), $\rho = 0.049 \text{ \AA}^{-3}$ (orange curve), and $\rho = 0.1 \text{ \AA}^{-3}$ (green curve). The dashed portion of each curve is a visual guide for the reader. It highlights the second shell region of the blue curve, as well as the corresponding “second shell” regions in the orange and green curves, when the change in density is accounted for.

4.5 Conclusions

We used PIMC simulations to study the EOS of solid *para*-H₂ at $T = 4.2 \text{ K}$ using *ab initio* pair and three-body interaction potentials. Periodic boundary conditions were accounted for during the simulations. The finite system size errors were corrected using potential tail corrections, and the Trotter factorization errors were accounted for either by extrapolation or by using a sufficiently small imaginary time step.

We found that the difference between the perturbative and full inclusion treatments of the three-body interaction is fairly small, even at the highest densities explored in this paper. The total energies and pressures in both cases are very similar. The full inclusion

of the three-body interaction in the PIMC sampling protocol causes noticeable changes in each of the kinetic, pair potential, and three-body potential energies. However, when added together, these individual changes largely cancel out, resulting in only a small decrease in the total energy. This behaviour can be explained by the fact that the inclusion of the three-body interaction has only minor effects on the structure of the solid. For example, the centroid bead distribution shows negligible change going from the perturbative to the full inclusion case. The radial and centroid radial pair distributions soften very slightly, with the majority of the change in both cases occurring around the first peak. The success of the perturbative method shows that the three-body interaction, and possibly also higher-order many-body interactions, can often be treated as a sort of density-dependent background energy.

In Sec. 4.4.3, we show how the centroid bead distribution changes as a function of density. It will be of interest to test the deconvolution procedure derived to connect centroid radial distributions to real space ones in the context of liquid *para*-H₂,¹⁵⁶ and see how these concepts can be applied to a quantum solid. The centroid bead distribution is a measure of the quantum delocalization of the molecules, and thus it could be used to help predict *a priori* the coupling matrix for numerical coarse-grained path-integral simulations of solid *para*-H₂.^{157, 158}

The pair potential on its own is too repulsive at high densities to successfully reproduce the EOS. The three-body interaction, when used as the only many-body correction energy term, overcorrects the effects of the repulsive wall of the pair interaction. The resulting pressure-density curve severely underestimates experimentally observed pressures. It is possible that, to perform accurate simulations, we need four-body and higher-order many-body interaction terms. It could also be the case that static interaction potentials on their own are inadequate for simulating solid *para*-H₂ at high densities. A final potential source of inaccuracy is the use of the AHR approximation where *para*-H₂ molecules are treated as point-like particles. One way to remedy this situation would be to explicitly include molecular rotations in our path integral description. Such an extension is possible via the use of established rotational path integral techniques.^{52, 53, 67, 71, 159–162}

Chapter 5

A neural network-based four-body potential energy surface for parahydrogen

This chapter is reproduced from Alexander Ibrahim and Pierre-Nicholas Roy, “A neural network-based four-body potential energy surface for parahydrogen” *Journal of Chemical Physics* **160**, 244308 (2024), with the permission of AIP Publishing.¹⁶³

5.1 Introduction

The topic of many-body intermolecular interactions has seen a rise in interest over the past few decades.^{14,15,164–167} Hydrogen, as the simplest molecular species, is an excellent subject for the study of many-body interactions. One of its nuclear spin isomers, parahydrogen (*para*-H₂) displays a variety of exotic quantum behaviors in many-body systems. For example, solid *para*-H₂ is a quantum molecular solid.^{1,2,4,6,7} The individual *para*-H₂ molecules that comprise it have relatively large zero-point energies.^{6–9} This gives each molecule very large amplitude vibrations about its nominal lattice site, which “inflates” the lattice and gives the solid a much larger molar volume than one would expect from a classical description.^{9,38}

It has become increasingly apparent that condensed many-body systems of *para*-H₂ molecules, especially at high densities, cannot be adequately modelled using only pairwise additive interactions.^{11,12,34,104} This appears strange at first, because three-body and

higher-order many-body interactions are strong only at very short ranges.⁸¹ However, these many-body interactions contribute significantly at much lower densities than one would expect classically, mainly because each *para*-H₂ molecule has such a delocalized probability distribution.

Introducing many-body interactions into these models is a challenging task. Pair interactions can often be reduced to a one-dimensional function of distance. However, interactions between three or more particles require several parameters to properly describe, and are necessarily multi-dimensional.¹⁴ Recent developments in multiple research fronts have made it possible to create potential energy surfaces (PES) for many-body systems. Advances in computational chemistry together with increased computing power make it possible to perform *ab initio* electronic structure calculations for large collections of particles. Also, the widespread availability of machine learning software makes it much easier to create higher dimensional PESs for many-body interactions.^{168,169}

Consider a collection of N *para*-H₂ molecules. Let the coordinate $\mathbf{R}_i$ describe the position and angular orientation of the i^{th} particle. The total interaction potential energy of the system V_{tot} can be expanded as a sum of many-body potentials using

$$V_{\text{tot}} = \sum_{i < j}^N V_2(\mathbf{R}_i, \mathbf{R}_j) + \sum_{i < j < k}^N V_3(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k) + \sum_{i < j < k < l}^N V_4(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k, \mathbf{R}_l) + \dots \quad (5.1)$$

where V_2 is the two-body interaction potential, V_3 is the three-body interaction potential, and so on.

Most research on interaction potentials between hydrogen molecules focuses on the pair potential. With hydrogen being the simplest molecule, there is no shortage of examples of two-body interaction potentials.^{16,25,62,66,97–102} Pair potentials for hydrogen molecules have several common features. They have a repulsive wall at short intermolecular distances, a somewhat deep attractive well (typically centred around 3.0 Å), and a weakly decaying attractive tail for the dispersion interaction. A common issue with several H₂ and *para*-H₂ pair potentials is that the repulsive wall is too strong. This leads to a severe overestimation of the pressure when modelling condensed systems of *para*-H₂ at high densities.^{13,42} At short intermolecular separations, pair potentials on their own become unreliable, and we require many-body interactions to accurately describe collections of *para*-H₂ molecules.^{11,12,104}

It is well known that the next order interaction, the three-body interaction potential for *para*-H₂, is overall attractive at short intermolecular distances.^{11,12,81,104} One common three-body analytic potential is the Axilrod-Teller-Muto (ATM) potential.^{17,18} It accurately models the dispersion relation at large distances, but incorrectly predicts the short-

range interaction to be net repulsive. In 2010, Manzhos *et al.* published a machine-learning based three-body PES for *para*-H₂.¹² Recently, our group published an isotropic 3D *ab initio* three-body PES for *para*-H₂.⁸¹ Its energies are calculated at the correlated coupled-cluster theory level, with single, double, and perturbative triple excitations (CCSD(T)). The calculations use an atom-centred augmented correlation-consistent triple zeta (AVTZ) basis set, supplemented by a (3s3p2d) midbond function.¹¹¹ We use the reproducing-kernel Hilbert Space method to fit the energies.^{36, 113, 139, 142}

Recently, we performed simulations of solid *para*-H₂ using the two-body Faruk-Schmidt-Hinde (FSH) potential¹⁶ and our recent three-body PES.¹²³ Due to its repulsive character at short distances, simulations that use only the FSH potential overestimate the pressure-density equation of state (EOS) at high densities.³⁷ Because the three-body PES is net attractive, its inclusion alongside the pair PES decreases the pressure. At low densities, this change slightly improves the agreement of the simulations with experiment. However, at high densities, the three-body PES overcorrects the repulsive effects of the pair PES, and greatly underestimates the pressure.

Wheatley *et al.* recently published an *ab initio* four-body PES for helium.¹⁷⁰ The energies were calculated at the CCSD(T) level, using an AVQZ atom-centred basis set. The interpolation for the PES is based on Gaussian processes.¹⁷¹ They performed path-integral Monte Carlo (PIMC) simulations using this potential to calculate the fourth virial coefficient of helium to reasonable agreement with experiment over a wide range of temperatures. The four-body PES for helium shares many qualitative features with the four-body PES for *para*-H₂ presented in this paper.

In this paper, we present a six-dimensional, isotropic four-body PES for *para*-H₂. The energies are calculated at the CCSD(T) level of theory, using an AVDZ atom-centred basis set, with an additional (3s3p2d) midbond function at the centre of mass of each calculation.¹¹¹ We use a rigid rotor approximation and average out the rotational degrees of freedom using the 6-point Lebedev quadrature³² to make the interactions isotropic. We model the PES using a multi-layer perceptron, trained using PyTorch.¹⁶⁸ At large intermolecular separations, we use an analytic four-body interaction potential to more accurately describe the dispersion interaction.^{19, 172} At short range, we fit the energies to an exponential decay, with some modifications for numerical stability.

The remainder of this paper is structured as follows: Sec. (5.2) goes over the creation of the PES, including the calculation of the interaction energies and how the degrees of freedom are managed. It also covers how the training data is generated, the transformations on the inputs and outputs that improve the effectiveness of the training, and the training process itself. Sec. (5.3) is a discussion and analysis of the PES. It includes an error analysis

of the electronic structure calculations and the methods used to manage the degrees of freedom. It also covers the short-range and long-range extrapolations of the PES, and ends with an analysis of the effects of the four-body interaction on a classical approximation of solid *para*-H₂.

5.2 Creating the Potential Energy Surface

5.2.1 Electronic Structure Calculations

We use the MRCC software program (version 2019)¹¹⁹ to carry out the electronic structure calculations for the (H₂)₄ energies. The calculations are performed at the CCSD(T) level of theory.¹²⁰ There is an AVDZ basis set centred at each of the eight hydrogen atoms. We supplement the basis with (3s3p2d) midbond functions¹¹¹ located at the centre of mass of the collection of atoms, as was done for the pair potential by Hinde²⁵ and this group’s recent three-body PES.⁸¹ An analysis of different basis sets and electronic structure methods is given in Sec. (5.3.2).

There are many settings available in the MRCC program that determine the precision of the energy calculations. All calculations provided in this paper were carried out using the default settings, except for `cctol`, which is set to 9, and `scftol` and `scfdtol`, which are both set to 7. Under these settings, the final iterations in the SCF and CCSD calculations change on the order of 10⁻¹¹ Hartrees and 10⁻¹⁰ Hartrees, respectively. Nearly all SCF and CCSD energy calculations converged within 9 and 20 iterations, respectively. After the conversion to wavenumbers and the sum of terms needed to find the interaction energy, we estimate a conservative upper bound of 10⁻⁴ cm⁻¹ on the convergence error of the interaction energies. As a test, we select a handful of calculations for the tetrahedron geometry (as presented in Sec. (5.3.2)), with side lengths ranging from 2.2 Å to 4.5 Å. We perform them once more, but with the value of `cctol` set to 10 instead of 9. Of the differences between the energies in our samples when `cctol=10` and `cctol=9`, the mean absolute difference was 7.2 × 10⁻⁶ cm⁻¹, whereas the maximum absolute difference was 2.7 × 10⁻⁵ cm⁻¹.

Let $\mathbf{R}_i = (\mathbf{r}_i, \rho_i, \mathbf{\Omega}_i)$ describe the centre-of-mass position, bond length, and space-fixed angular orientation of the hydrogen molecule with the label i , respectively. Define $E_{ijkl} = E_{ijkl}(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k, \mathbf{R}_l)$ to be the CCSD(T) energy with all molecules i , j , k , and l present. Define E_{ijk} to be the CCSD(T) energy with only molecules i , j , and k present, but still in the full four-body basis set. In other words, despite the nuclei and electrons of

molecule l being missing, the atom-centred basis of molecule l is still present. We define E_{ij} and E_i similarly, in that a missing index indicates that the molecule is missing but its basis set is still present. The counterpoise-corrected²⁸ interaction energy of four hydrogen molecules (with $(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k, \mathbf{R}_l)$ omitted for brevity) is given by

$$\begin{aligned}\Delta E = & E_{1234} \\ & - (E_{123} + E_{124} + E_{134} + E_{234}) \\ & + (E_{12} + E_{13} + E_{14} + E_{23} + E_{24} + E_{34}) \\ & - (E_1 + E_2 + E_3 + E_4).\end{aligned}\tag{5.2}$$

Each interaction energy requires 15 separate electronic structure calculations to produce.

5.2.2 Managing Degrees of Freedom

By modelling the molecules as a collection of four rotors in free space, we can describe the system as having 4 degrees of freedom from the bond lengths, 8 degrees of freedom from the angular orientations of the rotors, and 6 degrees of freedom from the distances between their centres of mass, for a total of 18 degrees of freedom.

We can use the rigid rotor approximation to fix the bond length of each molecule to the vibrationally averaged ground state length of $1.449 a_0$,¹¹ removing four degrees of freedom.

The eight degrees of freedom associated with the angular orientations are removed by performing a 6-point Lebedev quadrature.^{32,33,118} To perform this averaging, we first fix in space the centres of mass of each of the four hydrogen molecules. Next we align each rigid H_2 molecule along either the space-fixed x -, y -, or z -axis, giving each molecule 3 possible angular orientations. This alignment is done independently for each of the four hydrogen molecules, for a total of 81 different combinations of space-fixed angular orientations. Let n label each of the 81 different combinations. Performing the energy calculation described in Eq. (5.2) for the n^{th} combination gives the interaction energy $\Delta E^{(n)}$. The isotropic four-body interaction energy is the average of these 81 interaction energies.

$$V_4 = \frac{1}{81} \sum_{n=1}^{81} \Delta E^{(n)}(\{\mathbf{R}_i\}).\tag{5.3}$$

where $\{\mathbf{R}_i\} = (\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \mathbf{R}_4)$. This averaging process projects out the anisotropic components of the interaction energy. It is an approximation to performing the full integration over all the degrees of freedom. There are higher levels of Lebedev quadratures that

give more accurate averages, but require many more energy calculations to perform the average.³² In an analysis given in Sec. (5.3.1), we show that the next highest Lebedev quadrature give a negligible improvement in the accuracy.

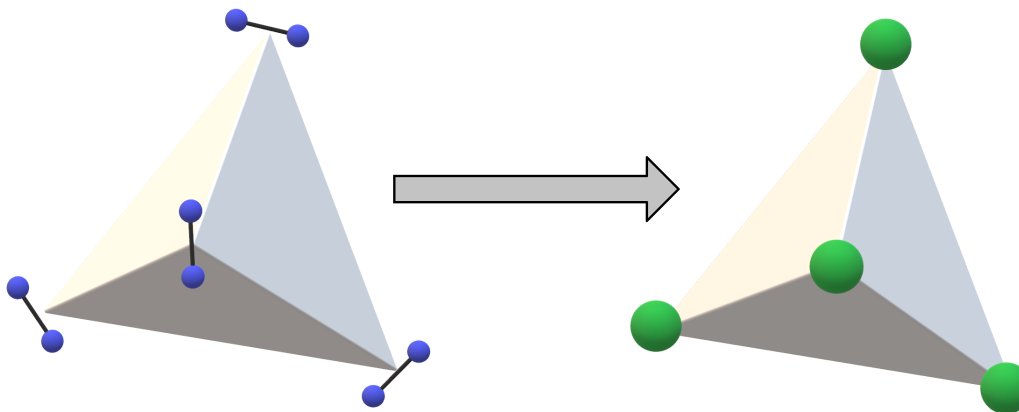


Figure 5.1: A graphical representation of how fixing the bond length of each hydrogen molecule, and averaging over the angular orientations of the system, means we now model the four hydrogen molecules as four points in space.

After removing the bond length and angular orientation degrees of freedom, we have reduced the description of our system from that of four interacting rotors to that of four interacting points in space (see Fig. 5.1). We are now left with only 6 degrees of freedom, corresponding to the relative distances between the centres of mass of each pair of hydrogen molecules. We can thus describe each four-body geometry with the six relative side lengths $\{r_{ij}\} = (r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34})$ where r_{ij} is the distance between the centres of mass of molecules i and j . For the remainder of the paper, when we refer to the “distance between two molecules,” we mean specifically the distance between each molecule’s centre of mass. Also, when we refer to the “interaction energy of a four-body geometry,” we refer to the isotropic interaction energy between four *para*-H₂ molecules, where the centres of mass of the molecules lie at the corners of the aforementioned four-body geometry.

Notice that each interaction energy is a linear combination of 15 separate CCSD(T) energies, each requiring its own electronic structure calculation. Moreover, each 6-point Lebedev quadrature is a linear combination of 81 separate interaction energies. This means that each averaged energy value requires a total of 1215 individual CCSD(T) energy values.

5.2.3 Sample Generation

To train the neural network and create the PES, we need to generate samples for training, testing, and validation. The input is the set of six side lengths $(r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34})$ and the outputs are the corresponding four-body interaction energy as given by Eq. (5.3). This PES is meant to be used in numerical simulations, including at high densities, and thus interaction energies at short distances must be well represented in the training data. At longer intermolecular separations, the *ab initio* CCSD(T) energies converge to the four-body dispersion relation (see Sec. (5.3.2)).^{19,172} Thus we do not have to sample energies from geometries with excessively large intermolecular separations, even for purposes where the weak long-range energies are important, such as simulations involving the gas phase.¹⁰⁹

The procedure for generating samples is as follows. First, we decide on an exponential decay sampling function

$$p(r) = Q \exp\{-Cr\}, \quad (5.4)$$

where C is a positive constant and Q is a constant chosen to normalize the distribution between $r_{\min}$ and $r_{\max}$ (mentioned later). To create an input data sample, we first sample 6 side lengths *i.i.d.* from $p(r)$ between $r_{\min}$ and $r_{\max}$, and label them $(r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34})$. We next attempt to construct a four-body geometry from the six side lengths via trilateration.¹⁷³ However, not every ordered set of 6 side lengths can be turned into a valid 3D shape. If the 6 side lengths form a valid four-body geometry, we accept the sample. If not, we generate an entirely new ordered set of 6 side lengths, and repeat the process of attempting to construct a valid 3D shape from the side lengths until it succeeds. In the supplementary material, we provide an example of the function we use to generate the four points from the six side lengths.

There are several reasons for settling upon this sampling strategy. First, it allows us to generate samples with different, uncorrelated side lengths, which is important because we do not know *a priori* the probability distribution of the *para*-H₂ molecules in the situation the PES is used in. Second, an exponential decay function allows us to generate samples to have shorter side lengths with a greater frequency than longer side lengths. This is desirable, because samples with excessively long pair distances have energies that are either negligibly weak or can be replaced with the analytic four-body dispersion interaction described in Sec. (5.3.4). By choosing the value of C , we can determine the distribution of side lengths, and therefore roughly determine the energies. For example, a large value of C causes Eq. (5.4) to generate smaller side lengths, which correspond to samples with larger energies. Looking at Fig. 5.2, it appears that we were successful in generating samples with a large range of energies, with a slight bias towards high-energy samples.

With $C_0 = 0.9209 \text{ \AA}^{-1}$, we sample 9000 points between $2.2 \text{ \AA}$ and $4.5 \text{ \AA}$ using $C = C_0$ (for a wide range of side lengths and energies), 1500 points between $2.2 \text{ \AA}$ and $4.5 \text{ \AA}$ using $C = 3C_0$ (for samples with short side lengths and large energies), 1500 points between $2.2 \text{ \AA}$ and $4.5 \text{ \AA}$ using $C = 6C_0$ (for samples with very short side lengths and very large energies), and 4000 points between $2.8 \text{ \AA}$ and $4.5 \text{ \AA}$ using $C = C_0/6$ (for samples with long side lengths and weak energies).

We supplement the training data using samples that are based specifically on the geometries of the *hcp* lattice. To generate these samples, we first select a reference molecule in the lattice. Then we consider all four-body geometries involving this reference molecule that satisfy two conditions; they have at least one side length equal to the lattice constant, and they have no side length that is more than twice the lattice constant. Geometries that disobey these conditions have a nearly negligible contribution to the total interaction energy, even at a lattice constant as low as $2.2 \text{ \AA}$. There are 83 different four-body shapes that satisfy these conditions. For each shape, we calculate the interaction energy for 47 different lattice constants, evenly spaced between $2.2 \text{ \AA}$ and $4.5 \text{ \AA}$, for a total of 3901 additional samples. A rescaling strategy for training described in Sec. (5.2.5) requires us to discard many of these *hcp* lattice-based samples, resulting in only 1610 of them being included in the training data.

