

**The Evaluation Of  
The ONIOM-EE Method For The  
QM/MM Hybrid Modeling of HF,  
CO and CO/HF Clusters**

by

Werner Crous

Thesis

submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

in

Chemistry

in the

Faculty of Science

at the

University of Stellenbosch

Supervisor: Prof. J.L.M. Dillen  
Co-supervisor: Dr. C. Esterhuysen

December 2006

## **Declaration**

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at an university for a degree.

Signature:.....

Date:.....

## Summary

Quantum mechanics is the method of choice when it comes to the accurate modeling of single molecules and clusters. The correlation energy is the single most important aspect when studying clusters computationally, and reproducing the correlation energy accurately poses a bigger challenge to the computational chemist than in the modeling of single molecules. Very high levels of theory and large basis sets need to be used.

Nevertheless, since the calculation of large systems, such as crystals and biological systems, is generally beyond the capacity of quantum mechanics, molecular mechanics is generally used for these systems. Unfortunately due to its nature, molecular mechanics cannot model important quantum effects, but this problem can be solved by a hybrid system in which one part of the system is treated by quantum mechanics and the remaining part by molecular mechanics.

In order to combine quantum mechanics with molecular mechanics one needs to optimize the parameters for the molecular mechanics part to allow it to function with the quantum mechanics. The research described in this work is based on the ONIOM-EE method, which is such a hybrid method.

In this work we investigate the applicability of the ONIOM-EE method in modeling hydrogen fluoride, carbon monoxide and CO/HF clusters. Most of the clusters' geometries in this work are not experimentally or computationally known. We therefore perform a computational analysis of all of the clusters by using various methods including Atoms in Molecules, Natural Bond Orbital analysis, Mulliken population analysis and the analysis of delocalized molecular orbitals to obtain information for the development of hybrid systems. During this process we look at different charge derivation schemes and at two different methods of optimizing force field parameters for these clusters. We develop a method to make force field optimization faster and better for specific hybrid systems. This method showed that in all cases the optimized parameters were an improvement on those of the Universal Force Field. We show the importance of an accurate description of the electrostatic interactions in HF, CO and CO/HF clusters and that this is the Achilles heel when attempting to optimize van der Waals parameters for force fields. We further show that atomic point charges are not a good approximation of a molecules' charge density in hybrid methods. In addition, we make suggestions on how the present method for ONIOM-EE can be improved to make the modeling of van der Waals clusters feasible.

## Opsomming

Kwantum meganika is die metode van keuse wanneer enkele molekule en molekulêre sisteme op rekenaar gemodeleer moet word. Dit is egter bekend dat die modelering van molekulêre sisteme 'n groter uitdaging stel aan die molekulêre modeleerder, aangesien baie hoë vlakke van teorie en groot basisstelle gebruik moet word om die korrelasie-energie, rekenkundig te produseer. Die akkurate herprodusering van die korrelasie-energie is seker die heel belangrikste vereiste waaraan voldoen moet word as molekulêre sisteme d.m.v. 'n rekenaar gemodeleer word.

Nietemin is dit onprakties om kwantum meganiese metodes te gebruik vir groot sisteme soos kristalle of biologiese molekule en juis om dié rede word molekulêre meganika meestal ingespan vir sulke gevalle. Molekulêre meganika is egter ondoeltreffend om belangrike kwantumeffekte te modeleer. Tog is daar 'n oplossing vir hierdie probleem in die vorm van 'n hibried sisteem waar een deel van die sisteem met kwantum meganika en die oorblywende deel van die sisteem met molekulêre meganika behandel word.

Om dit moontlik te maak om molekulêre meganika met kwantum meganika te kombineer, moet parameters vir die molekulêre meganika deel geoptimiseer word sodat dit saam met die kwantum meganiese deel kan funksioneer. Die navorsing wat in hierdie studie beskryf word is gebaseer op so 'n hibriedmetode wat bekend staan as ONIOM-EE.

In hierdie studie bestudeer ons die moontlikheid om ONIOM-EE te gebruik vir die modelering van molekulêre sisteme van waterstoffluoried, koolstofmonoksied en CO/HF sisteme. Die meeste van die sisteme, wat in hierdie studie behandel word, se strukture is onbekend, beide in terme van eksperimentele gegewens en molekulêre modelering. Ons voer dus 'n volledige analise van al die sisteme uit deur van verskeie metodes soos “Atoms in Molecules”, “Natural Bond Orbital” analise, Mulliken populasie analise en die analise van gedelokaliseerde molekulêre orbitale, gebruik te maak. Dit stel ons in staat om 'n hibriedsisteem te ontwikkel vir die molekulêre sisteme. Gedurende die proses ondersoek ons ook die gebruik van verskillende ladingsafleidings-sisteme en twee metodes word ondersoek waarop 'n kragveld vir 'n hibriedsisteem geoptimiseer kan word. Ons toon aan dat die geoptimiseerde parameters beter resultate lewer as die van die “Universal Force Field” en lig ook die belangrikheid daarvan uit dat die elektrostatiese interaksies se beskrywing 'n hibriedsisteem se Achilles hiel is indien van der Waals parameters geoptimiseer moet word. Ons toon aan dat die gebruik van puntladings op atome om die ladingsdigtheid in molekulêre sisteme te beskryf, 'n onakkurate benadering is. Sekere aanbevelings hoe om die ONIOM-EE metode sodanig te verbeter, dat dit wel gebruik kan word om van der Waals sisteme suksesvol te modeleer, word ook gemaak.

## Acknowledgements

- Professor J.L.M. Dillen, my supervisor, for allowing me to do this project. Thank you for pushing me beyond what I thought I could accomplish and challenging my every idea. Also thank you for allowing me to use a development version of the AIM program.
- Dr C. Esterhuysen, my co-supervisor, for her support and for believing in me. Thank you for being there and always listening to my complaints. Thank you for always making time to discuss my problems with me. I appreciate it.
- Professor J.L.M. Dillen and The National Research Foundation for their financial support.
- The Computational Chemistry Group at the University of Stellenbosch for their assistance and help, especially Gerhard Venter.
- The people who attempted to answer my questions about ONIOM, especially Jose Gascon from Yale University, Douglas J. Fox, Jim Hess and Dr Thom Vreven from Gaussian, Inc.
- My family and friends for their support.
- My mother for believing in me and who taught me never to give up, no matter what life throws at me. Thank you for supporting me right through this study.
- God, for giving me the strength to finish this project and for giving me the opportunity to broaden my horizons.

# Table of contents

Summary .....	i
Opsomming.....	i
Acknowledgements.....	iii
Table of contents.....	iv
Abbreviations and acronyms used in this work.....	x
Chapter 1:	
Introduction .....	1
1.1 Background .....	1
1.2 Aims .....	4
Chapter 2:	
Theoretical background on modeling methods for clusters .....	5
2.1 Introduction.....	5
2.2 Intermolecular forces .....	6
2.3 General introduction to computational methods for modeling clusters .....	8
2.4 Basis set problems .....	13
2.5 Post-HF methods for approximately solving the Schrödinger equation .....	15
2.6 Methods specifically pertaining to van der Waals clusters and intermolecular interactions .....	28
2.7 Molecular mechanics (MM) .....	31
2.8 Ways to represent a charge-distribution.....	34
2.9 QM/MM and the ONIOM hybrid methods.....	35
2.10 Force field parameterization for QM/MM.....	48
2.11 Accounts from the literature for the optimization of Lennard-Jones parameters for QM/MM .....	51
2.12 Summary and conclusions.....	52

Chapter 3:

Preparations for the molecular modeling of clusters for *ab initio* calculations and

ONIOM .....	54
3.1 Introduction.....	54
3.2 Software used in this work .....	54
3.3 Choice of method .....	56
3.4 Choice of basis set.....	57
3.5 General design of the QM/MM hybrid systems .....	62
3.6 Conclusion and summary.....	63

Chapter 4:

<i>Ab initio</i> calculations and analysis of hydrogen fluoride clusters .....	64
4.1 Introduction.....	64
4.2. Obtaining local minima on the Potential Energy Surface (PES) .....	64
4.3 Structural properties .....	66
4.4 Electronic interaction and binding energies .....	72
4.5 Analysis of bonding .....	77
4.6 Analysis of the electrostatics of the HF clusters .....	89
4.7 Summary, conclusions and future work .....	93

Chapter 5:

Optimization of the van der Waals parameters of the Universal Force Field (UFF) for

the use in ONIOM-EE optimizations of hydrogen fluoride clusters .....	95
5.1 Introduction.....	95
5.2 Computational details .....	95
5.3 Strategy to limit trials in optimization of force field (FF) –parameters .....	96
5.4 Criteria for assessing the quality of the force field (FF) .....	97
5.5 Optimization of FF-parameters .....	97
5.6 Optimizing the force field based on MP2 interaction energies .....	101
5.7 Summary and conclusions.....	105

Chapter 6:

<i>Ab initio</i> calculations and analysis of carbon monoxide clusters .....	107
6.1 Introduction.....	107
6.2 Computational details .....	108
6.3 Structural properties .....	110
6.4 Electronic binding and interaction energies of clusters .....	116
6.5 The importance of correcting for BSSE.....	118
6.6 Analysis of bonding .....	119
6.7 Investigating the electrostatics of the CO clusters .....	126
6.8 Oxocarbons vs CO clusters .....	130
6.9 Conclusions and future work.....	131

Chapter 7:

Optimization of the van der Waals parameters of the Universal Force Field (UFF) for use in ONIOM–EE optimizations of carbon monoxide clusters .....	132
7.1 Introduction.....	132
7.2 Computational details .....	132
7.3 Quality assessment of the force field.....	133
7.4 Optimization of the force field parameters .....	133
7.5 The applicability of using atom–atom contacts for quality assessment .....	136
7.6 Force field optimization based on the ONIOM steric energies .....	137
7.7 Summary, conclusions and future work.....	141

Chapter 8:

CO/HF clusters: <i>ab initio</i> calculations and force field optimization for hybrid optimizations .....	143
8.1 Introduction.....	143
8.2 Computational details .....	143
8.3 Structural properties .....	144
8.4 Analysis of electronic interaction and binding energies .....	148



8.5 Analysis of bonding .....	148
8.6 Atoms in molecules (AIM) analysis.....	151
8.7 Analysis of electrostatics.....	154
8.8 CO/HF force field optimization .....	154
8.9 Transferability of force field parameters and optimization for CO/HF clusters .....	157
8.10 Summary, conclusions and future work .....	159
Chapter 9:	
The ONIOM-EE method under the looking glass.....	160
9.1 Introduction.....	160
9.2 The electronic embedding procedure .....	161
9.3 The need for a better charge density description.....	163
9.4 Problem with the ESP derived charges .....	167
9.5 ONIOM-EE optimizations with rigid blocks .....	168
9.6 QM-QM interactions vs QM-MM interactions.....	169
9.7 A possible solution for problems encountered with ONIOM-EE.....	170
9.8 Summary, conclusions and future work .....	173
Chapter 10:	
Final conclusions and a summary of possible future work .....	174
10.1 Introduction.....	174
10.2 Main conclusions and brief summary .....	174
10.3 Future work .....	177
References.....	179
Appendix A.1 .....	186
Appendix A.2 .....	187
Appendix A.3 .....	187
Appendix A.4 .....	189

Appendix A.5 .....	190
Appendix B.1 .....	192
Appendix B.2.....	195
Appendix B.3.....	198
Appendix B.4.....	198
Appendix C .....	200
Appendix D .....	202

*As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.*

-ALBERT EINSTEIN (1879-1955)

# **Abbreviations and acronyms used in this work**

ADMP: Atom Centered Density Matrix Propagation molecular dynamics

AIM: Atoms in Molecules

ASEP: Average Solvent Electrostatic Potential

a.u. : atomic units

BSSE: Basis set superposition error

BCP: Bond critical point

CC: Coupled Cluster

CCP: Cage critical point

CCSD(T): Coupled cluster with singles and doubles; triples are estimated

CCSD: Coupled cluster with singles and doubles

CG: Conjugate gradient

CI: Configuration interaction

CP: Counterpoise

CPHF: Coupled perturbed Hartree-Fock

CPKS: Coupled perturbed Kohn-Sham

DFT: Density functional theory

DIM: Diatomics-in-molecules

DMA: Distributed multipole analysis

DQMC: Diffusion quantum Monte Carlo

DRF: Discrete Reaction Field

DZ: Double zeta

ESP: Electrostatic potential

FF: Force field

FMM: Fast multipole methods

GA: Genetic algorithm

GDIIS: Geometry directed in the iterative subspace

GGA: Generalized Gradient Approximation

GHO: Generalized hybrid orbital

GTO's: Gaussian primitives or Gaussian type orbital

HF: Hartree-Fock

IMOMM: Integrated molecular orbital molecular mechanics

IMOMO: Integrated molecular orbital molecular orbital

KS: Kohn-Sham  
LDA: Local density approximation  
MBAC: Many-body interaction analysis of clusters  
MBPT: Many-body or supermolecular perturbation theory  
MC: Monte Carlo  
MD: Molecular dynamics  
MKS: Merz-Kollman-Singh  
MM: Molecular mechanics or Molecular mechanics system  
MP: Møller-Plesset  
MP2: Møller-Plesset perturbation theory with corrections to the second order  
MP4: Møller-Plesset perturbation theory with corrections to the fourth order  
MP5: Møller-Plesset perturbation theory with corrections to the fifth order  
NAO: Natural atomic orbital  
NBO: Natural bond orbital  
ONIOM: Our own N-layered intergrated molecular orbital molecular mechanics  
ONIOM-EE: ONIOM with electronic embedding  
ONIOM-PCM: ONIOM with polarizable continuum model  
ONIOM-XS: ONIOM with exchange of solvents  
PES: Potential energy surface  
PBC: Periodic boundary conditions  
QM/MM: Quantum mechanics/Molecular mechanics hybrid method  
QM/SE: QM with semi-empirical method as MM  
QM: Quantum mechanics or Quantum mechanical system  
QMC: Quantum Monte Carlo  
RCP: Ring critical point  
RMSD: Root mean square deviation  
RS: Raleigh-Schrödinger  
SAPT: Symmetry adapted perturbation theory  
SCF: Self-consistent field  
STO's: Slater type orbitals  
TZ: Triple zeta  
UEG: Uniform electron gas  
UFF: Universal Force Field  
vdW: van der Waals

# Chapter 1

## Introduction

### *1.1 Background*

In 1978 the Nobel Prize winner Jean-Marie Lehn said, “*Just as there is a field of molecular chemistry based on the covalent bond, there is a field of supramolecular chemistry, the chemistry of molecular assemblies and of the intermolecular bond*” [Lehn, 1978]. The previous era of theoretical chemistry focused on the interactions between atoms and the formation of covalent bonds. The field Lehn pioneered goes beyond the molecule and emphasizes the need for a theoretical understanding of nonbonded interactions. An in-depth understanding of nonbonded interactions will lead to ultimate control at a molecular level, which will give us the ability to design molecular structures that can fulfill specific purposes.

When it comes to designing molecular structures for specific functions, Nature is a few steps ahead of us. Enzymes and proteins are realities where Nature utilizes nonbonded interactions to create fascinating “molecular machines”.

To understand and embrace this “new” era, we need sufficient models to explain and predict these nonbonded interactions. Progress in this field is unfortunately hampered by our limited computer technology, and the fact that most tools in quantum chemistry were originally designed for single molecules rather than systems of molecules.

There are many models to explain nonbonded interactions, especially models for hydrogen bonds, but every now and then, these models need to be modified as new experimental discoveries add to our understanding of intermolecular interactions. It is therefore understandable that new models need to be developed to comprehend the world beyond the molecule in order to lead us eventually to a unified theory of all nonbonded interactions. Computational chemistry can make a significant contribution in this field.

To model large molecular systems such as crystals, clusters and bio-molecules, theoretical chemists turn towards Newtonian physics for answers. They fit parameters to empirical equations in order to reproduce experimentally obtained data. This method of modeling large systems is called molecular mechanics (MM). This approach is reasonably accurate, but still, even if the fit to the experimental data is reasonable, quantum effects are not properly accounted for. Bond breaking and formation of bonds cannot be modeled by Newtonian physics; it is purely quantum mechanical in nature. On the one hand we can model small molecules, say up to a 100 atoms, reasonably accurately with quantum mechanics, but for

larger systems we have to turn to molecular mechanics. It seems as if we have reached a dead end. However, when it is known beforehand that a specific part of a system will form or partake in the breaking of bonds or exhibit other quantum effects, there is a partial solution to the problem – Quantum Mechanics/Molecular Mechanics (QM/MM).

In such a hybrid system, one part of the system is treated with a quantum mechanical method and the remaining part with molecular mechanics. In an enzyme, for example, a substrate and active site may be treated by quantum mechanics whereas the remaining part of the enzyme, by molecular mechanics. The only problem, of course, is combining these two totally dissimilar methods so that the lack of accuracy in the one does not influence the accuracy of the other too significantly. It can be compared to grafting a bamboo to a tree.

In 1976 [Warshel and Levitt, 1976], the first QM/MM system was constructed. It was remarkably good considering the technology at that point in time. We have come a long way since then, and according to a simple analysis by SciFinder Scholar 2006 of the results for the number of QM/MM method publications per year, last year (anno 2005) 312 articles were published with respect to 99 in 2000.

To enable the system treated with quantum mechanics (QM) to interact with the MM system, four factors should be considered:

1. The MM system should be able to polarize the QM system accurately;
2. The QM system should in return be able to polarize the MM system accurately;
3. The empirical equations and parameters for the MM system should be reparameterized for the specific QM system;
4. The boundary problem, where a plane between the QM and MM systems cuts a covalent bond, needs to be treated accurately so that overpolarization of the QM system does not take place.

Standard force fields are not optimized for QM/MM systems and therefore they always need to be optimized, therefore point 3 is crucial in all cases where the accurate use of QM/MM is required. Point 1 is default with most QM/MM systems except for ONIOM, *vide infra* and point 2 above has only been attempted by a limited number of researchers [Bakowies and Thiel, 1996; Jensen *et al.*, 2003, Kongsted *et al.*, 2003]. Most models for hybrid systems currently ignore this point as it slows down geometry optimizations tremendously.

In 1996, Morokuma and coworkers [Svensson *et al.*, 1996] developed a new way of doing a QM/MM calculation. The methodology they used to do QM/MM is part of their general methodology called ONIOM that stands for “*Our own N-layered Integrated molecular Orbital molecular Mechanics*” method. ONIOM is more general than QM/MM as it can easily be expanded to include more than two layers leading to variations such as

QM/QM/MM systems, which is very difficult or impossible with the QM/MM methodology. The problem with ONIOM, in terms of QM/MM systems, is that the QM system is not polarized as is done by default in most other QM/MM calculations. Although this might seem a problem at first, this approach makes geometry optimizations simpler as the QM and MM systems can be treated separately and stationary point charges can be used for the description of the interaction of two or more charge distributions. However, although this method is simpler it is not the preferred way to do QM/MM calculations. Therefore, in actual fact the original ONIOM method did not contribute much to the execution of QM/MM calculations in general, but contributed more to the extension of QM/MM systems to QM/QM and QM/QM/MM systems that were previously very difficult or impossible with the original QM/MM methodology.

In 2003, ONIOM-EE (ONIOM with electronic embedding) was developed whereby the QM system can now be polarized by point charges on the molecules in the MM system. In order to make this work, new optimization algorithms were also developed. One can predict that with the advent of ONIOM-EE methodology, this methodology will pave the way in developing better QM/MM methods rather than the original QM/MM methodology. Developments in QM/MM methodology can also be used in the ONIOM-EE methodology and vice versa so one expects to see synergy between the two methodologies.

Terminology might become confusing, so we emphasize that in this work we will use the ONIOM-EE methodology to do QM/MM calculations. ONIOM-EE is simply another way of doing a QM/MM calculation, but it can also be used for other variations as mentioned above for which the original QM/MM methodology cannot be used.

As mentioned already, in order to use QM together with MM, one should always optimize the force field for a specific QM system. To the best of our knowledge, a classical force field has never been optimized to use together with a correlated QM method to model HF, CO and CO/HF gas phase van der Waals clusters with QM/MM. Furthermore, the ONIOM-EE methodology has never been used for such a system. Combining QM with MM to model systems so challenging as van der Waals clusters will be insightful into what is needed to make the modeling of QM/MM systems as accurate as the modeling of pure QM systems. What would make such a study further challenging is the fact that very little is known, both experimentally and computationally, about CO and CO/HF clusters and that computationally it is very difficult to model these clusters accurately with standard methods. HF clusters are better known and simpler to model.



## ***1.2 Aims***

In light of the preceding background information the aims of this work are the following:

1. To have a sound theoretical understanding of all the methods used, especially ONIOM;
2. To generate *ab initio* data and geometries of reliable accuracy for CO, HF and CO/HF gas phase clusters;
3. To analyze and characterize these clusters in order to add to the understanding of the bonding in these clusters, which will hopefully aid in developing better hybrid methods;
4. To derive and validate atomic point charges for all the clusters to be used in QM/MM calculations;
5. To optimize force field parameters for the clusters to give quantitative data and good geometries when a force field is used for the molecular mechanics part in a QM/MM method;
6. To suggest improvements to the present ONIOM-EE methodology as applied to QM/MM systems, in order to make the modeling of clusters with this method more accurate in the future.

## Chapter 2

# Theoretical background on modeling methods for clusters

### 2.1 Introduction

As mentioned in **Chapter 1**, clusters pose a bigger challenge to the computational chemist than the modeling of single molecules. The phenomenon of electron correlation plays an important role in vdW forces and therefore simple theories such as Hartree-Fock (HF) are inadequate for the modeling of clusters.

We start this chapter with **Section 2.2** by introducing the current theory of intermolecular forces. In **Section 2.3**, we will give a brief introduction to computational chemistry and in **Section 2.4** we will discuss problems encountered when studying intermolecular forces utilizing limited basis sets. In **Section 2.5**, we will give an introduction to post-HF methods used in studying intermolecular interactions, while in **Section 2.6** two methods to calculate the interaction energy of a cluster will be discussed. This will conclude our review for the modeling of vdW clusters. Building up to our discussion of hybrid methods, this will be followed by a general introduction to molecular mechanics (MM) in **Section 2.7** and a discussion of a variety of schemes to derive charges from quantum mechanical data in **Section 2.8**. In **Section 2.9**, hybrid methods will be discussed with a special focus on information pertaining to our study. In order to graft the proverbial bamboo to the tree, it is necessary, as mentioned in **Chapter 1**, to optimize the parameters for force fields when using them in the context of hybrid systems. This will be discussed in **Section 2.10**. **Section 2.11** will give a brief overview of the literature pertaining to the optimization of Lennard-Jones parameters for hybrid methods. The chapter will be concluded with **Section 2.12** in which a summary of the chapter will be given.

Before we start, it is important to define some terminology since there is not always consensus about the meaning of some terms in the literature. In this work van der Waals (vdW) forces are regarded as the dispersion and exchange forces, whereas the term *nonbonded interactions* or *intermolecular forces* also includes the electrostatic and induction interactions. The term, “a vdW cluster”, includes both hydrogen bonded and other nonbonded clusters.

## 2.2 Intermolecular forces

### 2.2.1 The four fundamental intermolecular forces

When scientists attempt to discuss intermolecular forces, they are keen to abandon the difficult theory of quantum mechanics for classical models. This is understandable, as classical models have proved good in explaining many phenomena in the universe, such as gravity. However, sometimes things are different to what they seem. Taking gravity as an example, Einstein showed that gravity is a result of the bending of space-time rather than a classical force acting over a distance.

Common methods used to discuss intermolecular forces, use classical electrostatics to explain electrostatic phenomena and augment this further by accounting for quantum phenomena such as exchange and dispersion interactions, *vide infra*.

Intermolecular interactions are generally accepted to consist of the following individual interactions [Havenith, 2001]:

► **The electrostatic interaction:**

This interaction between monomers is caused by the interaction of the electron charge densities of permanent dipole or multipole moments. The interaction can be attractive or repulsive depending on the signs of the charges.

► **The induction interaction:**

If a permanent dipole or multipole moment on one monomer induces a dipole or multipole moment on another monomer in the cluster, it is called induction [Chęłasiński and Szcześniak, 1994]. Induction is an attractive interaction and it is represented by the polarizability, a second-order electric quantity. The polarizability is defined in terms of a transition moment from one quantum state to a new quantum state and an excitation energy [Magnasco, 2004].

► **The dispersion interaction:**

The dispersion energy has no classical analogue and is purely a quantum mechanical phenomenon due to long-range electron-electron correlation. If two molecules approach each other, instantaneous dipoles or multipoles can arise in both. This is called the dispersion interaction. It is not directed, such as is the case with the induction interaction, but if a multipole arises in one molecule the multipole in the other molecule will be directed towards the other molecule's multipole. The dispersion interaction is weakly attractive.

► **The exchange interaction:**

To counteract the dispersion interaction, there is the exchange interaction. The exchange interaction is a result of a quantum effect. Because the Pauli-exclusion principle must be obeyed, electrons with the same spin will repel each other and electrons with opposite spin will attract each other. The exchange interaction arises due to the repulsion of electrons with the same spin.

According to Magnasco [Magnasco, 2004], an intermolecular bond is formed when the small Pauli-repulsion, decreasing exponentially with the intermolecular distance, is offset by attractive interactions such as distortion interactions related to the electric properties of interacting molecules. These electric properties can include permanent or induced electric moments and polarizabilities.

### 2.2.2 *Many-body interactions*

In clusters, many-body interactions play a fundamental role in stabilization. Many-body interactions are interactions that are caused by the combination of the nonbonded interactions between molecules. For example, many-body interactions lead to a term called the dispersion-exchange interaction in weakly bonded trimers. Understanding bonding in large clusters becomes more and more difficult due to the many-body interactions.

According to Chałasiński and Szcześniak [Chałasiński and Szcześniak, 2000], the binding energy of a cluster can be written as:

$$\Delta E(\text{binding}) = \Delta E(1\text{-body}) + \Delta E(2\text{-body}) + \Delta E(3\text{-body}) + \dots + \Delta E(N\text{-body}) \quad (2.1)$$

The interaction energy is defined as the sum of the above terms while omitting the 1-body energy. The 1-body energy is equal to the distortion of the monomers' geometries from their gas phase geometries. Some authors call the binding energy the interaction energy, but it has been argued [Chałasiński and Szcześniak, 2000] that only the terms from the 2-body energy onwards should be seen as the interaction energy. In this work we call  $\Delta E(\text{binding})$  the binding energy and  $\Delta E(\text{binding})$  without  $\Delta E(1\text{-body})$ , the interaction energy.

## 2.3 General introduction to computational methods for modeling clusters

### 2.3.1 Ab initio methods

*Ab initio* methods attempt to approximate the exact solution of the many-electron Schrödinger equation for the ground state energy of a molecule or molecular system by deriving results from first principles without the use of explicit experimental data. However, most *ab initio* methods used, can only give approximate results that are not necessarily comparable to the results obtained by experiment. It is therefore important to understand these methods on a theoretical basis in order to make intelligent choices for modeling the system under study. Some methods are known to be better than others, but still the final litmus test is the reproduction of experimental results and trends.

### 2.3.2 The Schrödinger equation

The basis for quantum chemistry is the Schrödinger equation. Although simpler and more applicable to computational chemistry, this equation does not account for relativistic effects such as the Dirac equation does, but can nevertheless be used for the lighter atoms up to atomic number 36 on the periodic table<sup>1</sup>. It also does not recognize electron spin directly, but the Pauli-exclusion principle, which states that the wave function should be anti-symmetric with an exchange of two electrons, provides for the incorporation of spin in an indirect manner. This simplifies the solution of the Schrödinger equation, as the Hamiltonian does not have an effect on the spin and spin can therefore be removed from the equation and treated separately as an *ad hoc* function. The Schrödinger equation can be written as in equation 2.2,

$$\begin{aligned}\hat{H}\psi(\mathbf{R},\mathbf{r},\sigma) &= E\psi(\mathbf{R},\mathbf{r},\sigma) \\ s(\sigma)\hat{H}\phi(\mathbf{R},\mathbf{r}) &= s(\sigma)E\phi(\mathbf{R},\mathbf{r})\end{aligned}\quad (2.2)$$

where  $\hat{H}$  is the Hamiltonian operator and  $E$  the energy of the wave function.  $\psi(\mathbf{R},\mathbf{r},\sigma)$  denotes a spin orbital. A spatial orbital is denoted by  $\phi(\mathbf{R},\mathbf{r})$  where  $\mathbf{r}$  are the coordinates of an electron and  $\mathbf{R}$  the coordinates of a nucleus.  $s(\sigma)$  is a spin function with  $\sigma$  either equal to  $\alpha$  or  $\beta$  spin functions or linear combinations of them. In this study, we will always make use of spin orbitals.

---

<sup>1</sup>Relativistic effects for heavier atoms can be incorporated in an *ad hoc* manner by making use of effective core potentials (ECP's)

Although reasonably simple, the Schrödinger equation cannot be solved exactly for molecular systems larger than the hydrogen atom or  $\text{H}_2^+$ . One can however separate the movement of an electron from a nucleus if one assumes that nuclei, because of their heavier weight, will not move as much as electrons for a specific nuclear configuration. This is called the Born-Oppenheimer approximation. This approximation makes solving the Schrödinger equation for many-electron systems simpler, although an exact solution is still impossible due to the effect of electron correlation.

### 2.3.3 Hartree-Fock theory

Hartree-Fock (HF) theory is the simplest *ab initio* method and totally unsuitable for the accurate modeling of clusters; however for understanding the origin of the other methods used to model vdW clusters, a brief introduction is in order.

Hartree-Fock theory assumes that each electron in a many-electron atom or molecule only interacts with an average potential caused by all the other electrons in the atom or molecule. The direct interaction of one electron with another is therefore not accounted for. The interaction of electrons with each other is known as electron correlation. It has been shown that electrons are usually further apart in atoms and in molecules than predicted by the HF theory [Jensen, 2001]. When studying weak interactions such as intermolecular interactions, electron correlation needs to be accounted for as accurately as possible.

The HF theory approximates the true wave function of an atom or molecule with a single Slater determinant. Although a Slater determinant has the necessary anti-symmetry with respect to row and column exchange, it gives a quadratic form (see Fig. 2.1) of the wave function, which is not the true form for a many-electron atom or molecule.

