Accurate Interatomic Potentials for Simulations

Thesis by

James Joseph Gerdy

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California
1996
(Submitted December 1, 1995)

To Nathan Sugarman

Acknowledgements

I want to thank my advisor Bill Goddard who has been a great advisor. He has helped me to think critically and has provided excellent research resources for the group. Bill has kept the group exciting and rewarding both scientifically and socially. He hasn't done it alone, though. Thanks also go to his great secretaries Carol Scrivner and Debbie Chester who kept things running.

I also thank the friends I have made in the Goddard group and across campus. Thanks especially to my friends Kraig Anderson, Charles Musgrave, Michael Belmares, and Eric Kelson. You have made my time at Caltech a lot of fun.

I have also met my wife here at Caltech. Thanks, Amy, for all of your support. And I thank my parents for prodding me along to finish my work here.

I have dedicated this thesis to my general chemistry professor at the University of Chicago, Nathan Sugarman. He is the reason actually that I chose to study physical chemistry rather than developmental biology. After my freshman year there he was always available to offer advice on the many decisions I had to make. I and many others miss him.

Thesis Abstract

This thesis develops a rational foundation for the application of long-range forces to atomistic simulations. This area has lagged behind the other components of simulations because of two factors. First, the nonbond forces are difficult to probe experimentally. There are only a few materials for which properties clearly and directly correlate with van der Waals forces (such as molecular crystals) and then some of these cases are not relevant to common modeling applications (*e.g.* the hydrogen van der Waals forces in $H_{2(xtl)}$ are different from those in hydrocarbons). Second, more than for valence force field terms, van der Waals forces are difficult to determine by calculation. The forces are weak and require a large number of basis functions per atom to treat properly.

This thesis contains a method which has optimized a level of *ab initio* calculation on small clusters in order to extract a van der Waals potential. The size of the calculation is controlled by carefully optimizing the basis set. Moreover, unlike for previous calculations of this sort, it was recognized that the repulsive and attractive potentials (the monotonic potentials) which constitute the van der Waals potential can be calculated optimally with different basis sets, further accelerating the calculation for a given level of accuracy. Also the use of diffuse basis sets off of the atom centers is used here to make the basis sets more efficient. The method has been optimized for the case of nitrogen because it is both a closed shell case and relatively common in simulations. What results is a computational method which produces pair potentials for use in force field simulations. This is called the combination of monotonic potentials (COMP) method.

Subsequently, potentials have been calculated for the atoms H, He, C, N, O, F, Ne, Si, P, S, Cl, and Ar. In order to test the accuracy, the potentials are applied to test cases of molecular crystals and compared to other commonly used potentials. Important issues that are addressed are standard combination rules and the accuracy of using isotropic pair potentials. COMP potentials give a measure of accuracy of van der Waals potentials for any atom. This research has also yielded an

accurate functional form, a variant of the Morse potential, which has not been used in simulations but provides very accurate fits to the *ab initio* data. The relationships between different functional forms are analyzed so that the designer of force fields can make a judicious choice of both *ab initio* calculations to determine potentials and the appropriate functions with which to fit them.

Table of Contents

Chapter I. Introduction

This study concerns the determination of accurate intermolecular potentials from *ab initio* methods for application to atomistic simulations. Of the potentials needed for force field simulations the intermolecular van der Waals forces are often the most difficult to determine. There have been two approaches, empirical and computational. Empirical potentials are usually accurate since they are fit to material properties, but lack versatility and have not been determined for many atoms. The lack of versatility stems from the fact that the optimization is done to fit the properties of a material or class of materials while what is usually desired is a potential for an atom which may be present in other materials with a different hybridization or van der Waals contact geometry. In short, the dependence of empirical potentials on the materials chosen, or those available, for the fit limits their breadth of applicability. This is often not a severe problem, for empirical potentials have been incorporated into the DREIDING[1] force field with good results. The other problem with empirical potentials is that they can not be determined in any direct manner for some atoms such as Si, S, P, and metals. The problem is that there are not enough molecular crystals with these elements for which fits can be performed. So, *ab initio* methods are ideal for the problem. If a calculation can be devised to yield van der Waals energy it may be applicable to a variety of the elements including those overlooked by empirical fits.

The *ab initio* approach here looks at small (primarily tetraatomic) clusters and computes the energy at several geometries. If this is done so that the only difference between the geometries is the van der Waals energy it is possible to extract a potential. Issues from deciding on a methodology to interpreting the results for some elements are studied here. First, in order to develop an *ab initio* method reference is made to the accurate van der Waals potentials which have been determined for noble gas dimers. These methods cannot be simply copied to other, non-noble, elements for the methods used become far too expensive. The studies

serve as a guide, however, to what levels of accuracy can be achieved with *ab initio* calculations. The method is optimized for a particular case, nitrogen, by computing van der Waals energies for a variety of basis sets at the self-consistent field (SCF) and second-order Moller-Plosset perturbation (MP2) level. Second, with the optimized method available, potentials are calculated for atoms in the first and second row of the periodic table. The potentials are tested for accuracy by comparing to empirical potentials from past studies and by applying the potentials to force field simulations on a few test cases.

Third, the potentials are further analyzed with fits of functional forms. It is observed that some potential parameters can be fit to atomic properties such as the ionization potential (IP). Also the standard combination rules, arithmetic rules which are used to generate heterogeneous pair potentials from homogeneous cases, are shown to be accurate. Combination rules are a substantial aid to molecular modeling because the number of heterogeneous cases is, in general, proportional to the square of the number of homogeneous cases for a given set of elements. So, the task of determining and storing potentials is greatly reduced if combination rules are used.

The accuracy of physical interpretation of the potentials computed here arises from the basic physical understanding of van der Waals forces which comes from quantum mechanics. Within the van der Waals potential it is illuminating to understand that there are two forces, one repulsive and one attractive, which have a different origin and topological character. From gas imperfections and the energetics of crystals the simplest features of interatomic forces could be construed. First, it was clear that there is an inner wall which keeps atoms at a minimum distance and gives materials resistance to compression even when there are no bonds throughout the material (as with molecular crystals). The hardness of the inner wall can be seen in phenomena ranging from molecular beam scattering to the van der Waals Gas Law excluded volume constant. Molecular attraction can also be seen in more subtle properties such as the cohesion of molecular crystals. Nevertheless,

while such observations as gas imperfections hint at molecular forces, it is difficult to determine a potential energy surface from such studies. At the present time, there are extremely sensitive infrared and microwave techniques which can probe the energy of van der Waals clusters directly.[2-4] Early in the century, however, the precise nature and form of the van der Waals potential could only be described by wave mechanics. Accordingly, physical intuition about van der Waals forces comes from quantum mechanics.

The repulsive force arises from the Pauli exclusion principle which limits the number of electrons in an orbital. Basically, as nonbonded atoms approach their electron clouds must avoid overlapping by remaining orthogonal at the expense of increasing the curvature of the wavefunction and hence its energy.[5] This increase in energy or repulsive force can be approximated by the overlap of the nonorthogonalized orbitals from the interacting atoms and is, hence, exponential with separation. More precisely if two atoms cannot form a chemical bond as they are already bonded to other atoms within a molecule, possibly the same molecule, then they have a nonbond interaction. That they cannot bond means the atoms cannot form one or more new orbitals between them which have singlet paired electrons and a strong (over 10 kcal/mol) energy stabilization. Instead, the closer these atoms are to each other, the more overlap their valence orbitals have, but by the Pauli exclusion principle this overlap must be removed in a wavefunction solution. So a self-consistent solution of the Schrödinger equation which includes the Pauli exclusion principle, the Hartree-Fock (HF) equations, orthogonalizes the valence orbitals so that their overlap disappears. The result is that the energy of the valence orbitals on both atoms is elevated giving the repulsive pair potential.

In general an analytic calculation of the repulsive potential for a pair of atoms yields the product of an exponential times a scaled polynomial the order of which depends on the quantum states of the atoms.[6] The most prominent feature about this potential is that it is steep. Scattering experiments can use as a first approximation a hard sphere potential (Appendix A) which has an infinitely steep repulsive

function. It is this repulsive function which maintains the compressional rigidity of many materials such as molecular crystals and macromolecular complexes.

The attractive force is more difficult to describe theoretically and somewhat less important in models of materials. While the Pauli exclusion principle gives a direct understanding of the repulsive potential, the attractive potential was first described properly by London[7] who used virtual excitations to explain attraction based on instantaneous multipole interactions. Historically this was the essential step to a complete understanding, for all earlier theories did not give attractive forces between two neutral and nonpolar species. Only pure electrostatic forces and the induced dipole effect had been understood. The London theory dictates that all atoms, even noble gases (which have no multipole moments), have an attractive interaction with all other atoms. A more complete description of the dispersion force was offered by Slater and Kirkwood[8] who developed a precise notation for the London potential and higher order terms which arise from the virtual state formalism. Not only was this force more difficult to describe than the repulsive force, it has always been more difficult to calculate numerically.

The attractive force influences properties of materials such as the lattice spacing, cohesive energy, and compressibility of molecular crystals. Furthermore, it plays a crucial role in the ordering of macromolecular systems with a lot of low energy conformational barriers. Proteins are an excellent example. The primary structure is determined simply by chemical bonds. It is weak attractive forces, however, which shape the secondary and especially the tertiary structures. Also in processes such as polymerization and diffusion long-range attraction is a deciding factor in the time evolution of simulations. As current day simulations are now able to model these large systems the subtle effects of attractive potentials may be crucial.

The past twenty years have seen great progress in the simulation of materials and large molecules with atomistic, force field, simulations. Force field potentials are designed with the use of both empirical and *ab initio* data. Refinement of force

field potentials has come through fitting the potentials to experimental properties using nonlinear fitting techniques.[9-13] The most difficult part in these optimizations to determine accurately is the nonbonded potentials which include electrostatic and van der Waals potentials. The free parameter in the electrostatic part is the atomic charge. In the present thesis the concern is solely the van der Waals part. It is difficult to estimate these parameters empirically[14-15] and *ab initio* calculations are difficult because these are weak interactions which depend on correlation effects.

This study introduces an approach for an *ab initio* calculation which yields van der Waals potentials. In Chapter II the optimization process produces a basis set and level of calculation to yield accurate pairwise atomic potentials. A further strength of this approach is that it is applicable to any atom in the periodic table. It can be extended to provide potentials for all atoms from H to Lw. Chapter III applies it to a group of ten atoms and analyzes trends over the elements. Hydrogen is treated with special care as the *ab initio* approach is especially challenging to apply to it. Finally, Chapter IV summarizes the findings and discusses how this work and further efforts of this type will impact force field simulations.

References

1. S. L. Mayo, B. D. Olafson, and W. A. Goddard III, *J. Phys. Chem.*, **94**, 8897 (1990).
2. R. C. Cohen and R. Saykally, *J. Chem. Phys.*, **98** (8), 6007-6030 (1993).
3. R. G. Bone, *J. Phys. Chem.*, **98**, 3126-3138 (1994).
4. J. M. Hutson, *J. Chem. Phys.*, **96**(9), 6752-6767 (1992).
5. A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, McGraw Hill, New York, NY (1989).
6. C. C. J. Roothaan, *J. Chem. Phys.*, **19**(12), 1445-1458 (1951).
7. F. London., *Phyzikalische Zeitschrift*, **60**, 245-279 (1930).
8. J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids*, John Wiley & Sons, Inc., New York (1954).
9. S. Dasgupta and W. A. Goddard III, *J. Chem. Phys.*, **90**(12), 7207-7215 (1989).
10. N. Karasawa, S. Dasgupta, and W. A. Goddard III, *J. Phys. Chem.*, **95**(6), 2260-2272 (1991).
11. N. Karasawa and W. A. Goddard III, *Macromolecules*, **28**(20), 6765-6772 (1995).
12. N. L. Allinger, *J. Am. Chem. Soc.*, **99**, 8127 (1977).
13. W. A. Goddard III and N. K. Karasawa, *Submitted for publication.*
14. D. E. Williams, "Transferable Empirical Nonbonded Potential Functions" from R. M. Metzger, *Crystal Cohesion and Conformational Energies*, Springer-Verlag, NY (1981).
15. A .K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, *J. Am. Chem. Soc.*, **114**(25), 10025-10035 (1992).

Chapter II. The Development of the Method

II.1. Summary

A method has been developed to calculate accurate van der Waals potential energy surfaces from small quantum mechanical calculations. The *ab initio* potential function developed here is the combination of monotonic potentials (COMP) approach. This approach computes separately the repulsive and attractive potentials for a pair of atoms or a cluster. Then the repulsive, or Pauli, potential is added to the attractive, or correlation, potential to yield the total COMP potential. While this is automatically what happens in a standard correlation interaction (CI) or Moller-Plesset (MP) perturbation calculation what is different here is that the two monotonic potentials are computed separately with different basis sets and to different levels of wavefunction. This allows the accuracy to be optimized for each monotonic potential. Consequently, the COMP approach yields more accuracy per basis function (and, hence, per unit of computational cost) than standard correlated calculations. This is important because these calculations with an accurate level of electron correlation built in are expensive for heavy atoms. These calculations were done on clusters of two diatoms with D_{2d} symmetry. This symmetry gives a potential for the cluster which yields a pairwise potential for the atoms without assuming any functional form. Within the cluster there is only one unique nonbonded distance.

The optimal calculation calculates the repulsive curve at the HF level with a triple-zeta basis set plus double polarization (TZDP) which consists of $(4s3p2d1f)$ functions. This is also denoted as DZDP$^+$ because the valence functions are derived from a double zeta (DZ) containing a $(3s2p)$ set plus additional s and p functions which are scaled from it as opposed to the standard triple zeta (TZ) basis set with $5s3p$ functions which are optimized all together. Basis sets are derived from the correlation consistent basis sets of Dunning[1] except for special cases of added functions. The interaction energy is calculated at different geometries to a particular

Table II.1. N-N Pair Interaction Energy (kcal/mol) in the COMP Model

R(Å)	E_{rep}	E_{att}	E_{COMP}
2.722	2.429	-1.147	1.283
2.963	1.067	-0.699	0.368
3.205	0.459	-0.431	0.028
3.448	0.194	-0.269	-0.074
3.570	0.126	-0.213	-0.088
3.692	0.081	-0.170	-0.089
3.814	0.052	-0.136	-0.084
3.937	0.034	-0.110	-0.076
4.135	0.017	-0.078	-0.061
4.457	0.006	-0.047	-0.041
4.979	0.001	-0.022	-0.021

symmetry of the cluster.

Energies for a range of geometries define the van der Waals potential. The repulsive and attractive potentials (kcal/mol) are shown separately in the data, Table II.1, because they are both fit well by simple functions. R is the distance between nitrogen atoms (in Å). Because of the symmetry of the cluster, there is a single R value for each configuration. The repulsive energy is plotted in Figure II.1 with its best functional fit, an exponential function. The function is

$$Ae^{-CR} \qquad (II.1)$$

with $A = 39,649$ and $C = 3.5466$. The attractive monotonic potential, in this case MP2 correlation energy, is plotted in Figure II.2. It is also fit by an exponential with $B = 25.125$ and $D = 1.720$. All of the calculated energy points, upon which these exponential fits are based, are listed in Table II.1.

The COMP potential, Figure II.3, is given by the sum of the monotonic functions. It can be fit by several potential energy functions (see Appendix A for a discussion of details). First is the commonly used Lennard-Jones 12-6 function (LJ)

$$E(R) = D_e \left[\left(\frac{R_e}{R} \right)^{12} - 2 \left(\frac{R_e}{R} \right)^{6} \right]. \qquad (II.2)$$

The fit is improved by the more flexible exponential-6 (E6) function

$$E(R) = D_e \left[\left(\frac{\zeta}{\zeta - 6} \right) e^{\zeta(1-R/R_e)} - \left(\frac{\zeta}{\zeta - 6} \right) \left(\frac{R_e}{R} \right)^{6} \right]. \qquad (II.3)$$

The fit is improved with other functional forms based on the Morse potential

$$E(R) = D_e \left\{ e^{\zeta_1(1-R/R_e)} - 2e^{\frac{1}{2}\zeta_2(1-R/R_e)} \right\}, \qquad (II.4)$$

which has a much different shape to the right of R_e. One more improvement on the Morse potential which gives an even better fit is the generalization to different ζ's for the repulsive and attractive terms. This is the exponential-exponential (abbreviated exp-exp or EXEX) function

$$E(R) = D_e \frac{1}{\zeta_1 - \zeta_2} \left\{ \zeta_2 e^{\zeta_1(1-R/R_e)} - \zeta_1 e^{\zeta_2(1-R/R_e)} \right\}. \qquad (II.5)$$

All of these fits model well the COMP potential as shown in Figure B.2. The exp-exp provides the possibility of combining separately the attractive and repulsive potentials to give heterogeneous pair potentials (*i.e.* potentials between two different atoms, also known as off-diagonal). Moreover, while the Morse function has the restriction that $\xi = \zeta_1/\zeta_2 = 2$, this can have any value with the exp-exp function which allows separate adjustment of the curvatures of the regions to the left and to the right of R_e. Of the four functions above, the best fit is obtained with the exp-exp. The fitted parameters for each potential function with $(N_2)_2$ plus the rms error of the fits are in Table II.2. The R_e values in the table and following tables are given in units of Å and the D_e are in kcal/mol unless otherwise specified. Also scale

parameters such as ζ and ξ are unitless. A complete explanation of the meaning of these parameters and the details about the functions above is given in Appendix A.

It is clear from the rms error that the Morse and exp-exp fits are better quality than the LJ curve and even E6. This can be seen more clearly from Figure II.4 which plots the COMP data along with the potential curves with the fitted parameters. It can be seen that those functions with an R^{-6} term are unable to fit both the inner wall and attractive tail simultaneously. The Morse curve possesses the same shape as the data. Moreover, the exp-exp curve is versatile enough to have a nearly exact fit.

Table II.2. Potential Function Parameters Fit to the COMP Potential

Potential	R_e(Å)	D_e(kcal/mol)	ζ	ξ	RMS ($\times 10^3$)
LJ 12-6	3.6409	0.0859	...	...	10.579
Exp-6	3.6329	0.0886	14.7497	...	3.907
Morse	3.6561	0.0896	12.3808	...	3.074
Exp-Exp	3.6456	0.0891	12.389	2.640	0.543

II.2. Optimization of the Calculational Method

Levels of accuracy were balanced against computational expense to design a viable method for doing the cluster calculations to extract the van der Waals potentials. The calculation is refined by optimizing both the basis set and calculation level for accuracy of the interaction energy of a model cluster. The calculation was optimized specifically for the D_{2d} cluster of nitrogen, $(N_2)_2$. To achieve the highest accuracy with a small, double-zeta (DZ) or triple-zeta (TZ), basis set, some adjustments were made to the normal way in which van der Waals cluster calculations are done.

First, the repulsive and attractive energies were computed with two separate basis sets and levels of calculation, optimizing the efficiency of each calculation. Optimizing the repulsive calculation separately allows this potential to be computed without diffuse functions which actually compromise the accuracy of the basis set superposition (BSS) energy correction at long ranges (for $R > R_{eq}$). Furthermore, the repulsive calculation benefits from the use of extra valence s and p functions which make it an effective TZ basis set. Second, the attractive potentials were improved by using diffuse functions in places where there are no atoms, the diffuse function approach of Tao and coworkers.[2a-e]

This is a more efficient use of diffuse functions than the standard placement of an identical set on each atom. Since they need to be between pairs of atoms the functions are actually wasted to some extent when placed on all atoms, for each atom in the cluster has nonbonded atoms only on one side. Hence, more than a hemisphere is wasted as it is directed away from the other molecule. Tao and coworkers studied clusters of the type

$$X - A_2, \qquad (II.6)$$

where the X is a halogen atom and A is a monovalent atom such as H. The goal was to develop a more accurate method for calculating molecular dissociation energies. Tao and coworkers have found that the positioning of diffuse functions at the

geometric midpoint of the line segment connecting atom X with the bond midpoint of A_2 is essential for an accurate calculation. The extra diffuse functions improve the accuracy of the cluster interaction energy and the multipole moments of the constituent molecules. The dissociation energies of diatomic molecules with bond functions are calculated to be 98-99% of the experimental D_e while without bond functions only 90% is recovered. Furthermore, the dipole moment of the three polar diatomics Tao studied are improved significantly by bond functions (Table II.3).

Table II.3. Dipole Moments (DÅ) of Diatoms With and Without Midpoint Functions (From Tao and Coworkers[2e])

Molecule	Without	With	Expt.
HF	1.8499	1.7964	1.8030
HCl	1.2028	1.1148	1.1132

The total (COMP) potential then is constructed by combining a repulsive potential and a correlation potential computed with two different basis sets. This method is excellent for the analysis of the potentials because not only does it produce an accurate van der Waals curve, but also produces optimized monotonic potentials. The latter are important to gain insight into patterns across elements within a row or column of the periodic table. Furthermore, it is helpful to have rules for combining two potentials to determine a potential for the heterogeneous pair. Such rules are called combination rules and it is usually assumed that the potential parameters for the new pair will be some sort of average of the parameters for the homogeneous potentials. For instance given LJ parameters ${R_e}^{\mathrm{X}}$ and ${D_e}^{\mathrm{X}}$ for an atom X and the same for an atom Y, the heterogeneous pair XY would have parameters

$$\begin{cases} D_e^{\mathrm{X-Y}} = \sqrt{D_e^{\mathrm{X}} \times D_e^{\mathrm{Y}}} \\ R_e^{\mathrm{X-Y}} = \frac{1}{2}(R_e^{\mathrm{X}} + R_e^{\mathrm{Y}}) \end{cases} . \qquad (II.7)$$

Combination rules are tested in this study because the heterogeneous potential can be calculated directly by the COMP method. Several cases are shown below. The

combination rule in Equation II.7 is not perfect but adequate for simulations. Even more accuracy can be had by using the COMP potential. In fact, just one point of the heterogeneous potential can be used to scale the combination rule potential to yield exceptional accuracy. This works best if the repulsive and attractive potentials are combined separately.

A high level of calculation has been used, for small systems such as He_2 or Ne_2 , up to, for instance, MP4 or CI up to quadruples (QCI), and using a triple or quadruple-zeta (QZ) basis set. Klopper and Noga[3] have calculated the He_2 potential with an $(11s8p6d5f4g)$ basis set at the CISD and higher level. Their binding energy at equilibrium is converged well. With a smaller basis set, $(11s8p6d5g)$, the energy is 2.5% larger. With the larger $(11s8p6d5g4f3h)$ basis set they found the energy to differ by less than 0.01 K. The size of these three basis sets (N_{basis}) is 100, 136, and 169 basis functions per atom. Another entry in the annals of accurate noble gas dimer calculations is an Ar_2 calculation by Woon[4] which computes the cluster wavefunction at the MP4 level. The basis set is extensive, QZ and quintuple zeta (5Z), giving an equilibrium binding energy of 0.45mh compared to an experimental number of 0.42mh. Furthermore, Woon extrapolated a sequence of energies with different basis sets (DZ, TZ, QZ, and 5Z) to obtain an estimated binding energy of 0.45mh. Also the extrapolated potential makes only a 0.5% error in the equilibrium distance, $7.10a_0$, while the QZ and 5Z potentials are off by 2% and 1%. What is important for the design of the current calculational methods is that for noble gas systems, the accurate calculations commonly use QZ or higher basis sets in MP4 wavefunctions.

This level is, in fact, too slow to be practical for a series of calculations in systems with four or more atoms. There have been calculational studies on lone diatomics which yield van der Waals potentials,[5] but they are difficult to apply consistently to a varied set of atoms because these calculations rely on choosing a repulsive diatomic state that is representative of the van der Waals interaction. Potentials from clusters of diatomics or polyatomics are easier to interpret because

while the ground state of a diatom does not have van der Waals interactions, the ground state of the cluster does.

The drawbacks to the cluster approach are that there are in general several pairwise interactions within the cluster and it may be poorly defined how to extract a single pair potential and that these larger clusters are, of course, computationally more costly than single diatoms. The first problem is addressed by designing clusters of two diatoms with D_{2d} symmetry which have only one interatomic distance for van der Waals interactions. There are four such atom pairs with this distance, so the energy of the cluster need only be divided by 4 to yield a potential for the atom pair without assuming a functional form for the potential. For off-diagonal clusters, such as $N_2 - F_2$, the symmetry is C_{2v} which again has only one unique van der Waals distance.

The second problem, computational expense, is the main issue here. The basis set optimization and the additional refinements above provide an approach with reasonable speed which can be applied to any atom in the periodic table. For noble gases accurate calculations with QZ and higher basis sets are tractable because diatomic systems have only van der Waals interactions and provide all of the relevant information directly. For non-noble elements at least four atoms are required. This equates to a great increase in computational cost, for the speed of a calculation scales nonlinearly with the number of basis functions included. The speed of an MP2 calculation scales as $N_{basis}^{5.5}$. So, to go from a two atom system to a four atom system (assuming the same basis set is used in each calculation with the same number of functions per atom) means an increase of 45 times the computational time. The scaling is worse for higher levels of perturbation or CI calculation.

To study the van der Waals interaction some level of correlated wavefunction must be used. A Hartree-Fock (HF) calculation yields a purely repulsive potential for all the van der Waals clusters studied. Because of size consistency, MP perturbation calculations are a convenient way to determine higher-order properties.

The levels MP2, MP3, and MP4 have been used in studies of noble gas clusters. It is found that for He_2 , MP2 gives 80% of the equilibrium binding energy. MP3 provides 10% and the triple excitations in MP4 give 5%. The components of MP4 are less than 1%.

The compromise of accuracy against computational resources that works for noble gas dimers must be retrenched for the sake of more general calculations. It must be said, however, that while the calculation must be reduced the accuracy of these calculations is not necessarily less than for noble gases. As will be illustrated below, the error in the calculated potential is at an acceptable level. This indicates that perhaps the error associated with a particular approach (*i.e.* a particular choice of basis and wavefunction level) gives an error that is more a constant associated with that approach than a portion of the interaction energy. For instance in the case of the HeNe potential above, the calculation by Tau[2a] gave an interaction energy of 30.26 μh, off by about 35 μh from the exact number. The interaction energy in nitrogen is in excess of 500 μh so that such an error would be less detrimental to accuracy.

The computation time required for the $(N_2)_2$ potentials is measured per single point calculation (an entire potential has between 9 and 14 points). $(N_2)_2$ timing is typical of the first row cases and is used for example. A single MP2 point requires 32 to 54 cpu minutes on the HP workstations[6] (depending on the geometry as shorter R clusters take longer). Second, there are basis set superposition points which take equally long. Finally there are the HF/DZDP$^+$ points, both normal and BSS calculations, which provide the repulsive energy. These take between 14 and 42 minutes. A representative time is that for the point at the equilibrium geometry, 40 minutes for MP2/DZDPDZ and 20 for HF/DZDP$^+$. This gives a total of 120 minutes per point. From this it is safe to conclude that MP4 calculations are out of the question. The correlation part of the calculation takes 67% of the CPU time, or 80 minutes per point here. If this were done at the MP4 level a factor of about eight times more CPU time, making 640 minutes, would be required for the correlation

calculation. So, while a COMP point with MP2 requires 120 CPU minutes, using MP4 it requires 680 minutes.

Consequently, calculation of correlation energy is done at the MP2 level in the present study because of both the prohibitive cost and the diminishing returns in accuracy at higher levels. The $(N_2)_2$ potential was then studied extensively to choose optimal basis sets for the different parts of the COMP calculation. In all cases, the use of basis sets beyond the standard DZ or TZ sets is critical because the weak attractive energy in the van der Waals potential relies at large separations on the overlap of the most diffuse functions. The attractive force can be described accurately only with the use of diffuse functions because long-range attractive forces arise from correlation between excited electronic states on each center. Proper description of both these excited states and of the correlation requires diffuse functions to be added to the basis set. Furthermore, an excited state on one atom can interact with the excited state on another atom to form an unoccupied antibonding orbital and an occupied bonding orbital. For this to account for attractive forces there must be overlap between the diffuse orbitals on each center. This poses an additional challenge for the potential energy of interaction is calculated by performing single energy calculations at some particular level of correlation over a range of separation distances. In order to obtain accurate energies out to a maximum distance R_{max}, the correlation calculation with diffuse functions must be accurate and there must be sufficient orbital overlap at R_{max}. So, the larger R_{max} is required to be, the greater number of diffuse functions needed.

A range of R must be chosen over which to compute the potential. With the goal of doing atomistic simulations, the potential needs to be known for energies and distances relevant to modeling. Information near $R = R_e$, for instance, is more important than for very small R. This range (R_{min}, R_{max}) needs to define the repulsive wall for $R < R_e$, describe the curvature at the well bottom, and define the attractive tail for $R > R_e$. Here we look at the repulsive wall up to $E(R_{min}) = +D_e$ in energy and the attractive tail to $E(R_{max}) > -\frac{1}{2}D_e$. For $R < R_{min}$, $E(R)$ need

not be quantitatively accurate because simulations rarely access those regions of energy except for transients in relaxation or trajectories. The potential in this region must, however, have the correct asymptotic limits below

$$\begin{cases} \lim_{R\to 0+} E(R) = \infty \\ \lim_{R\to\infty} E(R) = 0 \end{cases}$$

This is important with modeling applications to be tolerant of initial guesses which have high energy close contacts (for $R < R_{min}$) by chance even when the simulation is intended to be at a lower energy. It is desirable to avoid this either through the choice of functional forms (Appendix A) or through simulation algorithms which check for catastrophe regions, pair potential functions $E(R)$ which fit well the COMP data on (R_{min}, R_{max}) but are not sufficiently repulsive at small R where there is no data. The most infamous case is the E6 function which fits a variety of both *ab initio* and experimental data but has the wrong limit at small R. Similarly the limit that $E(R)$ goes to 0 at long R is equally essential. While all of the functional forms discussed in this study have the proper long R limit, one can easily obtain a polynomial fit to the COMP data which diverges at long R.

The customary placement of diffuse functions is on the atom center. In clusters where the goal is to determine intermolecular attractive forces, this is not the most efficient approach. The problem with atom centered diffuse functions is that a large part of the function occupies space outside of the cluster whereas the midpoint functions are contained entirely within the cluster. For example, Tao and coworkers place diffuse functions at the geometric midpoint of noble gas dimers. This greatly reduces the number of basis functions required for the task compared with atom centered functions. At the MP4 level the attractive energy is 30.26 μh, 45% of the experimental value (66.47μh). The midpoint function used the uncontracted basis functions $\{3s3p2d\}$ (midpoint diffuse functions are denoted by putting the basis functions in braces while standard nucleus centered diffuse functions are shown in parentheses). The binding energy is 61.41 μh, over 90% of the experimental value. A further bonus with midpoint functions is their positional

independence. Tao has varied for several dimers the position of the functions along the dimer axis. For heterogeneous dimers the geometric center or axis midpoint does not give the strongest interaction. What's more the energy is very stable with respect to the position of the bond functions.

Table II.4 shows the results of F. M. Tao varying the position of the bond functions in a cluster of H_2 and He. The energies are MP4 energies with a $(10s4p2d/6s4p2d)$ basis on the H and He atoms plus diffuse functions near the center of the cluster with uncontracted $\{3s3p2d\}$ functions. Two parameters specify the positioning of the bond functions, r_G and θ_G. A vector $\mathbf{r_G}$ is defined from the H_2 bond midpoint to the He atom. Its length is r_G and its angle from the H_2 axis is θ_G. What can be learned from the table is that large displacements of the bond functions from the geometric center of the cluster change the interaction energy only little. For instance, altering r_G from its midpoint value of 3.25Å to 4.0Å changes the interaction energy by only 0.19 calories for the linear case and by 0.12 calories for the t-shaped case. Hence bond functions are forgiving in cluster calculations and can be placed at the geometric center of tetraatomic clusters here.

Tao also compared this basis set to other sets with functions on the atoms including $(6s2p)$, $(6s4p)$, $(6s4p2d)$, and $(6s4p2d1f)$. The three diffuse function sets including the one above were $\{3s3p\}$, $\{3s3p2d\}$, and $\{3s3p2d1f\}$. Eleven of the twelve possible combinations were tested. The largest case, with f functions on both the atoms and midpoint set, was omitted. The results are in Table II.5. First, the largest effect in the table is that of having bond functions. For the two highest quality atom-centered basis sets, the difference between no diffuse functions and the lowest level is over 10% in energy. Another important result is that having f functions is unnecessary, for the difference between having them on the atom and not is less that 1% . Similarly the difference between having the f function at the midpoint and not is between 1% and 2%. So, this comparison justifies the use of the $6s4p2d\{3s3p2d\}$ used for Table II.4.

Tao finds with noble gas dimers, HeNe and HeAr, similar results for an

optimal cluster basis set.[2a] With these clusters, a polarized basis set and midpoint diffuse functions yield 92% -95% of the experimental binding energy. Two atom centered basis sets, $(5s2p1d)$ and $(5s4p2d)$, are used with a $3s3p2d$ diffuse midpoint set as with H_2-He and give binding energies of 86.13 and 89.81 μh at the MP4 level. This compares well with the experimental energy of 94.03 μh.

Table II.4. Interaction Energy (cal/mol) of $H_2 - He$ for Different Bond Function Positions (Adapted from Tao [2b])

	Linear			T-shaped	
r_G(Å)	ΔE^{HF}	ΔE^{MP4}	θ_G	ΔE^{HF}	ΔE^{MP4}
3.25	25.51	29.61	0°	17.76	26.00
3.25	...	...	10°	17.76	25.90
3.25	...	...	20°	17.76	25.60
3.5	25.51	29.69	0°	17.76	26.06
4.0	25.50	29.80	0°	17.76	26.12
4.5	25.49	29.83	0°	17.76	26.10
5.0	25.49	29.74	...	...	...
No BFs	25.50	27.69	...	17.78	23.49

This versatility of the bond functions is especially encouraging for heterogeneous clusters. When the two diatoms in the cluster are different it is not clear where to place bond functions, for the geometric center may not allow both diatomic molecules to take advantage of the diffuse functions properly because of their differing sizes and orbitals. Tao's test cases, however, suggest that this is not a problem. Placing the functions at the geometric center as a rule will place them sufficiently close to whatever would be the ideal location given that the large displacements in Table II.4 caused such small changes in energy. Tao and coworkers have named these off-center diffuse functions midbond functions describing how they are placed

Table II.5. Interaction Energy (μh) Difference of H_2-He Between Linear and T-Shaped Geometries (Adapted from Tao[2b])

Diffuse fn.s	$6s2p$	$6s4p$	$6s4p2d$	$6s4p2d1f$
None	5.96	7.27	6.79	6.35
$3p2s$	5.75	5.99	5.78	5.78
$3p2s2d$	5.60	5.60	5.76	5.80
$3p2s2d1f$	5.59	5.58	5.83	...

at the approximate middle of some weak intermolecular van der Waals bond. Here they are instead called midpoint functions to distinguish another way to place off-center functions. Midbond functions, in the present study, are diffuse basis sets placed at the middle of a covalent bond in the cluster. For example, in the $(N_2)_2$ cluster, there is a single midpoint basis set at the geometric center of the tetraatomic cluster. Using midbond functions, however, would give two diffuse function sets, one at the midpoint of each N_2 molecule.

For test cases, basis sets with functions on the atoms and no midpoint functions were tried first. The basis sets are DZ or TZ. SZ is not accurate enough for interatomic forces. For reference, energies of single N_2 diatoms are listed in Table II.6 for comparison to cluster energies.

The first unknown is the optimal size of the basis set. A test of DZ and TZ basis sets on nitrogen compares the binding energy and BSS correction on the cluster of two dimers at equilibrium separation. There are three energy values given here to describe basis set optimization. The first, E_{clst}, is the cluster energy or energy of the cluster at any given geometry minus the energy of the free fragments, $E_{mon.}$, which are diatoms in this study.

$$E_{clst}(R) = E^{wavefn.}_{cluster}(R) - 2E_{mon.} \qquad (II.8)$$

Associated with each basis and calculational type set is a BSS error. This

Table II.6. N_2 Molecule Energy

Basis	Energy	Δ(mh)	Basis set
DZ	-109.10718417530	247.987	$3s2p$
DZSP	-109.25660324250	98.568	$3s2p1d$
TZ	-109.15514297577	200.028	$5s3p$
TZDP$^-$	-109.32520076106	29.970	$5s3p2d$
TZDP	-109.35517074457	0.000	$5s3p2d1f$

is the energy difference between the cluster with all electrons versus the energy with only one diatom present. The latter is a counterpoise (CP) correction energy, $E_{CP}(R)$. The energy is determined by calculating the cluster energy but with one of the monomers replaced by ghost atoms which have only the basis functions present. The nuclei and electrons are removed from the calculation. So, E_{BSS} is defined as

$$E_{BSS}(R) = 2E_{CP}(R) - 2E_{mon}. \qquad (II.9)$$

A factor of two doubles the CP energy for the two monomers. For a heterogeneous dimer, consisting of monomers F and G, there are two CP energies, $E^F_{CP}(R)$ and $E^G_{CP}(R)$. This energy indicates how complete the basis set is. The more complete the basis set is, the more each monomer is stabilized by its own basis set and is unaffected by the presence or absence of a basis set at another center. Here the DZ and TZ basis sets all have a significant amount of E_{BSS} because these basis sets are incomplete. For a more powerful calculation, on the other hand, if the basis set is complete, then E_{BSS} will be small because the wavefunction of each monomer is described well by its own basis set.

Finally, the interaction energy of the cluster is defined as the cluster energy corrected for the BSS energy

$$E_{int}(R) = E_{clst}(R) - E_{BSS}(R). \qquad (II.10)$$

Within the cluster calculation the BSS interaction leads to stabilization that would disappear with exact basis sets. Instead, E_{BSS} is treated as an error. The correction in Equation (II.10) subtracts the entirety of this energy from the cluster energy in an attempt to extrapolate to what a complete basis set, with no BSS energy, would give. Boys and Bernardi[7] have shown that CP corrections greatly improve the accuracy of calculating interaction energies.

It is possible that this correction overcompensates. The ghost atoms in one monomer in the BSS calculation cannot utilize their respective basis functions at all, but in the cluster calculation the BSS stabilization is balanced between the two monomers. So, it is likely that E_{BSS} is, in fact, an overestimate of the true BSS energy. Nevertheless, following the standard counterpoise correction method, Equation (II.10) is used to generate interaction energies. Tao and Pan[8] have determined for the Ne_2 diatom that the BSS correction is accurate. They compared counterpoise corrections over a range of calculations from MP2/DZSP to MP4/DZDP plus extensive diffuse functions. Any overcompensation is unimportant compared to the error made with the basis set on each monomer.

