November 2020

Development of Next-Generation, Fast, Accurate, Transferable, and Polarizable Force-fields for Heterogenous Material Simulations

Adam E. Hogan
University of South Florida

Follow this and additional works at: https://scholarcommons.usf.edu/etd

Part of the Accounting Commons, and the Chemistry Commons

Scholar Commons Citation

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Development of Next-Generation, Fast, Accurate, Transferable, and Polarizable
Force-fields for Heterogenous Material Simulations

by

Adam E. Hogan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Chemistry
College of Arts and Sciences
University of South Florida

Major Professor: Brian Space, Ph.D.
Tom Keyes, Ph.D.
Arjan van der Vaart, Ph.D.
Shengqian Ma, Ph.D.
H. Lee Woodcock, Ph.D.

Date of Approval:
November 12, 2020

Keywords: Monte Carlo, molecular dynamics, porous materials, gas sorption, metal-organic frameworks

Copyright © 2020, Adam E. Hogan
DEDICATION

In loving memory of Jonita Ann Hogan (1959-2006).
I would like to thank my family: Bill, Billy, and Eve Hogan for their continual love and support. I am eternally grateful to my mentor and research advisor Dr. Brian Space, without whom I would have never pursued graduate school. I would also like to thank everyone that has made the Space group what it is today: Dr. Keith McLaughlin, Dr. Chris Cioce, Dr. Douglas Franz, Dr. Tony Pham, Dr. Katherine Forrest, Meagan Mulcair, Matthew Mostrom, Brant Tudor, Shanelle Suepaul, Luciano Laratelli, Logan Ritter, Andres Saez, Alexis Johnson, Alejandro Navas, Zac Dyott, and Mak Djulbegovic. I am grateful to have known every one of you and count you as some of my best friends. Special thanks to Andres Saez for work done in collaboration investigating the veracity of various methods to calculate dispersion coefficients. The advice and assistance given by my committee, Dr. Tom Keyes, Dr. Arjan van der Vaart, Dr. Shengqian Ma, and Dr. H. Lee Woodcock, has been invaluable. Finally, I would like to thank Dr. Jon Belof for many enlightening scientific conversations.
# TABLE OF CONTENTS

List of Tables ........................................................................................................ i

List of Figures ........................................................................................................ iii

Abstract .................................................................................................................. v

Chapter One: Introduction ...................................................................................... 1
  1.1 Symmetry Adapted Perturbation Theory ......................................................... 4
  1.2 Many-Body Polarization & Many-Body van der Waals ............................... 10
  1.3 Force-field Parameterization .................................................................... 12

Chapter Two: Development of PHAHST Small Hydrocarbon Models ............. 16
  2.1 Introduction .............................................................................................. 16
  2.2 Methods .................................................................................................... 18
  2.3 Results and Discussion ........................................................................... 19
  2.4 Conclusions ............................................................................................. 22

Chapter Three: Development of PHAHST for the Noble Gases, N₂, and H₂ .... 24
  3.1 Introduction .............................................................................................. 24
  3.2 Methods .................................................................................................... 27
  3.3 Results and Discussion ........................................................................... 30
  3.4 Conclusions ............................................................................................. 37

Chapter Four: Development of PHAHST Models for Material Simulations .... 38
  4.1 Introduction .............................................................................................. 38
  4.2 Methods .................................................................................................... 42
  4.3 Results and Discussion ........................................................................... 45
  4.4 Conclusions ............................................................................................. 55

Chapter Five: Conclusions and Future Directions ............................................. 58

References ............................................................................................................. 60
LIST OF TABLES

Table 2.1  Fit parameters for the newly developed PHAHST C2 hydrocarbon models................................................................. 20

Table 2.2  Fit parameters for the newly developed PHAST* C2 hydrocarbon models........................................................................ 20

Table 2.3  Fit parameters for the newly developed PHAST C2 hydrocarbon models................................................................. 21

Table 4.1  Fit parameters and associated RMSE for each fitting process for the newly developed PHAHST material force-field.............. 46

Table 4.2  Binding energies from positions found in Figure 4.5 calculated with periodic DFT, the newly proposed PHAHST force-field, and a generic Lennard-Jones based force-field .............................................. 54

Table 4.3  Comparison of simulated and experimental isosteric heat of adsorption (Qst) ............................................................... 55
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>SAPT2 exchange energies for the N₂ dimer plotted on a log scale</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>Mean unsigned error between the PHAHST model and <em>ab initio</em> for each possible trimer</td>
<td>21</td>
</tr>
<tr>
<td>2.2</td>
<td>Mean signed error between the PHAHST model and <em>ab initio</em> for each possible trimer</td>
<td>22</td>
</tr>
<tr>
<td>3.1</td>
<td>Mixed dispersion coefficients for the noble gas pairs calculated via <em>ab initio</em> and geometrically mixed</td>
<td>30</td>
</tr>
<tr>
<td>3.2</td>
<td>Born-Oppenheimer energy surface for Argon computed via CCSDT(Q) and the resulting PHAHST model</td>
<td>31</td>
</tr>
<tr>
<td>3.3</td>
<td>Born-Oppenheimer energy surface for H₂ computed via CCSD(T) and the resulting PHAHST model</td>
<td>32</td>
</tr>
<tr>
<td>3.4</td>
<td>Born-Oppenheimer energy surface for N₂ computed via CCSD(T) and the resulting PHAHST model</td>
<td>33</td>
</tr>
<tr>
<td>3.5</td>
<td>Calculated and experimental second virial coefficients for H₂, N₂, He, Ne, Ar, Kr, and Xe</td>
<td>34</td>
</tr>
<tr>
<td>3.6</td>
<td>Relative density deviation, ( (\rho_{\text{calc}} - \rho) / \rho ), of calculated bulk pressure-density isotherms for H₂, N₂, He, Ne, Ar, Kr, and Xe at 298 K (400 K for Xe)</td>
<td>36</td>
</tr>
<tr>
<td>4.1</td>
<td>HKUST-1 fragments used in the material fitting process</td>
<td>44</td>
</tr>
<tr>
<td>4.2</td>
<td>Quality of the resulting PHAHST material force-field fit to reproduce DFT/VV10 interaction energies between small molecules and HKUST-1 fragments</td>
<td>47</td>
</tr>
<tr>
<td>4.3</td>
<td>Calculated and experimental isosteric heat of adsorption for H₂, N₂, Ne, Ar, Kr, and Xe in HKUST-1</td>
<td>50</td>
</tr>
<tr>
<td>4.4</td>
<td>Calculated radial distribution functions for Ne, Ar, Kr, and Xe with the center of the tetrahedral pores during an intermediate</td>
<td></td>
</tr>
</tbody>
</table>
loading (approximately 4 sorbates per tetrahedral pore) Grand Canonical simulation .......................................................... 51

Figure 4.5 Small molecule binding sites located via simulated annealing with the PHAHST force-field and MPMC......................................................... 52
ABSTRACT

Theoretical modeling is extremely useful in guiding to experiment; however quantitatively accurate modeling of energy-relevant small molecule sorption at the heterogeneous interfaces present in metal-organic materials (MOMs) is currently challenging. MOMs are an emerging class of materials consisting of inorganic clusters and organic linkers that offer great potential in the areas of gas storage, gas separation, and catalysis due to the possibility of large surface areas, complex heterogeneous surfaces, and rational designability. Efficient chemical separations involving these materials could reduce the US’s total energy consumption by approximately 10 to 15%. In this dissertation, the parameterization of small molecules and metal clusters with existing potentials as well as the development of a novel next-generation potential with associated small molecule and material parameters are discussed. The resulting force-fields, denoted PHAHST, represent a significant advance in the field and can quantitatively predict/retrodict binding sites and energetics at complex heterogeneous interfaces. This is accomplished via a physically grounded approach to all potential forms, mixing rules, and parameters as well as including explicit polarization in all environments. The resulting force-fields are extensible and a procedure for developing compatible models is established. All parameters and code associated with this work are freely available online at https://github.com/mpmccode/mpmc.
CHAPTER 1

INTRODUCTION

Quantitively accurate molecular modeling at the nanoscale is important for providing insights into the nature of intermolecular interactions in a wide variety of fields, including molecular biology, materials research, the energy/petroleum industry, and as inputs in multiscale modeling efforts. However, the parameters and potentials currently in widespread use suffer from many deficiencies, mainly related to compromises made early on in the history of molecular simulation due to the lack of fast multiprocessor computers, lack of commercially available programmable GPUs, and inefficient algorithms for solving the time-consuming equations present in both quantum mechanical and classical simulations.\textsuperscript{1-5} Improvement of any individual component of the current generation of force-fields is impossible without addressing the inherent error calculation present in many force-fields. This work and the resulting force-field, denoted PHAHST for Potentials with High Accuracy, High Speed, and Transferability, reexamines the underlying potentials and parameters in an accurate, fast, and, most importantly, transferable manner by physically grounded choices for all potential forms, mixing rules, parameters, and including explicit polarization in every environment.

The new PHAHST force-fields are applied to energy-relevant small molecules and metal-organic materials (MOMs) due to the pressing need to address climate
change, either by reducing demand for fossil fuels or by capturing CO$_2$ from flue gas or directly from the atmosphere. Metal-organic materials (MOMs) are a broad class of compounds containing inorganic clusters and organic ligands and comprise both discrete compounds, such as metal-organic polyhedra, as well as infinitely extended 1, 2, and 3-dimensional structures, such as the 3-dimensional metal-organic framework.$^6$ Permanent porosity, large surface areas, and complex heterogeneous interfaces are present in many MOMs, making them suitable candidates for gas storage, gas separation, and catalysis applications. Further, their parameters may be finely tuned by adjusting the composition of either the organic linkers or the inorganic nodes, creating materials tailored for a specific purpose.$^7$

For instance, chemical separations currently account for approximately half of the United States’ industrial energy consumption and approximately 10-15% of the United States’ total energy consumption.$^8$ Moving away from energy-intensive separation methods, e.g. cryogenic distillations in the case of separating acetylene from ethylene, can result in large (~90%) energy savings and highly-selective MOMs are one possible route to economically feasible membrane separation processes. Similarly, fossil fuel dependence may also be alleviated by switching to a “Hydrogen Economy,” using hydrogen as the primary means of energy storage and transport.$^9$ Traditional hydrogen storage methods have various deficiencies, however, including expensive cooling systems and extreme working conditions, and MOMs have been proposed as a potential solution. Alternatively, CO$_2$ may be captured from post-combustion flue gas or directly from the atmosphere using energy-efficient physisorption in MOMs.$^{10,11}$
Molecularly detailed simulations are extremely important in providing insights into the mechanisms behind gas sorption in MOMs, allowing the inherent tunability of MOMs to be exploited for specific applications once this has been uncovered. Molecular mechanics, including both molecular dynamics and Monte Carlo methods, is based upon using classical mechanics to model molecular systems. The potential energy and, in the case of molecular dynamics, the forces are assumed to be entirely dependent on the atomic nuclear coordinates, implicitly invoking the Born-Oppenheimer approximation. The collection of both the underlying potentials and associated parameters comprise a force-field.