The aforementioned sampling strategy for geometries has certain limitations. The range of side lengths is relatively constrained to limit the sampling of geometries with especially weak energies. This means that, within any given sample, the largest side length is at most about twice the length of any other side length. This leads to relatively “compact” geometries. Certain types of geometries, such as planar or linear geometries, are poorly represented, and if they appear at all, it is due to their presence in the *hcp* lattice-based samples.

5.2.4 Feature Transformations

For our neural network, the input data is the ordered set of six side lengths,

$$(r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}), \quad (5.5)$$

and the output is the corresponding interaction energy,

$$V_4 = V_4(r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}). \quad (5.6)$$

To simplify training and improve the performance of our model, we apply three transformations to the six side lengths. First, we take the reciprocals of each of the six side

lengths. Shorter pair distances generally correspond to greater energies, and taking the reciprocal of the distance turns smaller distances into larger input features. Next we linearly map all the features onto $(0, 1]$. We then perform a permutation of the six features in a way that preserves the symmetry of the four-body system.

Because the four *para*-H₂ molecules are identical, our model must be invariant to the permutation of the molecule indices. However, swapping indices changes the order of the six side lengths. If, for example, we swap particle indices 1 and 2 for the six side lengths in Eq. (5.5), the input becomes $(r_{12}, r_{23}, r_{24}, r_{13}, r_{14}, r_{34})$. The neural network thus interprets the permuted input as a completely different sample, and produces a different output energy. On a neural network trained on data without accounting for this permutation symmetry, an index swap can even change the output energy by tens of wavenumbers at short ranges.

Not all permutations are possible. Of the $6! = 720$ permutations of these elements, only 24 of them are allowed. For example, there is no way to swap the indices such that we only interchange the first two side lengths r_{12} and r_{13} , while leaving the other four side lengths unchanged. In particular, we cannot simply sort the six side lengths as a way of solving this permutation issue. As seen later in Table 5.2, it is possible to have two separate four-body geometries with the same six pair distances in different, permutationally-incompatible orders. In such a case, naively sorting the pair distances would make the two samples indistinguishable. In addition, not every six-tuple input of pair distances corresponds to a valid four-body geometry, and a naive sort could generate such six-tuples.

We decide to account for the permutation symmetry at the input transformation level by selecting the “smallest” permutation, as defined by lexicographic ordering (see the supplementary material). For readers familiar with the Python programming language, this is the kind of ordering that Python uses to compare two tuples of equal size of floating-point numbers. We have a slower, naive implementation, and a much faster lookup-based implementation.

In the slower naive implementation, we loop over all 24 allowed permutations of the input (after the first two transformations), and select the one with the lowest lexicographic order. However, under certain conditions, we can come up with a much faster implementation. In the cases where the shortest and second-shortest pair distances are unique, the permutation with the lowest lexicographic order can be uniquely determined using a lookup table. In particular, in a molecular simulation the six pair distances are floating point numbers calculated from four points in space that have been perturbed from their original positions. Thus, in this case, the six pair distances are almost always unique, and this much faster solution becomes viable.

Another considered solution was to use the permutations for data augmentation. However, as seen in Sec. (5.2.6), we propose the possibility of using small models during the simulation for improved runtime performance. When we make the model permutation invariant, the increased error from using a smaller model is easily manageable. However, when we train using permutations for data augmentation, the error from smaller models increases considerably, because there are too few parameters to fit the permutations.

5.2.5 Rescaling the Energies

The energies in the overall data set span about five orders of magnitude. Our model must be able to predict both short-range, strong-energy samples and long-range, weak-energy samples with high accuracy. Despite being much smaller in magnitude, at lower densities, the probability distribution of solid *para*-H₂ is predominantly in the long-range, weak-energy region of coordinate space. The mean-squared error (MSE) loss function does not prioritize the prediction of weak-energy samples during training. Other loss functions that might work well with outputs over a large range of values, such as the relative squared error, or the mean square log error, were considered, but provided unsatisfactory results.

Instead, we can take advantage of the analytically predicted behavior of the interaction potential to transform the energies in a way that makes training more effective. We create a rescaling function for the energies. This rescaling function is the sum of an exponential decay and an r^{-12} decay function

$$\phi(r) = A \exp\{-Br\} + Cr^{-12} \quad (5.7)$$

where $A = 3.1803 \times 10^6 \text{ cm}^{-1}$, $B = 4.623057 \text{ \AA}^{-1}$, $C = 4220.011 \text{ \AA}^{12} \text{ cm}^{-1}$, and r is the mean of the six side lengths of a given sample. The choice of using an exponential decay and a r^{-12} power decay was to make the rescaling function somewhat resemble the features of the true underlying *ab initio* potential (see Sec. (5.3.4)). This makes it easier to rescale the energies into a similar range. Importantly, the rescaling function is positive for the entire coordinate space of the samples. This means we can divide the *ab initio* energies by this rescaling function to recover a rescaled energy. The neural network is trained using these rescaled energy values as outputs. To recover the predicted energy from the output of a neural network, we multiply the output of the neural network by the rescaling function Eq. (5.7).

Certain weak-energy samples with large intermolecular separations from the *hcp* data set end up becoming very large when rescaled, due to the r^{-12} term in the denominator.

These samples have large enough side lengths that their energies will be completely overwritten by the Bade potential in the final PES anyways (see Sec. (5.3.2)). By filtering out all samples where the mean side length is greater than $4.5 \text{ \AA}$, we can remove the outsized influence of these samples on the loss function, and improve our rescaling strategy. This decreases the number of training samples from the *hcp* data set from 3901 to 1610.

The rescaling operation decreases the dynamic range of the outputs. To show this, we first take the output energies (after the aforementioned filtering), and divide them all by the sample with the smallest absolute energy. We then take the logarithm of the absolute value of the results. The same operations are performed on the rescaled energies. Normalized histograms of the results are shown in Fig. 5.2. Of the original energies, about 95% of the samples lay within a range of five orders of magnitude. After the rescaling, about 94% of the samples lay within a range of only two orders of magnitude. Reducing the dynamic range of the output values greatly improves the model’s prediction of the weak energy samples, which are important at lower densities.

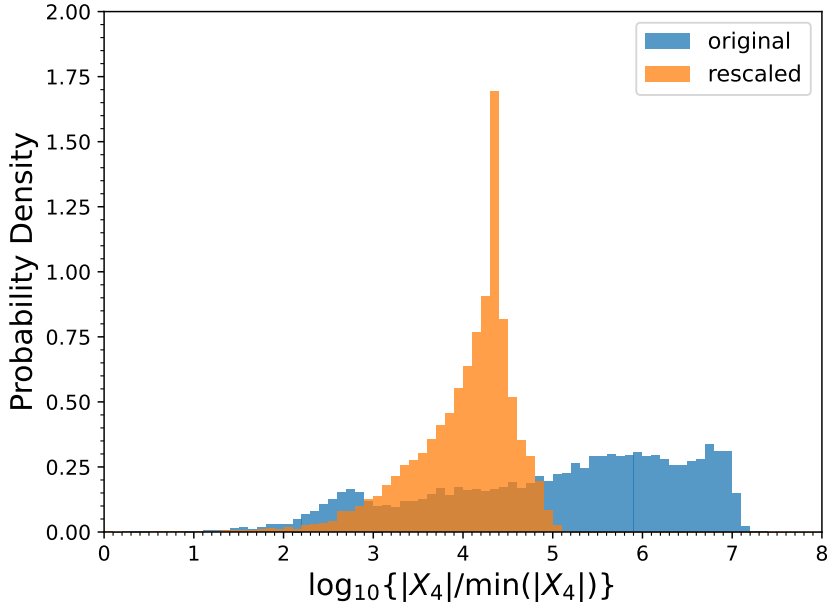


Figure 5.2: Normalized histograms of the sample outputs, after dividing them by the smallest value, then taking the absolute value followed by the logarithm. The variable X_4 is a stand-in for either the four-body interaction potential energy V_4 (blue) or the rescaled energy V_4/ϕ (orange), where ϕ is given by Eq. (5.7) and is evaluated at the appropriate coordinates.

5.2.6 Neural Network and Training

We train five neural networks in total. Each is a simple multilayer perceptron, created, trained, and evaluated using PyTorch.¹⁶⁸ One of the intended uses for this PES is for Monte Carlo simulations of large collections of *para*-H₂ molecules, where the evaluation of the four-body potential likely dominates the simulation’s runtime. In the recently published paper by Wheatley *et al.* on the four-body PES for helium,¹⁷⁰ they noted that their four-body potential requires much more computing time than the two-body or three-body potentials, which limited the temperatures for which they could perform their PIMC simulations. The model’s runtime performance is thus an important factor, and we make reasonable trade-offs concerning other aspects of the network to accommodate it. For example, in addition to large networks, we also create smaller networks that are faster to evaluate, even though they may have larger testing errors.

We create four multilayer perceptrons of different sizes. Each has four hidden layers, alongside the input layer of size 6 and the output layer of size 1. We place ReLU activation functions¹⁷⁴ between the hidden layers. The four networks are the 64-128-128-64 model (where the first hidden layer has a size of 64, the second has a size of 128, and so on), the 32-64-64-32 model, the 16-32-32-16 model, and the 8-16-16-8 model. A depiction of the 8-16-16-8 model is given in Fig. 5.3.

We also train a version of the largest network using shifted softplus¹⁷⁵ activation functions between the layers, which we call the 64-128-128-64-SSP model. Unlike the networks trained with ReLU activation functions, this model is continuously differentiable, for situations where forces are required.

The 16000 *ab initio* samples from the probability distributions are split randomly into training, validation, and testing sets using a 6 : 1 : 1 split, giving 12000 training samples, 2000 validation samples, and 2000 testing samples. The training data is further augmented with the remaining 1610 *ab initio* samples from the *hcp* lattice geometries, *i.e.* those remaining after filtering out all samples with a mean side length greater than 4.5 Å. The input features of each sample are passed through the transformations described in Sec. (5.2.4), and the output energies are rescaled using the transformation described in Sec. (5.2.5).

For training, we use a batch size of 64, a mean squared error loss function, and an Adam optimizer with a learning rate of 2×10^{-4} , supplemented with a learning rate scheduler that decreases the learning rate by 1% every 25 epochs past the first 100 epochs. The 64-128-128-64 and 64-128-128-64-SSP models were trained for 20000 epochs, and the other three were trained for 10000 epochs. We find that omitting weight decay entirely improves

the model’s accuracy, most likely due to the low-noise nature of the electronic structure energies. Both batch normalization and dropout were found to have a negligible effect on training, and so were omitted for the final model.

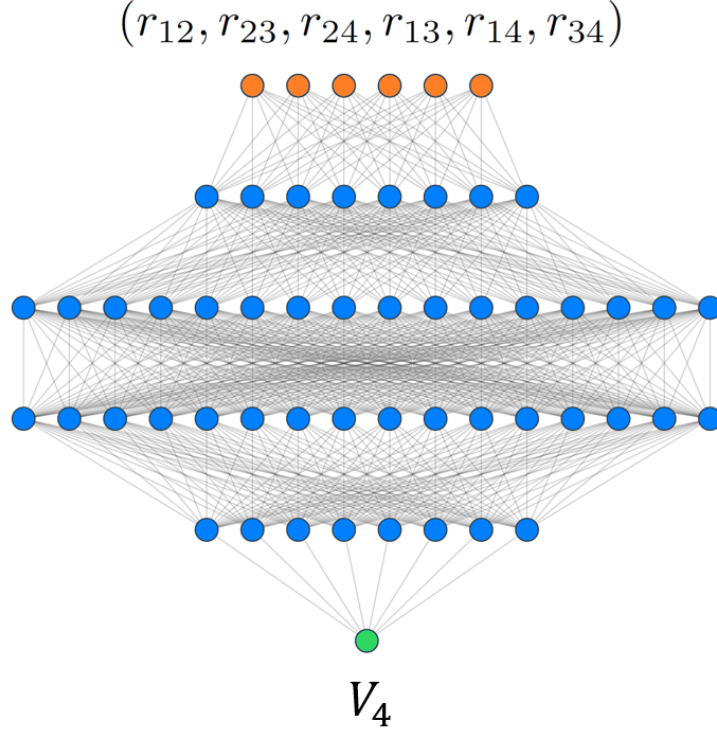


Figure 5.3: A depiction of the 8-16-16-8 model. The six orange circles represent the input features. The blue circles represent the fully-connected linear hidden layers. The green circle represents the output. There are a total of 4 ReLU layers, one after each linear layer. The results of the calculations propagate downwards in the figure. Not shown are the feature transformations performed on the six side lengths, or the rescaling transformation performed on the output. Diagram made with the assistance of Ref. [176].

In Fig. 5.4, we show the loss for the training of the 64-128-128-64 and 64-128-128-64-SSP models as a function of the epoch. Shown are the results for the training data set, the validation data set, and the training data set with the 1610 samples from the *hcp* lattice removed. In the supplementary material, we show similar figures for the smaller models.

For both models shown in Fig. 5.4, the MSE loss decreases rapidly for all data sets. The large amount of noise in the curves for the earlier epochs can be attributed to the fact

that we are viewing the logarithm of the MSE loss. As the scheduler decreases the learning rate, all error curves gradually settle. Both models reach a similar training error, but the use of the shifted softplus activation function gives a lower validation error.

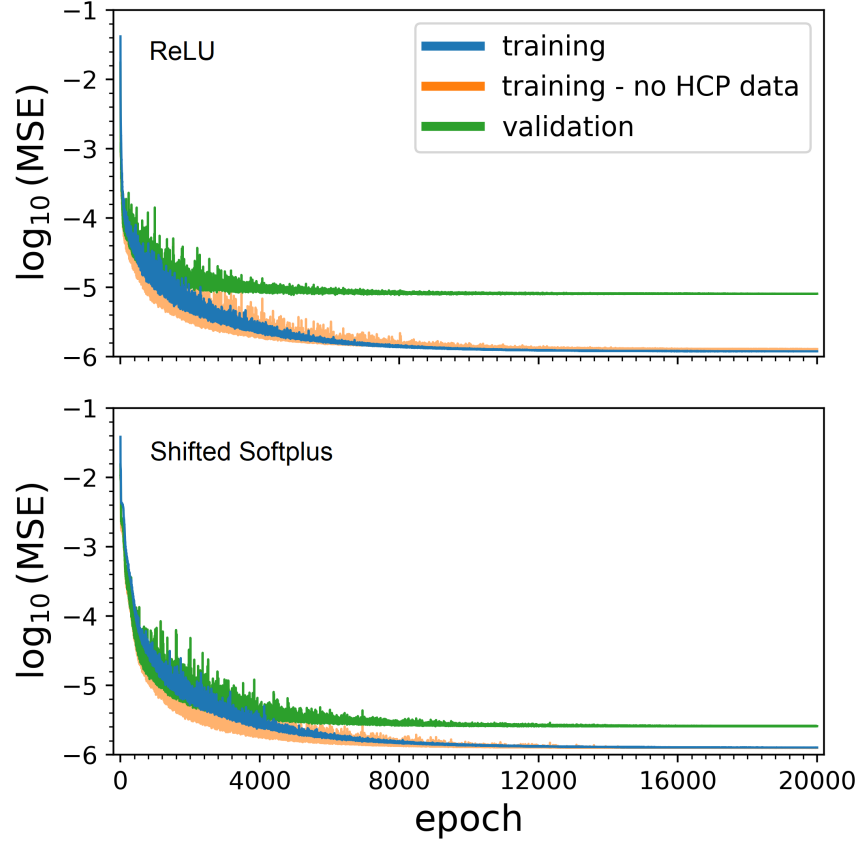


Figure 5.4: The base-10 logarithm of the mean squared error loss as a function of the epoch during the training of the 64-128-128-64- model (top), and the 64-128-128-64-SSP model (bottom). The former uses the ReLU activation function, while the latter uses a shifted softplus activation function. The plots are shown for the training data set (blue, solid), the validation data set (green, solid), and the training data set with the 1610 samples from the *hcp* lattice removed (orange, semi-transparent).

The models have a more difficult time fitting the *hcp* lattice samples. In both the 64-128-128-64 and 64-128-128-64-SSP models, the error of the training data exceeds the error of the training data with the *hcp* lattice samples removed, at earlier epochs. The

smaller models have even more trouble fitting the *hcp* samples. In the 16-32-32-16 model, the training and validation errors are nearly the same, and in the 8-16-16-8 mode, the training error is *higher* than both the validation error and the error of the training samples with the *hcp*-based samples removed.

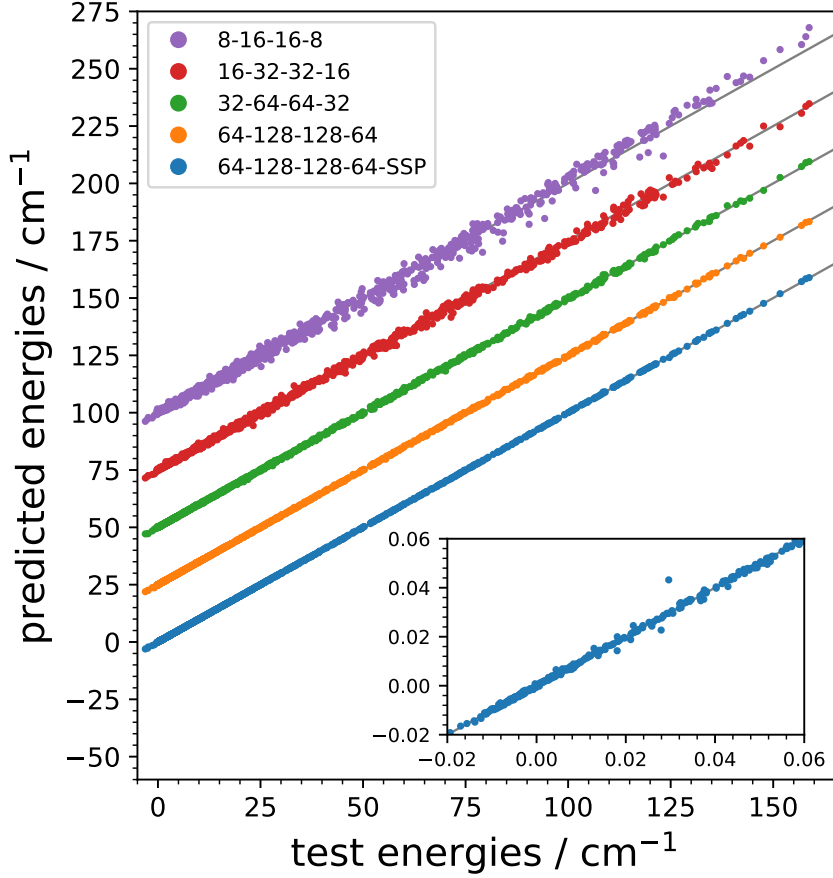


Figure 5.5: The four-body interaction energy predicted by each model, plotted against the corresponding true energy, for all samples in the test data set. To avoid overlapping the results, constant offsets are added to the predicted energies of each model. Shown are the outputs for the 64-128-128-64-SSP model (blue) [no offset], the 64-128-128-64 model (orange) [offset by 25 cm^{-1}], the 32-64-64-32 model (green) [offset by 50 cm^{-1}], the 16-32-32-16 model (red) [offset by 75 cm^{-1}], and the 8-16-16-8 model (purple) [offset by 100 cm^{-1}]. The inset shows the predictions for the 64-128-128-64-SSP model for a certain range of weaker energies (both axes in units of cm^{-1}).

At late stage training, the training error becomes nearly constant, and essentially all models beyond a certain epoch can be considered “ideal.” To avoid biasing the models to the test data, we arbitrarily choose each model’s weights corresponding to the very last epoch. The root mean squared errors (RMSE) of the models when compared against the test samples, starting with the 64-128-128-64 and decreasing in size, are 0.120 cm^{-1} , 0.254 cm^{-1} , 0.705 cm^{-1} , and 1.495 cm^{-1} . The RMSE of the 64-128-128-64-SSP model is 0.066 cm^{-1} . As expected, the smaller models produce worse predictions. Using shifted softplus over ReLU as an activation function for the largest model size roughly halves the RMSE.