Beyond this, the interaction energy can be improved two ways, by calculating a higher level wavefunction like MP3 or MP4 or by using diffuse functions. Because MP4 calculations are so time consuming and require more complete basis sets to be taken full advantage of the present study uses MP2 for correlation energies. Diffuse functions are essential for accurate van der Waals interaction potentials as the Tao and Pan study, among others, indicates. A diffuse d-function (with an exponent of $\alpha = 0.080$) is included in some test cases. Table II.7 shows the results and the number of basis functions in each case. These cases demonstrate the importance of having a TZ basis set rather than DZ to reduce the BSS error. In addition, polarization and diffuse functions are necessary for a converged interaction energy. All together the best basis set, TZDP^{-}d , has 116 functions for the cluster. This basis set is considered the minimum adequate for accurate van der Waals potentials in this study.

Table II.7. The $(N_2)_2$ Cluster Energy

Basis	E_{clst}	E_{BSS}	E_{int}	N_{basis}
DZSP	512	412	100	56
DZSP$\{d\}$	2,185	1,898	287	76
DZDP$^-\{d\}$	2,283	1,884	399	96
TZSP	576	466	110	76
TZDP$^-$	586	315	271	88
TZSP$\{d\}$	939	688	251	96
TZDP$^-\{d\}$	897	536	361	116

Subsequently, the d-function is optimized by varying the diffuse exponent while keeping the tight exponent fixed. A further refinement in the diffuse functions is to choose them to form with the polarization functions an even tempered set. Cases in Table II.8 have TZDP$^-$ on the atoms and even tempered d-functions with $\alpha_1^d/\alpha_2^d = \alpha_2^d/\alpha_3^d$. Here, the diffuse d function goes with α_3^d and the other two are polarization functions. It is beneficial to adjust the polarization functions for a given diffuse function. This is a considerable improvement over the TZDP^-d case in Table II.7 where α=0.080, for all of the cases in Table II.8 have a higher binding energy than 361μh. Moreover the best two cases, $\alpha_3^d = 0.12$ and 0.15, have a considerably lower BSS correction than the other two cases, $\alpha_3^d = 0.08$ and 0.05. The optimal α_3^d is 0.12. The choice is not critical, however, as a change of 25% in α to either 0.15 or 0.08 leads to a change in E_{int} of 2.5% . So, while $\alpha_3^d = 0.12$ is used here as the preferred exponent, the cluster interaction energy is very stable with respect to the diffuse function exponent.

After the normal functions are examined, the next concern is diffuse functions, and specifically midpoint diffuse basis sets. The first test of midpoint functions was done with an additional diffuse s (0.067), a diffuse p (0.052), and a diffuse

Table II.8. Cluster Energy with Diffuse d-Functions

$(\alpha_1^d, \alpha_2^d, \alpha_3^d)$	E_{clst}	$E_{BSS}/\mu h$	$E_{\rm int}/\mu h$
(1.654, 0.500, 0.15)	834	436	398
(1.654, 0.469, 0.12)	867	464	403
(1.654, 0.364, 0.08)	964	570	394
(1.654, 0.280, 0.05)	940	548	392

d (0.038), Table II.9. The use of two diffuse d-functions potentially augments the interaction energy by about 10% with a large addition of computational time. The addition of an f-function is also examined adding to the TZDP$^-$ basis set a diffuse d-function (optimal $\alpha = 0.12$) and a diffuse f- function. Alternatively, the f-function can be a polarization function rather than a diffuse function making the basis set TZDP (instead of TZDP$^-$) with a diffuse d-function. The TZDP $\{d\}$ basis set is actually more complete than TZDP$^-\{df\}$ for it gives a much lower basis set superposition energy having a full set of DP polarization functions. The change in E_{int}, however, favors TZDP$^-\{df\}$ by about 10% . Both of these are viable basis sets for accurate cluster interaction energies.

Table II.9. Selection of the Midpoint Function Type

Basis Set	E_{clst}	E_{BSS}	E_{int}	N_{basis}
TZDP$^-\{d\}$	897	536	361	116
TZDP$^-\{2d\}$	920	506	414	136
TZDP$^-\{sp2d\}$	1,149	723	426	152
TZDP$^-\{df\}$	1,157	688	469	144
TZDP$\{d\}$	877	452	425	144

Next, midpoint functions are optimized. A sequence of different basis functions on the midbond position are examined to determine the optimal choice. The first optimization of the bond functions uses the TZDP$^-$ on the nitrogen atoms and a single d-function for the bond function. This is a 101 basis function calculation the results of which are in Table II.10 for four different $\alpha_d^{\text{diffuse}}$ values. It is determined that $\alpha = 0.19$ is optimal. The use of s, p, and d functions together (with the same exponent) in an $\{spd\}$ midpoint set increases the binding energy considerably. Furthermore, optimization of this exponent gives the same minimum (Table II.11).

Table II.10. Optimization of the $\{d\}$ Midpoint (Atoms have TZDP)

d-exponent	$E_{int}/\mu h$
0.12	341.93168
0.17	352.63620
0.19	352.74500
0.24	347.58350

Table II.11. Optimization of the $\{spd\}$ Midpoint (Atoms have TZDP$^-$)

α	$E_{int}/\mu h$
0.09	428.52668
0.19	526.49295
0.27	515.72946

A single diffuse function is just the beginning. Better results are obtained with several functions. A pair of d-functions (0.3118, 0.1039), centered at 0.18, gives an interaction energy of 369.8110 h. More substantial improvement in the interaction energy comes from using several orbitals with different angular momentum. In

Table II.12 bond functions with the optimal exponent on *s*,*p*,*d*, and *f* functions are added to the TZDP$^-$ basis set which is on the atoms. This table demonstrates that a single *d*-function is not an adequate midbond function. The use of *s*, *p*, and *d* functions together improves the situation considerably. The best two levels, *spd* and *spdf*, are far above the rest, but the use of the *f*-function gives diminishing returns for the extra seven basis functions it relies upon. As a result, the *spd* midbond function is the standard one in this study.

Table II.12. Composition of the Midpoint Set (Atoms have TZDP$^-$)

Function	E_{int} (μh)	N_{basis}
$\{d\}$	352.74500	101
$\{sd\}$	445.24024	102
$\{pd\}$	409.90401	104
$\{spd\}$	500.34115	105
$\{spdf\}$	517.64755	112
$\{2s2p2d\}$	513.04335	114
$\{2s2p2df\}$	524.80365	121
$\{2s2p2d2f\}$	529.35811	128
$\{3s3p3d\}$	513.74184	123

Accuracy may be improved by using more than one independent exponent. Table II.12 compares SZ $\{spd\}$, DZ $\{2s2p2d\}$, and TZ $\{3s3p3d\}$ sets. The midpoint sets are all centered at $\alpha_c = 0.19$. That is, for a single *s* function $\alpha_1^s = \alpha_c$. For $2s$, $\alpha_1^s = \sqrt{3}\alpha_c$ and $\alpha_2^s = \alpha_c/\sqrt{3}$ and for $3s$, $\alpha_1^s = 3\alpha_c$, $\alpha_2^s = \alpha_c$, and $\alpha_3^s = \alpha_c/3$. First it is clear that all of these $\{nsnpnd\}$ levels are converged to 25 μh or 5%. Furthermore, the difference between $\{spd\}$ and the best energy, for $\{2s2p2d2f\}$, is only 3%. So, both $\{spd\}$ and $\{2s2p2d\}$ (which is in between the two) are viable

basis sets for the interaction energy at this geometry. Because the difference in energy between $\{2s2p2d1f\}$ and $\{2s2p2d2f\}$ is less than 1% it is also possible to use up to d functions but not f. All together, the range of bond functions in Table II.12 demonstrates that $\{spd\}$ bond functions give viable cluster energies in either a single, double, or triple-zeta implementation. Below it is demonstrated that the $\{2s2p2d\}$ or better level becomes necessary in order for the whole potential curve to be accurate with a single choice of exponents. Nevertheless, at a single point, even $\{spd\}$ is adequate.

Not only must diffuse functions be optimized, but the valence and polarization functions must be tested too. The next question is what atom-centered basis set is needed on the atoms for accurate interaction energies using a $\{spd\}$ or a $\{2s2p2d\}$ bond function. Table II.13 shows energies for three atom-centered basis sets all used in combination with the $\{2s2p2d\}$ midbond function. This shows that the DZDP basis and the TZDP basis may be viable basis sets. The most economical accurate basis set with midbond functions is DZDP plus $\{spd\}$ or $\{2s2p2d\}$. With the basis set optimized, the smallest viable basis set is DZSP on the atoms plus $\{spd\}$ at the midpoint for accurate energy at the midpoint.

Table II.13. Comparison of Atom-Centered Basis Sets with $\{2s2p2d\}$ Midbond Functions

Basis set	$E_{int}(\mu h)$	N_{basis}
DZSP	482.73287	65
DZDP	526.49295	113
TZDP	515.35615	133

To decide between basis sets it is more thorough to examine a whole potential over a range of R values than E_{int} at a single point. Over the important range of R, test potentials are computed with different basis sets to see at what level they converge on an accurate potential. For this purpose the monotonic potentials are

examined separately. The midpoint functions are optimized more carefully to give $\alpha_c^{\text{diffuse}} = 0.2355$ using a DZDPDZ basis.

A set of points near R_e was chosen to define the numerical COMP potential. Here again a balance is sought. On the one hand, enough points are needed to capture the topological details of the curve. The calculated table of energies must describe the second derivative in order to be useful for modeling and simulations. On the other hand, each point contributes to the calculational cost of the potential so a minimum number is desired. To use a small number of points most effectively the spacing between points is scaled. They are most dense near the minimum where the curvature is the greatest and increasingly sparse away from the minimum in both directions. Precisely the spacing is controlled by a power function spacing rule. This method works best, of course, if the value of R_e is known. In cases that it is not, it is estimated using empirical potentials. Then, points are added in as necessary after examining the first results.

The rule used for selecting points starts by choosing one point at R_e. Then points along the attractive tail are chosen at distances of

$$R_j^{att} = R_{min} + \sum_{i=o}^{j} \gamma^{(i-1)} d \qquad (II.11)$$

where d is some spacing such as 0.10 Å and γ is the golden ratio (the solution to $\gamma(\gamma - 1) - 1 = 0$, $\gamma = 1.618\cdots$). Points along the inner repulsive wall are chosen similarly with

$$R_j^{rep} = R_{min} - \sum_{i=o}^{j} \gamma^{\frac{(i-1)}{2}} d \qquad (II.12)$$

where $\sqrt{\gamma}$ is used instead of γ because of the steeper slope of the inner wall. The parameter d is adjusted to provide a certain number of points in a given region. Three points in the inner wall and three in the tail (plus the one at R_e) are considered a minimum for a well defined numerical potential. Adding the point at the minimum there are seven points. The parameter d is adjusted then so that the attractive points reach an energy of -D_e, halfway in energy from the well bottom

to the asymptotic limit. Also on the inner wall the three points should include two attractive points plus a repulsive point, R_3^{rep}, with $E(R_3^{rep}) \leq +D_e$. Values of d range from 0.10 Å to 0.12 Å. The number of points calculated was between 9 and 12 in this study. More than seven points were included in these calculations so that regions more distant from R_e could be examined. Seven points is considered to be a minimum number of points for application of this method to other cases.

The potential energy in these final tests is refined by correcting for electrostatic interaction energy. The nonpolar diatoms used in the present study all have nonzero quadrupole moments because of the electron density in the chemical bond. Within the cluster the quadrupoles have a repulsive interaction which should be removed from the van der Waals potential. This correction is done in the interest of producing transferable potentials. That is, the electrostatic interaction within the cluster used in the *ab initio* calculation is particular to that cluster. Electrostatics in a simulation will be different in accordance with the chemical environment an atom may have. The scheme of removing the electrostatic energy from the van der Waals interaction yields a potential which can be assigned to atoms which have different charges from the diatom and the new charge will, along with the van der Waals potential, give the correct interaction potential.

The electrostatic potential, $E_Q(R)$, is computed by fitting point charges to the *ab initio* quadrupole moment for each level of calculation. Point charges are placed at the nuclei and the bond midpoint giving charges of $+q, -2q, +q$ which can be adjusted to fit a quadrupole moment, Θ. An electrostatic potential is then generated simply by, at each R of the calculation, computing the electrostatic energy of the fitted charges at their respective positions in the cluster. The quadrupole moment comes from quantum mechanics and Table II.14 lists the calculated quadrupole moments for the diatoms studied in the present thesis. There are three calculational levels for which the moments are computed. The repulsive wavefunction (HF/DZDP$^+$) gives Θ^{rep} and the attractive wavefunction (MP2/DZDP) Θ^{cor}. Also there is an HF calculation with the DZDP basis which gives Θ^{HF}. Experimental

values are listed with some diatoms for comparison. Then the potential is corrected by the following equation

$$E_{vdW}(R) = E_{int}(R) - E_Q(R). \qquad (II.13)$$

While $E_Q(R)$ is taken from the quadrupole interaction alone, the presence of midpoint functions introduces the possibility of charge transfer. Fortunately, for the facility of this method, the transfer is minimal. The largest charge transfer with the MP2/DZDPDZ calculation is at equilibrium $\approx$ 3.65Å where $q = +0.01$ on the N atoms and $-4q$ on the midpoint. At longer and shorter R values, q is smaller. At $R = 2.609$ Å, $q = -0.0024$ and at $R = 4.389$ Å, $q = 0.0077$. It is understandable that this charge transfer is largest where the diffuse set is optimized, for it depends on orbital overlap between each monomer and the midpoint. This overlap is optimized for R_e in this procedure. All together, considering the small value of these charges, $E_Q(R)$ is taken purely from the quadrupole moment. Combining Equation II.13 with the correction equation for electrostatic potential energy above gives

$$E_{vdW}(R) = E_{cluster}^{wavefn.}(R) - 2E_{BSS}(R) - E_Q(R). \qquad (II.14)$$

This new potential, $E_{vdW}(R)$, is the final result of the *ab initio* calculation.

With points chosen to define the potential, van der Waals potentials are computed with different midpoint basis sets. The repulsive curve is covered in Tables B.1 to B.3. The first two sets of data, for DZSP and DZDP basis sets, show that using bond functions is a problem since the energy is attractive at long range. The 4.979 Å energy with both SZ and DZ functions is negative which is an obvious error. This is due to the excessive BSS energy with diffuse functions and an incomplete basis set. This problem is solved by going to TZ basis sets in Table B.3. The highest level shown is TZDP$^+$DZ which has 158 basis functions. Although this basis set has bond functions, its energy is positive because of its more complete basis set. TZDP$^+$NZ (or simply TZDP$^+$) is slightly more repulsive and has far

Table II.14. Quadrupole Moments of Diatomic Molecules

Diatom	Θ^{rep}	Θ^{HF}	Θ^{cor}	Θ^{expt}
H_2	x	x	x	x
C_2	+3.372	+3.298	+2.571	
N_2	-1.290	-1.314	-1.716	-1.52
O_2	+0.366	-0.370	-0.448	-0.59
F_2	+0.734	+0.810	+1.020	+0.88
Si_2	+1.5312	+1.5312	+1.6385	
P_2	+0.6735	+0.564	+0.1341	
S_2	+1.2988	+1.2922	+1.1703	
Cl_2	+2.1978	+2.2649	+2.2512	

fewer basis functions, only 140. For repulsive potentials it is clear that midpoint functions are unnecessary and, in fact, burden the calculation with respect to time.

For the sake of saving basis functions, a plain TZDP basis set was tried with no bond functions (NZ). From the few points calculated it appears to coincide with the positive potentials, TZDP$^+$, more than the ones that go negative, DZSP and DZDP. Also TZDPNZ has only 124 basis functions, but because it is less complete than the effective quadruple-zeta basis sets, the BSS error makes it noticeably more attractive. Finally, an effective triple-zeta basis set, DZDP$^+$NZ, is actually an excellent approximation of the TZDP$^+$ potentials. Moreover, this basis set is the smallest of those in Table B.2, with only 120 basis functions (that means each point is 2.3 times faster than for the 140 basis function TZDP$^+$ basis set). The DZDP$^+$ potential is repulsive out to 5.0Å and has an energy closer to TZDP$^+$ than TZDP does. Hence, the repulsive potentials in the COMP method are HF/DZDP$^+$ energies with no midpoint functions.

For the attractive potentials, in Tables B.4 through B.6, the importance

of diffuse functions is clear. Table B.4 shows large changes from DZSPNZ to DZSP{SZ} or {DZ}. The attractive energy at $R = 3.692$Å is 698 μh for DZSPNZ but rises to 1260 and 1283 μh for $\{spd\}$ and $\{2s2p2d\}$ midpoint functions. Similarly, the attractive energy at that distance rises over 36% using the DZDP basis set, in Table B.5, upon the addition of a $\{spd\}$ midpoint function set. Even the DZDP$^+$ basis set in Table B.6 has noticeably small energies when used without midpoint functions compared to any of the potentials with midpoint functions. So, some midpoint functions are essential for accuracy. The next question is what level is necessary. Near equilibrium there is not much difference between the energies with $\{spd\}$ and $\{2s2p2d\}$ midpoint functions. At shorter range, however, the difference becomes apparent. At $R = 2.722$Å, going from SZ to DZ midpoint functions increases the attractive energy by 5.3% for DZSP and 4.1% for DZDP. A DZ midpoint set is more suited to computing a potential curve because with two different diffuse ζ values it can be more accurate at distances other than that at which the functions are optimized.

The error when the attractive energy calculation is failing is that the energies are too repulsive. If the diffuse functions are not optimized, the energy tends toward the attractive energy with no diffuse functions. Tables B.4 and B.5 show that the DZSPSZ and DZDPSZ potentials become more repulsive with respect to their DZDP counterparts away from R_e, where the diffuse functions are optimized, than at R_e. With DZDP on the atoms, for instance, the DZDPSZ energy is 2.8% more repulsive than DZDPDZ at $R = 3.692$Å but at the endpoints of the data set, $R = 4.979$Å and $R = 2.722$Å, the difference is 5.2% and 4.1% . So, DZ is a better choice for potentials. The basis sets in Table B.5 have 104 functions for DZDPNZ and 9 extra functions for each ζ of midpoint functions giving 113 and 122 functions for DZDPSZ and DZDPDZ.

The attractive energy was also computed with a quadruple-zeta basis set, TZDP$^+$DZ . This basis set gives 158 basis functions for the cluster. For the COMP method this is just beyond what is desired and would be easily applicable to new

cases. This is especially important because third and fourth row atoms will have more basis functions. It is a good test case, however, because it is an accurate energy which the DZSP and DZDP test cases approach. Of all the attractive test cases listed, DZDPDZ is clearly the closest to TZDP$^+$DZ. What's more the increased attractive energy at the endpoints that DZDPDZ has over DZDPSZ is an asset for the latter basis set because it differs from the more complete TZDP$^+$DZ set by 0.5% and 1.7% . So, with only 122 basis functions, DZDPDZ is the preferred basis set for COMP potentials.

This method, calculating the repulsive potential with HF/DZDP$^+$ and the attractive with MP2/DZDPDZ, makes it possible to compute a potential curve in 1400 CPU minutes. There is additional time for the optimization of the midpoint function which is equal to or less than the time to compute the potential. Nevertheless, to perform the same calculation at the MP4/QZTP level, as is done with noble gas dimers, would take far longer. A QZTP basis set with diffuse functions has at least 256 basis functions. Using midpoint diffuse functions reduces that to 238. Just that basis set size alone makes the MP2 calculation time go from 80 minutes to 3,200 minutes. Beyond that, going to MP4 will lengthen the calculation by about a factor of 10. Each energy point would take over 2 CPU days at the MP2 level and over 3 CPU weeks at MP4. In comparison the speedup with the COMP approach is extremely helpful.

The basis sets used in the COMP method are based on the correlation consistent set of Dunning.[1] Within this, the DZDP part for the atoms is taken directly from combining the valence part, DZ of the cc-pVDZ basis set and the polarization part, DP of the cc-pVTZ basis set. Midpoint diffuse functions are optimized for each individual cluster. Table II.15 lists optimized exponents defining the diffuse midpoint sets for atoms in the first two rows of the periodic table. The extra functions for the DZDP$^+$ repulsive potential basis set are extrapolated from the s and p functions in the Dunning set. The α_s^+ and α_p^+ values are determined by diminishing the most diffuse α value by the ratio it has with the next most diffuse function. As

Table II.15. Optimized Midpoint Functions

Cluster	$\alpha_1^{diffuse}$	$\alpha_2^{diffuse}$
$(H_2)_2$	0.225	0.075
He_2	...	...
$(C_2)_2$	0.147	0.049
$(N_2)_2$	0.408	0.136
$(O_2)_2$	0.542	0.181
$(F_2)_2$	0.277	0.0923
Ne_2	0.26	0.0779
$(Si_2)_2$	0.3464	0.1155
$(P_2)_2$	0.1386	0.4619
$(S_2)_2$	0.189	0.063
$(Cl_2)_2$	0.2078	0.0693
Ar_2	0.3464	0.1155
C_2O_2	0.225	0.075
N_2F_2	0.520	0.173

a result, the three most diffuse s and p functions in DZDP$^+$ are even tempered.

For comparison the polarization functions from the Dunning basis set are listed in Table II.16. The midpoint functions have no precise relation to the diffuse functions in every case. For a few cases, though, C, N, O, S, P, and Ar, the midpoint $\alpha_1^{diffuse}$ is about $\frac{1}{3}$ of the SP exponent, α_d^{SP}. In other cases this relation is quite different. So, the midpoint basis set must be optimized. It is worth noting that this optimization need not be excessive. The dependence of binding energy on the binding function is fairly weak. Tables II.10 and II.11 show that small deviations in $\alpha^{diffuse}$ are tolerable.

A variation of the single midbond function adopted from Tao is instead a

Table II.16. Polarization Functions

Atom	SP	DP
H	0.727	1.407, 0.388, 1.057
He	1.275	3.044, 0.758, 1.965
C	0.550	1.097, 0.318, 0.761
N	0.817	1.654, 0.469, 1.093
O	1.185	2.314, 0.645, 1.428
F	1.640	3.107, 0.855, 1.917
Ne	2.544	4.014, 1.096, 2.544
Si	0.275	0.481, 0.159, 0.336
P	0.373	0.652, 0.216, 0.452
S	0.479	0.819, 0.269, 0.557
Cl	0.600	1.046, 0.344, 0.706
Ar	0.738	1.234, 0.410, 0.890

set of diffuse functions at the midbond of each diatom. This approach has the advantage this it is more easily transferable to off-diagonal clusters which combine two different diatoms. That is, whereas the normal midbond function must be reoptimized for an off-diagonal cluster, the bond functions on each diatom may be transferable to other clusters. Moreover, some clusters which will be important to calculate may in general have less symmetry and more bonds than the D_{2d} clusters which are the focus of this study. Clusters such as benzene dimer have more bonds than $N_2 - N_2$ and, moreover, there are other geometries of an $A_2 - A_2$ cluster which do not make obvious the placement of the midbond function. The bond midpoint diffuse function, however, is more easily transferable to these other environments. Midbond functions (as opposed to midpoint) have been optimized for some systems as well. Table II.17 shows the exponents for six cases. Only for fluorine did the

Table II.17. Optimized Midbond Functions

Pair	α_{tight}	$\alpha_{diffuse}$
H_2	0.225	0.075
C_2	0.147	0.049
N_2	0.408	0.136
O_2	0.542	0.181
F_2	0.240	0.080
Si_2	...	...
S_2	...	...
P_2	...	...
Cl_2	0.210	0.07

optimal function vary from the midpoint function established above.

Irrespective of whether midpoint or midbond functions are used, the COMP method with diffuse functions placed off of the atom centers is an improvement in calculational efficiency over standard calculations. The elements of the calculation which are essential are the BSS correction, diffuse functions of some sort, and a level of correlation. An accurate van der Waals potential calculation must use at least MP2 correlation energy, diffuse $\{2s2p2d\}$ functions, and at least a DZDP set on the atoms. The optimization above show that less will not suffice. A higher level, such as MP4/QZTPTZ, would provide excellent accuracy and may avoid the need for a separate repulsive calculation, but is not viable at the present time because of the computational time it would require. Reductions in time are obtained by nonstandard arrangements. The midpoint function helps considerably. Compared to atom centered basis sets this amounts to reducing the basis set size to about half and the accompanying speedup is over a factor of ten. While it may still be slow for larger atoms this is a viable method for accurate pair potentials. The separate

repulsive calculation allows the use of a lower level correlation basis set because the repulsive energy requires a higher level than the MP2 wavefunction. This is good because the repulsive HF calculation is very quick compared to MP2. The optimization here leads to separate repulsive (HF/DZDP^{+}) and attractive (MP2/DZDPDZ) potentials.

II.3. Results and Tests.

This section illustrates the application of the COMP potential to modeling. There are two major issues. First, the assumption for fitting the potentials in Section II.1 to the *ab initio* data is that the interaction is pairwise and isotropic. This is an approximation which ignores many body effects and anisotropy due to orientation and angular dependence of the electron density on each atom. The severity of this approximation can be judged by computing the actual *ab initio* potential of other orientations of the $(N_2)_2$ dimer and comparing these potentials to what would be predicted with the pairwise model. The comparison shows that the COMP potential can be used and gives an idea of the accuracy of the pairwise potential approximation. Moreover the angular dependence is analyzed in order to propose what would be necessary in a more accurate model. Second, the potential is used in modeling the nitogen crystal by analyzing the minimum structure. Moreover, other empirical potentials, which are fit to the properties of a wide range of nitrogen containing systems, are compared with the COMP potential.

The most fundamental issue is the anisotropy of the potential. If this causes inextricable inaccuracies then application to modeling is jeopardized. Potentials of atoms in their ground state are, of course, exactly isotropic because their wavefunctions are spherically symmetrical.[9] In molecules, however, the orbitals have a directionality imposed by bonds and, less strongly, by the electrostatic field of other atoms. The most important effect here is that the wavefunctions of valence orbitals in molecules are much less symmetrical. The valence orbitals in N_2, for instance, and other diatomics have some cylindrical symmetry about the bond axis, but, in general, have anisotropy with respect to angular displacements from the bond axis. This is a substantial effect and can be seen in the anisotropy of the polarizability of diatomic molecules.

Tables B.7, B.8 and B.9 show the COMP points for the three geometries. First, the D_{2d} geometry of the standard COMP potential in Section II.2 is included, called the cross (CRS) potential. The second is the linear dimer with the two

Table II.18. Optimized Exp-Exp Parameters for $(N_2)_2$

Cluster	R_e	D_e	ζ	ξ	RMS($\times 10^3$)
CRS	3.6464	0.0869	12.3269	2.5241	1.341
TEE	3.8602	0.0662	12.2520	2.6013	27.752
LIN	4.0041	0.04599	12.7953	3.0328	0.0601
CRS1	3.6464	0.0869	12.3269	2.5241	1.341
TEE1	3.7795	0.0903	12.0000	2.0012	1.4485
LIN1	3.85097	0.08626	12.3058	3.0331	0.0597

diatoms colinear (LIN). The third and last structure studied here is a t-shaped cluster (TEE) with one diatom pointing toward the second. The first is colinear with the center of the second while the second has its bond axis perpendicular to the first. These three sample the extremes of orientation if it is figured that the end-on (parallel) and side-on (perpendicular) van der Waals potentials differ.

The most straightforward manner in which to fit this data is to assume a van der Waals pair potential is between each atom of the cluster. Then the fit is done by fitting the sum of all four pair potentials in each geometry to the *ab initio* potential. The exp-exp function was used here with the optimized parameters listed in Table II.18. This suggests that no single potential gives perfect energies for the nitrogen atom. This may arise from fundamental shortcomings in the pair potential approximation. Nevertheless, for the purpose of simulations the COMP potential from the CRS geometry is adequate. van der Avoird and coworkers[10–12] have performed MP2 calculations on $(N_2)_2$ in different orientations and distances (139 geometries) to obtain a global potential for nitrogen. They found that the global minimum to the anisotropic potential is the crossed orientation with a distance of 3.50 Å and an equilibrium energy of 0.358 kcal/mol for the cluster (which is 0.0896 for a single N-N pair). Also nitrogen crystal has been modeled by van der Waals

potentials with potentials similar to the CRS case in Table II.18.

As an aside, the parameters in Table II.18 demonstrate the robustness of the off center diffuse function approach. The COMP potential from the midpoint function calculation yields the nitrogen potential in Table II.2. The parameters are nearly identical to Table II.18. For a more detailed comparison, the data is tabulated in Table B.11 for the midpoint calculation above and in Table B.7 for the midbond calculation. These two can be regarded as interchangeable and one should be chosen based on the geometry of the calculation and how the diffuse basis sets can be used most efficiently.

A model that has been used to reproduce anisotropic potentials is the dumbbell model where repulsive centers are shifted slightly off of the atom centers. A shift could conceivably be caused by the shape of the molecular orbitals which have some order of cylindrical symmetry. D. E. Williams[13] fit the dumbbell model to the repulsive energy of $(H_2)_2$ clusters. Four orientations (linear, t-shaped, parallel, and crossed) were fit with a simple exponential repulsion function with a variable shift in the position of the repulsive centers along the bond axis. The result is that the potential is fit by an optimal dumbbell shift of 0.07Å inwards which places the repulsive centers at a separation of 0.60Å.

In the present study, however, this approach does not solve the anisotropy problem. Fits were done with the CRS potential in the LIN and TEE clusters. Like with the dumbbell model of Williams, the position of the repulsive or attractive center was variable along the bond axis. Then the displacement of the repulsive or attractive center, ΔR, was optimized. This was done separately for the repulsive and attractive energies. Optimized values of the shift are in Table II.19. There it is seen that this shift is not consistent. Unfortunately, it is not clear how to implement this into molecular mechanics.

An alternative is to simplify the anisotropy. One way to do this is to formulate a general angular dependence of the potential. First it may be supposed that each atom has an equilibrium distance along the bond axis and a separate

Table II.19. Dumbbell Model Shifts for Nitrogen Clusters Using the CRS Potential

Cluster	ΔR^{rep}	ΔR^{att}
LIN	0.06	-0.04
TEE	0.02	0.00

one perpendicular to it. Then the $r_e = \frac{1}{2}R_e$ of any pair potential (half distances are used for the orientational treatment so that the orientation dependence of each atom is separate) would be computed by interpolating between $r_\perp$ and $r_\parallel$

$$r_e = r_\perp + sin(\theta)(r_\parallel - r_\perp). \qquad (II.15)$$

Second, the LIN potential is refit with a single EXEX potential on the two nearest atoms. This assumes that the other three pairs do not have van der Waals forces because of the intervening electron clouds. The fit is shown as LIN1 (for one atom pair) in Table II.18. This potential has a markedly different R_e and D_e than LIN.

The equilibrium distances for equation (II.15) are given by halving the appropriate R_e values from Table II.18

$$\begin{cases} r_\perp = \frac{1}{2}R_e^{(CRS)} \\ r_\parallel = \frac{1}{2}R_e^{(LIN)} \end{cases} . \qquad (II.16)$$

This gives $r_\perp = 1.8232$ Å and $r_\parallel = 1.9255$ Å. No CRS1 fit is needed as all four pairs in the CRS cluster can be counted as close contacts (and perpendicular at that). Finally, the $R_e^{(TEE)}$ value is determined by computing the r_e values for the two atom pairs, the two close contacts. For one dimer, the contact is nearly perpendicular and both atoms have van der Waals interaction. For the other, the contact is nearly parallel and only one atom has a contact. The angle of the contact, θ, is taken from the equilibrium structure with a value of 10° for the perpendicular monomer and 80° for the other one. θ tends to limit values of 90° and 0° for large R. Then equation (II.15) gives r_e values of 1.924 and 1.841 for the two monomers. Adding

them together

$$R_e = r_e^{mon.\#1} + r_e^{mon.\#2} \qquad (II.17)$$

gives $R_e^{(TEE)} = 3.765$ which compares well with the value for TEE1 in Table II.18. The TEE1 potential, like LIN1, was computed assuming that some pair interactions are discounted. In this case, the TEE cluster is assumed to have van der Waals contacts. This is the reduced orientationally averaged potential (ROAP) method.

It seems that no special treatment is needed for D_e values because they vary by less than 4% in the different orientations. The curvature of the TEE potential would not be predicted exactly with a combination of CRS and LIN. The prediction would be $\kappa \approx 75.8$ while the value for $\kappa = 12.00$ in the table is 72.0. All together, the ROAP result is encouraging. The implication for simulations is that a pair potential with no dependence on bond directionality might be inaccurate. For every system besides noble gases, the simulated atoms will have bonds. This effect is studied by looking at the orientational dependence of the *ab initio* potential of the nitrogen cluster. The three clusters examine the extreme orientations of the cluster. The LIN cluster has the lone pairs of the two nearest nitrogen atoms pointing toward each other while both the CRS and TEE clusters have lone pairs pointing away from each other in some way.

Figure II.6 shows the predicted potential for the LIN cluster if just the standard COMP potential (CRS geometry) is used. The biggest difference is that the R_e and R_0 values of the predicted potential are about 0.20Å too small. Since the goal of the COMP method is to develop potentials for simulations this can be regarded as a limited success. The COMP potential here is already useful. Since the LIN geometry is an extreme, errors made in simulations would be somewhat less than 0.20Å in R_0. This error can only be corrected by adding orientational information to simulations. On the other hand, the COMP potential is close to empirical potentials. It is seen in simulations that the COMP potential is accurate. It may be concluded that the discrepancy with the LIN geometry is not a severe problem.

A test of the standard COMP potential is to do simulations using it. Nitrogen crystal is a good case because its properties are well known. Nitrogen crystal has a cubic α-phase minimum structure which is used to test the COMP potential. Simulations are done with a rigid molecule and a $(-q, +2q, -q)$ charge scheme on the atoms to reproduce the experimental quadrupole moment,[14] Θ=-1.40 DÅ. Even for the COMP potential the experimental quadrupole moment is chosen because the idea with electrostatic correction in Equation II.14 is that whatever charge is in each wavefunction is corrected for. Then the resulting potential can be applied to N_2 molecules here with their experimental charges.

Table II.20. Properties of Nitrogen Crystal from Force Fields

Case	a (Å)	U_{coh} (kcal/mol)	B (GPa)
DREIDING	5.43	2.05	2.93
COMP	5.55	1.89	3.38
PQ	5.54	2.02	...
Expt.	5.644	1.808	3.29

Experimental properties[15,16] are modeled[17] well by the force field cases in Table II.20. The empirical potential of Williams [13] uses an E6 potential with parameters (3.6621, 0.0774, 13.843) and gives slightly poorer results than with COMP/Morse. C. S. Murthy and coworkers[18,19] also fit potentials to nitrogen crystal. Their results are close to those of Williams who actually fit azohydrocarbons and not nitrogen crystal itself. The study of Murthy and coworkers developed several potentials with different charge schemes. The potential that most closely resembles the COMP potential is their point quadrupole potential (PQ) which uses the experimental quadrupole moment (Table II.14) plus a Lennard-Jones pair potential. The potential parameters are (3.724, 0.0703). For comparison this gives a

cohesive energy of $U_{coh}^{PQ} = 2.02$ kcal/mol. This potential is close to the Williams potential. With its larger R_e value it may yield a more accurate unit cell geometry. The 0K optimized value is $a^{PQ} = 5.54$ Å.

It is worth noting that all of these parameter sets err in the same direction. While the optimized cells are close to experimental cells, the cell size is too small and the cohesive energy is too large. The error in R_0 made by the COMP potential on the LIN cluster is just the same. In this case (Figure II.6), R_0 is about 4.5% too small and D_e is 10% too large with the predicted curve. If it was assumed that the simulation results have this same sort of error then we would expect $a =$ 5.405 Å and $U_{coh} = 1.99$ kcal/mol. The actual simulation results are all between this extrapolated result and the experimental numbers. This suggests that the LIN and CRS potentials are being averaged in nitrogen crystal and while the empirical potentials are optimal for the pair potential approach, accuracy has room to be improved using *ab initio* data.

Williams performed a series of fits to properties of azohydrocarbon crystals to obtain a nitrogen van der Waals potential. He regards the nitrogen crystal as an atypical case and the study focuses on properties of eight azohydrocarbon crystals. The optimal E6 potential actually is close to the potentials used to fit nitrogen crystal. The E6 parameters are (3.662, 0.0774, 13.843). This is in the range of parameters used for molecular nitrogen crystal where R_e ranges from 3.60 to 3.81 Å and D_e from 0.059 to 0.087 kcal/mol. In tests with azohydrocarbons both COMP and Williams potentials perform well. These are less sensitive tests, however, since hydrogens dominate the nonbonded interactions as they have the closest contacts.

These tests on the crystals demonstrate the accuracy of the COMP approach. From *ab initio* data a useful van der Waals potential is obtained for use in simulations. What's more, the Morse function is an adequate representation even though the EXEX function has a considerably superior fit. This is fortunate for other researchers because the Morse potential is commonly available in simulation pack-

ages. The ROAP approach, introduced above, could be a boost to accuracy to the next generation of simulation programs. The COMP method is also a means to obtain a better understanding of orientational anisotropy and gives a measure of the accuracy of the isotropic pair potentials developed in the present thesis.