Historically force-fields have been fit to reproduce bulk phenomena such as bulk liquid densities, diffusion coefficients, or vapor-liquid equilibria while being informed by experimental spectroscopic results such as the experimentally measured bond lengths or quadrupole moment. This approach may be adequate for bulk simulations however it is entirely inadequate in heterogeneous simulations due to the fitting process selecting only the subset of configurations that appear in the bulk. For example, the most important bulk configurations of CO2 include the “T” and “slip-parallel” configurations due to its large quadrupole, leaving the “end-on-end” configuration poorly defined; however, this configuration is important in open-metal sorption due to the large partial negative charge on the oxygen atom. Further, most simple water models have permanent partial charges that produce a dipole of ~3 Debye, effectively incorporating polarization into the charges, while the gas phase dipole moment of a single water molecule is 1.85 Debye. While this may reproduce properties in rough agreement with bulk water
such a model will unable to reproduce the proper energetics in the highly
anisotropic environment present in heterogeneous materials.

More recently force-fields fit to reproduce \textit{ab initio} have come into their own
with the simultaneous advancement of computing power and efficient electronic
structure programs.\textsuperscript{2,5} Accurately calculated bond lengths, partial charges,
polarizabilities, bonded parameters, and even dispersion coefficients are now within
the grasp of even a desktop computer. It is now routine to pair coupled-cluster
methods and large basis sets on standard supercomputing nodes. This approach
will be taken throughout the rest of this work.

In the rest of this introduction, the form of intermolecular interactions will be
discussed in the framework of symmetry adapted perturbation theory. Next,
practical implementation of these potentials, mainly focused on many-body
interactions, in existing molecular simulation codes will be discussed. An overview
of fitting force-field parameters is then given. Finally, different electronic structure
methods are reviewed with a focus on accurate intermolecular interaction energies.

\textbf{1.1 Symmetry Adapted Perturbation Theory}
Symmetry adapted perturbation theory (SAPT) directly probes intermolecular
interactions and allows the assignment of physically meaningful labels to different
parts of the potential. SAPT may be used as an \textit{ab initio} method to calculate
interaction energies as well as a tool to better inform the functional dependence of
the potentials underlying classical force-fields. Perturbation theory starts with the
assumption of two atoms or molecules, \( A \) and \( B \), initially in their ground state, \( 0_A \)
and \( 0_B \), with a Hamiltonian, \( H^0 \), consisting solely of intramolecular forces. The
combined system is then expanded in powers of perturbing potential, \( H' \).
\[ H = H^0 + H' \]  

(1.1)

In the case of interacting atoms or molecules, the perturbing potential is the coulomb interaction between the opposing electrons and atomic nuclei.

\[ H' = k \frac{q_1 q_2}{r_{12}} \]  

(1.2)

The first-order correction to the energy, \( E^{(1)} \), is simply the expectation value of the perturbing potential with the ground states of both molecules. This is immediately identified as the permeant electrostatic interaction.

\[ E^{(1)} = \langle 0_A 0_B | H' | 0_A 0_B \rangle \]  

(1.3)

The second-order correction to the energy, \( E^{(2)} \), contains expectation values with a sum over states. \( E^{(2)} \) is usually divided into two different contributions, induction (or polarization) and dispersion.

\[ E^{(2)} = - \sum_{m \neq 0} \frac{|\langle 0_A 0_B | H' | m_A n_B \rangle|^2}{E_m^A + E_n^B - E_0^A - E_0^B} \]  

(1.4)

The induction energy for atom \( A \), \( E_{\text{ind}}^A \), contains the subset of the sum over states involving the excited states of \( A \) and the ground state for \( B \) and vice versa for the induction energy for atom \( B \).

\[ E_{\text{ind}}^A = - \sum_{m \neq 0} \frac{|\langle 0_A 0_B | H' | m_A 0_B \rangle|^2}{E_m^A - E_0^A} \]  

(1.5)

\[ E_{\text{ind}}^B = - \sum_{n \neq 0} \frac{|\langle 0_A 0_B | H' | 0_A n_B \rangle|^2}{E_n^B - E_0^B} \]

Classically induction can be viewed as the permanent electrostatics on one molecule inducing the opposing molecule partially into an excited state. Induction is always an attractive interaction unlike the electrostatic energy however it may also be strongly anisotropic, emphasizing the importance of its inclusion in heterogeneous molecular simulations. Later in this section the induction energy in
the dipole approximation will be presented and used throughout the rest of this work. The dispersion energy, $E_{\text{disp}}$, is likewise always attractive and contains the subset of the sum over states solely involving excited states.

$$E_{\text{disp}} = - \sum_{m \neq 0} \sum_{n \neq 0} \frac{|\langle 0_A 0_B | H' | m_A n_B \rangle|^2}{E_m^n + E_n^B - E_0^A - E_0^B} \quad (1.6)$$

The dispersion energy has no classical analog, being a purely quantum mechanical effect arising from electronic correlation. It is useful to expand the total electrostatic potential, $\nu$, as a series of multipole moments interacting with spatial derivatives of the electrostatic field, $V$, at a point (or series of points) to more fully understand the functional form and parameterization of dispersion interactions.

$$\nu = q^A V^A + \mu_\alpha^A V^A + \frac{1}{3} \Theta_{\alpha \beta}^A V^A + \ldots \quad (1.7)$$

In general, the multipole interaction tensor, $T$, is given by spatial derivatives of coulomb’s law.

$$T_{\alpha\beta - \nu} \equiv \frac{1}{4\pi \varepsilon_0} \nabla_\alpha \nabla_\beta \ldots \nabla_\nu \frac{1}{R} \quad (1.8)$$

An explicit definition of the dipole-dipole interaction tensor, $T_{\alpha\beta}$, is given as it will be used in both the derivation of the dipole-dipole dispersion energy and the explicit polarization energy.

$$4\pi \varepsilon_0 T_{\alpha\beta} = \frac{\partial^2}{\partial R_\alpha \partial R_\beta} \frac{1}{R} = \frac{3 \hat{R}_\alpha \hat{R}_\beta - \delta_{\alpha\beta}}{R^3} \quad (1.9)$$

Contraction of $T_{\alpha\beta}^{AB}$ with the dipole moment of atom $B$ results in the electric field $V^A_\alpha$ on atom $A$. Contraction of the dipole moment of atom $A$ with the electric field, $V^A_\alpha$, results in the dipole-dipole interaction energy. Expansion and factorization of terms results in,
Identifying the middle term as the dipole polarizability, $\alpha$, allows a large simplification in the polarization energy.

$$\alpha_{\alpha\beta}^A \equiv \sum_{n \neq 0} \frac{(0_A | \mu_{\alpha}^A | n_A) (n_A | \mu_{\beta}^A | 0_A)}{E_n^A - E_0^A}$$

The physical picture is an induced dipole, $\Delta \mu^B$, being created on atom $B$ by the electric field on atom $B$, $V^B$, scaled by the dipole polarizability on atom $B$, $\alpha^B$.

$$\Delta \mu^B_\gamma \equiv \alpha_{\beta\gamma}^B V^B_\gamma$$

The polarization is finally given by the dot product of the induced dipole with the permanent electric field resulting from the dipole on atom $A$.

$$E_{ind}^{B dip-dip} = -\frac{1}{2} \Delta \mu^B_\gamma V^B_\gamma$$

Such a clean factorization of the dispersion energy is not possible due to the double sum over states on both atoms.

$$E_{disp}^{dip-dip} = -\sum_{m \neq 0} \sum_{n \neq 0} \frac{(0_A | \mu_{\alpha}^A | m_A m_B) (m_A n_B | \mu_{\beta}^A | 0_A)}{E_m^A - E_0^A + E_n^B - E_0^B}$$

Making use of an integral identity,

$$\frac{1}{A + B} = \frac{2}{\pi} \int_0^\infty \frac{AB}{(A^2 + u^2)(B^2 + u^2)} du$$

, and the previous definition of the dipole polarizability $\alpha$ results in a formally exact expression for the dipole-dipole dispersion energy.
Each dipole-dipole interaction tensor, $T_{\alpha\beta}$, contributes a $1/R^3$ to the functional dependence on the energy, resulting in an overall $1/R^6$ dependence. For convenience in molecular simulations dispersion coefficients are defined as follows,

$$E_{\text{disp}}^{dip-dip} = -\frac{\hbar}{2\pi} T_{\alpha\beta} T_{\gamma\delta} \int_0^\infty \alpha^A_{\alpha\gamma}(iu) \alpha^B_{\beta\delta}(iu) \, du$$  \hfill (1.16)

Higher-order multipole moments contribute an additional $1/R$ dependence so that, e.g., the functional dependence of both the quadrupole-quadrupole dispersion energy and the dipole-octupole dispersion energy is $1/R^8$. Note that the odd-numbered dispersion coefficients, e.g. $C_7$ resulting from the dipole-quadrupole energy, are zero in isotropic atoms and molecules due to rotational symmetry. A slight modification to the functional form is needed at short distances otherwise the dispersion energy would tend towards negative infinity. This doesn’t occur in nature and would also be a problem in, e.g., grand canonical Monte Carlo simulations. Tang and Toennies found the form of the dispersion damping function for two hydrogen atoms is an incomplete gamma function.$^{17}$

$$E_{\text{disp}} = -f_6(bR) \frac{C_6}{R^6} - f_8(bR) \frac{C_8}{R^8} - f_{10}(bR) \frac{C_{10}}{R^{10}} - \ldots$$ \hfill (1.18)

$$f_{2n}(bR) = 1 - \left( \sum_{k=0}^{2n} \frac{(bR)^k}{k!} \right) e^{-bR}$$

The exponential range parameter, $b$, is usually assumed to be the same range parameter as the exponential repulsion because they both arise from the
same source, overlap of the wavefunction, so there is no need to fit an additional parameter for the damping functions.