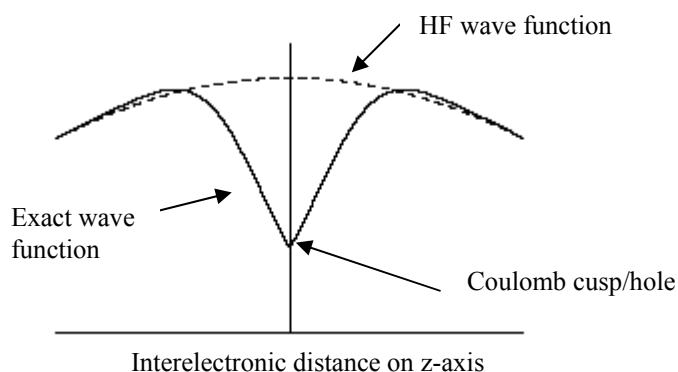


Fig. 2.1: Diagram illustrating the difference between the HF wave function's behavior with respect to the interelectronic distance, compared to the exact wave function's behavior. In this figure both electrons have the same x and y coordinates, only their z coordinates vary.

Although the HF wave function is a poor approximation of the true wave function, this difference is actually quite small and therefore HF theory is reasonably accurate for single molecules. However, quantitative work will require that the Coulomb cusp is accounted for more precisely [Knowles *et al.*, 2000]. This can be done by using many Slater determinants in the expansion of the true wave function, which is the essence of post-HF methods. Another problem with the true wave function is that it is discontinuous when the interelectronic distance is zero. Other than the singularity in the wave function, the Schrödinger equation in 2.2 does not contain any singularities on the right hand side. Therefore, the only singularity on the left hand side should be a singularity in the wave function. However, the left hand side contains another singularity, namely  $1/r_{12}$  where  $r_{12}$  is the interelectronic distance. Therefore, this singularity should be cancelled on the left hand side by another singularity to make both sides of 2.2 equal. This singularity can be shown to originate from the kinetic energy. This simply means that as electrons come very close to each other, their kinetic energy will increase to infinity leading to a fast movement of one electron away from another [Knowles *et al.*, 2000]. A method able to generate a wave function in the limit of a complete basis set that is linear when moving in any direction from  $r_{12}=0$ , will solve the Schrödinger equation exactly.

#### 2.3.4 Generating approximate wave functions

As the true wave function of a poly-electronic molecule or atom is not known, these wave functions should be approximated with linear combinations of other functions. Usually the wave functions are expanded in terms of one-electron basis functions. By expanding the true wave function, any arbitrary basis can be chosen and the wave function transformed to that

basis. Common basis functions that are used are Gaussian Type Orbitals (GTO's) and Slater Type Orbitals (STO's). A GTO can be written as follows,

$$\chi_{GTO} = Ne^{-\alpha r^2} x^l y^m z^n \quad (2.3)$$

where  $N$  is the normalization constant,  $\alpha$  is a constant known as the exponent and  $x$ ,  $y$  and  $z$  are Cartesian coordinates.  $l$ ,  $m$  and  $n$  are integral exponents of the Cartesian coordinates. The sum of  $l$ ,  $m$  and  $n$  is used analogously to the angular momentum quantum number for atoms. Therefore, if the sum is zero, it denotes an  $s$ -orbital. If the sum is 1, then it denotes a  $p$ -orbital etc. GTO's account both for the radial and angular components of a wave function. Another characteristic of GTO's is that they are all real functions whereas a wave function can be complex as well. However, taking linear combinations of complex functions can give real functions that are also solutions to the Schrödinger equation.

A single GTO cannot account for the increase in nodes in higher angular momentum wave functions or the correct form of an orbital. However, linear combinations of these GTO's can. A basis function, created by combining a series of GTO's in a linear combination, is known as a contraction. In most software, using GTO's as a basis for the expansion of molecular orbitals, all the exponents and coefficients defined for each GTO, also called a primitive, have been pre-optimized and cannot change during a calculation. Only the normalization constants in front of basis functions are allowed to change. Sometimes single GTO's are also used as basis functions, but this increases the number of constants that need to be determined.

Some basis sets are specifically optimized to converge quickly to the correlated wave function with addition of basis functions, such as the correlation consistent basis sets of Dunning [Woon and Dunning Jr., 1993; Kendall *et al.*, 1992; Dunning Jr., 1989; Peterson *et al.*, 1994; Wilson *et al.*, 1997]. The split valence (SV) sets of Pople and coworkers [Hehre *et al.*, 1972; Krishnan *et al.*, 1980; Frisch *et al.*, 1984] use a contraction of different numbers of primitives for core orbitals than for valence orbitals. For example, in the 6-31G basis set, 6 primitives are used for the core orbitals and a contraction of 3 primitives and another single primitive are used for the valence orbitals. To illustrate the notation for the Pople-type SV basis sets, we will give an example. In this work we used a basis set denoted by 6-311+G(d,p). This means that 6 GTO's are used to describe the core orbitals of the atom, while a contraction of 3 GTO's and two separate GTO's are used for the valence orbitals. The addition sign indicates that each orbital is augmented with a diffuse function of the same angular momentum. For example, a  $p$ -function, will be augmented with a  $p$  diffuse function. The terms in the parenthesis state that  $d$ -type polarization functions that are equal to 5 uncontracted  $d$ -type GTO's, must be added for the heavy atoms and  $p$ -type polarization functions that are equal to 3 uncontracted  $p$ -type GTO's, must be added to the hydrogen



atoms after the addition of the diffuse functions. In the notation used above, a comma always separates the polarization functions for the heavy atoms from the hydrogen atoms.

When modeling vdW clusters, it is important to include both polarization functions and diffuse functions [Chalański and Szcześniak, 1994]. Polarization functions are functions of higher angular momentum that can be added to the current basis set, to gear it towards the recovery of correlation energy. They cannot be derived from HF calculations. Diffuse functions are functions with small exponents and therefore decay slowly with the distance from the nucleus. Table 2.1 compares the two basis sets that were mainly used in this study for the carbon monoxide monomer.

Table 2.1: Information for the carbon monoxide monomer for the two basis sets used in this study. The contents of the parenthesis are read as follows: (*s*-type GTO's, *p*-type GTO's, *d*-type GTO's). For each *p*-type, 3 basis functions are added and for each *d*-type, 6 basis functions in the case of 6-31G(d) and 5 basis functions in the case of 6-311+G(d) are added. The numbers in the parenthesis are the number of primitives used in each contraction.

	Carbon	Oxygen
Atomic orbitals	1s, 2s, 2p	1s, 2s, 2p
6-31G(d)		
Approximation by GTO's	(631,31,1)	(631,31,1)
Total basis functions	15	15
Total GTOs	28	28
6-311+G(d)		
Approximation by GTO's	(63111,3111,1)	(63111,3111,1)
Total basis functions	22	22
Total GTOs	35	35

Basis sets can also be optimized based on the type of calculation to be performed. For example a basic basis set such as STO-3G will not be suitable for correlated calculations, due to the lack of polarization basis functions. In general, a good basis set should have the following features [Dunning Jr. *et al.*, 1998]:

1. It should be geared towards the recovering of the correlation energy.
2. It should converge quickly with the addition of polarization functions (*vide infra*).
3. It should satisfy the demands of all four fundamental interaction energy components: the exchange, dispersion, induction and electrostatic interactions [Chalański and Szcześniak, 1994].
4. It should have a small *basis set superposition error* (BSSE) when used for the modeling of vdW clusters [Chalański and Szcześniak, 1994].
5. The basis set must be as compact as possible.
6. The basis set must cover both the angular and radial spaces of the wave function in a consistent manner.

## 2.4 Basis set problems

The problem with basis sets is that they are centered on atoms. This means that intermolecular interactions involving larger distances than bond distances are not effectively reproduced, or, in other words, the convergence towards the true wave function is very slow with addition of basis functions. One way to remedy this is to make use of basis functions centered between molecules in addition to the basis functions centered on the atoms. This has been shown to give a faster convergence [Chalański and Szcześniak, 2000]. Another problem with basis sets is the so-called BSSE and basis set saturation error. These errors will now be briefly discussed.

### 2.4.1 Basis Set Superposition Error (BSSE)

As mentioned above, basis functions are usually centered on atoms. Unfortunately, if the basis sets used for the monomers are too small, errors can arise in the calculation of the interaction energy of a cluster. In order to minimize the energy, basis functions on one atom in a monomer can form linear combinations with basis functions on one atom in another monomer, resulting in the use of more primitives for the description of the wave function between the two monomers than should actually be the case. This gives rise to more stabilization and artificially lowers the interaction energy.

Simply stated, the BSSE is caused when a larger basis set is used to describe a system as a whole, than the basis set used to describe each monomer of the system individually. If a basis function on one atom is orthogonal to another basis function on another atom, which is not necessarily the case, they would not be able to “mix” and form direct products of nonzero value.

To correct for this error, a counterpoise (CP) correction can be attempted [Jensen, 2001]. In this correction, all the monomers’ energies are calculated with the basis set of the whole cluster. The sum of these energies is then subtracted from the energy calculated with the basis set of the whole cluster. This is then supposed to give an approximation of the error. The value so obtained is an upper limit of the true BSSE.

According to Chalański and Szcześniak [Chalański and Szcześniak, 1994], it is important when using the CP correction, that the magnitudes of the BSSE should not be used to judge the quality of the interaction energy. The BSSE calculated with the CP method is not related to the error in the interaction energy, but is purely an effect of a too small basis set for the problem at hand.

#### 2.4.2 Basis set saturation error

The interelectronic Coulomb cusp condition, mentioned in **Section 2.2.3**, is very slowly reproduced by an expansion of one-electron basis functions [Chalański and Szcześniak, 2000]. This effect is very serious for the dispersion interactions and demands high polarization functions. Even large correlation consistent basis sets such as d-aug-cc-V6Z<sup>2</sup> are not good enough. As mentioned earlier, one can remedy this problem by using bond functions in the middle of the vdW bond. A bond function is a basis function that is centered on a vdW bond rather than on an atom. One can also make the basis functions dependent on the interelectronic distance by explicitly including terms that are linear in the interelectronic distance. Methods incorporating the interelectronic distance directly such as MP2-R12 also exist [Chalański and Szcześniak, 2000].

#### 2.4.3 Obtaining the approximate equations of the molecular orbitals

In HF theory, molecular orbitals (MO's) are approximated by canonical orbitals. These canonical orbitals are pseudo-eigenfunctions of the Fock-operator. When the Fock-operator is applied to a canonical orbital, the corresponding eigenvalue is the molecular orbital energy of the corresponding canonical orbital. The energies of the canonical orbitals are minimized using undetermined Lagrange multipliers and chosen to be orthogonal and normalized. The optimization of these orbital functions is complete when the change in the energy is as close as possible or equal to zero.

HF theory works by diagonalization<sup>3</sup> of the Fock-matrix, containing Fock-operators and wave functions for each electron, to give a set of energies and molecular orbital coefficients. The new coefficients are substituted again in the equation of the wave function on which the Fock-operator operates to give a new energy and molecular orbital coefficients. This is done iteratively until the energy does not change significantly any more. As the equations must be solved self-consistently, the method is also known as self-consistent field (SCF).

Each eigenvalue in the diagonal matrix is the energy of a canonical orbital. When using basis functions the method is somewhat modified in the form of Roothaan-Hall equations that also need the overlap matrix of the basis functions in order to compute the energy. We will not discuss this in more detail here, but the theory can be found in any good textbook on computational chemistry, such as the book by Jensen [Jensen, 2001]. The Fock-operator is an

---

<sup>2</sup> This basis set is doubly augmented with diffuse functions, has six basis functions for each electron's description and is considered as extremely large.

<sup>3</sup> This is done by a unitary transformation by which the Fock matrix is diagonalized by multiplying with a unitary coefficient matrix on one side of the operator and by its inverse on the other side.

effective one-electron operator and can be written for a closed shell system where electrons are all paired as:

$$\hat{F}_i = \hat{h}_i + \sum_j^{N/2} (2\hat{J}_j - \hat{K}_j) \quad (2.4)$$

where  $\hat{h}_i$  is a sum of the average kinetic energy of the electron and the Coulomb interaction between the electron and all nuclei.  $N$  is the number of electrons.  $\hat{J}_j$  and  $\hat{K}_j$  are the Coulomb and exchange operators respectively for electron  $j$ . The exchange operator has no classical analogue and arises out of the constraint of the Pauli-exclusion principle on the wave function.

In computational chemistry, the canonical orbitals are usually expanded in GTO's. In a closed shell system, spin is incorporated as an *ad hoc* function, but for open shell systems, where basis functions are used to describe each spin orbital separately, spin functions can make a large difference to the energy.

The eigenvalues and canonical orbitals are determined based on the variation theorem. According to the variation theorem, the eigenvalue so obtained will be larger or equal to the true eigenvalue of the Hamiltonian when operating on this wave function. Configuration Interaction (CI) is another method based on the variation theorem.

## ***2.5 Post-HF methods for approximately solving the Schrödinger equation***

As quantum mechanics is based on probabilities instead of certainties, one can also say that there will always be a certain non-zero probability that an electron will be excited to an orbital with a higher main quantum number than the original one or that two electrons will be excited to orbitals due to electron correlation. When one electron is excited, it is called a single excitation or a single, for two electrons it is known as a double and for three it is known as a triple and so on. All of these excitations can be represented by Slater determinants that differ from the ground state Slater determinant. The sum of these different Slater determinants will therefore account for all the probabilities and so will reproduce a wave function that, if optimized, that incorporates all electron correlation in the best possible way in the limit of the basis set. This is how full-CI is done. It is important to know that even full-CI, with a complete basis set, will not give the exact electronic energy for a system, as the general CI wave function is not linear in  $r_{12}$ , but it will give the best answer for a specific one-electron basis set that can be obtained [Knowles *et al.*, 2000]. As can be expected, it is impractical to do full-CI for chemically significant systems. Therefore, other approximate methods were

developed that can give good results by attempting to use the most “important” Slater determinants in the expansion of an approximate wave function.

Perturbation theory attempts to approximate the energy and wave function by using Taylor expansions around a point where the energy and wave function are known. Perturbation theory is much more suitable to use for the description of molecular clusters since the energy term of each intermolecular interaction can be expanded individually, giving rise to zero, first and second-order perturbations (see **Section 2.6**).

In perturbation theory, we deal with an infinite series. Depending on where the series is truncated, will determine the value for the energy obtained. If the series is truncated wrongly, it might lead to divergence of the series for a specific problem [Olsen *et al.*, 1996; Leininger *et al.*, 2000]. An infinite expansion will give the exact energy and wave function. General perturbation theory is also known as Rayleigh-Schrödinger (RS) perturbation theory.

### 2.5.1 Møller-Plesset perturbation theory

In this study we made extensive use of the Møller-Plesset perturbation theory (MP-theory) developed by Møller and Plesset in 1934 [Møller and Plesset, 1934]. They proposed a perturbation treatment of atoms and molecules using the zero-order HF wave function as the unperturbed wave function. In this section we will briefly derive the equations illustrating this theory, but before we continue it would be important to clarify the terminology that we will use. With the term, *HF wave function*, we explicitly mean any Slater determinant that can be an eigenfunction of the Fock-operator. When we talk about the *zero-order HF wave function* it specifically means the lowest energy Slater determinant that is an eigenfunction of the Fock-operator or the wave function obtained when performing a HF calculation.

To illustrate MP-theory, consider the Schrödinger equation in equation 2.5 once again. The true wave function can be expanded into HF wave functions,  $\psi_i$ , of different order. A  $\psi_i^{(1)}$  wave function denotes the first-order of a HF wave function for example. It is important to realise that these higher-order wave functions are all linear combinations of different Slater determinants generated from the zero-order HF wave function, by operators, *vide infra*. The energy can also be expanded into energy contributions of different order where  $E_i^{(0)}$  is the zero-order energy. This energy is not equal to the HF energy.

$$\begin{aligned}\hat{H}\psi &= E\psi \\ \psi &= \psi_i^{(0)} + \lambda\psi_i^{(1)} + \lambda^2\psi_i^{(2)} + \lambda^3\psi_i^{(3)} + \lambda^4\psi_i^{(4)} + \dots \\ E &= E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \lambda^3 E_i^{(3)} + \lambda^4 E_i^{(4)} + \dots\end{aligned}\tag{2.5}$$

The Hamiltonian can also be expanded in terms of the zero-order Hamiltonian,  $\hat{H}^{(0)}$ , and a perturbation Hamiltonian,  $\hat{H}'$ .

$$\hat{H} = \hat{H}^{(0)} + \lambda \hat{H}' \quad (2.6)$$

$\lambda$  is a coefficient in the linear expansions and is usually set to equal 1 for a full perturbation.

In MP theory certain assumptions are made. First the  $\hat{H}^{(0)}$  is set equal to the sum of Fock-operators which, if operating on the zero-order HF wave function, will give the energy eigenvalues of all the canonical orbital wave functions.  $\hat{H}'$  is set equal to the difference between the true Hamiltonian and  $\hat{H}^{(0)}$ . Therefore,

$$\begin{aligned} \hat{H}^{(0)} &= \sum_{m=1}^N \hat{F}_i(m) \\ \hat{H} - \hat{H}^{(0)} &= -\hat{V}_{ee} = -\frac{1}{r_{12}} \end{aligned} \quad (2.7)$$

where  $\hat{F}_i$  is the Fock-operator and  $m$  shows that it is the Fock-operator of electron  $m$ .  $N$  is the number of electrons.  $-\hat{V}_{ee}$  is the electron repulsion operator and incorporates both the Coulomb and exchange interactions between two electrons in an average fashion.  $r_{12}$  is the distance between two individual electrons.

Now we can substitute the assumptions in 2.5 and 2.6 into the left hand side of the Schrödinger equation to obtain:

$$\sum_{m=1}^N \hat{F}_i(m) \psi_i^{(HF)} + \lambda \left( \sum_{m=1}^N \hat{F}_i(m) \psi_i^{(1)} - \hat{V}_{ee} \psi_i^{(HF)} \right) + \lambda^2 \left( \sum_{m=1}^N \hat{F}_i(m) \psi_i^{(2)} - \hat{V}_{ee} \psi_i^{(1)} \right) + \dots \quad (2.8)$$

To make this equation simpler we explicitly showed that the zero-order wave function is the Hartree-Fock wave function,  $\psi_i^{(HF)}$ . To simplify the equation further and obtain the various corrections to the energy, we can multiply with the complex conjugate of  $\psi_i^{(HF)}$  throughout and integrate over all space. This is a trick as we actually multiply with the complex conjugate of a different HF wave function,  $\psi_j^{HF}$ , but only terms that have  $i=j$  will survive.

$$\left\langle \psi_i^{(HF)} \left| \left( \sum_{m=1}^N \hat{F}_i(m) - \lambda \hat{V}_{ee} \right) \right| \psi_i^{(HF)} \right\rangle + \lambda \left\langle \psi_i^{(HF)} \left| \left( \sum_{m=1}^N \hat{F}_i(m) - \lambda \hat{V}_{ee} \right) \right| \psi_i^{(1)} \right\rangle + \lambda^2 \left\langle \psi_i^{(HF)} \left| \left( \sum_{m=1}^N \hat{F}_i(m) - \lambda \hat{V}_{ee} \right) \right| \psi_i^{(2)} \right\rangle + \dots \quad (2.9)$$

We can simplify the equation above even further by using a method called intermediate normalization. Let  $\psi_i^{(HF)}$  be the unperturbed wave function and  $\psi$  be the solution to the many-electron Schrödinger equation. When using intermediate normalization, we assume that  $\langle \psi | \psi_i^{HF} \rangle = 1$ . All the other corrections to the zero-order HF wave function are therefore orthogonal to this wave function. It does not matter if our assumption is wrong initially, as the integral above can be normalized by adding a constant. A constant will not influence the

solution of the Schrödinger equation. This simplifies equation 2.9 tremendously as now some terms will vanish. For example the first, second and third matrix element<sup>4</sup> and so on can be simplified significantly. As an example, we will show this for the second matrix element in equation 2.9.

$$\begin{aligned}
 \langle \psi_i^{(HF)} | (\sum_{m=1}^N \hat{F}_i(m) - \lambda \hat{V}_{ee}) | \psi_i^{(1)} \rangle &= \langle \psi_i^{(HF)} | (\sum_{m=1}^N \hat{F}_i(m) | \psi_i^{(1)} \rangle - \lambda \langle \psi_i^{(HF)} | \hat{V}_{ee} | \psi_i^{(1)} \rangle) \\
 &= \langle \psi_i^{(1)} | (\sum_{m=1}^N \hat{F}_i(m) | \psi_i^{(HF)} \rangle)^* - \lambda \langle \psi_i^{(HF)} | \hat{V}_{ee} | \psi_i^{(1)} \rangle \quad (2.10) \\
 &= \varepsilon^* \langle \psi_i^{(1)} | \psi_i^{(HF)} \rangle^* - \lambda \langle \psi_i^{(HF)} | \hat{V}_{ee} | \psi_i^{(1)} \rangle = 0 - \lambda \langle \psi_i^{(HF)} | \hat{V}_{ee} | \psi_i^{(1)} \rangle
 \end{aligned}$$

The first matrix element in this expansion is zero due to intermediate normalization as both  $\psi_i^{(1)}$  and  $\psi_i^{(HF)}$  are eigenfunctions of the Fock-operator and hence orthogonal to each other.

Using these results, we can now simplify the equation in 2.9 to give:

$$E_i^{(HF)} - \langle \psi_i^{(HF)} | \hat{V}_{ee} | \psi_i^{(1)} \rangle - \langle \psi_i^{(HF)} | \hat{V}_{ee} | \psi_i^{(2)} \rangle + \dots \quad (2.11)$$

Here we used the fact that the first matrix element in 2.8 is equal to the HF energy. The first matrix element in equation 2.11 is the matrix element for the second-order correction to the energy. The second matrix element is the third-order correction to the energy etc. Correcting the HF energy with a second-order correction is called MP2, which was the method predominantly used in this work. A correction to the third-order is called MP3, a fourth-order correction, MP4 and so on. For the rest of this derivation we will focus on the second-order correction to the HF energy. In this matrix element, there is one unknown wave function. This wave function can be expanded in terms of the zero-order HF wave function as follows:

$$\psi_i^{(1)} = \langle \psi_i^{(HF)} | -\hat{V}_{ee} \left| \sum_m a_m (\hat{T} \psi_i^{HF}) \right\rangle \quad (2.12)$$

where  $a_m$  is a coefficient in the expansion and  $\hat{T}$  is the coupled cluster operator. This operator generates all the excited determinants from the zero-order HF wave function.

It can be written as:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots + \hat{T}_n \quad (2.13)$$

where  $\hat{T}_1$  generates all singles from the HF wave function,  $\hat{T}_2$  generates all doubles etc. This only means that occupied orbitals in the HF determinant are replaced by virtual orbitals. These virtual orbitals have no physical meaning, are generated purely as linear combinations of the basis functions, and are not occupied with electrons. These linear combinations still conform to the symmetry point group of the system. They can be seen as analogous to

---

<sup>4</sup> A matrix element is an integral in bra-ket notation. Bra-ket notation is a short way of saying that the functions in the bra-ket should be integrated over all space for each function in the bra-ket. The left side of the bra-ket contains the complex conjugate of the function and the right hand side the function.

antibonding molecular orbitals; however, they are not real molecular orbitals. The larger the basis set, the larger the number of these orbitals and the more excitations that are possible and the better the possibility of approximating the electron correlation accurately. As these virtual orbitals are unoccupied, they will not contribute to the energy, unless they are forced to be occupied, such as when the coupled cluster operator operates on the HF wave function.

For a single excitation, one spin orbital is replaced by a virtual unoccupied spin orbital and for a double excitation two occupied spin orbitals are replaced by two virtual spin orbitals. Therefore, many single, double and third excitation Slater determinants etc. are generated by application of this operator. It can be shown that  $a_m$  is equal to:

$$a_m = \frac{\langle \hat{T}\psi_i^{HF} | -\hat{V}_{ee} | \psi_i^{HF} \rangle}{E_i^{HF} - \langle \hat{T}\psi_i^{HF} | \hat{H} | \hat{T}\psi_i^{HF} \rangle} \quad (2.14)$$

Substituting  $a_m$  into equation 2.12 and then substituting this equation for  $\psi_i^{(1)}$  in the matrix element for the second-order correction to the energy in 2.11 we obtain,

$$\begin{aligned} & \left\langle \psi_i^{(HF)} | -\hat{V}_{ee} \left| \frac{\langle \hat{T}\psi_i^{HF} | -\hat{V}_{ee} | \psi_i^{HF} \rangle}{E_i^{HF} - E_i^{ex}} (\hat{T}\psi_i^{HF}) \right. \right\rangle \\ &= \frac{\left| \langle \psi_i^{ex} | \hat{V}_{ee} | \psi_i^{HF} \rangle \right|^2}{E_i^{HF} - E_i^{ex}} \end{aligned} \quad (2.15)$$

where  $E_i^{ex}$  is equal to the total energy of all the excited determinants and  $\psi_i^{ex}$  is equal to the sum of all the Slater determinants that can be generated by the coupled cluster operator from the HF wave function. This equation can actually be simplified again as Condon-Slater rules state that only Slater determinants, which differ by at most 2 spin orbitals from the HF wave function can contribute to the energy [Levine, 2000]. Therefore, only doubles need to be incorporated in the equation and this gives,

$$\frac{\left| \langle \hat{T}_2\psi_i^{HF} | \hat{V}_{ee} | \psi_i^{HF} \rangle \right|^2}{E_i^{HF} - \langle \hat{T}_2\psi_i^{HF} | \hat{H} | \hat{T}_2\psi_i^{HF} \rangle} \quad (2.16)$$

For the rest of the derivation we will convert to explicit spin orbitals,  $i$  and  $j$ , for the unperturbed Slater determinant, and  $a$  and  $b$  for all of the doubly excited determinants.

$$\begin{aligned} & \sum_{i < j}^{occ} \sum_{a < b}^{vir} \frac{\left| \langle ij | \hat{V}_{ee} | ab \rangle \right|^2}{E_i^{HF} - E_i^{ex2}} \\ &= \sum_{i < j}^{occ} \sum_{a < b}^{vir} \frac{\left| \langle ij | \frac{1}{r_{12}} | ab \rangle - \langle ij | \frac{1}{r_{12}} | ba \rangle \right|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \end{aligned} \quad (2.17)$$



In equation 2.17 the double summations are over all the occupied and virtual orbitals.  $E_i^{ex2}$  is the total energy of all the doubly excited Slater determinants. In 2.17, we also substituted for the electron-electron repulsion operator. The  $\epsilon$ -values in the denominator in equation 2.17 are the eigenvalues of each orbital denoted by the subscripts. The denominator is simply equal to the difference in the energies of two Slater determinants.

The correction of the energy to the second-order can therefore be obtained by the following steps:

- ▶ Determine the ground state Hartree-Fock Slater determinant.
- ▶ Generate all the double excitations from this Slater determinant.
- ▶ Calculate the energy from the ground state Hartree-Fock Slater determinant by solving for the energy in the Schrödinger equation.
- ▶ Calculate the energies of all the doubly excited determinants.
- ▶ Work out the two-electron integrals between the Molecular orbital (MO) in the Hartree-Fock determinant and each MO in the doubly excited determinants. This is the most time consuming part of the MP2 method as two-electron integrals over the atomic orbitals should first be known [Jensen, 2001].
- ▶ The values obtained from the calculations above can then be substituted in equation 2.17 to determine the second-order correction to the HF energy.

In MP2 theory the Fock-operator is applied many times for various Slater determinants. MP theory is size extensive, which means that the method scales properly with the number of interacting fragments. As shown in Table 2.2, divergence can be the largest problem with higher order MP methods.

Table 2.2: The advantages and disadvantages of MP theory

ADVANTAGES OF MP theory	DISADVANTAGES OF MP theory
Size extensive (See text)	HOMO-LUMO gaps must be relatively large to avoid divergence [Knowles <i>et al.</i> , 2000]
Affordable incorporation of electron correlation	Divergence for higher MP-theories than MP2 must be monitored by a method such as CCSD(T) [Chałasiński and Szcześniak, 2000]
	Lower order MP-methods sometimes overbind van der Waals clusters [Jensen, 2001; Hopkins and Tschumper, 2004].
	Convergence can take extremely long if the HF wave function is far from the optimized wave function [Jensen, 2001].
	Basis set dependent. Can be divergent for some basis sets and for others not [Jensen, 2001].
	Parallel computing is not supported by all commercial software.
	The theory is not variational since the MP energy is not necessarily larger or equal to the true energy

### 2.5.2 Coupled Cluster (CC) theory

Truncated CC methods are used as a benchmark when calculating interaction energies for vdW clusters. They are all based on CC theory. CC theory is slightly different in its approach to MP theory. In truncated CC theory the corrections to the wave function in equation 2.5 only include specific excitations, such as doubles or singles. Equation 2.5 can be written as:

$$\psi = \psi_i^{(0)} + \lambda[t(\hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_N)\psi_i^{(0)}]^{(1)} + \lambda^2[t(\hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_N)\psi_i^{(0)}]^{(2)} + \dots \quad (2.18)$$

where  $t$  stands for each coefficient with which each excited Slater determinant need to be multiplied. The  $t$  variables are also called amplitudes. The zero-order wave function is usually chosen as in MP theory as the Hartree-Fock wave function.

MP theory truncates the series in equation 2.18 at a specific order, whereas truncated CC theory truncates the sum of the CC operators at a specific excitation. *Coupled Cluster with Singles and Doubles* (CCSD) is a common CC method that truncates the sum of the CC operators at the single and double excitations. CC theory always uses the series in 2.18 to infinity. CC theory is size-extensive due to the way in which the CC operators are used. The CC wave function is written as:

$$\begin{aligned} \psi_i^{CC} &= e^{\hat{T}} \psi_i^{HF} \\ e^{\hat{T}} &= 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \dots \end{aligned} \quad (2.19)$$

If only singles and doubles are considered, one obtains:

$$e^{\hat{T}} = e^{(\hat{T}_1 + \hat{T}_2)} = 1 + (\hat{T}_1 + \hat{T}_2) + \frac{1}{2}(\hat{T}_1 + \hat{T}_2)^2 + \dots \quad (2.20)$$

where  $\hat{T}$  is equal to the sum of all the operators used to generate the excited Slater determinants from the HF wave function.