References

1. The Correlation Consistent Polarized Valence Basis (cc-pVTZ) Functions for H, He, B, C, N, O, F, and Ne are in T. H. Dunning, Jr., *J. Chem. Phys.*, **90**, 1007 (1989). Functions for Al, Si, P, S, Cl, and Ar are in D. Woon and T. H. Dunning, Jr., *to be published.*
2. (a) F.-M. Tao, *J. Chem. Phys.*, **98**(4), 3049-3059 (1993). (b) F.-M. Tao, and Y-K. Pan, *J. Chem. Phys.*, **97** (7), 4989-4995 (1992). (c) F.-M. Tao, *J. Chem. Phys.*, **98**(3), 2481-2483 (1993). (d) F.-M. Tao, *J. Chem. Phys.*, **100**(7), 4947-4954 (1994). (e) F.-M. Tao, *J. Chem. Phys.*, **100**(5), 3645-3650 (1994).
3. W. Klopper and J. Noga, *J. Chem. Phys.*, **103**(14), 6127-6132 (1995).
4. D. E. Woon, *Chem. Phys. Lett.*, **204**(1,2), 29-35 (1993).
5. H.-J. Böhm and R. Ahlrichs, *J. Chem. Phys.*, **77**(4), 2028-2034 (1982).
6. Calculation times are listed for the Hewlett-Packard Apollo 9000 model 735 workstation with a clock speed of 100MHz. Calculations were also performed on model 735 computers with a speed of 125MHz giving a 25% speedup. Calculations on a Cray YMP (the Caltech/JPL machine) were between a factor of 2 and 2.16 faster.
7. S. F. Boys and F. Bernardi, *Mol. Phys.*, **19**(4), 553-566 (1970).
8. F.-M. Tao and Y.-K. Pan, *J. Chem. Phys.*, **95**(9), 3582-3588 (1991).
9. E. Merzbacher, *Quantum Mechanics*, John Wiley and Sons, NY (1961).
10. R. M. Berns and A. v. d. Avoird, *J. Chem Phys.*, **72**(11), 6109 (1980).
11. J. Tennyson and A. v. d. Avoird, *J. Chem Phys.*, **77** (11), 5664 (1982).
12. G. Brocks and A. v. d. Avoird, *Mol. Phys.*, **55** (1), 11-32 (1985).
13. D. E. Williams, *J. Chem. Phys.*, **43**(12), 4424-4426 (1965).
14. D. E. Stogryn and A. P. Stogryn, *Mol. Phys.*, **11**(4), 371-393 (1966).
15. J. Donahue, *The Structure of the Elements*, Robert E. Krieger Publishing Company, Malabar, FL, (1982).

16. J. K. Kjems and G. Dolling, *Phys Rev B*, **11**, 1639-1647 (1975).

17. The molecular mechanics and molecular dynamics calculations used POLYGRAF (V3.21 and V3.3) from Molecular Simulation Inc. of Burlington, Massachusetts.

18. C. S. Murthy, K. Singer, M. L. Kline, and I. R. McDonald, *Mol. Phys.*, **46**(6), 1387-1399 (1980).

19. C. S. Murthy, S. F. O'Shea, and I. R. McDonald, *Mol. Phys.*, **50**(3), 531-541 (1983).

Figures

Figure II.1 The repulsive potential of nitrogen (points) is fit by an exponential function and a RYD2 function.

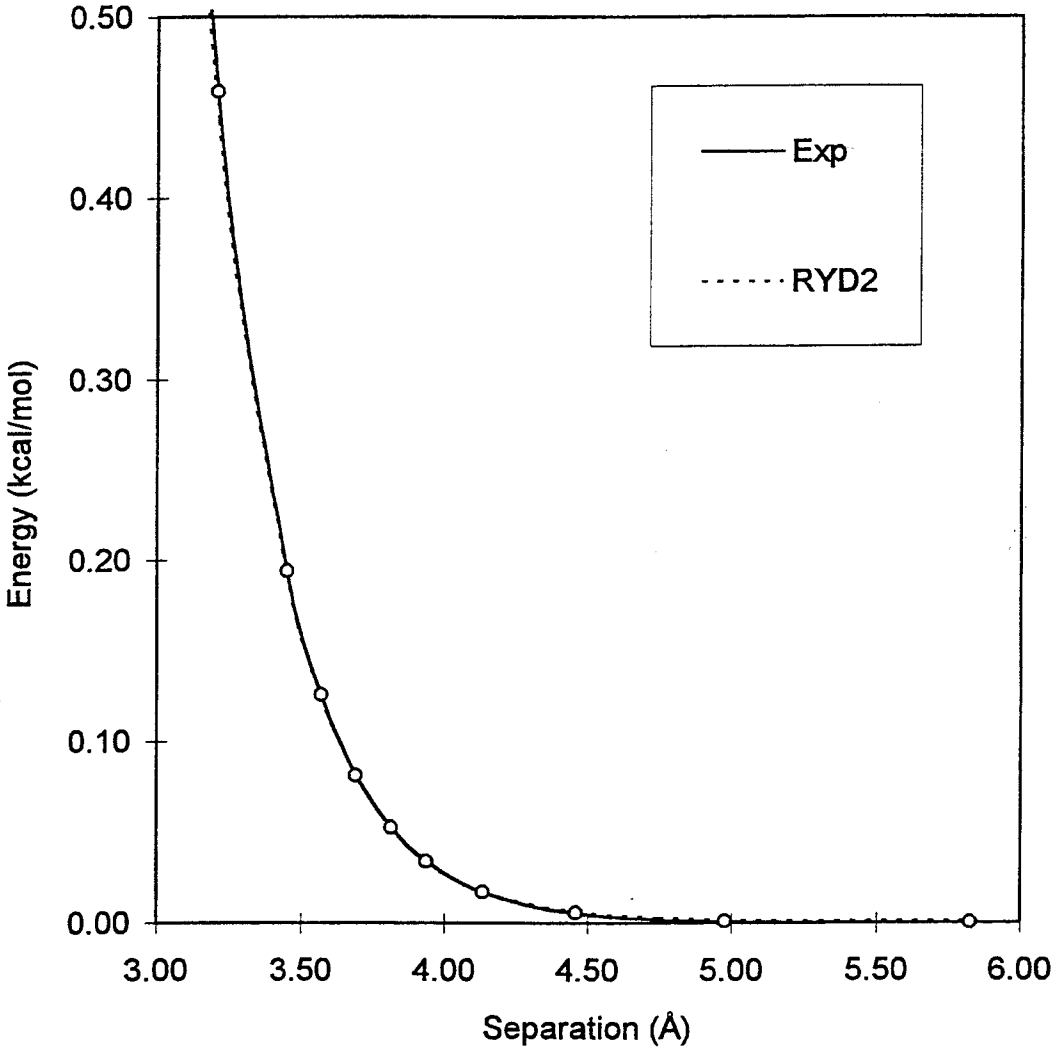


Figure II.2 The attractive potential of nitrogen (points) is fit by an exponential function and a RYD2 function.

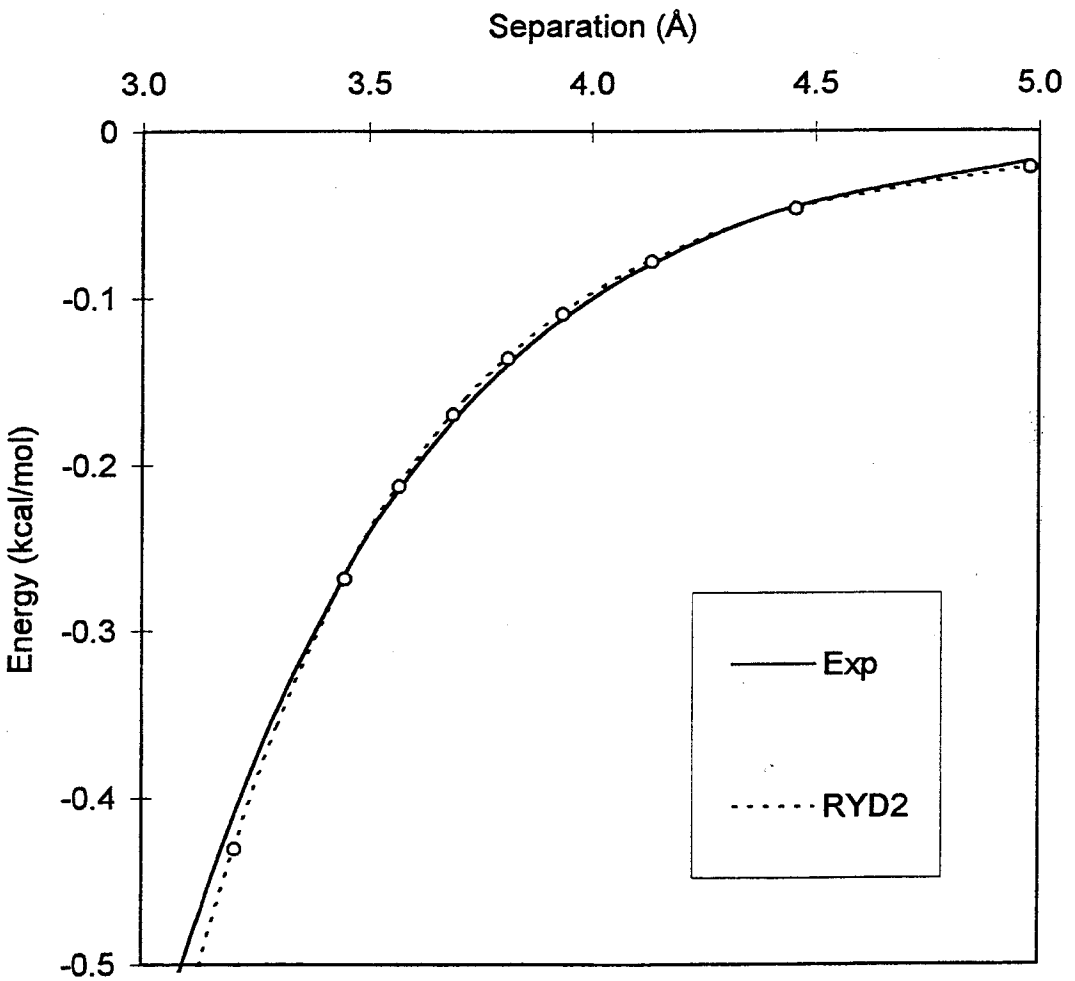


Figure II.3 Plot of *ab initio* nitrogen potential (circles) with a Rydberg function fit (line).

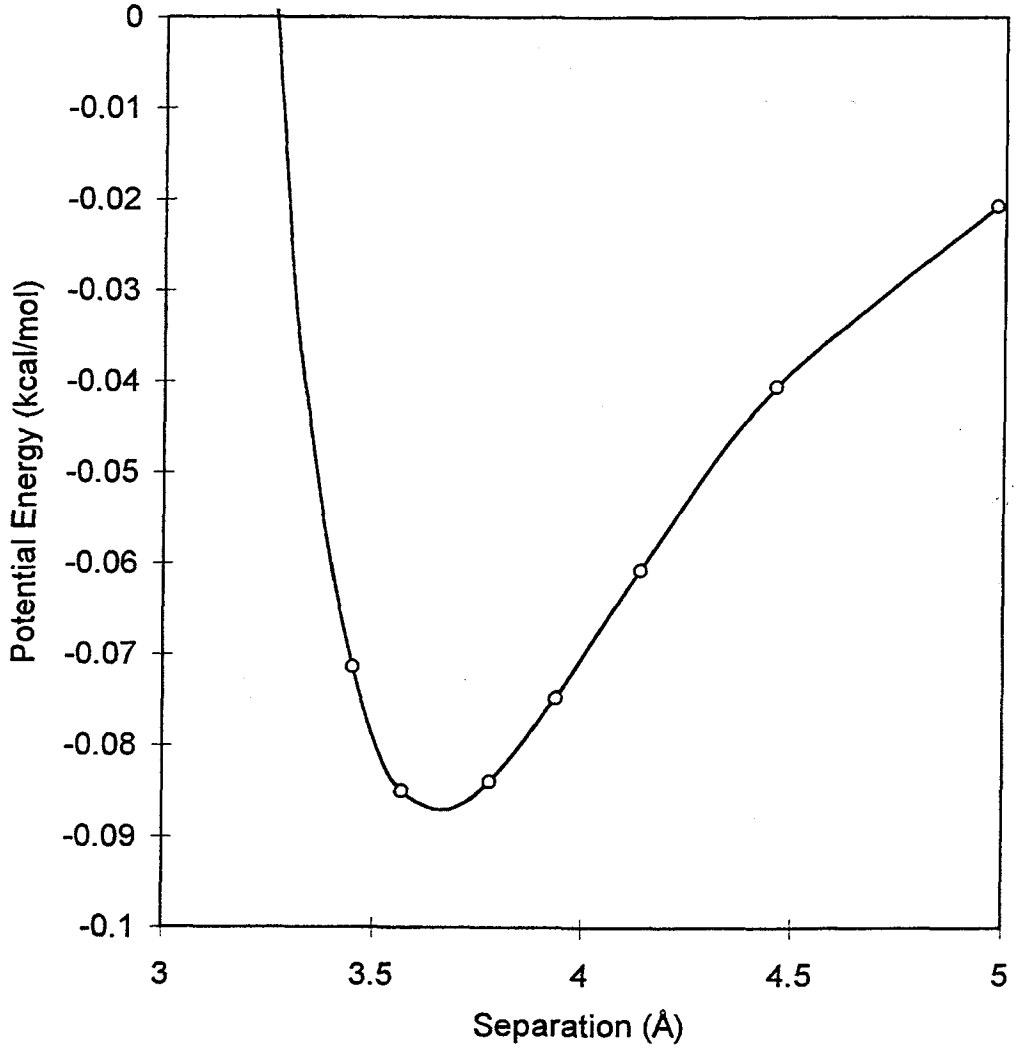


Figure II.4 The total COMP potential of nitrogen (points) is fit by potential functions.

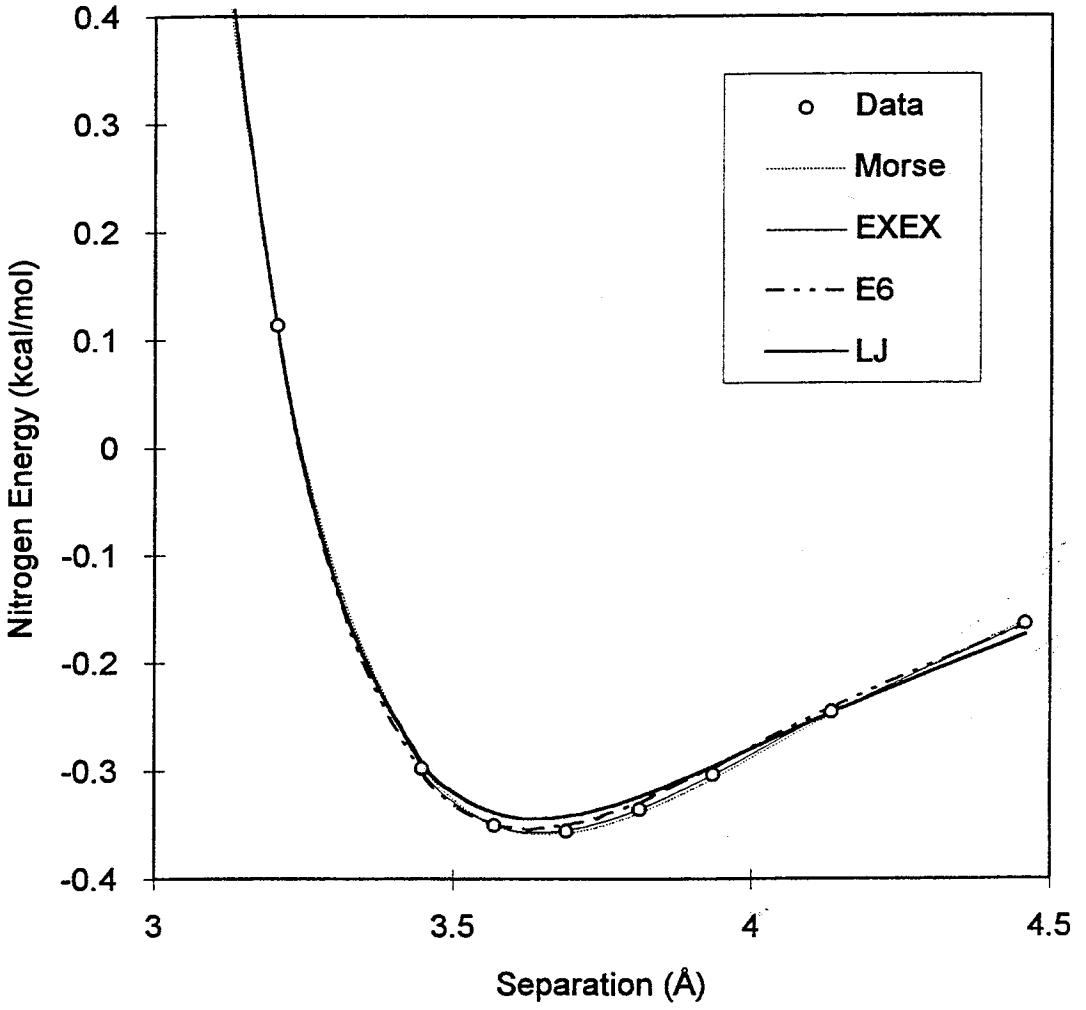


Figure II.5 The COMP potentials of nitrogen for three different geometries.

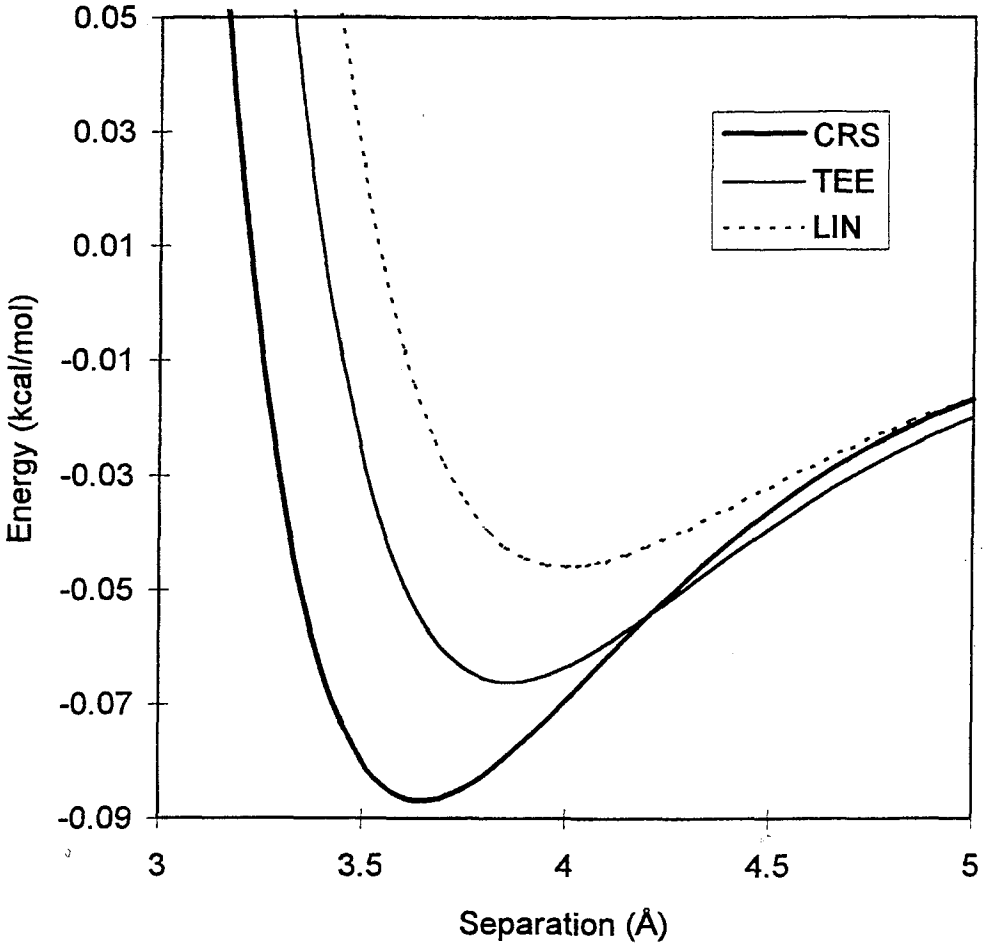
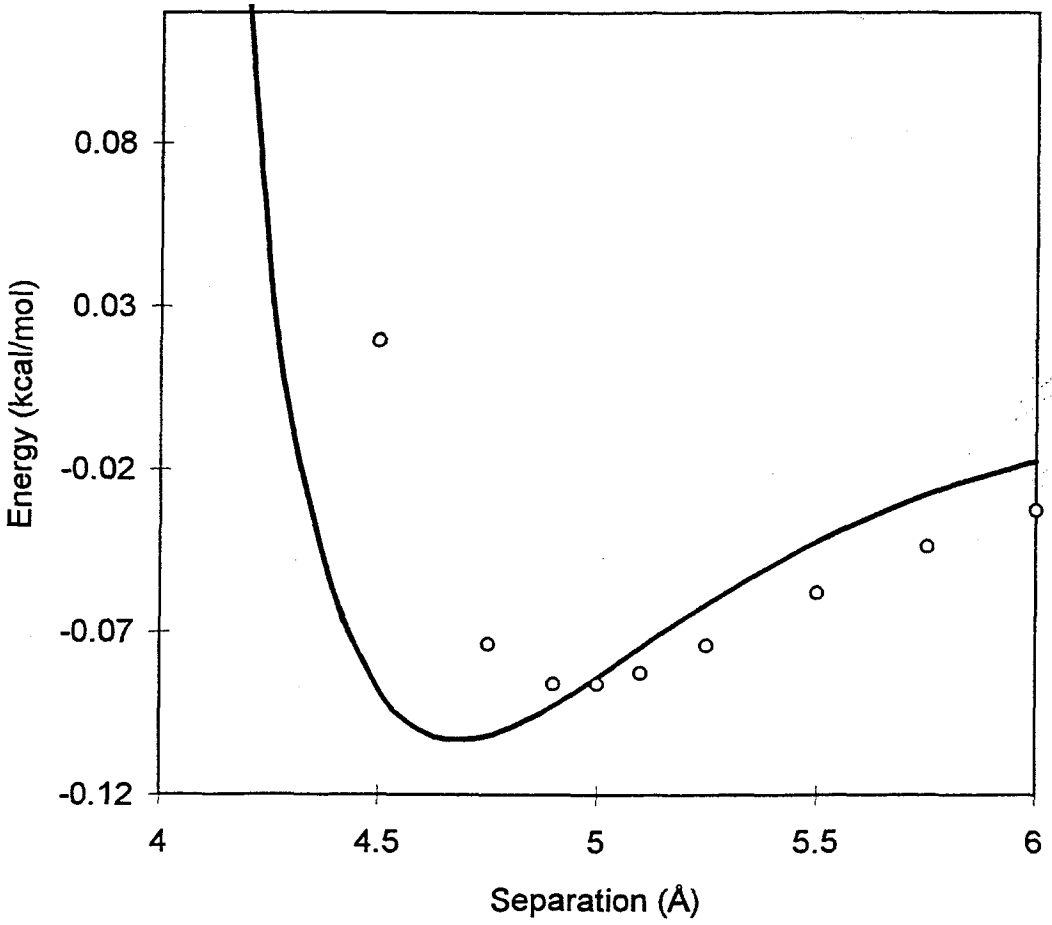


Figure II.6 Transfer of the COMP CRS potential to the LIN geometry. The solid line is the LIN potential calculated using CRS parameters from Table II.18 and the dots are *ab initio* data points for the LIN cluster.



Chapter III. Applications and Force Field Development.

III.1. Introduction

Here the COMP method is applied to several atoms in the first and second rows of the periodic table. Comparisons are made based on empirical potentials (for the first row especially) and based on force field calculations of properties of materials. The point of this chapter is that the COMP potential can be applied to a variety of atoms and, in the cases where it is available, match empirical fits. Topological features of these potentials are examined. It is found that contrary to usual force field design, the best functional fit to these *ab initio* results is an exponential-exponential (EXEX) function. Ths EXEX function and the Morse function, a special case, describe the COMP potentials better than the exponential-6 (E6) or Lennard-Jones (LJ) which are usually used for this purpose.

There are also intriguing trends in the data. The range parameters of the monotonic potentials are found by curve fits. They correlate with atomic properties, namely the ionization potential (IP). So, the IP can be used both to understand and to predict van der Waals parameters. Moreover, a model is proposed for varying the potential according to atomic charge in polarized systems. Since, for instance, hydrogens in hydrocarbons are always slightly positive and oxygens are negative, modifying the van der Waals potentials, which are based on neutral systems in the quantum mechanical model, can improve accuracy. Also the data here give a sound basis to standard combination rules. Some potentials were calculated for heterogeneous cases and compared to the prediction that would be made by standard combination rules. What is found is that the combination rules work best when applied to the monotonic potentials separately. The exponential curves combine in the standard manner.

III.2. Computational Details

The COMP method developed in Chapter I is applied directly to other atoms. Changes were made in the basis sets to accommodate atoms from the second row and hydrogen which use a different number of functions than the first row atoms. For each case the diffuse midbond functions were optimized. The valence functions come from Dunning[1] as before. Values for the optimized diffuse functions with the midpoint method are in Table II.15 for all of the elements discussed in this chapter. Also for hydrogen and the second row atoms, the valence and polarization functions were different than for the first row calculations. For hydrogen, DZ means $(5s/3s)$ and DP means $(2p1d)$. The midpoint function is $\{2s2p\}$ rather than $\{2s2p2d\}$ for the heavy atoms and the midbond function is $\{2s2p\}$ just like for the heavy atoms. Also DZDP$^+$ for hydrogen adds an extra s function only. For the second row DZ valence is $(12s8p/4s3p)$ rather than $(9s4p/3s2p)$ while the DP polarization set is still $(2d1f)$.

The computational cost for the second row is far worse than for the first row. Hydrogen, however, is exceptionally cheap. Each point (including the correlation and the repulsive energy calculations for both normal and BSS energies) takes 200 seconds on the HP workstations.[2] Each point for $(Cl_2)_2$, on the other hand, takes 660 minutes. Calculations on third row atoms and metals will probably be done with effective core potentials in the future considering the large increase in cost going from the first to the second row.

The electronic states of the diatoms[3,4] were investigated in order to choose the spin state of the calculation. The electronic state appropriate for N_2 in $(N_2)_2$ was the ground state $^1\Sigma_g^+$, which makes sense for any molecule with a closed shell ground state. Among the diatoms studied here H_2, N_2, F_2, P_2, and Cl_2 have closed shell ground states which are germane to van der Waals interactions in simulations. Moreover, these are all in the $^1\Sigma_g^+$ ground state. It is in general possible that different electronic states for a single diatom are relevant to different situations in simulations. For instance, a $^3\Sigma$ state in carbon may be relevant to sp-hybridized

systems while $^7\Sigma$ may be more closely related to saturated systems. Electronic states were chosen here to have the most relevance to a broad range of simulations.

For carbon the $^3\Sigma_u^+$ state C_2 was chosen because it has a π-electron system and is relevant to graphite for which a lot of experimental data is known. This state represents the orbitals in the C_2 biradical resulting from homolytic cleavage of C-H bonds in acetylene. An equilibrium bond distance of $r = 1.23$Å was used. For nitrogen the $^1\Sigma_g^+$ state was used, being the ground state for N_2. This is, of course, the appropriate state for modeling $N_{2(gas)}$ and N_2 diffusion, but for N in molecules, however, the hybridization of nitrogen orbitals may be different than in N_2. Oxygen does not have a closed-shell ground state. Nevertheless, the ground state, $^3\Sigma_g^-$ is used. This state is also the ground state of sulfur and is used here for S_2. For silicon, the open-shell ground state, $^3\Sigma_g^+$, is used here. The repulsive potential is calculated for both the triplet and septet states of Si_2 for comparison. The states used in the calculations here and their equilibrium bond lengths are listed in Table III.1.

Table III.1. Selected Electronic States of the Diatomics

Atom	State	l_{bond}(Å)
H_2	$X^1\Sigma_g^+$	0.74152
C_2	$c^3\Sigma_u^+$	1.23
N_2	$X^1\Sigma_g^+$	1.0977
O_2	$X^3\Sigma_g^-$	1.2074
F_2	$X^1\Sigma_g^+$	1.4119
Si_2	$X^3\Sigma_g^-$	2.246
P_2	$X^1\Sigma_g^+$	1.8931
S_2	$X^3\Sigma_g^-$	1.8892
Cl_2	$X^1\Sigma_g^+$	1.897

III.3. Quantum Potentials: Main Group Elements

The repulsive and attractive potentials were fit to monotonic functions, the first of which was the exponential

$$Ae^{-CR}. \qquad (III.1)$$

Exponential fit parameters illustrate trends in the topology of the potentials from atom to atom which can be traced to atomic properties such as the IP which gives the parameter C directly, atomic excitation energies to characterize the dispersion interaction, and the electron affinity to characterize the variation of these parameters with respect to atomic charge. Table III.2 is a compilation of parameters.

Table III.2. Exponential Fit Parameters for *ab initio* Potentials

Atom	A	C	B	D
H	2424.75	3.4012	9.3596	1.6944
C	5271.94	2.6173	144.93	1.5988
N	39,649.85	3.5466	158.90	1.7623
O	36,139.71	3.7863	161.01	1.9653
F	70,771.01	4.3552	23.57	1.7036
Si	40,023.44	2.5609	172.39	1.2771
P	18,937.75	2.5418	327.65	1.3706
S	73,757.02	3.0451	353.64	1.5097
Cl	298,507.6	3.2250	294.18	1.5980
Ar	191,695.9	3.5941	270.46	1.6598
C-O	17,440.86	3.2020	152.46	1.7710
N-F	66,931.51	4.0094	58.57	1.7338
P-Cl	57,862.48	2.9193	249.62	1.4801

The fits of monotonic potential data can be improved by using a Rydberg

function

$$E^{Rydberg}(R) = Ax^{-\zeta R}(\sum_{i=0}^{n} p_i R^i) \qquad (III.2)$$

to fit both the repulsive and attractive potentials. There is some curvature in the logarithm of the monotonic energies which arises from the two origins. First, the two curves are a culmination of interactions between molecular orbitals on each monomer. The repulsive energy arises from the Pauli repulsion between sets of orbitals on each monomer with a range of characteristic decays or ζ. So the repulsive energy has a composite of range parameters within it. Second, the energy is actually derived from the overlap which is itself, for atomic orbitals, a Rydberg function. There are angular components for orbitals with $l \neq 0$ from quantum mechanics, but these are ignored in this simpler treatment as the current thesis tests the application of isotropic pair potentials. The problem of curvature is even more pronounced with the attractive potentials because they arise from interactions of the more diffuse excited states. Hence, a smaller effective C or, equivalently ζ, means that R is farther from the asymptotic region where the highest order polynomial term in the overlap expression dominates.

The Rydberg function of order two (RYD2)

$$E^{\mathrm{RYD2}}(R) = Ae^{-CR}(1 + p_1(R + p_2 R)) \qquad (III.3)$$

is used to improve the fit of monotonic potentials. As is shown in Figures II.1 and II.2, the RYD2 fit improves the fit of the attractive potential, but not so much the repulsive, which is fit well by the pure exponential function. Since this function is difficult to use in simulations, however, and since the gain in accuracy over the pure exponential is modest, the latter is preferred here.

Functional fits of the total COMP potentials are obtained for the common potential functions LJ 12-6 and E6. The parameters and quality of the fit are summarized in Tables III.3 through III.6. In addition, unconventional functional forms are used in order to obtain a more accurate fit, the Morse and EXEX functions

Table III.3. LJ Parameters

Atom	R_e	D_e	RMS (x10^3)
C	3.894	0.0981	1.470
N	3.641	0.0859	3.615
O	3.328	0.0964	2.091
F	3.443	0.0328	6.956
Si	4.768	0.2230	9.861
P	4.196	0.4146	0.490
S	3.904	0.449	0.087
Cl	3.767	0.328	1.284
C-O	3.710	0.0870	22.976
N-F	3.468	0.0659	2.076
P-Cl	4.217	0.219	32.27

(Appendix A). Note how the error in the fit is improved at least threefold in each case from a Morse fit to a generalized Morse fit as was observed previously with nitrogen alone.

The potential function parameters are given in the series of tables for the four functional forms LJ, E6, Morse, and EXEX. Just as for nitrogen, the EXEX function fits better than any of the other functional forms. As a rule, the more parameters the better, for LJ 12-6 is the worst in all cases and the two three-parameter potentials are in the middle for rms error of the best fit. These fits are illustrated in the figures of Appendix B.

Table III.4. E6 Parameters

Atom	R_e	D_e	ζ	RMS (x10^3)
C	3.919	0.0997	13.597	0.825
N	3.633	0.0886	14.750	1.444
O	3.296	0.1030	15.604	0.974
F	3.325	0.0489	17.786	3.897
Si	4.736	0.2199	14.404	2.782
P	4.288	0.3933	11.877	0.019
S	3.908	0.456	14.192	0.020
Cl	3.705	0.396	16.724	0.091
C-O	3.718	0.0873	13.827	0.270
N-F	3.437	0.0701	15.444	1.437
P-Cl	4.246	0.217	13.702	3.47

The most illuminating part of the parametric fits is the patterns between elements and the universality of the topology of the binding curve. Figures III.1 and III.2 show the potential functions all together on the same scale for the first and second row elements. There are, of course, marked differences between them which are also apparent from a perusal of the parameters in Tables III.3 through III.6. Figures III.3 and III.4 show scaled curves all together and then it is apparent that scaling out the values of R_e and D_e of each curve gives a universal curve with $R_e, D_e = 1$. This curve can be scaled back to a real curve then, given parameters for a new atom pair. For these figures a third parameter, ζ, was scaled and set to 12 (see Appendix A).

Now a force field can be designed which uses the average of a set of scaled curves as a numerical universal curve and refers to a table of parameters to generate numerical potentials by rescaling the universal curve for a particular atom pair. The

Table III.5. Morse Parameters

Atom	R_e	D_e	ζ	RMS (x10^3)
C	3.994	0.0990	10.963	1.338
N	3.641	0.0896	12.381	1.227
O	3.328	0.1073	12.480	1.011
F	3.364	0.0483	13.535	0.630
Si	4.767	0.2230	12.243	0.698
P	4.308	0.3972	10.811	0.002
S	3.932	0.455	11.800	0.008
Cl	3.789	0.407	12.146	0.109
C-O	3.728	0.0898	12.272	0.495
N-F	3.468	0.0749	12.748	0.0835
P-Cl	4.238	0.226	15.523	1.04

parameters would be fit to data for atom pairs which are included in the initial data set. Then some rules can be used to generate parameters for unknown pairs. To illustrate the robustness of the EXEX fit, the parameters can be generated from (A,C,B,D) values of Table III.2 using Equation A.31. The resulting EXEX function is compared to the direct fit in Table III.6 in Figure III.5.

Böhm and Ahlrichs[5] (BA) calculated potentials for twelve atoms and observed systematic dependence of fitted functional parameters on the IP. They performed first-order SCF calculations to give repulsive potentials. This approach orthogonalizes the density matrices of each monomer in the cluster calculation, a way to treat BSS interactions. The potentials they calculated were done at four or five points and over a range in energy from about 0.1 kcal/mol to 24 kcal/mol. In keeping with the purpose of the present study they chose energies for chemical interest, for simulations at temperatures below 1000K. The curves were all fit by a

Table III.6. Exp-Exp Parameters

Atom	R_e	D_e	ζ	ξ	RMS (x10^3)
C	3.942	0.0995	11.456	2.998	0.371
N	3.646	0.0891	12.389	2.640	0.212
O	3.314	0.105	12.645	1.947	0.354
F	3.372	0.0497	13.359	1.633	0.080
Si	4.772	0.2231	11.834	1.879	0.074
P	4.306	0.3969	10.812	2.078	0.009
S	3.922	0.456	11.939	2.527	0.001
Cl	3.742	0.404	12.758	2.931	0.009
C-O	3.722	0.0880	12.046	2.742	0.118
N-F	3.467	0.0748	12.747	2.036	0.0783
P-Cl	4.237	0.228	12.617	1.846	0.930

simple exponential curve, Equation III.1, with an error of 0.025 kcal/mol. Fitted exponential parameters are listed in Table III.7.

Trends found in the first and second row cases studied here could be used to generate parameters for new atom pairs since trends have been found based on atomic properties. Fits of the exponential parameters, $A, B, C,$ and D to some function of IP have been know to give some correlation.

$$\begin{cases} A(\sqrt{IP}) = b_A + m_A\sqrt{IP} \\ C(\sqrt{IP}) = b_C + m_C\sqrt{IP} \end{cases}. \qquad (III.4)$$

This is especially so for the range parameters which corresponds with the size of orbitals. The repulsive one, C, is proportional to $\sqrt{IP}$ from elementary considerations.[6] The fits are performed using two IP values, the atomic IP and the valence averaged IP (VAIP) from Rappé and Goddard.[10] The VAIP (and the same goes for VAEA) is designed to represent the effective IP of an atom in a

Table III.7. Exponential Fit Parameters (Adapted from BA[5])

Atom	$A(kcal/mol)$	C(Å)
H	1631.526	2.630883
C	31814.76	2.761293
N	44992.47	3.239463
O	61433.23	3.711964
F	82956.82	4.192024
P	107931.7	2.453222
S	90173.19	2.844453
Cl	121485.9	3.163863
Ar	206074.3	3.589114

molecule. Chemical bonds alter the valence orbital structure of some atoms and, hence, the IP. For example, Be has an s^2 ground state. In some molecule like BeH_2 the hybridization changes to give Be a state more like $(sp)^2$ which has a different IP than the free atom. One result, as in Table III.9, is that the VAIP numbers are a monotonic function of atomic number even when IP is not. Since these clusters use molecules rather than free atoms (except, of course, for the noble gases) the VAIP should be tried. Linear correlation here can be used to extend the set of parameters known for a given row.

Unfortunately, the fits are flawed for the second row. If the fit is done only to P, S, and Cl as BA[5] have done then the fit yields a negative value for C of silicon. To avoid this absurdity, silicon may be included in the fit set, but then the problem is that both sets of IP values are monotonic in atomic number while the list of C values including silicon is not. The fit is irredeemably poor. Furthermore, this approach is limited because the fundamental energy expression which relates $\sqrt{IP}$ to the parameters are derived only for simple systems. The wavefunctions of the

Table III.8. Linear Fits of Parameters

Parameter	b	m	Avg. Error
A^{1strow}	-194,504	62.501	4589
A^{1strow}-VA	-194,504	61,930	6334
C^{1strow}	-1.258	1.344	0.260
C^{1strow}-VA	-2.498	1.600	0.093
D^{2ndrow}	0.3514	0.2738	0.0308
D^{2ndrow}-VA	0.3264	0.4297	0.0118

diatoms in the present thesis have more features than the systems in the textbook problem. For instance, there is angular dependence of the wavefunction around the bond axis and there are fluctuations in amplitude of different orbitals along the bond axis. This complicates the simple $\sqrt{IP}$ analysis and the present calculations, being more accurate than those of BA, demonstrate this intricacy.