In Fig. 5.5, we show the four-body interaction energy predicted by each model, against the corresponding true energy, for all samples in the test data set. The closer a point is to the diagonal line, the better the prediction. We see that the relatively large 64-128-128-64 and 64-128-128-64-SSP models do an excellent job of predicting the interaction energies in the test set. As we move to smaller models, the predictions become worse. In the inset of Fig. 5.5, we provide a closer look at the predictions of the 64-128-128-64-SSP model between -0.02 cm^{-1} and 0.06 cm^{-1} . Notably, these predictions are of energies below the model’s RMSE of 0.066 cm^{-1} . Training the models without the energy rescaling procedure described in Sec. (5.2.5) results in much worse predictions of the weak energy samples.

5.3 Discussion and Analysis

5.3.1 Lebedev Quadrature Comparison

To create the isotropic interaction energies in our PES, we use a 6-point Lebedev quadrature to average over the rotational degrees of freedom of each *para*-H₂ molecule.^{32,33,118} In a Lebedev quadrature, we calculate the interaction energies of the hydrogen molecules in specific combinations of angular orientations, then perform a weighted average of these energies. This weighted average only approximates the true integration over all the rotational degrees of freedom. The 6-point quadrature is described in Sec. (5.2.2).

Using a larger Lebedev quadrature scheme increases the accuracy of the approximation, but also increases the number of terms in the sum. The 6-point Lebedev quadrature scheme is the smallest possible scheme. It uses only 3 angular orientations per molecule, and with 4 molecules, we need $3^4 = 81$ energies to perform a single average. The next largest Lebedev scheme is the 14-point scheme. It uses 7 angular orientations per molecule, and thus requires $7^4 = 2401$ energies to perform a single average, nearly a 30-fold increase over the 6-point scheme.

To estimate the integration error, we select three tetrahedra of different side lengths, and calculate the four-body interaction energies for each one, using both the 6-point and 14-point Lebedev quadrature schemes. The results are shown in Table 5.1. In all three cases, the difference between the 6-point and 14-point averaged energies is acceptably small. Thus, the 6-point Lebedev quadrature scheme should be acceptable for the isotropic energies of our PES.

Table 5.1: Interaction energies of four *para*-H₂ molecules in a tetrahedron geometry for three chosen side lengths r , spherically averaged using the 6-point (V_4^{6p}) and 14-point (V_4^{14p}) Lebedev quadratures. Calculations are performed using an AVDZ basis set at the CCSD(T) level.

$r(\text{\AA})$	$V_4^{6p}(\text{cm}^{-1})$	$V_4^{14p}(\text{cm}^{-1})$	difference (%)
2.20	1.9917×10^2	1.9887×10^2	0.15
2.95	2.3478×10^0	2.3484×10^0	0.028
4.00	-5.9113×10^{-3}	-5.8332×10^{-3}	1.3

5.3.2 Basis Set and Method Comparison

The *ab initio* energies in the PES are calculated at the CCSD(T) level of theory, using an AVDZ atom-centred basis with a supplementary (*3s3p2d*) midbond basis set at the system’s centre of mass. Our choice of the calculation method and the size of the basis set takes into account both the quality of the resulting energies and the feasibility and expense of performing the calculations.

To compare the quality of calculations for different methods and basis sets, we calculate the four-body interaction energy of the tetrahedron geometry for side lengths between 2.2 Å and 4.5 Å. We do so for four different cases: at the HF level of theory using an AVDZ basis, at the MP2 level of theory using an AVDZ basis, at the CCSD(T) level of theory using an AVDZ basis, and at the CCSD(T) level of theory using an AVTZ basis. All calculations also use a supplementary (*3s3p2d*) basis set at the system’s centre of mass. We also show the four-body component of the four-body dispersion interaction potential, also known as the Bade potential (see Sec. (5.3.4)).

The results are shown in Fig. 5.6. We see that the CCSD(T) energies calculated using both the AVDZ and AVTZ basis sets are qualitatively and quantitatively very similar. The MP2 curve is qualitatively different from both CCSD(T) curves; it is more repulsive overall, and does not have the same “dip” around 3.8 Å. At short ranges, the MP2/AVDZ

and CCSD(T)/AVDZ energies differ from the CCSD(T)/AVTZ energies by around 10 % and 2 %, respectively. The HF curve is the most repulsive of all the curves, and shows no dispersion interaction whatsoever at long ranges. The analytic dispersive Bade potential is attractive for the entire range, and also much weaker. At long distances, the CCSD(T)/AVDZ and CCSD(T)/AVTZ curves converge to the Bade potential, which is a good indication of their quality.

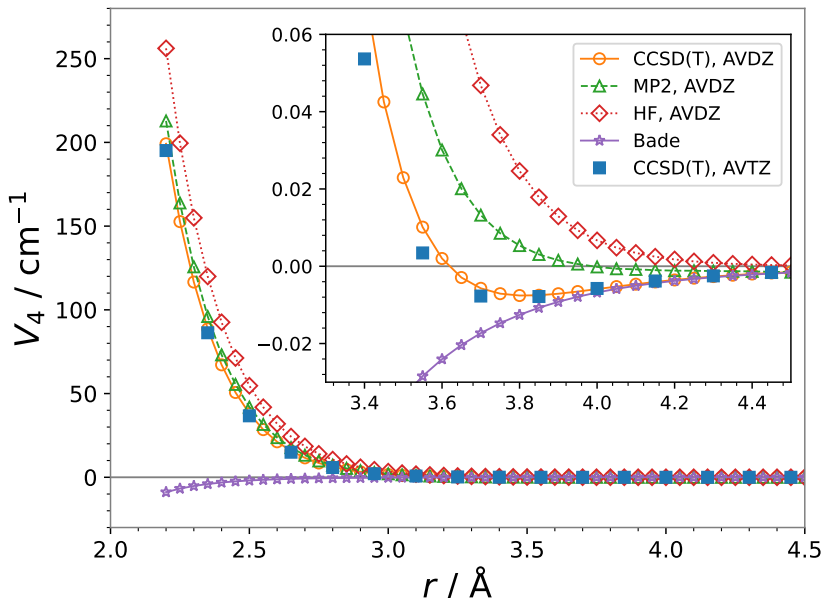


Figure 5.6: The four-body interaction potential energies of a tetrahedron, for side lengths between 2.2 Å and 4.5 Å. Shown are the energies calculated for {theory level, basis set} combinations of {CCSD(T), AVDZ} (orange circles), {CCSD(T), AVTZ} (solid blue squares), {MP2, AVDZ} (green triangles), and {HF, AVDZ} (red diamonds), alongside the Bade potential (purple stars). We use the Midzuno-Kihara-type estimate of the Bade coefficient in the Bade potential shown here, although the difference between that estimate and the other estimates isn’t visible at this scale.

In Fig. 5.7, we show the relative error of calculating the four-body interaction energy of a tetrahedron with an AVDZ basis set relative to an AVTZ basis set. The two curves are more similar in form at short distances and at long distances. The relative error naturally worsens around 3.6 Å, which is roughly where the interaction energies cross 0, as can be seen in Fig. 5.6.

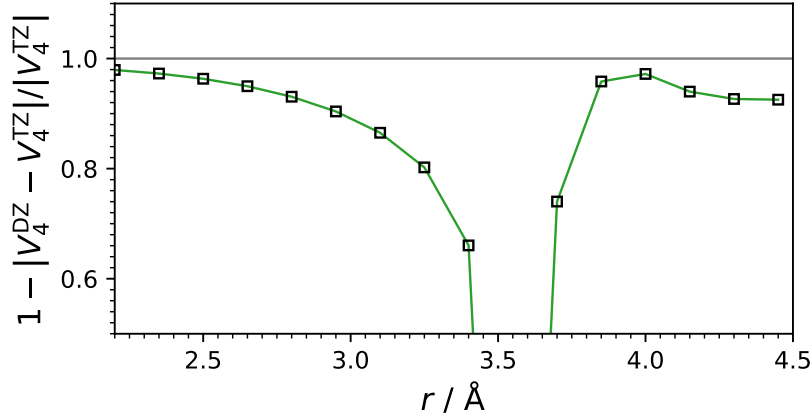


Figure 5.7: The relative error of calculating the four-body interaction energy of a tetrahedron with the AVDZ basis set relative to the AVTZ basis set as a function of the tetrahedron's side length r . The dip in the graph occurs around where the four-body energy crosses 0.

Because basis sets are finitely sized, all our calculations experience a systematic basis set size error (BSSE). One way to estimate the BSSE is to perform the same calculation with basis sets of several sizes, and extrapolate the energies to that of an infinitely large basis. We select two common extrapolation functions, the exponential decay function, given by

$$E_e(N) = E_e^{(\infty)} + c_e \exp\{-\alpha N\} \quad (5.8)$$

and the power decay function, given by

$$E_p(N) = E_p^{(\infty)} + c_p N^{-\beta}. \quad (5.9)$$

In both equations, N is the basis set cardinality (AVDZ = 2, AVTZ = 3, AVQZ = 4, etc.), $E_e^{(\infty)}$ and $E_p^{(\infty)}$ are the energies for an infinitely large basis set as predicted by the exponential and power decay fits respectively, and c_e , c_p , α , and β are constants.

In Fig. 5.8, we show the four-body interaction energy of tetrahedra of side lengths 2.2 Å and 2.95 Å as a function of the basis set cardinality, using both extrapolation functions. The fit constants are provided in the supplementary material. For a tetrahedron of side length 2.2 Å (2.95 Å), the AVDZ energy captures 97.9% (90.4%) of the AVTZ energy, 97.1% (88.5%) of the AVQZ energy, 96.6% (88.1%) of the exponentially extrapolated energy, and 95.9% (87.5%) of the power decay extrapolated energy. The power decay function

gives a much more pessimistic prediction of the convergence rate than the exponential decay function for the smaller tetrahedron.

The BSSE from using the AVDZ basis set is greater for larger geometries. Similarly, Wheatley *et al.* found that the difference between AVTZ and AVQZ calculations for four-body systems of helium, relative to the energy, increases with the size of the geometry.¹⁷⁰ The basis set size error from using an AVDZ basis set exceeds the fit error from the neural network for the 64-128-128-64 and 64-128-128-64-SSP models. We should also note that, compared to the AVDZ calculations, the AVTZ and AVQZ calculations respectively take an average of 25 and 420 times as long, and require about 4 and 30 times as much memory.

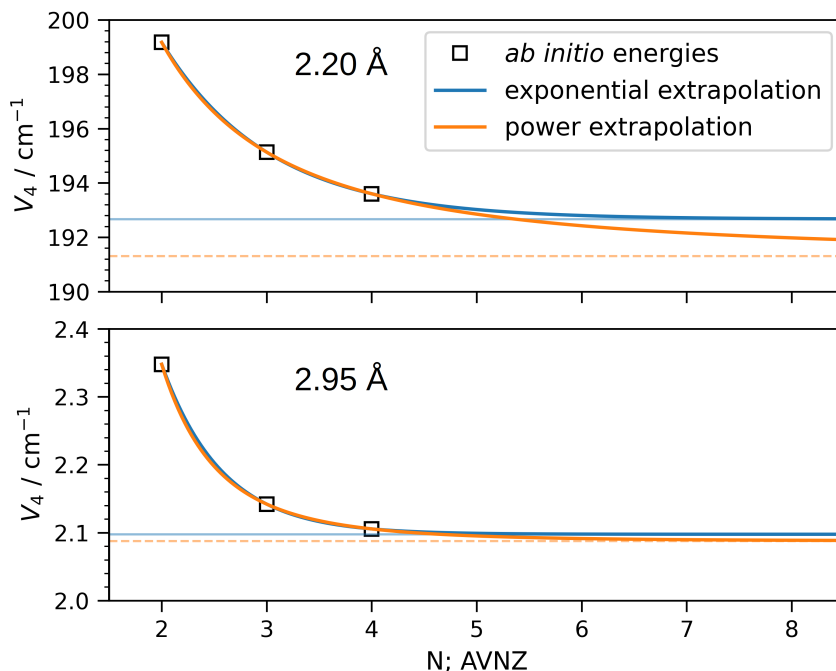


Figure 5.8: The four-body interaction energy of a tetrahedron of side length 2.2 Å (top) and of side length 2.9 Å (bottom), as a function of the basis set size (AVDZ = 2, AVTZ = 3, AVQZ = 4, etc.). The energies are extrapolated using the exponential decay extrapolation (blue curve) and the power decay extrapolation (orange curve). The horizontal lines represent the energies when extrapolated to a basis of infinite size.

To put these errors into perspective, one important use case for this PES is in numerical simulations of *para*-H₂ molecules alongside the FSH pair potential and the recent three-body PES. The FSH potential predicts a pair interaction energy of 1079.4 cm⁻¹ between

two *para*-H₂ molecules 2.2 Å apart. The recent three-body PES predicts a three-body interaction energy of -578.4 cm^{-1} between three *para*-H₂ molecules in an equilateral triangle of side length 2.2 Å. The four-body interaction energy for a tetrahedron with a side length of 2.2 Å is roughly 200 cm^{-1} . At 2.95 Å, the two-body, three-body, and four-body energies for the aforementioned shapes are 19.0 cm^{-1} , -16.4 cm^{-1} , and about 2.1 cm^{-1} , respectively. Keeping in mind that the four-body interaction is itself typically viewed as a correction term to the two-body and three-body interactions, the BSSE for this PES is probably tolerable for most applications, even at larger geometries.

5.3.3 Short Range: Exponential Decay

We have no training samples with a side length less than 2.2 Å, and thus we need a method to extrapolate our energies to smaller geometries.

Consider the function $V_4(\{sr_{ij}\})$, where $\{r_{ij}\}$ denotes the six side lengths with the condition that the shortest among them is 2.2 Å, and $s \approx 1$ is a scalar. By modifying s , we scale all the side lengths by the same factor. The four-body *para*-H₂ interaction energy $V_4(\{sr_{ij}\})$ generally scales exponentially as a function of s . We can see that this is the case for the tetrahedron geometry in Fig. 5.6. In fact, it is the case for all 8 of the highest energy *hcp* geometries shown in Fig. 5.9, and the majority of the lower-energy *hcp* geometries as well. Only some of the *hcp* geometries with very weak energies do not fit well as a function of s to the exponential decay function. However, because they are weak, the error from this poor fit is generally insignificant as long as the energies aren't extrapolated to extremely short distances. We use an exponential decay fit to extrapolate the four-body *para*-H₂ interaction PES to geometries where the shortest side length is less than 2.2 Å. The details are provided in the supplementary material.

5.3.4 Long Range: Bade Potential

The quadruple-dipole dispersion potential^{19,172} is

$$V_B(\{\mathbf{r}_i\}) = V_B^{(2)}(\{\mathbf{r}_i\}) + V_B^{(3)}(\{\mathbf{r}_i\}) + V_B^{(4)}(\{\mathbf{r}_i\}) \quad (5.10)$$

where the “pair component” is given by

$$V_B^{(2)}(\{\mathbf{r}_i\}) = -B_{12} \sum_{i < j}^4 r_{ij}^{-12}, \quad (5.11)$$

the “triplet component” is given by

$$V_B^{(3)}(\{\mathbf{r}_i\}) = -B_{12} \sum_{j=1}^4 \sum_{i \neq j}^3 \sum_{k>i, k \neq j}^4 \frac{(1 + (\mathbf{u}_{ij} \cdot \mathbf{u}_{jk}))^2}{r_{ij}^6 r_{jk}^6}, \quad (5.12)$$

and the “quadruplet component” is given by

$$V_B^{(4)}(\{\mathbf{r}_i\}) = -2B_{12}[f(1234) + f(1243) + f(1324)], \quad (5.13)$$

where

$$\begin{aligned} f(ijkl) = & (r_{ij}r_{jk}r_{kl}r_{li})^{-3} \Big[-1 \\ & + (\mathbf{u}_{ij} \cdot \mathbf{u}_{jk})^2 + (\mathbf{u}_{ij} \cdot \mathbf{u}_{kl})^2 + (\mathbf{u}_{ij} \cdot \mathbf{u}_{li})^2 \\ & + (\mathbf{u}_{jk} \cdot \mathbf{u}_{kl})^2 + (\mathbf{u}_{jk} \cdot \mathbf{u}_{li})^2 + (\mathbf{u}_{kl} \cdot \mathbf{u}_{li})^2 \\ & - 3(\mathbf{u}_{ij} \cdot \mathbf{u}_{jk})(\mathbf{u}_{jk} \cdot \mathbf{u}_{kl})(\mathbf{u}_{kl} \cdot \mathbf{u}_{ij}) \\ & - 3(\mathbf{u}_{ij} \cdot \mathbf{u}_{jk})(\mathbf{u}_{jk} \cdot \mathbf{u}_{li})(\mathbf{u}_{li} \cdot \mathbf{u}_{ij}) \\ & - 3(\mathbf{u}_{ij} \cdot \mathbf{u}_{kl})(\mathbf{u}_{kl} \cdot \mathbf{u}_{li})(\mathbf{u}_{li} \cdot \mathbf{u}_{ij}) \\ & - 3(\mathbf{u}_{jk} \cdot \mathbf{u}_{kl})(\mathbf{u}_{kl} \cdot \mathbf{u}_{li})(\mathbf{u}_{li} \cdot \mathbf{u}_{jk}) \\ & + 9(\mathbf{u}_{ij} \cdot \mathbf{u}_{jk})(\mathbf{u}_{jk} \cdot \mathbf{u}_{kl})(\mathbf{u}_{kl} \cdot \mathbf{u}_{li})(\mathbf{u}_{li} \cdot \mathbf{u}_{ij}) \Big]. \end{aligned} \quad (5.14)$$

In the above equations, B_{12} is a constant, and r_{ij} and $\mathbf{u}_{ij}$ are the distance and unit vector from position $\mathbf{r}_i$ to $\mathbf{r}_j$, respectively.

To create the long-range correction to our PES, we are only interested in the quadruplet component. This is because when calculating the four-body interaction energy in Eq. (5.2), the pair and triplet components are already subtracted out. Thus, our analytic long-range interaction will only be Eq. (5.13), which we refer to as the Bade potential in the remainder of the paper.

At long distances, the Bade potential decays as r^{-12} . We plot the Bade potential alongside the *ab initio* energies in Fig. 5.6 for tetrahedra of side lengths between 2.2 Å and 4.5 Å. We see that the CCSD(T) curves and the Bade potential converge at long distances. Unlike the *ab initio* energies, the Bade potential is net attractive and much weaker at short range. This discrepancy mirrors the case for the three-body interactions, wherein the ATM potential predicts the short-range energies to be weak and net repulsive, whereas the three-body *ab initio* energies are stronger and attractive.

Estimates of the B_{12} Coefficient

There is much less information in the literature on the interaction coefficient for the four-body dispersion interaction B_{12} than the C_6 coefficient for the pair interaction, or even the C_9 coefficient for the three-body interaction. We can approximate the C_6 , C_9 , and B_{12} using¹⁷²

$$C_6 \approx \frac{3}{4} \hbar \omega_0 \alpha^2 \quad (5.15)$$

$$C_9 \approx \frac{9}{16} \hbar \omega_0 \alpha^3 \quad (5.16)$$

$$B_{12} \approx \frac{45}{64} \hbar \omega_0 \alpha^4, \quad (5.17)$$

where α is the polarizability of molecular hydrogen, and $\hbar \omega_0$ is a coupling constant. Combining these equations gives

$$B_{12} = \frac{5}{3} \frac{C_9^2}{C_6}. \quad (5.18)$$

This relationship is similar to the Midzuno-Kihara approximation,¹⁷⁷ which relates C_9 and C_6 though $C_9/C_6 = 3\alpha/4$. Using values of $C_6 = 58203.6 \text{ cm}^{-1} \text{ \AA}^6$ from Ref. [31], and $C_9 = 34336.2 \text{ cm}^{-1} \text{ \AA}^9$ from Ref. [25], we find an estimate of $B_{12} = 33760.1 \text{ cm}^{-1} \text{ \AA}^{12}$. We can also approximate B_{12} directly from the *ab initio* CCSD(T) interaction energies for the tetrahedron geometry. The estimates of B_{12} from energies calculated with the AVDZ and AVTZ basis sets, are $31667.0 \text{ cm}^{-1} \text{ \AA}^{12}$ and $29492.8 \text{ cm}^{-1} \text{ \AA}^{12}$, respectively (see the supplementary material). The Midzuno-Kihara-like estimate of B_{12} is in surprisingly good agreement with the estimates from the *ab initio* energies, given that it is derived from a combination of three approximations.