The CCSD energy, or any other CC energy, can be written as in equation 2.21 [Jensen, 2001]:

$$E^{CCSD} = E^{HF} + \sum_i^{occ} \sum_a^{vir} t_i^a \langle \psi^{HF} | \hat{H} | \psi_i^a \rangle + \sum_{i < j}^{occ} \sum_{a < b}^{vir} (t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a) \langle \psi^{HF} | \hat{H} | \psi_{ij}^{ab} \rangle \quad (2.21)$$

where  $\psi_i^a$  means a single excitation, where the occupied orbital  $i$  in the zero-order HF Slater determinant was replaced by a virtual orbital  $a$  to create a new determinant.  $\psi_{ij}^{ab}$  is a double excitation where occupied orbitals,  $i$  and  $j$  in the zero-order HF Slater determinant are replaced by the virtual orbitals  $a$  and  $b$ . In this equation, we denote the zero-order HF wave function as  $\psi^{HF}$  and the HF energy as  $E^{HF}$ . The  $t$ -variables are the amplitudes for each excitation.

As can be deduced from equation 2.19, the  $e^{\hat{T}}$ -operator does not generate only these excitations and amplitudes. The number of amplitudes generated by  $e^{\hat{T}}$  is dependent on the truncation of the CC operator. Each amplitude is dependent on the other. Nonlinear equations, containing many other amplitudes for various excitations as generated from the zero-order HF wave function, can be written for each amplitude. In principle, the number of amplitudes to be determined is equal to the number of nonlinear equations [Levine, 2000]. Iterative methods are used to solve for the amplitudes. After all the amplitudes are solved, only those in equation 2.21 are substituted back to determine the energy and the wave function. The more excitations allowed, the better the accuracy of the amplitudes and eventually the energy and wave function.

The most common truncated CC theory used for vdW clusters is CCSD(T), which means that single excitations and double excitations are included for each occupied orbital, and triple excitations are approximated usually by an MP4 calculation [Levine, 2000]. The CC method is a large improvement over MP methods, but its applications are limited to small systems only such as single molecules or diatomic dimers. In Table 2.3 we list the advantages and disadvantages of CC theory.

Table 2.3: Advantages and disadvantages of the Truncated Coupled Cluster method

ADVANTAGES OF CC theory	DISADVANTAGES OF CC theory
CCSD(T) is seen as the benchmark method for determining accurate electronic energies	Very expensive in terms of hardware requirements
Size-extensive (See text on MP theory for definition)	Extremely long calculation times
	The theory is not variational since the CC energy is not necessarily larger or equal to the true energy
	Basis set dependent

### 2.5.3 Density functional theory (DFT)

Density functional theory (DFT) uses a different approach than the *ab initio*<sup>5</sup> methods discussed above. The theory of DFT comes from Hohenberg and Kohn [Hohenberg and Kohn, 1964] who proved in 1964 that the ground-state electronic energy of a molecule or atom is determined completely by its total electron density. The electron density at a point  $\mathbf{r}$  is according to the Born-interpretation, equal to the sum of the squares of all the spin orbitals at that point. The ground-state electronic energy for a system can be expanded into different energy terms, their values all dependent on the electron density of the ground-state. If we consider the system under study as nuclei suspended in a uniform electron gas (UEG) where electrons are not correlated and where electron exchange is not allowed, we can compute both the total kinetic energy and the total Coulomb energy for such a system. Fast multipole methods (FMM) make the scaling of the computation of the Coulomb energy directly proportional to the number of electrons [Jensen, 2001]. However, to obtain the total energy of the system, we still need to relate the electron density to the electron exchange and correlation energy in the exact system. To do this we need to derive functionals, or functions of the electron density that when the electron density is substituted into these functionals, will give the exchange and correlation energies exactly. This is the ultimate aim of DFT. If this is accomplished, we will be able to compute the exact energy or the exact solution to the many-electron Schrödinger equation by only knowing the system's electron density. This is the reason why Hohenberg and Kohn received the Nobel Prize for Chemistry in 1998. At present we are only able to approximate these functionals and there are a multitude of functionals, such as B3LYP available, that can be used for DFT calculations.

Similar to HF theory, DFT theory also employs a single Slater determinant for the total ground state wave function, but the canonical orbitals are of course different and are known as

<sup>5</sup> Some computational chemists regard DFT as an *ab initio* method. We will regard it as a method in its own class.

*Kohn-Sham* (KS) orbitals. The KS orbitals are eigenfunctions of the  $\hat{h}_{KS}$  operator and can be expanded in one-electron basis functions:

$$\hat{h}_{KS} = -\frac{1}{2}\nabla^2(\mathbf{r}_i) + \int \frac{\rho(\mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|} d\mathbf{r}_j + \frac{Z_k}{|\mathbf{R}_k - \mathbf{r}_i|} + \hat{V}_{xc} \quad (2.22)$$

The first term on the right hand side of the equation is equal to the kinetic energy of an electron  $i$  at position  $\mathbf{r}$ . Its position is therefore denoted by  $\mathbf{r}_i$ . The integral following this term is an integral over another electron  $j$ 's coordinates. This integral is equal to the Coulomb operator in HF theory except that here the electron density,  $\rho$  for electron  $j$ , is used instead of the wave function. The next term in the equation is the Coulomb interaction between electron  $i$  and a nucleus  $k$ .  $Z_k$  is the atomic number of that nucleus.  $V_{xc}$  is chosen as the derivative of the exchange-correlation energy with respect to the electron density of the KS orbital. The KS orbitals are solved just as with HF theory, by using undetermined Lagrange multipliers, to minimize the energy of the KS orbitals with the constraint that they should be orthogonal and normalized. To obtain the energies of the KS orbitals, equation 2.23 needs to be solved:

$$\hat{h}_{KS}\phi_i = \varepsilon_i\phi_i \quad (2.23)$$

In this equation,  $\phi_i$  is a KS orbital and  $\varepsilon_i$  is the energy of the KS orbital. This equation must be solved self-consistently, as the  $\hat{V}_{xc}$  operator is dependent on the electron density of the KS orbital.

In general, two approximations of the exchange and correlation functionals exist. In the *Local Density Approximation* (LDA), the electron density is seen as a UEG. In the *General Gradient Approximation* (GGA), the electron density is seen as a nonuniform electron gas and the exchange and correlation energies are made dependent not only on the electron density, but also on the derivatives of the electron density with respect to position. Exchange and correlation functionals can also be paired [Jensen, 2001].

To attempt to model vdW interactions with DFT, we note that the correlation and exchange components in the Hamiltonian are dependent on the interelectronic distance  $r_{12}$ . In the Hamiltonian this interelectronic distance can be expanded as:

$$\frac{1}{r_{12}} = \frac{1 - \text{erf}(\mu r_{12})}{r_{12}} + \frac{\text{erf}(\mu r_{12})}{r_{12}} \quad (2.24)$$

where  $\text{erf}$  is the standard error function. This separates the two-electron operator  $r_{12}$  into a short-range and long-range part.  $\mu$  is a parameter determining the ratio of these parts [Kamiya *et al.*, 2002]. The exchange functional can therefore also be separated into a long-range and short-range part.

With addition of a vdW correlation functional, the above approach is remarkably accurate for rare-gas dimers. The details are too complicated to discuss here, so the reader is referred the account by Sato and coworkers [Sato *et al.*, 2005]. They suggested that the reason why DFT fails to reproduce vdW bonds accurately is not so much the inadequate account of the electron correlation by vdW correlation functionals such as the ALL-functional [Andersson *et al.*, 1996], but the inadequate reproduction of the long-range exchange interactions that together with the short-range exchange interactions should counteract the correlation contribution. This may be the reason why standard DFT functionals cannot reproduce dispersion interactions and are unfortunately unsuitable for the modeling of vdW clusters. For the modeling of vdW clusters, we are therefore limited to *ab initio* methods such as MP2, which needs a significant number of computer memory and hard disk space, even for small systems. Table 2.4 summarizes the advantages and disadvantages of DFT for modeling intermolecular interactions and molecules.

DFT is size-consistent, meaning that one would obtain the same energy when calculating the energy of an H<sub>2</sub>-molecule, for example, and multiply that with 2, as when calculating the energy of 2 H<sub>2</sub>-molecules separated by 100 Å [Jensen *et al.*, 2001]. However, DFT is not necessarily size-extensive as it is usually accepted not to be able to model weakly bonded vdW clusters.

Table 2.4: Advantages and disadvantages of DFT

ADVANTAGES OF DFT	DISADVANTAGES OF DFT
Size-consistent (see text).	The functionals to use are dependent on the specific problem at hand.
Can be applied to large molecules and clusters.	Some functionals need to be parameterized for different atoms.
Not as basis set dependent as other methods.	Sometimes new functionals must be developed for new problems and this can take time.
Affordable incorporation of electron correlation and exchange.	Currently no functional exists that can be used routinely for the modeling of van der Waals clusters.
Can give results, comparable to higher level correlation consistent methods, in a much shorter time.	
Not as expensive as other correlated methods in terms of hardware specifications.	
Can be run in parallel.	

#### 2.5.4 Quantum Monte Carlo (QMC) methods

Quantum Monte Carlo can be used to derive a many-body wave function for a cluster, solid or molecule and should be used to add correlation to a wave function with insufficient correlation. It is also used to derive functionals for DFT. A trial wave function, subjected to insufficient correlation, can be obtained from an *ab initio* calculation with a basis set. This wave function is then used as the starting wave function for QMC, which will improve this wave function by adding the necessary correlation.

QMC is usually used for condensed media and solids, but has also recently been used to find the ground state wave function of He clusters [Bressanini *et al.*, 2000] and the HF dimer solvated in He [Sarsa *et al.*, 2002]. QMC scales better with size than CC theory, but incorporates almost all of the correlation energy. Currently QMC can be performed for systems with a thousand or more electrons. We will discuss one type of QMC briefly, namely *Diffusional Quantum Monte Carlo* (DQMC).

In DQMC, the imaginary time dependent Schrödinger equation (see equation 2.25) is solved.

$$-\frac{\partial \psi(\mathbf{R}, t)}{\partial t'} = \left(-\frac{1}{2}\nabla^2 + V(\mathbf{R}, t) - E_T\right)\psi(\mathbf{R}, t) \quad (2.25)$$

$t$  is a real variable measured in imaginary time.  $E_T$  is an energy offset determined by the state of the system at a particular time. It is important to note that in equation 2.25 the anti-symmetry principle is not obeyed and needs to be approximated with *a posteriori* methods, *vide infra*.

The idea is to solve the imaginary time dependent Schrödinger equation by making use of a Monte Carlo method whereby random walkers move in a  $3N$  coordinate electron and nucleus wave function space and diffuse randomly based on a diffusion equation. As they diffuse, they multiply or die out according to certain previously set rules that are based on the ratio of the probability density of the present step with the probability density of the previous step. This can be determined from the Boltzman distribution. The tempo at which the walkers multiply is proportional to  $V(\mathbf{R})$ , the potential energy at position  $\mathbf{R}$ . After an infinite time, it is expected that the walkers will move towards the density of the ground state wave function that is a solution to the imaginary time dependent Schrödinger equation. The Schrödinger equation is solved for each wave function at a specific timestep by Monte Carlo integration. When after several trials the energy cannot be lowered any further, the resulting energy is the exact solution of the many-electron time dependent Schrödinger equation without the use of the Born-Oppenheimer approximation. The only problem is that the Schrödinger equation does not recognize spin and so the Pauli-exclusion principle is not obeyed. Therefore, an approximation of the sign of the wave function has to be made *a posteriori*. One of the

approximations that can be used is the fixed node approximation [Foulkes *et al.*, 2001]. This however does not have a large influence on the final energy. Other properties such as accurate vibrational modes can also be determined. In Table 2.5 the advantages and disadvantages of DQMC are summarized.

Table 2.5: Advantages and Disadvantages of DQMC [Foulkes *et al.*, 2001]

ADVANTAGES OF DQMC	DISADVANTAGES OF DQMC
It gives an accurate wave function treatment of many-body effects.	If the starting guess wave function is too far off from the "correct" wave function then it will take longer for the walkers to accumulate towards the "correct" electron density.
It is a very general approach, which is applicable to solids and molecules and can determine almost any ground state expectation value.	Because the wave function does not obey the anti-symmetry principle, approximations need to be made, such as the fixed node approximation.
The computational cost scales with $N^3$ , where a method such as CCSD(T) scales with $N^7$ . $N$ is the number of electrons.	It is demanding to calculate first and second derivatives of the total energy with respect to the atomic positions.
It has low storage requirements and it benefits from being run on parallel architectures.	The results are obtained with a statistical error bar that decays with the inverse of the square root of the calculation time.
The DQMC method does not suffer from basis set effects, as no basis sets are used.	The fixed-node approximation may bias the results in some cases.
When tested on a $C_{10}$ -cluster DQMC recovered <i>circa</i> 95% of the correlation energy, whereas CCSD(T) only recovered about 75%.	Only limited information about excited states is available.
There is QMC-software available such as zori <sup>6</sup> .	For solids the results may suffer from finite size errors.

For more information, see Foulkes *et al.*, 2001.

### 2.5.5 Other methods

In the above discussion, we discussed a few well known and promising methods for the modeling of van der Waals clusters. There are also other methods to study intermolecular interactions such as the localized-MP2 (LMP2) [Saebo *et al.*, 1993; Schutz *et al.*, 1998] method that will probably compete with DFT in modeling vdW clusters in the future [Chalański and Szcześniak, 2000]. Due to lack of space and time, it is not possible to go into all of these methods or even discuss them in detail.

<sup>6</sup> Zori is open source and can be obtained from the Lester-group.  
<http://www.cchem.berkeley.edu/~walgrp/>



## ***2.6 Methods specifically pertaining to van der Waals clusters and intermolecular interactions***

The methods discussed in **Section 2.5** are standard methods that have been used for the modeling of vdW clusters before. When one models vdW clusters, it is of little value to compare electronic energies of clusters. The interaction energy and binding energy are usually used to give some insight into the stabilization of vdW clusters.<sup>7</sup>

In this section, we will discuss two methods that are specifically designed for determining the electronic interaction energies of clusters as accurately as possible within the limits of the basis set and method used.

### *2.6.1 The many-body supermolecular perturbation theory (MBPT)*

According to the *many-body supermolecular perturbation theory* (MBPT), one can calculate the electronic interaction energy of a cluster by subtracting the electronic energies of the individual monomers at infinite separation, constrained to their geometries in the cluster, from the overall electronic energy of the cluster and then correct this interaction energy for BSSE [Chalasiński and Szcześniak, 2000].

An approximate correction can be done by using a counterpoise correction [Boys and Bernardi, 1970]. To apply the many-body perturbation theory, no special software is needed; standard *ab initio* computational software can be used. In Table 2.6 the advantages and disadvantages of MBPT are summarized.

---

<sup>7</sup> This can also be used as a criterion for force field optimization.

Table 2.6: Advantages and disadvantages of many-body supermolecular perturbation theory

ADVANTAGES OF SUPERMOLECULAR PERTURBATION THEORY	DISADVANTAGES OF SUPERMOLECULAR PERTURBATION THEORY
Simple implementation	Basis set superposition error can only be approximated
No special software needed	The many body interaction can not be expanded in separate components
In general it gives good results	

### 2.6.2 Symmetry Adapted Perturbation theory (SAPT or I-MP)

*Symmetry Adapted Perturbation Theory* (SAPT) [Rybak *et al.*, 1991], is another method that is used for the determination of accurate interaction energies. Instead of writing the interaction energy as the difference between the energy of the cluster and the individual monomers, one can write it directly as the sum of individual terms for the different energy contributions.

$$E_{\text{int}} = E_{\text{pol}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{pol}}^{(2)} + E_{\text{exch}}^{(2)} + \dots \quad (2.26)$$

$E_{\text{pol}}^{(1)}$  is the classical electrostatic interaction energy.  $E_{\text{pol}}^{(2)}$  is a sum of classical induction and quantum mechanical dispersion energies.  $E_{\text{exch}}^{(1)}$  is the first-order exchange correction and  $E_{\text{exch}}^{(2)}$  is the second-order exchange correction. Both these terms are defined in the context of SAPT [Rybak *et al.*, 1991]. The true or total Hamiltonian for two molecules  $A$  and  $B$  can be expanded by using perturbation theory:

$$\hat{H} = \hat{F} + \zeta \hat{V} + \lambda \hat{H}'_A + \lambda \hat{H}'_B \quad (2.27)$$

where  $\hat{F}$  is the Fock-operator operating on the zero-order HF wave function and  $\hat{V}$  is the interaction operator and is defined as the difference between the true Hamiltonian and the true Hamiltonians of the composite monomers of the cluster.  $\hat{H}'_A$  and  $\hat{H}'_B$  is the perturbations on the HF energies of molecules  $A$  and  $B$  respectively and can both be expanded into a series containing different orders of the wave functions of the monomers.  $\zeta$  and  $\lambda$  are simply constants determining the number of perturbation and can be set to one for full perturbation. The interaction operator does not preserve the anti-symmetry principle, as electrons between the monomers are not forced to obey the anti-symmetry rule.

To create a space in which the operator can function, we therefore need a larger space. The tensor product can therefore be taken between the two Hilbert spaces of molecules  $A$  and  $B$  to give a space in which the interaction operator can operate. This is why SAPT is a symmetry

adapted perturbation theory. The Hilbert spaces for each molecule are subspaces of the new Hilbert space.

Taking all of this into consideration, all the different components of the interaction energy at the Hartree-Fock level and corrections to the energies of these interactions can be written as in equation 2.27. Each component can be expanded with respect to Rayleigh Schrödinger perturbation theory. As an example, the polarization energy can be written as:

$$E_{pol}^n = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} E_{pol}^{(nij)} \quad (2.28)$$

where  $E_{pol}^{(nij)}$  is of the  $n$ th order in  $\hat{V}$  and the  $i$ th order in  $\hat{H}'_A$  and  $j$ th order expansion in  $\hat{H}'_B$ .

SAPT has advantages over other *ab initio* methods such as the traditional MBPT method, but is also not foolproof, as convergence of the series still contributes to failures in some cases. It is therefore important to test the results with a supermolecular method [Chalasiński and Szcześniak, 2000]. Table 2.7 summarizes the advantages and disadvantages of SAPT. SAPT is usually used when very accurate quantum mechanically based potentials are derived for vdW clusters such as for the CO dimer [Vissers *et al.*, 2005].

Table 2.7 Advantages and disadvantages of SAPT [Rybak *et al.*, 1991]

ADVANTAGES OF SAPT	DISADVANTAGES OF SAPT
The interaction energy is calculated directly as the sum of physical contributions.	Divergence can take place as is general with perturbation theory.
The interaction energy so obtained is free from BSSE.	Special software is needed and the theory behind the method is complicated.
The components of the interaction energy are size extensive; therefore, it can be applied to polyatomic molecules.	
Different energy corrections exhibit different angular and radial dependencies so it would be possible to obtain excellent analytical fits in terms of physical interpretable parameters.	
The individual energy corrections show different basis set requirements and can therefore be calculated separately.	
The overall computational effort is smaller than in the case of the many-body supermolecular perturbation theory.	
Free software is available on the internet.	

## 2.7 Molecular mechanics (MM)

It would have been ideal if all computational problems could be solved by quantum mechanics. However, most systems are still too large today to be considered by this. When a computational problem is too large to be treated with quantum mechanics (QM), it is generally modeled by using molecular mechanics (MM).

MM is based on a Newtonian physics approach and makes use of force fields. It is a method whereby quantum mechanics are approximated with classical mechanical calculations in order to speed up calculations. MM makes extensive use of the Born-Oppenheimer principle discussed in **Section 2.3**, as the change in energy of a molecular system is made dependent only on the change in nuclear coordinates [Leach, 2001].

In QM, electrons experience different forces dependent on their environments. For example, a hydroxyl oxygen atom will interact differently to a carbonyl oxygen atom. As MM cannot account for the movement of individual electrons, various atom-types are used. In the above case, separate atom-types will be used for the hydroxyl oxygen and carbonyl oxygen atoms.

In this work, the Universal Force Field (UFF) [Rappé *et al.*, 1992] was used extensively. In UFF, the atom-types are written in the following way. The first two characters correspond to the chemical symbol. If only one character is used, it is followed by an underscore. The

chemical symbol is followed by a description of the hybridization or geometry: 1 = linear (sp-hybridized), 2 = trigonal (sp<sup>2</sup>-hybridized), R = resonant, 3 = tetrahedral (sp<sup>3</sup>-hybridized), 4 = square planar (dsp<sup>2</sup>-hybridized), 5 = trigonal bipyramidal (dsp<sup>3</sup>-hybridized) and 6 = octahedral (d<sup>2</sup>sp<sup>3</sup>-hybridized). Indicators such as the formal oxidation state or indicators that the atom is part of a zeolite lattice etc may follow this. To illustrate this; in this study, we used C\_1 as an atom-type in the carbon monoxide monomer, as this atom is sp-hybridized in the monomer [Rappé *et al.*, 1992].

Next, the interaction of the atom-types with each other must be defined. For each atom-type a number of parameters are derived based on experimental information or quantum mechanical calculations. These parameters are used as the reference values in the force field. The energy of the interaction between atom-types (also called steric energy) can be written in MM as the sum of a number of contributions:

$$E_{steric} = E_{stretch} + E_{bend} + E_{tors} + E_{inv} + E_{el} + E_{vdw} + E_{cross} \quad (2.29)$$

where  $E_{stretch}$  is the sum of all the energies for bond stretches from the reference value.  $E_{bend}$  is the sum of all the energies for all the angle bends from a reference angle.  $E_{tors}$  is the sum of all the torsional energies when dihedral angles are changed from a reference value.  $E_{inv}$  is the sum of all the inversion potentials of some molecules and corresponds to an out of plane deformation of an atom from a reference value.  $E_{el}$  is the electrostatic interaction and usually has no previously defined reference value in the force field.  $E_{vdw}$  is the vdW interaction between nonbonded atoms and is again based on the change from reference values. The last term in 2.29 is the energy for the cross-terms.

If the cross-terms are omitted, then the internal coordinates, are treated independently of each other, which is a severe approximation as there are, for example, molecules where a change in the valence angle can result in valence bond lengthening. The energy associated with the

cross-terms is the energy for the coupling between the internal coordinates. By including cross-terms, the force field's accuracy, in determining the steric energy, is increased. The derivatives of the energy with respect to the nuclear coordinates of the molecule(s) also become more accurate.

The terms in equation 2.29 have different mathematical expressions that are chosen to model the specific interaction considered. For example, a bond stretch can be modeled by either a Morse-potential, a Taylor expansion around the reference bond length or sometimes even the simple Hook's equation [Leach, 2001]. Parameters for these mathematical expressions are chosen to mimic experimental results or quantum mechanical results.

In this work, we are only concerned with the  $E_{el}$  and  $E_{vdw}$  terms. Therefore, in the rest of this section we will look in detail at these two terms.

### 2.7.1 The van der Waals term

The van der Waals term is based solely on two of the four intermolecular interactions discussed earlier in **Section 2.2**. The first is a repulsive interaction due to the effect of the electron-electron exchange and the second is an attractive interaction due to the dispersion interaction. The vdW interaction energy for the interaction between atoms  $A$  and  $B$ , can be written as in equation 2.30:

$$E_{\text{vdW}}(R^{AB}) = E_{\text{repulsive}}(R^{AB}) - \frac{C^{AB}}{(R^{AB})^6} \quad (2.30)$$

The positive constant  $C^{AB}$  is fitted to the attractive term of the van der Waals interaction between the two atoms based on quantum mechanical calculations and shows the same power-law relationship as a model such as the Drude model [Leach, 2001]. The repulsive term cannot be derived in a classical way, as it is solely based on a quantum effect with no classical analogue.

Many functions have been proposed to approximate this function. For example, one approximation for the repulsive part is an exponential function. The only problem that will be encountered with such a function is when the user places molecules accidentally too close to each other in their starting geometries. Then the vdW forces between the molecules will be too large and make them fly apart. The Buckingham type potential uses an exponential as the repulsive term. One can also use a Morse-potential [Jensen, 2001].

As these potentials cannot be calculated quickly, their calculation can slow down the calculation of the steric energy. Therefore, the Lennard-Jones potential is usually used. This potential is less accurate, but much quicker to calculate. Two flavors of the Lennard-Jones potential are most often used, as shown in equations 2.31 and 2.32:

$$E_{12-6}^{LJ}(R) = \varepsilon \left[ \left( \frac{R_0}{R} \right)^{12} - 2 \left( \frac{R_0}{R} \right)^6 \right] \quad (2.31)$$

$$E_{9-6}^{LJ}(R) = \varepsilon \left[ 2 \left( \frac{R_0}{R} \right)^9 - 3 \left( \frac{R_0}{R} \right)^6 \right] \quad (2.32)$$

where  $\varepsilon$  is the reference well depth or the dissociation limit in kcal/mol and  $R_0$  is the reference van der Waals distance in Å between two atoms.  $R$  is the distance between the atoms at any time during the optimization or simulation. The 12-6 Lennard-Jones potential rises more quickly for short intermolecular distances than the 9-6 potential.

### 2.7.2 *The electrostatic energy*

In standard force fields the electrostatic energy is determined by the Coulomb interaction between point charges centered on atoms or in some force fields by the interaction between dipole moments of molecules. The use of stationary atomic point charges to represent a charge distribution is in most cases a severe approximation, as two approaching charge distributions will influence each other and lead to a change in the initial atomic point charges. Furthermore, charge distributions of polar molecules are anisotropic and not entirely spherically symmetrical and can therefore not be accurately represented by atomic point charges.

For clusters, the electrostatic interaction term should incorporate the remaining two of the four nonbonded interactions, which are the induction and electrostatic interactions. To include the induction interaction, the charges used in the force field should be allowed to polarize other atoms, changing their charges or dipoles to minimize the energy. Force fields that are able to do this are called polarizable force fields and may play an important role in the development of better QM/MM methods. See, for example, [Kaminski *et al.*, 2002] for information on a polarizable force field. The charge equilibration method of Goddard and coworker [Rappé and Goddard III, 1991] is a way of polarizing classical force fields; however, it is not as accurate as polarizable force fields.

## **2.8 Ways to represent a charge-distribution**

### 2.8.1 *Merz-Kollman-Singh (MKS) charges*

The well-known Mulliken charges are not accurate in describing a charge distribution, as they are basis set dependent. Mulliken charges depend on the square of the coefficients of the constituting basis functions of the molecular orbitals and the overlap populations of the basis functions as described by the overlap matrix [Levine, 2000; Jensen, 2001]. A more accurate representation of charges is based on the electrostatic potential (ESP). The ESP at a point is defined as the force acting on a unit of positive charge placed at that specific point. The electrostatic potential is a property of the wave function.

The ESP is determined at a set of chosen points. Which points are chosen depends on the charge derivation scheme used. A least squares fitting procedure is employed to determine the set of partial atomic charges that will reproduce the ESP at the specific chosen points. The error between the true electrostatic potential at a point and the electrostatic potential

calculated at that point, is minimized under the constraint that only charges that sum to the charge of the entire system, are used [Leach, 2001]. In this way charge is conserved.

In order to choose the points on the ESP to be sampled for the determination of point charges, the Merz-Kollman-Singh (MKS) [Singh and Kollman, 1984] charge-scheme first determines the overlap surface of the vdW radii of all the atoms in the system. Then points are sampled either on the border of this surface or outside the surface. The region outside the surface is divided into layers determined by creating several surfaces around the overlap surface of the vdW radii. Usually the layers are obtained by scaling the vdW radii of the atoms. CHELPG charges [Breneman and Wiberg, 1990] are also based on the ESP, but this will not be discussed here.

## 2.9 QM/MM and the ONIOM hybrid methods

In this study, we made extensive use of the “*Our own N-layered Integrated molecular Orbital molecular Mechanics*” (ONIOM) method developed by Keiji Morokuma and co-workers in 1996 [Svensson *et al.*, 1996] to do QM/MM calculations. This method is more versatile than other methods to do QM/MM, and can for example be extended to calculations not only incorporating a QM and MM region, but also 2 QM regions. In this section we will mainly discuss the theory behind performing a QM/MM calculation utilizing the ONIOM methodology. However, for completion, we will also include a brief discussion of other hybrid methods and their applications and show the versatility of the ONIOM methodology in extending the QM/MM system to more complicated hybrid systems.

### 2.9.1 Introduction to QM/MM

The aim of QM/MM is to treat one part of a molecular system with quantum mechanics (QM) and another that may be too large for QM, with MM. In general, the total energy of a system can therefore be written as:

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM} \quad (2.33)$$

where  $E_{total}$  is the total energy of the entire system.  $E_{QM}$  is the energy for the part of the system treated with a QM method and  $E_{MM}$  is the energy of the system treated with a MM method.  $E_{QM/MM}$  is the energy of the interaction between the MM and QM systems. A QM/MM calculation can be done with the original QM/MM methodology or in the framework of the ONIOM methodology.



### 2.9.2 The original QM/MM methodology

In the original QM/MM methodology the term  $E_{QM/MM}$  is written as in equation 2.34 [Cui and Karplus, 2000].

$$E_{QM/MM} = \sum_{i=1}^N \sum_{j=1}^m \langle \psi_i | \frac{q_j}{R_{ij}} | \psi_i \rangle + E_{vdW} + \sum_{k=1}^n \sum_{j=1}^m \frac{q_j Z_k}{R_{jk}} \quad (2.34)$$

The first double sum in 2.34 is a summation over  $N$  electrons in the QM system and  $m$  point charges on the atoms in the MM system. The QM wave function  $\psi_i$ , is polarized by the point charges in the MM system, each one denoted by  $q_j$ .  $R_{ij}$  is the distance between an electron and a point charge.  $E_{vdW}$  is the van der Waals interaction between the two systems, described by a potential such as the Lennard-Jones potential. The last term in 2.34 is the interaction between the charge of a nucleus,  $k$ , in the QM system and a point charge  $q_j$  in the MM system. There are  $n$  nuclei in the QM system and  $Z_k$  is the atomic number of nucleus  $k$ .  $R_{jk}$  is the distance between a nucleus in the QM system and a point charge in the MM system. Although 2.34 is by default used for QM/MM calculations based on the original methodology, we will see that this type of calculation can be regarded in the ONIOM methodology as ONIOM utilizing electronic embedding, since the MM atoms are allowed to polarize the wave function of the QM system.