A curious fact that is discovered with these fits is that while the range parameter, C, is difficult to fit for the second row, the attractive range, D, is fit well. While the equivalent fit for the first row is poor, only the second row fit is listed in Table III.8.

When this fitting is successful, or in cases that calculation may augment the $\sqrt{IP}$ dependence, these relationships provide a way to account for the effect of polarization on the van der Waals potential. It is known from calculations that the changes in atomic charge change, in turn, the van der Waals radius of an atom. Hart and Rappé[7] have observed with H-A molecules that when A is more electronegative than H, the van der Waals radius of H is diminished. Accordingly, when A is electropositive, such as an alkali metal, the radius of H is increased. This makes sense in that the removed electron density in a partially positive H atom decreases the amplitude and range of the valence wavefunction, making R_e decrease.

Table III.9. IP and EA Values for Selected Atoms.

Atom	IP (eV)	VAIP	EA(eV)	VAEA
H	13.59844	13.59844	0.754	0.754
He	24.58741	24.58741	...	...
C	11.26030	8.298	1.2629	0.2797
N	14.53414	10.406	-0.07	1.0188
O	13.61806	12.778	1.4621	2.0593
F	17.42282	15.422	3.399	3.399
Si	8.15169	5.986	1.385	0.681
P	10.48669	7.655	0.7465	1.463
S	10.36001	9.453	2.0771	2.442
Cl	12.96764	11.413	3.617	3.617
Ar	15.75962	13.510	...	...

Roughly then the linear dependence of C on $\sqrt{IP}$ gives a way to predict this change. One can construct an effective IP for a charged atom which is greater than the neutral atom IP (IP^0) for $+q$ and less than IP^0 for $-q$.

$$IP(q) = IP^0 + \frac{1}{2}(IP - EA)q \qquad (III.5)$$

This linear equation gives an effective IP(q) in terms of atomic properties, the IP and the EA. The coefficient of charge in Equation III.5 is called the idempotential, $\eta = \frac{1}{2}(IP - EA)$. The potential can be modified by the charge by calculating C for $IP(q)$. It is convenient to relate potential parameters to an atomic property such as IP about which much is known.[8-10] This charge modified van der Waals (CMVDW) approach can be implemented in simulations software in the future to yield more realistic potentials.

Combination rules are an important aspect of the *ab initio* potentials. They are frequently used in simulation because they allow the amount of stored param-

eters to be conserved. That is, only homogeneous potentials need to be known and heterogeneous potential parameters can be determined quickly by the likes of Equation II.7. Tables III.3 through III.6 give the functional fits to the heterogeneous potentials calculated with the COMP method. With both the homogeneous and heterogeneous COMP potentials available, the test is to apply combination rules to a pair of homogeneous potentials and compare it to the heterogeneous case. It can be seen from the parameter tables (Tables III.3 through III.6) that the parameters do not combine very well. What does combine better are the monotonic potentials. Actually they combine separately and the final curve can be assembled from the combined potentials.

The first step is obtaining the mean exponential parameters with the following combination rule

$$\begin{cases} \bar{A} = \frac{1}{2}(A_1 + A_2) \\ \bar{C} = \sqrt{C_1 C_2} \end{cases} \qquad (III.6)$$

where $\bar{A}$ and $\bar{C}$ are the parameters for the combined monotonic potential. The same treatment is applied to $\bar{B}$ and $\bar{D}$.

Table III.10. Combination of Exponential Fit Parameters

Atom	$\bar{A}$	$\bar{C}$	$\bar{B}$	$\bar{D}$
C-O	13,803.12	3.2108	152.76	1.7821
N-F	52,972.25	3.9509	52.26	1.7330
P-Cl	78,857,80	3.0725	290.44	1.4668
(C-O)	17,462.18	3.2108	152.76	1.7821
(N-F)	55,460,64	3.9509	61.20	1.7330
(P-Cl)	110,608.32	3.0725	235.90	1.4668

What Table III.1 shows is that by combining the exponential parameters a closer match is obtained between the mean values and the actual parameters from

the *ab initio* potentials. What is also striking is that the match in exponential prefactors, especially the repulsive one, is rather poor while the match with exponential range parameters is far better. Since there is really one problem parameter, A, a refinement of this monotonic combination rule (MCR) is to rescale $\bar{A}$ with a single value from the calculated repulsive curve. The idea is that the calculation of the combined energy at just one point is a small price to pay for accurate combined potentials. If this is viable then it is still far better than explicitly calculating the combined potential for every heterogeneous pair.

Here it seems that only the A parameter needs any special treatment. This indicates that the only heterogeneous *ab initio* calculation that needs to be done is a repulsive energy. This is encouraging because the primary expense of COMP calculations is the correlation energy. The HF energy takes a small fraction of the time (between about 10% and 30%). The scaling works in a straightforward way. The scaling point R_{scal} is chosen to be R_e or the closest calculated point. A scaling coefficient is developed from the repulsive energies

$$\lambda = E_{rep}^{\mathrm{X-Y}}(R_{scal})/\bar{E}_{rep}(R_{scal}), \qquad (III.7)$$

where the $\bar{E}$ potential is not an *ab initio* potential but, rather is computed using the combined parameters above (Table II.10). This factor is then applied to the $\bar{E}_{rep}$ potential. The fits give much more accuracy as the scaling brings the monotonic potential much closer to the *ab initio* heterogeneous potential.

The second set of parameters in Table III.10 is scaled by λ from Equation III.7. Figure III.6 shows how this improves the fit significantly. This scaling can be applied to the attractive potential as well as the repulsive (Figure III.7). In the case of C-O, $\bar{B}$ is left unscaled because the combination rule potential is already very close to the *ab initio* potential (as would be seen by the one energy point comparison), but it is scaled for the other two cases. The scaled A and B values in Table III.10 actually differ from the *ab initio* values in Table III.2 because the scaling does not necessarily correct the prefactor entirely from the comparison of

a single point. What happens is that because the range parameters differ, the prefactor should not match exactly. Perhaps a refinement of this scaling would use two points of the heterogeneous *ab initio* potential to correct both parameters simultaneously. For now, however, this simple one point scaling works remarkably well.

Table III.11. Combination of Total Potentials.

Case	R_e	D_e	ζ	ξ
CO	3.726013	0.092811	12.54803	1.80802
NF	3.47013	0.079677	12.93884	2.312555
PCl	4.255881	0.226158	12.51106	1.972366
CO	3.584948	0.113837	12.11037	1.796696
NF	3.492046	0.069082	12.92211	2.27991
PCl	3.950458	0.46215	11.86009	2.094759
CO	3.750569	0.084743	12.66986	1.796696
NF	3.483954	0.074643	12.89216	2.27991
PCl	4.29069	0.227893	12.88154	2.094759

The actual total potentials resulting from the combination are described by the EXEX parameters in Table III.11. There are three sets of data within the table. First, a recombination of monotonic parameters gives EXEX parameters with the use of Equation A.31. These match well the parameters fit directly to the COMP data in Table III.6. Second, data from given by Equation A.31 once again but using monotonic parameters given by the straight combination rule in Equation III.6. These deviate from the desired numbers, especially in D_e. Third, the scaled monotonic parameters are used to produce EXEX parameters which give an outstanding match to the direct fits (Figure III.8).

The results of this section are very pertinent to simulations and the choice

of potential functions. The overall result is that the EXEX function is shown to be the best for fitting the COMP potentials. This is tied to the basic monotonic, exponential potentials, because the EXEX function, more than the others, is built from two exponential functions. The exponential parameters are themselves correlated to atomic properties. The monotonic potentials are also valuable for combination rules because combining is more obvious with the monotonic curves. Moreover, accurate combination rules are obtained if the scaling procedure is applied.

III.4. Quantum Potentials: Hydrogen

Hydrogen is treated differently from the other elements in this study. For one, its calculation was set up differently than the others because it needs a different basis set. More importantly, the van der Waals potential of hydrogen (and of helium) is qualitatively different from those of heavy elements. The equilibrium binding energy and long range dispersive tail are both exceptionally weak. Accordingly atomic hydrogen has a low polarizability (Table III.12). Because of these properties, the regular COMP method has problems with hydrogen.

Table III.12. Polarizabilities (Å^3) of Some Atoms and Diatoms[11]

Atom	α_{atom}	α_{diatom}
H	0.666793	0.8023
He	0.2050	...
N	1.10	1.7403
O	0.802	1.5812
F	0.557	1.38
Ne	0.3956	...
Cl	2.18	4.61
Ar	1.6411	...
Br	3.05	7.02

COMP potentials for hydrogen were calculated in the usual manner using HF/DZDP$^+$ for $E^{rep}(R)$ and MP2/DZDPDZ for $E^{att}(R)$. Table B.21 shows the results over the relevant range of R. Functional fits yield parameters in Table III.13 and are illustrated in Figure B.9. The equilibrium van der Waals distance is around 3.64Å. In contrast, both the empirical potential of Williams and the SA potential from *ab initio* put R_e near 3.2 Å. For hydrogen more than for the heavy elements there is a problem with the COMP approach. The hydrogen potential

is a valid potential for the D_{2d} geometry of the cluster, but this geometry and the chemical environment of the H atoms within the cluster may not be relevant to other systems. For example, hydrogen is charged in all molecules besides H_2. Hart and Rappé have demonstrated[7] how this charge modifies the van der Waals R_e significantly. Attempts to overcome this problem with the standard COMP approach are to modify the COMP potential according to the H charge and to examine potentials for different orientations of the diatoms in the $(H_2)_2$ cluster.

Table III.13. Parameters of Potential Function Fits to the Hydrogen COMP Potential

Potential	R_e	D_e	ζ	ξ	RMS (x10^3)
LJ 12-6	3.6294	0.00860	...	...	6.643
Exp-6	3.6305	0.00876	14.385	...	3.282
Morse	3.6498	0.00893	13.306	...	2.955
Exp-Exp	3.6407	0.00885	12.276	2.605	11×10^{-6}

Because orientational anisotropy might be crucial in modeling the hydrogen potential properly, more potentials were calculated just as for $(N_2)_2$ in section II.3. Three geometries of $(H_2)_2$ were treated with the COMP method to give potentials for different relative orientations of the bonds. Just as for nitrogen, in Section II.3, the CRS, TEE, and LIN geometries are studied (Figure III.9). The resulting potentials (Table III.14) are quite different. The CRS potential has a longer R_e and a smaller D_e compared to the other two. It is likely that the TEE and LIN potentials are more relevant for simulations given the orientation of C-H bonds in many materials. For instance, in polymers and macromolecules there is a backbone of carbon with hydrogens pointing outward. This structure predisposes pairs of H atoms on neighboring molecules to have their respective H-C bonds in a configuration more resembling LIN or TEE than CRS (at least for close contacts such as the closest

Table III.14. EXEX Parameters for $(H_2)_2$ Clusters

Cluster	R_e	D_e	ζ	ξ
CRS	3.6407	0.00885	12.2761	2.6052
TEE	3.3869	0.01496	10.4586	2.7794
LIN	3.2942	0.01153	10.7902	2.3149
CRS1	3.6407	0.008850	12.27605	2.6052
TEE1	3.2765	0.02388	10.12076	2.7802
LIN1	3.0831	0.02780	10.10413	2.3150

H...H pairs).

An accurate van der Waals approach might correct the problem by developing a dependence on atomic charge. It is possible that the COMP potential for hydrogen is inaccurate because it is for neutral H and H found in real molecules is charged (usually positive). For instance, hydrogen atoms bound to carbon have a charge of about +0.15.[12,13] According to the CMVDW model above this charge should modify the van der Waals potential, in this case making R_e diminish. Taking the exponential fit of the repulsive He_2 potential, the repulsive range parameter is fit to $\sqrt{IP}$. The result is $C = m\sqrt{IP} + b$ with $m = 0.806783$ and $b = 0.426176$. For $q_H = +0.15$, the $IP(q)$ is increased from 13.598 to 15.525 eV and the range parameter is increased, in turn, from 3.400 to 3.605. The final effect this has on the potential is found using the EXEX parameter equations from section A.1. The R_e is reduced to near the DREIDING value. The final parameters with CMVDW are (3.20, 0.0220, 11.17). For simplicity only the C parameter is charge modified because it has the clearest dependence on $\sqrt{IP}$ and has the most influence on the equilibrium distance.

Hydrogen is an interesting case because there is exact data on its electronic states. Kolos and Wolniewicz[14] (KW) have calculated the energy and first derivative

of H_2 over a large range of distances for the ground state, $X^1\Sigma_g^+$, and the b state, $^3\Sigma_u^+$. These potentials provide a basis for comparison with the COMP curve because a van der Waals potential can be constructed from the KW data. Any pair of hydrogen atoms which have nonbonded interactions will each be bonded to some other atom (free atoms are ignored in the present study). Statistically the hydrogen atoms will be coupled as either a singlet or a triplet. In fact, they will be a singlet $\frac{1}{4}$ of the time and triplet the other $\frac{3}{4}$. So, combining the exact potentials for the X state, $E^{S=0}(R)$, and the b state, $E^{S=1}(R)$ in the proportions

$$E(R) = \frac{1}{4}(E^{S=0} + 3E^{S=1}) \qquad (III.8)$$

makes a spin averaged (SA) van der Waals potential. The approximation is that the KW data is for atoms bonded to each other while in reality the bonds are of a different nature (most simulations will be done on systems other than $H_{2(xtl)}$ and they are oriented in any number of different ways. Most importantly, the valence orbitals of hydrogen atoms will be shaped differently in materials than in the H_2 molecule.

Be that as it may, this provides an excellent opportunity to test a molecular potential against a van der Waals potential. The SA potential is fit by a Morse potential to give parameters of $(3.229, 0.026, 9.819)$. Williams, on the other hand, has $(3.195, 0.0152, 12.382)$ with an E6 function.[15] Also Karasawa and Goddard have fit an E6 hydrogen potential using the properties of polyethylene[13] with the result of $(3.167, 0.020, 11.2)$. In comparison to the empirical potentials, the SA method is possibly a viable technique.

Because tests with diatomic fluorine were both difficult and inaccurate at the CASSCF level, this approach is not pursued here, but suggested as an alternative to COMP. It is possible that SA potentials will eventually be easier than calculations of the COMP type because only two atoms are involved. It also makes sense that the forces probed by an electronic structure calculation of a single diatom will reveal all that is necessary to model the pair interaction in simulations. The effort of refining

this approach is uninviting at the present time for two reasons. First, optimization would require a higher level of wavefunction to obtain accurate energies in general. Second, and more forbidding, is the choice of electronic states to study. Already with fluorine there are several symmetries of triplet state which are relevant. Also with heavy atoms there are more spin states to average than just the singlet and triplet states of hydrogen. Furthermore, since many diatoms have open shell ground states, the choice of wavefunctions to average over is somewhat arbitrary.

The superiority of the COMP approach is that the interaction energy of the clusters is directly related to a van der Waals interaction. Also the refinement of the SA approach necessary to obtain an accurate method are considerable. There are also inherent limitations in the SA approach that are not present with COMP. For instance, it is possible, and even assured in some cases, that the pair potentials need to be modified by information about the orientation of bonds because van der Waals interactions in materials are anisotropic. This cannot be done in any straightforward manner with SA calculations. Second, there is less ambiguity in the choice of electronic states with COMP calculations because the molecules which comprise the cluster are in their ground state (or a judiciously chosen excited state) while with the SA calculations, it is not clear how to choose the states relevant to simulations.

Force field tests[16] with the COMP hydrogen potentials show that the standard potential, from the twist geometry, is not appropriate for hydrocarbons. With $R_e \approx 3.90$ Å it gives the wrong structure for polyethylene crystal. The computed density is too low. For comparison, the experimental structure of polyethylene crystal[17] has a *Pnma* cell with $(a, b, c) = (7.1946\ \text{Å}, 2.5443\ \text{Å}, 4.7989\ \text{Å})$ and all angles are 90°. Optimizing the structure with the COMP potential gives a unit cell volume of 122.74 Å^3 compared to 87.84 Å^3 for the observed structure. This enlargement of the unit cell is almost evenly divided into a 14% stretch in a and a 17% stretch in c. These are the two directions dominated by van der Waals forces while b is the chain axis and is defined both by van der Waals forces and by the

valence forces which define the hydrocarbon backbone. This cell parameter is too large as well, by 4.7%. That error is also due to the fact that the valence force field, DREIDING,[18] has not been optimized for these van der Waals potentials. Assuming that optimization could be done, and the work of Karasawa and Goddard[13] demonstrates that this is a straightforward task, the b lattice parameter can always be correct for the right valence potentials. Then the effective error made by the COMP potentials is not 40% but rather 33%. Nevertheless, this van der Waals potential for hydrogen is not optimal.

The orientation of the diatoms in the *ab initio* cluster actually changes the potential quite a lot. EXEX functions are fit to the LIN and TEE structures in Table III.14. Those are much closer to the Williams values in Table III.14 and the structure of polyethylene crystal is simulated more accurately by these than the standard COMP potential. When optimized with molecular mechanics[17] the LIN potential gives polyethylene cell parameters of $(7.377, 2.552, 4.935)$ and TEE gives $(8.141, 2.669, 5.649)$. While the density is still too low, the forces are clearly more accurate if several orientations are used in the *ab initio* modeling instead of just one. Moreover, the default orientation in the present study is the worst one for hydrogen. Since hydrogen has a low atomic polarizability there is a large difference between the polarizability parallel and perpendicular to the bond in H_2. This difference is also present when hydrogen bonds to heavy atoms. The result is that the van der Waals attraction is stronger than *ab initio* calculations predict and equilibrium distances are, hence, shorter.

It is a snag in the *ab initio* methodology of the present thesis that hydrogen is the most problematic case because it is, in simulations, the most prevalent. There are few material simulations which do not have hydrogen atoms. More importantly, molecules tend to have hydrogen atoms on the exterior so that contacts between molecules involve hydrogen disproportionately. These contacts, being the closest atom pairs, dominate the intermolecular interactions in macromolecule and material systems. Hence the hydrogen potential is the most important van der Waals

potential in a molecular mechanics force field. For the future, in order to make the *ab initio* COMP approach more useful as a tool in the design of force fields, the hydrogen van der Waals potential should be improved. On the other hand, the problem is manageable at the level of the present study because hydrogen is a special case. If hydrogen must be treated *ad hoc* with a special potential, an empirical potential, or a specially constructed *ab initio* potential such as the SA one above, that is acceptable because the COMP potentials can still be used for the remaining elements.

Another peculiarity of hydrogen which is very important for modeling is the hydrogen bond. This also is neglected by the COMP calculations. The structure of a hydrogen bond has hydrogen in a polar covalent bond, A-H, where A is nitrogen, oxygen, or even fluorine. Then if the hydrogen is directed at a lone pair on another atom from this same set a hydrogen bond can form. This involves a larger bond energy than is typical for van der Waals interatomic attraction, $\approx$ 1 kcal/mol, about tenfold larger than van der Waals D_e values. Also the distance of the hydrogen atom to the acceptor is smaller than R_e. Typical distances are approximately 2.6 to 3.0 Å between the heavy atoms. The nonbonded distance from hydrogen to the lone pair is about 1.5 to 2.0 Å.

Because van der Waals energies become repulsive at separations about 10 to 15 % less than R_e (*Cf.* Appendix A) the usual potentials cannot model this at all. The problem may be resolved by modifying the van der Waals potential for charge. If the amount of charge transferrred in a hydrogen bond is taken into account it may reduce R_e sufficiently to allow the close approach observed in hydrogen bonded structures. The hydrogen bond, however, is more problematic to model than merely tuning the potential to have the correct R_e and D_e. In weakly associated molecules with hydrogen bonds, the hydrogen moves rapidly between the heavy atoms. Each time it switches the identity of the atom to which it is bonded and the hydrogen bond acceptor. On the one hand, since molecular modeling programs are not designed to break bonds and change connectivity in simulations, the hydrogen bond is not

easy to model. On the other hand, since this oscillation leaves the distance between the heavy atoms unchanged it may be acceptable for structural analysis to model hydrogen bonds while enforcing the approximation that the chemical A-H bond does not break.

Be that as it may, the facile treatment of hydrogen bonding is to declare, as is usually done in physical chemistry, that it is in a category separate from and somewhere in between chemical bonding and van der Waals bonding. That is, the *ab initio* calculation of COMP van der Waals potentials must be narrow in focus and relegate hydrogen bonding to another computational study. A potential which actually fits polyethylene crystal better is CMVDW. Unfortunately, a fit of parameters to $\sqrt{IP}$ is not as solid for H as for first or second row atoms because the H row only has two elements. So, while an accurate fit is assured, its meaning for Equation III.5 is open to question. The fit yields m_C=0.806783 and m_C=0.426176 as above. Given the IP and EA of H and He from Tables III.9 and III.10, the IP for hydrogen in a hydrocarbon is $IP(+0.15)$=15.5246 eV, elevated from the neutral atom number of 13.598 eV. This, in turn, reduces the repulsive range parameter through Equation III.5 and gives a Morse potential with parameters (3.20, 0.0220, 11.17) which is excellent for polyethylene.

The modeling of hydrogen is indeed the most difficult case in the present thesis. Moreover, it is likely the most difficult case in the periodic table. For even if the representation by a pair potential is improved, any errors are compounded because hydrogen is so prevalent and influential in simulations. Just the fact alone that most close contacts in molecules and materials involve hydrogens makes this potential crucial. The status of this thesis is that a pair potential exists which will give accurate modeling results. Unlike in Section III.3 where the COMP potential directly provides the information, however, here several approaches are used to converge on the solution. Because of the limitations of the SA approach the COMP approach, perhaps with orientational sampling, is more promising to develop further. For now the CMVDW potential is used for further test cases.

III.5. Modeling Test Cases

The potentials for carbon, nitrogen, oxygen, and fluorine were tested by implementing them in molecular mechanics simulations of crystals. For comparison, the parameters of the van der Waals energy in the DREIDING[18] force field are listed in Table III.15. Most of these are taken directly from the empirical fit work of Williams and coworkers,[15] but the atoms which have $\zeta = 12.000$ were determined by the designers of the force field, Mayo *et al.* Comparing the DREIDING R_e and D_e parameters from Table III.15 with any of the parameters sets for COMP from Tables III.3 to III.6 shows similarities. The R_e and D_e values for C, N, O, and F have the same ordinal relationship for the COMP potentials and DREIDING. The zigzag trend of D_e, increasing from C to N and from O to F but decreasing from N to O, is present in the COMP and DREIDING parameters.

Graphite is modeled well by the COMP carbon potential. The equilibrium spacing from x-ray crystallography[19] is reproduced to within 0.02Å. The experimental value of $c = 6.67$ Å is given upon minimization. The cohesive energy which provides a test of the well depth of the COMP potential is not available from experiments. Our potential yields 1.170 kcal/mol. Another property controlled directly by nonbond forces is the elastic constant C_{33} which is the compressibility along the crystal axis direction. COMP gives 63.8 GPa.

The minimum energy phase of molecular oxygen and of molecular fluorine both have similar structures. These are the α phases and consist of closest packed layers of diatoms parallel to each other (in $F_{2(xtl)}$ the molecules are tilted from the plane by 18°). Then the planes are stacked so that the lone pairs of molecules point toward each other.

While fluorine crystal is difficult to model with pair potentials, a more important test is fluorocarbons. Real simulations will be concerned with molecules in which fluorine is bonded to something other than another fluorine and most commonly bonded to carbon. Williams found his fluorine potential by fitting the structures of a few aromatic and anti-aromatic fluorocarbon crystals. One is chosen

Table III.15. DREIDING Force Field van der Waals Parameters

Atom	R_e	D_e	ζ
H	3.19500	0.01520	12.38200
B	4.02000	0.09500	14.23000
C	3.89830	0.09510	14.03400
N	3.66210	0.07740	13.84300
O	3.40460	0.09570	13.48300
F	3.47200	0.07250	14.44400
Al	4.39000	0.31000	12.00000
Si	4.27000	0.31000	12.00000
P	4.15000	0.32000	12.00000
S	4.03000	0.34400	12.00000
Cl	3.95030	0.28330	13.86100

Table III.16. Oxygen Potential Test With $O_{2(\mathrm{xtl})}$

Force field	a(Å)	b(Å)	c(Å)	β	ρ(g/cm^3)
DREIDING	5.724	3.282	5.094	135.03°	5.495
COMP	5.705	3.280	5.095	134.32°	5.448
COMP(E6)	5.551	3.214	5.124	135.19°	5.769
Expt	5.403	3.429	5.086	132.53°	5.352

here as a test case, octafluorocyclooctatetraene, which has the benefit of having a varied distribution of angles between bonds. The structure[20] is monoclinic, $P2_1/c$, with cell parameters $a = 9.96(1)$Å, $b = 7.04(1)$Å, $c = 12.66(1)$Å, $\beta = 111.6(1)°$. The comparison is done holding the molecule geometry rigid but letting the unit

cell geometry vary. Also charges are assigned following the prescription of Karasawa and Goddard[21] with a charge of -0.17 on the fluorine atoms and a balancing charge of +0.17 on the carbons. The mechanics is done with a cutoff of 9Å. The minimized structure with the COMP force field using a Morse function was $a = 9.94$Å, $b = 7.08$Å, $c = 12.65$Å, $\beta = 111.8°$. In contrast, with DREIDING, the result is $a = 9.99$Å, $b = 7.18$Å, $c = 12.71$Å, $\beta = 111.6°$. In this case the accuracy of the cutoff was tested by repeating one calculation with a cutoff of 15Å. This increased the length of calculational time by a factor of three, but did not change the result much. The force changed from 13.9987 Å-kcal/mol to 13.9982 and the rms cell gradient changed from 2.8885 to 2.9187. Both force fields yield good results with this structure. The result with COMP is excellent considering that DREIDING was parameterized to fit cases including this one while COMP is purely *ab initio*.

Several established force fields have chosen van der Waals parameters for some of the elements here. These are all based on empirical data. The AMBER and CHARMM force fields were developed for the modeling of biological molecules such as proteins and DNA. As discussed by Rappé and coworkers[22,23] the emphasis of these force fields was on accurate valence parameters and van der Waals parameters were determined using phenomenological approximation or extrapolation from known values. Among these three, listed in Table III.17, Rappé shows that DREIDING performs superiorly, thus demonstrating the importance of accurate van der Waals potentials, perhaps over valence spring parameters. The COMP potentials are quite applicable and useful to force field design. In keeping with the goal of this study the calculation is straightforward with an acceptable computational cost. The accuracy is also more than acceptable considering this is basically a judiciously designed MP2/TZ calculation.

So, the approach is versatile enough to be refined to higher accuracy by directly increasing the wavefunction level and basis set completeness. Because of fundamental accuracy problems such as anisotropy and charge dependence of the van der Waals potential and even the accuracy of the pairwise treatment itself, the

Table III.17. Parameters from Different Force Fields

Atom	AMBER	CHARMM	DREIDING
H	3.08, 0.01	2.936, 0.0045	3.195, 0.0152
H-bond	2.00, 0.02	1.60, 0.0498	3.915, 0.0001
C	3.60, 0.06	3.60, 0.0903	3.8983, 0.0951
N	3.50, 0.16	3.50, 0.238	3.6621, 0.0774
O	3.20, 0.20	3.20, 0.1591	3.4046, 0.0957

higher accuracy COMP calculations are not very important at the present time. For if the accuracy is reduced to, say, less than 0.1%, then with these other issues unsolved there is an inherent inaccuracy which far exceeds the COMP energy accuracy.

References

1. The Correlation Consistent Polarized Valence Basis (cc-pVTZ) Functions for H, He, B, C, N, O, F, and Ne are in T. H. Dunning, Jr., *J. Chem. Phys.*, **90**, 1007 (1989). Functions for Al, Si, P, S, Cl, and Ar are in D. Woon and T. H. Dunning, Jr., *to be published.*
2. Calculation times are listed for the Hewlett-Packard Apollo 9000 model 735 workstation with a clock speed of 100MHz. Calculations were also performed on model 735 computers with a speed of 125MHz giving a 25% speedup. Calculations on a Cray YMP (the Caltech/JPL machine) were between a factor of 2 and 2.16 faster.
3. G. Herzberg, *Molecular Spectra and Molecular Structure I. Diatomic Molecules*, van Nostrand Reinholdt, NY (1953).
4. G. Herzberg and R. Huber, *Molecular Spectra and Molecular Structure IV*, van Nostrand Reinholdt, NY (1971).
5. H.-J. Böhm and R. Ahlrichs, *J. Chem. Phys.*, **77**(4), 2028-2034 (1982).
6. D. A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, CA (1983).
7. J. R. Hart and A. K. Rappé, *J. Chem. Phys.*, **97**(2), 1109-1115 (1992).
8. J. L. Reed, *J. Phys. Chem.*, **95**, 6866-6870 (1991).
9. S. W. Benson, *J. Phys. Chem.*, **93**, 4457-4462 (1989).
10. A. K. Rappé and W. A. Goddard III, submitted for publication. This work is prepared as a two paper series titled, "Generalized Mulliken-Pauling Electronegativities."
11. D. R. Lide, *The CRC Handbook of Chemistry and Physics, 75th ed.*, CRC Press, Boca Raton, FL (1994).
12. A. K. Rappé and W. A. Goddard III, *J. Phys. Chem.*, **95**(8), 3358-3363 (1991).
13. N. Karasawa, S. Dasgupta, and W. A. Goddard III, *J. Phys. Chem.*, **95**(6),

2260-2272 (1991).

14. W. Kolos and L. Wolniewicz, *J. Chem. Phys.*, **43**(7), 2429-2441 (1965).
15. (a) D. E. Williams and D. J. Houpt, *Acta. Cryst.*, **B40**, 404-417 (1984). (b) D. E. Williams and D. J. Houpt, *Acta. Cryst.*, **A37**, 293-301 (1981). (c) D. E. Williams and D. J. Houpt, *Acta. Cryst.*, **B42**, 286-295 (1986).
16. The molecular mechanics and molecular dynamics calculations used POLYGRAF (V3.21 and V3.3) from Molecular Simulation Inc. of Burlington, Massachusetts.
17. G. Avitable, *et al.* , *J. Polym. Sci.*, **13**, 351 (1975).
18. S. L. Mayo, B. D. Olafson, and W. A. Goddard III, *J. Phys. Chem.*, **94**(26), 8897-8909 (1991).
19. J. Donahue, *The Structure of the Elements*, Robert E. Krieger Publishing Company, Malabar, FL, (1982).
20. B. B. Laird and R. E. Davis, *Acta. Cryst.*, **B38**, 678-680 (1982).
21. N. Karasawa and W. A. Goddard III, *Macromolecules*, **28**(20), 6765-6772 (1995).
22. A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, *J. Am. Chem. Soc.*, **114**(25), 10025-10035 (1992).
23. C. Casewit and A. K. Rappé, *Force Fields Across Molecular Mechanics, to be published.*

Figures

Figure III.1 A composite of C, N, O, and F potentials plotted with the same distance and energy scale.

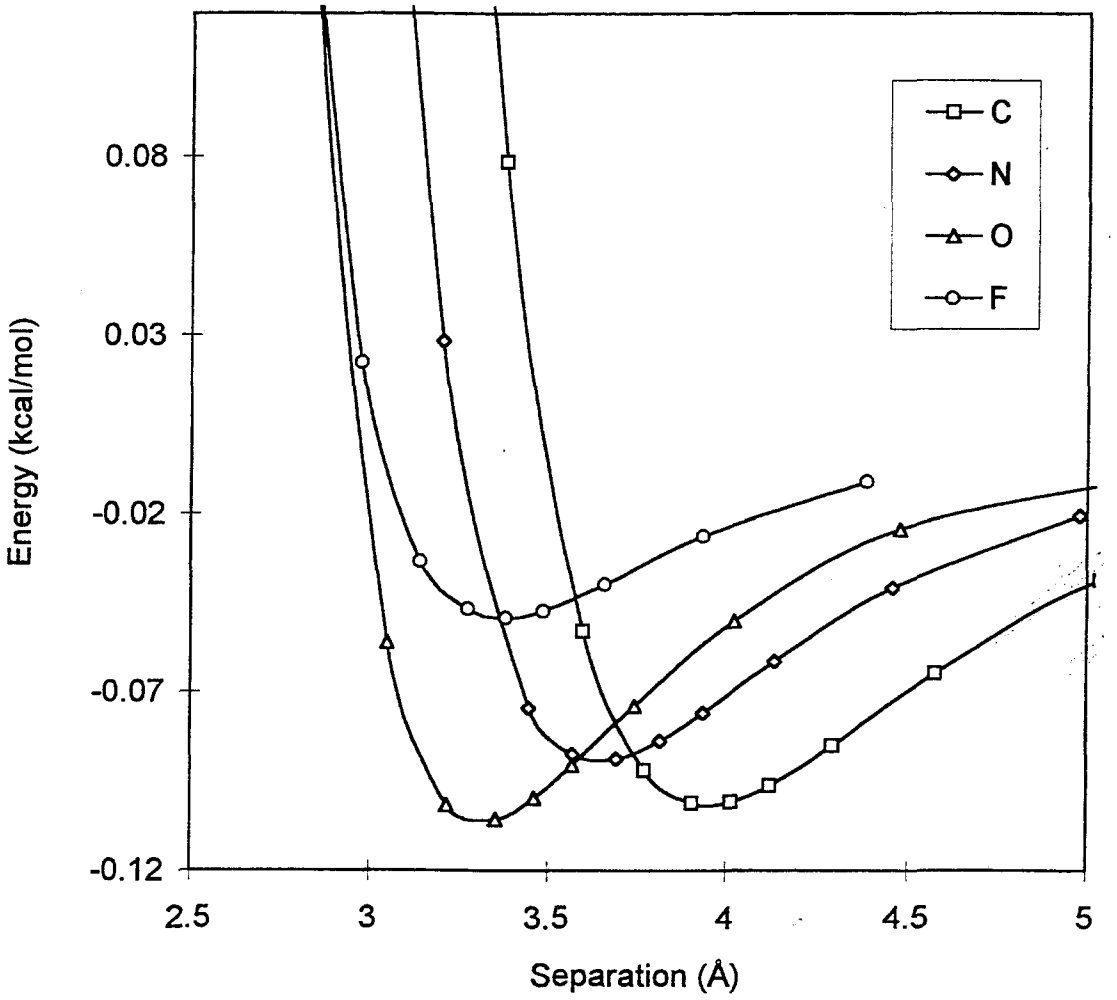


Figure III.2 A composite of Si, P, S, and Cl potentials plotted together with the same energy and distance scale.

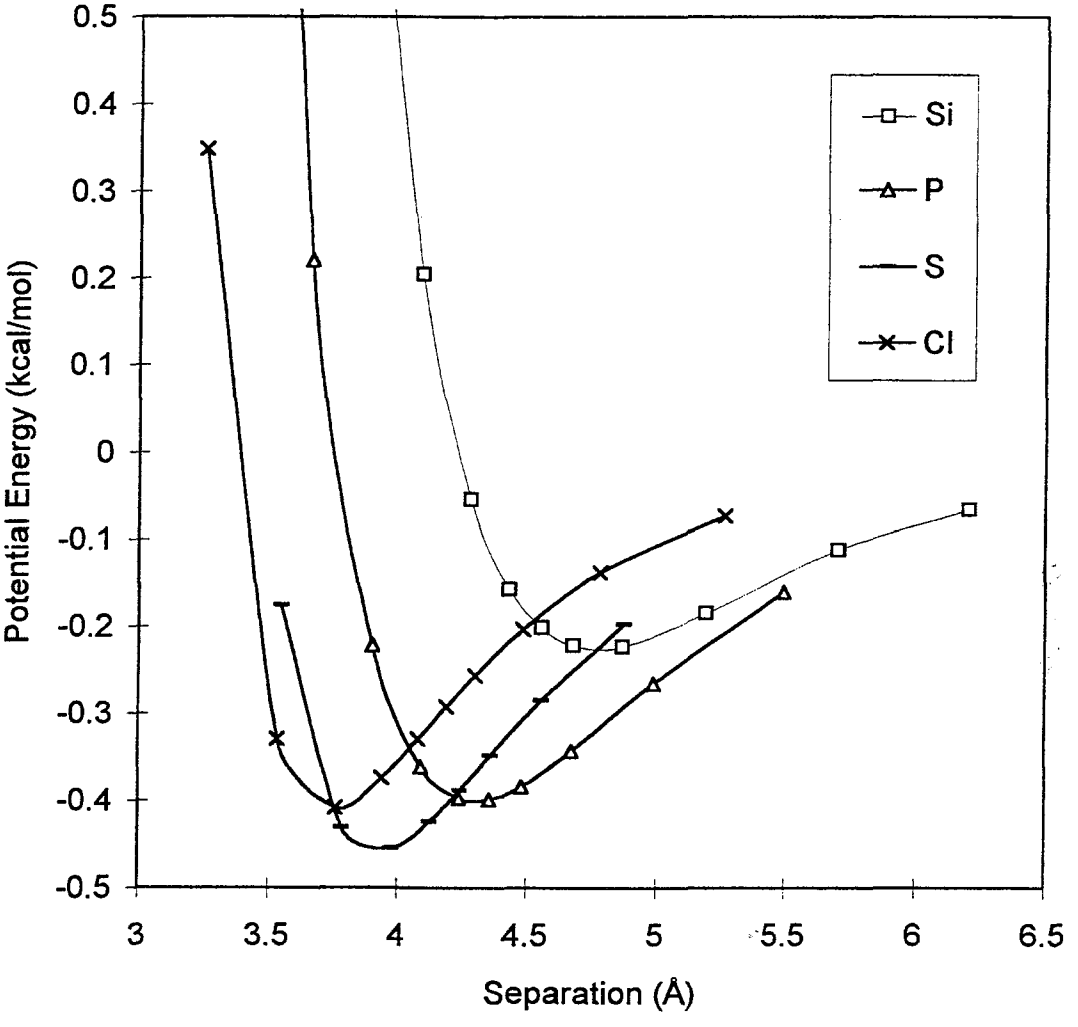


Figure III.3 A composite of scaled potential energy points for C, N, O, and F. The energy and distance are unitless and the curvature is scaled to 72. The points are plotted against a Morse function (line) for comparison.