Accounting for the antisymmetry of the wavefunction upon exchange of electrons produces an additional exchange term associated with each level of correction. In the SAPT framework, the exchange energy constitutes most of the molecular repulsive energy due to the Pauli exclusion principle arising from exchange symmetry. Plotting the exchange energy on a log scale demonstrates the strongly exponential nature of intermolecular repulsions, see Figure 1.1. Another way to think about repulsion is considering that the decrease of electron density in the region of molecular overlap. This descreening of the nuclei is proportional to the overlap integrals, which are approximately exponential.¹⁸

![Figure 1.1: SAPT2 exchange energies for the N₂ dimer plotted on a log scale.](image-url)
1.2 Many-Body Polarization & Many-Body van der Waals

In order to properly simulate in polar, heterogeneous materials including explicit polarization in classical simulations are essential to capturing the underlying interactions. Following the physical interpretation from SAPT, $\mu^A$ is the induced dipole on atom $A$, $\alpha^A$ is the static dipole polarizability, $V^A$ is the permanent electrostatic field, and $V'^A$ is the induced electrostatic field resulting from interactions with induced dipoles on all other atoms.

$$\mu^A = \alpha^A(V^A + V'^A) \quad (1.19)$$

Expanding the induced field in terms of dipole-dipole interaction tensor, $T$, results in:

$$\mu^A = \alpha^A(V^A + \sum_{A \neq B} T_{\beta\gamma} \mu^B) \quad (1.20)$$

Defining a new matrix, $A$, results in a supermatrix equation solvable with standard linear algebra libraries, the polarization energy is then calculated from equation (1.13). In single-threaded applications, Gauss-Seidel iterations coupled with a final Palmo-Krimm correction can provide extremely accurate energies, while Preconditioned conjugate gradient methods provide a route amenable to multithreaded applications and GPU acceleration while maintaining good convergence criteria.\textsuperscript{19-21} Thole-type exponential damping, effectively modifying $T$, is employed to prevent infinite polarization energies at short distances.\textsuperscript{22} Long-range corrections in the electric field can either be handled by the Wolf or Smooth Particle Mesh Ewald methods.\textsuperscript{23}
Many-body van der Waals models may be naturally constructed from the same framework by replacing the static dipole polarizability with the frequency dependent Drude polarizability and observing that van der Waals interaction are present in the absence of an external field.

$$\alpha \rightarrow \frac{\alpha \omega_0^2}{\omega_0^2 - \omega^2}$$  \hspace{1cm} (1.22)

$$\vec{V} = 0$$

One then obtains an eigenvalue equation for the 3N normal mode frequencies containing the same A supermatrix that may be solved with conventional linear algebra methods.

$$A\vec{u} = \frac{\omega^2}{\omega_0^2 \alpha} \vec{u}$$  \hspace{1cm} (1.23)

The energy is then obtained via the normal finite temperature expression for the energy of a harmonic oscillator, $$\left(\frac{h \omega}{2}\right) \coth \left(\frac{h \omega}{2kT}\right)$$. A Cholesky decomposition may be used to obtain a symmetric A supermatrix in the case of a heterogenous simulation. Exploiting the properties of the trace, the numerical scaling may be further reduced from $$N^3$$ to $$N^2$$ by a binomial expansion of the square root of A.  \(^24\)
1.3 Force-field Parameterization

After appropriate potential forms have been chosen the next step is to parameterize small molecule sorbates and material force-fields. There are a few objectives to keep in mind going forward; the entire point of these force-fields is to model complex heterogeneous interfaces so transferability is paramount. This is mainly achieved by staying as close to physical theory as possible and the use of appropriate mixing rules as well as including explicit polarization in every environment. Grand canonical simulations typically require millions of potential energy evaluations so computational tractability is also a concern. However, explicit polarization, which is a many-body potential, constitutes the vast majority (>90%) of computational time so small changes to the pair potentials won’t contribute much to the overall computation time. Additionally, arbitrary pair potentials may be tabulated and interpolated with constant scaling. 

In order to facilitate widespread adoption there is also an incentivized to keep the potential form as simple as possible while still retaining the correct physics. Finally, accuracy is also a paramount concern – this is accomplished by fitting to the highest level \textit{ab initio} calculations possible, in many cases going beyond what is considered the “gold standard” of chemistry, i.e. coupled-cluster with single, doubles, and perturbative triples with a tz/qz extrapolation of the correlation energy.

Generally, all parameterization involves minimizing a loss function. Typically, the loss function is the sum of the squares of the residuals, this is also known as L2 loss. L2 loss heavily punishes outliers and may potentially distort parameterization in cases where the fitting data is potentially unreliable. An alternative loss function is the sum of the absolute values of the residuals, also known as L1 loss. L1 loss
may have trouble converging the parameter set near the minimum as it is linear in the residuals. A compromise, known as smooth L1 loss and presented below where \( r \) is the residual, is suitable for most applications as it is quadratic in the residuals near zero and linear at larger residuals.

\[
\text{Smooth } L1 = \sum \begin{cases} 
0.5 r^2 & |r| < 1 \\
|r| - 0.5 & |r| \geq 1
\end{cases}
\] (1.24)

The level of theory used to fit the models determines the overall level of accuracy of the resultant force-fields. From both an accuracy and computational cost perspective the hierarchy of \textit{ab initio} methods is coupled cluster > MP2 \( \approx \) SAPT > DFT+D > Hartree-Fock. Coupled cluster methods are held as the “gold standard” of quantum chemistry, with other methods being benchmarked against it however they typically scale as \( N^7 \). Next, MP2 and SAPT, both being based on perturbation theory, have approximately the same accuracy for intermolecular calculations and scale slightly better at \( N^4 \sim N^5 \). DFT and Hartree-Fock both scale as \( N^3 \), however DFT is almost always more accurate. Including dispersion corrections in DFT is of paramount concern as DFT energies are mostly based on the local value or gradient of the density.

Two methods for calculating dispersion coefficients from \textit{ab initio} are used in this work and a brief overview is provided next. The first method is the Williams-Stone-Misquitta method and is based on an exact relationship between the dispersion coefficients and polarizabilities at imaginary frequencies.

First, atom-based polarizabilities at 10 different frequencies are fit to reproduce the total molecular polarizability tensor as well as many thousands of point-to-point polarizabilities between points chosen randomly on the van der Waal's surface.
Then atom-based dispersion coefficients are obtained by numerically integrating these polarizabilities. This process is entirely ab initio and parameter-free.

The second method to calculate dispersion coefficients comes from the framework of XDM. XDM assumes that dispersion interactions are related to multipole moments of interacting exchange holes. Dispersion coefficients are then obtained via expectation values of these exchange holes over an atom and the polarizability of the atom. Typically, XDM gets the polarizability of an atom by scaling the free atom’s polarizability by the free atom’s volume. This usually results in poor polarizabilities for metal atoms, where there is a large difference between the polarizability of different formal charges, so this method has a caveat as well as being based on an approximation.

Mixing rules allow the calculation of unlike parameters from like parameters the interaction potential between two unlike atoms or molecules from the two interaction potentials of the like atoms or molecules. Mixing rules reduce the total number of parameters needed for a molecular mechanics simulation from \( O(N^2) \) to \( O(N) \).

For systems where quantum effects may be important, such as low-temperature \( \text{H}_2 \) simulations, semi-classical corrections such as the Feynman-Hibbs correction may be employed. The Feynman-Hibbs correction includes higher spatial derivatives of the energy, effectively a local approximation of the nuclei’s density becoming nonlocal due to quantum effects.

\[
E_{FH} = E + \frac{\beta \hbar^2}{24\mu} \left( E'''' + \frac{2}{r} E' \right) + \frac{\beta^2 \hbar^2}{1152\mu^2} \left( \frac{15}{r^3} E' + \frac{4}{r^4} E'' + E''' \right)
\]  

(1.25)

The rest of this work is organized as follows. In chapter 2 a series of small hydrocarbon models are developed. Chapter 3 describes the development of noble
gas, H$_2$, and N$_2$ models. Chapter 4 describes the development of a force-field for the metal-organic framework HKUST-1 and application of previously developed small molecule models to heterogenous sorption simulations. Finally, chapter 5 contains a brief overview and some concluding thoughts.
CHAPTER 2

DEVELOPMENT OF PHAHST SMALL HYDROCARBON MODELS

2.1 Introduction

There is a pressing need for realistic modeling at the microscale that requires improvements in force-fields and atomistic modeling via molecular dynamics and Monte Carlo simulations. Molecularly detailed simulations can model diverse processes from bulk fluid properties and separations to sorption in porous materials. This level of chemically useful description is possible if the force-fields are carefully constructed with predictive and explanatory value. Here, a systematic approach to developing an accurate and transferable potential energy function is applied to produce atomistically detailed models for the C2 hydrocarbons. The intermolecular force-field is constructed with the plan to model complex molecular mixtures in diverse environments including high pressures and temperatures, at interfaces, and in complex heterogeneous environments. This builds on our efforts to provide a next-generation framework for precision modeling of sorption in porous materials.\textsuperscript{16,20,27–30}

The C2 hydrocarbons (ethane, ethylene, and acetylene) are among the most common chemicals used in industrial applications. Ethane is the second most common chemical, behind methane, in natural gas and ethylene, a primary reagent in the production of many plastics, is currently the most produced organic
compound in the world with global production has been estimated at 158 million metric tons.\textsuperscript{31} It is mainly produced from petroleum by thermal cracking or catalytic cracking over zeolites. Ethylene is used for a wide variety of purposes, chief among them is the polymerization reaction to polyethylene and the synthesis of the precursors to polystyrene and polyvinyl chloride. Additionally, it is a very important plant hormone that regulates the ripening of fruit, among other biological processes. Acetylene is a delirious byproduct of the cracking process. If left in the ethylene stream it can poison the polymerization catalysts downstream. Above a partial pressure of 15 psi acetylene will even explosively decompose into hydrogen and carbon. It is therefore essential to remove all traces of acetylene from the ethylene stream obtained by cracking.

Thus, it is sought to construct a robust model of two carbon hydrocarbons as a first in a series of PHAHST (potentials with high accuracy, high speed, and transferability) that will include a suite of sorbent molecules, with a focus on energy-relevant species, and a prescription for creating porous material compatible force-fields. Here, two different families of models are pursued for ethylene, differing in the choice of repulsion/dispersion potential. The PHAST models, including those previously published\textsuperscript{16,32}, are based on the Lennard-Jones equation while a new family, introduced as PHAHST herein, are based on the first three asymptotically correct dispersion coefficients, damped with an incomplete gamma function, as proposed by Tang and Toennies (1984) and exponential repulsion.\textsuperscript{17} Additionally models excluding explicit polarization were fit within the PHAST family of models to preserve compatibility with existing force-fields as well as being able to serve as a basis for comparison in future studies, and in that case use of an
asterisk (*) indicates the inclusion of explicit polarization. The total potential energy, \( U \), is taken as a sum of the permanent electrostatic interactions, \( U_{el} \), many-body polarization as detailed in the following section, \( U_{pol} \), and the repulsion/dispersion potential, \( U_{rd} \).