For the PES published with this paper, we decide to use the AVTZ-based estimate for B_{12} . We should emphasize that the neural network presented with this paper does not depend on the chosen value of B_{12} . Instead, the PES adjusts the interaction energy of a sample at long ranges either after or instead of evaluating the neural network. For cases where the long-range accuracy is especially important, the value of B_{12} can be easily replaced in the code without having to retrain the model.

Transitions Between Different Parts of the PES

With no intervention, there are small discontinuities between the short-range, long-range, and multilayer perceptron parts of the PES. These discontinuities are typically very small, and in situations where only energies are required, are probably not too concerning.

However, there are many situations where it would be desirable to calculate forces with the PES. The Python functor for the PES provided in the accompanying repository wraps around the different parts of the PES introduces continuous transitions between them. When used with the 64-128-128-64-SSP model, this PES is continuously differentiable. The conceptual details of the transition between the short-range, long-range, and multilayer perceptron parts of the PES are provided in the supplementary material.

5.3.5 Interactions in Solid Parahydrogen

Solid *para*-H₂ is a hexagonal close-packed (*hcp*) crystal.⁹ Even though solid *para*-H₂ is a quantum solid whose molecules are very delocalized about their nominal lattice sites, we can gain several insights from analyzing its classical frozen lattice configuration.

Earlier, we mentioned generating training data based on the geometries of the *hcp* lattice. We select a reference molecule in the lattice, and generate all four-body geometries involving this reference molecule such that (1) at least one side length is equal to the lattice constant, and (2) no side length is more than twice the lattice constant. Geometries that do not obey these conditions have a nearly negligible contribution to the total interaction energy, even at a lattice constant as low as 2.2 Å. There are 83 distinct four-body shapes that satisfy these conditions, each occurring with a certain frequency. A subset of the geometries are presented in Table 5.2. It shows the six side lengths $\{r_{ij}\}$ of the geometry normalized by the lattice constant, and the number of times it appears subject to the aforementioned conditions N_c . The geometries are listed in order of increasing average side length. The full table with all 83 geometries, including sample energy values is provided in the supplementary material.

Consider a classical frozen *hcp* lattice with a lattice constant of 2.2 Å. For each geometry, we take its four-body *para*-H₂ interaction energy $V_4(\{r_{ij}\})$, multiply it with the number of times that geometry appears N_c , and divide by 4 to account for multiple counting. This gives us $V_4^{(\text{tot})}$, the total contribution of each geometry to the average four-body *para*-H₂ interaction potential energy per particle in the *hcp* lattice. In Fig. 5.9, we show $V_4^{(\text{tot})}$ for each geometry ID in Table 5.2.

We see that nearly all the geometries provide a repulsive contribution for the four-body *para*-H₂ interaction energy. We also see that the majority of geometries have nearly no contribution. Of the 8 geometries with a total contribution greater than 300 cm⁻¹, 7 of them have the lowest geometry ID numbers (and thus the shortest average side lengths). However, the average pair distance itself cannot explain the average interaction energy on its own. Consider geometries 15 and 16 in Table 5.2. These two geometries have the same

six side lengths but different physical structures (none of the 24 allowed index permutations turn either geometry into the other). At a lattice constant of $2.2 \text{ \AA}$, they have drastically different total contributions $V_4^{(\text{tot})}$ of 3.99 cm^{-1} and 589.18 cm^{-1} , respectively.

The four-body interaction potential for *para*-H₂ shares many features with the four-body interaction potential of helium.¹⁷⁰ Both are repulsive at short range. At larger intermolecular separations, the energies are a mixed of repulsive and attractive, although the repulsive energies tend to be stronger and cover a larger amount of the coordinate space.

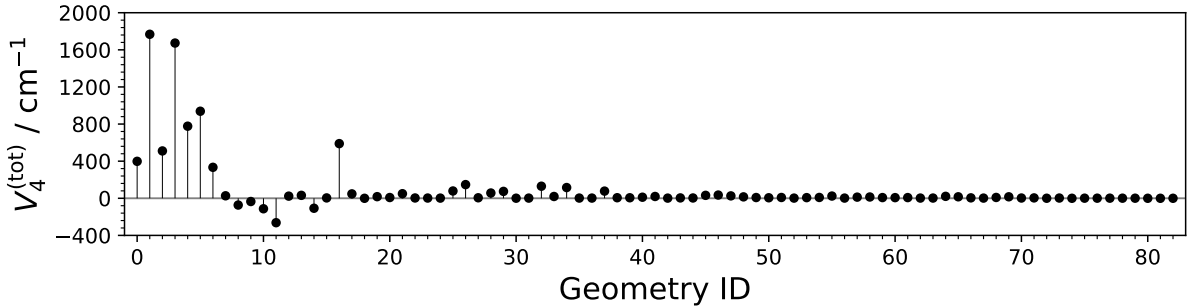


Figure 5.9: The total contribution of each geometry to the average four-body *para*-H₂ interaction potential energy per particle in an *hcp* lattice, when the lattice constant is $2.2 \text{ \AA}$ ($\rho \approx 0.133 \text{ \AA}^{-3}$). The corresponding geometry IDs are provided in the supplementary material. The geometries are labelled in order of increasing average side length.

We can look at the impact of many-body interactions in rare gas solids, such as solid helium-4, neon, argon, krypton, and xenon, for insights on how they behave in solid *para*-H₂. Like solid *para*-H₂, the rare gas solids are molecular solids with weak intermolecular dispersion interactions, and are common subjects in quantum path integral simulations.¹⁷⁸

There appears to be a common pattern in many-body interactions of rare gas atoms. The overall interaction energy at short intermolecular separations switches signs based on the parity of the interaction order. At high densities in solid helium-4, Tian *et al.* found that the 2-body, 4-body, and 6-body interactions have net repulsive contributions, while the 3-body and 5-body interactions have net attractive contributions.²⁰ The magnitude of the contributions also decrease with every increasing order. Similarly, at high densities in solid krypton, the 2-body and 4-body interactions have overall repulsive contributions, while the 3-body interaction is attractive.¹⁷⁹ In solid neon, argon, and xenon, the three-body interaction energy is known to be overall attractive.¹⁸⁰ It appears that many-body

interactions between *para*-H₂ molecules also follow this pattern.

Table 5.2: The geometries of the *para*-H₂ molecules in an *hcp* lattice. The first column shows the geometry ID label. The geometries are labelled in order of increasing side length. The second column shows N_c , the number of times that geometry appears in the first two shells. The remaining six columns show the side lengths of the geometries, normalized by the lattice constant. This table only shows 17 of the 83 geometries that satisfy the conditions described in the text. A full table is provided in the supplementary material.

ID	N_c	r_{12}	r_{13}	r_{14}	r_{23}	r_{24}	r_{34}
0	8	1	1	1	1	1	1
1	48	1	1	1	1	1	$\sqrt{2}$
2	12	1	1	1	1	1	$\sqrt{8/3}$
3	36	1	1	1	1	1	$\sqrt{3}$
4	12	1	1	$\sqrt{2}$	$\sqrt{2}$	1	1
5	144	1	1	1	1	$\sqrt{2}$	$\sqrt{3}$
6	48	1	1	1	1	$\sqrt{2}$	$\sqrt{11/3}$
7	24	1	1	1	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{8/3}$
8	72	1	1	1	1	$\sqrt{3}$	$\sqrt{3}$
9	48	1	1	1	1	$\sqrt{8/3}$	$\sqrt{11/3}$
10	48	1	1	1	1	$\sqrt{3}$	$\sqrt{11/3}$
11	96	1	1	1	1	$\sqrt{3}$	2
12	24	1	1	$\sqrt{2}$	$\sqrt{2}$	1	$\sqrt{11/3}$
13	24	1	1	1	$\sqrt{2}$	$\sqrt{2}$	2
14	24	1	1	1	1	$\sqrt{11/3}$	$\sqrt{11/3}$
15	48	1	1	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{3}$
16	144	1	1	$\sqrt{2}$	$\sqrt{3}$	1	$\sqrt{3}$

For the equation of state, we calculate the energy per molecule ϵ as a function of the density ρ for a solid *para*-H₂ *hcp* lattice with no zero-point motion. We can then calculate the pressure P as a function of density using

$$P = \rho^2 \left. \frac{\partial \epsilon}{\partial \rho} \right|_T \quad (5.19)$$

For these calculations, we use the FSH two-body potential,¹⁶ the recent isotropic three-body PES,⁸¹ and the 64-128-128-64 four-body PES presented in this paper.

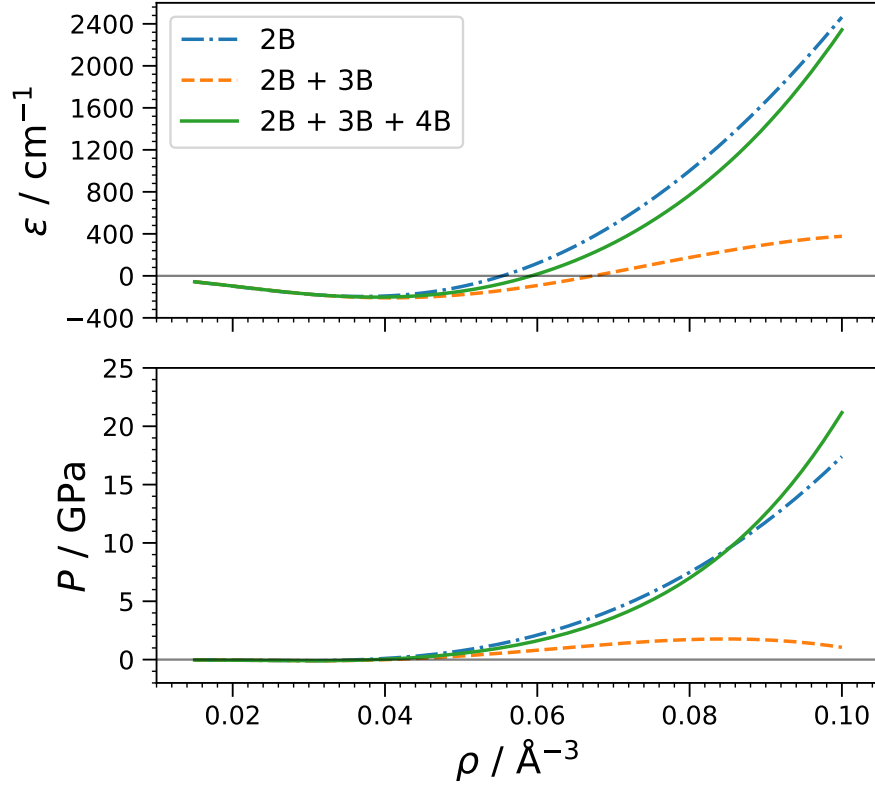


Figure 5.10: We show the classical energy-density curves (top) and classical pressure-density curves (bottom), for a frozen *hcp* lattice whose interactions are made up of the two-body potential on its own (blue, dash-dotted), the two-body and three-body potentials together (orange, dashed), and the two-, three-, and four-body potentials together (green, solid). The top and bottom subfigures both share the same horizontal axis. The highest density shown here, $\rho = 0.1 \text{ \AA}^{-3}$, corresponds to a lattice constant of $a = 2.42 \text{ \AA}$.

In Fig. 5.10, we show the classical energy-density curves (top) and classical pressure-density curves (bottom), for a lattice whose interactions are made up of the two-body potential on its own, the two-body and three-body potentials together, and the two-, three-, and four-body potentials together. We see that, below about $0.08 \text{ \AA}^{-3}$, the addition of the four-body interaction moves the pressure-density curve below the case where the two-body PES is used on its own, and above the case where the two-body and three-body PESs are present. Qualitatively, these results are promising. In an earlier study of PIMC simulations of solid *para*-H₂, we found that the pair potential overestimates the pressure-density curve

compared to experiment.³⁷ However, the inclusion of the three-body potential causes an overcorrection at higher densities,¹²³ causing the simulation to underestimate the pressure-density curve. Except at very high densities, the four-body potential should place the pressure-density curve between these two former cases, bringing it closer to experiment.

5.4 Conclusion

We have presented an isotropic *ab initio* four-body interaction PES for *para*-H₂. The electronic structure calculations are performed at the CCSD(T) level with an AVDZ atom-centred basis set, supplemented by a (3*s*3*p*2*d*) midbond function. The energies are fit to a multilayer perceptron. The procedure used to construct the PES has been provided in detail. This includes the methods of sample generation, the transformations on the input side lengths and the output energies to help improve the training process, and the training of the neural network itself.

We have also discussed how the interaction energy is extrapolated outside the training data’s regime, using an analytic dispersion interaction at long intermolecular separations, and an exponential decay when the *para*-H₂ molecules are close together. The four-body *para*-H₂ interaction energy is overall repulsive at high densities. To put this into context, we know from previous PIMC simulations that the *ab initio* two-body FSH PES on its own is too repulsive at high densities,³⁷ and that the inclusion of the *ab initio* three-body PES is too attractive at high densities.¹²³ Using a classical approximation, we show that the additional inclusion of the four-body PES will likely improve the agreement of the simulation results with experiment, except at very high densities.

Chapter 6

Conclusion

Over the course of this thesis, I create *ab initio* three-body and four-body PESs for *para*-H₂, and use them alongside an existing *ab initio* two-body PES to perform PIMC simulations of solid *para*-H₂.

In chapter 2, I show the results of performing PIMC simulations of solid *para*-H₂ using the FSH potential.³⁷ This is a 1D *ab initio* two-body PES, created by taking Hinde’s 6D two-body PES and using the AHR method to account for the rovibrational degrees of freedom. We account for the finite system size error by extrapolating the results to an infinite number of *para*-H₂ molecules, and the Trotter error by extrapolating the results to an infinite number of imaginary time slices. These simulations offer a good prediction of the equilibrium density. They also provide a good the first ever *a priori* prediction of the vibrational matrix shift of solid *para*-H₂, in good agreement with experimental results. However, we find that the FSH potential overestimates the pressure-density relationship at high densities; its short-range interactions are too repulsive on their own.

In chapter 3, I create an isotropic *ab initio* three-body PES for *para*-H₂.⁸¹ The energies are calculated at the CCSD(T) level of theory, using an AVTZ atom-centred basis set with additional (3s3p2d) midbond functions. The rigid rotor approximation removes the bond length degrees of freedom, and a 6-point Lebedev quadrature integrates over the angular orientation degrees of freedom. These operations leave us with an isotropic PES whose only parameters are the three relative pair distances between the centres of mass of the three *para*-H₂ molecules. We convert the three relative pair distances into rescaled Jacobi coordinates, which are convenient to use, mutually orthogonal, and physically insightful. To create the PES from the energy data, we use the reproducing kernel Hilbert space toolkit by Unke *et al.* The three-body PES is overall attractive at short range, following

an exponential decay trend. At long range, it is overall weakly repulsive, converging to the dispersive analytic Axilrod-Teller-Muto potential. We find that only a small number of geometries in the *hcp* lattice are responsible for the majority of the total three-body interaction energy.

In chapter 4, I show the results of performing PIMC simulations of solid *para*-H₂ using both the FSH potential and the three-body PES constructed in the previous chapter.¹²³ We use tail corrections to account for the finite system size error of our simulation box. We correct the Trotter error both by extrapolation and by performing our simulations with a sufficiently large number of beads. The three-body PES is accounted for in one of two ways. In the full treatment, we use the two-body and three-body PESs for both sampling and estimating. In the perturbative treatment, only the two-body PES is used for sampling, while both potentials are used for estimating. Both the full and perturbative treatments give very similar pressure-density relationships. The three-body PES introduces minor changes to the structure of the solid. The kinetic, two-body, and three-body potential energy components each change when the three-body PES is used for sampling. However, the changes in each component partially cancel each other out, and the change in the total energy is fairly small. The inclusion of the three-body PES improves the agreement of the simulation results with experiment at low densities. However, at high densities, the attractive interactions from three-body PES overcorrect the repulsive interactions from two-body PES, resulting in an underestimation of the pressure-density curve.

In chapter 5, I create an isotropic *ab initio* four-body PES for *para*-H₂.¹⁶³ The energies are calculated at the CCSD(T) level of theory, using an AVDZ atom-centred basis set with additional (*3s3p2d*) midbond functions. As was done for the three-body PES, we use the rigid rotor approximation and the 6-point Lebedev quadrature to reduce the number of degrees of freedom to only the relative pair distances between the *para*-H₂ molecules. A multilayer perceptron is used to create a PES from the energy data. To account for the permutation symmetry of the molecule labels, we use a lexicographic sorting transformation on the permutations of the input side lengths. We use a rescaling function transformation inspired by the approximate form of the potential, to ensure that the MLP is able to fit energies over five orders of magnitude with similar relative accuracy. The final PES is continuously differentiable everywhere. The four-body PES is overall repulsive at short range, following an exponential decay trend. At long range, it is overall weakly attractive, converging to the dispersive Bade potential. Similar to the case for the three-body energies, only a small number of geometries in the *hcp* lattice are responsible for the majority of the total four-body interaction energy. A classical calculation of the density-pressure relationship shows that the inclusion of the four-body PES will likely improve the agreement of simulation results with experiment up to some higher density.

6.1 Future work

6.1.1 Using the four-body PES in PIMC simulations

Although I have created the four-body PES, I have not yet used it in a PIMC simulation of solid *para*-H₂. I anticipate that its inclusion will improve the agreement of the simulations with experimental data to higher densities. However, similar to our experiences with the three-body PES, at high enough densities the four-body PES will likely no longer be sufficient, and higher-order interactions, such as the five-body and six-body interactions, will become relevant.

One challenge we face is that the PIMC simulations become considerably slower when higher-order many-body interactions are included. The four-body PES itself is fairly slow to evaluate, because passing inputs through a multilayer perceptron involves calling activation functions and performing matrix multiplications repeatedly in sequence. We compare this with the two-body PES, which uses only 1D linear interpolation, and the three-body PES, which we can refit to use trilinear interpolation. There is also the combinatorial nature of evaluating the four-body interaction, which scales with the number of particles in the simulation box as $\mathcal{O}(N^4)$. Certain strategies and tradeoffs will have to be created to make the simulations computationally tractable.

6.1.2 Many-body force evaluations

The three-body PES has been developed with the intention of being used only for energy calculations. We can refit the energies using another machine learning method to allow the PES to provide forces. This change would allow us to use the three-body PES in PIMD simulations.

6.1.3 Higher-order many-body PESs

We expect the inclusion of the four-body PES in our simulations to only accurately model solid *para*-H₂ until some particular high density. Beyond this density, we expect to need higher-order many-body interactions, such as the five-body or six-body interactions, to be able to match the experimental results.

There are several challenges involved with calculating the five-body interaction energy. The number of calculations required for a single isotropic energy data point increases drastically with the number of particles involved. For example, a single *ab initio* four-body

potential energy value is calculated using a linear combination of 15 separate energy values (see Eq. (1.3)) and to make the energy isotropic, we take the average of $3^4 = 81$ interaction energies in the Lebedev quadrature, for a total of 1215 separate energy calculations. However, a single *ab initio* five-body potential energy value requires 31 separate energy values, and the Lebedev quadrature requires $3^5 = 243$ such interaction energies, for a total of 7533 separate energy calculations.

Certain compromises would have to be made to create the five-body PES. The calculations might have to be done using a smaller basis set, or lower-quality electronic structure methods. We might need to reduce the total amount of training data for the model, such as by reducing the number of weak energy samples by ignoring geometries with large pair distances. We might also find a more optimal way to sample the training data, such as by sampling geometries directly from state snapshots from a PIMC simulation.

6.1.4 PIMC simulations with more degrees of freedom

The two-body, three-body, and four-body PESs used in this thesis have all been reduced to isotropic potentials using various approximations such as the AHR method, the rigid rotor approximation, and Lebedev quadratures. However, we might find different results if we were to perform our PIMC simulations with more degrees of freedom intact.