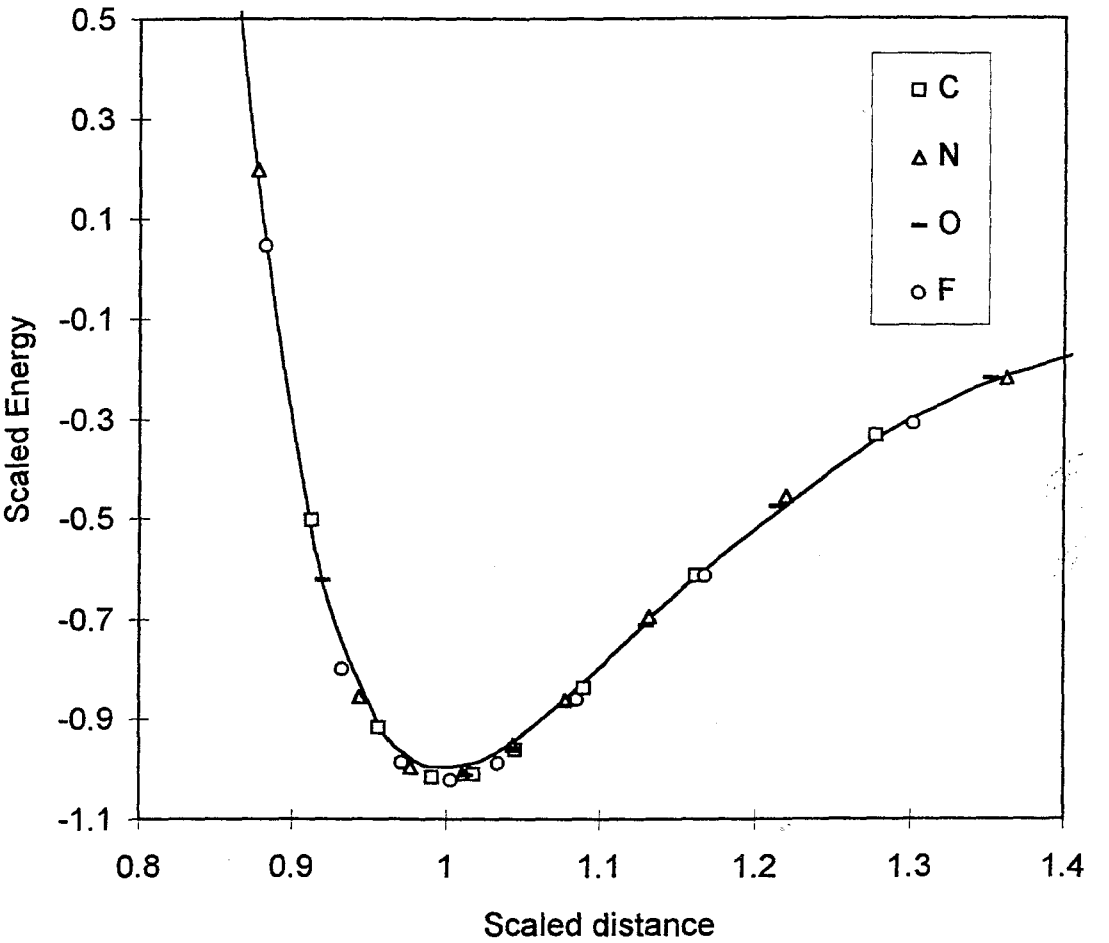


Figure III.4 A composite of scaled potential energy points for Si, P, S, and Cl. The energy and distance are unitless and the curvature is scaled to 72. The points are plotted against a Morse function (line) for comparison.

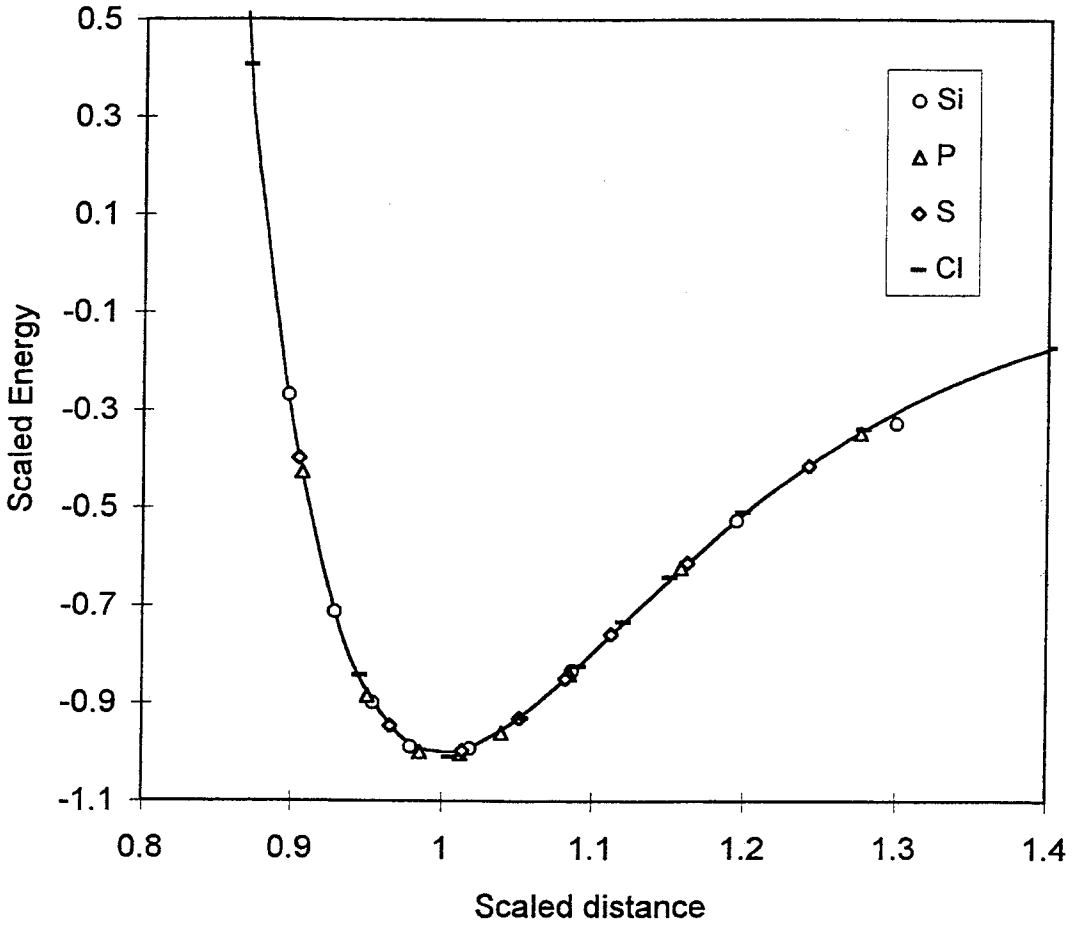


Figure III.5 The EXEX parameters from direct fits are show with solid lines (data points indicated by markers). Also EXEX curves generated from monotonic potential fits are shown for comparison (dashed line).

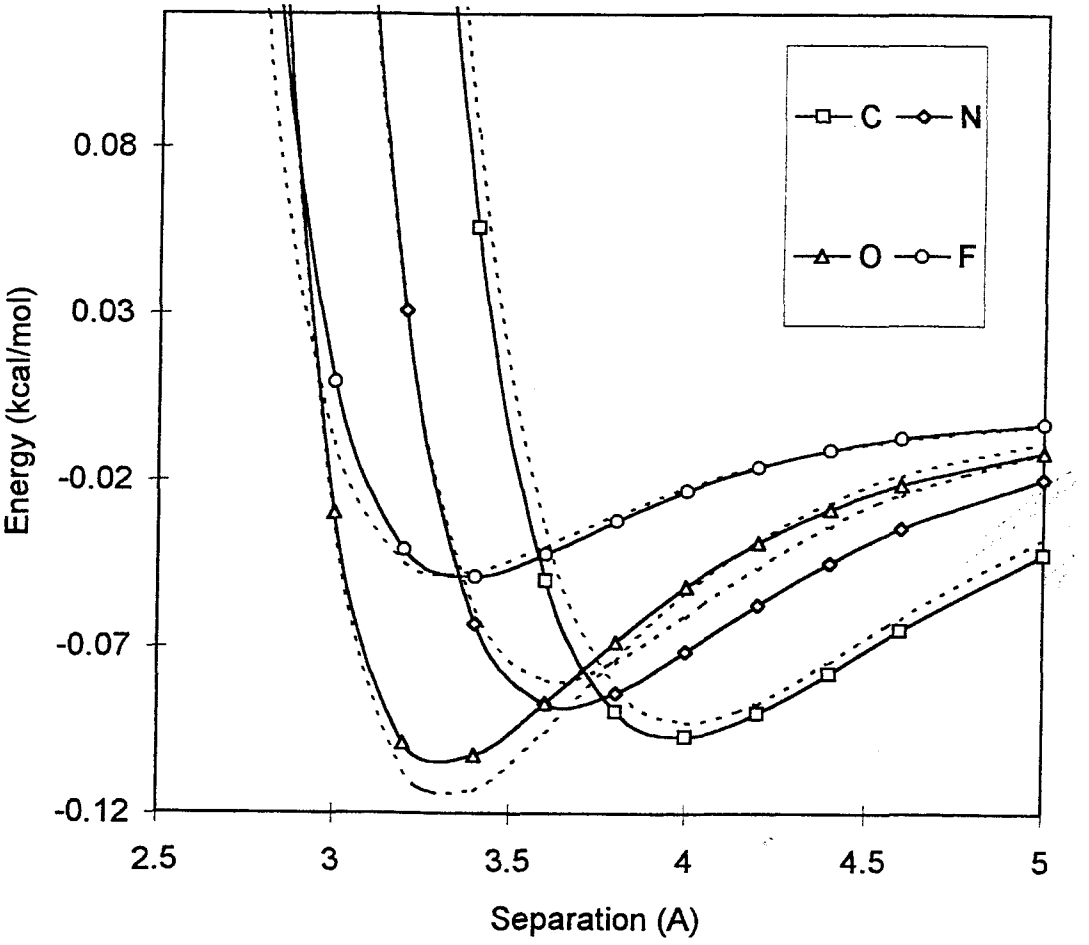


Figure III.6 The combination of repulsive potentials is examined. N and F potentials are shown in the same scale along with the N-F data (points). Both the combination (dashes) and the scaled combination (gray line) are included.

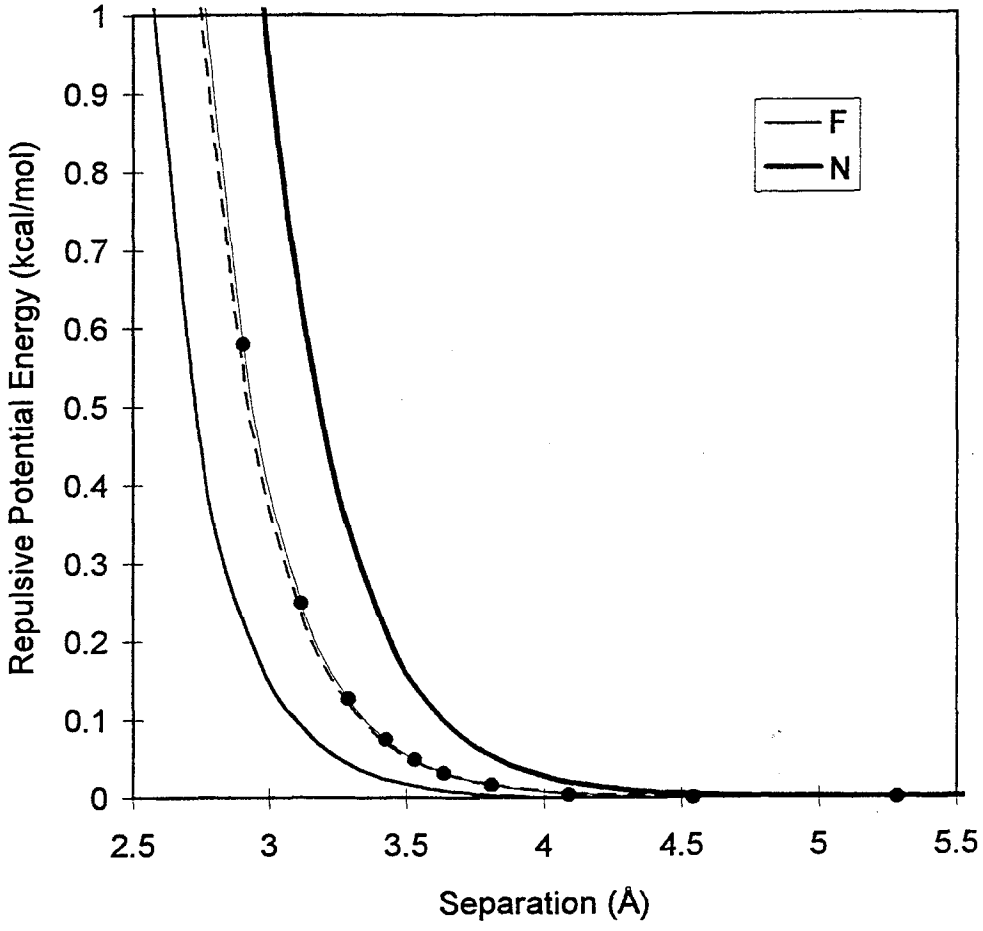


Figure III.7 The combination of attractive potentials is examined. N and F potentials are shown in the same scale along with the N-F data (points). Both the combination (dashes) and the scaled combination (gray line) are insluded.

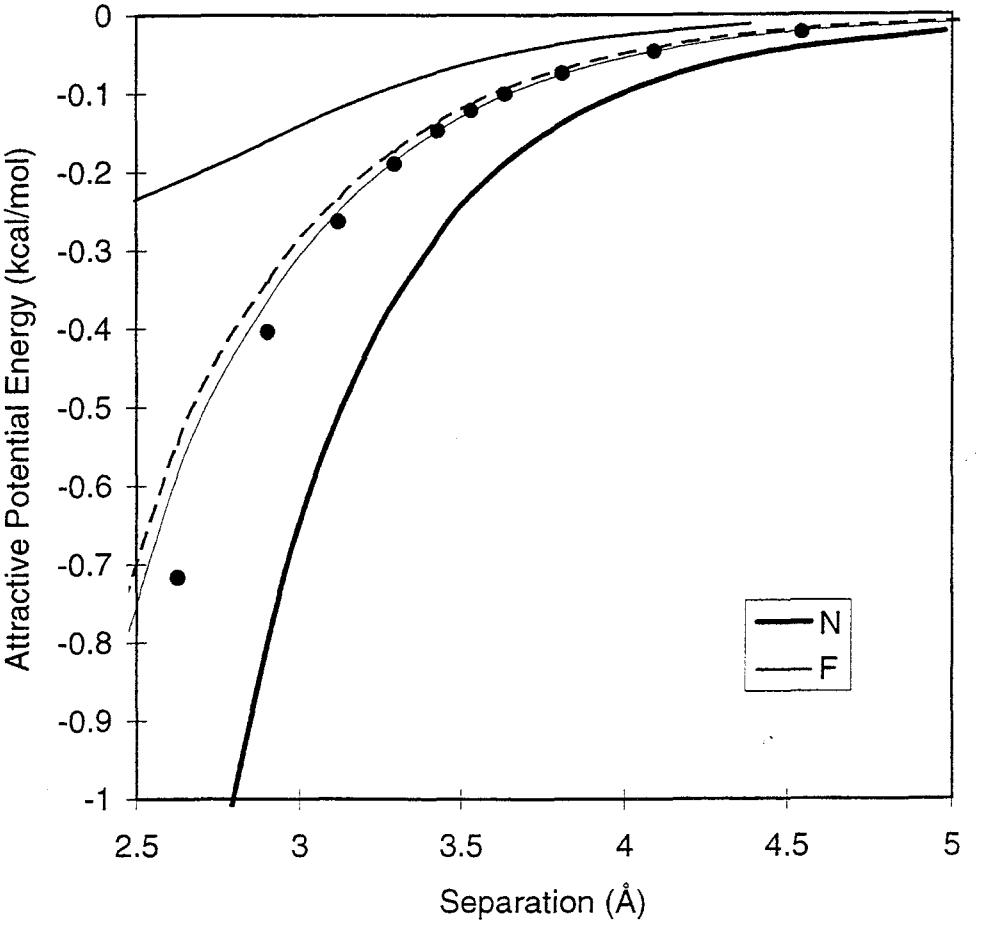


Figure III.8 The total potentials of N and F are plotted next to the heterogeneous N-F potential. In addition the combination rule potential is plotted (dashed line) and the scaled version (gray line).

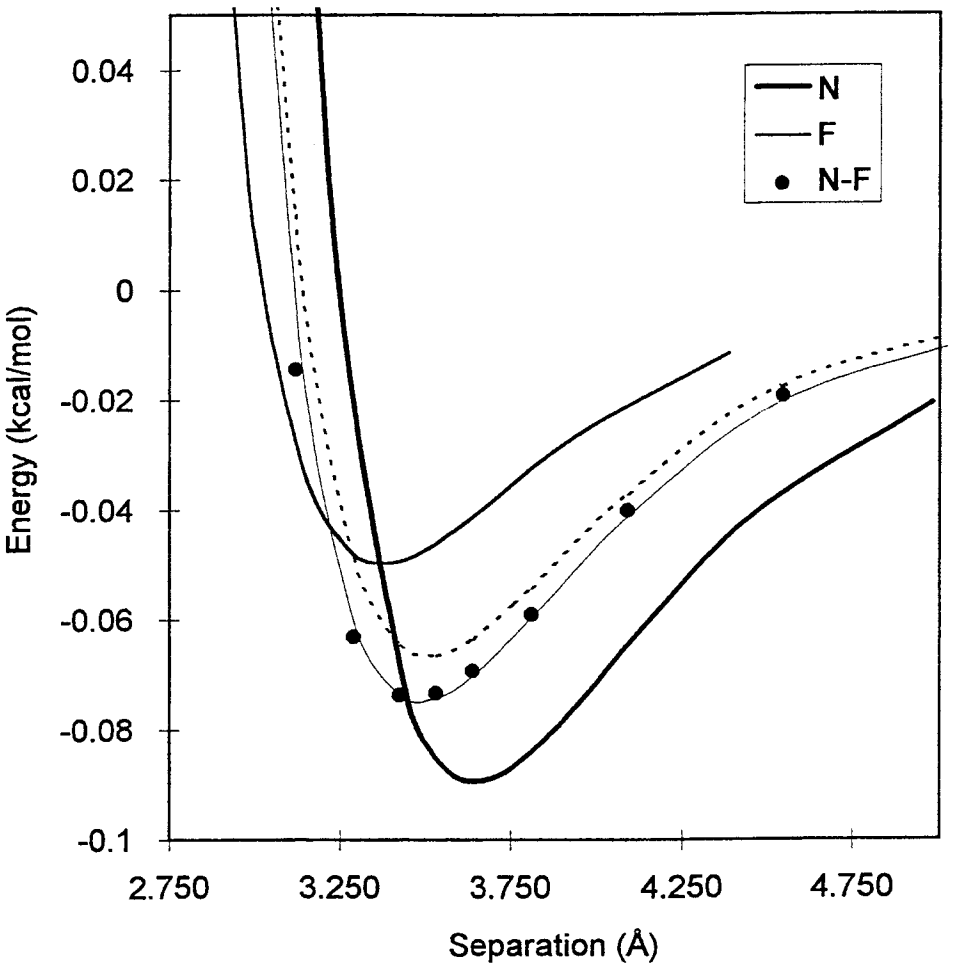
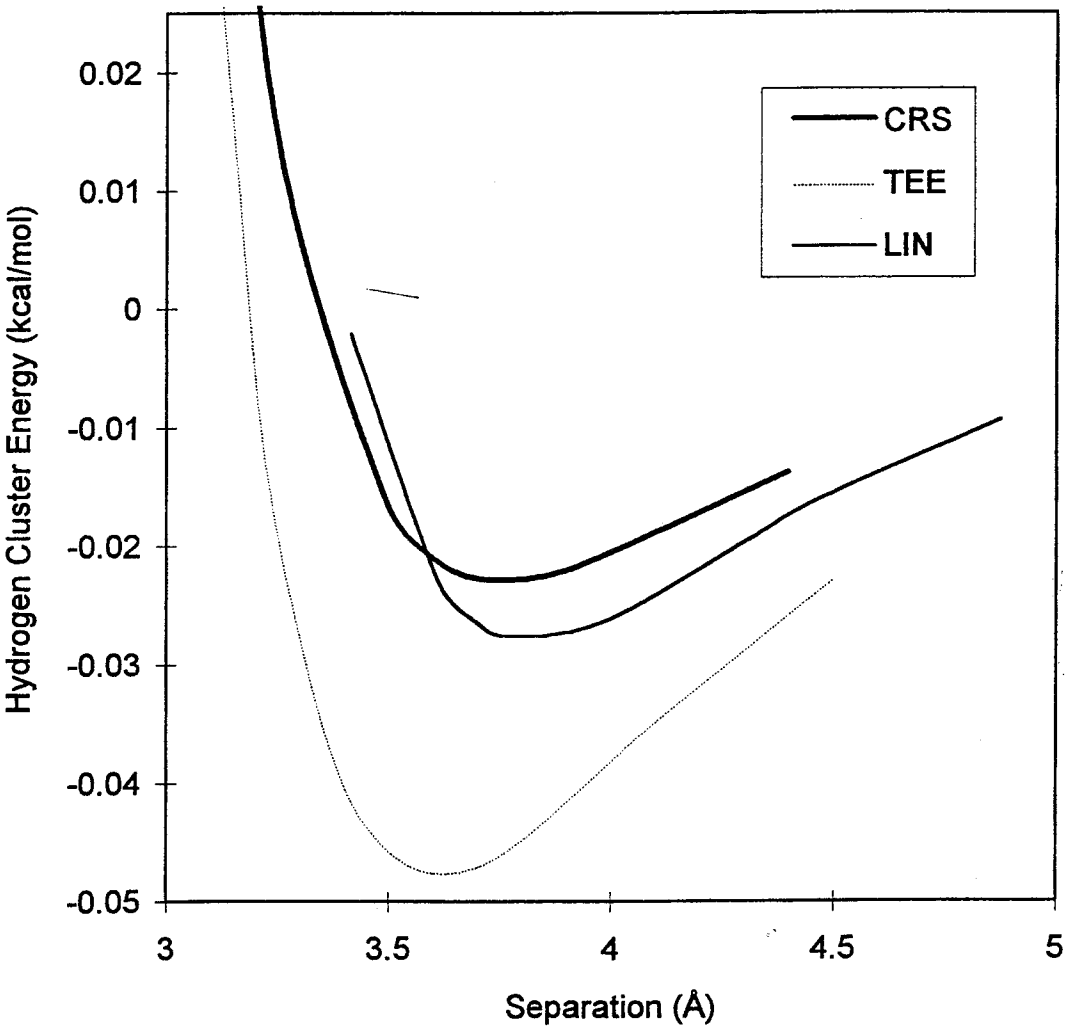


Figure III.9 Orientations of the hydrogen cluster are compared to test the anisotropy of the van der Waals potential.



Chapter IV. Conclusion and the Future of this Approach.

The COMP method of computing van der Waals potentials establishes an accurate and direct way to determine pair potentials for simulations. The strengths of this method are that it has been optimized to be accurate with a small basis set and it can be applied to any pair of atoms. The DREIDING[1] force field shows that the fundamental design principle in constructing an accurate force field is that high quality van der Waals potentials are essential.[2] Since there are inextricable difficulties in obtaining adequate potentials for many atoms, *ab initio* potentials promise to be very important in the development of the next generation of force fields. Furthermore, this is the true challenge in developing newer and more accurate force fields since valence terms, such as bond and angle spring constants, are more directly obtainable.

The midbond and midpoint functions are a great asset to the calculation because they impart the necessary accuracy with an acceptable number of basis functions. The midpoint function is the most efficient with only one diffuse function basis set at the geometric center of the cluster. Midbond functions, on the other hand, placed at the bond midpoint of each bond, are important because they can be used in all sorts of settings. While the midpoint functions cannot be generalized in a transparent manner to arbitrary molecular clusters, midbond functions can be applied clearly.

The COMP method is the first step in producing *ab initio* potentials for simulations. The future will see two advancements over the present study. First, the method will be applied either directly or at a lower level of accuracy to other atoms. For larger atoms this may only be practical with the availability of ample computer resources. The potentials of large atoms such as Br, Pb, and other third and fourth row atoms is very costly with the current method. Effective core potential calculations may be accurate enough to allow calculations on third-row atoms to be

done as easily as the second-row calculations here.

Second, the COMP method may be reiterated on first and second row cases with greater accuracy. This would also require increased computational power. The use of MP4 level calculations with more complete basis sets, such as QZ, would give more accurate potentials. Most importantly, the effective range of the calculation could be extended to a larger R_{max}. That is, for DZSP basis sets, the calculation is only accurate very close to the point where it is optimized, R_e in this case. For MP2/TZDP like was used here, the potential is accurate over a large enough range to define a potential. For MP4/QZTP, the value of R_{max} may be twice as far from R_e as with MP2/TZDP. With a potential computed over such a large range of R it may be easier to specify the long-range behavior of the potential. In Chapter III it appears that the attractive tail has an exponential topology since it is fit well by a Morse potential. This could be made more certain by extending the potential.

So, the approach is versatile enough to be refined to higher accuracy by directly increasing the wavefunction level and basis set completeness. Because of fundamental accuracy problems such as anisotropy and charge dependence of the van der Waals potential and even the accuracy of the pairwise treatment itself, the higher accuracy COMP calculations are not very important at the present time. For if the accuracy is reduced to, say, less than 0.1% then with these other issues unsolved there is an inherent inaccuracy which far exceeds the COMP energy accuracy.

The method has also been shown to be versatile enough to produce potentials for several orientations. This is made easier since the diffuse function sets (either midpoint or midbond) are accurate for different geometries even if they are optimized only in one single case. Here, the functions were optimized for the CRS geometry and were adequate for the other geometries studied. These potentials open the possibility to include orientation in simulations. The anisotropy of van der Waals potentials is inherent in the nature of the electron density of the diatoms. At some time, when simulation programs have more computational resources avail-

able, orientational potentials will be implemented in generic force fields.

An issue with the implementation of COMP potentials in simulations is the use of functional forms. The potentials can be implemented in a purely numerical manner which assumes no functional form. Most simulation programs, however, are more suited to potential functions. What's more, these functions are ideal for analyzing properties of van der Waals potentials by inspection. For instance, a perusal of Table III.3 gives one an idea of trends from atom to atom and of the topology of the potential of any single atom. For a while, at least, potential forms will be prominent in the discussion of van der Waals potentials. The optimal functional forms here are the exp-exp potential and Rydberg potential. These are a great departure from the usual van der Waals functions in simulations which have an R^{-6} term in accordance with the London potential. The exp-exp potential, however, is basically a short-range potential. So, as discussed in Appendix A, it totally lacks the London potential, but describes the attractive energy with an exponential function only. This is indeed an advantage for simulations because the exp-exp function has no long-range sum. For instance, in a lattice, the exp-exp interaction can be ignored beyond the third shell. If this is accurate for materials, the result would be faster simulations.

The COMP method lays the foundation for the next generation of generic force fields which will rely on accurate van der Waals potentials. A few important steps have been taken here. First, the technique for computing accurate potentials has been optimized and tested on ten atoms. For these potentials the accuracy of combination rules and isotropic pair potentials has been appraised. Second, the relation of the van der Waals potential topology has been related to atomic properties. Third, nonstandard functional forms are used to give the optimal fits of the computed potentials.

References

1. S. L. Mayo, B. D. Olafson, and W. A. Goddard III, J. Chem. Phys., **94**(26), 8897-8909 (1991).
2. C. Casewit and A. K. Rappé, *Molecular Mechanics Across Chemistry, in press.*

Appendix A. Potential Functions in Simulations.

A.1. The Topology of van der Waals Functions.

This Appendix reviews the types of functional forms which are used to describe van der Waals interactions both in this study and elsewhere. There have been a variety of such functions developed throughout the history of physical chemistry and most of them are still used currently in some particular setting. They all develop parametrically an energy function of R, interatomic separation. The functions studied here are isotropic pair potentials. There are, of course, anisotropic interatomic potentials such as the Axilrod-Teller[1] three-body potential or spherical harmonic approaches which build in angular structure of orbitals. Here we look only at simple $E_{vdw}(R)$ pair potentials which are the most widely used in simulations.

Pair potentials all share a few characteristics fundamental to the van der Waals potential. First, there is a steep inner wall representative of Pauli repulsion. Second, there is an equilibrium distance R_e at which the energy is a minimum, $E(R_e) = -D_e$. Finally there is an attractive tail on the segment (R_e, ∞) where the energy is nonpositive. Beyond that, some combination of exponential functions and rational functions is composed into a potential function. The potentials are then differentiated by topological characteristics such as the curvature near equilibrium. Also, as a rule of thumb, the topological character of a potential function near equilibrium is far more important than away from equilibrium. At chemical energies, $T \approx 1000K$, it is only important that the inner wall be repulsive. The precise energy above 1000K is not important, it is only important for simulations that the potential is repulsive at small R. At long R it is less important to precisely specify or fit the potential function. This is somewhat a matter of practicality as there is no data which specifically determines, either from theory or from experiment, an interatomic potential at long range, such as 30 Å, for anything but a pair of atoms in a vacuum treated by quantum mechanics. That is, besides the London dispersion potential of two atoms there is no guide as to the interatomic potential energy of a pair of atoms

in a molecule or in a crystal. At long R the asymptotic limit of the potential has an unknown importance. It is often assumed that a van der Waals potential should give the London dispersion interaction as its limit. This is not definitely known and without direct experimental evidence to substantiate the assumption it is possible that the induced multipole forces which form the basis for dispersion are somehow either diminished or unimportant in molecules and materials due to screening by other atoms.

Be that as it may, what is important in examining potential functions is the character near R_e. A simple way to characterize a potential is with three quantities, R_e, D_e, and the curvature at R_e. For comparison a unitless scaled curvature is calculated

$$\kappa = \left(R_e^2/D_e\right) E''(R_e). \tag{A.1}$$

For van der Waals potentials $\kappa \approx 72$. What may be more illuminating than the curvature at the minimum is the relative position of two characteristic points, the zero (R_0) and the inflection point (R_i), with respect to the minimum (R_e). These points are defined in the usual way

$$\begin{cases} E(R_0) = 0, & \text{the zero point} \\ E'(R_e) = 0, & \text{the minimum} \\ E''(R_i) = 0, & \text{the inflection point} \end{cases} \tag{A.2}$$

Stipulating, of course, that $E''(R_e) > 0$ for potentials with more than one critical point. The scaled versions of these points, $\rho_0 = R_0/R_e$ and $\rho_i = R_i/R_e$, have values near 0.90 and 1.10. Small changes in these values along with changes in κ represent the fundamental differences between potential functions. These values describe the function over the relevant range of R.

Two more characteristics of each curve which should be quantified are the asymptotic behavior at short and long range. Here we define $\bar{A}$ and $\bar{B}$ to achieve this

$$\begin{cases} \bar{A} = \lim_{R\to 0} E(R) \\ \bar{B} = \left(\frac{1}{2D_e R_e^6}\right) \lim_{R\to\infty} R^6 E(R) \end{cases}. \tag{A.3}$$

The quantity $\bar{A}$ has only limited values it can assume. The limit is either finite, infinite, or pathological (*i.e.* negative). The quantity $\bar{B}$, on the other hand, is more valuable. It defines the long range behavior of the potential function which is related to the dispersion potential. The limit is scaled by R^6 because the London dispersion potential and the attractive terms of several potentials are BR^{-6}. This $\bar{B}$ is of limited value for potentials which lack such a term. The comparison is useful, however, because in the application of potentials to lattice sums in simulations, a dispersion term can be important. A lattice sum of an infinite BR^{-6} attraction gives a simulation which is sensitive to the value of B. Hence, $\bar{B}$ is useful to compare potentials which are being fit at the same time to pair potential data and to simulated crystal properties.

There are some very simple potentials which lack the basic features listed above. These potentials are both mathematically pathological and unsuited for accurate simulations due to their simplistic construction, but are mentioned because of their widespread usage in physical chemistry. Furthermore, it is a striking demonstration as to the minimal nature of characteristics which define interatomic potentials in many situations that these potentials are useful for the many simulations they have been applied to. The hard sphere potential and the square well potential are used as crude approximations in scattering problems. The hard sphere, the simpler of these two, is

$$E^{hard-sphere}(R) = \begin{cases} \infty, & \text{for } R < R_0 \\ 0, & \text{for } R \geq R_0 \end{cases} . \qquad (A.4)$$

This two-valued potential has no equilibrium distance and curvature properties. The square well potential adds to the hard-sphere an attractive region

$$E^{square-well}(R) = \begin{cases} \infty, & \text{for } R < R_0 \\ -\epsilon, & \text{for } R \geq R_{in} \\ 0, & \text{for } R \geq R_{out} \end{cases} . \qquad (A.5)$$

There is still no way to define R_e, R_0 , and R_i. Both of these pathological potentials have the advantage of simplifying integrals in scattering calculations to constant

values, and they have the correct limits as $R \rightarrow 0$ and $R \rightarrow \infty$. Another pathological potential of this type is the Sutherland potential

$$E(R) = \begin{cases} \infty, & R < R_0 \\ -C_m R^{-m}, & R \geq R_0 \end{cases} . \qquad (A.6)$$

All of the following potentials, on the other hand, are not only well defined on $(0, \infty)$ but are also C^1. Among potentials which are $C^1(0, \infty)$, most are $C^\infty(0, \infty)$ the only exceptions being potentials which are piecewise constructions yielding a C^n potential such as the Smith-Thakkar potential. This function interpolates between a Morse function and a $C_6 R^{-6} + C_8 R^{-8}$ term at long range with a Hermite polynomial in between. Also the Aziz helium potential[2] consists of a spline interpolation between a Morse and an R^{-6} term. This Morse-spline-van der Waals (MSV) and such potentials will not be examined here since the focus is rather to apply simple functions, which are not piecewise constructed, to simulations relying on a minimum number of parameters. Moreover the application of simple functional forms constrains the topological character. Specifically, with a single function, the topology of one region of the curve is inextricably linked with that of other regions. In the case that this leads to inaccuracies, piecewise constructions are able to dodge the problem easily. Since such functions are both more difficult to implement in simulations and more complicated to interpret, the simple functions are still very useful. In this Appendix, the focus is on the simple functions.

There are several well behaved potential functions which are used commonly both to fit calculated potentials and empirical properties of materials.[3-5] Considering their usefulness in modeling, simulations, and simply comparing potentials of different systems it is important to explore the topological properties and versatility of these functions. Three, the ones implemented in the POLYGRAF[6] program, are included here. These are the Lennard-Jones (LJ) 12-6 potential, the exponential-6 (E6), and the Morse function. Two others are also included for a balanced survey, the generalized Lennard-Jones (GLJ) function, which is a generalization of several LJ type functions used in force fields, and the exponential-exponential (EXEX)

function which is introduced in the present thesis. All but the GLJ function are listed in Section II.1.

The simplest continuous potential function is the Lennard-Jones 12-6 function which has two free parameters

$$E^{LJ}(R; A, B) = AR^{-12} - BR^{-6}. \tag{A.7}$$

It can also be written in terms if its characteristic distances.

$$E^{LJ}(R; R_0, D_e) = D_e \left[\left(\frac{R_0}{R} \right)^{-12} - \left(\frac{R_0}{R} \right)^{-6} \right] \tag{A.8}$$

which is a very symmetrical expression in terms of the scaled distance $(\frac{R}{\sigma})$. It is also often scaled by R_e instead, giving the canonical form

$$E^{LJ}(R; R_e, D_e) = D_e \left[\left(\frac{R_e}{R} \right)^{-12} - 2 \left(\frac{R_e}{R} \right)^{-6} \right]. \tag{A.9}$$

The parameters for these functions are written in n-tuples. In standard form this is (A, B) and in canonical form (R_e, D_e). Canonical form may be assumed in the text. For functions below with more parameters these would be triples or quadruples. Also the units are assumed to be Å for distance and kcal/mol for energy unless otherwise specified.

The two forms of the LJ potential, Equations A.8 and A.9, are both used. The only difference between them is a factor of two in the attractive term arising from the fact that $R_e = 2^{\frac{1}{6}} R_0$ or $\rho_0 = 2^{-\frac{1}{6}}$. Also, $\rho_i = (\frac{13}{7})^{\frac{1}{6}}$. This equates to a nearly even spacing of R_0, R_e, and R_i, for

$$\rho_0 \rho_i = \left(\frac{13}{14} \right)^{\frac{1}{6}} \approx 0.988. \tag{A.10}$$

For the LJ potential, $\bar{A} = \infty$ and $\bar{B} = 1$. So, at short range the potential has the correct behavior as the energy increases without limit. This is ideal for molecular simulations so that atoms far within R_e of each other always have a strongly repulsive interaction. At long range there is only one value of $\bar{B}$ available

which means that once R_e and D_e are specified, the dispersion potential is also specified with this potential.

There is one simple generalization of the Lennard-Jones function which is occasionally used, the LJ $n-m$ function, or the $\nu-\mu$ function where μ and ν are real. This yields, of course, the properties of the LJ 12-6 function for the case that $\mu = 12$ and $\nu = 6$. It can be written in standard form as

$$E^{GLJ}(R; A, B) = AR^{-\nu} - BR^{-\mu} \qquad (A.11)$$

where $\nu > \mu$. In canonical form it is

$$E^{GLJ}(R; R_e, D_e) = D_e \left[\mu \left(\frac{R_e}{R} \right)^{-\nu} - \nu \left(\frac{R_e}{R} \right)^{-\mu} \right] \left(\frac{1}{\nu - \mu} \right). \qquad (A.12)$$

This is a more versatile function with ν and μ free to vary. Special cases are seen in the literature. Usually these use integral values, n and m. For example the 9-6, 9-3, 10-6, 12-10, and the n-6 (for $n > 12$) potentials have been built into force fields.[7] The main flaw these have in the present study is that they fit the COMP data poorly. As is shown in Chapter II the COMP data has a definite Morse character and power functions have the wrong topology for the attractive tail. Furthermore the GLJ function is fundamentally incorrect in the repulsive region since the Pauli repulsive force is exponential in nature and no power function can match an exponential completely.