### 2.2 Methods

The C2 hydrocarbons are approximated as rigid molecules. Note that it is trivial to relax this condition and introduce bonding parameters but this approach is not taken here.\(^{33-35}\) The equilibrium geometry used in all the models were obtained from a geometry optimization at the CCSD(T)/aug-cc-pvqz level of theory with the electronic structure program ORCA.\(^{36,37}\) Partial charges, located on atom-centered sites, were obtained using the CHELPG method of fitting the electrostatic potential using the orbital optimized CCSD/aug-cc-pvqz electronic density as implemented in ORCA.\(^{38,39}\) Atom-centered static dipole polarizabilities were obtained with the default procedure implemented by the CamCASP program. Briefly, the molecular orbitals of a gas phase monomer were calculated at the PBE0/aug-cc-pvqz level of theory with the CS00 asymptotic correction as implemented in NWChem.\(^{37,40,41}\) The total molecular polarizability and point-to-point polarizabilities between 2000 points randomly located on the 2x-4x van der Waal surface are then calculated with the CKS propagator. These molecular polarizabilities are then transformed to atom-centered point polarizabilities using the Williams–Stone–Misquitta methodology.\(^{2,25,26}\) Additionally, frequency-dependent polarizabilities were fit using this same procedure at 10 imaginary frequencies and the three leading atom-centered asymptotically correct dispersion coefficients, i.e. C6, C8, and C10, were calculated via the Casimir-Polder formula and Gauss-Legendre quadrature.\(^{42,43}\) For
instance, the formula for the leading asymptotically correction dispersion coefficient, C6, is given by the following, where $\alpha_i$ is the dipole polarizability of atom $i$:

$$C6 = \frac{3\hbar}{(4\pi\varepsilon_0)^2} \int_0^\infty \alpha^A_{\beta\gamma}(iu)\alpha^B_{\beta\delta}(iu) du$$ \hspace{1cm} (2.1)

The remaining repulsion/dispersion parameters (either Lennard-Jones or exponential repulsion) were fit to reproduce ab initio single point energies of 2,000 randomly generated trimer configurations containing all possible combinations of C2 hydrocarbons. The energies were computed at the DLPNO-CCSD(T) level of theory utilizing a aug-cc-pvdz/aug-cc-pvtz extrapolation and basis set superposition error corrected by the counterpoise method with the quantum chemistry program ORCA. Simulated annealing was then used to optimize atom-centered Lennard-Jones parameters, keeping all other previously calculated parameters constant, with the Monte Carlo simulation package MPMC.

2.3 Results and Discussion

The resulting parameter sets are shown in Table 2.1, Table 2.2, and Table 2.3. Mean unsigned and mean signed errors, see Figure 2.1 and Figure 2.2, were initially evaluated for all possible trimer combinations. Previous generation PHAST* and PHAST models, where an asterisk (*) indicates the inclusion of explicit polarization, are systematically too repulsive across all trimers. New generation PHAHST models have a lower overall unsigned error compared to previous generation models as well as having no systematic bias in the signed error. At the time these models were developed compatible PHAHST material force fields, presented in Chapter 5 of this work, were not yet available; however the PHAST*
and PHAST models have achieved success in predicting binding sites and binding energies across a large number of metal-organic materials.  

<table>
<thead>
<tr>
<th>Site</th>
<th>Molecule</th>
<th>$a$ (Å$^3$)</th>
<th>C6 (a.u.)</th>
<th>C8 (a.u.)</th>
<th>C10 (a.u.)</th>
<th>$\beta$ (Å$^{-1}$)</th>
<th>$\rho$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>$C_2H_6$</td>
<td>0.69670</td>
<td>13.53161</td>
<td>301.2267</td>
<td>21554.47</td>
<td>4.00769</td>
<td>3.17248</td>
</tr>
<tr>
<td>H</td>
<td>$C_2H_6$</td>
<td>0.47580</td>
<td>3.89683</td>
<td>0.0</td>
<td>0.0</td>
<td>3.20240</td>
<td>1.97467</td>
</tr>
<tr>
<td>C</td>
<td>$C_2H_4$</td>
<td>1.63040</td>
<td>40.50952</td>
<td>1325.417</td>
<td>56051.68</td>
<td>3.03860</td>
<td>3.47290</td>
</tr>
<tr>
<td>H</td>
<td>$C_2H_4$</td>
<td>0.19000</td>
<td>1.17702</td>
<td>0.0</td>
<td>0.0</td>
<td>4.77867</td>
<td>1.74237</td>
</tr>
<tr>
<td>C</td>
<td>$C_2H_2$</td>
<td>1.5514</td>
<td>39.44099</td>
<td>1123.91</td>
<td>43535.55</td>
<td>3.06161</td>
<td>3.40728</td>
</tr>
<tr>
<td>H</td>
<td>$C_2H_2$</td>
<td>0.1448</td>
<td>0.76016</td>
<td>0.0</td>
<td>0.0</td>
<td>4.71283</td>
<td>1.71998</td>
</tr>
</tbody>
</table>

Table 2.2: Fit parameters for the newly developed PHAST* C2 hydrocarbon models.

<table>
<thead>
<tr>
<th>Site</th>
<th>Molecule</th>
<th>$a$ (Å$^3$)</th>
<th>$\varepsilon$ (K/K_b)</th>
<th>$\sigma$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>$C_2H_6$</td>
<td>0.69670</td>
<td>130.67671</td>
<td>3.32434</td>
</tr>
<tr>
<td>H</td>
<td>$C_2H_6$</td>
<td>0.47580</td>
<td>0.36053</td>
<td>2.92205</td>
</tr>
<tr>
<td>C</td>
<td>$C_2H_4$</td>
<td>1.63040</td>
<td>40.56623</td>
<td>3.53935</td>
</tr>
<tr>
<td>H</td>
<td>$C_2H_4$</td>
<td>0.19000</td>
<td>11.89036</td>
<td>2.10037</td>
</tr>
<tr>
<td>C</td>
<td>$C_2H_2$</td>
<td>1.5514</td>
<td>40.49720</td>
<td>3.45111</td>
</tr>
<tr>
<td>H</td>
<td>$C_2H_2$</td>
<td>0.1448</td>
<td>8.06410</td>
<td>2.12597</td>
</tr>
</tbody>
</table>
Table 2.3: Fit parameters for the newly developed PHAST C2 hydrocarbon models.

<table>
<thead>
<tr>
<th>Site</th>
<th>Molecule</th>
<th>$\alpha$ (Å$^3$)</th>
<th>$\varepsilon$ (K/$K_B$)</th>
<th>$\sigma$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>C$_2$H$_6$</td>
<td>0.69670</td>
<td>161.83770</td>
<td>3.26722</td>
</tr>
<tr>
<td>H</td>
<td>C$_2$H$_6$</td>
<td>0.47580</td>
<td>0.10775</td>
<td>3.23633</td>
</tr>
<tr>
<td>C</td>
<td>C$_2$H$_4$</td>
<td>1.63040</td>
<td>45.10014</td>
<td>3.50743</td>
</tr>
<tr>
<td>H</td>
<td>C$_2$H$_4$</td>
<td>0.19000</td>
<td>10.33654</td>
<td>2.12769</td>
</tr>
<tr>
<td>C</td>
<td>C$_2$H$_2$</td>
<td>1.5514</td>
<td>43.36264</td>
<td>3.42613</td>
</tr>
<tr>
<td>H</td>
<td>C$_2$H$_2$</td>
<td>0.1448</td>
<td>15.63427</td>
<td>1.96363</td>
</tr>
</tbody>
</table>

Figure 2.1: Mean unsigned error between the PHAHST, PHAST*, and PHAST models and ab initio energies for each possible trimer.
2.4 Conclusions

Next-generation PHAHST models for the C2 hydrocarbons (ethane, ethylene, and acetylene) were fit and validated against high quality \textit{ab initio} trimer energies. A novel, physically based repulsion/dispersion potential and mixing rules are the main innovation over the current generation force-fields. The PHAHST models demonstrated improved unsigned and signed errors relative to Lennard-Jones based force-fields, which were systemically too repulsive. The PHAST* and PHAST models have been used to accurately predict binding sites and binding energies in a large...
number of metal-organic materials however the absence of compatible material force fields for PHAHST models have limited their application.
CHAPTER 3

DEVELOPMENT OF PHAHST MODELS FOR THE NOBLE GASES, N\textsubscript{2}, AND H\textsubscript{2}

3.1 Introduction

There is a pressing need to development next-generation models for many energy-relevant atoms and molecules, including the noble gases, N\textsubscript{2}, and H\textsubscript{2}. The noble gases are relevant to many industrial processes, including the separation of hazardous nuclear waste, as well as being an excellent test bed for the newly developed repulsion/ dispersion due to their lack of permanent electrostatic interactions.\textsuperscript{54–56} N\textsubscript{2}, being the primary component of our atmosphere, is ubiquitous on Earth and thus is involved in real life applications of separations technology.\textsuperscript{32} The promise of a “hydrogen economy” and the resulting flurry of interest in H\textsubscript{2} separations and storage technologies have made next-generation H\textsubscript{2} models a priority as well. This series of next-generation small molecule and material models, dubbed PHAHST, is designed with transferability, accuracy, and predictive power in mind.