The first QM/MM hybrid method was described in 1976 [Warshel and Levitt, 1976]. This was applied to a very simple system, using a semi-empirical method for the QM part, and its covalent linkages, and a force field for the MM part. The vdW interactions and electrostatic interactions were modeled with a 6-12 Lennard-Jones potential and a Coulomb sum respectively. As technology improved, Hartree-Fock and later DFT methods were used to describe the QM part. There are few literature accounts where other *ab initio* methods such as MP2 or CCSD were used for the QM part [Kongsted *et al.*, 2003; Osted *et al.*, 2004; Muños-Losa *et al.*, 2003].

QM/MM calculations are usually done by interfacing a QM program with an MM program. The QM program does the QM calculations while the MM program does the MM calculations. The  $E_{QM/MM}$  part is computed by a separate program or by additional code to an existing program. Problems with the QM/MM methodology starts to arise when one wants to determine properties such as NMR chemical shifts for the QM system and normal modes of vibration for the entire system. We will later see that using the ONIOM framework for QM/MM can solve all of these problems efficiently.

### 2.9.3 “Our own *N*-layered Integrated molecular Orbital molecular Mechanics” (ONIOM) method

The ONIOM method can be regarded as an extrapolation scheme. For a QM/MM system this extrapolation can be seen as a way to approximate the total energy of the system at the QM level of theory by adding the energy, for the QM system only, to an energy correction that is determined at the MM level of theory. However, due to the nature of ONIOM the lower-level of theory need not necessarily be MM. All of these concepts will become clearer as we progress with this discussion.

The framework of ONIOM divides the system under consideration into a set of  $N$  layers, rather than just 2 layers as in the original QM/MM methodology. The layers are treated with different levels of theory. Usually  $N$  is equal to two or three. The user can choose these layers arbitrarily. Whereas QM/MM systems treat the  $E_{QM/MM}$  term in equation 2.33 as a separate term, in ONIOM this term is divided between the calculations of the energies of the different layers. ONIOM has a unique way of defining these layers and in Fig. 2.2 we illustrate a three layered ONIOM scheme.

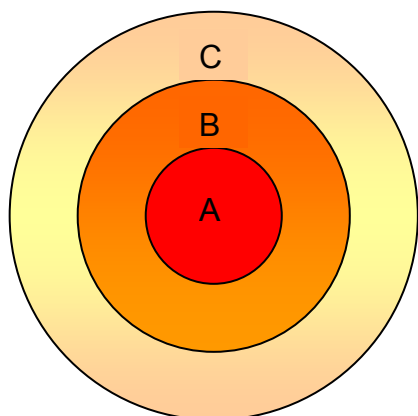


Fig. 2.2: ONIOM explained graphically: A=model system, A+B=intermediate model system, A+B+C=real system.

In ONIOM terms, if the word “high” appears next to a term or property, it means the highest-level calculation of the chosen methods must be performed on this part of the system. If the word “medium” appears next to a term, it means that a medium-level calculation should be used, i.e. a method that is less accurate than the method used in the previous part of the system. If the word “low” appears, it means a low-level method should be used or a method less accurate than in the part of the system treated with a medium-level calculation. For example, the term  $E_{\text{model}}^{\text{low}}$  means a low-level calculation on the model system or  $A$  in Fig. 2.2.

Before attempting an ONIOM calculation, one must first decide which methods are going to be used. For example, one can use CCSD(T) for the high-level calculation, MP2 for the

medium-level calculation and UFF for the low-level calculation. The ONIOM extrapolated energy for a three-layered system is defined as:

$$E(\text{ONIOM}) = E_{\text{model}}^{\text{high}} + E_{\text{intermed}}^{\text{med}} - E_{\text{model}}^{\text{med}} + E_{\text{real}}^{\text{low}} - E_{\text{intermed}}^{\text{low}} \quad (2.35)$$

ONIOM(2) with a QM method in one layer and a MM method in another layer is similar to performing a QM/MM calculation. The energy for such a system can simply be written as:

$$E(\text{ONIOM}) = E_{\text{model}}^{\text{high}} + E_{\text{real}}^{\text{low}} - E_{\text{model}}^{\text{low}} \quad (2.36)$$

In the general ONIOM framework, also called ONIOM with mechanical embedding, stationary charges are used for the QM and MM atoms. The user can specify the charges for the atoms in the QM and MM systems in the input or can utilize charge equilibration charges [Rappé and Goddard III, 1991]. ONIOM with mechanical embedding for QM/MM is known as IMOMM (Integrated molecular orbital molecular mechanics) [Masseras and Morokuma, 1995]. In 2003 a new development, ONIOM-EE or ONIOM with electronic embedding was introduced in the ONIOM framework. This allows the atomic point charges on the atoms in the MM system to polarize the QM wave function.

To compute the ONIOM extrapolated energy for a QM/MM system, three calculations are necessary, one calculation for each of the terms in equation 2.36. Depending on the use of mechanical or electronic embedding, the calculations for the energy for the terms will vary. As we only used electronic embedding in this work, we will discuss the calculations based on the ONIOM-EE scheme. For ONIOM-EE we can write for the model system calculated at the high-level of theory:

$$\hat{H}_{\text{model}}^{\text{high}} \psi_i = (\hat{H} + \sum_{i=1}^N \sum_{j=1}^m \frac{q_j}{R_{ij}}) \psi_i + \sum_{k=1}^n \sum_{j=1}^m \frac{q_j Z_k}{R_{jk}} \psi_i \quad (2.37)$$

where  $\hat{H}$  is the standard electronic Hamiltonian for an isolated QM system,  $\psi_i$  is the wave function,  $N$  is the number of electrons in the QM system and  $m$  is the number of point charges,  $q_j$  in the MM system.  $R_{ij}$  is the distance between electron  $i$  and MM atom  $j$ .  $n$  is the number of nuclei in the QM system and  $R_{jk}$  is the distance between a nucleus in the QM system and a point charge in the MM system. Now we multiply with the complex conjugate of  $\psi_j$  and integrate over all space. Only the terms where  $i=j$  will survive to give:

$$\begin{aligned} E_{\text{model}}^{\text{high}} &= \langle \psi_i | \hat{H} + \sum_{i=1}^N \sum_{j=1}^m \frac{q_j}{R_{ij}} | \psi_i \rangle + \sum_{k=1}^n \sum_{j=1}^m \frac{q_j Z_k}{R_{jk}} \langle \psi_i | \psi_i \rangle \\ &= E_i + \sum_{i=1}^N \sum_{j=1}^m \langle \psi_i | \frac{q_j}{R_{ij}} | \psi_i \rangle + \sum_{k=1}^n \sum_{j=1}^m \frac{q_j Z_k}{R_{jk}} \end{aligned} \quad (2.38)$$

where  $E_i$  is the electronic energy obtained when the electronic Hamiltonian operates on the polarized wave function. The interaction between a charge of a nucleus and a point charge in the MM system does not depend on the wave function and is therefore a constant for a specific geometry.

Turning our attention to the other terms in equation 2.36, we can write:

$$E_{\text{real}}^{\text{low}} = E_{\text{vdW}} + E_{\text{el}}^{\text{MM}} \quad (2.39)$$

$$E_{\text{model}}^{\text{low}} = E_{\text{model}}^{\text{MM}} + (E_{\text{el}}^{\text{MM}}) \quad (2.40)$$

where  $E_{\text{vdW}}$  is the total van der Waals interaction energy of the entire system and  $E_{\text{el}}^{\text{MM}}$  is the electrostatic interaction energy based on the initial point charges given for the QM atoms and the MM atoms in the input. For ONIOM-EE and general ONIOM, equation 2.39 is the same, but for ONIOM-EE the term in parenthesis in equation 2.40 also need to be added. The reason for this is that if charges are given for the QM atoms in the input,  $E_{\text{el}}^{\text{MM}}$  need to be subtracted from the total energy when applying equation 2.36. If it is not subtracted the electrostatic energy between the QM and MM systems will be counted twice - first at a QM level of theory and second at the MM level of theory.

#### 2.9.4 Geometry optimizations for QM/MM

One of the big problems currently encountered for QM/MM systems is how to optimize such a system in an efficient time [Vreven *et al.*, 2003]. Optimizing the whole system in Cartesian coordinates will be laborious as the quantum mechanical system will not be able to be optimized quickly in Cartesians.<sup>8</sup> Optimizing the whole system in redundant internal coordinates will also be inefficient, since the MM system that may contain thousands of atoms, will be impossible to optimize in a reasonable time. Therefore, today, most QM/MM methods by default use separate optimization algorithms for both systems. Although this complicates the optimization of a QM/MM system, this approach to geometry optimizations is well known to be more efficient [Vreven *et al.*, 2003] than otherwise. When the systems are optimized separately, the optimization of the QM system is known as macro-iterations and the optimization of the MM system is known as micro-iterations. When one system is optimized, the other system is held stationary. This is how ONIOM optimizations are done in *Gaussian 03* [Frisch *et al.*, 2003].

We will discuss 3 types of optimization strategies, from the literature, in order to explain the optimization algorithms implemented in *Gaussian 03*. Before we continue, we should note that optimization algorithms are applicable to both the original QM/MM methodology and the ONIOM methodology.

Assuming we use micro-iterations for the MM system, we can divide the entire system's coordinates into two sets of coordinates: coordinates for the QM system and coordinates for the MM system. The total first derivative (gradient) of the energy for an entire system can

---

<sup>8</sup> Calculating the energy for a QM system can take a long time and Cartesian coordinate optimizers use more optimization steps than redundant internal coordinate optimizers do.

therefore be determined for the coordinates ( $q_m$ )<sup>9</sup> of the QM system where all the derivatives involving MM coordinates ( $q_i$ ) therefore becomes equal to zero. This calculation can be repeated for the MM coordinates where the derivatives involving the QM coordinates will of course vanish. This is illustrated in equation 2.41.

$$\begin{aligned} \frac{\partial E_{total}}{\partial q_m} &= \frac{\partial E_{QM}}{\partial q_m} + 0 + \frac{\partial E_{QM/MM}}{\partial q_m} \\ \frac{\partial E_{total}}{\partial q_i} &= 0 + \frac{\partial E_{MM}}{\partial q_i} + \frac{\partial E_{QM/MM}}{\partial q_i} \end{aligned} \quad (2.41)$$

As can be expected when the MM system is optimized in the micro-iterations, the total gradient with respect to  $q_i$  should be determined for each optimization step. This will slow down the micro-iterations since Cartesian optimization algorithms, as already mentioned, use much more steps for a geometry optimization than redundant internal coordinate optimizers do. As QM/MM is usually used for large systems that cannot be optimized with general *ab initio* methods, this will create problems depending on how large the MM system is. Therefore, in order to speed up the micro-iterations, approximations of the gradient need to be made, of which the approximation of Yang and coworkers [Yang *et al.*, 2000] is the most severe, although, according to them, was still very efficient for their purposes. Instead of using the exact gradient in the micro-iterations, they removed the part of the exact gradient involving the forces between the MM and QM atoms, due to the polarization of the MM point charges on the wave function, and replaced this with a new gradient containing an approximation of the electrostatic energy. To approximate the electrostatic energy they derived ESP atomic point charges based on the polarized QM systems' electron density and calculated the Coulomb energy for the interaction between these charges and the charges in the MM system. When the QM system was optimized, the exact gradient was used, but when the MM system was optimized, the approximate gradient was used. This is only a good approximation if one assumes that the MM system does not move significantly during the micro-iterations with respect to the QM system. Vreven *et al.* [Vreven *et al.*, 2003] commented that this method is inaccurate as the potential energy surface (PES) for optimizing the QM system and the PES for optimizing the MM system is not the same. Friesner and coworkers, [Murphy *et al.*, 2000] used a better method for optimizing the MM system in the micro-iterations. They used the exact gradient or force with respect to the MM coordinates both to optimize the QM and the MM systems, but instead of calculating the exact forces for the MM system at each optimization step while it moved with respect to the QM system in the micro-iterations, they used an approximation. They derived the ESP point charges as based on the electron density of the QM system and used the change in the Coulomb energy between

---

<sup>9</sup> Note that  $q$  is now used for coordinates rather than charges as previously.

these charges and the stationary charges on the atoms in the MM system to approximate the change in the exact force during the micro-iterations. The original gradient at step  $n$  in the micro-iterations is therefore perturbed as illustrated in equation 2.42:

$$\frac{\partial E_{total}}{\partial q_i}(n) = \frac{\partial E_{MM}}{\partial q_i}(0) + \frac{\partial E_{QM/MM}}{\partial q_i}(0) + \left( \frac{[E_{el}^*(n)]}{R(n)} - \frac{[E_{el}^*(0)]}{R(0)} \right) \quad (2.42)$$

where the terms in parenthesis are an approximation of how the original force should change from step 0 or at the start of the micro-iterations to step  $n$  in the micro-iterations.  $E_{el}^*$  is the electrostatic energy as determined by a Coulomb interaction between the stationary charges on the MM atoms and the ESP charges on the QM atoms.  $R$  is the distance between the ESP derived charge and a stationary charge in the MM system. Eventually as the electrostatic force on the MM atoms exerted by the QM system becomes zero, the force of the QM system on the MM system becomes exact. However, there is one problem with this method. When the exact electrostatic force exerted by the QM system on the MM system initially differs largely from the approximated electrostatic force that should direct the optimization, the approximate change in the exact force will be negligible and the force of the QM system on the MM system will not become exact. We will discuss this further in **Chapter 9**. In Friesners' scheme [Murphy *et al.*, 2000] one optimization step is taken for the QM system after which the MM system is optimized with a stationary QM system. The process is then repeated until the two systems are self-consistent. Vreven and coworkers, [Vreven *et al.*, 2003] suggested that this method can be improved since the MM system will not be optimized when another optimization step is taken for the QM system. The new method is implemented in *Gaussian 03*. They suggested that after the micro-iterations the wave function should be reevaluated to see how the charges of the QM system should be updated after the micro-iterations and then a new set of micro-iterations should follow. An optimization step for the QM system is only taken after the MM system is optimized. Before an optimization step can be taken for the QM system, the Hessian (second-derivative matrix) has to be updated with the derivative of the forces of the optimized MM system on the unoptimized QM system. After an optimization step is taken in the QM system, a new set of micro-iterations is employed unless the MM system is still optimized after the optimization step in the QM system. The process is repeated for each step in the QM system until the QM and MM systems are self-consistent and the net force on the entire system is as close to zero as possible. Sometimes it might be that the QM optimization step takes the system out of the potential energy well obtained after the MM optimization. *Gaussian 03* makes sure that the QM optimization step stays in the correct potential energy well. For more information, the reader is referred to the article by Vreven and coworkers [Vreven *et al.*, 2003].

It is still possible to use a redundant internal coordinate optimization algorithm in *Gaussian 03* for ONIOM-EE geometry optimizations, but this optimization algorithm has not been adapted to include proper electronic embedding. A proper treatment is only available when separate optimization algorithms are used for the QM and the MM systems.

#### 2.9.4 QM/MM system optimizations with constraints

Since *Gaussian 03* allows the user to use constraints in redundant internal coordinates for the optimization of the QM system and constraints in Cartesian coordinates for the optimization of the MM system, one is allowed flexibility in the optimization of the QM system, but not much flexibility in the optimization of the MM system. In the optimization of the QM system, modified redundant internal coordinates can be added, redundant internal coordinates can be stepped to find transition states and internal coordinates can be frozen. In the MM system however, there are only two options: the Cartesian coordinates can either be frozen or a specified set of Cartesian coordinates can be made to behave as a rigid block. We will now briefly discuss these two options as both have been used in this work.

When the Cartesian coordinates of atoms in the MM system are frozen, the optimization should be slightly modified. When systems are optimized without any frozen atoms in the MM system, the whole QM system is usually frozen while the MM system is optimized. However, when atoms in the MM system are frozen these atoms can no longer position themselves with respect to the QM system as is generally done during the micro-iterations. In this case the QM system should be allowed to move to position itself with respect to the MM system. Vreven *et al.* [Vreven *et al.*, 2003] discuss various methods in how to approach this problem. They suggest three solutions to the problem:

1. The frozen Cartesians can be added to the coordinate system of the QM system and the optimization for the QM system can be performed in Cartesians, which would take very long.
2. The QM system's internal coordinates can be added to the MM system in the micro-iterations and the micro-iterations performed in redundant internal coordinates, which would also be laborious and inefficient.
3. One can convert the internal coordinates of the QM system to Cartesians and add these coordinates, together with the coordinates of the mass center or geometrical center of the QM system in Cartesians, to the MM system. The micro-iterations are then used to position the QM system, treated as a rigid block, with respect to the frozen MM atoms as well as the relaxed MM atoms. The QM system is then later optimized as usual.

Method 3 is used in *Gaussian 03* and it can be represented by Scheme A in Fig. 2.3.

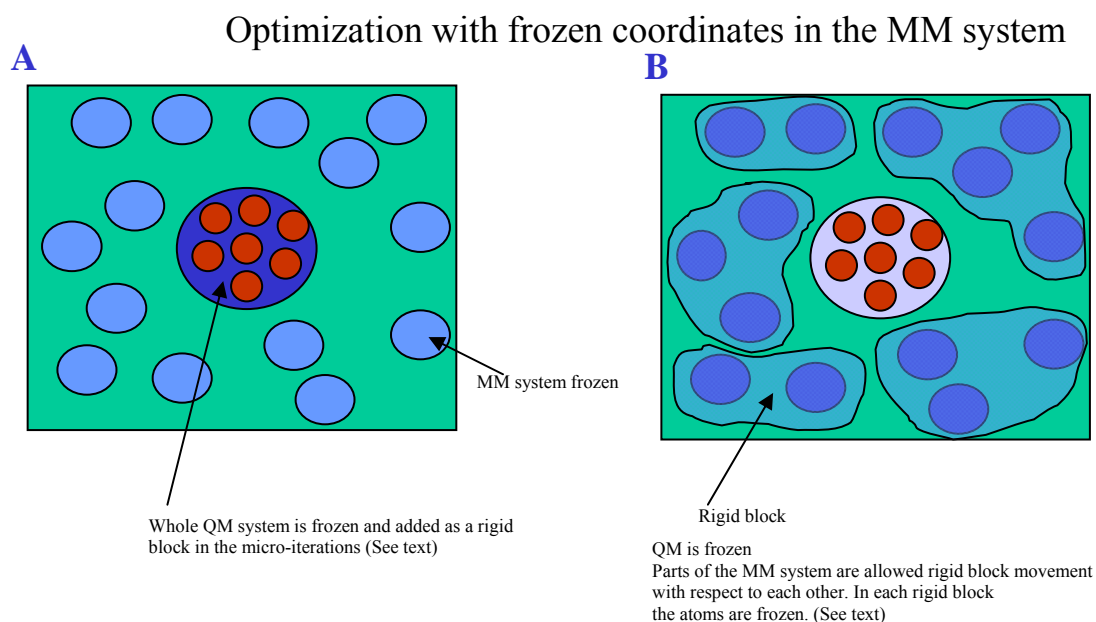


Fig. 2.3: Diagram illustrating two schemes when using frozen coordinates in the MM system.

It is also possible to make selected atoms in the MM system behave as rigid blocks. This is illustrated in Scheme B in Fig. 2.4. However, the larger the number of rigid blocks, the longer the optimization will take. The reason is that mass centers or geometrical centers, need to be added to the Cartesian coordinate system in the micro-iterations for each of these blocks. In an optimization, these blocks will each have to position themselves with respect to the QM system while simultaneously positioning themselves with respect to the other rigid blocks in the MM system. When mechanical embedding is used the optimization is not that difficult, but with electronic embedding, the force on the MM atoms by the QM system will have to be evaluated and updated each time a rigid block moves for all the atoms in the rigid block. The movement of the rigid blocks is usually scaled with respect to their translation and their rotation. The limitations of this method are that the MM rigid blocks are not fully optimized as no internal coordinates are available that would make the MM blocks rotate around their geometrical center or mass center. Therefore, the MM rigid blocks are limited to only basic movements. For our work we had to use rigid block movement, as most of our monomers in our clusters were restricted to the MM system. *Gaussian 03* does not provide a solution for this problem and according to Dr Thom Vreven [Vreven, personal communication], an optimization algorithm will not be available soon, although he said that in the meantime it is reasonable to attempt to do the optimizations by using rigid blocks.

### 2.9.5 Problems at the boundary



When a covalent bond stretch from an atom in the QM system to an atom in the MM system, it may be difficult to model as the QM and the MM systems will have dangling bonds at the boundary between the two systems. The boundary is much better defined in the ONIOM methodology than in other QM/MM methodologies and it is therefore here that ONIOM outperforms standard QM/MM methodologies. Various schemes to solve the boundary problem have been developed for QM/MM systems based on the traditional QM/MM methodology, but they are much more complicated than in the case of the ONIOM methodology.

In ONIOM, the whole system under study is defined in one coordinate system; the coordinate system of the real system, see Fig. 2.2. Therefore using link atoms, *vide infra* to partially solve this problem, is well defined in the real system. For example, for 1-phenylethanone in Fig. 2.4 the real system is defined as the entire molecule and the model system is defined as acetaldehyde. The model system is treated with quantum mechanics and the MM system, which is the remaining part of the real system, with a force field. The link atom is shown surrounded by a red circle in Fig. 2.4 and has coordinates  $\mathbf{r}_2$ . The link atom can be chosen to be any atom that can represent the part of the bond in the QM system, but is usually chosen to be a hydrogen atom or to be isolobal to the substituent in the MM system [Dapprich *et al.*, 1999].

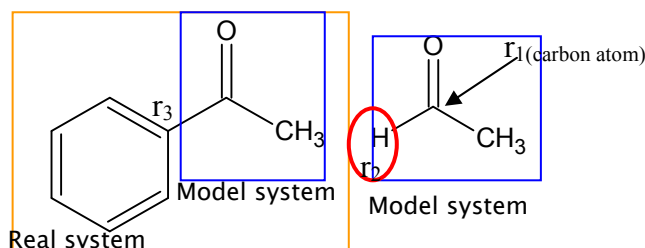


Fig. 2.4: Diagram illustrating the use of a link atom in an ONIOM calculation where a bond is partially QM and partially MM.

In *Gaussian 03* the position of link atoms can be scaled with respect to the distance of the bond between the MM and QM systems. In Fig. 2.4 we give labels to each of the atoms involved to describe the bond between the MM and QM systems. The position of  $\mathbf{r}_2$  can be scaled by varying  $g$  in equation 2.43.

$$\mathbf{r}_2 = \mathbf{r}_1 + g(\mathbf{r}_3 - \mathbf{r}_1) \quad (2.43)$$

Using the ONIOM methodology, all of the atoms in the entire system are defined in terms of the coordinates of the real system, which makes it very simple to determine normal modes of vibrations, as well as infrared (IR) and Raman intensities for the whole system under study. With other QM/MM methodologies such calculations can be very difficult and too time consuming for large systems. The calculation of derivatives for ONIOM will be discussed in **Section 2.9.6**.

The link atom method is not foolproof and for the ONIOM-EE methodology much research is being done to improve the electrostatics of a system containing link atoms. The main problem seems to be the overpolarization of the QM system by the link atom. Truhlar and coworkers have developed new methods to make the treatment of electrostatics in systems using link atoms more accurate [Truhlar *et al.*, 2005]. Other methods used to saturate the boundary in QM/MM systems are the *Generalized hybrid orbital* (GHO) [Pu *et al.*, 2004] method and frozen orbitals [Murphy *et al.*, 2000].

### 2.9.6 ONIOM vs other QM/MM methodologies

Regarding our discussion above, the reader will now appreciate the advantage that ONIOM has over traditional QM/MM methods. As also mentioned earlier, to derive properties for a system treated with ONIOM, even with link atoms, is very simple. The computation of the second derivative or Hessian matrix for the entire system can be written as [Dapprich *et al.*, 1999]:

$$\frac{d^2 E_{tot}}{dq^2} = [J^T \cdot \frac{d^2 E_{model}^{high}}{dq^2} \cdot J] + \frac{d^2 E_{real}^{low}}{dq^2} - [J^T \cdot \frac{d^2 E_{model}^{low}}{dq^2} \cdot J] \quad (2.44)$$

where  $E_{tot}$  is the total ONIOM energy of the real system and  $q$  are the Cartesian coordinates of a specific subsystem such as the model system. In the first bracket,  $J$  ( $J^T$ ) is the Jacobian (transpose of the Jacobian) necessary to project the second derivatives for the link atoms in the model system at the high-level of theory to the Hessian of the model system at the high-level of theory. In the third bracket, the Jacobian and its transpose project the second derivatives of the link atoms in the model system at the low-level of theory to the Hessian of the model system at the low-level of theory. Equation 2.44 then gives a Hessian that has the correct degrees of freedom as the link atoms are defined in the real system [Dapprich *et al.*, 1999].

When the QM and MM systems are not bonded, the Jacobian is a unit matrix. This procedure is implemented in both *Gaussian98* and *Gaussian 03*.

To derive second derivatives for an entire system, not even including link atoms, based on the traditional QM/MM methodologies, can be cumbersome. When the QM method is Hartree-Fock or DFT, the second derivatives can be derived based on the general second derivatives for these energies. To compute these derivatives the coupled perturbed Hartree-Fock (CPHF) or coupled perturbed Kohn Sham equations (CPKS) need to be solved for the entire system including the MM system [Cui and Karplus, 2000] and this complicates the derivation of the Hessian for such systems, especially when the systems are very large. Another advantage of ONIOM or

ONIOM-EE is that it is much more versatile than traditional QM/MM methodologies. For example, all the principles to derive derivatives, NMR chemical shifts etc. are in the ONIOM methodology not only possible for QM/MM systems, but also for QM/QM and QM/QM/MM and QM/QM/QM systems. In a CCSD/MP2 system for example, the interaction between molecules in the CCSD layer and the molecules in the MP2 layer is treated at the MP2 level of theory. We are not certain if ONIOM-EE has been applied for such systems, but using ONIOM with mechanical embedding has been done for a multitude of systems. There are still many improvements necessary before ONIOM and ONIOM-EE can be accepted as standard methods. We will list six:

1. Sometimes the QM wave function is overpolarized by the link atom and, as mentioned, has received attention in the literature [Truhlar *et al.*, 2005].
2. The treatment of the electronic embedding in ONIOM-EE can be made more accurate (see **Chapter 9**).
3. The MM system should not only be able to polarize the QM wave function, but the QM wave function needs to be able to polarize the MM system. This has been achieved based on the traditional QM/MM methodology, either by including induced dipoles [Bakowies and Thiel, 1996] in the QM/MM equation, or by using the *Discrete Reaction Field* (DRF) [Jensen *et al.*, 2003] method. Geometry optimizations are however not feasible for the systems of interest.
4. It is essential to develop better optimization algorithms, to incorporate mixed redundant and Cartesian coordinates in the micro-iterations so that bonds, valence angles and torsion angles can be restrained in the MM system.
5. Polarizable force fields can be used to incorporate the polarization of the molecules in the MM system.
6. As far as we know, molecular dynamics (MD) has not been applied for a system described by the ONIOM methodology. MD has been performed for QM/MM systems based on the traditional QM/MM methodology. See for example the article by Martín *et al.* [Martín *et al.*, 2002].

Despite these problems with the ONIOM and ONIOM-EE methodologies, both methodologies have been applied to solve many problems that would otherwise have been impossible to solve due to the size of the systems. Some applications for ONIOM/ONIOM-EE are:

- ▶ The modeling of enzyme catalysis [Sklenak *et al.*, 2004].
- ▶ The modeling of host-guest systems [Casadesùs *et al.*, 2003].
- ▶ Surface adsorption [Martins *et al.*, 2004].
- ▶ Nanotechnology [Ricca and Drocco, 2002; Froudakis, 2001].
- ▶ Molecular dynamics such as ONIOM-ADMP (ONIOM Atom centered density matrix propagation [Rega *et al.*, 2004].
- ▶ ONIOM(XS) (ONIOM with exchange of solvents) [Kerdcharoen and Morokuma, 2001].
- ▶ ONIOM-PCM (ONIOM with a polarizable continuum model) [Vreven *et al.*, 2001].
- ▶ Organometallic chemistry [Yamanaka and Mikami, 2002].
- ▶ RAFT polymerization [Izgorodina and Coote, 2006].
- ▶ Reactions in zeolites [Sangthong *et al.*, 2005].

As far as we know, the ONIOM methodology has never been used to study gas phase van der Waals clusters as in this work. In fact, the only study that we could find is where HF and other halides solvated in Ar were studied with a hybrid method. For the QM system the researchers used MP2 and for the MM system, Diatomics-in-molecules (DIM) [Bochenkova, 2004]. This hybrid system should be regarded as a QM/SE (quantum mechanical semi-empirical) system rather than a pure QM/MM system.

#### 2.9.7 *The accuracy of hybrid methods*

ONIOM and QM/MM methods are reasonably well developed, but no hybrid calculations and geometry optimizations can be blindly accepted as being accurate and have to be verified against experimental results if available. In the literature, it is also evident that one program's answer might differ from another program's answer, purely because different algorithms are implemented in different software packages [Friesner and Guallar, 2005]. When using methods such as IMOMO [Humbel *et al.*, 1996], the S-value test [Morokuma *et al.*, 2001] should aid in choosing good combinations of methods for the layers.

## ***2.10 Force field parameterization for QM/MM***

As far as we could ascertain, force fields have never been optimized for the ONIOM-EE scheme, since this option only became available in 2003 with the release of *Gaussian 03*, but there are accounts in the literature for force field optimizations utilizing the original QM/MM methodology. In this section we will look at various ways in which force fields can be optimized for QM/MM systems.

To modify a force field, such as UFF, to be used for QM/MM calculations and optimizations, the nonbonded interactions usually receive priority. Only when a brand new force field is developed, should one parameterize torsions, angle bends and bond stretches.