On the practical side, however, LJ and GLJ functions are quite useful for simulations. They are quick to compute (power functions being faster than exponentials) and the asymptotic limits are correct. There is no catastrophe region or critical point which is a maximum as long as $\nu > \mu$. A final bonus with the LJ 12-6 function is that it has only two parameters. The more versatile functions below have three or four. The LJ 12-6 function is easier to fit to data because of fewer parameters and because its two free parameters, R_e and D_e, are the easiest van der Waals parameters to determine experimentally.

To complete the discussion of the GLJ potential, the characteristic points of the GLJ are determined. They are variable, depending on μ and ν. Just as for the LJ 12-6 function, they are asymmetrically disposed around $\rho_e = 1$, for $\rho_0^{GLJ} = (\mu/\nu)^{\frac{1}{\nu-\mu}}$ and

$$\rho_i^{GLJ} = \left(\frac{\nu+1}{\mu+1}\right)^{\frac{1}{\nu-\mu}}. \tag{A.13}$$

As for the LJ 12-6 function, $\bar{A} = \infty$. $\bar{B}$ has three possible values depending on the value of μ.

$$\bar{B}^{\text{ GLJ}} = \begin{cases} \infty, & \text{for} \mu < 6 \\ 1, & \text{for} \mu = 6 \\ 0, & \text{for} \mu > 6 \end{cases}. \tag{A.14}$$

This is not too informative except to say that the potential violates the London dispersion relation except when $\mu = 6$. For brevity, this three valued $\bar{B}$ is referred to simply as $\bar{B}^{GLJ}$. The remaining characteristic quantity, κ, is informative for the GLJ function,

$$\kappa^{GLJ} = \nu\mu. \tag{A.15}$$

So, the parameters μ and ν directly control the curvature at the minimum along with the characteristic points ρ_0 and ρ_i above.

The exponential-6 (E6) function is the next most used for simulations after the LJ 12-6 function. It is more accurate than LJ because it has another free parameter which can adjust the curvature independent of R_e and D_e. The E6 function is, in standard form,

$$E^{E6}(R; A, B, C) = Ae^{-CR} - BR^{-6}. \tag{A.16}$$

In canonical form a third parameter ζ is introduced which controls the curvature

$$E^{E6}(R, R_e, D_e, \zeta) = D_e \left[\left(\frac{6}{\zeta-6}\right) e^{\zeta(1-\frac{R}{R_e})} - \left(\frac{\zeta}{\zeta-6}\right)\left(\frac{R_e}{R}\right)^6\right] \tag{A.17}$$

Indeed, the curvature at the minimum is controlled directly by the scaling parameter ζ, for

$$\kappa^{E6} = 6\zeta\left[\frac{\zeta-7}{\zeta-6}\right]. \tag{A.18}$$

The value of κ is equal to 72 for the LJ 12-6 function, but equals that value only for the case of $\zeta = 13.772$ with the E6 function. Furthermore, $\bar{B}$ is also controlled directly by ζ

$$\bar{B}^{E6} = \frac{\zeta}{2(\zeta - 6)}. \tag{A.19}$$

The condition of $\bar{B}^{LJ} = \bar{B}^{E6}$, however, is only met when $\zeta = 12$. This shows one fundamental difference between the two functions.

The characteristic points, ρ_0 and ρ_i, are given by transcendental equations because of the additive combination of an exponential term with a power term

$$\begin{cases} e^{\zeta(1-\rho_0)} = \left(\frac{\zeta}{6}\right)\rho_0^{-6} \\ e^{\zeta(1-\rho_i)} = \left(\frac{7}{\zeta}\right)\rho_i^{-8} \end{cases}. \tag{A.20}$$

Values of ρ_0 and ρ_i are shown in Table A.1 for selected values of ζ.

Table A.1. Characteristic Points of the Exponential-6 Function for Several ζ Values

ζ	ρ_0	ρ_i	$\bar{B}$
11	0.867139	1.129930	1.100
11.5	0.871917	1.125274	1.045
12	0.876101	1.121157	1.000
12.5	0.879829	1.117466	0.962
13	0.883198	1.114120	0.929
13.5	0.886273	1.111058	0.900
13.772	0.887841	1.109495	0.886
14	0.889104	1.108236	0.875
14.5	0.891727	1.105620	0.853
15	0.894170	1.103182	0.833

The E6 function has a repulsive term which is based on exponential Pauli repulsion. The exponential nature can be deduced from the quantum mechanics

of diatoms.[1,5] The analytical expression is modified in most cases by a polynomial arising from the $n \geq 2$ atomic orbitals and the overlap, S, itself.[8] A Rydberg function can fit, in general, an overlap expression, but for the *ab initio* data above the simple exponential seems to suffice. Also it has the London dispersion force, R^{-6}, for the attractive term.

While this function is chosen for physical reasons it has a pathology for small R. Because of this the E6 function has a critical point, R_c, which is a maximum, and its limit at small R is $-\infty$. This means that $\bar{A}^{E6} = -\infty$. This also means that that even if a pair of atoms in a simulation were at a separation of $R < R_c$ then there would be a very strong attractive force between them which will definitely disrupt the simulation. Since $E^{E6}(R_c)$ has a very large value, dynamics at temperatures below 1000K will not bring atoms within R_c. Even at normal temperatures, however, there could be a problem with initial structures or pathologically constrained structures. For example, if dynamics is begun with a structure containing a very close contact, $R < R_c$, then a problem arises. In this event, dynamics will bring the two atoms to $R = 0$ because of the strong attractive force from the $-BR^{-6}$ potential at short range. A simple ploy to circumvent this problem is to have a switching algorithm which replaces the E6 function with a repulsive function for $R < R_c$. A convenient function to switch with is the $-E^{E6}(R)$, making R_c no longer a maximum of the switched function (this is done in the POLYGRAF program[6]). Nevertheless, it is important to beware of such pathologies anywhere in force field potentials.

Finally there is the Morse function which was developed as a simple model for the quantum mechanics of diatomic molecules.[9,10] It is rarely used for van der Waals interactions, in fact, because although it has the correct asymptotic limit of 0 at long R, it lacks an R^{-6} term for London dispersion. Despite this it has a superior fit to the COMP potentials and can be very useful for simulations. Moreover, since it lacks a power function such as R^{-6}, the Morse function is a short-range potential. So, it

does not give infinite sums in lattice simulations, but can be truncated accurately. In standard form it is written

$$E^{Morse}(R; A, B, C) = Ae^{-CR} - Be^{-\frac{1}{2}CR} \tag{A.21}$$

or in terms of a scaled exponential

$$E^{Morse}(R) = \chi^2 - \chi, \text{where } \chi = Ae^{\zeta(1-R/R_e)}. \tag{A.22}$$

In canonical form ζ appears as a curvature scaling parameter

$$E^{Morse}(R; R_e, D_e, \zeta) = D_e \left\{ e^{\zeta(1-R/R_e)} - 2e^{\frac{1}{2}\zeta(1-R/R_e)} \right\}. \tag{A.23}$$

The Morse function has κ which is controlled by its scale parameter, $\kappa = \frac{1}{2}\zeta^2$. $\bar{B}$, on the other hand, is independent of any parameters as the limit is always 0. So, κ and $\bar{B}$ alone show definite differences between the LJ, E6, and Morse potentials. The $\bar{A}$ parameter for Morse is unremarkable, being a large finite number, $\bar{A} = A$. Finally the characteristic points are unlike those of the LJ and E6 functions in that they are precisely evenly spaced around R_e.

$$\begin{cases} \rho_0^{Morse} = 1 - \frac{2ln2}{\zeta} \\ \rho_i^{Morse} = 1 + \frac{2ln2}{\zeta} \end{cases}. \tag{A.24}$$

A generalization of the Morse potential is to let the two exponential terms, the repulsive and the attractive, vary independently. This adds a fourth parameter so that the exponential range parameters are not constrained to be related by a factor of two. This generalized Morse curve is called an exponential-exponential or exp-exp. The exp-exp (EXEX) function adds a fourth parameter which does two things. First in the standard representation

$$E^{EXEX}(R; A, B, C, D) = Ae^{-CR} - Be^{-DR}. \tag{A.25}$$

It is a composite of two exponentials. This is appropriate for representing the COMP potentials which are seen, in Chapter I, to be composed of an attractive

and a repulsive curve, both of which are fit well by a single exponential function. So, the exp-exp makes sense considering how the two monotonic curves may vary independently between cases. Second, in canonical form there are two scaling parameters, ζ_1 and ζ_2.

$$E^{EXEX}(R; R_e, D_e, \zeta_1, \zeta_2) = D_e \left\{ \zeta_2 e^{\zeta_1(1-R/R_e)} - \zeta_1 e^{\zeta_2(1-R/R_e)} \right\} \frac{1}{\zeta_1 - \zeta_2}. \quad (A.26)$$

For both Morse and exp-exp, the spacing of ρ_0, 1, and ρ_i is even. That is, the scaling parameters cannot change the identity

$$\frac{1}{2}(\rho_0 + \rho_i) = 1. \quad (A.27)$$

What is important with the general Morse is that ρ_0 and ρ_i are independent of κ (see Table A.2). Notice that for $\zeta_2 = \frac{1}{2}\zeta_1$, when the exp-exp function becomes a Morse, its values in Table A.2 become identical to those for Morse. So, the curvature at the potential minimum and the characteristic points along the inner wall and the tail can be adjusted independently for exp-exp. This is a very versatile function which, just like Morse, is short-range.

There is another form which shows better the commonality between the Morse function and the exp-exp, where the scaling parameters are transformed into ζ and ξ where

$$\begin{cases} \zeta = \sqrt{\zeta_1 \zeta_2} \\ \xi = \zeta_1/\zeta_2 \end{cases}. \quad (A.28)$$

These appear symmetrically in the canonical form which shows how the EXEX potential is a generalized Morse function.

$$E^{EXEX}(R; R_e, D_e, \zeta, \xi) = D_e \left\{ e^{\zeta\sqrt{\frac{\xi}{2}}(1-R/R_e)} - \xi e^{\zeta\sqrt{\frac{1}{2\xi}}(1-R/R_e)} \right\} \frac{1}{\xi - 1}. \quad (A.29)$$

Just like for the Morse potential, the curvature is related to ζ by $\kappa = \frac{1}{2}\zeta^2$.

The topological features of all four potentials are summarized in Table A.2 and since the exp-6 functions does not have a closed form for the points ρ_0 and ρ_i values are given specifically for $\zeta = 13.772$. For convenience the characteristic

Table A.2. Topological Measurements of the Main Potential Functions

Function	κ	ρ_0	ρ_i
LJ 12-6	72	0.890899	1.108686
Exp-6	$6\zeta\left(\frac{\zeta-7}{\zeta-6}\right)$	0.887841	1.109495
Morse	$\frac{1}{2}\zeta^2$	$1-2ln2/\zeta$	$1+2ln2/\zeta$
Exp-Exp	$\zeta_1\zeta_2$	$1-\frac{ln(\zeta_1/\zeta_2)}{(\zeta_1-\zeta_2)}$	$1+\frac{ln(\zeta_1/\zeta_2)}{(\zeta_1-\zeta_2)}$
Exp-Exp	$\frac{\xi}{(\xi^2-1)}s^2$	$1-\frac{ln(\xi)}{s}$	$1+\frac{ln(\xi)}{s}$

points of the EXEX function are given in terms of the spread, s, between the two ζ values

$$\begin{cases} s = \zeta_1 - \zeta_2 \\ \xi = \zeta_1/\zeta_2 \end{cases} . \qquad (A.30)$$

The parameters ζ and s could, in fact, be used as parameters in place of ζ and ξ or ζ_1 and ζ_2. Anyhow, it is clear that the characteristic points are even distributed about the minimum, $\rho = 1$, for both the Morse curve and the EXEX curve. So, the fourth parameter, in fact, does not give EXEX more control over these points. What is different is that κ can change for given characteristic points with EXEX but not with Morse. From the fits in section III.4 this seems to be crucial for accurately modeling the *ab initio* potentials.

The canonical parameters (R_e, D_e, ζ, ξ) show by inspection the basic topology of the potential. The standard parameters (A, B, C, D) on the other hand are useful for comparison to the monotonic potentials. For unlike with the other functional forms, since EXEX has four parameters one may see the the dependence of the canonical parameters on the standard parameter values. This can illustrate the influence of changes in monotonic parameters on the characteristic points of the EXEX function. A system of equations defines the transformation from standard to canonical parameters. The reverse transformation may be done by inspection,

comparing Equations A.25 with either A.26 or A.29.

$$\begin{cases} R_e = ln(AC/BD)/(C-D) \\ D_e = A(\xi-1)e^{-\zeta\sqrt{\frac{\xi}{2}}} = B(1-\frac{1}{\xi})e^{-\zeta\sqrt{\frac{1}{2\xi}}} \\ \zeta = R_e\,(2CD)^{1/2} \\ \xi = C/D \end{cases} . \qquad (A.31)$$

It can be shown using Equations A.31 that the repulsive and attractive range parameters have two different degrees of influence over the canonical parameters. The focus is on the position of the well bottom (R_e, D_e). This is useful to know for two reasons. First it is important to know how an error in a standard parameter causes errors in the canonical parameters. Second, when modifying parameters in correction schemes such as CMVDW (section III.4), it is important to know how much effect a parameter has on properties. For example, an EXEX function with canonical parameters $(3, 5, 0.1, 12, 2)$ is taken. Then the standard parameters are varied independently. Then the modified standard parameter sets are transformed back into canonical parameters and compared with the reference set.

Table A.3 gives six cases of this variation and transformation. The variations in the standard parameters are done independently so in each case only one parameter is varied from its original value. The quadruple represents the factor by which each parameter is multiplied. Of course, only a single value in the quadruple is different from one in each case. The canonical parameters are simply listed in the usual way to be compared with the reference line. Some basic facts about the EXEX function are illustrated. First, it is shown that changes in A and B are the simplest. They alter the well depth the most and the proportional change in D_e is twice as large for the change in B than for the same proportional change in A. Second, the attractive parameters (B, D) are less important for the equilibrium distance than the repulsive parameters (A, C). The first four cases show that changes in C alter R_e by twice as much as the equivalent proportional change in D. Third, changes in the scaling parameters are directly proportional to changes made in range parameters (C, D) and insensitive to changes in the prefactors (A, B).

Finally, values of the zero point, R_0, can be determined from Table A.3. This is important for simulations of structures because this value defines the hard wall of a van der Waals potential and can indicate the size or packing of molecular crystals. The scaled version, ρ_0, is insensitive to changes in all six cases. The original case has $\rho_0 = 0.0884475$ and the six cases here vary between 0.883 and 0.886. These variations are insignificant compared to the variation in R_e, especially in the first four range parameter cases. The conclusion is then that the zero of the potential is controlled just the same as the minimum, mostly by the range parameters C and D. Furthermore, the influence of C is over double that of D.

Table A.3. Relationship Between EXEX Standard and Canonical Parameters

Standard	Canonical
(1,1,1,1)	(3.500, 0.100000, 12.000, 2.00)
(1,1.02,1,1)	(3.376, 0.075873, 12.194, 2.04)
(1,1,1,1.02)	(3.560, 0.121539, 11.882, 1.96)
(1,0.98,1,1)	(3.634, 0.122088, 11.879, 1.96)
(1,1,1,0.98)	(3.443, 0.081460, 12.122, 2.04)
(1.02,1,1,1)	(3.512, 0.0980, 12.041, 2.00)
(1,1,1.02,1)	(3.488, 0.1041, 11.959, 2.00)

Of the functions examined above, the Morse and the EXEX functions are suitable for use as van der Waals potential functions in simulations. They combine the desired mathematical properties with the best fits to the COMP data. Both the Morse and EXEX functions have no pathologies. They are $C^\infty(0,\infty)$. Also they have the correct asymptotic limits, for $\bar{A} = A$ and $\bar{B} = 0$. What's more, the limit in $\bar{B}$ goes to zero very rapidly since both functions are based on exponentials and contain no power functions. This makes lattice sums facile. In fact, a Morse

van der Waals function is almost like having a small van der Waals cutoff. Only the nearest neighbors and perhaps the second and third neighbors (depending on the level of precision) count in the energy expression for an extended system. This means that both of these functions are faster than either LJ 12-6 or E6. The only possible drawback is that an R^{-6} term is often called for in the van der Waals function to match London theory. In the second part of this Appendix, section A.2, the implications of this on lattice simulations is examined.

In closing the present section mention should be made of some other valuable potential functions. Because of the appeal of functions which have London dispersion as long R, piecewise constructions have been made using accurate fits and short R plus dispersion at long R. What these functions all have in common is they have a few different functional forms and cutoff points, $\{R_i^{cut}\}_{i=1,2,\ldots,n-1}$ for n functions, at which the functions are pieced together. This is a constructed potential function (COPF)

$$\begin{cases} E(R) = f_1(R), \ \text{for} R \in (0, R_1) \\ E(R) = f_2(R), \ \text{for} R \in (R_1, R_2) \\ \vdots \\ E(R) = f_n(R), \ \text{for} R \in (R_{n-1}, \infty) \end{cases} \tag{A.32}$$

Proper constructions are at least $C^1(0,\infty)$ so that all matching conditions at cutoff points are C^1 or higher. These functions, as they are applied to simulations, can be treated by summarizing the basic types. Usually only $n \leq 3$ is needed. That gives a function to fit near the well bottom such as a Morse potential and f_n is some sort of dispersion function. This dispersion function can be any sort of multipole interaction term

$$E^{disp}(R) = \sum_{n=0}^{N} B_n R^{6+2n} \tag{A.33}$$

where N ranges from 0 to 2. These two functions are then splined together in the middle with a cubic polynomial, f_2, to impose C^1 matching conditions at the cutoff points. The function in the binding region can be a Morse function, to give an MSV function or anything else with a repulsive wall and some attractive region.

A simpler alternative is that the Morse and van der Waals functions can form a COPF with no intervening polynomial. So, f_1 is a Morse function and f_2 is a van der Waals function. The first application of this function was by Konowalow and Zakheim.[11] Their study used what they called a Morse-6 function which here is the COPF just described with only the R^{-6} term present in the f_2 dispersion function. The necessity of using a COPF is illustrated by this case. While f_1 may totally disappear at long R, f_2 is unbound for $R \rightarrow 0$ and the construction circumvents the imbalance. This sort of potential is, however, quite limited. For if the purpose is to include London dispersion, then the parameter in $f_2(R; B_0)$ is determined by polarizability. In order for matching criteria, two parameters of $f_1(R; R_e, D_e; \zeta)$ must be fixed. This compromises accuracy and even if the matching criterion is reduced to $C^0(R_1)$, for then still one parameter of f_1 is fixed. Nevertheless, this Morse-6 functions is being used currently in the design of a new force field.[12]

In this discussion the COPF is treated as ancillary because if the properties of two or more functions work well, each in different regions (R_i, R_{i+1}), then they can be used in a COPF for a more accurate potential. What is more important here is rather to understand the properties of simple functions. COPF is important to the designer of a force field if there is no other way to obtain accurate potentials. Either way, a systematic understanding of simple functions is precursory to the enlightened design of a COPF. A final type which is commonly used, but is quite different from the MSV and the like, is simply a spline fit to accurate data. This would be a COPF with a large N to fit data such as scattering or van der Waals cluster spectroscopic data. In order to have the correct limits, especially for force field simulations, it may be important to impose limit functions. For instance f_1 could be an exponential repulsion function and f_N a dispersion function, Equation (A.32).

Another accurate function is the Rydberg function, Equation (III.2), which can have the correct limit values for a potential function. It can also be very accurate. As demonstrated in Section III.3, the RYD2 function fits well the monotonic

potential data from *ab initio*. An excellent fit to the whole curve is obtained from a Rydberg function with powers from -2 to 2. While these functions can be very accurate, given their versatility, they are not considered ideal for the present thesis. One reason is that $\bar{A}$ is a finite number which may be small or even negative. This may be accommodated in the same way that $\bar{A}^{E6}$ is. More importantly, however, the values of the coefficients in the polynomial term are unstable to changes in the data. While the quality of the fit does not waiver, small changes in characteristic points lead to changes in coefficients. For instance, a 1% change in either R_e or D_e can lead to a change of over an order of magnitude or sign in a coefficient. Although this is not necessarily detrimental to simulations, it makes interpretation and finding patterns between elements very unclear. The values of parameters in Tables III.3 through III.6 directly give properties of the potentials. Rydberg parameters, on the other hand, are less useful.

All together the COPF approach is an aside and the main analysis here shows how a critical specification of potential function properties aids in the differentiation between the different functional forms. The analysis using characteristic points and asymptotic limits clearly characterizes the functions. Furthermore this section provides insight into the interpretation of functional fits to van der Waals potentials such as those in Tables III.3 through III.6. There are issues beyond the topological features which arise in simulations. Sums of potentials in extended systems further contrast the potentials with and without the long range London term. The next section applies the topological characterization above to extended systems.

A.2. Comparison of Functions in Simulations.

A further important issue with these potential functions besides topology is differences in their lattice sums. For simplicity lattice sums on an closest packed or face centered cubic (FCC) lattice are examined here but the trends seen with the long-range and the short-range potentials are representative of the lattice sums in materials. There is an interesting discrepancy in the fit of an E6 function to a Morse in two different ways. First, if the Morse potential over a certain range similar to those over which COMP potentials were determined is treated as data and the E6 is fitted to that the E6 function resembles the Morse near R_e. At long range, however, the E6 behaves as BR^{-6} whereas the Morse vanishes. Second, if crystal properties of an FCC noble gas lattice are computed with a Morse potential and then the properties are fitted with an E6 a different E6 function results. This is because the long-range behavior of the E6 function is being averaged over the lattice in the energy expression which determines the crystal properties.

If a Morse potential is chosen with $R_e = 3.5$Å, $D_e = 0.100$kcal/mol, and $\zeta = 12$ then the best fit E6 on the range (3.0Å, 5.0Å) has parameters $(R_e, D_e, \zeta) = (3.470, 0.099, 14.231)$ with an rms deviation of 0.003 kcal/mol between the curves. Here the E6 function is very similar to the Morse, differering significantly only in the curvature. If, however, the E6 is fit to the properties given by the Morse function on an FCC lattice, then it has the parameters $(3.500, 0.0845, 14.295)$. The result is that the D_e value is reduced. That is because while the Morse potential is significant for its nearest neighbors, 12 in an FCC lattice, the E6 attractive tail must be summed over the entire lattice. For the R^{-6} function, this lattice sum gives an attractive energy of effectively 14.45 nearest neighbors.[1] The quotient 12/14.45 happens to be very close to the factor by which D_e is reduced from the Morse parameter set to the latter E6 set.

The E6 fit to a Morse function over the range (3Å, 5Å) is actually nonlinear in ζ. The fitted parameters depend on the particular Morse ζ chosen. Values for R_e and D_e are merely representative. They can be scaled out to unitless values of

1 or scaled to other values. The value of ζ is the most important as it controls the relation between the fit E6 parameters and the Morse parameters. Table A.4 shows fitted E6 parameters for Morse κ values ranging from 58 to 84. The value of ζ^{E6} cannot be determined simply by matching the κ value it gives to the κ from the Morse ζ. For instance, with Morse $\zeta = 12$ above, the Morse curve has $\kappa = 72$ but the fitted E6 curve has $\kappa = 75.014$. Furthermore, the best fit of an E6 curve to a Morse is actually for a ζ value which is never found in a COMP potential.

Table A.4 shows the optimal case with $\kappa^{Morse} = 35.481$ which minimizes the rms error with ζ treated as a free parameter. Real potentials, however, have ζ values near 72. This mismatch may be a basic weakness of the E6 function for accurate van der Waals energies in small systems where the long range behavior of the E6 function can be ignored (or in any system where long range interaction is negligible). If most potentials are like the COMP potentials from about 3Å to 5Å, then the E6 function will always fit more poorly than Morse or EXEX. For simulations, though, it does not need to be ruled out, for the rms error in these fits and, hence, the suggested inaccuracy is rather small. Comparing Tables III.4 and III.6 it is seen that the best fits have a lower error than E6 fits by a factor of three or greater. The absolute rms error with E6, however, is small. It is 1% of D_e for every potential except for fluorine for which it is 8%. The superiority of the EXEX function is of interest to the designers of force fields. For extended systems it is especially important to evaluate the ramifications of including the R^{-6} attractive term.

The fit of an E6 function to the crystal properties of the constructed FCC lattice is even more nonlinear in ζ than the simpler fit above. Table A.5 shows the fitted parameters. The two tables together can be used to convert potential functions. For instance, when E6 functions are fitted to crystal properties as Williams[13] has done, then from that a Morse pair function can be determined using Table A.5. If this is regarded as the best approximation of the COMP potential, Table A.4 can be used to find the equivalent short range E6 function. The Morse potential

Table A.4. E6 Parameters from Fits to the Morse Potential

κ^{Morse}	R_e	D_e	ζ^{E6}	rms error
35.48052	3.500604	0.10023	9.032252	0.067308
58	3.483963	0.099310	11.85337	0.753330
62	3.482686	0.099012	12.41642	0.890329
66	3.481666	0.098670	12.98374	1.032120
68	3.481245	0.098483	13.26778	1.104751
70	3.480879	0.098286	13.55151	1.178468
72	3.480567	0.098078	13.83458	1.253200
74	3.480306	0.097861	14.11666	1.328869
76	3.480096	0.097634	14.39747	1.405392
80	3.479820	0.097155	14.95428	1.560649
84	3.479726	0.096643	15.50333	1.718263
88	3.479803	0.096102	16.04333	1.877549
92	3.480039	0.095536	16.573305	2.037870

is essential for this kind of analysis to go between a long-range potential and its short-range equivalent, because it has no long range interaction. What is found is that short range potentials have larger D_e values while R_e is changed by a very small amount.

Another problem with using pair potentials is that in materials there are, unavoidably, many body interactions and screening over long distances. It will be difficult for *ab initio* methods to get a handle on this because any studies will require calculations on systems considerably larger than the four atom clusters in the present study. Nevertheless, some insight can be gained by examining the properties of a noble gas crystal such as argon and its corresponding pair potential about which much is known. The idea is that there is a pair potential which is accurate for a real

Table A.5. E6 Parameters Fit to FCC Lattice Properties With a Morse Potential

κ	ζ_{Morse}	R_e	D_e	ζ
58	10.77033	3.574252	0.082649	11.04212
62	11.13553	3.549434	0.083824	11.95026
66	11.48913	3.530262	0.084676	12.95736
68	11.6619	3.522314	0.084995	13.49442
70	11.83216	3.515282	0.085275	14.04894
72	12.0000	3.509014	0.085497	14.62337
74	12.16553	3.503411	0.085669	15.21683
76	12.32883	3.498391	0.085805	15.82753
80	12.64911	3.489843	0.086027	17.08976
84	12.96148	3.482847	0.086137	18.41736

pair of argon atoms as would be found from an *ab initio* calculation on Ar_2 or from scattering experiments. This potential then would predict certain properties in an atomistic simulation of the extended crystal of FCC argon. Conversely there is a potential which fits the properties of the crystal and it may differ from the real pair potential. Hence, a noble gas crystal can shed light on the fundamental rectitude of the pair potential approach.

A high level, MP4/QZ, calculation by Baker[14] gives a pair potential which can be fit by a Morse function with the parameters (R_e, D_e, ζ) = (3.8440, 0.2450, 12.2739). The properties of solid argon[15,16] are, at 0K, the nearest neighbor distance a = 3.756 Å, the cohesive energy U_{coh} = 1.846 kcal/mol, and the bulk modulus B = 2.86 GPa. These properties are fit with an E6 function with parameters (3.83, 0.227, 15.07). To compare with the pair potential the same properties of the crystal are fitted with a Morse potential giving (3.820, 0.266, 12.221). Between the

crystal and the pair Morse potentials the R_e and ζ values match well. The major discrepancy is with the potential well depth. The crystal fit value is 9% larger than the pair potential value.

This extra binding energy in the crystal represents some combination of three body interactions and long range dispersion over the lattice. Precisely, though, the long range dispersion given by a London potential does not match the van der Waals energy here. Using the polarizability (Table III.12) and the ionization potential (Table III.9) of argon in the London equation gives $B^{London} = 737.8\text{Å}^6\text{kcal/mol}$ while the parameters of the E6 function fit to the crystalline properties gives $B^{E6} = 1190.49$ in the same units. Clearly the interpretation of the attractive term in the E6 function as representing London dispersion is not accurate. Indeed it cannot be expected to be accurate because of the interdependencies of the E6 parameters. As is detailed in section A.1, the attractive tail is parameterized by a quantity, $\bar{B}$, which is one-to-one with the parameter ζ. The parameter B comprises dependencies on R_e and D_e besides on $\bar{B}$. In a comparison of a lattice potential to a pair potential, as above, however, the latter two cannot deviate much from the pair potential values. The only parametric control which exists over B is by ζ. The conflict is that ζ also is adjusted to fit the short range part of the potential. The question is where the balance is struck.

Curiously in the case of argon, there is no reasonable value of ζ which can give

$$B^{E6} = B^{London}. \qquad (A.34)$$

That is because $2R_e^6 D_e = 1,433$. So, in order for equation A.34 to be satisfied, $\bar{B}$ must be 0.514 which is not listed in Table A.1. In fact, this corresponds to a $\zeta = 215.7$ which is outrageous for a van der Waals potential. This suggests that the parameters of the E6 function may represent short range energetics more than long range. Moreover it appears that there is some extent of long range interaction in the crystal potential. It can be fit by Morse parameters with a slightly larger well

depth than the pair potential.

The main point is that the energetics of materials and extended systems may be modeled more accurately if the van der Waals pair potential binding energy is increased by some small factor between 1 and 1.08. While a simple consideration of R^{-6} sums in a lattice suggests a factor of 1.22, the smaller factor determined from the properties of $Ar_{(xtl)}$ implies that there is screening reducing the effective range of dispersion interactions.

Alternatively, in order to understand empirical van der Waals parameters from fits to properties of materials it may be useful to use the same factor to reduce the well depth and estimate a pair potential. That is, since empirical potentials are fit to materials, it is reasonable to conclude that pair potentials can be extrapolated using the FCC lattice comparisons. The DREIDING van der Waals parameters serve as an example. They are the result of fits using the E6 function and properties of crystalline solids.

If such severe screening is assumed then any parameters fit to crystal properties would be in error if they are fit using a lattice sum for all pair interactions. The biggest and most obvious error is underestimating D_e. The lattice sum overestimates the number of pair interactions so that to fit the crystal properties correctly, the optimized value of D_e must be artificially reduced. With perfect screening, however, the D_e of a potential fit to the lattice should be equal to an experimental pair potential.

This screening hypothesis can be tested with *ab initio* potentials and specifically with the COMP method. The level of calculation developed in Chapter II does not give the required accuracy. A test was done on linear Ar_3. Here the potential is defined by *ab initio* data from about 3 to 4 Å. The triatom had both near neighbor distances at 3.91 Å. The question is how the triatom energy compares to the sum of two E(3.91 Å) energies. It is expected that the triatom energy is in between that sum and that sum plus E(7.82 Å). If it is the former value then the potential is screened and if the latter value then the screening hypothesis is mostly invalidated.

Any intermediate value may define or allow one to define a screening coefficient by which next neighbor potentials are reduced.

The result with the COMP method instead demonstrates the limit of accuracy. One fact is that there is no data at the long distance, the *ab initio* data ends at much shorter R. So, a comparison was made to the energy at 7.82 Å based on the functional fit. The repulsive energy was compared first because the functional fit of the *ab initio* data, and, hence, the extrapolation is the most accurate. The trimer repulsive energy exceeds the two pair prediction by 0.67% . This repulsive energy of the two pairs is actually 0.260 kcal/mol. A problem is clear since the extrapolated repulsive energy of a pair at 7.82 Å is less than 10^{-7} kcal/mol. Moreover, any three body effect is suggested only for correlation energies and not for the repulsive energy at all. This elevation of the repulsive energy is purely an artifact of extending the COMP calculation to a trimeric cluster. This comparison of the repulsive energy actually precludes the need to compare total or attractive energy since this discrepancy is purely a result of the incompleteness of the COMP calculation.

Specifically, the trimer has two midpoint functions. For the outer pair, which is effectively being compared to the pair energy extrapolation, there are two midpoint functions and another atom (which can double as a midpoint set for stabilization purposes). With the presence of a BSS correction, the importance of this problem in designing the calculation is difficult to ascertain. What is clear is that for discriminating energies at the sub 1% level and at long R the capability of the current COMP approach is inadequate. Extension to this level of accuracy should be obtained with a more complete calculation such as an MP4/QZTP level. What is most important beyond that is diffuse functions. The actual effect that is in question is in the correlation energy and for a triatomic cluster investigation such as this to be illuminating some requirements must be satisfied. First, the COMP correlation energy must be meaningful at $2R$ for some value of R for comparison to the triatom. This necessitates using larger diffuse functions to increase the value of R_{max} as discussed in Section II.2. Second, the actual calculation must be correct

to this level of accuracy which means addressing the 0.67% problem which appears here. It must further be accurate enough to calculate the small correlation energy at $2R$. This extension of the COMP method to this level and the further basis optimization involved will become realistic with increased computational resources. This comparison will be a direct way to ascertain screening and is a valuable goal for the designer of a force field since, as shown above, screening is critical for modeling the energy and forces in a lattice.

Despite not having a clear criterion for screening and the value of using the London dispersion term in force fields, the above discussion provides some guidelines for applying van der Waals potential functions to materials. First, the accuracy of an E6 function is actually variable. It depends on characteristic points of the potential. For COMP data in particular the accuracy suggested by Table A.4 varies by over a factor of two depending on the curvature of the potential. Second, there are well defined ways to transform between the E6 and Morse potentials based on the lattice sum. The most important rule is that the well depth must be reduced in the E6 compared to the Morse because of the presence of R^{-6} energies in the sum. Third, there are some materials for which optimal potentials may be compared to pair potentials giving a clue to the forces really present in extended systems. It seems that the London dispersion, while theoretically appropriate for some isolated systems, is either inappropriate or unnecessary in simulations of systems both small and large.

References to Appendix A

1. J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids*, John Wiley & Sons, Inc., New York (1954).
2. R. A. Aziz, *et al.*, *J. Chem. Phys.*, **70**(9), 4330-4342 (1979).
3. P. W. Atkins, *Physical Chemistry*, W. H. Freeman and Company, San Francisco, CA (1978).
4. R. S. Berry, S. A. Rice, and J. Ross, *Physical Chemistry*, John Wiley and Sons, NY (1980).
5. G. C. Maitland, M. Rigby, E. B. Smith, and W. A. Makeham, *Intermolecular Forces: Their Origin and Determination*, Clarendon Press, Oxford (1981).
6. The molecular mechanics and molecular dynamics calculations used POLYGRAF (V3.21 and V3.3) from Molecular Simulation Inc. of Burlington, Massachusetts.
7. C. J. Casewit and A. K. Rappé, *Molecular Mechanics Across Chemistry, in press.*
8. R. S. Mulliken, C. A. Rieke, D. Orloff, and H. Orloff, *J. Chem Phys.*, **17**(12), 1248 (1949).
9. P. M. Morse, *Phys. Rev.*, **34**, 57 (1929).
10. G. Herzberg, *Molecular Spectra and Molecular Structure I. Diatomic Molecules*, van Nostrand Reinholdt, New York (1949).
11. D. D. Konowalow and D. S. Zakheim, *J. Chem. Phys.*, **57**(10), 4375-4387.
12. A. K. Rappé , *private communications.*
13. (a) D. E. Williams and D. J. Houpt, *Acta. Cryst.*, **B40**, 404-417 (1984). (b) D. E. Williams and D. J. Houpt, *Acta. Cryst.*, **A37**, 293-301 (1981). (c) D. E. Williams and D. J. Houpt, *Acta. Cryst.*, **B42**, 286-295 (1986).
14. J. A. Baker, *J. Chem. Phys.*, **79**(12), 6306-6307 (1983).
15. M. L. Klein and M. V. Bobetic, *Rare Gas Solids*, Pergamon, Oxford, UK (1980).

16. M. S. Anderson and C. A. Swenson, *J. Phys. Chem. Solids*, **36**, 145 (1975).

Appendix B. Data Tables for COMP Potentials.

The tables in this Appendix contain data from single point calculations using the COMP method described in the text. The left-hand column in each table is a separation of some sort. R_{cm} is the distance between centers of the diatoms in the cluster. Some tables also list the separation between atom centers, R. Because of symmetry most of the clusters have only a single interatomic distance. Some special configurations were done on N_2 and H_2 for which there are two or three interatomic distances in each cluster. For these, R is not listed. Columns to the right list the energy calculated at each distance. E_{Rep} and E_{Att} are monotonic (repulsive and attractive) energies and E_{COMP} is the total potential, the sum of the two monotonic energies. Unless otherwise specified all distances are in Ångstroms. Energies are in μh/mol for the nitrogen basis set optimization data, Tables B.1 through B.6. In all following tables, for COMP data, the energy is in kcal/mol.

In all figures included at the end of this Appendix the legend is uniform unless otherwise specified in the table legend. There are four functions plotted: EXEX (gray line), Morse (gray dashed line), E6 (dashed line), and LJ (solid line). Data points are plotted as hollow dots.