The critical difference from our previous PHAST family of potentials is a new repulsion-dispersion potential consisting of the three leading asymptotically correct dispersion coefficients paired with exponential repulsion. The reasoning is as follows: intermolecular repulsive forces are the result of a decrease in the electron density in the region of intermolecular overlap, de-screening the nuclei, due to the
Pauli principle and the deformation energy is well described by an exponential function in the region of typical intermolecular separations. The leading dispersion term, \( C_6/r^6 \), contributes only about 60% of the total dispersion energy at typical equilibrium separations, necessitating the inclusion of higher terms to reproduce \textit{ab initio} interaction surfaces with high accuracy. \( D^9 \) Damping functions, being useful to prevent singularities in the dispersion potential at small internuclear distances in addition to being physically grounded, reduce the need to include very high order dispersion coefficients. \( ^{17} \) This choice is theoretically well grounded, enables the use of accurate mixing rules, and creates the opportunity to decrease the parameter space encountered during the fitting process by restraining dispersion coefficients to their experimental or calculated values. While highly accurate potentials with many terms fit computationally have an important role, the goal here is physically driven minimalism that will facilitate transferability. \(^3\)

An important contribution from our work is the choice of the form of each potential energy contribution. While several reasonable choices for each part -- repulsion, dispersion, electrostatics and polarization -- exist, the combinatorial space of possible potential forms is immense. Further, there is an even larger space of possibilities for fitting parameters to a chosen model. Here an excellent candidate is chosen with demonstrated efficacy, computational tractability and a recipe for the fitting of parameters that is achievable given current computation. The published literature reflects simulation successes, while this paper represents twenty years of some success and an extensive data base of simulation failures using simple empirical potentials for sorption studies in our group. The total form of our potential
energy includes contributions from the repulsion-dispersion, permanent electrostatic energies, and many-body polarization,

\[ U = U_{rd} + U_{el} + U_{pol} \]  

(3.1)

The permanent electrostatic energy is calculated via Coulomb's law between point partial charges, usually located on the nuclear coordinates, although this is a choice not a limitation.

\[ U_{el} = \sum_{i \neq j} k_e \frac{q_i q_j}{r_{ij}} \]  

(3.2)

The repulsion-dispersion potential takes the following form where \( \beta_{ij} \) is the exponential range parameter, \( \rho_{ij} \) is the distance at which the repulsive force is equal to \( F_0 \), chosen as \( 10^{-3} \) a.u. to give practical intermolecular distance parameters, and \( C_{2n} \) is the \( 2n \)'th asymptotically correct dispersion coefficient.

\[ U_{rd} = \sum_{i \neq j} F_0 e^{-\beta_{ij}(r_{ij}-\rho_{ij})} + \sum_{n=3}^{5} f_{2n}(\beta_{ij}r_{ij}) \frac{C_{2n}}{r_{ij}^{2n}} \]  

(3.3)

The choice of damping function, \( f_{2n}(\beta_{ij}r_{ij}) \), is that of an incomplete gamma function derived from the semiclassical Drude model and first proposed by Tang and Toennies.\(^{17}\) The range parameter, \( \beta_{ij} \), is assumed to be the same \( \beta_{\{ij\}} \) appearing in equation 3.4 as both characterize the range of electronic overlap.

\[ f_{2n}(\beta_{ij}r_{ij}) = 1 - e^{-\beta_{ij}r_{ij}} \sum_{k=0}^{2n} \frac{(\beta_{ij}r_{ij})^k}{k!} \]  

(3.4)
The dispersion coefficient mixing rules used herein are the usual geometric average.\textsuperscript{1,60} The repulsion mixing rules arise from a distortion model where the electronic densities of each atom are constrained to terminate at an approximately planar surface intermediate between them. Minimizing the total distortion energy results in equal and opposite restoring forces.\textsuperscript{61} The form of the repulsion used herein simplifies the chosen mixing rules to an arithmetic average of the distance parameter $\rho_{ij}$ and a harmonic average of the range parameter $\beta_{ij}$.

\[
\rho_{ij} = \frac{1}{2} (\rho_{ii} + \rho_{jj}) \\
\beta_{ij} = \frac{2 \beta_{ii} \beta_{jj}}{\beta_{ii} + \beta_{jj}}
\]  

Even though the repulsion/dispersion potential does not constitute a significant fraction of computational time during a typical polarizable simulation improvements in speed may be obtained via implementing the pair potential as an interpolated lookup table or utilizing faster approximate exponential functions in the Tang-Toennies damping function.\textsuperscript{62} Doing so reduces the computational cost of the repulsion/dispersion potential to that of a typical Lennard-Jones simulation.

\textbf{3.2 Methods}

As discussed in Section 3.1 the new PHAHST models require the determination of multiple parameters for each atom type, including two repulsion parameters, three dispersion coefficients, the static dipole polarizability, and the partial atomic charges. Due to a large number of parameters per atom and aided by our choice of physically grounded van der Waals parameters, it is our strategy to fix as many parameters as possible to either their observable or \textit{ab initio} calculated values. This
is most easily seen in the choice of geometry and partial charge distribution, however, recent advancements have calculated higher-order dispersion coefficients efficient enough to be routine as well.\textsuperscript{25,63} While it may be theoretically possible to calculate the exponential repulsion parameters via examination of, e.g., the long-range behavior of the monomer electron density, due to the large cancellation of attractive and repulsive energies resulting in the final interaction energy it is our choice to freely fit the repulsive parameters to reproduce the intermolecular potential surface as accurately as possible with the given form of the potential.

For all sorbates (the noble gases, H\textsubscript{2}, and N\textsubscript{2}), the static dipole polarizability and the two leading dispersion coefficients were calculated using the Williams-Stone-Misquitta (WSM) technique as implemented in the CamCASP suite of programs. Briefly, isotropic multipolar polarizabilities at the static and 10 imaginary frequencies were fit to minimize the square of the residuals of the point-to-point response between 2000 points on the 2x-4x Van der Waal surface. The point-to-point responses were calculated at the PBE0/aug-cc-pvqz level of theory and dispersion coefficients were then obtained via numerical integration of the polarizability as a function of the imaginary frequency. More details are available elsewhere.\textsuperscript{2,25,64} Due to the high multipolar polarizabilities that would otherwise be involved the C10 coefficient is taken as the ratio of the C6 and C8 coefficients, $C_{10} = \frac{49}{40} \times \frac{C_{8}^2}{C_{6}}$, as derived from isotropic quantum harmonic oscillator calculations. While dimensioned ratios derived from quantum harmonic oscillator calculations are considered unreliable dimensionless ratios have been demonstrated to be reliable.\textsuperscript{65} Atomic partial charges were evaluated using the orbital-optimized
coupled-cluster singles and doubles density with the standard CHELPG method as implemented in the electronic structure code ORCA.\textsuperscript{36–39}

The remaining exponential repulsion parameters, $\beta_{ii}$ and $\rho_{ii}$, were fit to reproduce previously published high-quality dimer surfaces using the surface fitting routine of the Monte Carlo simulation package MPMC in a least-squares manner.\textsuperscript{27,29,55,56,66–69} To remain consistent with the published datasets, the zero-point vibrationally averaged bond length of 1.1014 Å was used for N\textsubscript{2} and the experimental rotational spectra bond length of 0.742 Å was used for H\textsubscript{2} throughout. Configurations with an interaction energy above $1000 \, \frac{K}{k_B}$ were not included in the fitting process except in the case of He where a cutoff of $100 \, \frac{K}{k_B}$ was used instead due to the relatively small dimer binding energy of He. It was found that including repulsion parameters on the center of mass (COM) site of H\textsubscript{2} drastically improved the quality of the fit for H\textsubscript{2}, however, including repulsion parameters on the COM site of N\textsubscript{2} did not significantly impact the fitting process. This choice can be thought of as an ad hoc method to include anisotropy in the repulsion potential of H\textsubscript{2} while maintaining compatibility with pair based molecular simulation codes. Note that it would be trivial to extend the PHAHST models with bonded parameters via existing parameterization methods, e.g. QuickFF and similar codes, however in the primary focus of our work, adsorption simulations in relatively rigid MOFs, it is believed to be of limited importance and would require an order of magnitude of additional Monte Carlo sampling for converged properties.
3.3 Results and Discussion

The appropriateness of the dispersion coefficient mixing rule, a geometric mean rule, was initially tested and is shown in Figure 3.1. Dispersion coefficients were calculated via the Williams-Stone-Misquitta method both for homo-nuclear and hetero-nuclear dimers. Geometrically mixed dispersion coefficients were compared to directly calculated hetero-nuclear coefficients and excellent agreement was found. The best agreement was with the mixed C6 coefficients, all coming within one or two percent of \textit{ab initio}. Slightly worse agreement in the higher order terms is tolerable as these coefficients a minority of the total dispersion energy.

![Figure 3.1: Mixed dispersion coefficients for the noble gas pairs calculated via \textit{ab initio} and geometrically mixed.](image)

Next, Born-Oppenheimer surfaces were compared for the models and reference \textit{ab initio} energies. Figure 3.2 shows the Born-Oppenheimer surface for a representative noble gas, Ar; Figure 3.3 and Figure 3.4 shows the Born-Oppenheimer surfaces for the small molecules H$_2$ and N$_2$ respectively. Excellent
agreement is found for all noble gas dimers, with the small molecules have slightly higher fitting errors.

Figure 3.2: Born-Oppenheimer energy surface for Argon computed via CCSDT(Q) and the resulting PHAHST model. Almost perfect agreement is seen by only fitting the repulsive part of the potential. All other noble gas energy surfaces look similar.
Figure 3.3: Born-Oppenheimer energy surface for $\text{H}_2$ computed via CCSD(T) and the resulting PHAHST model. Very good agreement is seen after introducing repulsion parameters on the COM site, an ad hoc method of introducing repulsion anisotropy.
The second virial coefficient is a convenient initial validation of the developed models as it is an experimental observable sensitive to the two-body potential. The second virial coefficients were calculated by numerically integrating the Mayer f-function, orientationally averaging in the case of H$_2$ and N$_2$.\textsuperscript{70}

\begin{equation}
B_2(T) = -2\pi \int_0^\infty \left[ \langle e^{-\beta U(r)} \rangle - 1 \right] r^2 dr
\end{equation}

In the case of He a first-order quantum correction was applied, however, this correction did not significantly alter the second virial coefficient above
approximately 10 K. Figure 3.5 contains the calculated and experimental second virial coefficients as a function of temperature. Good agreement with the experiment is achieved across all sorbates and temperature ranges as likely a necessary albeit not sufficient evaluation of the potentials.