In UFF [Rappé *et al.*, 1992], the electrostatic interaction is a simple Coulomb interaction and therefore there are no parameters that can be optimized. The Coulomb interaction will depend on the charges of the MM and QM atoms. The vdW parameterization will be dependent on the electrostatic interactions.

When optimizing any force field one needs a reference system. One then parameterizes the force field to reproduce data that is known for the reference system, such as its geometry, its thermodynamical properties etc. The data can either be determined by experiment or when experimental data is lacking, by *ab initio* calculations [Leach, 2001].

The vdW equation is nonlinear and therefore methods of solving linear equations cannot be used. It can be extremely time-consuming to optimize a new force field, depending on the method and reference data used [Leach, 2001]. Force field optimization strategies can be divided into two groups: manual and automatic methods. Depending on the method and the criteria used for optimizing the parameters of the force field will determine if geometry optimizations are necessary or not.

### 2.10.1 Manual force field optimization methods

#### 2.10.1.1 Trial and error

The name says it all. This method is extremely time consuming and involves a significant number of human intervention. Parameters for the force field are intelligently guessed or based on previously optimized parameters. If the force field optimization method includes a geometry optimization, the computational chemist can specify these parameters in an input file for the computational software package that can then perform a geometry optimization after which the chemist can change the parameters and perform another geometry optimization. This process is then continued until the geometry is as close as possible to the reference geometry. The trial and error method was one of the methods that we attempted in this work to optimize force field parameters.

The quality of the force field can be assessed by making use of a *root mean square deviation* (RMSD), *vide infra*. The trial and error method might be unsuitable for some systems where geometry optimizations can be laborious.

#### 2.10.1.2 Automatic methods

The easiest way of optimizing force field parameters is by using a force field optimization program to vary the parameters. However, such a program should be able to perform QM/MM geometry optimizations or it should be able to interface with a separate program, such as *Gaussian 03* that can perform the optimizations. The latter is of course the most convenient option. By stepping the force field parameters by a line search and using numerically calculated derivatives, the force field optimization program should be able to optimize the force field parameters without human intervention. Such a program was not available during this work, unfortunately.

The majority of force fields for QM/MM systems are optimized for biological systems. In one account [Riccardi *et al.*, 2004] the hydrogen bond energy and bond length for several peptide dimers were used as a reference system in the optimization of force field parameters by an automatic optimization method. This group made use of a genetic algorithm (GA), *vide infra*, where they gave certain priorities to the fitting of various data. For example, they gave a higher priority to reproducing the hydrogen bond energy than to reproducing the hydrogen bond length.

In our work we also used a GA under certain circumstances, which will be described in the appropriate chapters. Other methods can also be used, such as simulated annealing or general derivative based methods. The advantage of using nonderivative-based methods is that these

methods are designed not to get stuck on a local minimum as is the case with general derivative based methods such as Newton-Raphson or quasi-Newton methods. The nonderivative-based methods can be used to find a global minimum on an optimization surface.

### 2.10.2 Genetic algorithms (GA's)

GA's are based on genetics and the principle of Darwin's survival of the fittest. In GA's, random models are generated and stored as so-called chromosomes. For example, if one wants to optimize 4 vdW parameters, then a set of chromosomes, the number that is specified by the user, will contain "genetic" sequences representing each of these parameters. The random parameters are encoded in binary bits. For example if a parameter is 9 it will be represented by 1001 in binary notation. These chromosomes are then allowed to "*crossover*" meaning that they exchange bits to form hybrid chromosomes that form the next generation. A fitness function is used to give a certain "fitness" to each chromosome in each generation. The fitness function is based on the ability to reproduce the reference model's results. Only the chromosomes with the highest "fitness" in a generation are allowed to "*crossover*" to form the new generation. The user can also allow a certain percentage of mutations and the type of crossover can also be specified. The process is repeated with the new generation and is continued for a certain number of generations that can be set by the user. Eventually as the generations progresses, some models die out and after a certain number of generations, set by the user, the chromosome with the highest "fitness" is given as the output. This is then the set of optimized parameters.

Results obtained with GA's should always be reoptimized with a gradient optimization algorithm [Jensen, 2001]. A GA can only move into a global minimum well, but to step towards the global minimum, gradient optimization algorithms should be used. The other problem with GA's are that they take much longer than gradient optimization methods to find minima and if all the GA's parameters, such as mutation probability, are not specific for a particular optimization under study, then GA's might give worse results than with gradient optimization methods. GA's can find both minima, maxima and can be set to optimize parameters to obtain a specific value such as the energy, for example. All variables can be constrained to certain limits.

### 2.10.3 Force field quality assessment

In order to assess the quality of a force field, statistical methods can be used. One of the simplest, and also the choice in our work, is the root mean square deviation (RMSD), which is defined as:

$$RMSD = \sqrt{\frac{\sum (\chi_{\text{reference}} - \chi_{\text{estimated}})^2}{N}} \quad (2.45)$$

where  $\chi_{\text{reference}}$  is the value for a property of the reference model,  $\chi_{\text{estimated}}$  is the value obtained for the same property after a trial run with the force field and  $N$  is the number of values for a properties being measured. For example, in our work we used nonbonded distances determined through a MP2-optimization as  $\chi_{\text{reference}}$  and measured the nonbonded distances obtained with our force field as  $\chi_{\text{estimated}}$ .

## 2.11 Accounts from the literature for the optimization of Lennard-Jones parameters for QM/MM

There are relatively few reports in the literature regarding the optimization of Lennard-Jones parameters for hybrid methods. Martín and coworkers [Martín *et al.*, 2002] optimized Lennard-Jones parameters for the simulation of water molecules. They made use of an ASEP (Average Solvent Electrostatic Potential) approach for QM/MM/MD (quantum mechanics-molecular mechanics-molecular dynamics) simulations and were able to improve previous results for liquid water obtained with this method.

Murphy and coworkers [Murphy *et al.*, 2000] optimized the Lennard-Jones parameters for the twenty natural amino acids. They managed to improve the QM/MM binding energies to within 0.6 kcal/mol of the desired target, which showed a significant improvement in the binding energies with respect to the default Lennard-Jones parameters, which gave grossly incorrect binding energies.

Lennard-Jones parameters were optimized for the modeling of bimolecular systems representing hypothetical interactions between a QM active site of a protein and a MM protein environment [Freindorf *et al.*, 2005].

Lennard-Jones parameters were developed for the interaction of  $\text{Na}^+$  and  $\text{Cs}^+$  with calix[4]arene-bis-crown-6 (BC6) in a QM/MM molecular dynamics simulation [Golebiowski *et al.*, 2001]. Lennard-Jones parameters were also optimized for the QM/MM simulation of liquid hydrogen fluoride, which was simulated by using a QM/MM molecular dynamics method [Wierzchowski *et al.*, 2003].



## 2.12 Summary and conclusions

As is evident from the discussion in this chapter, the task of eventually optimizing force field parameters for UFF to be used with ONIOM-EE for vdW clusters is challenging. The coin has two sides. The modeling of nonbonded interactions using *ab initio* methods is a challenge on its own. This is evident from the literature, which states that to obtain results comparable to what we know about these systems from experiments, large correlation consistent basis sets and high levels of theory such as CCSD(T) should be used. Modeling hydrogen bonds is simpler and methods to model hydrogen bonds include DFT, which yields good to excellent results when using a large correlation consistent basis set such as the triple zeta aug-cc-pVTZ basis set of Dunning. We also mentioned the importance of using post-HF methods when doing calculations on clusters.

Other than with the optimization of single molecules, the electronic energy for a cluster does not give valuable information and in the literature the binding energy or interaction energy is almost always quoted. We saw that in order to report accurate interaction energies or binding energies it is important to correct for the BSSE. One can use either MBPT theory or SAPT. The main advantage of SAPT is that no BSSE is included in the calculation of the interaction energy. The advantage of the supermolecular theory is its simplicity.

The vdW clusters used in this study are usually studied computationally as very little experimental information, giving insight in their geometries and properties, exists. In **Section 2.5** we mentioned many methods that can be used in modeling vdW clusters. During this study MP2, DFT and QMC were all considered as possible methods for modeling large vdW clusters quickly and accurately.

In this chapter, we also discussed CC theory, which is used as a benchmark for vdW clusters if no experimental information is available. We also briefly mentioned other methods that are used to optimize vdW clusters, such as the localized-MP2 method, which is a promising method in giving accurate answers in a quicker time than conventional MP2. It is however not used by many in the field.

The other side of the coin is the use of a hybrid method to model the clusters. Hybrid methods should be able to decrease the optimization time of larger clusters and even crystals significantly. We reviewed the background information necessary to understand our application of the ONIOM-EE methodology in this work. We also compared the standard and more well known QM/MM methods with the ONIOM methodology of Morokuma and coworkers and discussed the advantages and limitations of utilizing the ONIOM-EE method compared to achievements in the field of QM/MM, for example the mutual polarization of the

MM and QM systems. We briefly mentioned optimization algorithms that can be used to optimize a hybrid system in order to explain the addition of geometry constraints as used in this work. We discussed methods in which force fields can be optimized for application in hybrid methods. Finally, we reviewed a selection of literature regarding optimization of Lennard-Jones parameters for various hybrid systems.

It should be evident from this chapter, that many strides have been made towards developing hybrid methods since 1976, and that we are getting closer and closer by the day to developing quick and robust hybrid methods able to model systems that are too large for general *ab initio* calculations.

## Chapter 3

# Preparations for the molecular modeling of clusters for *ab initio* calculations and ONIOM

### *3.1 Introduction*

In order to do hybrid calculations of van der Waals clusters, various preparations have to be done and various decisions have to be made. The quality of the results will depend on these preparations and decisions. In this chapter we will concentrate on these.

Usually a variety of methods and different basis sets can be used to model different molecular clusters. Therefore, to model hydrogen fluoride clusters, a DFT functional would be more efficient, for example, in terms of the time needed for optimizations than post-HF *ab initio* methods. For the modeling of carbon monoxide clusters, however, standard DFT cannot be used due to its well-known inability to model dispersion interactions, and we are left with *ab initio* methods. In addition, hydrogen fluoride clusters can be modeled with smaller basis sets than in the case of carbon monoxide clusters, because the correlation energy does not play such an important role in the interaction energy of hydrogen fluoride clusters than in carbon monoxide clusters. When developing a force field for a hybrid system that should be able to model both these systems as accurately as possible, one needs to decide on a suitable method and basis set that is applicable for the modeling of both HF and CO clusters.

The chapter is divided as follows: In **Section 3.2** we will give a list of the software used in this work. In **Section 3.3** and **Section 3.4** we will illustrate the strategy used to find a suitable method and basis set. In **Section 3.5** we will discuss the design of the hybrid systems in this work. **Section 3.6** will conclude this chapter with a general summary and conclusions.

### *3.2 Software used in this work*

In this work we made use of a variety of software. Instead of specifying the software used for each application, we thought it more appropriate to list them in this section together with their references. Therefore, at any time, if the reader is interested in knowing how certain

calculations were performed, he/she can consult this section. In the following chapters we will not give the references to the software again in order to make the text simpler to read.

### 3.2.1 *Ab initio calculations, geometry optimizations, and vibrational analysis*

*Gaussian 03* [Frisch *et al.*, 2003] and *Gaussian98* [Frisch *et al.*, 1998] are well-known commercial computational chemistry packages able to do mainly *ab initio* and DFT types of calculations. *Gaussian98* is also able to do ONIOM with mechanical embedding. *Gaussian98* was mainly used in this work to optimize some of the smaller hydrogen fluoride and carbon monoxide clusters.

*Gaussian 03* is the newest version of the Gaussian software and is able to do ONIOM-EE. It also allows the user to modify force field parameters in the input. It has three built-in force fields, namely UFF [Rappè *et al.*, 1992], AMBER [Cornell *et al.*, 1995] and Dreiding [Mayo *et al.*, 1990]. We used UFF due to its general applicability to molecules and molecular systems. Dreiding could also have been used, but it is not as versatile as UFF has atom-types for almost all the atoms on the periodic table. AMBER is a force field for the modeling of large biological systems such as proteins.

*Gaussian 03* was used for the optimization of the larger clusters and for ONIOM-EE calculations. It is also possible to do force field single point calculations and optimizations with this software. In some cases, *Gaussian 03* was therefore also used for general force field optimizations. *Gaussian 03* also allows the user to estimate the BSSE of a calculation as determined by the counterpoise correction [Boys and Bernardi, 1970]. For clusters containing many monomers, such as in this work, this feature is very helpful.

### 3.2.2 *Visualization of results*

In order to visualize optimized molecules and animate optimizations and normal modes, Gaussview [Dennington II *et al.*, 2003] was used. Molden 4.0 [Schaaftenaar and Noordik, 2000] was also used to visualize some of the optimizations and normal modes of vibrations. Gaussview was also used to visualize molecular orbitals and plots of the electrostatic potential (ESP) mapped onto the electron density.

### 3.2.3 Analysis of wave functions

#### 3.2.3.1 Atoms in Molecules (AIM)

For the *Atoms in Molecules* method, used in some cases in this work, AIM-PAC 95<sup>10</sup> was used. This set of programs made it possible to calculate atomic charges, draw contours of the electron density and second derivatives of the electron density, determine gradient paths of the electron density and determine critical points and their properties.

#### 3.2.3.2 Natural Bond Orbital (NBO) analysis

NBO 3.0 [Glendening *et al.*, 2001] was used in the analysis of the wave functions of selected clusters and to obtain atomic charges for comparison to charges derived by other charge derivation schemes.

#### 3.2.3.3 Force field geometry optimizations

Cerius<sup>2</sup> [MSI, 1998] was used to do basic force field based geometry optimizations and general calculations. It gives a decomposition of the steric energy, whereas *Gaussian 03* does not. Cerius<sup>2</sup> was also used to assess the force field results obtained with *Gaussian 03*.

#### 3.2.3.4 Graphical representation

Images in this work were generated by Gaussview, Molekel 4.3 [Flükiger *et al.*, 2000] and WebLab ViewerLite 4 [MSI, 2000].

## 3.3 Choice of method

In order to obtain accurate geometries for the clusters, it was necessary to use post-HF methods. Our first choice was Møller-Plesset perturbation theory with energy corrections to the second order (MP2) as it is known as a method that incorporates electron correlation in an affordable way. However, it is known that DFT can be used to optimize systems in times comparable to that taken for HF theory optimizations while incorporating electron correlation. We therefore investigated the use of DFT for weakly bonded clusters such as CO clusters and found one outstanding account in the literature where functionals were designed that gave

---

<sup>10</sup> [www.chemistry.mcmaster.ca/aimpac/](http://www.chemistry.mcmaster.ca/aimpac/)

very good results for rare-gas clusters [Sato *et al.*, 2005]. However, in investigating the possibility of using these functionals in our work, rather than MP2, we learned that since they are brand new, they have not been incorporated into *Gaussian 03*.

After reading about Quantum Monte Carlo in an article on the modeling of large hydrogen clusters [Carmichael *et al.*, 2004], we investigated this method and learned that it is a faster method than CCSD, and can generate correlated wave functions of extremely high accuracy. Furthermore, no Slater determinants need to be stored on disk, making the hardware requirements much less severe than for *ab initio* methods. The other advantage is that it does not use basis sets, so BSSE and other basis set problems are absent with this method. To perform QMC however, one needs a dedicated parallel architecture of sufficient size, and this was not available at the time of this work.

On this basis we eventually opted for MP2. It was therefore necessary to choose a suitable basis set for the calculations.

### ***3.4 Choice of basis set***

In order to decide which basis set to use, certain criteria were set:

1. MP2 optimizations using the basis set had to be able to reproduce experimentally determined geometries, which had to be energy minima. For CO clusters, for example, it is currently accepted that there are two stable dimers and both are slipped antiparallel. The C-bonded one is the global energy minimum dimer [Surin *et al.*, 2003]. It was also important that the values for the bond lengths of the monomers for the energy minima structures had to be reasonably close to experiment.
2. The basis set had to have a minimum number of BSSE, as predicted by the counterpoise method [Boys and Bernardi, 1970].
3. The basis set had to perform well for both hydrogen fluoride and carbon monoxide clusters. We assumed that if this were the case, the basis set would also be suitable for CO/HF clusters.
4. The basis set size had to be suitable for the calculations that had to be done in the time frame for this project and suitable for the available hardware at our disposal.

Priority was given to criteria 1, 3 and 4, as criterion 2 is usually difficult to accomplish unless extremely large correlation consistent basis sets are used. In order to incorporate as much of the electron correlation as possible, polarization functions are recommended [Jensen, 2001]. After priorities 1, 3, 4 and possibly 2 were satisfied, we compared the bond lengths of an

optimized monomer in the gas phase with the experimental values. All the optimizations were done under very tight optimization conditions. For the HF dimer and CO dimers, no symmetry constraints were placed on the optimization. All the dimers were evaluated for being energy minima by doing a vibrational analysis and confirming that no imaginary frequencies were found.

Table 3.1 gives a summary of the properties of the HF dimer obtained by using a variety of basis sets. For the definition of the geometrical parameters used in Table 3.1, the reader is referred to Fig. 3.1.

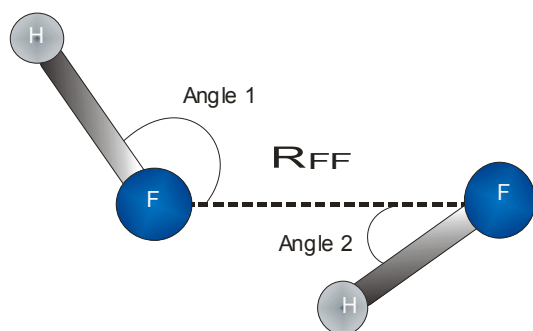


Figure 3.1 : Definitions for the geometrical parameters used in Table 3.1.  $R_{FF}$  is the distance between the two fluorine atoms.

Table 3.1: Results for the optimization of the HF dimer with MP2 utilizing a variety of basis sets. See Fig. 3.1 for the definition of the geometrical parameters.

Basis set	Interaction energy with BSSE-correction (kcal/mol)	Interaction energy without BSSE -correction (kcal/mol)	BSSE (kcal/mol)	%BSSE	Angle 1 (°)	Angle 2 (°)	$R_{FF}$ (Å)
6-31++G(d)	-4.4	-5.6	1.2	21.4	114.4	7.0	2.76
6-31+G(d,p)	-4.2	-5.0	0.7	14.9	115.5	7.4	2.78
6-31++G(d,p)	-4.2	-5.0	0.8	15.0	115.1	7.7	2.78
6-311G(d)	-4.5	-7.1	2.6	36.8	100.2	17.7	2.68
6-311G(d,p)	-3.8	-6.3	2.5	39.4	98.3	15.5	2.71
6-311+G(d,p)	-3.8	-4.7	0.9	19.8	121.3	6.1	2.79
6-311++G(d,p)	-3.8	-4.8	1.0	20.3	121.6	6.0	2.79
Aug-cc-pVDZ	-4.0	-4.7	0.7	14.4	110.2	6.5	2.75
6-31++G(2d,2p)	-4.1	-4.9	0.8	16.1	111.3	5.8	2.73
Experimental [Howard <i>et al.</i> , 1984]					117±6	10±6	2.72±0.03

From Table 3.1 we can make a number of observations:

1. The BSSE corrected interaction energy of the split valence (SV) basis sets seems strongly dependent on the number of Gaussian primitives (GTO's) used to describe the valence shell. The Dunning correlation consistent double

zeta (DZ) basis set [Dunning *et al.*, 1993; Kendall *et al.*, 1992; Peterson *et al.*, 1994] with augmented diffuse functions, Aug-cc-pVDZ, gives a smaller BSSE corrected interaction energy than the SV basis sets. This is because this basis set is much larger than the SV basis sets. One must however still remember that it is only DZ.

2. With the addition of *s* and *p* diffuse functions on the fluorine atom, the BSSE decreases. When diffuse functions are added to the hydrogen atoms however, the BSSE increases slightly in value. Also note that the values for the  $R_{FF}$  distance and the two angles, listed in Table 3.1, get closer to the experimental values.
3. From the experimental results in Table 3.1 it is evident that it is important to add higher order angular momentum polarization functions to both the fluorine and hydrogen atoms. It seems that the distance between the monomers gets closer to the experimental results if more polarization functions on the fluorine and hydrogen atoms are added to a SV basis set. This is probably a sign that correlation energy plays a large role in determining the experimental geometry of the HF dimer.

Based on our results, the following basis sets were accepted as suitable to model the hydrogen fluoride dimer: 6-31++G(d,p), 6-31++G(2d,2p), 6-311+G(d,p), aug-cc-pVDZ, 6-311++G(d,p).

Nevertheless, it is well known that post-HF methods other than DFT are expensive for large clusters. The augmented DZ basis set of Dunning<sup>11</sup> as well as the 6-311++G(d,p) basis set were therefore discarded and only used as a standard for comparing to the results obtained with the smaller basis sets. In any case, the 6-311++G(d,p) basis set also did not improve much on the percentage BSSE and the interaction energy of the 6-311+G(d,p) basis set. From these results we conclude that hydrogen does not play such a significant role in binding the two monomers together, and that fluorine is the major player.

The next step was to find a suitable basis set from our selected set, *vide supra*, to model the CO clusters. As we mentioned earlier, according to the literature there are two antiparallel dimers for carbon monoxide. It is accepted that the C-bonded dimer is the global minimum and the O-bonded dimer the local minimum. The O-bonded dimer is expected to have a larger BSSE than the C-bonded dimer because oxygen has more valence electrons than carbon and

---

<sup>11</sup> In *Gaussian 03* the Dunning correlation consistent basis sets have had redundant functions removed and have been rotated [Davidson *et al.*, 1996] in order to make calculations faster.



the basis set should therefore be less suitable for the description of the valence shell in oxygen than in carbon, leading to a higher BSSE.

Table 3.2: Table of the energies and percentage BSSE for the optimized CO dimer. See text.

Basis set	Energy in hartree (not BSSE corrected)	BSSE (kcal/mol)	Interaction energy with BSSE correction (kcal/mol)	Interaction energy without BSSE correction (kcal/mol)	%BSSE	Distance between mass centers of monomers (Å)
6-31++G(d,p)	-226.059160	0.4	-0.2	-0.6	71.7	3.72
6-311+G(d,p)	-226.157205	0.4	-0.1	-0.5	78.7	3.64
6-31++G(2d,2p)	-226.125357	0.3	-0.3	-0.6	51.2	3.70
Experiment [Surin <i>et al.</i> , 2003]						4.0

The first observation from Table 3.2 is that the BSSE is alarmingly high for the 6-311+G(d,p) basis set. However, the size of the BSSE estimated with the counterpoise method is only an upper limit. It is possible that the true BSSE is less than this approximation. BSSE should also not be used as a measure of the error in the interaction energy [Chalasiński and Szcześniak, 1994]. The only way to determine the correct number of BSSE would be to do the same calculations with SAPT. This however, is beyond the scope of this study.

As the valence-TZ basis set used is designed to give a better representation of the valence shell, than valence-DZ basis sets, this basis set was not discarded. We eliminated the 6-31++G(d,p) basis set as it gave a T-shaped O-bonded local energy minimum instead of the accepted antiparallel local energy minimum. Since the 6-31++G(2d,2p) basis set gave the correct local energy minimum, basis sets that were therefore chosen were 6-31++G(2d,2p) and 6-311+G(d,p).

In a final elimination round, we tested these two basis sets by optimizing the C-bonded dimer with each of these basis sets to yield local energy minima. We found the 6-31++(2d,2p) basis set to be inadequate as it gave a T-shaped structure as a local energy minimum. The 6-311+G(d,p) basis set, on the other hand gave a C-bonded slipped antiparallel structure as a local energy minimum with a distance between its monomers' mass centers of 4.37 Å. This is in very good agreement with the experimental value of 4.4 Å [Surin *et al.*, 2003]. Based on our results we have to conclude that 6-311+G(d,p) is the best for our application amongst the basis sets tested. It satisfies 3 of the 4 criteria set previously. However, it was necessary to do one more test. Although the 6-31++G(2d,2p) basis set gave a T-shaped C-bonded dimer, it might also have been that the antiparallel geometry might just be another local minimum on

the PES and that the optimization just did not converge to this minimum. In order to obtain certainty about this, the final optimized geometry obtained with the 6-311+G(d,p) basis set was used as the starting conformation for a single point calculation of the normal modes of vibration with the 6-31++G(2d,2p) basis set. This calculation confirmed that the antiparallel dimer is a first order transition state, rather than a minimum, with the 6-31++G(2d,2p) basis set.

In the literature we also found accounts where the 6-311+G(d,p) basis set was used with MP2 for both the modeling of carbon monoxide dimers [Han and Kim, 1997] and hydrogen fluoride clusters [Chandler *et al.*, 1995] as well as for CO/HF mixed clusters [Seung-Hoon, 1998].

We are reasonably sure that the reason for the poor performance of some of the tested basis sets is BSSE, however for us it is more important to reproduce experimentally determined geometries than to have a low estimate for the BSSE. At one stage during our research we did do a few optimizations of the dimers on a counterpoise corrected PES with the 6-31G(d) basis set. The geometries changed significantly from their original 6-31G(d) optimized geometries during the optimizations, proving that when modeling these clusters, it is important to use adequate basis sets.

As we mentioned earlier, we did do optimizations of both the HF and CO isolated monomers in the gas phase and the values for their calculated bond lengths agree well with the literature values. See Table 3.3.

Table 3.3: Comparison between bond lengths determined with MP2/6-311+G(d,p) and the experimental bond lengths of isolated CO and HF monomers.

	MP2/6-311+G(d,p)	Experimental
Bond length CO (Å)	1.1400	1.1280 <sup>(a)</sup>
Bond length HF (Å)	0.9167	0.9168 <sup>(a)</sup>

a) Huber and Herzberg, 1979

### 3.5 General design of the QM/MM hybrid systems

Other than in one case in **Chapter 5 (Section 5.3)**, all the hybrid QM/MM systems in this work were designed in a specific manner. Atoms 1 and 2 were always chosen as the monomer in the QM system and the rest of the monomers were confined to the MM system. The molecule in the QM system was the only one allowed to change its bond length. Each monomer in the MM system was allowed to move as a rigid block with respect to the monomer in the QM system and the other monomers in the MM system. In this way it was possible to freeze all the bond lengths of the MM molecules.

The high-level method for the QM system was MP2 with the 6-311+G(d,p) basis set. The low-level method for the MM system was UFF [Rappé *et al.*, 1992]. The QM system was allowed to be polarized by point charges on the MM molecules as according to the ONIOM-EE methodology, see **Section 2.9**. Charges for both the QM and MM atoms were given in the input.<sup>12</sup> All the atomic point charges used in the input files for *Gaussian 03* were determined from single point calculations of the MP2 optimized geometries of the clusters. Although in some cases, other charge-schemes were also used to derive point charges, Merz-Kollman-Singh (MKS) charges based on the MP2 density were mainly used for the charges on the QM and MM atoms in the ONIOM-EE calculations. The convergence criterion for the energy of the self-consistent field (SCF) was set to  $1 \times 10^{-12}$  for the single point calculations that were used to derive the point charges.

*Gaussian 03* allows one to choose between two types of atomic point charges for the charges that are derived by the software for the QM system during the ONIOM-EE calculation. These charges are Mulliken and MKS charges. We used MKS charges, which are also the default for ONIOM-EE calculations. For all MKS charges the default radii for charge fitting were used. MKS charges based on the MP2 electron density are used by default when MP2 is the high-level method. The MKS charges could not be constrained to the dipole moment during the ONIOM-EE calculations.

The internal set of UFF parameters was used, but the vdW parameters were given in the input for each force field type. Leaving the setting of the parameters to the program did not yield the same answers as when it was explicitly specified in the input.<sup>13</sup> All ONIOM-EE

---

<sup>12</sup> It is not necessary to put any charges on the QM atoms. These charges are only used in the MM parameter definition and eventually will play no role in the energy calculation or the optimizations, due to the way in which ONIOM-EE is designed (see **Chapter 2**).

<sup>13</sup> These discrepancies are related to the ONIOM-EE method as implemented in *Gaussian 03*. We confirmed this by doing general MM single point calculations with *Gaussian 03* with and without explicitly setting the van der Waals parameters in this regard and obtained the expected answers for the force field energy. When doing the same for an ONIOM-EE calculation, different answers were obtained for the force field energy in each of the two above scenarios. It was not possible to obtain helpful assistance in this regard when a question was posted on CCL.

optimizations were performed with the micro-iteration scheme for ONIOM-EE. No symmetry constraints were used on the ONIOM-EE optimizations. This was necessary, as the point group of the model system in each case was not the same as the point group of the real system.<sup>14</sup> At one stage, we considered using periodic boundary conditions (PBC) similarly to Treusekol and coworkers [Treusekol *et al.*, 2001] in order to simplify the calculations of the hybrid systems, but *Gaussian 03* does not allow PBC to be used with ONIOM. PBC has been used in QM/MM molecular dynamics simulations of liquid hydrogen fluoride before [Muñoz-Losa *et al.*, 2003]. Cartesian coordinates were used throughout for all ONIOM-EE optimizations and calculations.<sup>15</sup> Great care was taken to make sure that these Cartesian coordinates were the same as the coordinates for the MP2 optimized clusters.<sup>16</sup>

### 3.6 Conclusion and summary

We started this chapter by giving an overview of all the software that was used in our study. In this chapter we also reported our strategy in choosing a method for the modeling of the hydrogen fluoride, carbon monoxide and mixed hydrogen fluoride and carbon monoxide clusters, and gave reasons for deciding on MP2. In addition, we reported the strategy in finding a suitable basis set. This resulted in deciding on the 6-311+G(d,p) basis set. We then reported the setup of the hybrid systems and software related problems that had to be taken care of. In order to make our work reproducible we kept to some simple rules, such as specifying the van der Waals parameters explicitly in the input and using the exact coordinates of the MP2 optimized geometries for all our ONIOM-EE calculations and geometry optimizations.

---

<sup>14</sup> *Gaussian 03* does a symmetry check and switches off the symmetry constraints automatically if the point group of the model system does not conform to the real system's point group.