References

- ¹ J. van Kranendonk, *Solid hydrogen: Theory of the properties of solid H₂, HD, and D₂*. Boston, MA: Springer, 1983.
- ² H. P. Gush, W. F. J. Hare, E. J. Allin, and H. L. Welsh, “The infrared fundamental band of liquid and solid hydrogen,” *Can. J. Phys.*, vol. 38, p. 176, 1960.
- ³ C. S. Barrett, L. Meyer, and J. Wasserman, “Crystal structure of solid hydrogen and deuterium, and of neon–hydrogen and neon–deuterium mixtures,” *J. Chem. Phys.*, vol. 45, p. 834, 1966.
- ⁴ O. Bostanjoglo and R. Kleinschmidt, “Crystal structure of hydrogen isotopes,” *J. Chem. Phys.*, vol. 46, p. 2004, 1967.
- ⁵ O. Schnepp, “One-phonon excited states of solid H₂ and D₂ in the ordered phase,” *Phys. Rev. A*, vol. 2, no. 6, p. 2574, 1970.
- ⁶ F. Fernandez-Alonso, C. Cabrillo, R. Fernández-Perea, F. J. Bermejo, M. A. González, C. Mondelli, and E. Farhi, “Solid *para*-hydrogen as the paradigmatic quantum crystal: Three observables probed by ultrahigh-resolution neutron spectroscopy,” *Phys. Rev. B*, vol. 86, p. 144524, 2012.
- ⁷ L. H. Nosanow, “Theory of quantum crystals,” *Phys. Rev.*, vol. 146, p. 120, 1966.
- ⁸ M. Dusseault and M. Boninsegni, “Atomic displacements in quantum crystals,” *Phys. Rev. B*, vol. 95, p. 104518, 2017.
- ⁹ I. F. Silvera, “The solid molecular hydrogens in the condensed phase: Fundamentals and static properties,” *Rev. Mod. Phys.*, vol. 52, p. 393, 1980.

- ¹⁰ T. R. Prisk, R. T. Azuah, D. L. Abernathy, G. E. Granroth, T. E. Sherline, P. E. Sokol, J. Hu, and M. Boninsegni, “Zero-point motion of liquid and solid hydrogen,” *Phys. Rev. B*, vol. 107, p. 094511, 2023.
- ¹¹ R. J. Hinde, “Three-body interactions in solid parahydrogen,” *Chem. Phys. Lett.*, vol. 460, pp. 141–145, 2008.
- ¹² S. Manzhos, K. Nakai, and K. Yamashita, “Three-body interactions in clusters CO–($p\text{H}_2$) $_n$,” *Chem. Phys. Lett.*, vol. 493, pp. 229–233, 2010.
- ¹³ T. Omiyinka and M. Boninsegni, “Pair potentials and equation of state of solid *para*-hydrogen to megabar pressure,” *Phys. Rev. B*, vol. 88, p. 024112, 2013.
- ¹⁴ M. J. Elrod and R. J. Saykally, “Many-body effects in intermolecular forces,” *Chem. Rev.*, vol. 94, p. 1975, 1994.
- ¹⁵ M. Alkan, P. Xu, and M. S. Gordon, “Many-body dispersion in molecular clusters,” *J. Phys. Chem. A*, vol. 123, pp. 8406–8416, 2019.
- ¹⁶ N. Faruk, M. Schmidt, H. Li, R. J. Le Roy, and P.-N. Roy, “First-principles prediction of the Raman shifts in parahydrogen clusters,” *J. Chem. Phys.*, vol. 141, p. 014310, 2014.
- ¹⁷ B. M. Axilrod and E. Teller, “Interaction of the van der Waals type between three atoms,” *J. Chem. Phys.*, vol. 11, p. 299, 1943.
- ¹⁸ Y. Muto, “Force between nonpolar molecules,” *J. Phys. Math. Soc. Jpn.*, vol. 17, p. 629, 1943.
- ¹⁹ W. L. Bade, “Drude-model calculation of dispersion forces. III. The fourth-order contribution,” *J. Chem. Phys.*, vol. 28, pp. 282–284, 1958.
- ²⁰ C.-L. Tian, F.-S. Liu, F.-Q. Jing, and L.-C. Cai, “Five- and six-body effects on equation of state of solid ^4He ,” *J. Phys. Condens. Matter*, vol. 18, pp. 8103–8112, 2006.
- ²¹ M. Kállay, P. R. Nagy, D. Mester, Z. Rolik, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, B. Hégyel, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and Á. Ganyecz, “The MRCC program system: Accurate quantum chemistry from water to proteins,” *J. Chem. Phys.*, vol. 152, p. 074107, 2020.

- ²² MRCC, a quantum chemical program suite written by M. Kállay and P. R. Nagy and D. Mester and L. Gyevi-Nagy and J. Csóka and P. B. Szabó and Z. Rolik and G. Samu and J. Csontos and B. Hégyel and Á. Ganyecz and I. Ladjánszki and L. Szegedy and B. Ladóczki and K. Petrov and M. Farkas and P. D. Mezei and R. A. Horbáth. See www.mrcc.hu.
- ²³ A. Szabo and N. S. Ostlund, *Modern quantum chemistry: Introduction to advanced electronic structure theory*. Dover Publications, 2 ed., 1996.
- ²⁴ T. D. Crawford and H. F. S. III, “An introduction to coupled cluster theory for computational chemists,” in *Reviews in computational chemistry volume 14*, pp. 33–136, John Wiley and Sons Inc., 2000.
- ²⁵ R. J. Hinde, “A six-dimensional H₂–H₂ potential energy surface for bound state spectroscopy,” *J. Chem. Phys.*, vol. 128, p. 154308, 2008.
- ²⁶ T. H. Dunning Jr, “Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen,” *J. Chem. Phys.*, vol. 90, no. 2, pp. 1007–1023, 1989.
- ²⁷ R. A. Kendall, T. H. Dunning Jr, and R. J. Harrison, “Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions,” *J. Chem. Phys.*, vol. 96, no. 9, pp. 6796–6806, 1992.
- ²⁸ S. F. Boys and F. Bernardi, “The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors,” *Mol. Phys.*, vol. 19, no. 4, p. 553, 1970.
- ²⁹ H. Li, P.-N. Roy, and R. J. Le Roy, “An “adiabatic-hindered-rotor” treatment allows *para*-H₂ to be treated as if it were spherical,” *J. Chem. Phys.*, vol. 133, p. 104305, 2010.
- ³⁰ C. Schwartz and R. J. Le Roy, “Nonadiabatic eigenvalues and adiabatic matrix elements for all isotopes of diatomic hydrogen,” *J. Mol. Spectrosc.*, vol. 121, pp. 420–439, 1987.
- ³¹ M. Schmidt, J. M. Fernández, N. Faruk, M. Nooijen, R. J. Le Roy, J. H. Morilla, G. Tejeda, S. Montero, and P.-N. Roy, “Raman vibrational shifts of small clusters of hydrogen isotopologues,” *J. Phys. Chem. A*, vol. 119, p. 12551, 2015.
- ³² V. I. Lebedev, “Quadratures on a sphere,” *Zh. Vychisl. Mat. Mat. Fiz.*, vol. 16, no. 2, pp. 10–24, 1976.

- ³³ A. D. Becke, “A multicenter numerical integration scheme for polyatomic molecules,” *J. Chem. Phys.*, vol. 88, no. 4, p. 2547, 1988.
- ³⁴ D. M. Ceperley, “Path integrals in the theory of condensed helium,” *Rev. Mod. Phys.*, vol. 67, p. 279, 1995.
- ³⁵ R. P. Feynman and A. R. Hibbs, *Quantum mechanics and path integrals*. Mineola, New York: Dover Publications, emended ed., 2010.
- ³⁶ O. T. Unke and M. Meuwly, “Toolkit for the construction of reproducing kernel-based representations of data: Application to multidimensional potential energy surfaces,” *J. Chem. Inf. Model.*, vol. 57, p. 1923, 2017.
- ³⁷ A. Ibrahim, L. Wang, T. Halverson, R. J. Le Roy, and P.-N. Roy, “Equation of state and first principles prediction of the vibrational matrix shift of solid parahydrogen,” *J. Chem. Phys.*, vol. 151, p. 244501, 2019.
- ³⁸ M. Zoppi and M. Neumann, “PIMC simulations of solid parahydrogen,” *Phys. Rev. B*, vol. 43, no. 13, p. 10242, 1991.
- ³⁹ T. Momose, D. P. Weliky, and T. Oka, “The stimulated raman gain spectrum of the $Q_{10}(0)$ transition of solid parahydrogen,” *J. Mol. Spectrosc.*, vol. 153, pp. 760–761, 1992.
- ⁴⁰ T. Oka, “High-resolution spectroscopy of solid hydrogen,” *Ann. Rev. Phys. Chem.*, vol. 44, pp. 299–333, 1993.
- ⁴¹ M. Zoppi, M. Neumann, and M. Celli, “Microscopic structure factor of liquid parahydrogen,” *Phys. Rev. B*, vol. 65, p. 092204, 2002.
- ⁴² F. Operetto and F. Pederiva, “Diffusion Monte Carlo study of the equation of state of solid para- H_2 ,” *Phys. Rev. B*, vol. 73, p. 184124, 2006.
- ⁴³ T. Lindenau, M. L. Ristig, K. A. Gernoth, J. Dawidowski, and F. J. Bermejo, “properties of liquid h_2 ,” *Int. J. Mod. Phys. B*, vol. 20, pp. 5035–5046, 2006.
- ⁴⁴ S. Grebenev, B. G. Sartakov, J. P. Toennies, and A. F. Vilesov, “Evidence for superfluidity in para-hydrogen clusters inside helium-4 droplets at 0.15 Kelvin,” *Science*, vol. 289, p. 1532, 2000.
- ⁴⁵ S. Grebenev, E. Lugovoi, B. G. Sartakov, J. P. Toennies, and A. F. Vilesov, “Spectroscopy of OCS-hydrogen clusters in He droplets,” *Faraday Disc.*, vol. 118, p. 19, 2001.

- ⁴⁶ S. Grebenev, B. G. Sartakov, J. P. Toennies, and A. F. Vilesov, "The structure of the OCS-H₂ van der Waals complex embedded inside ⁴He/³He droplets," *J. Chem. Phys.*, vol. 114, p. 617, 2001.
- ⁴⁷ S. Grebenev, B. G. Sartakov, J. P. Toennies, and A. F. Vilesov, "Effect of the symmetry of H₂ molecules on their rotations around an OCS molecule in superfluid ⁴He droplets," *Phys. Rev. Lett.*, vol. 89, p. 225301, 2002.
- ⁴⁸ S. Grebenev, B. G. Sartakov, J. P. Toennies, and A. F. Vilesov, "High-resolution infrared spectra of the OCS-H₂, -HD, and -D₂ van der Waals complexes in liquid helium droplets," *J. Chem. Phys.*, vol. 118, p. 8656, 2003.
- ⁴⁹ F. Paesani, R. E. Zillich, Y. Kwon, and K. B. Whaley, "OCS in *para*-hydrogen clusters: Rotational dynamics and superfluidity," *J. Chem. Phys.*, vol. 122, p. 181106, 2005.
- ⁵⁰ S. Grebenev, B. G. Sartakov, J. P. Toennies, and A. F. Vilesov, "Spectroscopic investigation of OCS-(*p*-H₂)_{*n*} (*n* = 1 – 16) complexes inside helium droplets: Evidence for superfluid behaviour," *J. Chem. Phys.*, vol. 132, p. 064501, 2010.
- ⁵¹ H. Li, R. J. Le Roy, P.-N. Roy, and A. R. W. McKellar, "Molecular superfluid: Nonclassical rotations in doped *para*-hydrogen clusters," *Phys. Rev. Lett.*, vol. 105, p. 133401, 2010.
- ⁵² T. Zeng, H. Li, and P.-N. Roy, "Simulating asymmetric top impurities in superfluid clusters: A *para*-water dopant in *para*-hydrogen," *J. Phys. Chem. Lett.*, vol. 4, p. 18, 2013.
- ⁵³ T. Zeng, G. Guillon, J. T. Cantin, and P.-N. Roy, "Probing the superfluid response of *para*-hydrogen with a sulfur dioxide dopant," *J. Phys. Chem. Lett.*, vol. 4, p. 2391, 2013.
- ⁵⁴ O. N. Osychenko, R. Rota, and J. Boronat, "Superfluidity of metastable glassy bulk *para*-hydrogen at low temperature," *Phys. Rev. B*, vol. 85, p. 224513, 2012.
- ⁵⁵ D. P. Weliky, K. E. Kerr, T. J. Byers, Y. Zhang, T. Momose, and T. Oka, "High-resolution infrared spectroscopy of isotopic impurity q₁(0) transitions in solid parahydrogen," *J. Chem. Phys.*, vol. 105, p. 4461, 1996.
- ⁵⁶ Y. Zhang, T. J. Byers, M.-C. Chan, T. Momose, K. E. Kerr, D. P. Weliky, and T. Oka, "High-resolution infrared spectroscopy of the *j* = 1 H₂ pair in parahydrogen crystals," *Phys. Rev. B*, vol. 58, p. 218, 1998.

- ⁵⁷ T. Momose, H. Honshina, M. Fushitani, and H. Katsuki, “pure hcp structure of pure solid h₂,” *Vib. Spectrosc.*, vol. 34, pp. 95–108, 2004.
- ⁵⁸ M. E. Fajardo, “Co in ph₂ matrix,” *J. Phys. Chem. A*, vol. 117, p. 13504, 2013.
- ⁵⁹ G. Tejeda, J. M. Fernández, S. Montero, D. Blume, and J. P. Toennies, “Raman spectroscopy of small para-H₂ clusters formed in cryogenic free jets,” *Phys. Rev. Lett.*, vol. 92, no. 22, p. 223401, 2004.
- ⁶⁰ R. D. Goodwin and H. M. Roder, “eos-lowpressure,” *Cryogenics*, vol. 3, p. 12, 1963.
- ⁶¹ S. C. Durana and J. P. McTague, “isotherms at 4.2k,” *J. Low Temp. Phys.*, vol. 21, p. 21, 1975.
- ⁶² I. F. Silvera and V. V. Goldman, “The isotropic intermolecular potential for H₂ and D₂ in the solid and gas phases,” *J. Chem. Phys.*, vol. 69, p. 4209, 1978.
- ⁶³ A. Driessen, J. A. de Waal, and I. F. Silvera, “Experimental determination of the equation of state of solid hydrogen and deuterium at high pressures,” *J. Low Temp. Phys.*, vol. 34, p. 255, 1979.
- ⁶⁴ A. F. Schuch, R. L. Mills, and D. A. Depatie, “fcc-hcp-convert,” *Phys. Rev.*, vol. 165, p. 1032, 1968.
- ⁶⁵ E. Cheng and K. B. Whaley, “Pimc of pure and li doped h₂,” *J. Chem. Phys.*, vol. 104, p. 3155, 1996.
- ⁶⁶ U. Buck, F. Huisken, A. Kohlhasse, D. Otten, and J. Schaefer, “State resolved rotational excitation in D₂ + H₂ collisions,” *J. Chem. Phys.*, vol. 78, p. 4439, 1978.
- ⁶⁷ T. Zeng and P.-N. Roy, “Microscopic molecular superfluid response: theory and simulations,” *Rep. Prog. Phys.*, vol. 77, p. 046601, 2014.
- ⁶⁸ J. van Kranendonk, “Rotational and vibrational energy bands in solid hydrogen,” *Physica*, vol. 25, pp. 1080–1094, 1959.
- ⁶⁹ J. van Kranendonk and G. Karl, “Theory of the rotational and vibrational excitations in solid parahydrogen, and frequency analysis of the infrared and raman spectra,” *Rev. Mod. Phys.*, vol. 40, p. 531, 1968.
- ⁷⁰ K. Pachucki and J. Komasa, “nonadiabatic corrections to rovibrational levels of H₂,” *J. Chem. Phys.*, vol. 130, p. 164113, 2009.

- ⁷¹ T. Zeng, N. Blinov, G. Guillon, H. Li, K. P. Bishop, and P.-N. Roy, “MoRiBS-PIMC: A program to simulate molecular rotors in bosonic solvents using path-integral Monte Carlo,” *Comput. Phys. Commun.*, vol. 204, pp. 170–188, 2016.
- ⁷² H. Li, N. Blinov, P.-N. Roy, and R. J. Le Roy, “Path-integral monte carlo simulation of ν_3 vibrational shifts for CO_2 in $(\text{He})_n$ clusters critically tests the He-CO_2 potential energy surface,” *J. Chem. Phys.*, vol. 130, p. 144305, 2009.
- ⁷³ L. Wang, D. Xie, R. J. Le Roy, and P.-N. Roy, “A new four-dimensional *ab initio* potential energy surface for $\text{N}_2\text{O-He}$ and vibrational band origin shifts for the $\text{N}_2\text{O-He}_N$ clusters with $N = 1 - 40$,” *J. Chem. Phys.*, vol. 137, p. 104311, 2012.
- ⁷⁴ Y. Yan and D. Blume, “Path integral Monte Carlo ground state approach: Formalism, implementation, and applications,” *J. Phys. B*, vol. 50, p. 223001, 2017.
- ⁷⁵ L. Wang, D. Xie, H. Guo, H. Li, R. J. Le Roy, and P.-N. Roy, “Pimc-he-n2o,” *J. Mol. Spectrosc.*, vol. 267, p. 136, 2011.
- ⁷⁶ I.-C. Yeh and G. Hummer, “Pbc of md,” *J. Phys. Chem. B*, vol. 108, p. 15873, 2004.
- ⁷⁷ H. Nakano and A. Terai, “Pbc of haldane,” *J. Phys. Soc. Jpn.*, vol. 78, p. 014003, 2009.
- ⁷⁸ M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. Oxford University Press, 2 ed., 2017.
- ⁷⁹ F. D. Murnaghan, “the compressibility of media under extreme pressures,” *Proc. Nat. Acad. Sci. (US)*, vol. 30, p. 244, 1944.
- ⁸⁰ B. G. Udovidenko and V. G. Manzhelii, “Isothermal compressibility of solid parahydrogen,” *J. Low Temp. Phys.*, vol. 3, p. 429, 1970.
- ⁸¹ A. Ibrahim and P.-N. Roy, “Three-body potential energy surface for *para*-hydrogen,” *J. Chem. Phys.*, vol. 156, p. 044301, 2022.
- ⁸² P. Sindzingre, D. M. Ceperley, and M. L. Klein, “Superfluidity in clusters of *p*- H_2 molecules,” *Phys. Rev. Lett.*, vol. 67, no. 14, p. 1871, 1991.
- ⁸³ Y. Kwon and K. B. Whaley, “Nanoscale molecular superfluidity of hydrogen,” *Phys. Rev. Lett.*, vol. 89, p. 273401, 2002.
- ⁸⁴ M. Boninsegni, “Superfluid response of parahydrogen clusters in superfluid ^4He ,” *J. Low Temp. Phys.*, vol. 201, p. 193, 2020.

- ⁸⁵ E. L. Wehry and G. Mamantov, "Matrix isolation spectroscopy," *Anal. Chem.*, vol. 51, pp. 643–656, 1979.
- ⁸⁶ A. J. Barnes, *Matrix isolation spectroscopy: Technique and Applications*, vol. 21, p. 351. Berlin, Heidelberg: Springer-Verlag, 1994.
- ⁸⁷ V. E. Bondybey, A. M. Smith, and J. Agreiter, "New developments in matrix isolation spectroscopy," *Chem. Rev.*, vol. 96, pp. 2113–2134, 1996.
- ⁸⁸ T. Momose and T. Shida, "Matrix-isolation spectroscopy using solid parahydrogen as the matrix: Application to high-resolution spectroscopy, photochemistry, and cryochemistry," *Bull. Chem. Soc. Jpn.*, vol. 71, pp. 1–15, 1998.
- ⁸⁹ T. Momose, M. Fushitani, and H. Hoshina, "Chemical reactions in quantum crystals," *Int. Rev. Phys. Chem.*, vol. 24, pp. 533–552, 2005.
- ⁹⁰ K. Yoshioka, P. L. Raston, and D. T. Anderson, "Infrared spectroscopy of chemically doped solid parahydrogen," *Int. Rev. Phys. Chem.*, vol. 25, no. 3, pp. 469–496, 2006.
- ⁹¹ M. Bahou, C.-W. Huang, Y.-L. Huang, J. Glatthaar, and Y.-P. Lee, "Advances in use of p-H₂ as a novel host for matrix IR spectroscopy," *J. Chin. Chem. Soc.*, vol. 57, pp. 771–782, 2010.
- ⁹² M. E. Fajardo, S. Tam, and M. E. Derosé, "Matrix isolation spectroscopy of H₂O, D₂O, and HDO in solid parahydrogen," *J. Mol. Spectrosc.*, vol. 695-696, pp. 111–127, 2004.
- ⁹³ M. E. Fajardo and C. M. Lindsay, "Crystal field splitting of rovibrational transitions of water monomers isolated in solid parahydrogen," *J. Chem. Phys.*, vol. 128, p. 014505, 2008.
- ⁹⁴ M. Miki and T. Momose, "Rovibrational transitions and nuclear spin conversion of methane in parahydrogen crystals," *Low. Temp. Phys.*, vol. 26, pp. 661–668, 2000.
- ⁹⁵ Y. Miyamoto, M. Fushitani, D. Ando, and T. Momose, "Nuclear spin conversion of methane in solid parahydrogen," *J. Chem. Phys.*, vol. 128, p. 114502, 2008.
- ⁹⁶ M. E. Fajardo, C. M. Lindsay, and T. Momose, "Crystal field theory analysis of rovibrational spectra of carbon monoxide isolated in solid parahydrogen," *J. Chem. Phys.*, vol. 130, p. 244508, 2009.