Figure 3.5: Calculated and experimental second virial coefficients for H2, N2, He, Ne, Ar, Kr, and Xe. Experimental data are taken from Dymond and Smith (1969) and interpolated.
As further validation, pressure-density isotherms in the supercritical region were calculated for all sorbates using the LAMMPS molecular dynamics code for computational convenience. The novel pair potential was implemented using a lookup table and polarization was handled naively in LAMMPS via Drude oscillators.
functionally identical results are obtained with reference to our implementation. For the noble gases, a 2 fs timestep, and a Nose-Hoover thermostat/barostat was used with a damping parameter of 1000 fs and all other parameters default. For H\textsubscript{2} and N\textsubscript{2} a 0.5 fs timestep was used along with a Nose-Hoover barostat and a Langevin thermostat, with the Drude oscillator degrees of freedom thermostated to 1 K. Figure 3.6 contains the deviations from experimental bulk pressure-density isotherms at 298 K (400 K for Xe). Overall, good agreement with experiment is observed. Larger deviations occur with an increasing number of electrons, both with the noble gases as well as H\textsubscript{2} and N\textsubscript{2}, possibly reflecting increasing uncertainty in their associated atomic parameters (e.g. C6, C8, etc) or the underlying potential itself (e.g. the possible necessity of including higher electrostatic multipole moments). The deviations may also be explained by the absence of many-body effects beyond polarization in our model; Vogel et al. found the inclusion of a 56 parameter three-body potential containing both short and long-range components necessary to accurately reproduce bulk densities in the supercritical region of Ar.\textsuperscript{71} For the sake of simplicity and transferability, this approach is not taken in this work; additional investigation into the effect of many-body potentials and practical inclusion in heterogeneous environments is needed. Indeed, a planned extension of this work is replacing the inverse sixth part of the potential with many-body van der Waals interactions that use a formalism isomorphic to the polarization potential with zero external field and an ad hoc form Drude form for the polarizability leading to distinct matrix equations that need to be solved.\textsuperscript{29,76} These methods can handle the many-body dispersion interactions accurately and systematically, without additional adjustable parameters. They are computationally feasible although
nontrivial and less effort has been taken to improve the efficiency of solving the requisite equations. Still, standard matrix manipulations are required that can be accessed in optimized forms for parallel computer execution. These interaction energies are generally substantial and important even though they are typically neglected. They can contribute $\approx 30\%$ of interaction energies between nanostructures.\textsuperscript{24}

**3.4 Conclusions**

Next-generation PHAHST models for a series of energy-relevant small molecules have been developed covering a wide range of potential use cases. Explicit polarization and physically based repulsion/dispersion potential were included in every environment. The models were fit to high quality \textit{ab initio} calculations, including CCSD(T) and explicit calculation of dispersion coefficients via the Williams-Stone-Misquitta method.\textsuperscript{25,64} Validation was performed by comparison to experimental second virial coefficients and bulk pressure-density isotherms. Excellent agreement was found with all experimental data. Further work needs to be done to validate the models in heterogenous material simulations.
CHAPTER 4

DEVELOPMENT OF PHAHST MODELS FOR MATERIAL SIMULATIONS

4.1 Introduction
The Space group has been part of a group of developers of empirical potentials that can predict and explain / retrodict energetics and dynamics of small molecules interacting with complex materials with high fidelity.\textsuperscript{4,12,13,57,77–83} A principle thrust of ours has been the study of metal-organic frameworks (MOFs), a porous three dimensional subset of metal organic materials. MOFs are an emerging class of materials that show great potential in the areas of gas storage, separations, and catalysis due to their large surface areas, pore volumes, and the ability of their pore chemistry to be rationally designed and modified for specific applications.\textsuperscript{84–86} Theoretical modeling is extremely useful in providing guidance to experiment; however, such studies are particularly challenging due to the heterogeneous interfaces present in MOFs. Traditionally, empirical potentials have been designed to work best in homogeneous media like neat liquids where effects like induced dipoles can be included in a mean field fashion in the permanent point charges typical of most models. An illustrative example is simple water molecules where permanent point charges produce a dipole of \( \approx 3 \) debye while the gas phase value is \( \approx 1.85 \) debye; the dipole is increased to represent an average liquid state value. This works remarkably well in the bulk but is entirely inadequate in interfacial or
sorption modeling. For another example of strong, specific interactions that highlights uncertainty in traditional empirical potentials, consider the recycling of used nuclear fuel results in off-gas streams containing radioactive Kr isotopes and nonradioactive yet chemically similar Xe isotopes. Modeling can aid in designing separations from each other for safe and economical storage; recently MOFs have been proposed as a solution. Empirical simulations of distinct atomic / molecular species are ubiquitously handled by mixing like species interaction potentials using prescribed “mixing rules”. There is no exact way to construct them and theory guides reasonable choices. It should be noted that the typical Lennard-Jones mixing rules for attractions, using an geometric mean for the well depth, often denoted as \( \epsilon \), is not the theoretically sound choice which would be the geometric mean of the inverse sixth power term of the potential. The choice of mixing rules in a simulation becomes increasingly important as the disparity in van der Waals parameters grows yet the typical mixing rules used fail spectacularly when calculating cross interactions involving these heavy noble gases. This work addresses the failings associated with current generation materials simulations, with an emphasis on the currently challenging areas of classically describing polarization derived phenomena like open-metal site interactions and dispersion mixing of heavy noble gases, to develop a series of next generation potentials in a highly accurate and transferable manner without sacrificing computational tractability.

HKUST-1 was chosen for the development of potential energy surfaces due to the large quantity of high-quality experimental data and previous theoretical studies allowing for a basis of comparison; it is also a non-trivial material with a
complicated topology and a diversity of binding sites. Open-metal sites are present along with cages where sorbate/sorbent interactions are strong. While dispersion interactions are typically thought of as long-range contributions adding to the cohesive energy, the large inverse powers associated with these phenomena are strong and quickly varying at short distances and serve to balance to typically exponential repulsions in many sorption systems.

HKUST-1, first synthesized in 1999 by Chui et al., consists of “copper paddlewheel” secondary building units (SBUs) linked by 1,3,5-benzenetricarboxylate (BTC) ligands into three distinct pores: a 13 Å spherical pore primarily decorated with open Cu$^{2+}$ sites, a slightly smaller 11 Å spherical pore primarily decorated with 1,3,5-benzenetricarboxylate ligands and a significantly smaller, roughly 5 Å tetrahedral pore bounded by four benzenetricarboxylate linkers with access provided by four trianularly shaped windows leading to the larger spherical pore.$^{89}$ The diversity of pore shapes has largely been overshadowed by the presence of open-metal sites, with experimental studies focusing on the binding of, e.g., H$_2$ to the exposed Cu$^{2+}$ sites, however the small tetrahedral pores are especially important in the case of noble gas adsorption and separations.$^{54,90-92}$

Often metal-organic materials simulations in the literature are restricted to generic force-fields based on the Lennard-Jones equation with an inverse twelfth power repulsion and inverse sixth power attraction. Typical examples include DREIDING, Universal Force-field (UFF), or Optimized Potentials for Liquid Simulations-All Atom (OPLS-AA).$^{78,93-95}$ The electrostatics are represented by atomic point charges in this class of models. The Lennard-Jones equation, originally chosen mainly due to the computational efficiency of obtaining the repulsion term by
squaring the dispersion term, is well known to be overly repulsive at close distances and the typically employed Lorentz-Berthelot mixing rules have been shown to underestimate well depth parameters between mixed noble gases by as much as 80\%.\textsuperscript{1,57}

While a review of force-field development is out of the scope of this article due to the large amount of published work. Considering biomolecular applications and development of force-fields based on the symmetry adapted perturbation theory (SAPT) and density functional theory (DFT), readers are directed elsewhere.\textsuperscript{2,4,5,80} Recent developments as directly pertains to metal-organic force-field development will be briefly summarized. Broadly, work has proceeded focusing on either the intramolecular part of the potential, reproducing experimental crystal structures to improve the current understanding of flexible systems, or the intermolecular part of the potential, focusing on reproducing binding sites and energies. Three recent advances in intramolecular potentials are QuickFF, MOF-FF, and VMOF. QuickFF, an automated software package primarily for the parameterization of covalent terms, has been extended to periodic systems and validated against experimental structural parameters of four MOFs.\textsuperscript{33,96} MOF-FF, based on the MM3 van der Waals potential and using genetic algorithms to improve the fitting process, has successfully reproduced experimental lattice constants for five different materials.\textsuperscript{97} Similarly, VMOF, also based on the MM3 potential, has been shown to accurately reproduce phonon spectra of a wide variety of MOFs.\textsuperscript{98} Many groups have made important contributions to developing intermolecular potentials for material simulations, including Smit, Sholl, Schmidt, and Space. Smit, et al. used van der Waals corrected DFT to parameterize CO\textsubscript{2}, H\textsubscript{2}O, and CH\textsubscript{4}
interactions with the M-MOF-74 series and showed that generic force-fields based on UFF inadequately describe both the energetics of open-metal site interaction and uptake isotherms.\(^8^2\) The Sholl group have published a large number of papers parameterizing CO\(_2\), CH\(_4\), noble gases, and a wide range of hydrocarbons for use with a number of materials, including MOFs and zeolitic imidazolate frameworks (ZIFs), showing that generic Lennard-Jones based force-fields inadequately describe binding site energetics yet also highlighting the difficulties in modeling noble gas adsorption accurately.\(^8^1,8^1,9^9–1^0^2\) The Schmidt group has developed both small molecule, mainly N\(_2\) and CO\(_2\), and material force-fields fit to reproduce SAPT calculations and they have achieved excellent agreement across a wide range of materials.\(^1^0^3,1^0^4\) Work in the Space group has previously focused on the development of highly accurate small molecule potentials, denoted PHAST, fit to highly accurate couple cluster interaction energies and with explicit polarization for use alongside generalized Lennard-Jones material force-fields.\(^1^6,2^9,3^2\)