<sup>15</sup> When working with a two atom QM system in *Gaussian 03* it is not possible to give the input as a Z-matrix, as *Gaussian 03* automatically assumes that the first three atoms will be in the QM system.

<sup>16</sup> During the process of ONIOM-EE optimizations, it seemed that if the starting coordinates differed at the fifth decimal number, slightly different answers were obtained. We are not sure if this is a problem in the software or is related to the fact that we used rigid body motion for all the MM monomers in the micro-iterations. More information on this and other Gaussian-related problems can be found in **Appendix B**.

## Chapter 4

# *Ab initio* calculations and analysis of hydrogen fluoride clusters

### 4.1 Introduction

The aim of this chapter is to report the optimization and analysis of several energy minima of hydrogen fluoride clusters. Most of the data reported here is used to setup a QM/MM hybrid system. This hybrid system will be described in detail in **Chapter 5**.

In order to optimize the clusters, the second order Møller-Plesset perturbation theory (MP2) is utilized. The optimization of the clusters to their energy minima with this method will be described in **Section 4.2**. The optimized geometries are then analyzed in terms of their structural properties (**Section 4.3**) and energies (**Section 4.4**) and compared to results found in the literature, where applicable. In the next part of this chapter, the bonding of the clusters will be analyzed (**Section 4.5**) and in **Section 4.6** the electrostatic properties based on the ESP will be analyzed in terms of their Electrostatic Potential (ESP) and atomic point charges will be determined based on this ESP. These charges will be used to represent the charge distribution of the molecular mechanical system (MM) in the QM/MM calculations and optimizations. A new way of looking at hydrogen bonding, described in this chapter, might aid the development of better hybrid methods. The chapter will, as usual, be concluded with the most important results, conclusions and the scope for future work in **Section 4.7**.

### 4.2. Obtaining local minima on the Potential Energy Surface (PES)

#### 4.2.1 General computational details

*Gaussian98* and *Gaussian 03* were used to optimize HF clusters at the MP2 level of theory using both the 6-31G(d) and 6-311+G(d,p) basis sets. All the geometries of the HF clusters were optimized by using a modified version of the original Berny optimization algorithm [Shlegel, 1982] as incorporated in the Gaussian software. In difficult geometry optimization cases, the *Geometry Directed Inversion in the Iterative Subspace* (GDIIS) optimization algorithm was also utilized. The optimized structures obtained with the 6-31G(d) basis set were used as starting points for final geometry optimizations of the clusters with the larger 6-

311+G(d,p) basis set. This was done to ensure that a slow convergence in energy, with respect to each optimization step, was avoided. The MP2 series is notorious for slow convergence when the reference wave function differs too much from the correct optimized MP2 wave function for a specific set of nuclear coordinates [Jensen, 2001].

#### 4.2.2 Pre-optimization of clusters with the 6-31G(d) basis set

The starting points for the pre-optimization of the HF clusters were obtained from an article by Hodges and coworkers [Hodges *et al.*, 1998]. Twelve energy minima<sup>17</sup> were chosen and reoptimized with the 6-31G(d)<sup>18</sup> basis set under very tight optimization conditions<sup>19</sup> as specified in the software. GDIIS was used for the tetramers, pentamers and hexamers. See **Appendix A.1**. Geometry optimizations were done by constraining the symmetry of each cluster during the calculations to its lowest symmetry point group. No further constraints on the geometries of the clusters were applied. Energy minima were confirmed by a lack of imaginary frequencies in the vibrational analysis<sup>20</sup>. If imaginary frequencies were found, the eigenvector of the specific mode of vibration was added or subtracted from the Cartesian coordinates of the transition state and the cluster reoptimized.

Transition state structures were obtained in the case of the HF trimer. The linear trimer (see **Appendix A.1**) converted to a chain structure and after reoptimization, where force constants were explicitly recalculated at each optimization step, to a cyclic structure. In all other cases, the very tight optimization conditions were effective enough to eliminate transition states. In some cases, such as the trimer discussed above, the geometries converged to indistinguishable structures, reducing the original number of clusters to nine. These nine clusters, consisting of one dimer, one trimer, two tetramers, two pentamers and three hexamers were used as the starting points for the optimization with the larger 6-311+G(d,p) basis set.

---

<sup>17</sup> The starting geometries can be found in **Appendix A.1**

<sup>18</sup> The 6-31G(d) basis set usually makes use of six *d* Cartesian Gaussian primitives (See **Section 2.3**) instead of five. The consequence of this “error” is an addition of an extra *s*-type function to the basis set. [Jensen, 2001] This error is not of much concern.

<sup>19</sup> Very tight optimization conditions aim to keep the net force on a molecule to 0.000002 hartrees/bohr or less.

<sup>20</sup> The normal modes of vibration are derived from a mass weighted Cartesian Hessian matrix. An imaginary vibration occurs when the eigenvalue of this specific vibrational mode is negative, indicating a maximum for the energy of this mode.

#### 4.2.3 Final optimization of clusters

The same optimization conditions were used as with the 6-31G(d) basis set. Symmetry constraints were used for all clusters during the optimizations. Only for the HF dimer was it necessary to remove symmetry constraints on the cluster during the optimization in order to achieve convergence. Two of the pre-optimized hexamers converged to distinct structures and eventually eight distinct minima were obtained, as confirmed by the lack of imaginary frequencies in their vibrational analysis. These structures can be found in **Appendix A.1**.

### 4.3 Structural properties

The optimized structures for the 6-311+G(d,p) basis set are shown in Fig. 4.2. Some of the structural parameters for each cluster are summarized in Table 4.1. The angles mentioned in Table 4.1 refer to the angles defined in Fig. 4.1. We will refer to these angles again when discussing the geometries of the clusters.

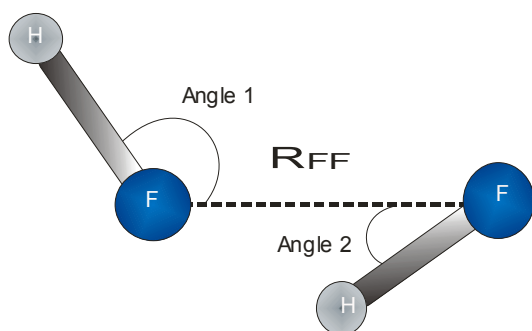


Fig. 4.1: Diagram defining Angle 1 and 2 and  $R_{FF}$  as used hereafter. (This diagram was given in **Chapter 3**, but is given here for convenience.)

Table 4.1: Structural properties of the HF-clusters optimized with 6-311+G(d,p). Bonds in Ångström and angles in degrees. The experimental gas phase value of the HF monomer is 0.9168 Å [Collins *et al.*, 1995].

	I	II	III	IV	V	VI	VII	VIII
MP2/6-311G+(d,p)								
F-H	F1-H2 : 0.921 F3-H4 : 0.920	All: 0.929	All: 0.937	F1-H2 : 0.925 F3-H4 : 0.935 F5-H6 : 0.930 F8-H7 : 0.921	All: 0.941	F1-H2 : 0.932 F3-H4 : 0.946 F5-H6 : 0.939 F8-H7 : 0.936 H9-F10 : 0.921	All:0.942	F1-H2 : 0.942 F3-H4 : 0.942 F5-H6 : 0.942 F7-H8 : 0.942 F9-H10 : 0.942 H12-F11 : 0.918
R <sub>FF</sub>	All:2.788	All: 2.668	All: 2.583	F1-F3: 2.803 F1-F5: 2.667 F3-F5: 2.593 F8-F3: 2.756	All: 2.548	F1-F8:2.600 F8-F5:2.565 F5-F3:2.519 F3-F1:2.653 F10-F3:2.772	All:2.535	All in ring: 2.547 F11-F7: 3.013
FH...F	All:1.874	All:1.865	All: 1.681	F1-H6: 1.861 F5-H4: 1.730 H2-F3: 2.135	All: 1.616	F3-H2:1.777 H4-F5:1.595 H6-F8:1.659 H7-F1:1.704 H9-F3:1.852	All:1.595	All in ring: 1.612
Angle 1	All:121.3	All:84.8	All: 102.6	F1F3H4: 77.3 F3F5H6: 88.8 F5F1H2: 93.0 F8F3H4: 127.6	All: 114.4	F1F3H4:99.7 F3F5H6:104.0 F5F8H7:103.7 F8F1H2:104.0	All:122.6	All in ring: 112.7
Angle 2	All:6.1	All:24.8	All: 12.6	F1F3H2: 36.5 F5F3H4: 18.3 F1F5H6: 24.4	All: 6.4	F5F3H4:9.6 F8F5H6:12.3 F1F8H7:13.4 F3F1H2:16.1	All:2.6	All in ring: 5.5

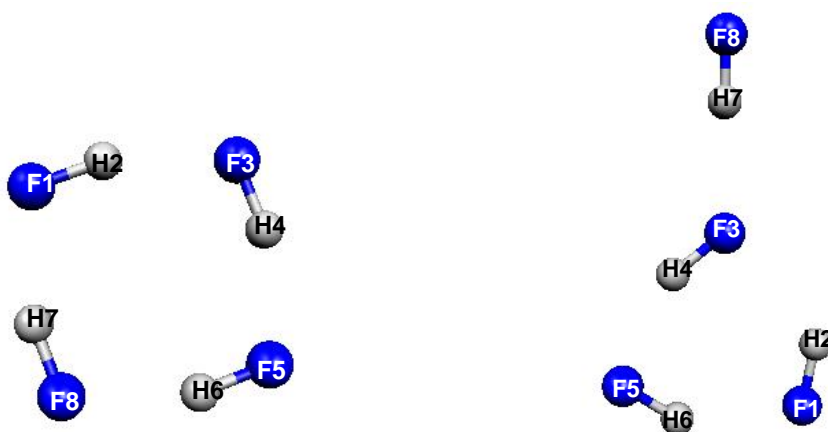


MPW1PW91 /aug-cc-pVDZ [Guedes <i>et al.</i> , 2003]								
F-H	0.927	0.943	0.958		0.963		0.964	
F-F	2.723	2.561	2.487		2.459		2.459	
FH...F	1.804	1.712	1.548		1.548		1.482	
Experimental [Collins <i>et al.</i> , 1995]								
F-F	2.72±0.03							
Angle 1	117±6							
Angle 2	10±6							



Cluster I:  $C_s$

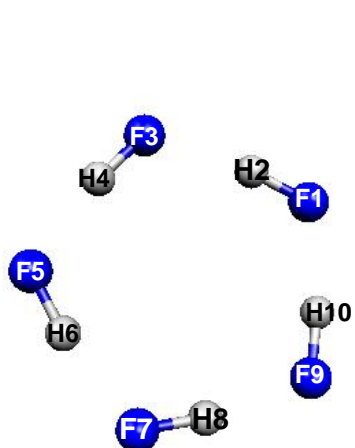
Cluster II:  $C_{3h}$



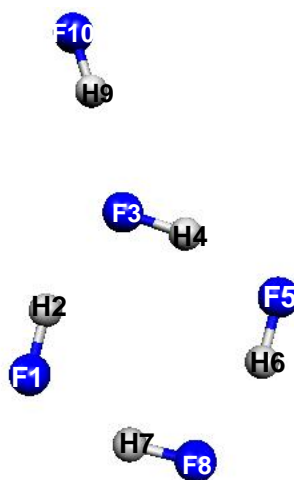
Cluster III:  $C_{4h}$

Cluster IV:  $C_s$

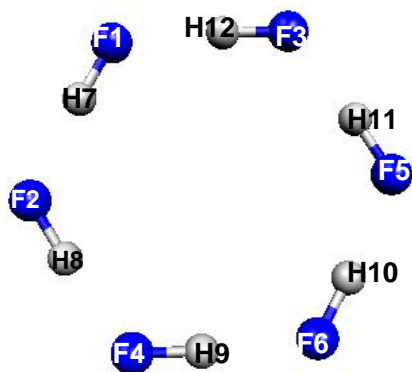
Fig. 4.2 Continued...



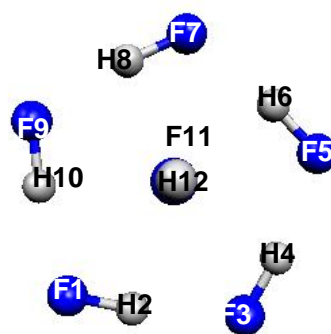
Cluster V:  $C_{5h}$



Cluster VI:  $C_s$



Cluster VII:  $C_{6h}$



Cluster VIII:  $C_s$  almost  $C_5$  (very slightly out of plane)

Fig. 4.2: Optimized geometries of  $(HF)_{2-6}$  clusters with MP2/6-311+G(d,p). The symmetry point group is also shown.

HF vapor contains many clusters of different sizes, but the stable clusters larger than the dimers such as clusters II, III, V and VII, in Fig. 4.2, are well-established by experiment to be cyclic and planar [Quack and Suhm, 1997]. The less stable clusters are usually not stable enough under the experimental conditions to be detected; however, by growing the clusters in helium nano-droplets, metastable species can be formed and detected by infrared laser spectroscopy. Cluster IV was recently detected in this way, but cluster VI could not be stabilized in helium nano-droplets [Douberly and Miller, 2003], although it has been calculated in various *ab initio* studies before [Abu-Awwad, 2002; Douberly and Miller, 2003; Suhm, 1995]. Cluster VIII, to the best of our knowledge, has not yet been found experimentally, and computationally there is only one account where it was optimized with Hartree-Fock with the 6-31G basis set, which is highly inferior to MP2 theory (see **Chapter 2**), and no detailed account was given for this cluster [Abu-Awwad, 2002].

Based on the available literature and experimental results, we conclude that our structural results, reported in Table 4.1, are very good. For clusters such as IV and VI, where one monomer interacts with a cyclic cluster, or where the monomer hovers above a cyclic cluster, such as in cluster VIII, the bond lengths differ significantly from each other. In cluster IV, for example, the bond length of the external monomer is identical to a monomer in the HF dimer (Cluster I), whereas the other bond lengths vary between 0.925 and 0.935 Å. The same type of behavior is found in cluster VI, but in cluster VIII the bond length of the external monomer is convincingly shorter than the bond length of the monomer in the HF dimer and borders on the bond length of the HF monomer of 0.9167 Å<sup>1</sup>. Another reason why Cluster VIII is different from the other clusters is that the bond lengths of all the monomers in the ring of this cluster are almost identical to the bond lengths of the cyclic pentamer (cluster V). This is evidence that the interaction between the external monomer and each monomer in the ring in this cluster are of a weak vdW nature. It is also important to mention that in the original pre-optimized cluster (see **Appendix A.2**) the monomer was lying horizontally above the cyclic cluster and therefore had a chance to form a hydrogen bond with one of the fluorine atoms in the ring. Instead, it deliberately “chose” to form several weak interactions by positioning the fluorine atom of the external monomer in such a way that its lone pairs are available to the ring. As we discussed previously (see **Section 4.2**) we used very tight optimization conditions and therefore we expect that this cluster is the closest local minimum to the starting geometry (see **Appendix A.2**) on the PES.

Lastly, it is worth mentioning that Chandler and coworkers [Chandler *et al.*, 1995] obtained the same geometries for the cyclic HF clusters (II, III, V and VII) as we did when using exactly the same level of theory and basis set. They also optimized the dimer, but our

---

<sup>1</sup> This value was obtained at the MP2 level of theory with the 6-311+G(d,p) basis set. See **Chapter 3** (Table 3.3). The experimental value is 0.9168 Å.

results do not agree with their hydrogen bond length of 1.827Å and it is unclear how this value was obtained. We confirmed that except for this hydrogen bond length, all their structural parameters are identical to ours. We should note that the dimer was recalculated by Seung-Hoon [Seung-Hoon, 1998] with the same basis set and method and their results agree with ours.

#### ***4.4 Electronic interaction and binding energies***

As the definition of binding and interaction energies were already discussed in **Section 2.1**, we will only briefly explain how these parameters are defined.

The binding energy is defined as the energy that is gained by distorting monomers from their gas phase bond lengths and associating them with other monomers to form a cluster. The interaction energy is the sum of all the many-body energies in the cluster and does not involve the destabilization of stretching the bond lengths from the monomers' gas phase values. The difference between the binding energy and the interaction energy is called the one-body energy [Chalasiński and Szcześniak, 2000] as the two-body, three-body etc. energies are incorporated in the interaction energy. In Table 4.2 we summarize the energies and the frequencies of the stretching vibrational modes as obtained for each HF cluster.

Table 4.2: Energies (kcal/mol) and selected vibrations (cm<sup>-1</sup>) of the clusters as calculated with MP2/6-311+G(d,p). Zero-point vibration energy\* corrections were not applied. All the energies except the electronic energy were corrected for BSSE by the counterpoise method [Boys and Bernardi, 1970].

Cluster	I	II	III	IV	V	VI	VII	VIII
Electronic interaction energy (kcal/mol)	-3.8	-11.6	-21.6	-15.0	-30.1	-24.6	-37.4	-32.1
Electronic binding energy (kcal/mol)	-3.8 (-3.04,-4.6) <sup>(a)</sup>	-11.3	-20.5	-14.6	-28.2	-23.3	-34.9	-30.0
BSSE	0.9	2.8	5.5	3.9	7.9	6.7	9.9	9.6
MP2 one-body energy (b)	0.0	0.3	1.1	0.4	2.0	1.4	2.5	2.1
H-F Stretching vibrations (Scaled by 0.9502) <sup>(c)</sup>	v1: 3901 v2: 3954	v1: 3718 v2: 3796 v3: 3796	v1:3493 v2 :3622 v3 :3622 v4 :3678	v13629 v2 3772 v3 3864 v4 3898	v1 3380 v2 3519 v3 3519 v4 3609 v5 3609	v1 3393 v2 3566 v3 3648 v4 3734 v5 3910	v1 3347 v2 3473 v3 3473 v4 3575 v5 3575 v6 3607	v1 3355 v2 3499 v3 3499 v4 3593 v5 3593 v6 3971

Cluster	I	II	III	IV	V	VI	VII	VIII
Experimental stretch vibrations HF in He nanodroplets <sup>(d)</sup>	3862	3709	3438	N/A		v1 3174 v2 3385 v3 3489 v4 3593 v5 3876		

\* Due to Heisenberg's uncertainty principle, molecules never have zero energy at their vibrational ground states. The electronic energy determined by *ab initio* software does not include this energy and to obtain a more accurate binding or interaction energy, the electronic energy should be corrected for this. Keeping it simple, we did not correct for this explicitly.

a) The first value in parenthesis is the experimental binding energy ( $D_e$ ). The second value is the experimental binding energy without the zero-point vibration correction ( $D_0$ ) [Quack and Suhm, 1995].

b) The one-body energy is the difference between the binding and interaction energies

c) No scaling factor for the exact basis set could be found. This scaling factor is the one determined for the 6-311G(d,p) basis set. [NIST Computational Chemistry Comparison and Benchmark Database, 2005]

(d) [Douberly and Miller, 2003]

A reasonably good value, compared to the experimental value for the binding energy, was obtained. The reason why the value for the binding energy is not better is mainly because of the method and basis set used and the fact that the binding energy was not corrected for the zero-point vibration energy.

We also see that the experimental stretching vibrational modes are not in the same range as the calculated ones. This is because the scaling factor that we used might not be accurate, as it is based on a different basis set. Another problem can be that molecular vibrations are approximated by a harmonic oscillator function and in real life molecular vibrations show some level of anharmonicity. QMC can be used to correct for the anharmonicity in the vibrational modes. However, the difference in each experimental and smallest computational frequency varies at most by  $55\text{ cm}^{-1}$ , a difference of *circa*  $0.2\text{ kcal/mol}$  in energy.

In Fig 4.4 we plot the average interaction energy obtained for a monomer in each cluster against the cluster size in order to show the influence of cluster size on the interaction energy, and therefore the average many-body interactions on each monomer in the clusters.

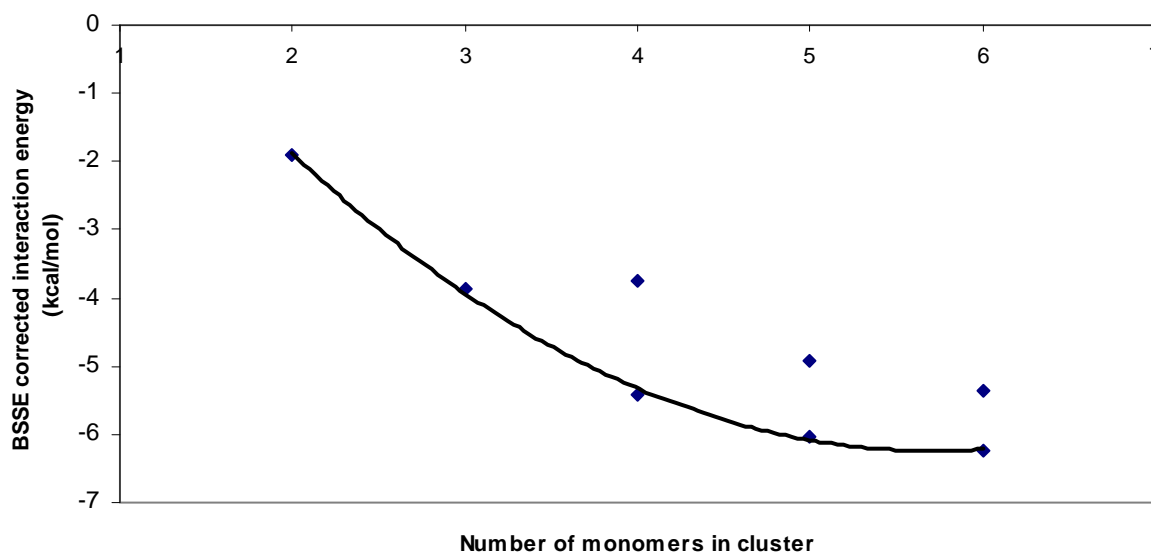


Fig. 4.4: Graph showing the trend of the average BSSE-corrected interaction energy per monomer with the increase in the cluster size. A third order polynomial can be fitted to the data representing the global minima obtained in the optimizations.

We can see from Fig. 4.4 that the change in interaction energy per monomer declines gradually with an increase in cluster size. The clusters above the trend line in Fig. 4.4 are the clusters with monomers attached to rings above or in the same plane of the rings, such as cluster IV, VI and VIII. Note that the interaction energy per monomer for the clusters with monomers attached to rings gets increasingly closer to the interaction energy per monomer for the same size cyclic clusters. Based on these results, we therefore conclude that the chance of



monomer insertion into a cyclic structure declines with cluster size. Work by Douberly and coworker agrees with this statement [Douberly and Miller, 2003].

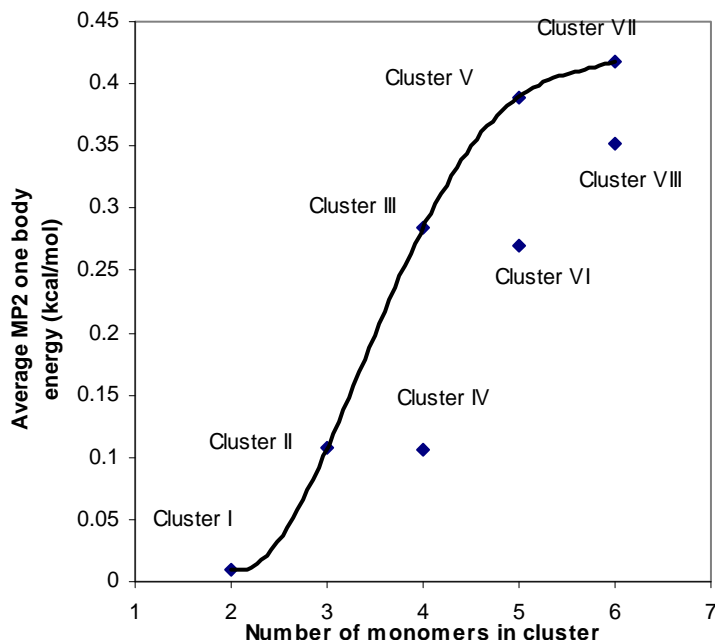


Fig. 4.5 : Graph showing the change in average MP2 one-body energy per monomer with an increase in cluster size. The data points not on the trend line represent the average one-body energies of the local energy minima. A fourth order polynomial can be fitted to the data.

Fig. 4.5 illustrates the trend in the average MP2 one-body energy per monomer with an increase in cluster size. It is evident from our results that in the case of HF clusters the one-body energy plays a fundamental role in the binding energies of the clusters.<sup>1</sup> This one-body energy is well balanced by the many-body interactions in the clusters, resulting in an overall increase in stability of the monomers in the cluster relative to when they are isolated in the gas phase. To investigate this, a many-body interaction analysis of clusters (MBAC) [Kulkarni *et al.*, 2004] can be done or SAPT can be used to obtain more information, however, a detailed analysis is not relevant to this work.

The trend in the one-body energies with respect to cluster size differs from the trend in the many-body energies with respect to cluster size since the curve drawn in Fig. 4.5 shows a sigmoidal relationship instead of a hyperbolic relationship with cluster size. We see that a similar flattening in the curve for the average one-body energies of the dimer and hexamer are responsible for this. Taking into consideration data from Table 4.1 it might be able to show that the one-body energies are related to the size of Angle 1 (see Fig. 4.1), since in the dimer

<sup>1</sup> We will see later with CO clusters that one-body energies for these clusters are insignificantly small

(Cluster I) and the cyclic hexamer (Cluster VII), this angle is similar. For the clusters with external monomers, we see that as the cluster size increases, their one-body energies become very similar in magnitude to their cyclic analogues, except for cluster VIII, which shows a significant decrease in average one-body energy from its cyclic pentamer analogue. This is due to the almost insignificant bond stretch of the external monomer from the bond length for an isolated monomer in the gas phase.

#### ***4.5 Analysis of bonding***

A common theory of hydrogen bonding is based on results obtained with *Natural Bond Orbital* (NBO) [Reed and Weinhold, 1988] analysis. According to this theory, a donor atom has a lone pair, which is donated into an antibonding molecular orbital on the molecule containing the acceptor atom. Chemically this is not a poor model, but we found that NBO analysis failed to describe the bonding, based on the MP2 density in cyclic and other clusters of HF<sup>2</sup>. When the SCF density was used there were no problems so we figured that this might have something to do with the more localized nature of the SCF density compared to the MP2 density. NBO uses a localized approach to bonding. This does not mean that NBO analysis cannot describe correlated wave functions, as it is indeed possible to do an NBO analysis on the MP2 density of the HF monomer. It simply means that the only way to apply the tools of NBO analysis to the HF cyclic clusters is by using the SCF density.

When we did a second-order perturbation analysis of the energy of off-diagonal elements in the Fock-matrices of some of the cyclic HF clusters based on the SCF density, we obtained the expected result: a donation of electrons from the fluorine lone pairs into the  $\sigma^*$  antibonding orbital on the adjacent monomer. There was no difference in this behavior as the cluster size is increased, except for a larger stabilization due to a larger charge transfer. This effect is identified as the cooperative hydrogen bond effect due to delocalization of the electron density [Rincón *et al.*, 2001]. However, we have found other effects related to electron delocalization that cannot be described by NBO analysis, forcing us to adopt a new model based on a correlated, rather than a single, determinant wave function.

To design such a model we make use of the MP2 wave function. In single molecules, this wave function does not lend itself to a simple chemical explanation in terms of bonds, due to its delocalized nature, and NBO analysis is indispensable for such cases. For delocalized clusters however, we found that the delocalized wave function can give a new perspective on hydrogen bonds and probably also on other nonbonded interactions.

---

<sup>2</sup> Doing an NBO analysis and explicitly specifying that the density to be used should be MP2, failed.

Delocalized molecular orbitals are totally constructed from quantum mechanical calculations and give the wave function with the lowest energy in the limits of the method and basis set used. Another reason why we prefer to analyze the exact delocalized wave function is that we are of the opinion that molecules should not be treated as consisting out of atoms; rather as individual entities. We think that a projection of the intramonomer electron density onto atoms of a monomer, such as is done in NBO analysis, is unnatural. *Atoms in Molecules* (AIM), another method by which molecules are divided into atoms, has recently been criticised as being a noumenon<sup>3</sup> [Parr *et al.*, 2005]. In this section, we propose a new way of understanding hydrogen bonds. We hope that in the future this model will be able to explain proper, improper, and other types of hydrogen bonding as well. We will show that the topology of the electron density obtained by the well known method of *Atoms in Molecules* supports the idea.

#### 4.5.1 Intermolecular orbitals

The nature of the hydrogen bond and weak van der Waals bonds is still not completely understood. Galvez *et al.* [Galvez *et al.*, 2001], for example, commented that in the hydrogen fluoride dimer stabilization is caused by interpenetration of the electron densities of the two monomers, rather than charge transfer, a mainstream idea [van der Vaart and Merz, 2002; Weinhold, 1997]. This comment by Galvez *et al.* was made based on *Atoms in Molecules* results. Upon bringing one monomer closer to the other they showed that electron densities interpenetrate. Some researchers such as Espinosa and coworkers [Espinosa *et al.*, 2002] have shown that a type of electron-sharing interaction takes place when two HF monomers are brought closer together to form the dimer. Dunitz and Gavezzotti [Dunitz and Gavezzotti, 2005] summarize the dilemma of the nature of the intermolecular bond as follows: “*What, then, should one do about distinguishing genuine intermolecular bonds from indiscriminate atom-atom contacts? Where should one stop talking and thinking about bonds?*” Although we are not absolutely confident that we have the answers to these questions, we believe that going back to the basis of quantum mechanics, namely the wave function, is a logical choice to start looking for them.

When the MP2 wave function of the HF dimer is visualized (see Fig. 4.6) one clearly observes what we would like to call for convenience, *bonding and antibonding intermolecular orbitals*. We will use these terms throughout this work and we emphasize that it should not be mistaken to mean bonding and antibonding MO's. We define a *bonding*

---

<sup>3</sup>The Oxford English dictionary defines a *noumenon* as, “An object knowable by the mind or intellect, not by the senses; specifically (in Kantian philosophy) an object of purely intellectual intuition”.

*intermolecular orbital* as a molecular orbital that has no nodes in the intermonomer space and an *antibonding intermolecular orbital* as an MO that has one node in the intermonomer space. The use of these terms will become clearer later. When talking about an entire wave function, we will refer to these orbitals as *molecular orbitals*, however when explicitly talking about the intermonomer space between two monomers, we will call these orbitals, *intermolecular orbitals*.