Table B.1. Nitrogen Repulsive Potential Tests I: DZSP Basis Set

R_{cm}	DZSPSZ	DZSPDZ	DZSPNZ
2.484	33900.3	33891.8	34933
2.722	14925.8	14935.5	15500
2.963	6402.01	6422.28	6675.1
3.205	2677.04	2683.98	2792.7
3.448	1084.33	1076.35	1134.6
3.570	678.145	666.734	715.06
3.692	416.923	404.866	447.09
3.814	250.317	239.505	277.22
3.937	145.248	136.52	170.28
4.135	53.176	47.659	75.558
4.457	1.02111	-1.9069	18.102
4.979	-8.7747	-10.936	0.3553

Table B.2. Nitrogen Repulsive Potential Tests II: DZDP Basis Set

R_{cm}	DZDPSZ	DZDPDZ	DZDPNZ
2.484	34933	34048.2	
2.722	15097.1	15154.2	
2.963	6532.86	6549.13	
3.205	2758.52	2765.1	2762.8
3.448	1132.06	1135.6	1122.2
3.570	715.561	716.69	705.76
3.692	447.069	445.93	439.97
3.814	275.184	272.59	271.85
3.937	166.046	162.85	166.42
4.135	68.774	65.665	
4.457	10.1324	7.6074	
4.979	-6.0392	-7.8152	

Table B.3. Nitrogen Repulsive Potential Tests III: TZDP Basis Set

R_{cm}	DZDP$^+$NZ	TZDPNZ	TZDP$^+$DZ	TZDP$^+$NZ
2.484	34620.8		34711.5	
2.722	15525.8	15511	15534.8	
2.963	6825.71		6821.1	6881.4
3.205	2944.43	2942.4	2944.43	2983.2
3.448	1249.86		1252.91	1276
3.570	811.197	806.18	813.85	831.35
3.692	526.125	518.86	527.598	540.84
3.814	341.576	332.41	341.623	351.7
3.937	222.39	211.98	221.171	
4.135	112.381	101.5	109.912	115.32
4.457	39.0834		36.3984	39.869
4.979	9.34758		7.80251	9.8165

Table B.4. Nitrogen Attractive Potential Tests I: DZSP Basis Set

R_{cm}	DZSPSZ	DZSPDZ	DZSPNZ
2.484	-10545	-11329	-7611
2.722	-6573.4	-6922.8	-4426
2.963	-4078.5	-4233.7	-2572.2
3.205	-2537	-2605.4	-1508
3.448	-1588.3	-1619.5	-898.1
3.570	-1260.2	-1283.3	-698.22
3.692	-1002	-1021	-545.9
3.814	-798.72	-815.88	-429.49
3.937	-638.68	-655	-340.25
4.135	-448.58	-463.9	-237.35
4.457	-258.98	-271.95	-138.61
4.979	-113.5	-122.39	-65.421

Table B.5. Nitrogen Attractive Potential Tests II: DZDP Basis Set

R_{cm}	DZDPSZ	DZDPDZ	DZDPNZ
2.484	-11536		
2.722	-7105.3	-7395.5	
2.963	-4366.9	-4509.7	
3.205	-2690.6	-2782.7	-2123
3.448	-1681.6	-1737.7	-1279
3.570	-1338.5	-1380.6	-1000
3.692	-1069.9	-1101.8	-786.6
3.814	-858.75	-883.68	-622.4
3.937	-692.12	-712.43	-495.9
4.135	-493.15	-508.79	
4.457	-292.63	-303.75	
4.979	-135.45	-142.85	

Table B.6. Nitrogen Attractive Potential Tests III: TZDP Basis Set

R_{cm}	DZDP^{+}NZ	TZDP^{+}DZ
2.484	-10297	-12129
2.722	-6084.6	-7356.7
2.963	-3597.6	-4476.9
3.205	-2147.7	-2752
3.448	-1305.9	-1716.7
3.570	-1027.9	-1364.8
3.692	-815.01	-1090.3
3.814	-651.25	-875.35
3.937	-524.58	-706.44
4.135	-376.03	-505.2
4.457	-227.88	-301.74
4.979	-110.49	-140.92
5.575	-40.585	

Table B.7. Nitrogen Cluster Potential in the Crossed Geometry

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.359	2.483925			
2.609	2.722477	9.614243	-4.42622227	5.14993
2.859	2.962918	4.178358	-2.70313775	1.46551
3.009	3.107906	2.502166	-2.02264805	0.48230
3.109	3.204822	1.767839	-1.67179984	0.10414
3.359	3.447881	0.742387	-1.03300931	-0.2933
3.484	3.569770	0.464104	-0.83127414	-0.3583
3.700	3.780873	0.209255	-0.56351292	-0.3490
3.734	3.814152	0.184574	-0.53573388	-0.3564
3.859	3.936608	0.116325	-0.42506433	-0.3074
4.061	4.134818	0.056378	-0.29885120	-0.2457
4.389	4.457389	0.017996	-0.17413956	-0.1617
4.918	4.979129	0.004319	-0.07848304	-0.0802

Table B.8. Nitrogen Cluster Potential in the T-shaped Geometry

R_{cm}	R_{in}	R_{out}	E_{rep}	E_{att}	E_{COMP}
3.30	2.804461	3.889087	4.456224	-1.862588	2.59364
3.50	3.000833	4.087175	2.197402	-1.27776	0.91964
3.75	3.246922	4.335032	0.884571	-0.796066	0.0885
4.00	3.493566	4.583121	0.340071	-0.516102	-0.1760
4.12	3.612118	4.702276	0.209801	-0.422817	-0.2130
4.25	3.740655	4.831408	0.121021	-0.342806	-0.2218
4.37	3.859391	4.950646	0.069336	-0.284058	-0.2147
4.50	3.988107	5.079862	0.046970	-0.233126	-0.1862
5.00	4.483860	5.577186	-0.004639	-0.115094	-0.1197
5.50	4.980462	6.074949	-0.007443		

Table B.9. Nitrogen Cluster Potential in the Linear Geometry

R_{cm}	E_{rep}	E_{att}	E_{COMP}
4.00	2.060554	-0.778600	1.28195
4.25	0.866865	-0.521656	0.34521
4.38	0.559414	-0.428287	0.13113
4.50	0.377987	-0.358498	0.01949
4.75	0.176264	-0.249909	-0.0736
4.90	0.116610	-0.202431	-0.0858
5.00	0.090455	-0.176329	-0.0859
5.10	0.071439	-0.153899	-0.0825
5.25	0.051871	-0.125987	-0.0741
5.50	0.033150	-0.091234	-0.0581
5.75	0.023186	-0.066952	-0.0438
6.00	0.017295	-0.049803	-0.0325

Table B.10. Carbon Potential

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.609	2.750151	3.943418	-2.16569	1.777727
2.976	3.100488	1.592777	-1.09425	0.498525
3.264	3.377891	0.749769	-0.67132	0.07845
3.49	3.596742	0.410144	-0.46305	-0.0529
3.668	3.769705	0.255348	-0.34741	-0.09206
3.808	3.906061	0.176759	-0.27797	-0.10121
3.918	4.013374	0.133012	-0.23375	-0.10074
4.028	4.120829	0.100601	-0.19694	-0.09634
4.206	4.294984	0.06482	-0.14993	-0.08511
4.494	4.577389	0.03301	-0.09779	-0.06478
4.96	5.035678	0.01233	-0.05111	-0.03878
5.714	5.779814	0.003377	-0.02022	-0.01685
10	10.03775	0	-0.00061	-0.00061

Table B.11. Nitrogen Potential

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.359	2.483925	5.420714		
2.609	2.722477	2.429156	-1.1466	1.282553
2.859	2.962918	1.066616	-0.69873	0.36789
3.109	3.204822	0.459125	-0.43071	0.028411
3.359	3.447881	0.19416	-0.2686	-0.07444
3.484	3.56977	0.125656	-0.21324	-0.08758
3.609	3.691867	0.081188	-0.17003	-0.08884
3.734	3.814152	0.052444	-0.13624	-0.0838
3.859	3.936608	0.033916	-0.10973	-0.07582
4.061	4.134818	0.016873	-0.07823	-0.06136
4.389	4.457389	0.005615	-0.04657	-0.04096
4.918	4.979129	0.001172	-0.02179	-0.02062
5.775	5.827146	0.000268		

Table B.12. Oxygen Potential

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.045	2.216771	8.029864	-2.13271	5.897155
2.412	2.559257	2.236027	-1.10186	1.134162
2.7	2.832322	0.797384	-0.63515	0.162231
2.926	3.048528	0.350766	-0.4067	-0.05593
3.104	3.219762	0.183318	-0.2851	-0.10178
3.244	3.354935	0.110085	-0.21567	-0.10558
3.354	3.461411	0.073729	-0.17347	-0.09974
3.464	3.568101	0.049316	-0.13986	-0.09054
3.642	3.741151	0.025559	-0.09939	-0.07383
3.93	4.022058	0.008569	-0.05856	-0.04999
4.396	4.478489	0.001303	-0.02572	-0.02442
5.15	5.220589	0.000152	-0.0089	-0.00875

Table B.13. Fluorine Potential

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.287	2.495444	1.294007	-0.23551	1.058500
2.575	2.761792	0.408743	-0.18806	0.220688
2.801	2.97363	0.167668	-0.14556	0.022111
2.979	3.141865	0.082084	-0.11462	-0.03254
3.119	3.27491	0.045983	-0.0932	-0.04722
3.229	3.379839	0.02874	-0.0784	-0.04966
3.339	3.485081	0.017693	-0.06543	-0.04773
3.517	3.655976	0.007784	-0.04811	-0.04033
3.805	3.933815	0.001866	-0.02844	-0.02658
4.271	4.38615	0.00018	-0.01166	-0.01148

Table B.14. Silicon Potential

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
3.50	3.843	2.167	-1.288	0.879
3.77	4.091	1.137	-0.933	0.204
3.98	4.285	0.670	-0.724	-0.054
4.14	4.434	0.438	-0.595	-0.157
4.27	4.556	0.307	-0.507	-0.200
4.40	4.678	0.212	-0.433	-0.221
4.60	4.866	0.117	-0.339	-0.222
4.94	5.189	0.041	-0.225	-0.185
5.48	5.705	0.008	-0.120	-0.112

Table B.15. Phosphorous Potential

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.669	2.986	-5.303	10.955	5.652
3.084	3.362	-3.247	5.611	2.364
3.409	3.663	-2.165	2.386	0.220
3.666	3.903	-1.557	1.336	-0.221
3.867	4.092	-1.198	0.837	-0.361
4.026	4.243	-0.972	0.575	-0.398
4.150	4.361	-0.826	0.428	-0.398
4.275	4.480	-0.701	0.317	-0.384
4.476	4.672	-0.539	0.196	-0.343
4.802	4.985	-0.355	0.090	-0.265
5.329	5.495	-0.185	0.025	-0.160

Table B.16. Sulfur Potential

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
3.286	3.547	1.502	-1.678	-0.176
3.544	3.787	0.724	-1.155	-0.431
3.746	3.977	0.407	-0.861	-0.454
3.905	4.127	0.259	-0.684	-0.425
4.03	4.246	0.182	-0.571	-0.388
4.155	4.364	0.129	-0.477	-0.348
4.357	4.557	0.074	-0.359	-0.284
4.685	4.872	0.031	-0.228	-0.197

Table B.17. Chlorine Potential

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.935	3.254	-1.629	1.977	0.348
3.245	3.536	-1.025	0.696	-0.329
3.489	3.762	-0.710	0.302	-0.408
3.681	3.940	-0.532	0.159	-0.374
3.832	4.082	-0.425	0.096	-0.329
3.95	4.193	-0.357	0.065	-0.292
4.069	4.305	-0.300	0.044	-0.255
4.26	4.486	-0.227	0.024	-0.203
4.57	4.781	-0.147	0.009	-0.138
5.072	5.263	-0.075	0.002	-0.073

Table B.18. Carbon-Oxygen Heterogeneous Pair Potential

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
3	3.121578	0.795775	-0.62136	0.174416
3.12	3.338394	0.396822	-0.41543	-0.0186
3.225	3.510649	0.228702	-0.30221	-0.07351
3.403	3.645547	0.148881	-0.23592	-0.08704
3.542	3.750567	0.106715	-0.19488	-0.08816
3.65	3.858675	0.075781	-0.16042	-0.08464
3.761	4.031389	0.043797	-0.11828	-0.07448
3.938	4.312178	0.017679	-0.07344	-0.05576
4.225	4.767701	0.0036	-0.03575	-0.03215

Table B.19. Nitrogen-Fluorine Heterogeneous Potential Data

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.106	2.288268	6.311824	-1.55327	4.758551
2.473	2.629955	1.678642	-0.7168	0.961846
2.761	2.902423	0.579936	-0.40436	0.175578
2.987	3.118189	0.248547	-0.26297	-0.01442
3.165	3.289097	0.126437	-0.18956	-0.06312
3.305	3.424027	0.073899	-0.14751	-0.07361
3.415	3.53032	0.048267	-0.12159	-0.07332
3.525	3.636834	0.031251	-0.10054	-0.06929
3.703	3.809612	0.015363	-0.07437	-0.059
3.991	4.090112	0.004358	-0.04644	-0.04208
4.457	4.545964	-4.3E-05	-0.02292	-0.02296
5.211	5.287292	-0.00045	-0.00865	-0.0091

Table B.20. Phosphorous-Chlorine Heterogeneous Potential Data

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.4	2.764931	16.47068	-4.46147	12.0092
2.8	3.118468	6.272185	-2.59741	3.67478
3.08	3.372128	3.058813	-1.76293	1.295885
3.3	3.574192	1.703277	-1.29529	0.407989
3.5	3.759634	0.987997	-0.9767	0.011293
3.64	3.890301	0.672111	-0.80091	-0.1288
3.75	3.993413	0.496247	-0.68506	-0.18881
3.86	4.096882	0.366605	-0.58587	-0.21926
4.04	4.266901	0.224132	-0.45101	-0.22688
4.34	4.551972	0.100155	-0.2934	-0.19325
4.82	5.011711	0.028681	-0.14994	-0.12126

Table B.21. Hydrogen Potential

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.373	2.430006	0.604532	-0.18651	0.418023
2.624	2.675664	0.268419	-0.1163	0.15212
2.822	2.870102	0.139719	-0.08013	0.059591
2.977	3.022636	0.083496	-0.05995	0.02355
3.099	3.142865	0.054921	-0.04744	0.007482
3.195	3.237565	0.039616	-0.03966	-4.4E-05
3.291	3.332339	0.028523	-0.03311	-0.00459
3.446	3.485501	0.016794	-0.02482	-0.00803
3.520	3.55868	0.013013	-0.02167	-0.00866
3.600	3.637829	0.009879	-0.01875	-0.00887
3.698	3.734837	0.007058	-0.01574	-0.00868
3.800	3.835857	0.004989	-0.01317	-0.00818
3.900	3.934946	0.003567	-0.01109	-0.00752
4.000	4.03408	0.002568	-0.00938	-0.00681
4.104	4.137223	0.001842	-0.00791	-0.00607
4.250	4.282091	0.001179	-0.00627	-0.00509
4.500	4.53032	0.00059	-0.00428	-0.00369
4.760	4.788674	0.000322	-0.00294	-0.00262

Table B.22. Hydrogen Cluster I. CRS

R_{cm}	R	E_{rep}	E_{att}	E_{COMP}
2.624	2.675664	1.047157	-0.37897	0.668184
2.822	2.870102	0.54324	-0.25365	0.289594
2.977	3.022636	0.323097	-0.18731	0.13579
3.195	3.237565	0.154607	-0.12411	0.030501
3.291	3.332339	0.111636	-0.10408	0.007555
3.500	3.538898	0.055082	-0.07177	-0.01669
3.600	3.637829	0.039428	-0.06042	-0.02099
3.700	3.736817	0.028345	-0.05106	-0.02271
3.850	3.885396	0.017484	-0.03994	-0.02246
4.000	4.034080	0.011001	-0.03151	-0.02051
4.400	4.431004	0.003699	-0.01743	-0.01373

Table B.23. Hydrogen Cluster II. TEE

R_{cm}	E_{rep}	E_{att}	E_{COMP}
2.624	1.203285	-0.49650	0.706787
2.822	0.607494	-0.33124	0.276250
2.977	0.349378	-0.24249	0.106884
3.195	0.154660	-0.15840	-0.00374
3.291	0.106091	-0.13209	-0.02600
3.446	0.055860	-0.09936	-0.04350
3.698	0.016845	-0.06400	-0.04715
4.104	-0.00113	-0.03357	-0.03470
4.500	-0.00345	-0.01936	-0.02282

Table B.24. Hydrogen Cluster III. LIN

R_{cm}	E_{rep}	E_{att}	E_{COMP}
3.250	0.305638		
3.416	0.159126	-0.16122	-0.0021
3.610	0.091213	-0.11414	-0.02293
3.698	0.071637	-0.09806	-0.02642
3.763	0.060240	-0.08781	-0.02757
3.978	0.035213	-0.06166	-0.02644
4.300	0.017772	-0.03746	-0.01969
4.475	0.013049	-0.02900	-0.01596
4.877	0.007374	-0.01673	-0.00935

Figures

Figure B.1 Carbon COMP potential and functional fits.

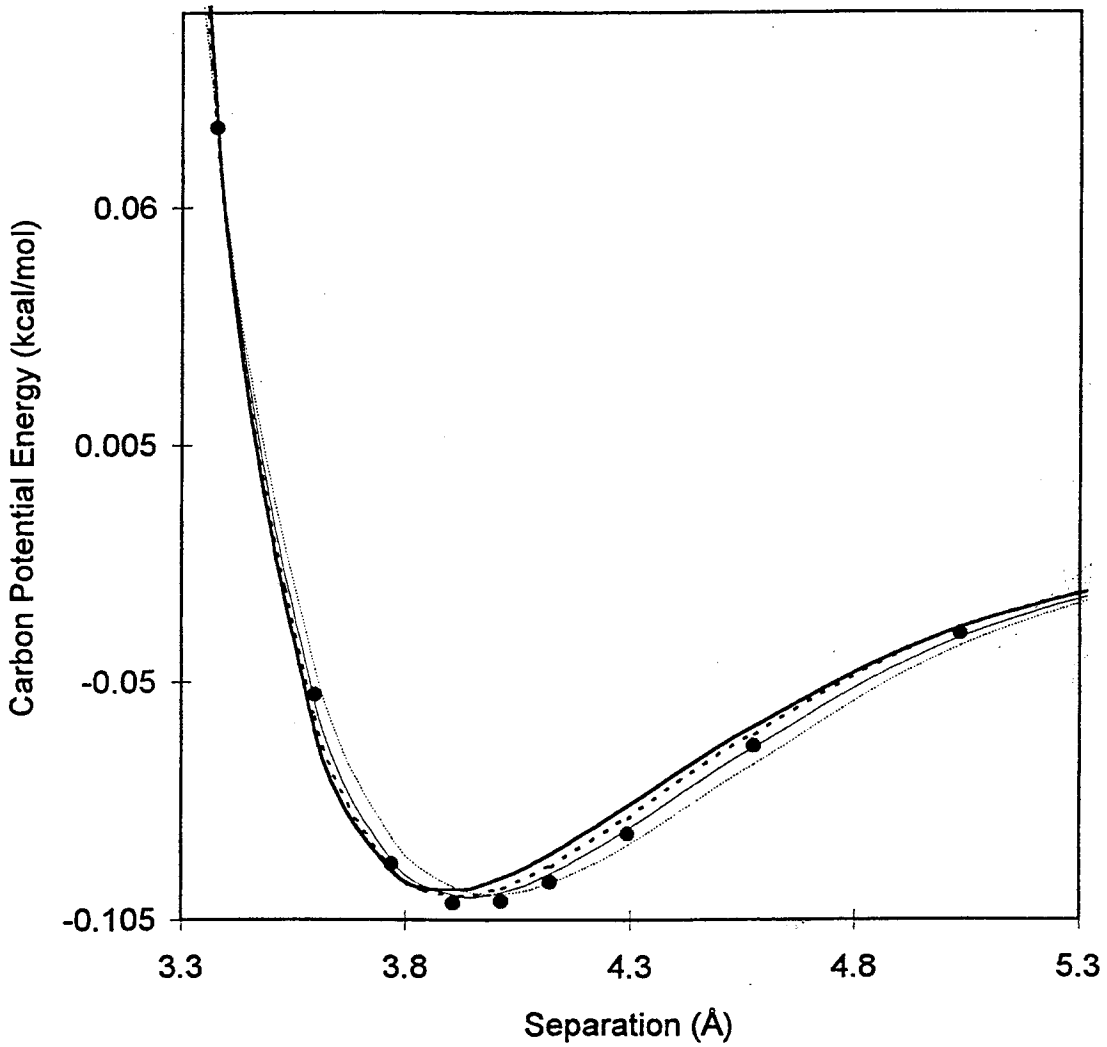


Figure B.2 Nitrogen COMP with functional fits.

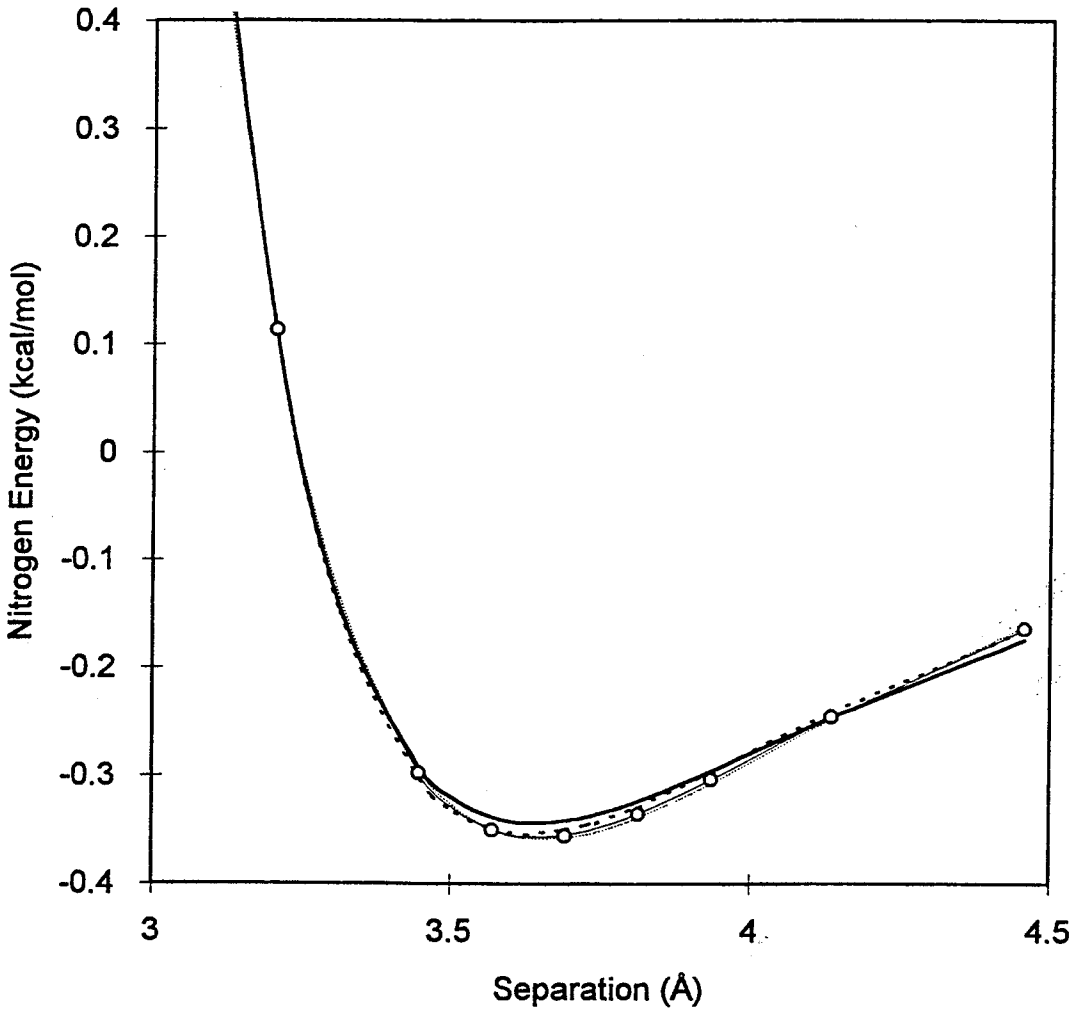


Figure B.3 Oxygen COMP potential plus functional fits.

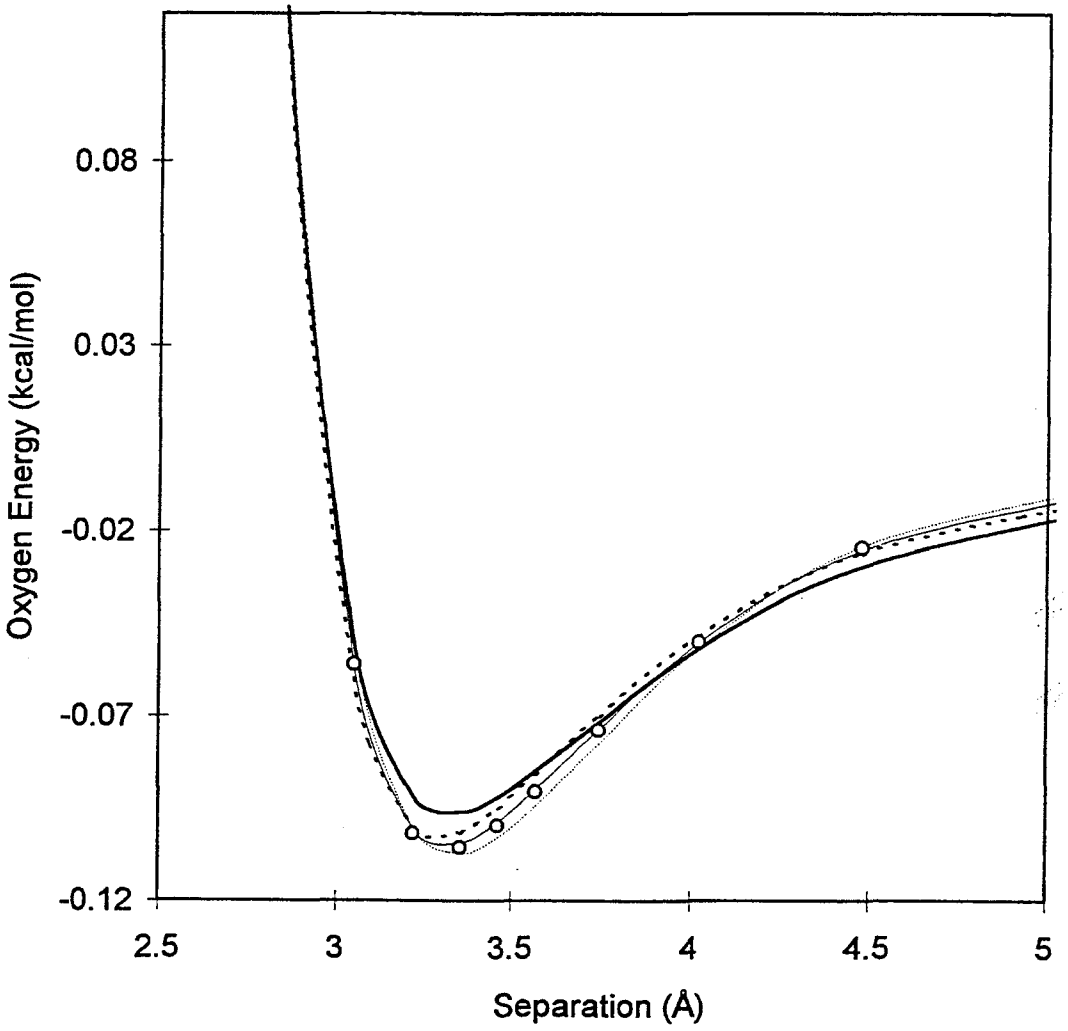


Figure B.4 Fluorine COMP potential with functional fits.

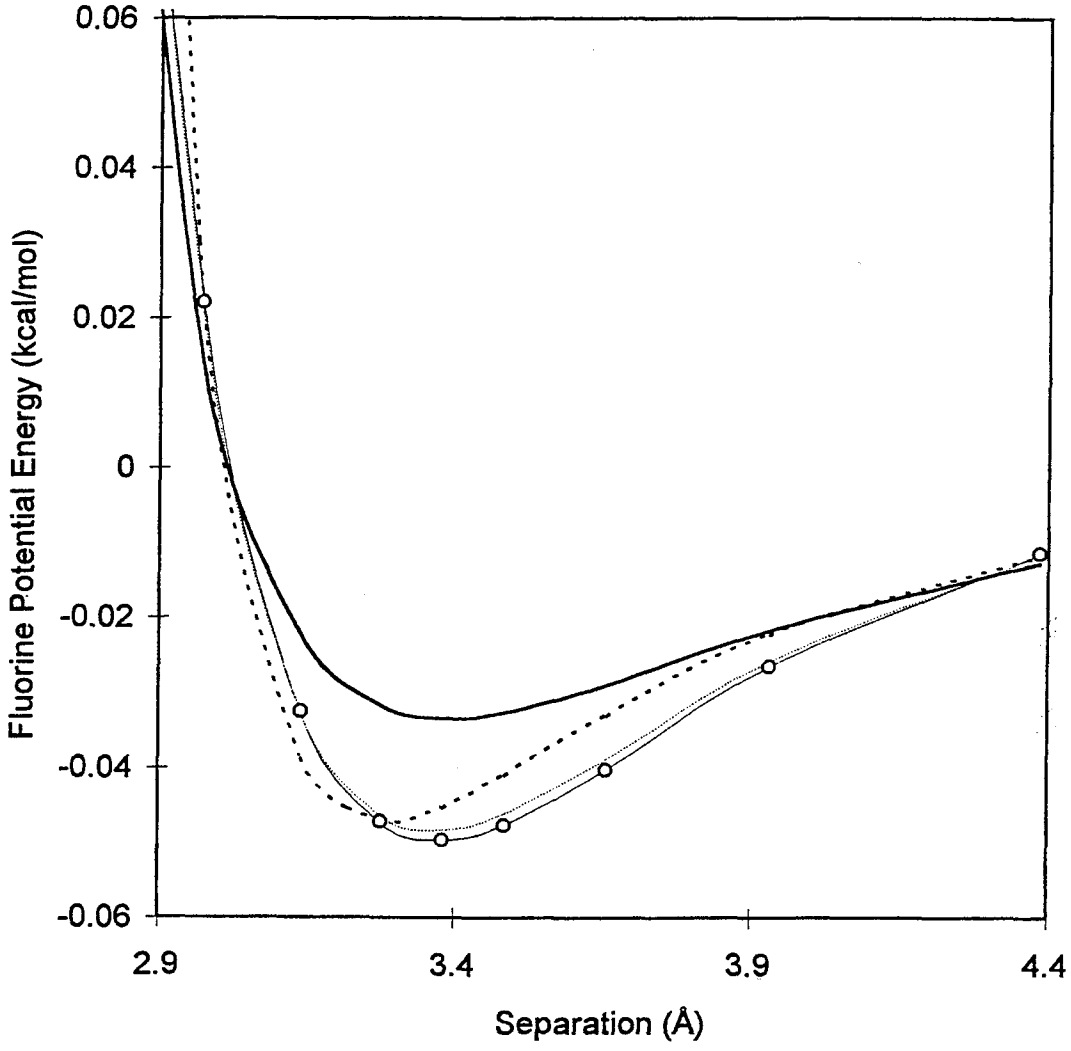


Figure B.5 Silicon COMP potential and functional fits.

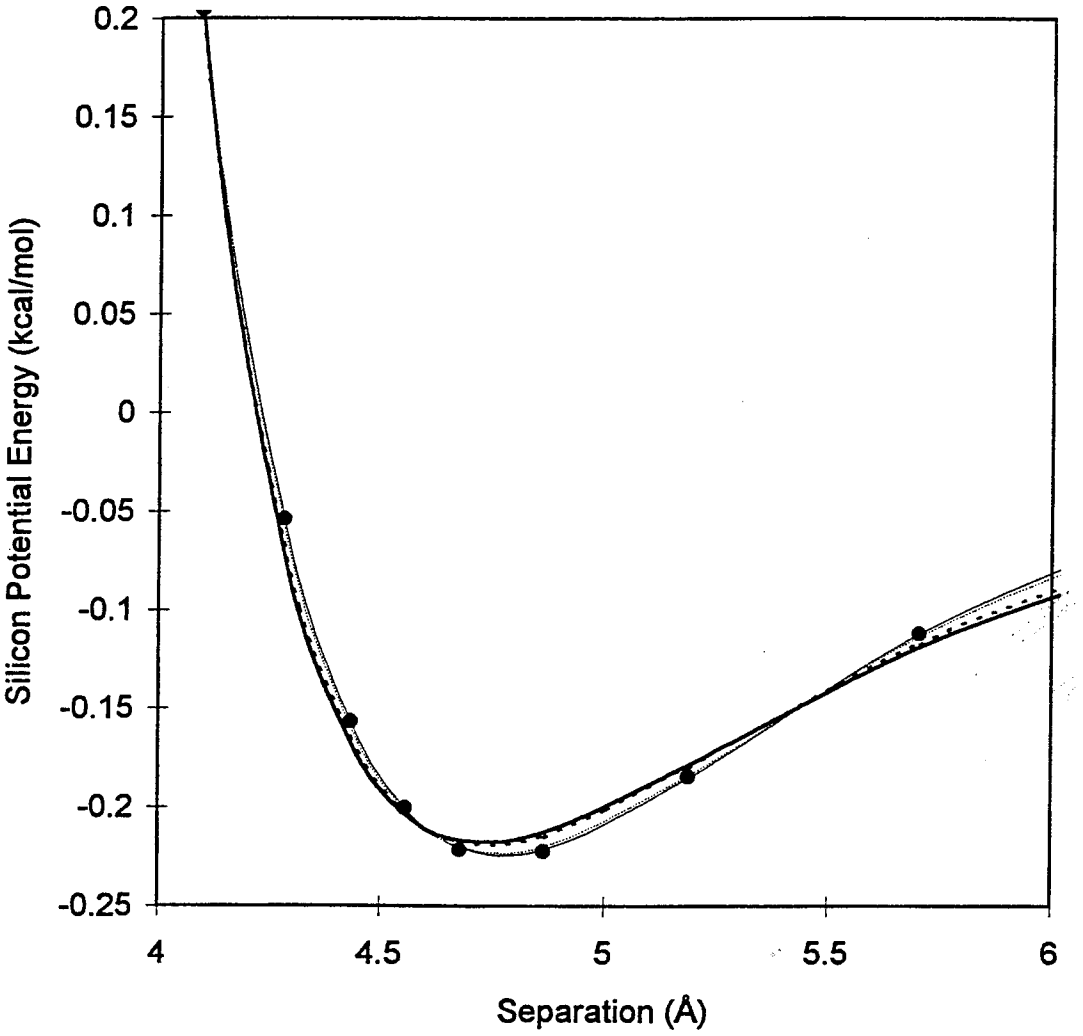


Figure B.6 Phosphorous COMP potential and functional fits.

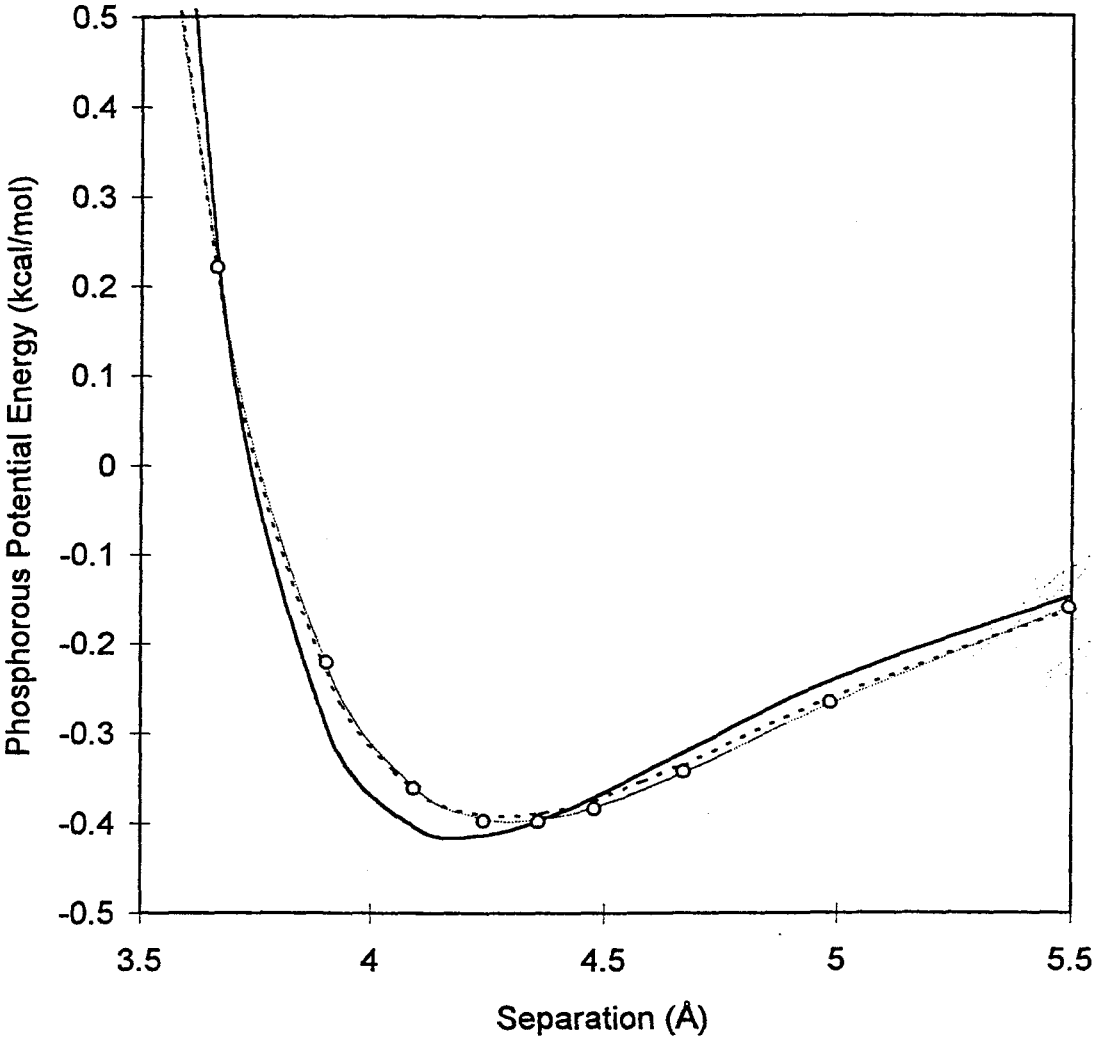


Figure B.7 The sulfur COMP potential and functional fits.