4.2 Methods

For the determination of the HKUST-1 parameters five representative fragments were chosen from the experimental X-ray single-crystal structure ranging in size from Cu\(_2\)(HCO\(_2\))\(_4\) to Cu\(_6\)(HCO\(_2\))\(_9\)(C\(_9\)H\(_3\)O\(_6\)) with truncated bonds being replaced by a single hydrogen atom (see Supporting Information). These fragments were chosen to preserve functional groups (in the case of HKUST-1, “copper paddlewheels” and BTC ligands) as much as possible while minimizing the disruption caused by the truncation of bonds. Note that the choice of using fragment-based methods to parameterize the nonbonding parameters is arbitrary, previous work has shown that periodic DFT based methods with configurations generated either by identifying
approach paths or by simulating using a generic classical forcefield successfully reproduce experimental uptakes.\textsuperscript{81,82} Atomic partial charges were evaluated at the PBE0/def2-tzvpp level of theory with the CHELPG method as implemented in ORCA.\textsuperscript{36} Due to the size of the fragments chosen, the Williams-Stone-Misquitta methodology is not feasible for the calculation of atomic multipole polarizabilities and dispersion coefficients. Instead, exchange-hole dipole moment (XDM) calculations at the PBE0/def2-qzvpp level of theory were used to evaluate the atomic multipole moment integrals, \( \langle M_l^2 \rangle \), from which dispersion coefficients may easily be obtained in combination with the static dipole polarizability, e.g. \( C_6 = \frac{1}{2} \alpha \langle M_l^2 \rangle \).\textsuperscript{63} Constraining the dispersion coefficients in this way results in three fitting parameters, the two exponential repulsion parameters, \( \beta \) and \( \rho \), and the static dipole polarizability, \( \alpha \). The fitting set consisted of 200 randomly orientated configurations per pair of the five HKUST-1 fragments with the sorbates Ne, H\(_2\), and N\(_2\) (for 3000 total configurations) within the 0.75x-1.5x Van der Waal surface. Interaction energies were calculated with the VV10 nonlocal dispersion corrected density functional, chosen due to its superior description of dispersion bound complexes, and the def2-tzvpp basis set, correcting for BSSE.\textsuperscript{105-107} The parameters for each MOF atom were then fit in a least-squares manner, holding the parameters of the sorbates constant to their previously determined values, using the surface fitting routine of MPMC with a cutoff energy of 1000 K/k\(_B\).\textsuperscript{77} Table 4.1 contains all parameters of the developed force-field as well as the root mean square error (RMSE) of the final fitting process for each sorbate as well as the material fitting process.
Grand Canonical Monte Carlo simulations of N$_2$, H$_2$, Ne, Ar, Kr, and Xe were performed in HKUST-1 with the MPMC code (see the Supporting Information). Each simulation was run for 1,000,000 equilibration steps, which consist of either a combined translation and rotation move on a single molecule or an
insertion/deletion move, and averages were calculated over 6,000,000 production steps. All internal bonds were constrained to be rigid during the simulation and explicit polarization was included in every condition. Ne simulations were carried out at 60 K, Ar simulations at 140 K, Kr simulations at 200 K, Xe simulations at 260 K, and H$_2$ and N$_2$ simulations were carried out at 303 K and 77 K. Simulation isosteric heats of adsorption (Qst) were determined from fluctuation theory and the reproduced experimental data was determined with the Clausius-Clapeyron equation and experimental adsorption isotherms over a range of temperatures.$^{92,108}$ Figure 4.3 contains the calculated and experimental isosteric heats of adsorption for the sorbates in HKUST-1. Binding sites were determined by analyzing simulation trajectories and refining single-molecule positions via simulated annealing in the case of N$_2$ and H$_2$. Binding site energies were calculated with the MPMC code and a single sorbate molecule located at the refined position from the previous step.

4.3 Results and Discussion
The fit parameters and associated RMSE for the material force field is presented in Table 4.1.
Table 4.1: Fit parameters and associated RMSE for each fitting process for the newly developed PHAHST material force-field.

<table>
<thead>
<tr>
<th>Site</th>
<th>Molecule</th>
<th>(\alpha (\text{Å}^3))</th>
<th>C6 (a.u.)</th>
<th>C8 (a.u.)</th>
<th>C10 (a.u.)</th>
<th>(\beta (\text{Å}^{-1}))</th>
<th>(\rho (\text{Å}))</th>
<th>RMSE (K/(k_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>He</td>
<td>He</td>
<td>0.2002</td>
<td>1.4072</td>
<td>11.136</td>
<td>107.96</td>
<td>4.6845</td>
<td>2.3838</td>
<td>11.2</td>
</tr>
<tr>
<td>Ne</td>
<td>Ne</td>
<td>0.3823</td>
<td>6.2127</td>
<td>67.986</td>
<td>911.38</td>
<td>4.9943</td>
<td>2.8024</td>
<td>16.6</td>
</tr>
<tr>
<td>Ar</td>
<td>Ar</td>
<td>1.6550</td>
<td>65.460</td>
<td>1438.9</td>
<td>38745</td>
<td>3.8853</td>
<td>3.6862</td>
<td>27.2</td>
</tr>
<tr>
<td>Kr</td>
<td>Kr</td>
<td>2.4970</td>
<td>130.1</td>
<td>3981</td>
<td>149225</td>
<td>3.5289</td>
<td>4.0348</td>
<td>21.8</td>
</tr>
<tr>
<td>Xe</td>
<td>Xe</td>
<td>4.026</td>
<td>288.4</td>
<td>11390</td>
<td>551047</td>
<td>3.2469</td>
<td>4.4752</td>
<td>33.7</td>
</tr>
<tr>
<td>H</td>
<td>H₂</td>
<td>0.34325</td>
<td>2.8847</td>
<td>38.972</td>
<td>644.96</td>
<td>3.1006</td>
<td>1.8594</td>
<td>4.0</td>
</tr>
<tr>
<td>COM</td>
<td>H₂</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>3.6278</td>
<td>2.6645</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>N₂</td>
<td>0.85092</td>
<td>17.805</td>
<td>416.32</td>
<td>11925</td>
<td>3.8537</td>
<td>3.3151</td>
<td>38.6</td>
</tr>
<tr>
<td>COM</td>
<td>N₂</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>MOF</td>
<td>0.29252</td>
<td>6.9696</td>
<td>262.83</td>
<td>13951</td>
<td>8.8235</td>
<td>2.7385</td>
<td>90.1</td>
</tr>
<tr>
<td>C</td>
<td>MOF</td>
<td>0.71317</td>
<td>11.890</td>
<td>547.52</td>
<td>27318</td>
<td>4.0015</td>
<td>3.3593</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>MOF</td>
<td>1.6806</td>
<td>27.701</td>
<td>709.36</td>
<td>19821</td>
<td>3.8954</td>
<td>3.2387</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>MOF</td>
<td>0.02117</td>
<td>0.16278</td>
<td>5.03239</td>
<td>202.99</td>
<td>36387</td>
<td>1.8745</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.2: Quality of the resulting PHAHST material force-field fit to reproduce DFT/VV10 interaction energies between small molecules and HKUST-1 fragments.

Radial distribution functions, see Figure 4.4, were calculated for Ne, Ar, Kr, and Xe from the center of the tetrahedral cages at intermediate loadings (approximately 4 sorbates per tetrahedral cage). Finally, the locations of the identified binding sites are shown in Figure 4.5.

In agreement with previous X-ray powder diffraction and neutron powder diffraction studies by Zeric et al. Ar, Kr, and Xe were observed to bind primarily to the tetrahedral cage and virtually no binding to the open-metal site was observed. The most favorable site calculated for Xe (27.8 kJ/mol) is located near the center of the tetrahedral cage, maximizing van der Waals contact with all four adjacent
linkers. This site is approximately 0.7 Å away from the exact center of the tetrahedral cage, resulting in four symmetric yet overlapping binding sites, only one of which may be filled due to the Van der Waals radii of Xe. Four additional symmetric binding sites (25.4 kJ/mol) are located at the window sites of the tetrahedral cages, resulting in five Xe per tetrahedral cage at high loadings. A drop of about 5 kJ/mol occurs in both the simulation and experimental Qst after filling all available center sites (approximately 0.83 mmol/g). The most favorable site for Kr (22.0kJ/mol) is found to be at the window site while the center site (20.5 kJ/mol) is slightly farther off the exact center (about 1.2 Å) than it was in the case of Xe. The center site for Kr can be thought of as moving out farther to maximize van der Waals contact with three adjacent aromatic linkers, due to the slightly smaller van der Waals radii of Kr over Xe. Each tetrahedral cage has four symmetrically identical window and center sites; similarly to Xe, only a single center site may be occupied at any one time, however, in the case of Kr sorption at a center site blocks sorption at the adjacent window site and vice versa, resulting in only four Kr per tetrahedral cage. The binding sites for Ar are intermediate of the center and windows sites, 2.1 Å off the exact center in the direction of the window sites. However, Ar has a small enough van der Waals radii to not impede adsorption at the other binding sites in the cage, resulting in four Ar per cage at higher loading. Both Kr and Ar are observed to behave similarly in the simulation Qst, dropping off after filling 4 atoms per tetrahedral cage (approximately 3.3 mmol/g). Ne is observed to have no preferred binding site as its van der Waals radii is relatively small compared to the shape of the tetrahedral pores, this is evident in both the relatively flat and featureless simulation Qst and broad radial distribution inside the tetrahedral pores. In all cases
of noble gas adsorption, the polarization energy contributes a very small (<3% for Ne, <2% for all others) portion of the total interaction energy due to the relatively nonpolar surface of the linkers.

$N_2$ and $H_2$ display no preference towards distinct binding sites at 303 K, owing to the high thermal energy (2.5 kJ/mol) relative to the difference in binding site energy, resulting in the relatively flat experimental $Q_{st}$ at all loadings at this temperature. At 77 K, the $Q_{st}^0$ for $H_2$ increases to 7.5 kJ/mol and $H_2$ begins to preferentially sorb to the open-metal site with a calculated binding energy of 10.4 kJ/mol, with the next best binding site located at the window site of the tetrahedral cages (8.5 kJ/mol). The binding energy at the open-metal site is dominated by electrostatic interactions (41% permanent electrostatics, 15% polarization) while the window site is dominated by repulsion/dispersion interactions (76%) with a small polarization contribution (3%), emphasizing the importance of including explicit polarization in heterogeneous, i.e. open-metal site, adsorption. Without explicit polarization the predicted binding energies for the open-metal site and the window site would be roughly equivalent. The open-metal site binding energy is within experimental uncertainty (10.1 ± 0.7 kJ/mol) of a temperature dependent IR study on $H_2$ adsorption in HKUST-1. At 77 K $N_2$ is fully loaded in HKUST-1. The lowest two binding sites found for $N_2$ are both located in the tetrahedral cage, either in the window site, maximizing charge-quadruple interactions with the carboxylate oxygens, or deeper in the cage, maximizing van der Waals contact with the BTC linkers. The open-metal site binding energy of a single $N_2$ (13.3 kJ/mol) is slightly lower than the average energy per particle (13.7 kJ/mol), consistent with observations of relatively weak $N_2$-$Cu^{2+}$ interactions. Attempts to optimize the
position of $N_2$ interacting with $Cu^{2+}$ consistently resulted in $N_2$ adjusting to a position adjacent to the carboxylate oxygens.