An antibonding intermolecular orbital can be interpreted as an effect of the Pauli-exchange repulsion effect between the electron densities of the monomers in the intermonomer space. The bonding intermolecular orbitals can be interpreted as the result of the distortion (electrostatic and induction) and dispersion interactions of the electron densities of the monomers. Also, observe that the bonding intermolecular orbitals are in pairs of antibonding and bonding intermolecular orbitals, similar to molecular orbitals for single molecules, except that there is “slight bonding” character in HOMO-1. In **Chapter 8** we will give a possible reason why such extra bonding intermolecular orbitals occur.

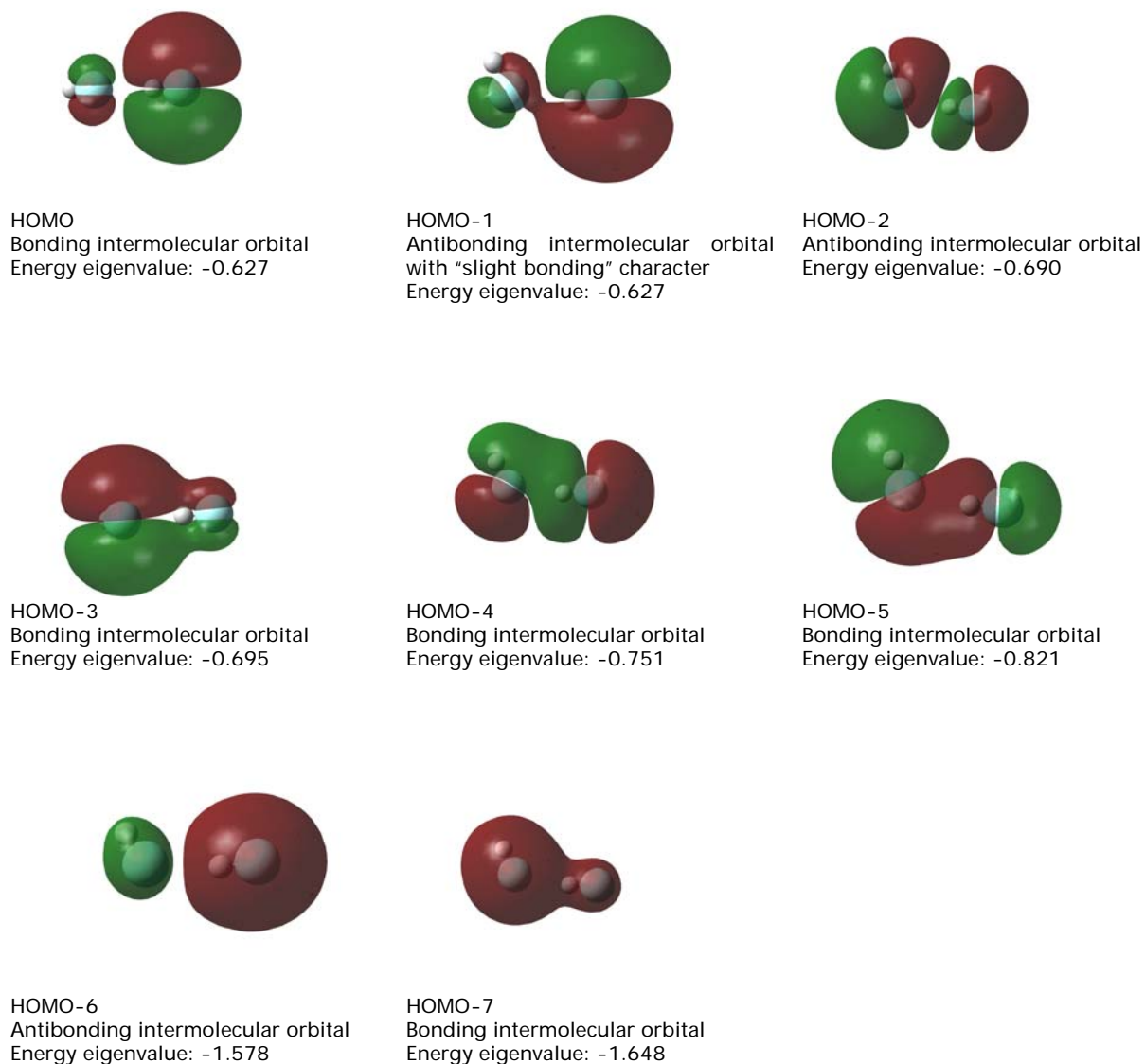


Fig. 4.6: Bonding and antibonding intermolecular orbitals as found in the HF dimer. The orbitals are arranged from highest to lowest with respect to their energy. The calculations were done with 6-311+G(d,p) as basis set. All molecular orbitals were visualized with a wave function cutoff of 0.004 a.u. The nonbonding orbitals centered on the fluorine atoms are not shown. Energy eigenvalues are given in a.u.

In Fig. 4.7 and 4.8 we show selected delocalized orbitals for clusters II and VII.

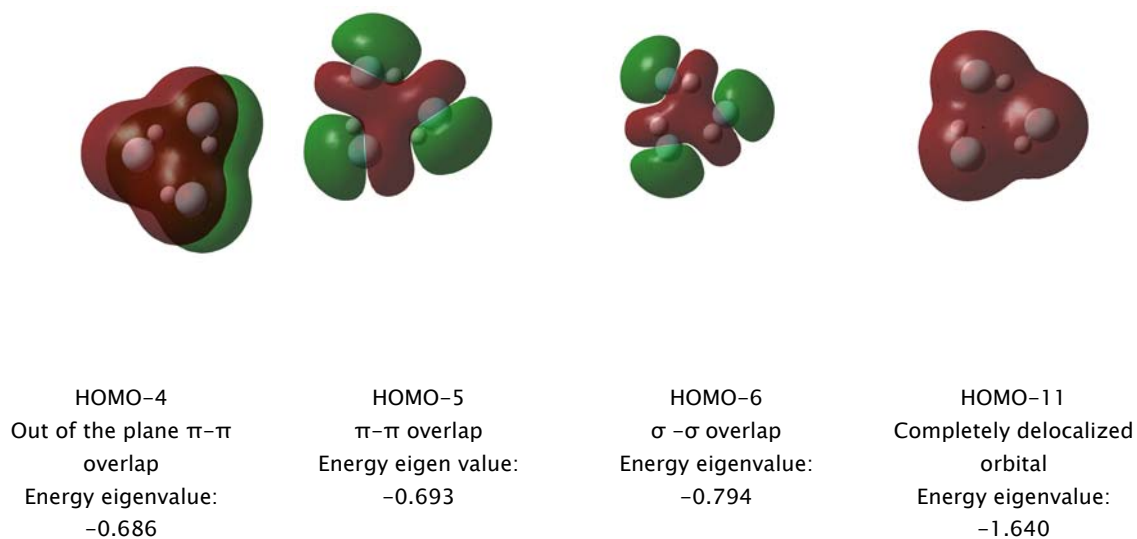


Fig. 4.7: Selected delocalized orbitals visualized for the HF trimer.

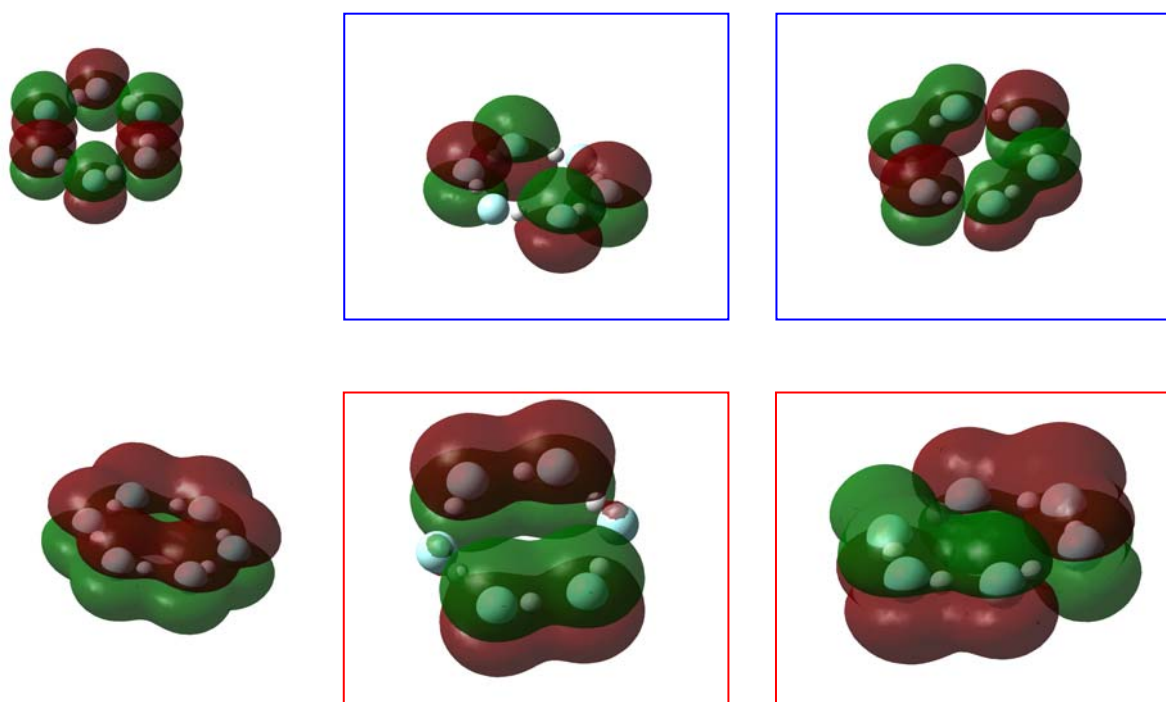


Fig. 4.8 Selected intermolecular orbitals for the cyclic HF hexamer. The point group is  $C_{6h}$  and we therefore expect degenerate states (shown here with frames of the same color) similar to that found for benzene with symmetry  $D_{6h}$ . The orbitals are arranged to decrease in energy as one moves clockwise from the top.

The wave functions that we derive are all time independent; therefore we cannot directly observe wave function evolution by general *ab initio* calculations. By looking at how the electron populations change from a single monomer wave function to the the wave function for a monomer in a cluster, we can however obtain some idea of the movement of electron

density. For convenience that will become clear later, we define the *relative electron population* as the change in the electron population from one quantum state to another. Here specifically the *relative electron population* is calculated as the difference between the electron population for a specific monomer in a cluster and the electron population for an isolated monomer. For example, if the electron population is 1.2 a.u. in a specific orbital in a monomer and in the cluster it is 0.8 a.u., then the *relative electron population* is equal to -0.4 a.u. for this orbital. We will use this term often throughout this work and its definition will therefore not be repeated again. We will call the method of determining the relative electron populations, *relative electron population analysis*. To obtain the electron population for the isolated monomer and the monomer in a cluster in order to compute the *relative electron populations*, we use Mulliken population analysis. In computational chemistry Mulliken population analysis is known to give erroneous electron populations in some cases such as when using diffuse functions as we did [Jensen, 2001], however it was used here as it is solely based on the basis functions and can give an indication of how the basis functions are divided amongst the molecule(s) in a cluster compared to the isolated molecules. NBO analysis, for example, uses natural atomic orbitals and is more localized so that the nature of the MP2 density is not reproduced efficiently. In order to obtain useful information, we decided to determine only the electron populations of each single shell.

As basis functions are centered on atoms, and because we are working with molecules, we would not expect to find similar electron populations in the “atomic centered orbitals” compared to those in the isolated atoms. In our analysis, the “atomic centered orbitals” are divided into *s* and *sp* shells. This division is similar to the division in some basis sets. The *d*-orbitals are treated separately. Each *sp* or *s*-shell has a different main quantum number. In Fig. 4.5 we show the *relative electron populations* for the fluorine atom in an HF monomer in a selected number of clusters. The shells and the *3d* orbitals are shown separately. In Fig. 4.6 we repeat this analysis for the hydrogen atom in the HF monomer. Notice that here, due to the simplicity of this part of the wave function, we decided to treat the *2s* and *2p* orbitals separately instead of treating it as one shell.

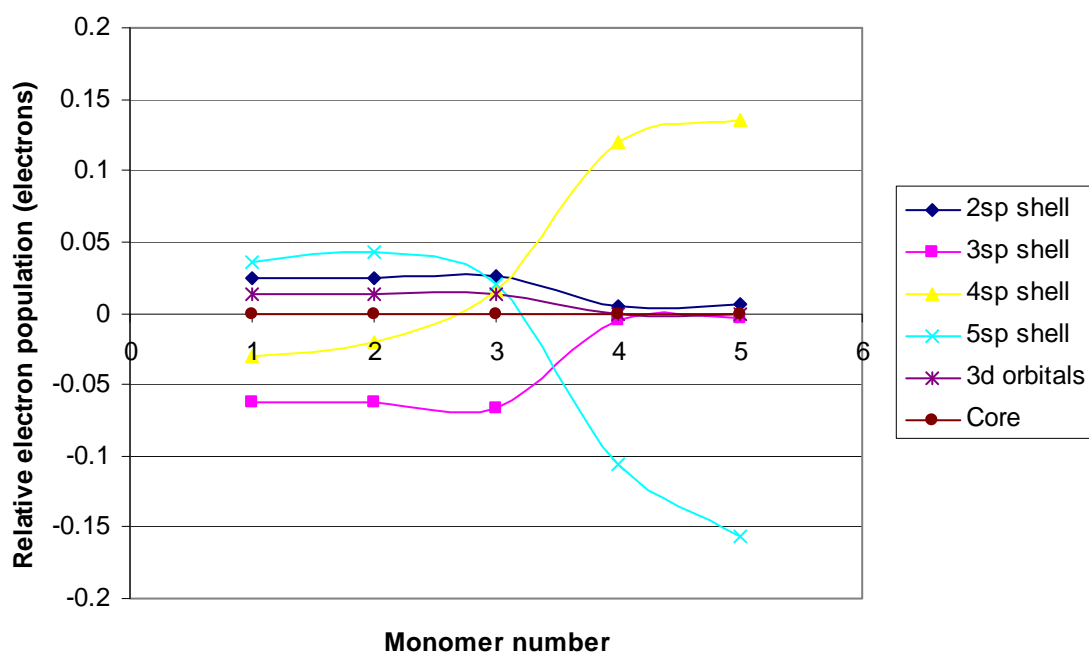


Fig. 4.5: *Relative electron populations* for the F-atoms found in each selected cluster. The monomer numbers are designated as follows:

1 refers to the acceptor monomer in the HF dimer, 2 refers to the donor monomer in the HF dimer and 3, 4 and 5 refers to a monomer in the cyclic trimer (II), the cyclic tetramer (III) and the cyclic pentamer (V) respectively.

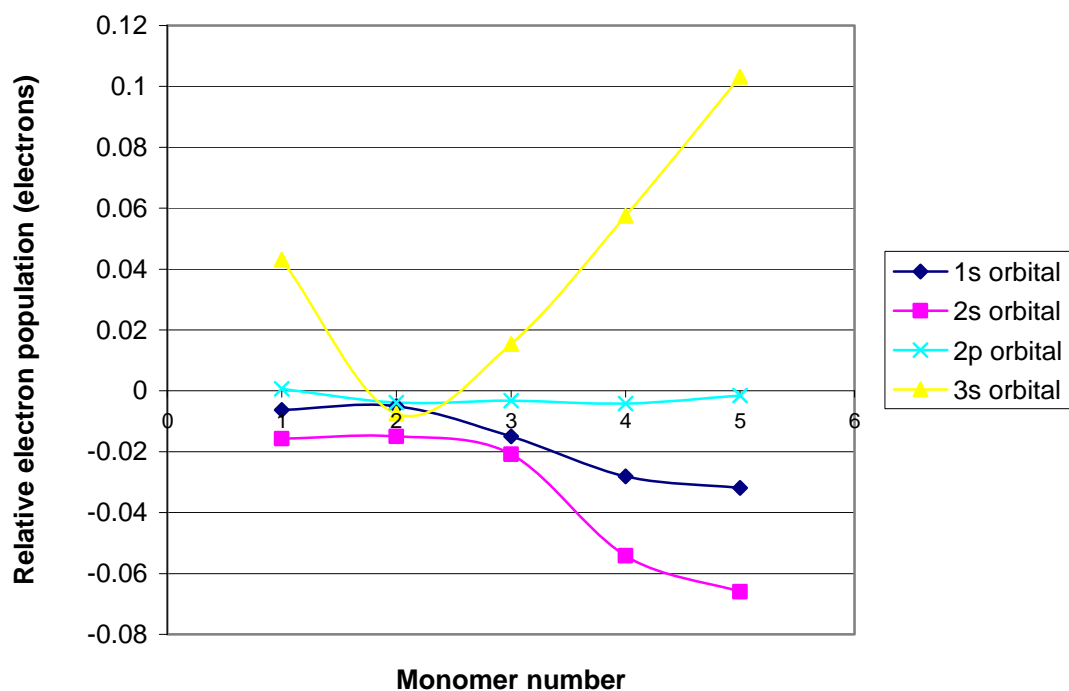


Fig. 4.6: *Relative electron populations* for the H-atoms found in each selected cluster. The monomer numbers are designated as follows:

1 refers to the acceptor monomer in the HF dimer, 2 refers to the donor monomer in the HF dimer and 3, 4 and 5 refers to a monomer in the cyclic trimer (II), the cyclic tetramer (III) and the cyclic pentamer (V) respectively.



It is clear from both Fig. 4.5 and Fig. 4.6 that the *relative electron populations* in the clusters are slightly different to that in an isolated HF monomer, otherwise all the data points would lie on the ordinate axis. Therefore, there is definitely a change in the population of energy levels when the monomer finds itself in the electric field of another monomer, such as in a cluster. This is a beautiful illustration of polarization in action. In **Section 2.2** we mentioned that polarization can be defined in terms of the transition moment from one quantum state to another and an excitation energy. We see similar behavior with metal complexes where the field of the ligands changes the population of the electrons in such a way as to give rise to different terms at different field strengths.<sup>4</sup>

Using correlated wave functions, the mixing of atomic orbitals with atomic orbitals of higher main quantum numbers results in a lowering of the energy of the orbitals. The hybrids formed in this way contain a larger contribution by higher lying spin orbitals to their character than in Hartree-Fock theory. These orbitals are more diffuse (larger) and can overlap better with orbitals on adjacent monomers in clusters. We see that the fluorine atom donates electron density to the acceptor monomer's  $2p$  orbitals on the hydrogen atom when the acceptor and donor monomers are compared. This is in accordance with the NBO analysis discussed at the start of **Section 4.5**, where a lone pair is donated to an antibonding orbital centered mainly on the hydrogen of the acceptor monomer. However, when thinking in terms of orbitals delocalized over various monomers, we can no longer discuss molecular orbitals as each belonging entirely to a single monomer. According to the delocalized intermolecular orbital hypothesis<sup>5</sup>, suggested earlier it would be incorrect to say that electrons are transferred from one molecular orbital on one monomer to another on a different monomer. We have to explain this behavior based on the donation of electrons into a new bonding delocalized orbital that is delocalized over all the atoms in the system. The electron density is just higher around the accepting monomer than if it was isolated. We also see that for clusters larger than the dimer, more and more electrons are removed from the core orbitals of the hydrogen atom into higher shells on the hydrogen atom. It therefore seems as if both the fluorine and the hydrogen atoms play a role in the donation in a delocalized orbital that is delocalized over the entire cluster. The movement of valence and sometimes core electrons (see *relative electron population analysis* in Fig. 4.6 for the hydrogen atom) to the intermolecular orbitals is controlled by the amount of polarization that one monomer's electric field effects on another monomer's wave function, as well as electron correlation. This should result in the weakening of the valence bonds of the monomers as we indeed found earlier (compare Table 4.1).

When studying the graphs in Fig. 4.5 and 4.6 we see that there are a significant change in the relative electron population of the  $4sp$  and  $5sp$  shells of the fluorine atom between the

---

<sup>4</sup>However, relativistic effects also play a role in metal ligand bonding.

<sup>5</sup> It is only a hypothesis as it has not been proven except by an indirect method.

trimer, tetramer and pentamer. In the trimer the  $4sp$  and  $5sp$  shell almost have the same relative electron populations, but in the tetramer and pentamer it is different as if these two shells are reversed and also “unbalanced”<sup>6</sup> with respect to the isolated monomer. It might be that for the clusters larger than the trimer the electrons in the  $5sp$ -shell, which are lower in energy than the electrons in the  $4sp$ -shell (relative to an isolated monomer) in these clusters, are excited into the  $4sp$ -shell. This would lead to a slight destabilization in the larger cyclic clusters. The significance of these results will become clear, but first we have to draw a comparison between the arrangement of the electrons in the HF trimer and the arrangement of electrons in benzene. Benzene has a strong ring current and is regarded as aromatic, whereas cyclobutadiene is known to be anti-aromatic [Rehaman *et al.*, 2006]. When one studies the electron configuration of benzene, one sees that all the bonding  $\pi$ -orbitals are filled and this is most probably the reason for its aromaticity, as other monocyclic systems not having  $4n+2$   $\pi$ -electrons are not aromatic [Cotton, 1990]. In the HF trimer, the  $4sp$  and  $5sp$ -shells have more electrons than in the monomer, whereas the other cyclic clusters have “unbalanced” electron configurations such as in cyclobutadiene. It is therefore possible that this is the reason for the weak ring current<sup>7</sup> that is known to be present in the HF trimer [Rehaman *et al.*, 2006] and not in the other cyclic HF clusters. The ring current might be related to the fact that fewer electrons need to be excited from the  $5sp$ -shell to the  $4sp$ -shell or vice versa, depending on the energies of the shells, as the  $4sp$  and  $5sp$ -shells already contains sufficient electrons. The electron-flow will therefore not be interrupted by an excitation as might be possible in the larger cyclic clusters. Again, this shows the usefulness of the intermolecular orbital model for describing bonding in HF cyclic clusters. The reader might argue that these electrons are not  $\pi$ -electrons such as in benzene. However, it is now realized that aromaticity might include sigma-delocalized systems as well [Rehaman *et al.*, 2006].

Our model can also explain cooperativity, just as in the NBO theory for hydrogen bonding, as we see in both Fig. 4.5 and 4.6 that with cluster size the electron density donated from the donor to the acceptor monomer increases and the “back-donation” is also more significant. Back-donation is again a localized concept. According to our intermolecular orbital explanation, it should rather be explained as a donation of the acceptor monomer’s electron density into the delocalized intermolecular orbital. Both the polarizability of the atoms in the monomers and their electron negativity will influence their interaction with each other.

---

<sup>6</sup> They are “unbalanced” as the one shell loses more electrons than in the isolated monomer and the other gains much more electrons than in the isolated monomer. In the trimer both the  $4sp$  and  $5sp$  shells have on average more electrons than in the isolated monomer irrespective of when the measurement is done.

<sup>7</sup> NICS (Nuclear independent chemical shift) values are calculated by determining the chemical shift of a ghost atom containing only basis functions and no nucleus, in the magnetic field of a molecule. These values can be used as a sign of aromaticity or anti-aromaticity in molecules.

The effect of basis set/atomic orbital superposition, described so far, is well known but as far as we know have never been used to describe intermolecular interactions; only valence bonds.

Thinking in terms of intermolecular orbitals, strange hydrogen bonds such as the hydrogen bond between HF and acetylene [Wojtulewski and Grabowski, 2002] can be easily explained as a  $\pi$ -orbital on the acetylene overlapping with molecular orbitals, probably  $\sigma$ , on the HF monomer.

By plotting the molecular orbitals of all the other clusters it was evident that the molecular orbitals show delocalization of both  $\pi$  and  $\sigma$  electron density. In clusters IV, VI and VIII the wave functions are delocalized from the external monomers to the cyclic rings.

#### 4.5.2 Atoms in Molecules (AIM)

To complete our understanding of the bonding in these clusters, we made use of the theory of *Atoms in Molecules* (AIM) [Biegler-König *et al.*, 1982]. AIM is a method where the wave function of a molecule is used to derive its electron density. The electron density is a three-dimensional function and has many maxima, minima and saddle points. One of the aims of AIM is to find critical points on this hyper-surface. These critical points are all maxima in one or two directions on the electron density hyper-surface. A Newton-Raphson optimization algorithm is used to find these maxima on the surface of the electron density. Nuclear attractors are points in the electron density that are three-dimensional maxima on the hyper-surface. They represent the core electrons or the K-shell electrons. By taking the second derivatives of the electron density with respect to position, the signature of the local maxima can be determined on the electron density hyper-surface. The signature gives us information on the nature of the local maxima. For example a bond critical point (BCP) has a signature of (3,-1) meaning that there are 3 directions of local curvature and two of these local curvatures have a negative second derivative and one a positive second derivative. The electron density is therefore a maximum in 2 directions and a minimum in one at this critical point.

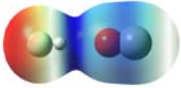
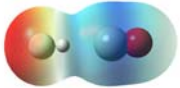
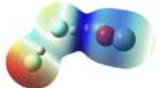

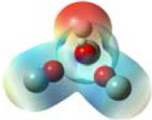
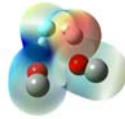
To investigate the movement of electrons in the clusters we used AIM to determine the electron densities ( $\rho$ ) at various critical points. These results are reported in Table 4.3. We do not report the values for the Laplacians or second derivatives as all of them, except for cluster VIII, are negative as typical for hydrogen bonds.

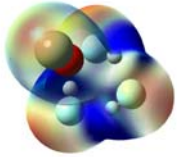




Table 4.3: AIM analysis of the HF clusters. BCP stands for Bond Critical Point and RCP stands for Ring Critical Point. Mon denotes: monomer.

	Single Monomer	I	II	III	IV	V	VI	VII	VIII
<b><math>\rho</math> of BCP (e.bohr<sup>-3</sup>)</b>	0.370	0.362 0.364	0.350	0.337	External:0.362 Mon1:0.339 Mon2:0.349 Mon3:0.356	0.331	External:0.363 Mon1:0.333 Mon2:0.323 Mon3:0.344 Mon4:0.338	0.329	Mon in ring: 0.330 External Mon: 0.366
<b>Relative shift in BCP away from fluorine atom in monomer (%)</b>	0.00	0.34 0.25	0.57	0.80	External 0.36 Mon1 : 0.48 Mon2: 0.79 Mon 3:0.66	0.87	External: 0.32 Mon1:0.72 Mon2:0.91 Mon3:0.86 Mon4:0.80	0.90	External Mon: 0.16 All monomers in ring: 0.88
<b>RCP of ( e.bohr<sup>-3</sup>)</b>			0.008	0.002	0.007	0.0005	0.002	0.0001	0.004

- a) The relative position is calculated by dividing the distance of the critical point from a hydrogen atom by the bond length of the monomer and multiplying by 100 and subtracting this result from the result obtained for the same calculation for a single monomer. It is therefore the percentage movement of the BCP from its position in the monomer.

Table 8.5: ESP-maps including charges derived with various charge schemes. All the values are in a.u.

ESP -maps	SCF density based MKS charges	MP2 density based MKS charges [AIM charges]	MP2 density based NAO charges	ESP-maps	SCF density based MKS charges	MP2 density based MKS charges [AIM charges]	MP2 density based NAO charges
	F1:-0.374 H2:0.256 O3:0.110 C4:0.009	F1:-0.350[-0.721] H2:0.231[0.718] O3:0.233[-1.131] C4:-0.114[1.134]	F1:-0.546 H2:0.544 O3:-0.463 C4:0.464		F1:-0.387 H2: 0.225 C3: 0.364 O4:-0.202	F1 : -0.369[-0.740] H2 :0.210[0.714] C3 :0.231[1.101] O4 :-0.072[-1.075]	F1:-0.562 H2:0.544 C3:0.397 O4:-0.379
	O1:0.116 C2 :0.017 H3:0.244 H4:0.430 F5:-0.333 F6:-0.474	O1:0.058 C2:-0.049 H3:0.296 H4:0.253 F5:-0.282 F6:-0.275	O1:-0.478 C2:0.481 H3:0.563 H4:0.552 F5:-0.555 F6:-0.563		F1:-0.380 H2:0.246 C3:0.296 O4:-0.183 C5:0.126 O6:-0.105	F1:-0.358 H2:0.224 C3:0.173 O4:-0.058 C5:0.006 O6:0.014	F1-0.565 H2:0.536 C3:0.400 O4:-0.381 C5:0.409 O6:-0.409
	O1:0.057 C2:-0.001 C3:-0.001 O4:0.058 C5:0.006 O6:0.047 H7:0.184 F8:-0.350	O1:0.169 C2:-0.114 C3:-0.113 O4:0.168 C5:-0.108 O6:0.160 H7:0.159 F8:-0.321	O1:-0.445 C2:0.444 C3:-0.445 O4:0.444 C5:0.444 O6:-0.445 H7:0.548 F8:-0.546		F1:-0.437 H2:0.412 O3:-0.009 C4:0.038 H5:0.292 F6:-0.363 C7:0.034 O8:0.033	F1:-0.416 H2:0.389 O3:0.105 C4:-0.077 H5:0.261 F6:-0.330 C7:-0.083 O8:0.151	F1:-0.565 H2:0.552 O3:-0.447 C4:0.446 H5:0.562 F6:-0.549 C7:0.467 O8:-0.467

	F1:-0.458 H2:0.449 F3:-0.463 H4:0.453 H5:0.448 F6:-0.461 O7:-0.002 C8:0.033	F1:-0.433 H2:0.425 F3:-0.438 H4:0.429 H5:0.424 F6:-0.436 O7:0.120 C8:-0.090	F1:-0.568 H2:0.568 F3:-0.568 H4:0.568 H5:0.568 F6:-0.567 O7:-0.432 C8:0.430		F1:-0.386 H2:0.481 O3:-0.117 C4:0.101 H5:0.254 F6:-0.278 C7:0.156 O8:-0.210	F1:-0.369 H2:0.447 O3:0.039 C4:-0.058 H5:0.249 F6:-0.272 C7:0.016 O8:-0.052	
	H1:0.167 F2:-0.319 C3:0.052 O4:-0.026 F5:-0.434 H6:0.388 C7:0.375 O8:-0.204	H1:0.144 F2:-0.294 C3:-0.072 O4:0.097 F5:-0.413 H6:0.365 C7:0.246 O8:-0.073			C1:0.130 O2:-0.107 O3:-0.017 C4:0.130 F5:-0.383 H6:0.279 C7:0.227 O8:-0.169	C1:0.011 O2:0.011 O3:0.011 C4:0.011 F5:-0.356 H6:0.248 C7:0.113 O8:-0.047	
	H1:0.408 F2:-0.461 H3:0.137 F4:-0.253 F5:-0.456 H6:0.4000 C7:0.420 O8:-0.195	H1:0.382 F2:-0.439 H3:0.106 F4:-0.217 F5:-0.433 H6:0.373 C7:0.294 O8:-0.065					

Again as in both **Chapters 5** and **7**, the  $\epsilon$ -values were not changed. To save time, only the first six clusters were used. We believe that this is still adequate for a force field optimization.