- ⁹⁷ J. M. Farrer and Y. T. Lee, "Intermolecular potentials from crossed beam differential elastic scattering measurements. VII para-H₂ + para-H₂," *J. Chem. Phys.*, vol. 57, no. 12, p. 5492, 1972.
- ⁹⁸ E. A. Mason and W. E. Rice, "The intermolecular potentials of helium and hydrogen," *J. Chem. Phys.*, vol. 22, no. 5, p. 522, 1954.
- ⁹⁹ J. O. Hirschfelder, R. B. Bird, and E. L. Spotz, "The transport properties for non-polar gases," *J. Chem. Phys.*, vol. 16, no. 10, p. 968, 1948.
- ¹⁰⁰ M. J. Norman, R. O. Watts, and U. Buck, "A spherical potential for hydrogen from solid state and scattering data," *J. Chem. Phys.*, vol. 81, p. 3500, 1984.
- ¹⁰¹ P. Diep and J. K. Johnson, "An accurate H₂–H₂ interaction potential from first principles," *J. Chem. Phys.*, vol. 112, no. 10, p. 4465, 2000.
- ¹⁰² K. Patkowski, W. Cencek, P. Jankowski, K. Szalewicz, J. B. Mehl, G. Garberoglio, and A. H. Harvey, "Potential energy surface for interactions between two hydrogen molecules," *J. Chem. Phys.*, vol. 129, p. 094304, 2008.
- ¹⁰³ M. M. Szcześniak and G. Chałasiński, "Ab initio calculations of nonadditive effects," *J. Mol. Struct. (Theochem)*, vol. 261, pp. 37–54, 1992.
- ¹⁰⁴ P. Wind and I. Røeggen, "Ab initio calculation of three-body interaction in the (H₂)₃ trimer," *Chem. Phys.*, vol. 211, pp. 179–189, 1996.
- ¹⁰⁵ T. I. Sachse, K. T. Tang, and J. P. Toennies, "A simple damping function for the three-atom dispersion energy," *Chem. Phys. Lett.*, vol. 317, pp. 346–350, 2000.
- ¹⁰⁶ O. A. von Lilienfeld and A. Tkatchenko, "Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids," *J. Chem. Phys.*, vol. 132, p. 234109, 2010.
- ¹⁰⁷ Y. Huang and G. J. O. Beran, "Reliable prediction of three-body intermolecular interactions using dispersion-corrected second-order Möller-Plesset perturbation theory," *J. Chem. Phys.*, vol. 143, p. 044113, 2015.
- ¹⁰⁸ M. Moraldi, "Effective pair potential for solid molecular hydrogen at high pressures," *J. Low Temp. Phys.*, vol. 168, pp. 275–284, 2012.
- ¹⁰⁹ G. Garberoglio, "On the contribution of non-additive three-body interactions to the third virial coefficient of *para*-hydrogen," *Chem. Phys. Lett.*, vol. 557, p. 26, 2013.

- ¹¹⁰ W. B. J. M. Janssen and A. van der Avoird, "Dynamics and phase transitions in solid ortho and para hydrogen and deuterium from an *ab initio* potential," *Phys. Rev. B*, vol. 42, pp. 838–848, 1990.
- ¹¹¹ F.-M. Tao and Y.-K. Pan, "Møller-Plesset perturbation investigation of the He₂ potential and the role of midbond basis functions," *J. Chem. Phys.*, vol. 97, no. 7, p. 4989, 1992.
- ¹¹² T.-S. Ho and H. Rabitz, "A general method for constructing multidimensional molecular potential energy surfaces from *ab initio* calculations," *J. Chem. Phys.*, vol. 104, no. 7, p. 2584, 1996.
- ¹¹³ T.-S. Ho and H. Rabitz, "Reproducing kernel Hilbert space interpolation methods as a paradigm of high dimensional model representations: Application to multidimensional potential energy surface construction," *J. Chem. Phys.*, vol. 119, no. 13, p. 6433, 2003.
- ¹¹⁴ T.-S. Ho, T. Hollebeek, and H. Rabitz, "A global H₂O potential energy surface for the reaction O(¹D) + H₂ → OH + H," *J. Chem. Phys.*, vol. 105, p. 10472, 1996.
- ¹¹⁵ T.-S. Ho, T. Hollebeek, H. Rabitz, S. D. Chao, R. T. Skodje, A. S. Zyubin, and A. M. Mebel, "A globally smooth *ab initio* potential surface of the 1A' state for the reaction s(¹d) + H₂," *J. Chem. Phys.*, vol. 116, p. 4124, 2002.
- ¹¹⁶ J. C. Castro-Palacio, T. Nagy, R. J. Bemish, and M. Meuwly, "Computational study of collisions between O(³P) and NO(²π) at temperatures relevant to the hypersonic flight regime," *J. Chem. Phys.*, vol. 141, p. 164319, 2014.
- ¹¹⁷ O. T. Unke, J. C. Castro-Palacio, R. J. Bemish, and M. Meuwly, "Collision-induced rotational excitation in N₂⁺(²σ_g⁺, ν = 0) – Ar: Comparison of computations and experiment," *J. Chem. Phys.*, vol. 144, p. 224307, 2016.
- ¹¹⁸ X.-G. Wang and T. Carrington. Jr., "Using Lebedev grids, sine spherical harmonics, and monomer contracted basis functions to calculate being energy levels of HF trimer," *J. Theor. Comp. Chem.*, vol. 2, no. 4, pp. 599–608, 2003.
- ¹¹⁹ Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki, and M. Kállay, "An efficient linear-scaling CCSD(T) method based on local natural orbitals," *J. Chem. Phys.*, vol. 139, p. 094105, 2013.
- ¹²⁰ K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories," *Chem. Phys. Lett.*, vol. 157, no. 6, pp. 479–483, 1989.

- ¹²¹ R. J. Hinde, “Vibrational dependence of the H₂–H₂ C₆ dispersion coefficients,” *J. Chem. Phys.*, vol. 122, p. 144304, 2005.
- ¹²² F. Lekien and J. Marsden, “Tricubic interpolation in three dimensions,” *Int. J. Numer. Methods Eng.*, vol. 63, pp. 455–471, 2005.
- ¹²³ A. Ibrahim and P.-N. Roy, “Equation of state of solid parahydrogen using *ab initio* two-body and three-body interaction potentials,” *J. Chem. Phys.*, vol. 157, p. 174503, 2022.
- ¹²⁴ A. Bhandari, A. P. Rollings, L. Ratto, and J. D. Weinstein, “High-purity solid parahydrogen,” *Rev. Sci. Instr.*, vol. 92, p. 073202, 2021.
- ¹²⁵ F. M. Mutunga, K. M. Olenyik, A. I. Strom, and D. T. Anderson, “Hydrogen atom quantum diffusion in solid parahydrogen: The H + N₂O → *cis*-HNNO → *trans*-HNNO reaction,” *J. Chem. Phys.*, vol. 154, p. 014302, 2021.
- ¹²⁶ S. Sheludiakov, D. M. Lee, V. V. Khmelenko, Y. A. Dmitriev, J. Järvinen, J. Ahokas, and S. Vasiliev, “Purely spatial diffusion of H atoms in solid normal- and para-hydrogen films,” *Phys. Rev. B*, vol. 105, p. 144102, 2022.
- ¹²⁷ A. I. Strom, A. Gutiérrez-Quintanilla, M. Chevalier, J. C. C. Crépin, and D. T. Anderson, “Matrix isolation spectroscopy and nuclear spin conversion of propyne suspended in solid parahydrogen,” *J. Phys. Chem. A*, vol. 124, pp. 4471–4483, 2020.
- ¹²⁸ M. Boninsegni, “Computer simulations of supercooled liquid hydrogen mixtures and the possible crystallization slowdown,” *Phys. Rev. B*, vol. 95, p. 104518, 2017.
- ¹²⁹ J. W. Leachman, R. T. Jacobsen, S. G. Penoncello, and E. W. Lemmon, “Fundamental equations of state for parahydrogen, normal hydrogen, and orthohydrogen,” *J. Chem. Phys. Ref. Data.*, vol. 38, no. 3, pp. 721–748, 2009.
- ¹³⁰ A. L. Barnes and R. J. Hinde, “Effect of three-body interactions on the zero-temperature equation of state of HCP solid ⁴He,” *J. Chem. Phys.*, vol. 146, p. 094510, 2017.
- ¹³¹ R. A. Aziz, F. R. W. McCourt, and C. C. K. Wong, “A new determination of the ground state interatomic potential for He₂,” *Mol. Phys.*, vol. 61, pp. 1487–1511, 1987.
- ¹³² W. Cencek, K. Patkowski, and K. Szalewicz, “Full-configuration interaction calculation of three-body nonadditive contribution to helium interaction potential,” *J. Chem. Phys.*, vol. 131, p. 064105, 2009.

- ¹³³ H. Li, P.-N. Roy, and R. J. Le Roy, “Analytic Morse/long-range potential energy surfaces and predicted infrared spectra for CO₂-H₂,” *J. Chem. Phys.*, vol. 132, no. 21, p. 214309, 2010.
- ¹³⁴ L. Wang, D. Xie, R. J. Le Roy, and P.-N. Roy, “A new six-dimensional potential energy surface for H₂-N₂O and its adiabatic-hindered-rotor treatment,” *J. Chem. Phys.*, vol. 139, p. 034312, 2013.
- ¹³⁵ H. Li, X.-L. Zhang, R. J. Le Roy, and P.-N. Roy, “Analytic Morse/long-range potential energy surfaces and predicted infrared spectra for CO-H₂ dimer and frequency shifts of CO in (*para*-H₂)_N $N = 1 - 20$ clusters,” *J. Chem. Phys.*, vol. 139, p. 164315, 2013.
- ¹³⁶ T. Zeng, H. Li, R. J. Le Roy, and P.-N. Roy, ““Adiabatic-hindered-rotor” treatment of the parahydrogen-water complex,” *J. Chem. Phys.*, vol. 135, p. 094304, 2011.
- ¹³⁷ A. Marr, T. Halverson, A. Tripp, and P.-N. Roy, “Vibrational Raman shifts of spin isomer combinations of hydrogen dimers and isotopologues,” *J. Phys. Chem. A*, vol. 124, no. 34, pp. 6877–6888, 2020.
- ¹³⁸ L. D. Caballero, C. Schran, F. Briec, and D. Marx, “Neural network interaction potentials for para-hydrogen with flexible molecules,” *arXiv preprint arXiv:2206.08251*, 2022.
- ¹³⁹ T. Hollebeek, T.-S. Ho, H. Rabitz, and L. B. Harding, “Construction of reproducing kernel Hilbert space potential energy surfaces for the 1 A” and 1 A’ states of the reaction $N(^2D) + H_2$,” *J. Chem. Phys.*, vol. 114, p. 3945, 2001.
- ¹⁴⁰ H. Zhai and S. Y. Lin, “A fast hybrid method for constructing multidimensional potential energy surfaces from *ab initio* calculations: A new global analytic PES of NH₂ system,” *Chem. Phys.*, vol. 455, pp. 57–64, 2015.
- ¹⁴¹ D. Koner, J. C. S. V. Veliz, R. J. Bemish, and M. Meuwly, “Accurate reproducing kernel-based potential energy surfaces for the triplet ground states of N₂O and dynamics for the $N + NO \leftrightarrow O + N_2$ and $N_2 + O \rightarrow 2N + O$ reactions,” *Phys. Chem. Chem. Phys.*, vol. 22, pp. 18488–18498, 2020.
- ¹⁴² O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, “Machine learning force fields,” *Chem. Rev.*, vol. 121, pp. 10142–10186, 2021.

- ¹⁴³ M. Boninsegni, N. V. Prokof'ev, and B. V. Svistunov, "Worm algorithms for continuous-space path integral Monte Carlo simulations," *Phys. Rev. Lett.*, vol. 96, p. 070601, 2006.
- ¹⁴⁴ M. Boninsegni, N. V. Prokof'ev, and B. V. Svistunov, "Worm algorithm and diagrammatic Monte Carlo: A new approach to continuous-space path integral Monte Carlo simulations," *Phys. Rev. E*, vol. 74, p. 036701, 2006.
- ¹⁴⁵ P. Attard, "Pair-hypernetted-chain closure for three-body potentials: Results for argon with the Axilrod-Teller triple-dipole potential," *Phys. Rev. A*, vol. 45, no. 6, pp. 3659–3669, 1991.
- ¹⁴⁶ F. del Río, E. Díaz-Herrera, O. Guzmán, J. A. Moreno-Razo, and J. E. Ramos, "Analytical equation of state with three-body forces: Application to noble gases," *J. Chem. Phys.*, vol. 139, p. 184503, 2013.
- ¹⁴⁷ R. E. Cohen, O. Gülseren, and R. J. Hemley, "Accuracy of equation-of-state formulations," *Am. Mineral.*, vol. 85, pp. 338–344, 2000.
- ¹⁴⁸ H. Mao and R. J. Hemley, "Ultrahigh-pressure transitions in solid hydrogen," *Rev. Mod. Phys.*, vol. 66, pp. 671–692, 1994.
- ¹⁴⁹ S.-Y. Chang and M. Boninsegni, "*Ab initio* potentials and the equation of state of condensed helium at high pressure," *J. Chem. Phys.*, vol. 115, pp. 2629–2633, 2001.
- ¹⁵⁰ R. A. Aziz, V. P. S. Nain, J. S. Carley, W. L. Taylor, and G. T. McConville, "An accurate intermolecular potential for helium," *J. Chem. Phys.*, vol. 70, pp. 4330–4342, 1979.
- ¹⁵¹ L. W. Bruch and I. J. McGee, "Calculations and estimates of the ground state energy of helium trimers," *J. Chem. Phys.*, vol. 59, pp. 409–413, 1973.
- ¹⁵² A. R. Janzen and R. A. Aziz, "An accurate potential energy curve for helium based on *ab initio* calculations," *J. Chem. Phys.*, vol. 107, pp. 914–919, 1997.
- ¹⁵³ M. J. Cohen and J. N. Murrell, "An analytic function for the three-body potential of He_3 ," *Chem. Phys. Lett.*, vol. 260, pp. 371–376, 1996.
- ¹⁵⁴ C. J. Pickard and R. J. Needs, "Structure of phase III of solid hydrogen," *Nat. Phys.*, vol. 3, pp. 473–476, 2007.

- ¹⁵⁵ E. Gregoryanz, C. Ji, P. Dalladay-Simpson, B. Li, R. T. Howie, and H.-K. Mao, “Everything you always wanted to know about metallic hydrogen but were afraid to ask,” *Matter Radiat. Extremes*, vol. 5, p. 038101, 2020.
- ¹⁵⁶ N. Blinov and P.-N. Roy, “Connection between the observable and centroid structural properties of a quantum fluid: application to liquid para-hydrogen,” *J. Chem. Phys.*, vol. 120, no. 8, pp. 3759–3764, 2004.
- ¹⁵⁷ P.-N. Roy and N. Blinov, “Centroid dynamics with quantum statistics,” *Isr. J. Chem.*, vol. 42, no. 2-3, pp. 183–190, 2002.
- ¹⁵⁸ W. H. Ryu, Y. Han, and G. A. Voth, “Coarse-graining of many-body path integrals: Theory and numerical approximations,” *J. Chem. Phys.*, vol. 150, p. 244103, 2019.
- ¹⁵⁹ T. Cui, E. Cheng, B. J. Alder, and K. B. Whaley, “Rotational ordering in solid deuterium and hydrogen: A path integral Monte Carlo study,” *Phys. Rev. B*, vol. 55, p. 12253, 1997.
- ¹⁶⁰ D. Marx and M. H. Müser, “Path integral simulations of rotors: Theory and applications,” *J. Phys. Condens. Matter*, vol. 11, p. R117, 1999.
- ¹⁶¹ N. Blinov, X. G. Song, and P.-N. Roy, “Path integral Monte Carlo approach for weakly bound van der Waals complexes with rotations: Algorithm and benchmark calculations,” *J. Chem. Phys.*, vol. 120, p. 5916, 2004.
- ¹⁶² N. Blinov and P.-N. Roy, “Effect of exchange on the rotational dynamics of doped helium clusters,” *J. Low Temp. Phys.*, vol. 140, p. 253, 2005.
- ¹⁶³ A. Ibrahim and P.-N. Roy, “A neural network-based four-body potential energy surface for parahydrogen,” *J. Chem. Phys.*, vol. 160, p. 244308, 2024.
- ¹⁶⁴ G. R. Medders, A. W. Götz, M. A. Morales, P. Bajaj, and F. Paesani, “On the representation of many-body interactions in water,” *J. Chem. Phys.*, vol. 143, p. 104102, 2015.
- ¹⁶⁵ Z. Fan, W. Chen, V. Vierimaa, and A. Harju, “Efficient molecular dynamics simulations with many-body potentials on graphics processing units,” *Comput. Phys. Commun.*, vol. 218, pp. 10–16, 2017.
- ¹⁶⁶ A. O. de-la Roza, L. M. LeBlanc, and E. R. Johnson, “What is “many-body” dispersion and should I worry about it?,” *Phys. Chem. Chem. Phys.*, vol. 22, pp. 8266–8276, 2020.

- ¹⁶⁷ G. A. Cisneros, K. T. Wikfeldt, L. Ojamäe, J. Lu, Y. Xu, H. Torabifard, A. P. Bartók, G. Csányi, V. Molinero, and F. Paesani, “Modeling molecular interactions in water: From pairwise to many-body potential energy functions,” *Chem. Rev.*, vol. 116, pp. 7501–7528, 2016.
- ¹⁶⁸ A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: an imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, Curran Associates, Inc., 2019.
- ¹⁶⁹ M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattemberg, M. Wicke, Y. Yu, and X. Zheng., “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from tensorflow.org.
- ¹⁷⁰ R. J. Wheatley, G. Garberoglio, and A. H. Harvey, “Four-body nonadditive potential energy surface and the fourth virial coefficient of helium,” *J. Chem. Eng. Data.*, vol. 68, pp. 3257–3264, 2023.
- ¹⁷¹ R. S. Graham and R. J. Wheatley, “Machine learning for non-additive intermolecular potentials: Quantum chemistry to first-principles predictions,” *Chem. Commun.*, vol. 58, p. 6898, 2022.
- ¹⁷² W. L. Bade, “Drude-model calculation of dispersion forces. I. General theory,” *J. Chem. Phys.*, vol. 27, pp. 1280–1284, 1957.
- ¹⁷³ B. T. Fang, “Trilateration and extension to global positioning system navigation,” *J. Guid. Control Dyn.*, vol. 9, pp. 715–717, 1986.
- ¹⁷⁴ K. Fukushima, “Cognitron: A self-organizing multilayered neural network,” *Biol. Cybernetics*, vol. 20, pp. 121–136, 1975.
- ¹⁷⁵ K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, “SchNet: A deep learning architecture for molecules and materials,” *J. Chem. Phys.*, vol. 148, p. 241722, 2018.