Figure 4.3: Calculated and experimental isosteric heat of adsorption for $H_2$, $N_2$, Ne, Ar, Kr, and Xe in HKUST-1. Noble gas experimental data is taken from Forster et al. and $H_2$ and $N_2$ experimental data is taken from Staudt et al.$^{92,108}$
Figure 4.4: Calculated radial distribution functions for Ne, Ar, Kr, and Xe with the center of the tetrahedral pores during an intermediate loading (approximately 4 sorbates per tetrahedral pore) Grand Canonical simulation. Peaks near 0 Å are assigned the center site, 3-4 Å the window site, ≈8 Å the larger, spherical pores, and 11-12 Å the window site of adjacent tetrahedral pores. Kr and Xe displays clear differentiation between the center and window site, with the distinction decreasing as a function of atomic size. Ar displays only a single, intermediate binding site and Ne lacks any well-defined binding sites.
Figure 4.5: Small molecule binding sites located via simulated annealing with the PHAHST force-field and MPMC.
For the purposes of comparison, binding site energies were evaluated at the previously identified locations via PHAHST, the typical settings for a Lennard-Jones simulation in MPMC and the QuickStep module of the periodic DFT code CP2K. The rVV10 functional, similar to the VV10 functional used in the initial material parameterization scheme with a slight modification to allow for efficient implementation in plane wave based codes, was chosen due to its accuracy in reproducing noble gas dimer curves, among other data sets.\textsuperscript{105,109,110} Within CP2K, a cutoff of 900 Ry, 5 integration grids, a relative cutoff of 60 Ry, a nonlocal Van der Waals cutoff of 200 Ry, and a convergence criterion of $10^{-6}$ was used. The basis set TZVP-MOLOPT-GTH was used for C, H, and O, TZVP-MOLOPT-SR-GTH for Cu and DZVP-MOLOPT-SR-GTH was used for the noble gases. All other parameters were left at their default. For the Lennard-Jones calculations in MPMC OPLS-AA Lennard-Jones parameters were used for C, H, and O, with UFF Lennard-Jones parameters used for Cu. The Darkrim-Levesque model was used for H\textsubscript{2} and the TraPPE model for N\textsubscript{2} was used.\textsuperscript{13,15} Noble gas Lennard-Jones parameters were taken from the literature.\textsuperscript{100,111} Polarization contributions were not included in the Lennard-Jones calculations. PHAHST consistently outperformed the generic Lennard-Jones force-field with errors of approximately 1 kJ/mol with respect to the periodic DFT calculations vs >3 kJ/mol for the generic Lennard-Jones force-field. Additionally, the generic Lennard-Jones based force-field significantly underestimated the magnitude of H\textsubscript{2} binding to the open-metal site and errors in noble gas binding energies were systemically overly attractive, with the singular exception of Xe binding in the window site, likely due to the inaccurate Lorentz-Berthelot mixing rules.
Table 4.2: Binding energies from positions found in Figure 4.5 calculated with periodic DFT, the newly proposed PHAHST force-field, and a generic Lennard-Jones based force-field.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Site</th>
<th>rVV10/CP2K</th>
<th>PHAHST/MPMC</th>
<th>OPLSA-AA/UFF/MPMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ne</td>
<td>Center</td>
<td>-5.2</td>
<td>-4.5</td>
<td>-6.3</td>
</tr>
<tr>
<td></td>
<td>Window</td>
<td>-5.7</td>
<td>-5.3</td>
<td>-6.5</td>
</tr>
<tr>
<td>Ar</td>
<td>Center</td>
<td>-16.5</td>
<td>-16.1</td>
<td>-17.5</td>
</tr>
<tr>
<td>Kr</td>
<td>Center</td>
<td>-21.0</td>
<td>-20.5</td>
<td>-24.4</td>
</tr>
<tr>
<td></td>
<td>Window</td>
<td>-20.5</td>
<td>-22.0</td>
<td>-21.0</td>
</tr>
<tr>
<td>Xe</td>
<td>Center</td>
<td>-29.0</td>
<td>-27.8</td>
<td>-33.3</td>
</tr>
<tr>
<td></td>
<td>Window</td>
<td>-24.5</td>
<td>-25.4</td>
<td>-21.6</td>
</tr>
<tr>
<td>H₂</td>
<td>Metal</td>
<td>-12.4 (10.1 exp.)</td>
<td>-10.4</td>
<td>-3.4</td>
</tr>
<tr>
<td></td>
<td>Window</td>
<td>-7.8</td>
<td>-8.5</td>
<td>-8.1</td>
</tr>
<tr>
<td>N₂</td>
<td>Center</td>
<td>-18.3</td>
<td>-17.8</td>
<td>-17.3</td>
</tr>
<tr>
<td></td>
<td>Window</td>
<td>-17.4</td>
<td>-17.7</td>
<td>-16.3</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>-11.4</td>
<td>-13.4</td>
<td>-11.4</td>
</tr>
<tr>
<td>RMSD</td>
<td></td>
<td></td>
<td>1.08</td>
<td>3.22</td>
</tr>
</tbody>
</table>
Table 4.3: Comparison of simulated and experimental isosteric heat of adsorption (Qst).

<table>
<thead>
<tr>
<th>MOF</th>
<th>Sorbate</th>
<th>Simulation Qst (kJ/mol)</th>
<th>Experiment Qst (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOTT-100</td>
<td>H₂</td>
<td>-7.9</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>N₂</td>
<td>-13.4</td>
<td>-5.1</td>
</tr>
<tr>
<td></td>
<td>Ar</td>
<td>-12.4</td>
<td>-5.0</td>
</tr>
<tr>
<td></td>
<td>Kr</td>
<td>-16.1</td>
<td>-14.5</td>
</tr>
<tr>
<td></td>
<td>Xe</td>
<td>-20.4</td>
<td>-18.6</td>
</tr>
<tr>
<td>NOTT-101</td>
<td>H₂</td>
<td>-7.4</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>N₂</td>
<td>-11.5</td>
<td>-14.4</td>
</tr>
<tr>
<td></td>
<td>Ar</td>
<td>-11.2</td>
<td>-8.1</td>
</tr>
<tr>
<td></td>
<td>Kr</td>
<td>-15.8</td>
<td>-14.1</td>
</tr>
<tr>
<td></td>
<td>Xe</td>
<td>-22.1</td>
<td>-21.1</td>
</tr>
<tr>
<td>PCN-14</td>
<td>H₂</td>
<td>-7.5</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>N₂</td>
<td>-14.1</td>
<td>-14.2</td>
</tr>
<tr>
<td></td>
<td>Ar</td>
<td>-13.6</td>
<td>-10.2</td>
</tr>
<tr>
<td></td>
<td>Kr</td>
<td>-19.2</td>
<td>-16.5</td>
</tr>
<tr>
<td></td>
<td>Xe</td>
<td>-29.4</td>
<td>-17.9</td>
</tr>
</tbody>
</table>

### 4.4 Conclusions

A new family of models, denoted PHAHST, has been developed for H₂, N₂, the noble gases, and a prototypical metal-organic material with next-generation potentials and fit to *ab initio* calculations. A procedure is described that is amenable for
inclusion of additional sorbates and sorbents in the PHAHST force-field family.

Accuracy and transferability are emphasized with the use of minimalist theoretically grounded potentials and mixing rules. These models are designed from the outset to have predictive accuracy in material simulations. All models were verified against experimental second virial coefficients, bulk pressure density data and known binding site behavior. Agreement was excellent among the bulk datasets and with respect to the material simulations. The known deficiency of Lennard-Jones based potentials in describing sorption systems with any reasonable accuracy was overcome by using the PHAHST potentials. Indeed, PHAHST potentials have the desired predictive and retrodictive power in describing materials sorption for rare gases and other non-charged, repulsion/dispersion dominant sorbates.

Future studies will be focused on expanding the library of parameters as well as further exploring the importance of many-body effects in material simulations. Metal-organic materials may contain a wide variety of linkers and metal nodes, requiring an equally wide variety of atomic parameters and robust training sets to ensure transferability to novel materials. Flexibility, where important, may be incorporated using already established methods of fitting parameters to reproduce ab inito Hessians. Many-body Van der Waals / dispersion interactions are known to contribute significantly to the many-body energy and their application to next-generation potentials has not yet been explored. Many-body van der Waals methods are still computationally expensive and advances in algorithms may be needed for their inclusion to become routine however. Additionally, including anisotropy in the repulsion potential may prove important in producing highly accurate small molecule potentials, such as CO₂.
In sum, a new generation of potential functions is presented to supersede the outdated Lennard-Jones based parameterization paradigm. A minimal set of potential energy functions was chosen from amongst the many possible, along with a recipe for their parameterization that is computationally distinct yet tractable for sorbates and sorbents; the resulting force-field is shown to have accuracy comparable to tractable electronic structure calculations on the same systems and is in excellent agreement with extant experimental data. Thus, the number of interfacial and sorbate/sorbent systems that is amenable to high fidelity exploration via empirical molecular modeling is now greatly expanded. The insights provided here ought to apply equally well to force-field development for chemical biology simulations that face similar molecular heterogeneity with complex interfaces and charged moieties throughout. The potential form is amenable to evaluation in the LAMMPS molecular dynamics code and is implemented in our publicly available MPMC Monte Carlo code.
CHAPTER 5
CONCLUSIONS AND FUTURE DIRECTIONS

A new generation of fast, accurate, and transferable force-fields for molecular simulations, PHAHST, has been developed from first principles. The force-field is designed from the outset to capture all essential physics of intermolecular interactions, with physically justified functional forms, mixing rules, and parameter sets fit to extremely high-quality \textit{ab initio} calculations. Several small hydrocarbons, H\textsubscript{2}, N\textsubscript{2}, the noble gases, and a prototypical class of metal-organic materials have been modeled with results closely matching experiment and periodic DFT calculations. The force-field is extensible by construction and additional small molecule and material parameters may be easily parameterized. Additionally, many-body van der Waals potentials may be incorporated for increased accuracy in close contact binding involving multiple sorbates or dense supercritical states. Our simulation code is open source and freely available at 
https://github.com/mpmccode/mpmc. Alternatively, the force-field may be implemented using the native pair styles present in the open source LAMMPS molecular dynamics simulation package.

Future directions are mainly focused on improving the potential library, both with new small molecule and material models, and additional compatible improvements to the underlying potentials. High priority small molecules include CO\textsubscript{2}, methane, and the C3 hydrocarbons as they are all relevant to the energy
industry and global climate change. On the potential side, fast many-body van der Waals variants are currently being developed and tested for accuracy. Future improvements important in small molecule sorption may include introducing a degree of anisotropy in the repulsion potentials and improving the description of electrostatics, either in the small molecules with higher multipole moments or in the materials by the electrostatic potentials directly. It is of the utmost important that any additional extensions to the PHAHST force-fields maintain the currently excellent levels of transferability, accuracy, speed, and predictive power.
REFERENCES


(17) Tang, K.; Toennies, J. P. An Improved Simple Model for the van Der Waals Potential Based on Universal Damping Functions for the Dispersion Coefficients. 


(39) Sherrill, C. D.; Krylov, A. I.; Byrd, E. F.; Head-Gordon, M. Energies and Analytic Gradients for a Coupled-Cluster Doubles Model Using Variational


(70) Mayer, J. E.; Mayer, M. G. *Statistical Mechanics*; John Wiley & Sons, Inc.: New York, United States, 1940.


(73) *NIST Thermophysical Properties of Fluid Systems*.