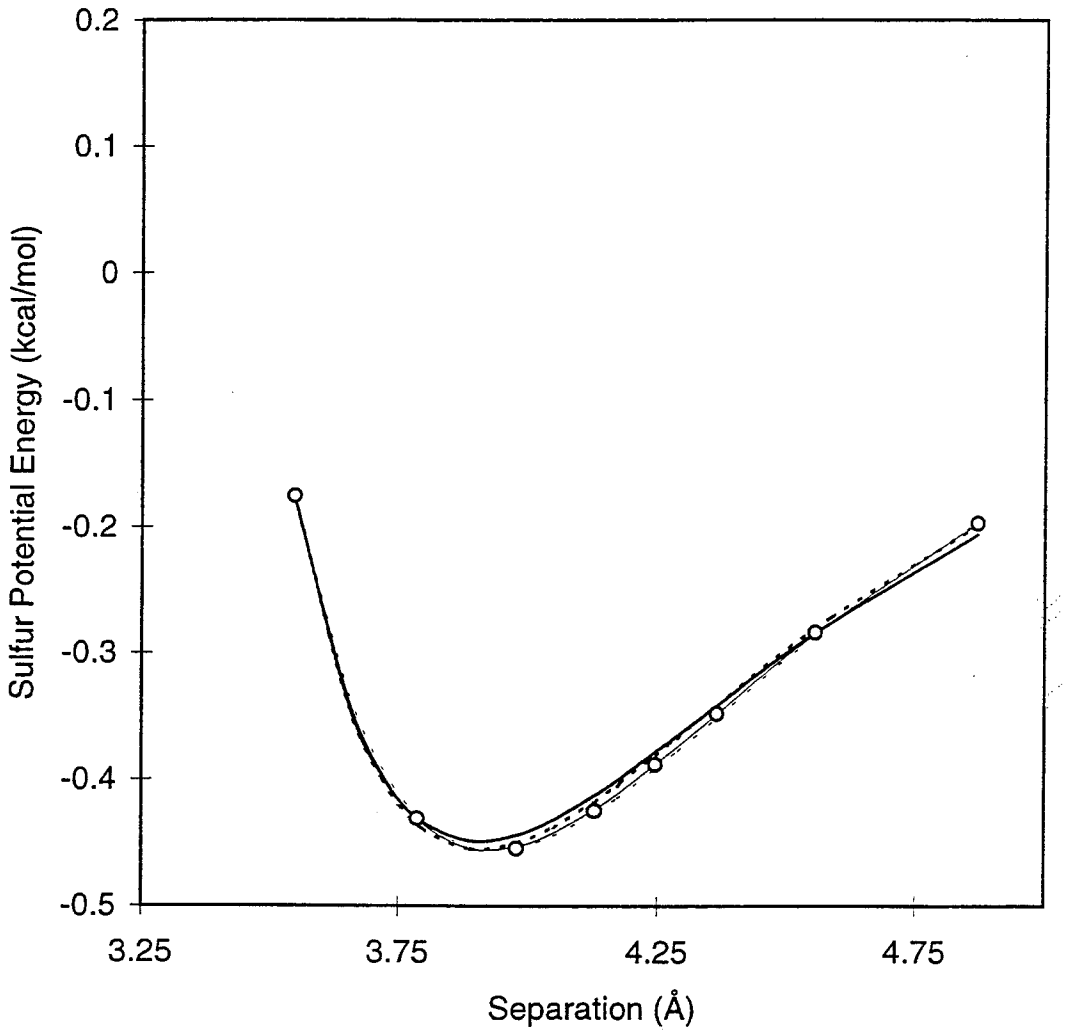


Figure B.8 Chlorine COMP potential and functional fits.

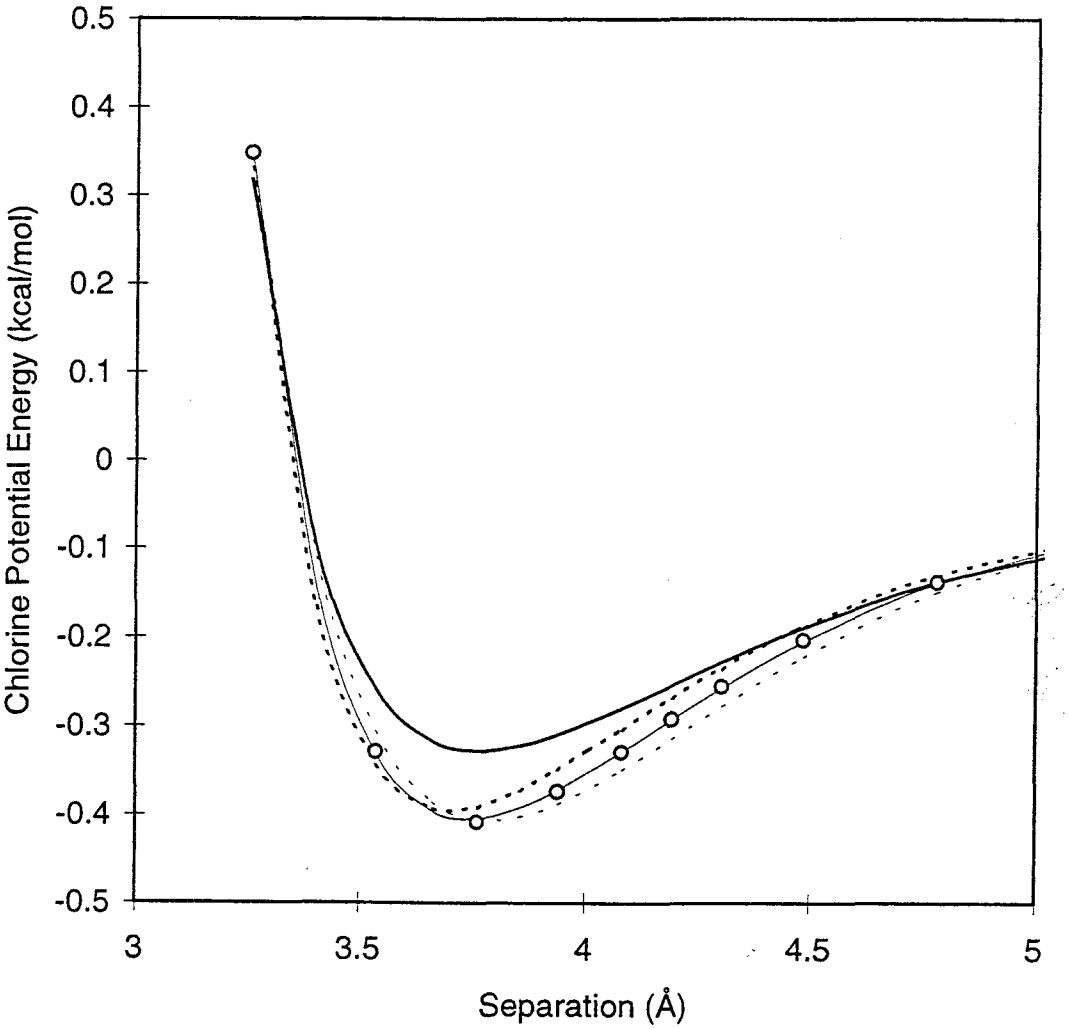


Figure B.9 Hydrogen COMP potential with functional fits.

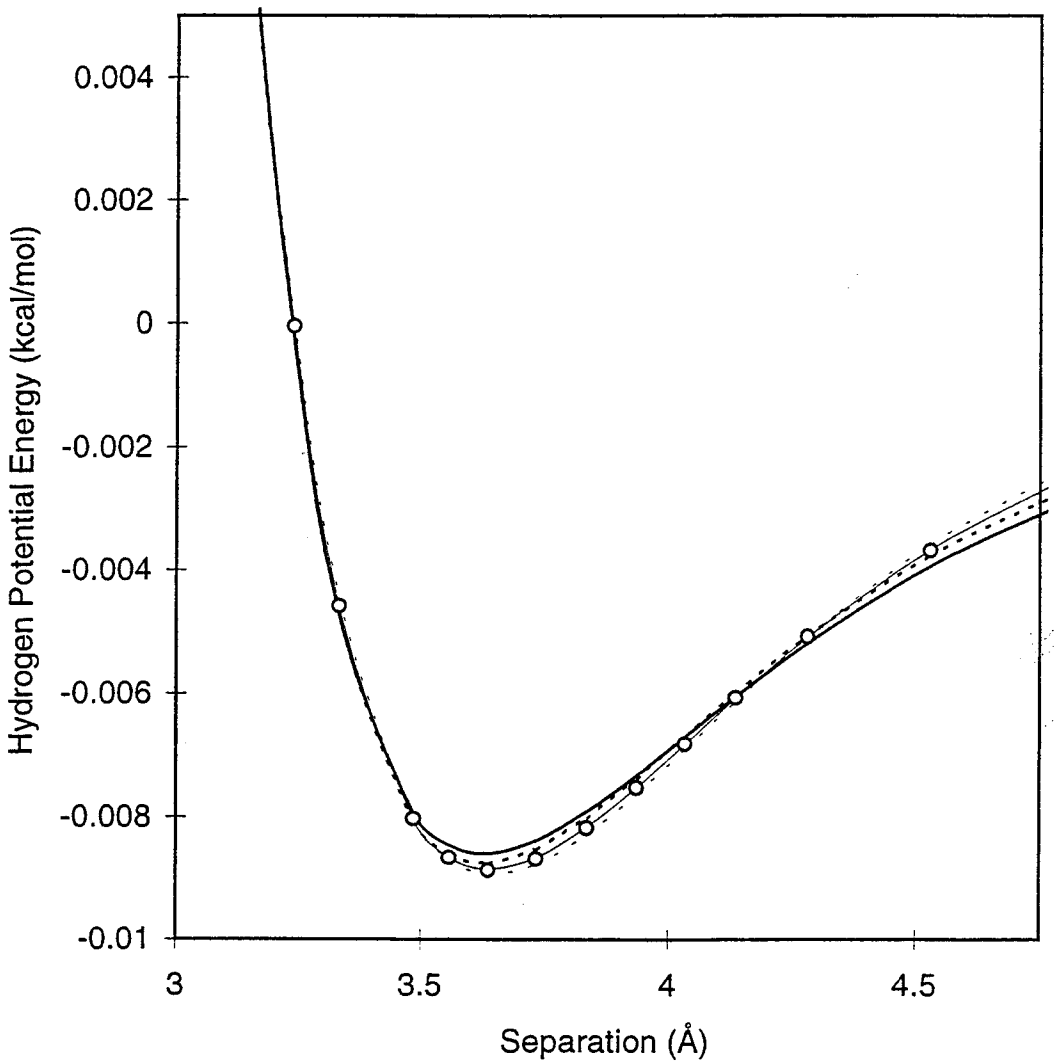


Figure B.10 C-O heterogeneous COMP potential and fits.

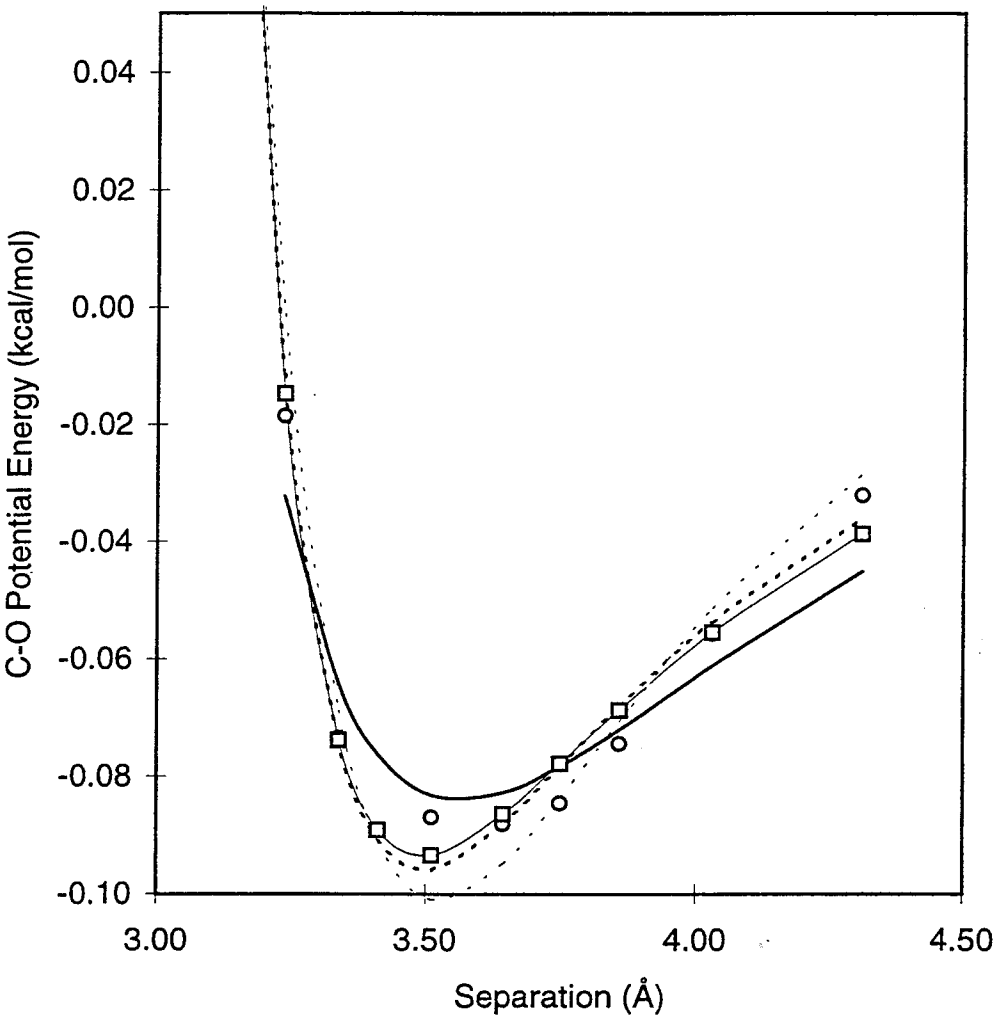


Figure B.11 N-F heterogeneous COMP potential with fits.

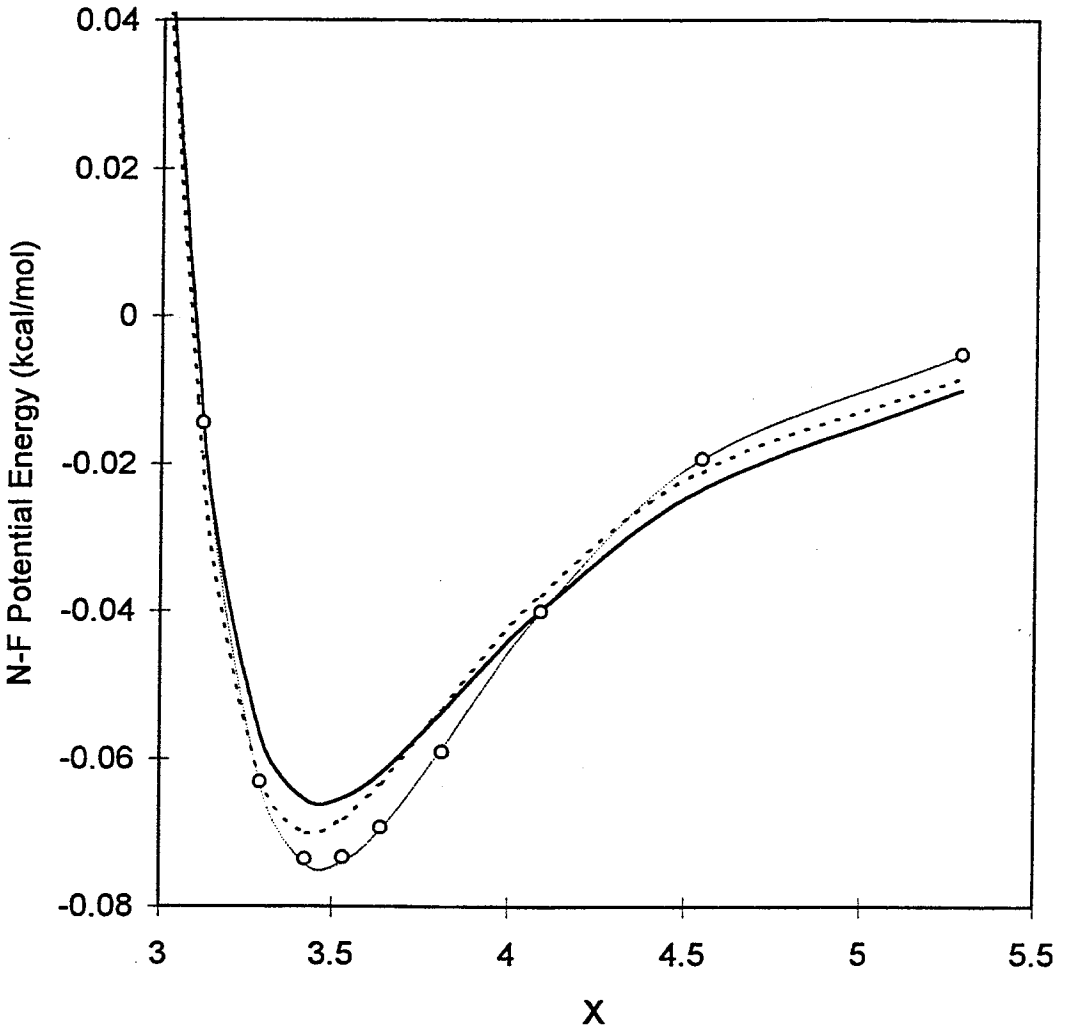
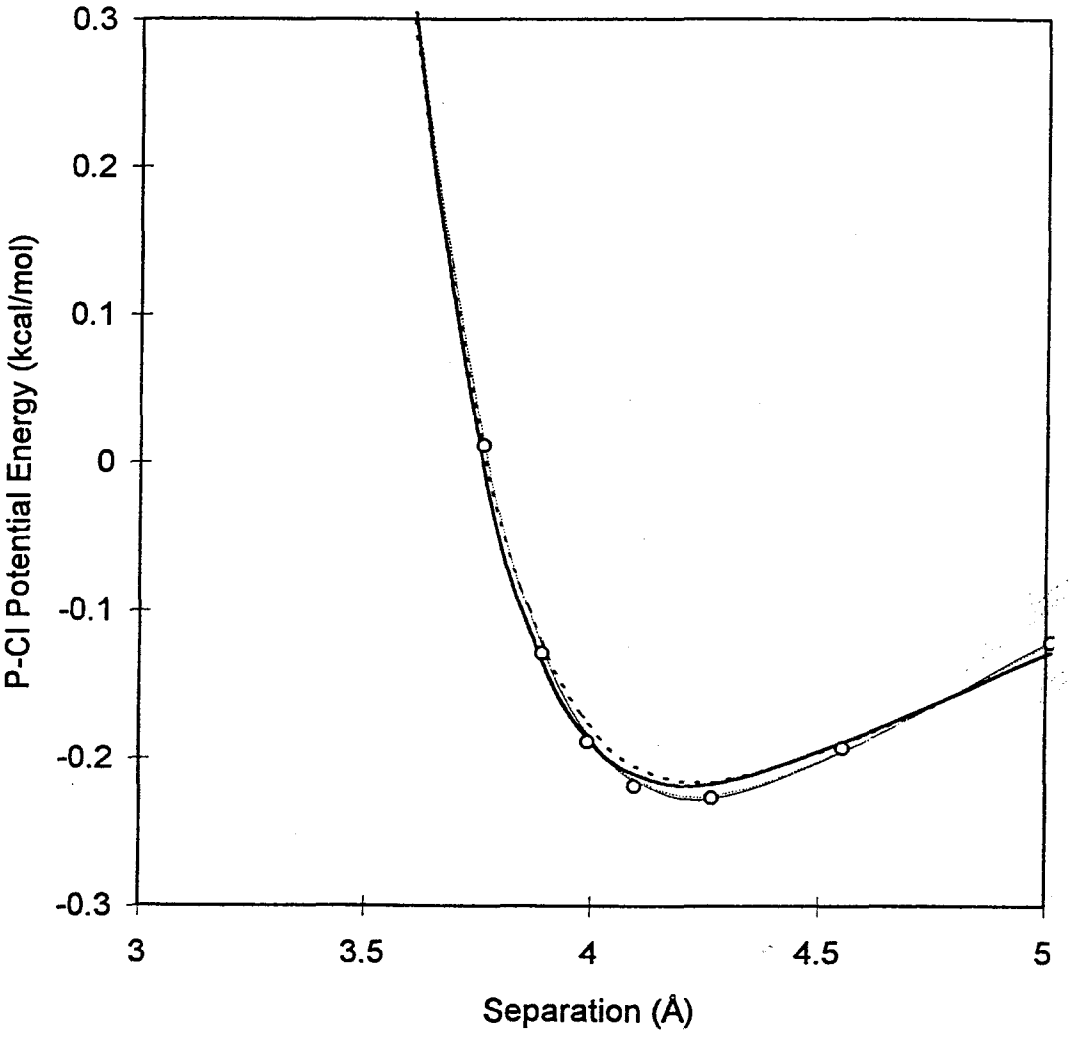


Figure B.12 P-Cl heterogeneous COMP potential and fits.



Appendix C. Femtochemistry.

The included material below consists of two papers in the effort of the Zewail group to study the dynamics of the chemical bond through femtosecond spectroscopy, femtochemistry. These two papers deal with the principal issues of the femtochemistry of both bound and unbound states. The first paper, *Femtosecond Selective Control of Wave Packet Population*, explains the beats or resonance patter in the probe intensity as a function of time. This work followed about a year of study of the iodine system experimentally and at the time the beats were part of the intriguing results which also led to the ability to invert a femtochemistry result into frequency space to obtain a potential energy surface of the excited state.

The second paper, *Femtosecond Real-Time Probing of Reactions 6. A Joint Experimental and Theoretical Study of Bi_2 Dissociation*, is part of Professor Zewail's effort on diatoms. Bi_2 is amenable to femtochemistry because of the large nuclear mass and, hence, the rather slow dissociation speed. The femtosecond spectroscopy in this case elucidated the topology of dissociative states.

Appendix D. Atomistic Simulations on 2-D Hydrocarbon Systems.

One example of the effect of van der Waals forces on molecular ordering is the self-assembled monolayer (SAM) class of materials. These materials consist of alkyl chains which organize into a two-dimensional crystalline lattice while affixed to a metal surface (usually gold) by an α-thiol functional group. The following paper studies the nature of the bonding between the sulfur atoms in response to a previous x-ray structure paper. What is also examined here is the energetics of different packing symmetries. The packing of straight chain alkanes is understood in terms of crystalline polyethylene which has an optimal structure and a set of stable structures of different symmetry.

Besides addressing an important issue in SAM science, this paper has aided a wider effort in the Goddard group to characterize alkyl tail packing in monolayers of more general types. Specifically the scheme of considering a monolayer to be, in an abstract sense, a slab sectioned from crystalline polyethylene gives a clear way to study the energetics. Because the energetics of polyethylene crystal can be investigated further for new symmetries, a method of extrapolating to a monolayer of tails can help find stable structures. The basic problem is that the nearest neighbor spacing of a monolayer is constrained by the surface binding sites and is usually some distance away from the ideal spacing in the crystal. The compromise, which often involves some sort of leaning or canting, comes out of the work below.

Included below is a preprint of the paper prepared this year by W. A. Goddard III and the author.

3/23/95

. *The Journal of the American Chemical Society*

The Atomistic Structure for Self-Assembled Monolayers of Alkane Thiols on Au (111) Surfaces

James J. Gerdy and William A. Goddard III*

Materials and Process Simulation Center, Beckman Institute (139-74)
Division of Chemistry and Chemical Engineering (CN 9070)
California Institute of Technology, Pasadena, California 91125

Abstract

Using force fields based on quantum mechanics we calculate the stable crystal structure for decanethiol dimers on Au (111). We find that the optimum structure leads to an X-ray diffraction pattern identical with that recently determined by Eisenberger and coworkers. This confirms the $c(4 \times 2)$ cell determined by He and X-ray diffraction, providing the first atomistic structural description of a self-assembled monolayer (SAM). The excellent agreement between theory and experiment suggests that theory could be used to predict structures for new SAM's, allowing the design to be carried out in advance of experiment.

*To whom correspondence should be addressed.

D-3

The Atomistic Structure for Self-Assembled Monolayers of Alkane Thiols on Au (111) Surfaces

James J. Gerdy and William A. Goddard III

Materials and Process Simulation Center, Beckman Institute (139-74)
Division of Chemistry and Chemical Engineering (CN 9070)
California Institute of Technology, Pasadena, California 91125

Self-assembly of molecules provides one of the most promising approaches to synthesizing nanoscale devices (spatial sizes of 10 nm or below).[1] Of particular interest are the self-assembled monolayer (SAM) two-dimensional nanostructures.[2–7] Starting with early work involving alkanethiols on Au (111),[8–10] simple straightforward methods have been developed for preparing well ordered monolayers and multilayers. There remain, however, a number of uncertainties about the structural characteristics of these films. Indeed, for the original alkanethiols, the chemical character of the sulfur atom at the gold surface was not established. The starting material is an alkanethiol (1a). It was widely thought that the thiols deprotonate at the surface, leading to thiolates (1b),[1,2,8] but there was little direct evidence for this view.

$$RSH \tag{1a}$$

$$RS^{-} \tag{1b}$$

$$RSSR \tag{1c}$$

where

$$R = -(CH_2)_{n-1}(CH_3) \tag{1d}$$

Some vibrational evidence indicated that the thiols might form disulfides (1c) at the surface (using EELS, Nuzzo, and coworkers[9] observed $S-S$ stretch and $C-S-S$ bend vibrations); however, the results were not consistent over samples.

A breakthrough concerning the structure of these systems was reported recently (December 1994) in Science by Eisenberger and coworkers.[14] Using grazing-angle X-ray diffraction of well annealed overlayers of decanethiols on Au [111], they were able to establish that these thiols form *dimers*, (1c), packed into a two-dimensional crystal structure commensurate with the Au [111] substrate. This leads to a $c(2 \times 4)$ unit cell with the alkane tails canted by $\theta = 37° \pm 2°$ from the surface normal. Although the data does not determine the atomic details, it does establish that the all-trans alkane tails form a pseudo closest packed array with a structure *different* than the normal orthorhombic polyethylene crystal (confirming earlier work by Scoles and coworkers[19]).

These experiments stimulated us to examine the detailed atomic arrangements using molecular dynamics and molecular mechanics simulations. For these studies we started with the DREIDING[15] force field (FF) but modified the $S-S$ dihedral potential to match the rotational barrier calculated from quantum chemistry (QC) studies[16] of diethyldisulfide (this is denoted D/S). Second, we modified the alkane part of this force field to match one developed[17] for polyethylene crystal. This is denoted as D/S,PE. We expect S to prefer binding to three-fold hollow sites on the surface. This was incorporated into the force field by using a bond interaction from each sulfur atom to a gold atom one layer down from the surface.

First, we examined the isolated alkanethiol dimer (1c). We found the parallel structure in Figure 1 where the top view shows that the two alkane tails are parallel and in contact while the bottom view (looking *along* the direction of the alkyl tail) shows that the two parallel all-trans chains have perpendicular CC planes. The two chains have average $C-C$ contacts of 4.4Å (the contact distances between interior hydrogens is about 2.9Å). Forcing the $C-C$ planes to be *parallel* and minimizing the structure leads to an energy 3.0 kcal/mol higher than Figure 1, and we did not consider this further.

The second important structure for (1c) has the all-trans chains splayed away from each other. The van der Waals contacts favor the parallel structure (Figure

1), stabilizing it by 0.67 kcal/mol for each carbon added to both chains. On the other hand, the splayed structure minimizes strain in the $C-S-S-C$ turn while the parallel structure has 3.0 kcal/mol of strain energy in the turn. For chain lengths longer than 13, the parallel structure is favored. For the chain length of 10 used here and in the X-ray study,[14] the splayed structure is favored by 1.85 kcal/mol. However, the parallel structure should lead to much greater bond energy to the Au surface, stabilizing it over the splayed structure (which is also disfavored by the small saturation coverage). In addition, the packing energy of the monolayer stabilizes the parallel dimer by an additional 18 kcal/mol per dimer.

In order to understand how these disulfide units pack to form a dense monolayer, we first examined the structures and energetics for various packings of infinite all-trans alkyl chains to form polyethylene crystals. As indicated in Figure 2, there are two simple classes:

1. L with all CC planes parallel, and
2. T with half the CC planes perpendicular to the other half.

Each arrow in Figure 2 represents the orientation of the alkyl tail. The CC plane lies in the plane indicated by the arrow and the arrowhead shows the phase of a CH_2 in this plane. These calculations indicate a number of L and T type structures with comparable energies (see Table 1). T is the observed structure of PE. Including zero point energy the experimental cohesive energy at 0K for PE is 2.02 kcal/mol,[17] in good agreement with these calculations (2.07 kcal/mol).

The restriction that pairs of adjacent chains have perpendicular CC planes (in order to attach to the $S-S$ linkage) restricts SAM considerations to the T and T^* structures. The energies of the optimized structures for various packings of thiol dimers on Au (111) are given in Table 2. T^* is lowest and the optimum T^* structure is shown in Figure 3. The alkane tails leads to a $c(4 \times 2)$ pattern with pseudo close packing. This compares well with the X-ray diffraction[14] and He diffraction[19] studies which determined that a monolayer of alkanethiol on Au (111) has the $c(4 \times 2)$ packing corresponding to our T^* structure. The theory finds

the alkane thiols to be canted at an angle to the surface normal of $\theta = 28.3 \pm 1.5°$ (the canting is parallel to the a axis). Reference 14 reports $\theta = 37 \pm 2°$ with a setting angle $\phi = 21°$; however, details of how this is obtained from the data are not provided.

We show in Figure 4a the X-ray diffraction intensities predicted from the T^* structure (this pattern is for an out-of-the-plane momentum of $Q_z = 0.1/\text{A}$ in order to compare directly with experiment). The experimental diffraction pattern is shown in Figure 4c where we see an excellent fit with the predictions for T^*. In particular, both theory and experiment find systematic extinctions for (1,0), (0,1), and (1,2). Due to the canting of the alkyl tails (and the nonzero Q_s), the theory gives slightly different intensities in the four quadrants of Q_x, Q_y. Averaging over these four quadrants leads to Figure 4b which has intensities in excellent agreement with experiment (Figure 4c). The slight differences might be because the theory is for $T = 0\text{K}$ while the experiment is for $T = 300\text{K}$.

Thus we conclude that the lowest energy T^* structure predicted from the theory is identical to the fully annealed structure observed by experiment. This is very encouraging because it indicates that theory might be used to predict the most stable structures *in advance of experiment.* This would be essential for designing new nanostructures where one might choose support, structure of the overlayer, and additives to optimize the properties.

The procedure for predicting the structure for a new SAM can be illustrated for the case considered herein. In the T structure, each alkane take has six neighbors (pseudo close packing) with four at 4.35Å and two at 4.81Å. In T^* there are four neighbors at 4.38Å and two at 4.87Å. For Au (111) the nearest neighbor spacing is 2.87Å, too small for the alkane thiols. The smallest spacing of Au (111) compatible with the alkane spacing is 4.97Å. Thus we would expect the dimer to orient so that the two neighbors preferring 4.81Å or 4.87Å increase this spacing slightly to attain 4.97Å. This might require less strain for T^*, perhaps explaining its preference over T. With some spacings increased from 4.87 to 4.97 the others should decrease from

4.38 to 4.35Å (assuming constant area per alkane). In order to match the spacings of Au (111), this requires canting of the alkane thiols. The amount of canting can be predicted by assuming the spacing of 4.35Å. This leads to a projected spacing of 3.77 $= 4.35 \times\sqrt{3}/2$ (where the geometric factor corrects for the distance perpendicular to the canting plane). The spacing of Au atoms is $4.97 \times\sqrt{3}/2 = 4.30$. Thus the canting angle is $\cos\theta = 3.77/4.30$ or $\theta = 28.7°$. The optimum angle calculated from the simulations is $\theta = 28.3°$, in excellent agreement. Starting with this packing we would carry out molecular dynamics studies to obtain the optimum structure and properties. To consider a different alkane thiol, say polypropylene (PP), polyvinyl chloride (PVC), polytetrafluoroethylene (PTFE), or nylon, we would first consider the packing of the isolated chains (from theory or experiment). Using these spacings we would consider the minimum spacing for attachment to the support. Based on this we would terminate the alkyl chain with a suitable group for self-assembled attachment to the support. The modification of tail spacings to fit the actual spacing of the support can then be predicted along with the canting angle. With these starting structures we would carry out molecular mechanics and molecular dynamics calculations to obtain more accurate predictions of structure and properties (cohesive energy, vibrational frequencies, entropy, specific heat). This would be followed by experiments focusing on the best cases.

Acknowledgements

The research was funded by DOE-AICD. The facilities of the MSC are also supported by grants from NSF (CHE 94-13930 and ASC 92-100368). Allied-Signal Corp., Asahi Chemical, Asahi Glass, BP CHemical, Chevron Petroleum Technology, Chevron Oronite, Xerox, Hughes Research Lab., and Beckman Institute.

References

1. A. Ulman, R. in *Ultrathin Films* (Wiley Interscience, New York, 1989).
2. A. Ulman and R. P. Scaringe, Langmuir **8**, 894-897 (1992).
3. L. H. Dubois and R. G. Nuzzo, Ann. Rev. Phys. Chem. **43**, 437 (1993).

4. J. D. Swalen, D. L. Allara, J. D. Andrade, E. A. Chandross, S. Garoff, J. Isrealachvili, T. J. McCarthy, R. Murray, R. F. Pease, J. F. Rabolt, K. J. Wynne, and H. Yu, Langmuir **3**, 932-950 (1987).
5. S. J. Stranick. A. N. Parikh, Y-T. Tao, D. L. Allara, and P. S. Weiss, J. Phys. Chem. **98**, 7637 (1994).
6. J. A. M. Sandag-Huethorst, C. Schonenberger, and L. G. J. Fokkink, J. Phys. Chem. **98**, 6826 (1994).
7. J. Xu, H-L. Li, and Y. Zhang, J. Phys. Chem. **97**, 11497 (1993).
8. (a) L. Strong and G. M. Whitesides, Langmuir **4**, 546-558 (1988). (b) C. D. Bain, H. A. Biebuyck, and G. M. Whitesides, Langmuir **5**, 723-727 (1989).
9. R. G. Nuzzo, F. A. Fusco, and D. L. Allara, J. Am. Chem. Soc. **109**, 2358-2368 (1987).
10. D. L. Allara and R. G. Nuzzo, Langmuir **1**, 45-52 (1985).
11. R. G. Nuzzo, L. H. Dubois, and D. L. Allara, J. Am. Chem. Soc. **112**, 558 (1990).
12. R. G. Nuzzo, B. R. Zegarski, and L. H. Dubois, J. Am. Chem. Soc. **109**, 733 (1987).
13. (a) C. Chidsey and D. Loiacon, Langmuir **6**, 682 (1990). (b) M. D. Porter, T. B. Bright, D. L. Allara, and C. E. D. Chidsey, J. Am. Chem. Soc. **109**, 3559 (1987).
14. P. Fenter, A. Eberhardt, and P. Eisenberger, Science **266**, 1216-1218 (1994).
15. S. L. Mayo, B. D. Olafson, and W. A. Goddard III, J. Phys. Chem. **94**, 8897-8909 (1990).
16. C. H. Goerbitz, J. Phys. Org. Chem. **7**, 259-267 (1994).
17. N. Karasawa and W. A. Goddard III, J. Phys. Chem. **95**, 2260 (1991). The MC form was used for the valence interactions, the Lennard-Jones 12-6 form was used for van der Waals, and charges were ignored.
18. A. R. Rappè, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, J. Am. Chem. Soc. **114**, 10024 (1992).

19. N. Camillone III, C. E. D. Chidsey, G. Lu, and G. Scoles, J. Chem. Phys. **98**, 3503 (1993).

20. The molecular mechanics and molecular dynamics calculations used POLYGRAF (V 3.3) and the X-ray diffraction calculations used CERIUS^2 both from Molecular Simulation Inc. of Burlington, Massachusetts.

Table 1. Cohesive energy (kcal/mol per CH_2) of polyethylene crystals.

	Cohesive Energy		Structure			
Lattice	D/S[d]	D/S,PE[d]	a Å	b Å	γ deg	ϕ[c] deg
$L11$	2.14	2.08	4.27	4.74	120	35.9°
$L12$	1.99	2.02	4.96	8.06	120	10.9°
$L22$	2.03	2.06	8.03	10.28	120	6.2°
T[a]	2.09	2.07[b]	7.19	4.81	90	42.0°
T^*	1.98	2.00	7.29	9.73	90	43.3°

[a]This corresponds to the observed orthohombic structures of PE.

[b] The experimental value 2.02 kcal/mol (see reference 17).

[c]Setting angle.

[d]The force field, see text.

Table 2. Cohesive energy and structure of alkylsulfide dimer monolayers.

Structure	Cohesive[a] Energy	Unit Cell	Canting[b] (degree)
$L12$	17.16	(1×2)	$26.2° \pm 1.2°$
$L22$	15.43	(2×2)	$26.5° \pm 2.2°$
T	18.42	(1×2)	$26.7° \pm 0.5°$
T^*	18.64	$c(2 \times 4)$	28.3 ± 1.5

[a]kcal/mol per dimer.

[b]Calculated by using the last four carbons of each chain.

Figure Captions

Figure 1. The optimum *parallel* structure of decane sulfide dimer. Here the two alkyl tails are parallel but the CC planes of these tails orient perpendicularly with an average spearation of 4.4Å. The bottom view is rotated from the upper view by 90° about the horizontal axis.

Figure 2. Stable packings of infinite alkyl chains. The arrows represent the CC planes of the alkyl tails and the direction of the arrow is the phase of the tail. The structures in (a) have all planes parallel and are labeled L. The structures in (b) have half the CC planes perpendicular to the other half and are denoted as T. In these structures the setting angle (ϕ) of one CC plane is taken in the a direction. This setting angle (see Table 2) is optimized in each case to obtain the optimum structure. The structure T in (b) corresponds to the observed orthohombic structure of PE (observed $\phi = 41°$ for $T = 300$K, calculated $\phi = 42°$ at 0K).

Figure 3. The optimum structure of the T^* SAM. (b) shows the side view oriented to show the two dimers per unit cell with their CC planes parallel or perpendicular to the view. (a) is a top view but oriented *along the alkyl chains* (at a canting of 28.3° from the normal) to compare with the idealized structures in Figure 2.

Figure 4. The X-ray diffraction patterns. (a) shows the predicted diffraction pattern[20] for the optimized monolayer with the T^* structure, (b) averages the predicted pattern over the four quadrants of (a). (c) is the experimental pattern adapted from reference 14.