- ¹⁷⁶ A. LeNail, “NN-SVG: Publication-ready neural network architecture schematics,” *J. Open Source Softw.*, vol. 4, p. 747, 2019.
- ¹⁷⁷ Y. Midzuno and T. Kihara, “Non-additive intermolecular potential in gases I. van der Waals interactions,” *J. Phys. Soc. Jpn.*, vol. 11, no. 10, pp. 1045–1049, 1956.
- ¹⁷⁸ C. P. Herrero and R. Ramirez, “Path-integral simulation of solids,” *J. Phys. Condens. Matter*, vol. 26, p. 233201, 2014.
- ¹⁷⁹ C.-L. Tian, N. Wu, F. Liu, S. K. Saxena, and X. Zheng, “Four-body interaction energy for compressed solid krypton from quantum theory,” *J. Chem. Phys.*, vol. 137, p. 044108, 2012.
- ¹⁸⁰ K. Rościszewski, B. Paulus, P. Fulde, and H. Stoll, “*Ab initio* coupled-cluster calculations for the fcc and hcp structures of rare-gas solids,” *Phys. Rev. B*, vol. 62, no. 9, p. 5482, 2000.

Appendix A

Supplementary material for Chapter 5

A.1 Trilateration

In the sample generation procedure, we sample the six side lengths of the geometry

$$(r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34})$$

and use a trilateration procedure to generate the Cartesian coordinates.

The function `SixSideLengthsToCartesian` below accepts the six side lengths of the four-body geometry, and a small number $\varepsilon > 0$ to help account for a corner case involving the position of point 3 (described later). If the six side lengths represent a valid four-body geometry, the function returns four Cartesian coordinates that generate the geometry. Otherwise, the function returns **None**.

function SIXSIDELENGTHSTOCARTESIAN($r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}, \varepsilon$)

▷ calculate the relevant x positions

$$x_2 \leftarrow r_{12}$$

$$x_3 \leftarrow (r_{12}^2 + r_{13}^2 - r_{23}^2)/2r_{12}$$

$$x_4 \leftarrow (r_{14}^2 - r_{24}^2 + r_{12}^2)/2r_{12}$$

▷ calculate the y positions, accounting for a corner case and floating-point errors

$$\tilde{y}_3^2 \leftarrow r_{13}^2 - x_3^2$$

```

▷ indication that the geometry is impossible
if  $\tilde{y}_3^2 < 0$  and  $|\tilde{y}_3^2| > \varepsilon$  then
    return None
end if

if  $\tilde{y}_3^2 > \varepsilon$  then
     $y_3 \leftarrow \sqrt{\tilde{y}_3^2}$ 
     $y_4 \leftarrow (r_{14}^2 - r_{34}^2 + r_{13}^2 - 2x_3x_4)/2y_3$ 
else
     $y_3 \leftarrow 0$ 
     $y_4 \leftarrow 0$ 
end if

▷ calculate  $z_4$ , accounting for floating-point errors
 $\tilde{z}_4^2 \leftarrow r_{14}^2 - x_4^2 - y_4^2$ 

▷ indication that the geometry is impossible
if  $\tilde{z}_4^2 < 0$  and  $|\tilde{z}_4^2| > \varepsilon$  then
    return None
end if

 $z_4 \leftarrow \sqrt{\max(0, \tilde{z}_4^2)}$ 

▷ All coordinates needed to generate the points have been calculated
 $(p_1, p_2, p_3, p_4) \leftarrow \{(0, 0, 0), (x_2, 0, 0), (x_3, y_3, 0), (x_4, y_4, z_4)\}$ 
return  $(p_1, p_2, p_3, p_4)$ 
end function

```

Using this procedure, the generated points satisfy certain properties. Point 1 is always at the origin. Point 2 lies on the positive x -axis. Point 3 lies in the region $y \geq 0$ and $z = 0$. Point 4 lies in the region $z \geq 0$.

The two conditional statements that cause the function to return **None** are those in which the six side lengths do not correspond to a valid geometry. They indicate that the six side lengths violate the triangle inequality rule in one or more ways.

In the general case where point 3 lies in the xy -plane but not on the x -axis, the procedure can assign point 4 a definite, unique position. However, if point 3 lies on the x -axis and $y_3 = 0$, the calculation for y_4 blows up. In this special case, points 1, 2, and 3 all lie on

the x -axis, and point 4 may lie anywhere on a ring of radius $\sqrt{r_{14}^2 - x_4^2}$ that lies parallel to the yz -plane and is centred at x_4 on the x -axis. We may arbitrarily select $y_4 = 0$ without modifying the relative pair distances of the system.

Calculating y_3 and z_4 each involve taking a square root, and in cases where y_3 and z_4 are very close to or essentially 0, floating-point errors can cause the arguments to the square root to be negative. The above function accounts for these cases.

A.2 Lexicographic Ordering

Lexicographic ordering is a way to compare two ordered sequences based on their elements. For readers familiar with the Python programming language, this is the kind of ordering that Python uses to compare two tuples of numbers.

When comparing two sequences, we compare elements at the same index in either sequence, from left to right. If the two elements are equal, we move on to compare the elements in the next index. As soon as we hit an index where one element is lower than the other, the sequence of the lower element is considered the lower sequence overall. In terms of pseudocode, this can be represented as the following function **LexLowest**, which takes two sequences, compares them lexicographically, and returns the smallest of the two (or either if they are equal).

```
function LEXLOWEST( $A, B$ )
   $n \leftarrow$  length of  $A$ 
  for  $i \leftarrow 1$  to  $n$  do
    if  $A[i] < B[i]$  then
      return  $A$ 
    else if  $A[i] > B[i]$  then
      return  $B$ 
    end if
  end for
  return  $A$                                 ►  $A$  and  $B$  are equal, return either
end function
```

For example, suppose we have input samples

$$\begin{aligned} s_0 &= (2, 3, 4, 5, 6, 7) \\ s_1 &= (2, 3, 4, 7, 1, 9) \\ s_2 &= (1, 9, 11, 25, 13, 15). \end{aligned}$$

When comparing these samples, we would find that $\text{LexLowest}(s_0, s_1) = s_0$, because of the comparison of 5 and 7 in the fourth position, $\text{LexLowest}(s_0, s_2) = s_2$, because of the comparison of 2 and 1 in the first position, and $\text{LexLowest}(s_1, s_2) = s_2$, again because of the comparison of 2 and 1 in the first position.

A.3 Basis Set Superposition Error Fit Constants

In the Discussion and Analysis section of the main text, in the subsection titled “Basis Set and Method Comparison”, we take the AVDZ, AVTZ, and AVQZ energies for tetrahedra and extrapolate them to infinite basis size using the the exponential decay function

$$E_e(N) = E_e^{(\infty)} + c_e \exp\{-\alpha N\} \quad (\text{A.1})$$

and the power decay function

$$E_p(N) = E_p^{(\infty)} + c_p N^{-\beta} . \quad (\text{A.2})$$

For the tetrahedron of side length 2.2 Å, the fit constants are $E_e^{(\infty)} = 192.668 \text{ cm}^{-1}$, $c_e = 45.341 \text{ cm}^{-1}$, $\alpha = 0.9707$, $E_p^{(\infty)} = 191.311 \text{ cm}^{-1}$, $c_p = 26.999 \text{ cm}^{-1}$, and $\beta = 1.780$. For the tetrahedron of side length 2.95 Å, the fit constants are $E_e^{(\infty)} = 2.0976 \text{ cm}^{-1}$, $c_e = 7.954 \text{ cm}^{-1}$, $\alpha = 1.729$, $E_p^{(\infty)} = 2.0875 \text{ cm}^{-1}$, $c_p = 3.775 \text{ cm}^{-1}$, and $\beta = 3.858$.

A.4 Estimates of Bade Coefficients

In Fig. A.1, we plot transformed versions of the Bade potential alongside the *ab initio* CCSD(T) AVDZ and AVTZ interaction energies for a tetrahedron as a function of the tetrahedron side length r . For the transformations, we multiply the interaction energies by r^{12} , and rescale all the curves by the same constant such that the transformed Bade potential is equal to -1 . The figure makes it easier to see that the *ab initio* energies have the expected r^{-12} decay rate at large intermolecular separations. It is from where these rescaled curves level out that we estimate the interaction coefficients B_{12} derived from the AVDZ and AVTZ *ab initio* energies.

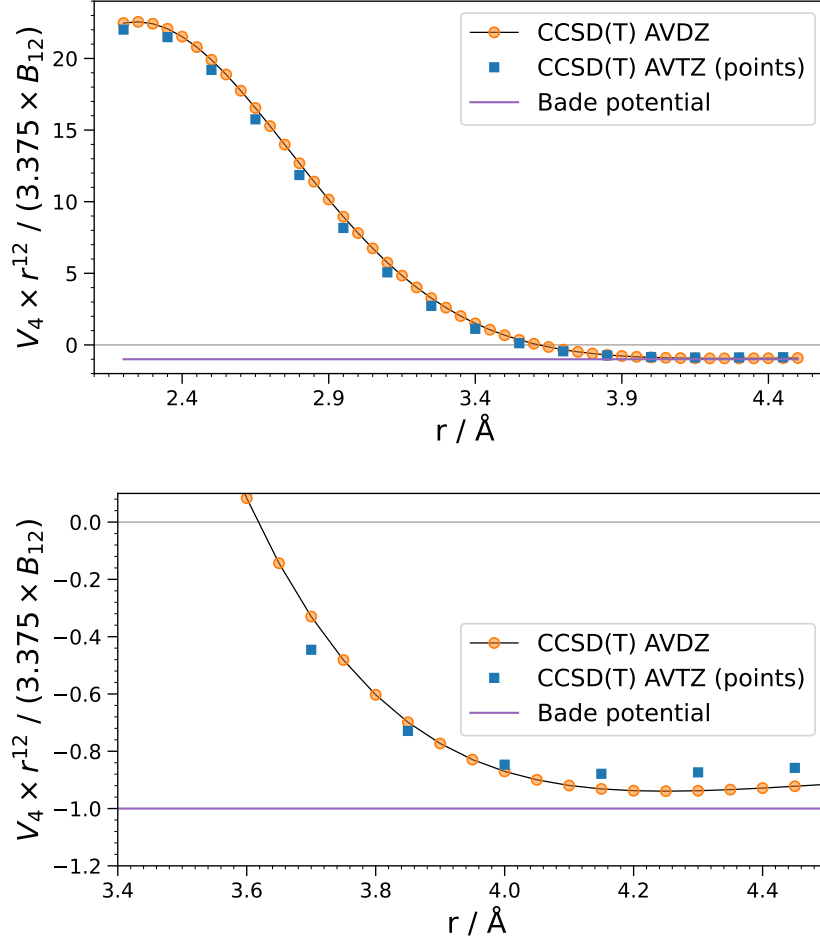


Figure A.1: The rescaled four-body interaction energy for a tetrahedron geometry, as a function of the tetrahedron's side length r . Shown are the CCSD(T) energies calculated using the AVDZ basis (orange circles, black line), and the AVTZ basis (blue squares), as well as the Bade potential (purple line). The term in the y -axis label, B_{12} , is the Bade coefficient as estimated using the Midzuno-Kihara-inspired approximation. The top subfigure shows the results for the entire range of energies calculated for the tetrahedron, while the bottom subfigure emphasizes the region where the long-range r^{-12} decay begins.

A.5 Training Error Curves for Smaller Models

The main text contains the training error of the neural network as a function of epoch, for the 64-128-128-64 model. Here we show the training curves for the 32-64-64-32 model (Fig. A.2), the 16-32-32-16 model (Fig. A.3), and the 8-16-16-8 model (Fig. A.4).

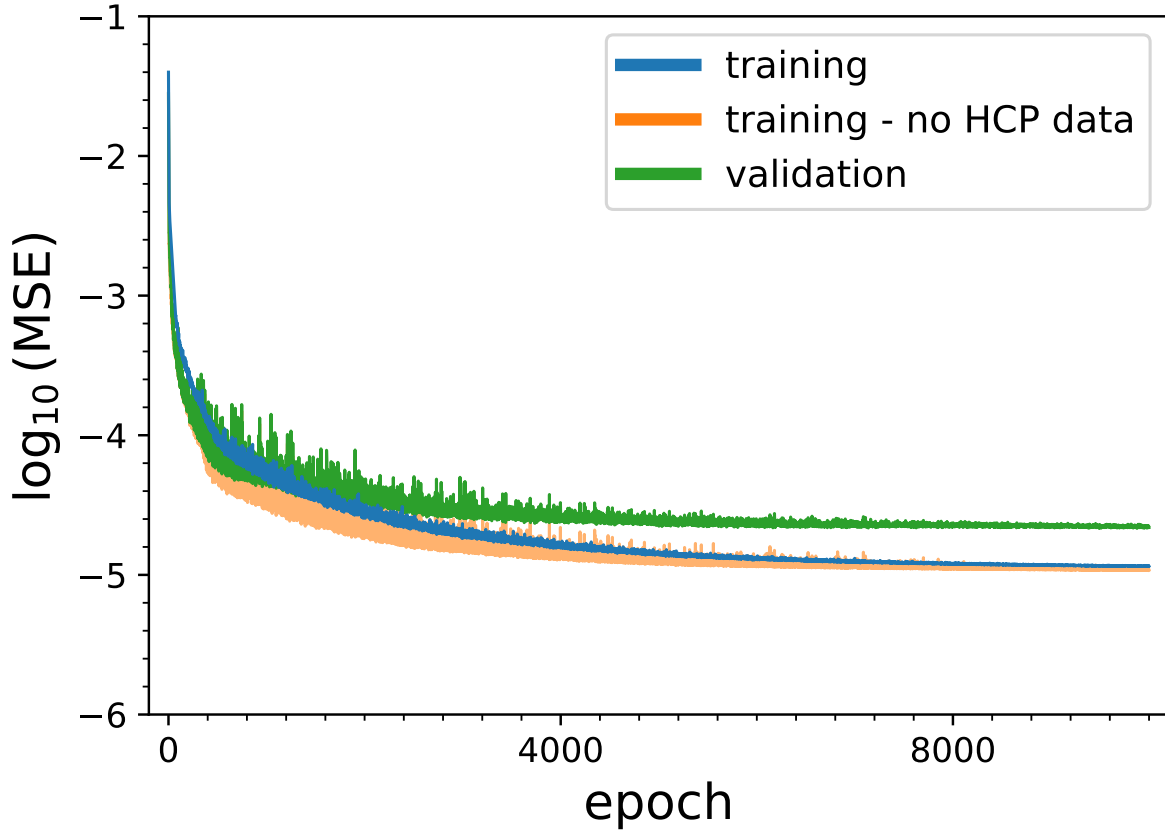


Figure A.2: The base-10 logarithm of the mean squared error loss as a function of the epoch during the training of the 32-64-64-32 model. The plots are shown for the training data set (blue, solid), the validation data set (green, solid), and the training data set with the 1610 samples from the *hcp* lattice removed (orange, semi-transparent).

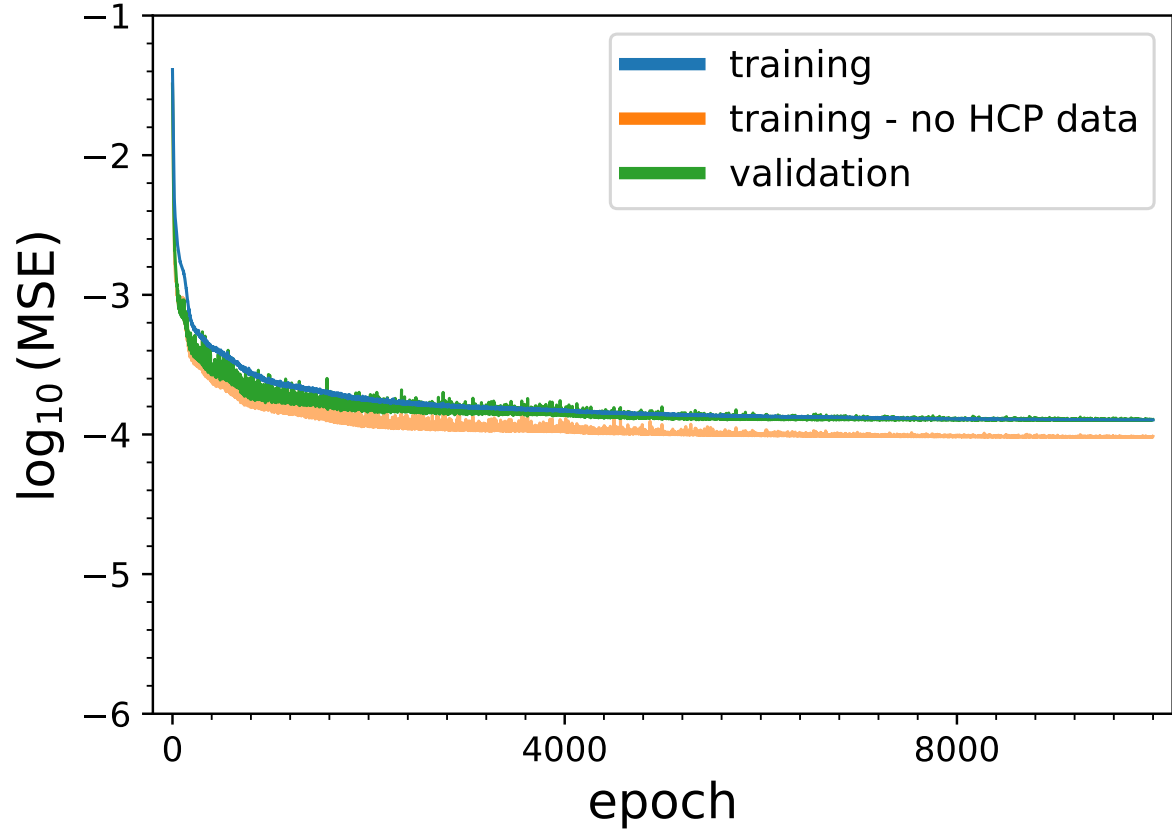


Figure A.3: The base-10 logarithm of the mean squared error loss as a function of the epoch during the training of the 16-32-32-16 model. The plots are shown for the training data set (blue, solid), the validation data set (green, solid), and the training data set with the 1610 samples from the *hcp* lattice removed (orange, semi-translucent).

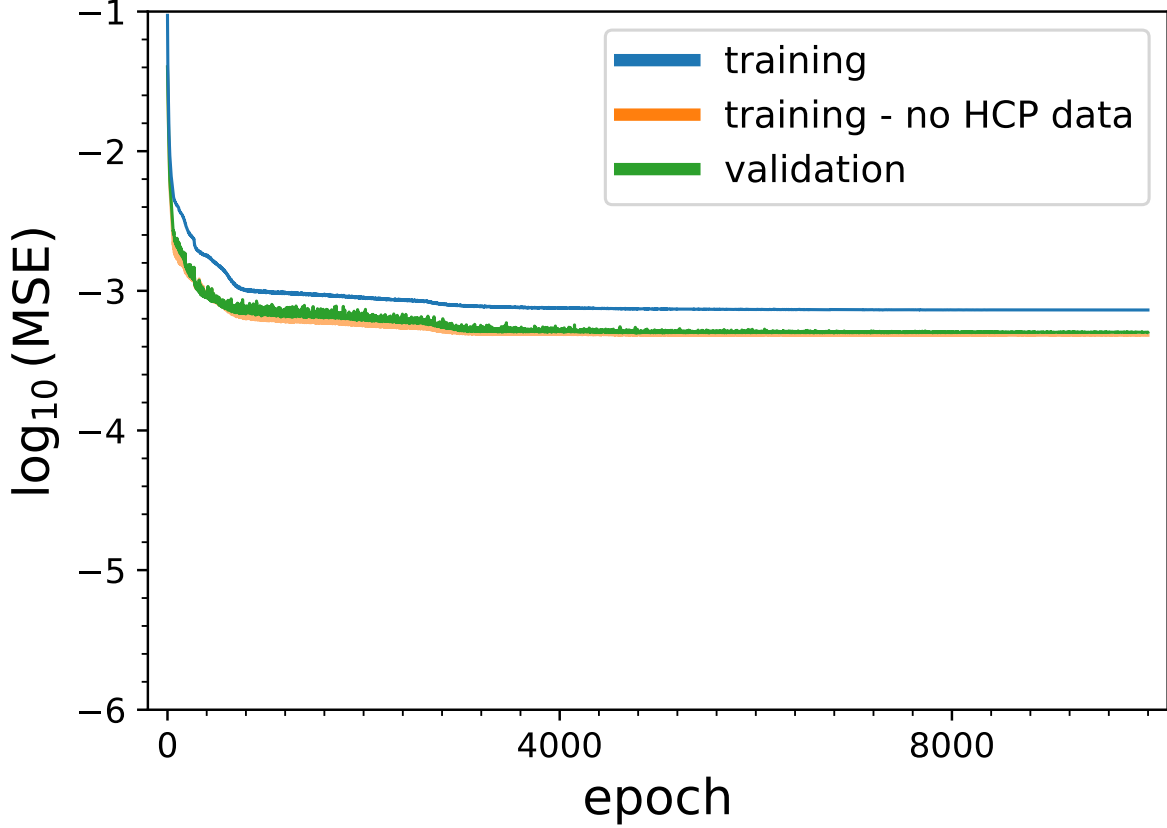


Figure A.4: The base-10 logarithm of the mean squared error loss as a function of the epoch during the training of the 8-16-16-8 model. The plots are shown for the training data set (blue, solid), the validation data set (green, solid), and the training data set with the 1610 samples from the *hcp* lattice removed (orange, semi-transparent).

A.6 Short Range: Exponential Decay Fit

For four-body geometries with at least one side length shorter than $2.2 \text{ \AA}$, we use an extrapolation function that is a linear combination of an exponential decay fit and a linear fit. For well-behaved cases, the extrapolation function is entirely the exponential decay fit. For cases where the exponential decay fit would behave poorly or blow up, the linear fit becomes more prominent.

Suppose we have a geometry g , made up of the six side lengths $g = \{r_{ij}\}$, where at least one of the side lengths is less than $2.25 \text{ \AA}$. Define $\Delta R = 0.01 \text{ \AA}$ and $\tilde{R}_U = 2.25 \text{ \AA}$, where the subscript is chosen in the context of a later section in the supplementary material. Let R_0 be the shortest side length in the input. We calculate the scaling factors $s_0 = \tilde{R}_U/R_0$ and $s_1 = (\tilde{R}_U + \Delta R)/R_0$, and define two new geometries $g_0 = \{s_0 r_{ij}\}$ and $g_1 = \{s_1 r_{ij}\}$.

We now calculate the exponential coefficient

$$c = \frac{1}{\Delta R} \ln \left| \frac{V_4(g_1)}{V_4(g_0)} \right|. \quad (\text{A.3})$$

The energy for the original geometry g is now calculated using the exponential decay fit

$$V_4^{(\text{ex})}(g) = V_4(g_0) \exp\{-c(\tilde{R}_U - R_0)\}. \quad (\text{A.4})$$

However, there are cases where c is very large, and Eq. (A.4) blows up very quickly. On closer inspection, this typically only happens in cases where $V_4(g_0)$ is very close to 0, which means a small absolute change in energy results in a large relative change in energy. To prevent this, we smoothly transition from the exponential fit to a less physically realistic, but more numerically stable linear fit.

The full equation is

$$V_4(g) = \omega(c; c', c'') V_4^{(\text{ex})}(g) + [1 - \omega(c; c', c'')] V_4^{(\text{li})}(g) \quad (\text{A.5})$$

where $V_4^{(\text{li})}(g)$ is a linear extrapolation below $r = r_0$, $c' = 6$, $c'' = 8$, and $\omega(x; a, b)$ is given by

$$\omega(x; a, b) = \begin{cases} 0 & \text{if } x \leq a, \\ \frac{1}{2}[1 - \cos(\pi k)] & \text{if } a < x < b, \\ 1 & \text{if } x \geq b \end{cases} \quad (\text{A.6})$$

where $k = (x - a)/(b - a)$. If c is below c' , the fit is entirely exponential; if c is above c'' , the fit is entirely linear; and if c is between them, it is a finite linear combination of the two. The constants c' and c'' are chosen by hand after investigating certain cases where the exponential decay fit behaved poorly.

The linear fit purposely sacrifices accuracy for stability. However, because the calculation of the exponential coefficient only blows up when the energies involved are weak, the error from this linear adjustment is generally insignificant as long as the energies aren't extrapolated to extremely short distances.

The decision to set $\tilde{R}_U$ to $2.25 \text{ \AA}$ rather than $2.2 \text{ \AA}$ is done in anticipation of the use of a transition function described in the next section.

A.7 Transitions between different parts of the PES

Description

The PES, using the MLP trained with the shifted softplus activation function, is continuously differentiable.

Transition between *ab initio* and dispersive parts

Conceptually, the PES is split into an “*ab initio* part,” a “dispersive part,” and a “mixed (*ab initio*/dispersive) part.” The *ab initio* part is made of the multilayer perceptron (MLP) and the exponential decay parts, described in further detail later. The dispersive part is made of the Bade potential. The mixed (*ab initio*/dispersive) part is made of a linear combination of the *ab initio* and dispersive parts.

Each four-body geometry falls into one of these three categories based on its average side length $\tilde{r}$. Define a “lower” distance $\tilde{r}_L$ and an “upper” distance $\tilde{r}_U$, where $\tilde{r}_L < \tilde{r}_U$. The part of the PES that a sample falls into is described by

$$\text{part of PES} = \begin{cases} \textit{ab initio} & \text{if } \tilde{r} \leq \tilde{r}_L, \\ \text{mixed } (\textit{ab initio}/\text{dispersive}) & \text{if } \tilde{r}_L < \tilde{r} < \tilde{r}_U, \\ \text{dispersive} & \text{if } \tilde{r} \geq \tilde{r}_U \end{cases} \quad (\text{A.7})$$

In the mixed (*ab initio*/dispersive) part, the energy of a sample is a linear combination of the *ab initio* and dispersive parts. The fraction of the output that comes from the dispersive part is given by

$$\omega(\tilde{r}; \tilde{r}_L, \tilde{r}_U) = \begin{cases} 0 & \text{if } \tilde{r} \leq \tilde{r}_L, \\ \frac{1}{2}[1 - \cos(\pi k)] & \text{if } \tilde{r}_L < \tilde{r} < \tilde{r}_U, \\ 1 & \text{if } \tilde{r} \geq \tilde{r}_U \end{cases} \quad (\text{A.8})$$

where $k = (\tilde{r} - \tilde{r}_L)/(\tilde{r}_U - \tilde{r}_L)$. In the current PES, $\tilde{r}_L = 4 \text{ \AA}$ and $\tilde{r}_U = 4.5 \text{ \AA}$.

The decision to make the transition between the *ab initio* and dispersive parts depend on the average side length, and the choice of $\tilde{r}_L$ and $\tilde{r}_U$, were based on observations of the *hcp* lattice geometries. By looking at the energies of geometries in the *hcp* lattice as a function of the lattice constant, it appeared that when their average side length was around $4 \text{ \AA}$, the energy was sufficiently converged to the Bade potential.

Transition within the *ab initio* part

The *ab initio* part is further split into a part where the energies are determined directly by the MLP (the “mid-range part”), by the exponential decay extrapolation (the “short-range part”), and by a linear combination of the two (the “mixed (short/mid)-range part”).

The exponential decay fit is described earlier in the supplementary material. With no mixed (short/mid)-range part to transition between the two, there is a small discontinuity between the mid-range and the short-range parts. Depending on how small we make the “scaling step size” (ΔR) for the exponential decay fit, we can make the discontinuity very small. For example, setting $\Delta R = 0.01 \text{ \AA}$ introduces a discontinuity at the short-range/mid-range transition on the order of 0.1% for geometries with large energies (such as the tetrahedron, and other “compact” geometries), and on the order of 1 – 3% for geometries with much smaller energies (on the order of 1 cm^{-1}).

For situations where only the energies are needed, this discontinuity is likely unimportant. For situations where the PES must be continuously differentiable, we can remove this discontinuity by using the same transition function and strategy we used for the transition between the *ab initio* and dispersive regions.

A four-body geometry with an average side length less than $\tilde{r}_U$ falls into either the short-range, mid-range, or mixed (short/mid)-range parts, based on its shortest side length $\tilde{R}$. Define two distances $\tilde{R}_L$ and $\tilde{R}_U$, where $\tilde{R}_L < \tilde{R}_U$. The part of the *ab initio* part that a sample falls into is described by

$$\text{part of the } ab \text{ initio PES} = \begin{cases} \text{short} & \text{if } \tilde{R} \leq \tilde{R}_L, \\ \text{mixed (short/mid)} & \text{if } \tilde{R}_L < \tilde{R} < \tilde{R}_U, \\ \text{mid} & \text{if } \tilde{R} \geq \tilde{R}_U \end{cases} \quad (\text{A.9})$$

In the mixed (short/mid)-range part, the energy of a sample is a linear combination of the short-range and mid-range parts. The fraction of the output that comes from the mid-range part is given by

$$\omega(\tilde{R}; \tilde{R}_L, \tilde{R}_U) = \begin{cases} 0 & \text{if } \tilde{R} \leq \tilde{R}_L, \\ \frac{1}{2}[1 - \cos(\pi k)] & \text{if } \tilde{R}_L < \tilde{R} < \tilde{R}_U, \\ 1 & \text{if } \tilde{R} \geq \tilde{R}_U \end{cases} \quad (\text{A.10})$$

where $k = (\tilde{R} - \tilde{R}_L)/(\tilde{R}_U - \tilde{R}_L)$. In the current PES, $\tilde{R}_L = 2.2 \text{ \AA}$ and $\tilde{r}_U = 2.25 \text{ \AA}$.

The energies from the short-range and mid-range parts of the PES are already very close to one another in the mixed (short/mid)-range region, and so the transition is fairly clean.

Limitations

Certain kinds of samples are not treated well with this mixing of, and transition between, *ab initio* and dispersive energies. For example, consider a geometry where molecules 1 and 2 are less than 2.2 Å apart, molecules 3 and 4 are far away from them (molecules 3 and 4 may be close to one another as well). If the average side length is between $\tilde{r}_L$ and $\tilde{r}_U$, the PES calculates a physically unrealistic mix of a short-range exponential wall and a long-range dispersive decay. If the average side length is greater than $\tilde{r}_U$, the PES categorizes this sample to be entirely dispersive. No samples in the training data exist for such geometries, and the energy produced by the PES for both cases is physically unrealistic. However, based on extrapolations of certain geometries in the *hcp* lattice training data, the four-body energies of such extreme samples are very small (typically on the order of 10^{-3} cm^{-1} to 1 cm^{-1}), and the *ab initio* total two-body energy is very large (typically on the order of 10^3 cm^{-1} to 10^5 cm^{-1}). Moreover, although physically unrealistic, the transition is still continuously differentiable.

The short-range exponential decay fit is based on the observation that, for small geometries, the energy of a four-body geometry varies exponentially as all its side lengths are scaled by the same factor. The exponential coefficient for the fit is determined by *ab initio* data that only goes down to side lengths of 2.25 Å. The short-range energy predictions made by this PES for four-body geometries with side lengths much less than 2.2 Å are unlikely to be accurate. In cases where the Boltzmann weight of such small four-body geometries is non-negligible, such as in solids at very high densities, such a limitation might become significant.

A.8 Contributions to the *hcp* lattice for each geometry

We supplement the training data using samples that are based specifically on the geometries of the *hcp* lattice. The procedure for generating the samples is as follows. First, we select a reference molecule in the *hcp* lattice. Then we consider all four-body geometries involving this reference molecule such that (1) they have at least one side length equal to the lattice constant, and (2) they have no side length more than twice the lattice constant. We find that 83 different four-body geometries satisfy these conditions. Information about all 83 geometries is provided in Table A.1.

Table A.1: Information about particular four-body geometries in the hcp lattice, their combinatorial frequency, and their contributions to the average four-body interaction potential energy per particle. The six side lengths in each geometry are expressed in the permutation with the lowest lexicographic order. The geometries are then ordered among themselves using their average side length, with the lexicographic order between their side lengths used as a tie-breaker. The first column is an ID number assigned to the geometry, for identification. The second column contains N_c , the number of times that geometry appears under the aforementioned conditions for which they were selected. Columns 3 to 8 are the ratio of the six side lengths to the lattice constant. Column 9 is the total contribution of the geometry to the average four-body interaction potential energy per particle (units of cm^{-1}), when the lattice constant is $2.2 \text{ \AA}$.

ID	N_c	r_{12}	r_{13}	r_{14}	r_{23}	r_{24}	r_{34}	$V_4^{(\text{tot})}$
0	8	1	1	1	1	1	1	398.3489
1	48	1	1	1	1	1	$\sqrt{2}$	1767.6561
2	12	1	1	1	1	1	$\sqrt{8/3}$	510.3478
3	36	1	1	1	1	1	$\sqrt{3}$	1673.2181
4	12	1	1	$\sqrt{2}$	$\sqrt{2}$	1	1	777.0374
5	144	1	1	1	1	$\sqrt{2}$	$\sqrt{3}$	938.2455
6	48	1	1	1	1	$\sqrt{2}$	$\sqrt{11/3}$	333.7082
7	24	1	1	1	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{8/3}$	26.8497
8	72	1	1	1	1	$\sqrt{3}$	$\sqrt{3}$	-72.8998
9	48	1	1	1	1	$\sqrt{8/3}$	$\sqrt{11/3}$	-34.8335
10	48	1	1	1	1	$\sqrt{3}$	$\sqrt{11/3}$	-112.8734
11	96	1	1	1	1	$\sqrt{3}$	2	-263.0247
12	24	1	1	$\sqrt{2}$	$\sqrt{2}$	1	$\sqrt{11/3}$	23.2044
13	24	1	1	1	$\sqrt{2}$	$\sqrt{2}$	2	31.7028
14	24	1	1	1	1	$\sqrt{11/3}$	$\sqrt{11/3}$	-107.4177
15	48	1	1	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{3}$	3.9874
16	144	1	1	$\sqrt{2}$	$\sqrt{3}$	1	$\sqrt{3}$	589.1818
17	48	1	1	$\sqrt{2}$	$\sqrt{8/3}$	1	$\sqrt{11/3}$	47.4894
18	48	1	1	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{11/3}$	-0.1676
19	48	1	1	$\sqrt{2}$	1	$\sqrt{3}$	$\sqrt{11/3}$	17.5318
20	24	1	1	1	$\sqrt{8/3}$	$\sqrt{3}$	$\sqrt{3}$	8.1791

Table A.1: Continued from previous page...

ID	N_c	r_{12}	r_{13}	r_{14}	r_{23}	r_{24}	r_{34}	$V_4^{(\text{tot})}$
21	12	1	1	$\sqrt{3}$	$\sqrt{8/3}$	$\sqrt{2}$	$\sqrt{2}$	49.4868
22	8	1	1	1	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	3.3276
23	24	1	1	$\sqrt{3}$	1	$\sqrt{3}$	$\sqrt{3}$	2.5747
24	24	1	1	$\sqrt{2}$	1	$\sqrt{11/3}$	$\sqrt{11/3}$	1.9198
25	48	1	1	$\sqrt{2}$	$\sqrt{11/3}$	1	$\sqrt{11/3}$	78.6803
26	48	1	1	$\sqrt{8/3}$	$\sqrt{3}$	1	$\sqrt{11/3}$	147.4124
27	48	1	1	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{3}$	5.5033
28	12	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{2}$	1	56.9016
29	48	1	1	$\sqrt{11/3}$	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{8/3}$	74.2656
30	48	1	1	$\sqrt{8/3}$	1	$\sqrt{11/3}$	$\sqrt{11/3}$	0.8395
31	48	1	1	$\sqrt{3}$	1	$\sqrt{3}$	2	1.9873
32	48	1	1	$\sqrt{3}$	$\sqrt{3}$	1	2	130.0325
33	96	1	1	$\sqrt{3}$	$\sqrt{2}$	$\sqrt{2}$	2	19.3653
34	96	1	1	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{2}$	$\sqrt{3}$	115.7225
35	48	1	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{11/3}$	2.2902
36	48	1	1	$\sqrt{3}$	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{11/3}$	2.2277
37	48	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{3}$	1	77.3423
38	24	1	1	$\sqrt{3}$	$\sqrt{8/3}$	$\sqrt{3}$	$\sqrt{3}$	6.0805
39	24	1	1	$\sqrt{2}$	2	$\sqrt{3}$	$\sqrt{3}$	5.6200
40	48	1	1	2	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{3}$	11.4372
41	48	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{3}$	2	1	19.5027
42	24	1	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{8/3}$	1.7976
43	48	1	1	$\sqrt{2}$	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{11/3}$	4.0230
44	48	1	1	$\sqrt{11/3}$	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{11/3}$	3.2812
45	48	1	1	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{2}$	$\sqrt{11/3}$	31.6711
46	48	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	1	34.7712
47	48	1	1	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{8/3}$	$\sqrt{3}$	25.8993
48	12	1	$\sqrt{8/3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	$\sqrt{8/3}$	1	15.2368
49	24	1	1	2	$\sqrt{8/3}$	$\sqrt{3}$	$\sqrt{3}$	8.0599
50	48	1	$\sqrt{2}$	$\sqrt{8/3}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{2}$	5.6212

Table A.1: Continued from previous page...

ID	N_c	r_{12}	r_{13}	r_{14}	r_{23}	r_{24}	r_{34}	$V_4^{(\text{tot})}$
51	24	1	1	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{3}$	8.5996
52	48	1	1	$\sqrt{8/3}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	1.9287
53	96	1	1	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	2	7.0128
54	24	1	1	2	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	8.9254
55	48	1	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	2	1	24.0374
56	24	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{2}$	$\sqrt{11/3}$	2.1125
57	48	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{2}$	$\sqrt{3}$	12.3522
58	48	1	$\sqrt{2}$	2	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{2}$	13.9028
59	48	1	1	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{11/3}$	7.6466
60	48	1	1	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{11/3}$	7.0239
61	24	1	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	1	7.5321
62	48	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	2.3053
63	24	1	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{2}$	2.4616
64	48	1	$\sqrt{8/3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	1	20.9388
65	24	1	$\sqrt{3}$	$\sqrt{11/3}$	2	$\sqrt{3}$	1	16.0475
66	48	1	$\sqrt{2}$	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{8/3}$	$\sqrt{3}$	4.6222
67	24	1	1	$\sqrt{8/3}$	2	$\sqrt{11/3}$	$\sqrt{11/3}$	1.9854
68	12	1	$\sqrt{3}$	2	2	$\sqrt{3}$	1	8.9282
69	48	1	$\sqrt{2}$	2	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{2}$	16.0005
70	48	1	1	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	2.8300
71	48	1	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{2}$	4.0850
72	12	1	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{8/3}$	0.3714
73	48	1	1	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	2	2.6144
74	24	1	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{11/3}$	$\sqrt{11/3}$	2	0.0388
75	48	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{11/3}$	1.0870
76	48	1	$\sqrt{2}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	$\sqrt{3}$	1.1112
77	24	1	1	$\sqrt{3}$	2	2	2	1.2307
78	48	1	$\sqrt{8/3}$	$\sqrt{3}$	$\sqrt{11/3}$	$\sqrt{3}$	$\sqrt{3}$	1.0489
79	48	1	$\sqrt{3}$	$\sqrt{11/3}$	2	$\sqrt{3}$	$\sqrt{3}$	0.8478

Table A.1: Continued from previous page...

ID	N_c	r_{12}	r_{13}	r_{14}	r_{23}	r_{24}	r_{34}	$V_4^{(\text{tot})}$
80	24	1	$\sqrt{11/3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	$\sqrt{11/3}$	$\sqrt{3}$	0.0701
81	24	1	$\sqrt{3}$	$\sqrt{3}$	2	2	2	-0.0329
82	96	1	$\sqrt{3}$	$\sqrt{11/3}$	2	$\sqrt{11/3}$	$\sqrt{11/3}$	0.0262

DATA AVAILABILITY

The data that support the findings of the study presented in Chapter 5 are openly available at https://github.com/AlexanderIbrahim1/nn_fourbody_potential/. The raw inputs and outputs involved in performing the electronic structure calculations for the *ab initio* energies are available at DOI <https://zenodo.org/doi/10.5281/zenodo.11272855>.