### 8.9 Transferability of force field parameters and optimization for CO/HF clusters

In Table 8.6 we show how transferable the different force field parameters, as optimized for the CO and HF clusters, are to the CO/HF clusters. It is an important feature of good force field parameters that they are transferable to other systems containing the same atom-types [Leach, 2001]. We use the RMSD's for the fitting of the ONIOM steric energies to the MP2 BSSE corrected interaction energies as a measure of the quality of the force field parameters.

Table 8.6: Table illustrating the transferability of force field parameters. The quality of the force field is assessed based on ONIOM steric energies. See text.

	MP2 interaction energies (kcal/mol)	Energies obtained for the geometrically optimized CO and HF parameters as reported in <b>Chapters 5</b> and <b>7</b> (kcal/mol)	Energies for parameters obtained for the atom-types in <b>Chapters 5</b> and <b>7</b> with a frozen geometry method (kcal/mol)	Energies for best optimized force field parameters for CO/HF clusters as reported in this chapter (kcal/mol)	Energies obtained for the default UFF parameters (kcal/mol)
I	-1.05	3.75	3.34	3.33	7.03
II	-3.03	3.92	4.04	3.82	17.03
III	-5.25	-1.04	-1.50	-1.90	25.62
IV	-3.91	2.81	3.08	2.70	18.01
V	-2.11	10.84	-1.63	1.75	0.11
VI	-5.42	1.34	-3.44	-2.78	-3.93
R <sub>0</sub> -values (Å)	N/A	F <sub>-</sub> : 3.340 H <sub>-</sub> : 1.280 C <sub>-1</sub> : 3.700 O <sub>-1</sub> : 4.500	F <sub>-</sub> : 3.386 H <sub>-</sub> : 0.481 O <sub>-1</sub> : 3.687 C <sub>-1</sub> : 2.948	F <sub>-</sub> : 2.936 H <sub>-</sub> : 1.454 O <sub>-1</sub> : 3.403 C <sub>-1</sub> : 3.446	F <sub>-</sub> : 3.364 H <sub>-</sub> : 2.886 C <sub>-1</sub> : 3.851 O <sub>-1</sub> : 3.500
Total RMSD (kcal/mol)	N/A	7.61	4.77	4.53	17.84

Studying the RMSD's of the ONIOM steric energies with respect to the MP2 BSSE corrected interaction energies, we see clearly that the default UFF parameters are not suitable for the force field, as only cluster VI gives a negative ONIOM steric energy for the MP2 optimized

geometry. Using the parameters optimized with geometry optimizations (**Chapters 5 and 7**), we obtain a slightly better result; probably because the H\_ and F\_ R<sub>0</sub>-values have already been shown to give excellent results (see **Chapter 5**). When parameters optimized with a frozen geometry method (**Chapters 5 and 7**) are used to calculate the ONIOM steric energies of the selected CO/HF clusters, we see that the values for the ONIOM steric energies are slightly closer to the MP2 BSSE corrected interaction energies than those obtained when using geometry optimizations to obtain the force field parameters (see column 4 in Table 8.6). When the parameters for F\_, H\_, O\_1 and C\_1 are optimized independently, specifically for the six CO/HF clusters as described earlier in this chapter, we find that the RMSD in the total ONIOM steric energy is only slightly improved. In Fig. 8.4 we see that the ONIOM steric energies for each cluster are still not a good fit to the MP2 BSSE corrected interaction energies for each respective cluster.

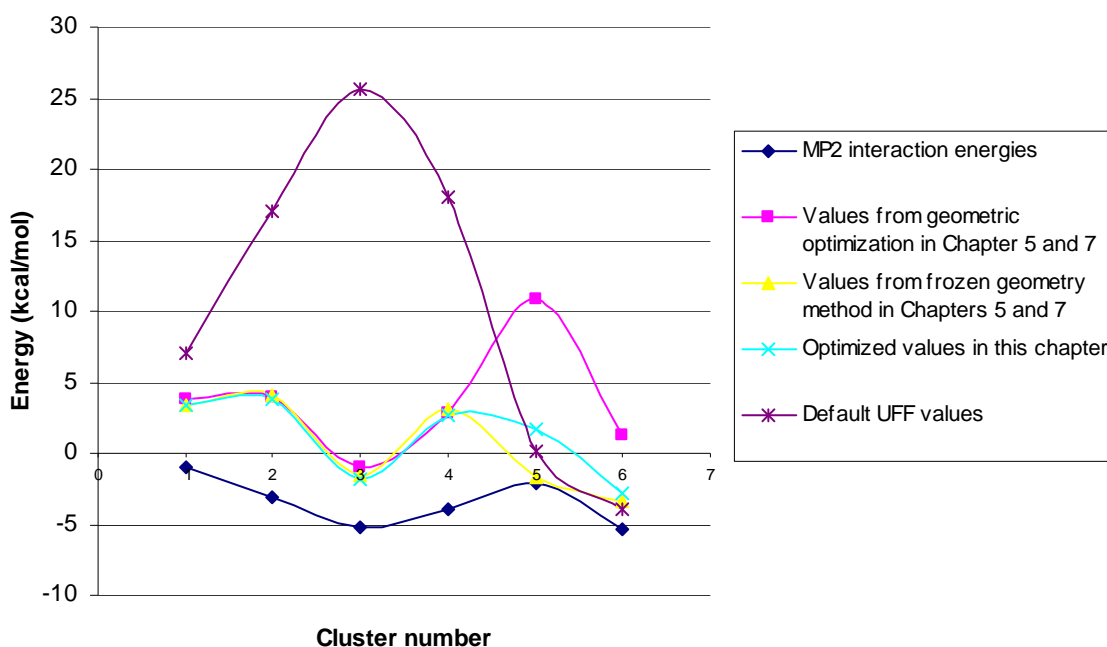


Fig. 8.4: The ONIOM steric energies/MP2 interaction energies shown for the first six CO/HF clusters. A cluster number of 3 refers to cluster III etc.



Even when the optimization was performed for the two dimers (Cluster I and II) and the van der Waals energy was explicitly minimized to a global minimum<sup>1</sup>, the ONIOM steric energies were still positive, as the electrostatic energy is too positive to be cancelled out by the van der Waals energy. Therefore, for the dimers it would only be possible to obtain a negative energy if the electrostatic energy was negative, which is not the case when MKS point charges are used on the MM atoms. We see that the trend for the optimized parameters is basically the same as for the transferred parameters, except for cluster V and VI. The parameters for Cluster V and VI are better for the values obtained in **Chapter 5** and **7** than for the newly optimized values. The best solution to the optimization of the force field, in terms of energy at least, is therefore not adequate. The main, and possibly the only, reason is the inadequate reproduction of the electrostatic energy for these clusters, because point charges are used instead of a more elaborate description of the electrostatic interactions.

### ***8.10 Summary, conclusions and future work***

In this chapter we reported the optimization of eleven CO/HF clusters with the 6-311+G(d,p) basis set at the MP2 level of theory. Various energies and geometrical parameters were determined and can be found in Table 8.1 and 8.2. Bonding was briefly analyzed for the O···H and C···H bonded dimers, based on AIM analysis and a delocalized molecular orbital model. An AIM analysis shows a larger interpenetration of the electron densities of the two monomers in the C···H bonded dimer than the O···H bonded dimer and this might be the main reason why the O···H bonded dimer is the local energy minimum and the C···H bonded dimer the global energy minimum. During the force field optimizations for the QM/MM hybrid system, we could not find an adequate fit of the ONIOM steric energies to the BSSE corrected MP2 interaction energies. There are two possible reasons for this: the atomic point charges alone are not sufficient to fully describe the electrostatic energy in the clusters and other methods are necessary; or the values of the point charges are inaccurate. Future work could include analyzing the different nonbonded interactions in the other CO/HF clusters by using AIM analysis, and the optimization of the force field using a more elaborate description of the electrostatic interactions in the clusters. In the next chapter we will discuss ways in which the hybrid method as a whole can be improved, to be suitable for the accurate modeling of clusters. A detailed vibrational analysis at the MP2 level of theory of all the CO/HF clusters might also prove valuable.

---

<sup>1</sup> When a UFF force field geometry optimization was performed with the optimized parameters while excluding electrostatic interactions, the geometry did not change from the geometry of the energy minimum at the MP2 level of theory.

## CHAPTER 9

# The ONIOM-EE method under the looking glass

### *9.1 Introduction*

The combination of quantum mechanics with molecular mechanics is an artificial way to allow for accurate calculations on systems that are usually considered too large for standard quantum mechanical methods. However, in the case of small clusters as discussed in this work, QM/MM might not be the ideal choice as certain quantum effects such as electron correlation cannot be taken into consideration for both the QM and MM systems and, as we have illustrated especially in **Chapter 6**, electron correlation may play a significant role in intermolecular interactions. Therefore it would be better to use a QM/QM method where the interaction between the two QM systems can be calculated at a low level of theory. Hydrogen bonds in alcohol dimers, for example, have been studied by IMOMO [Tschumper and Morokuma, 2002]. However, as this work specifically entails using QM/MM we will try to give suggestions on how QM/MM can be modified to make calculations for clusters more accurate. All our assumptions in this chapter are based on ONIOM-EE as implemented in *Gaussian 03*.

During the use of ONIOM-EE in this work, we experienced several problems related to electronic embedding. In **Section 9.2** we will describe these problems and finally give possible reasons for them. In **Section 9.3** we will give reasons for the need for a better charge density description of the MM system. In **Section 9.4** we will discuss problems with geometry optimizations for systems utilizing rigid blocks in the MM system and how these problems can be solved, and in **Section 9.5** we will show that Gaussian does not necessarily derive suitable ESP charges for ONIOM-EE. In **Section 9.6** we will discuss the difference in hydrogen bond distances for QM-QM and QM-MM interactions. In **Section 9.7** we will present a novel method of increasing the accuracy for the modeling of clusters with ONIOM-EE and **Section 9.8** will conclude this chapter.

## 9.2 The electronic embedding procedure

During this work we noticed when modeling the HF clusters with ONIOM-EE that the energy at the high level of theory for the model system was much larger in absolute value than is obtained when the same monomer, in the model system, was isolated and its energy calculated at the MP2 level of theory. This result was found when either MKS or NAO charges were used on the atoms in the MM system. To simplify the following discussion we will define two terms. The ONIOM electrostatic energy is defined as the sum of  $E_{MM}^{el}$  and  $\Delta P$ . These terms are defined in **Section 5.6**. The classical electrostatic energy is the same as  $E_{el}$  that was also defined in **Section 5.6**. We report the results for both MKS and NAO point charges on the atoms in the MM system as well as the MP2 interaction energies in Table 9.1.

Table 9.1: Results for the calculations of the ONIOM electrostatic energies and classical electrostatic energies for six of the HF clusters. The results are reported for both MKS and NAO charges on the MM atoms. The MP2 interaction energies (BSSE-corrected) are also shown.

	MKS charges		NAO charges		
HF clusters	Total ONIOM electrostatic energy (kcal/mol)	Total classical electrostatic energy (kcal/mol)	Total ONIOM electrostatic energy (kcal/mol)	Total classical electrostatic energy (kcal/mol)	MP2 interaction energy (kcal/mol)
I	-78.7	-4.7	-117.6	-6.5	-3.8
II	-149.5	-13.4	-270.0	-20.6	-11.6
III	-210.2	-24.4	-413.1	-41.5	-21.6
IV	-234.6	-17.7	-390.3	-27.5	-15.0
V	-282.3	-35.1	-567.4	-62.4	-30.1
VI	-285.7	-26.6	-542.0	-48.2	-24.6

In Table 9.1 we see that the ONIOM electrostatic energies are always much larger than either the classical electrostatic energies or the MP2 interaction energies, no matter which charge scheme was used. We also see that the trend in the ONIOM electrostatic energies with cluster size is different from the trend for the MP2 interaction energies when MKS charges are used, whereas when NAO charges are used, it is the same. For example, when MKS charges are used, Cluster IV is lower in absolute ONIOM electrostatic energy than Cluster III, which does not correspond to the trend in the MP2 interaction energies, but when NAO charges are used,

then Cluster III is more stable than Cluster IV. We conclude that in this particular case NAO charges should be a better option as they give a more accurate description of all the clusters' charge densities. We also plotted the ONIOM electrostatic energies and classical electrostatic energies for the CO clusters as they vary with cluster size and this is presented in Fig. 9.1. For these calculations, we only used MKS charges on the MM atoms.

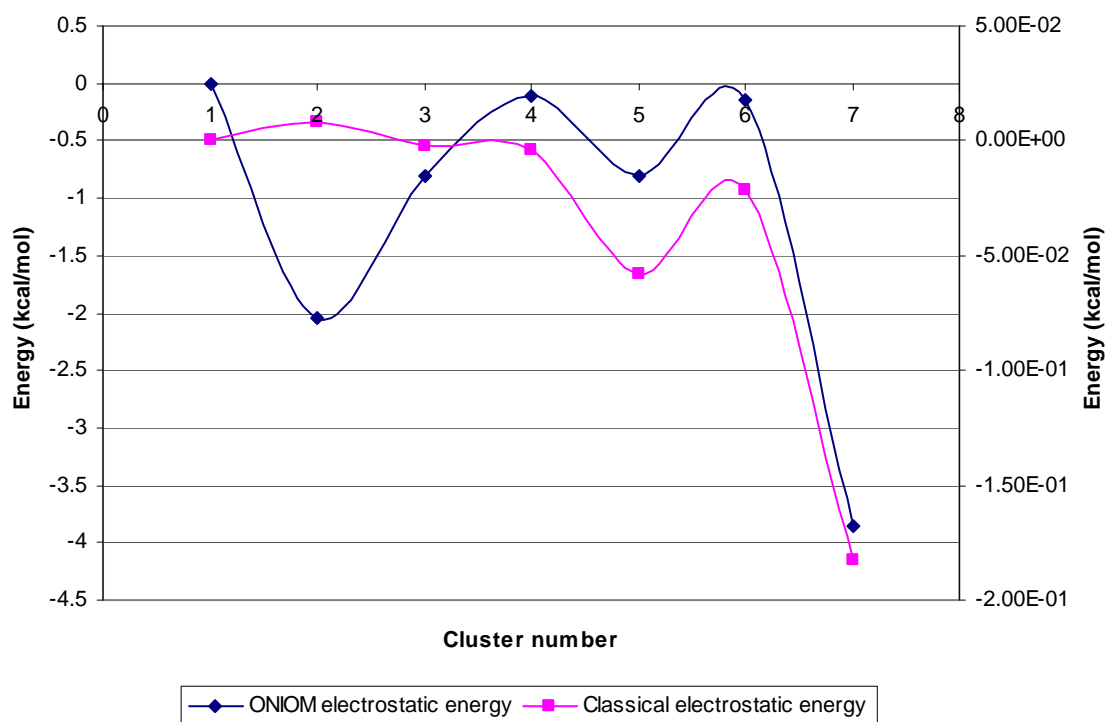


Fig. 9.1: The ONIOM electrostatic energy for each cluster is shown on the primary axis and the classical electrostatic energy is shown on the secondary axis. The cluster numbers are in the order in which the clusters were presented in **Chapter 6**. The cluster numbers for Cluster I and Cluster VII for example are 1 and 7 respectively.

In Fig. 9.1 we see that the trend in ONIOM electrostatic energies and classical electrostatic energies are only similar for Clusters IV to VII. This result shows that approximating the polarized QM wave function's density by ESP charges is not always a good approximation. In this case the approximation seems to be better for the larger clusters.

### 9.3 The need for a better charge density description

In **Section 9.2** we reported that the absolute ONIOM electrostatic energy is too large in magnitude when compared to the MP2 interaction energies. In this section we will try to elucidate the reason for this.

We attempted to obtain an output file containing the polarized wave function of the QM system for AIM analysis, but to no avail and instead had to use another method to obtain information on the polarized QM wave function. We used a method that we introduced in **Chapter 4**, which we coined *relative electron population analysis*. We did relative electron population analysis for the monomer in the QM system in the QM/MM system and for the same monomer in a dimer calculated with MP2. The latter will be called the *MP2 dimer*. In both the ONIOM single point and MP2 single point energy calculations, the MP2 energy minimum geometries were used.

The *relative electron populations* for the fluorine atom are shown in the first graph (Fig. 9.2) and those for the hydrogen atom are shown in the second (Fig. 9.3). Note that the  $z$ -axis is in the plane of the main  $C_s$  symmetry axis, therefore it is almost in line with the hydrogen bond and the  $x$  and  $y$ -axes are arranged as in a right-handed coordinate system.

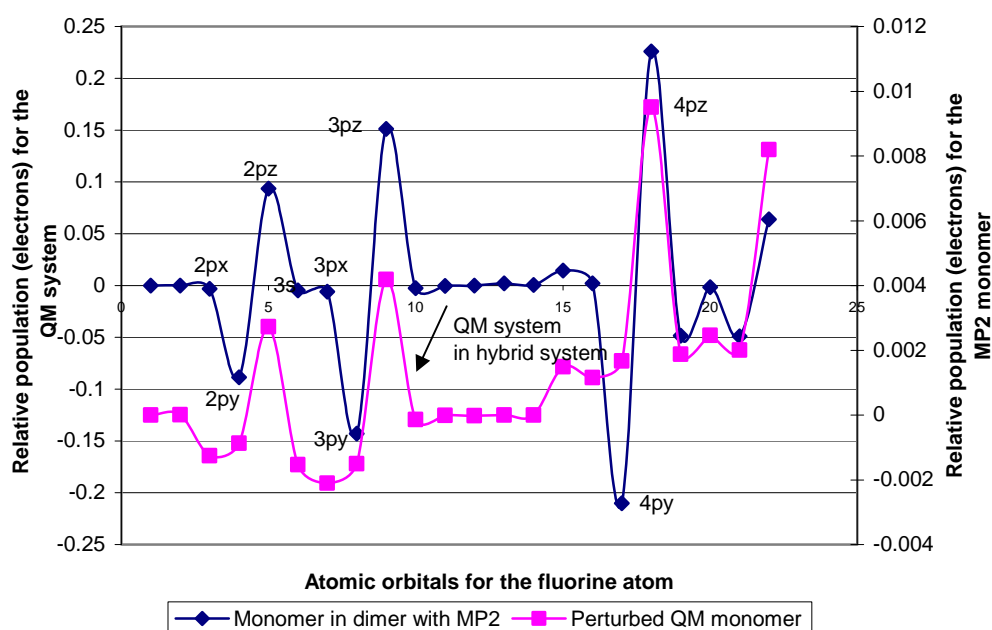


Fig. 9.2: Graph showing the difference in the relative populations of the electrons in each orbital for the fluorine atom in the QM system. The orbitals are numbered in the sequence of their main quantum numbers. The graph therefore starts with the  $1s$ -orbital and ends with the  $5p_z$ -orbital. The relative electron populations for the monomer in the MP2 optimized dimer are shown on the primary axis and the relative electron populations for the monomer in the QM system in the QM/MM hybrid are shown on the secondary axis.

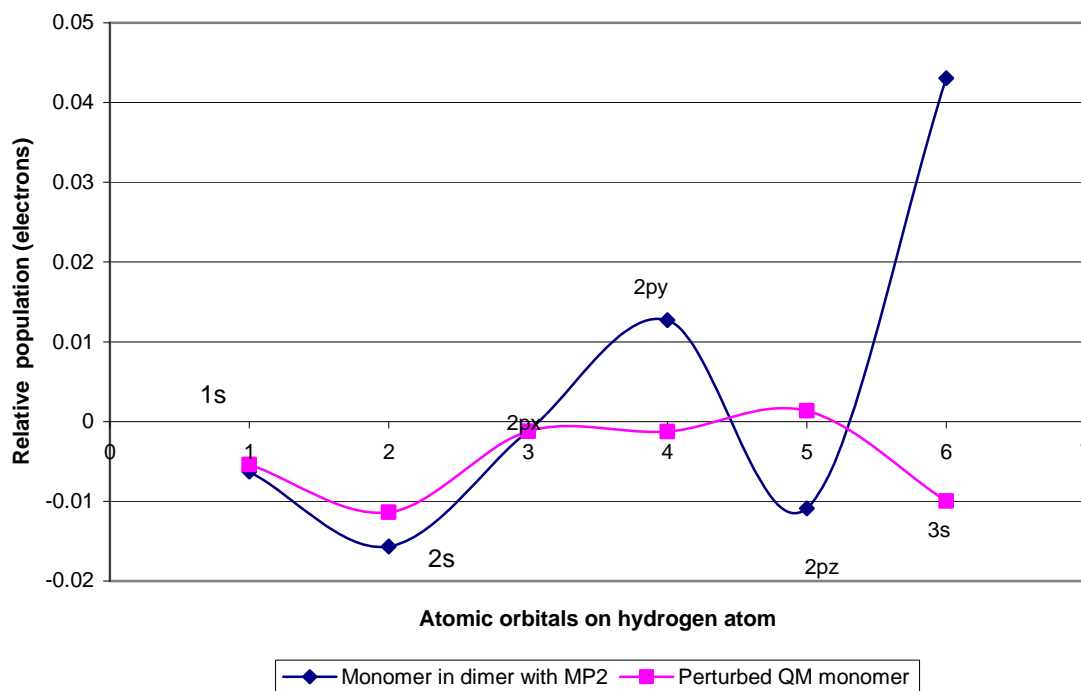


Fig. 9.3: Graph showing the difference in the relative electron populations of the electrons in each orbital for the hydrogen atom in the QM system. The orbitals are numbered in the sequence of their main quantum numbers. The graph therefore starts with the 1s-orbital and ends with the 3s-orbital.

Regarding both Fig. 9.2 and 9.3, we see that the perturbation of the point charges on the QM system is inaccurate compared to a pure MP2 calculation of the HF dimer. When we consider that the decrease in relative electron populations is related to the increase in the energy of the energy levels of the orbitals,<sup>2</sup> we can make the following assumption: if an atomic orbital is highly occupied it means that it is stable and if it is not, then it is unstable. For example, if we find a positive relative electron population for an atomic orbital then it means that this orbital is lower in energy and hence more stable than in the monomer.

If the wave function was perturbed “correctly” by the point charges of the MM system, the relative electron populations for the QM system and for the monomer in the MP2 dimer should be similar. However, they definitely differ significantly, as can be seen in Fig. 9.2 and Fig 9.3 for the fluorine and hydrogen atoms respectively.

In Fig. 9.2 we see that the  $2p_z$ ,  $3p_z$ ,  $4p_z$  and  $5p_z$ -orbitals show a similar trend in the hybrid system to the MP2 dimer. For the fluorine atom, we see a different trend in the relative electron populations for the orbitals with a maximum electron density in the y-direction with respect to the same orbitals of a monomer in the MP2 dimer. These orbitals are the  $p_y$ -orbitals.

<sup>2</sup> This can also be seen when the energy values for the orbitals are compared to an isolated monomer. The orbitals are in general higher in energy than in the monomer.

In the MP2 dimer these orbitals clearly show a destabilization, but we do not see this destabilization for the monomer in the QM/MM system. Putting negative charges in line with the lobes of the  $p_y$ -orbitals on the fluorine atom in the QM system will result in the movement of more electrons into the  $p_z$ -orbitals and so the perturbation of the QM system will be more like in the MP2 dimer. If the electrons move into the  $p_z$ -orbitals, they will be repelled by the negative point charge on the fluorine atom in the MM system. This will result in the electrons mainly spending time as far away as possible from the point charge on the fluorine atom in the MM system, which will lead to a larger bulging of the electron cloud on the right of the fluorine atom in the QM system. From Fig. 9.2 this also seems to be the case for the fluorine atom in the MP2 dimer.<sup>3</sup>

We can now give a possible explanation for the ONIOM electrostatic energy being unrealistically large in absolute value. The reason might be that because there are more electrons in the  $y$ -direction for the fluorine atom in the QM system than in the MP2 dimer, these electrons are closer to the positive point charge on the hydrogen atom in the MM system. The attractive interaction between these electrons and the positive charge can therefore lead to an unrealistic stabilization energy. If there were more electrons in the  $z$ -direction then there would have been a larger repulsive interaction between the negative point charge on the fluorine atom in the MM system and the electrons, which will lead to a destabilization. This in turn might lead to a more realistic ONIOM electrostatic energy value.

For the hydrogen atom, we see that the relative electron populations do not follow the same trend as in the MP2 dimer. We see that the  $2p_y$ ,  $2p_z$  and  $3s$ -orbitals on the hydrogen atom have different electron populations in the hybrid system to the MP2 dimer, but other than for the fluorine atom, they are more comparable and a secondary axis is not necessary when presenting these values graphically in Fig. 9.3. The difference in the relative electron populations with the  $2p_y$ -orbitals of the hydrogen atom show that two positive point charges should be added above and beneath the hydrogen in the QM system in order to attract electrons towards these orbitals. The magnitude of these charges should be small, as the correction that should be applied to the QM/MM system to make the hydrogen atom behave more as it does in the monomer in the MP2 dimer, is very subtle.

A possible solution to correct for both the relative electron population for the fluorine and the hydrogen atom is to add a quadrupole to the MM system. In Fig. 9.4 we show the two atomic point charges and the quadrupole that need to be added to the MM system. It is important that all the point charges on the MM system should be constrained so that the net charge of zero for the entire system is conserved.

---

<sup>3</sup> We see that in the MP2 dimer the fluorine atom in the monomer has a larger stabilization in the  $z$ -direction than the fluorine atom in the monomer in the QM system. This can only mean that more electrons are found on the right of the fluorine atom in the MP2 dimer.

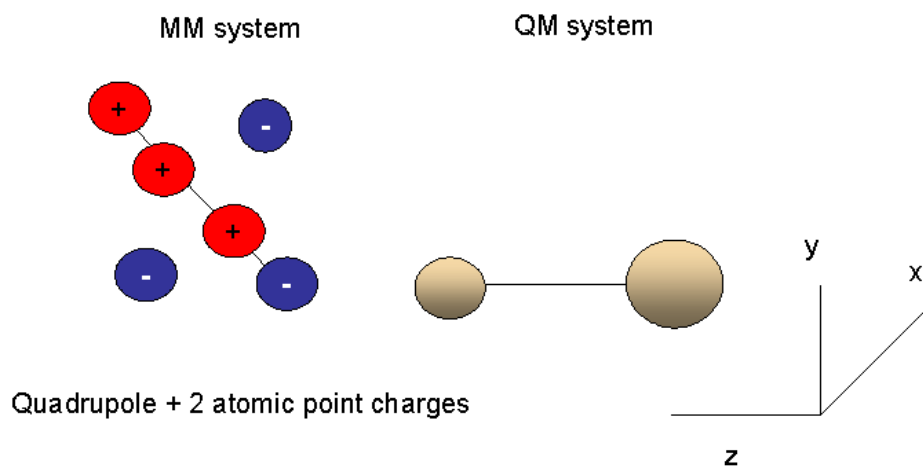


Fig. 9.4: Superposition of 2 monopoles and a quadrupole in the MM system to improve the polarization of the QM wave function.

The two negative point charges in the quadrupole should be chosen to have a significant repulsive effect on the electrons in the  $p_y$ -orbitals on the fluorine atom in the QM system. The positive charges of the quadrupole should be small in magnitude, as they will attract electrons into the  $p_y$ -orbitals of the hydrogen atom.

We have deliberately said nothing about the large difference in the relative electron populations for the  $3s$ -orbital on the hydrogen atom in the QM system compared to that of this atom in the MP2 dimer. The reason why this difference is obtained is due to charge transfer. In the MP2 dimer, charge is transferred from the fluorine atom in the donating monomer to the hydrogen atom in the accepting monomer, but in the QM/MM system this is of course not possible. There would also be a certain amount of energy associated with this charge transfer, but we believe that this might not have such a large effect on the ONIOM electrostatic energy.

Based on our results in this section we conclude that there are possibly two reasons for the absolute values of the ONIOM electrostatic energy being so large. The reasons are:

1. The polarization of the wave function cannot be accurate without using a better charge density description for the MM system, such as using a quadrupole in addition to the two atomic point charges;
2. Charge transfer cannot be modeled by QM/MM systems and the lack of proper charge transfer will have an effect on the electron density in the QM system.



### 9.4 Problem with the ESP derived charges

When we compare the MKS charges that are derived to represent the QM system's charge density for Cluster IV of the HF clusters during the first set of micro-iterations, we see that we obtain a zero total charge for the QM system when taking the sum of the derived charges. See Table 9.2. However, when we add all the charges for the entire system, including the point charges on the MM atoms, we do not obtain a total charge of zero for the system. Our results were obtained by performing ONIOM and MP2 single point calculations.

Table 9.2: Results for the point charges derived for the QM system (F1 and H2) and the true MKS charges for these atoms in the same cluster calculated with MP2. The charges are given in au.

	<b>MKS (ONIOM)</b>	<b>MKS (true)</b>
<b>F1</b>	<b>-0.488</b>	-0.430
<b>H2</b>	<b>0.488</b>	0.449
F3	-0.451	-0.451
H4	0.437	0.437
F5	-0.419	-0.419
H6	0.441	0.441
H7	0.430	0.430
F8	-0.457	-0.457
Total:	-0.019	0.000

Therefore, according to our results in Table 9.2, when ESP charges are derived the system is not neutral anymore. We suggest that the constraint on the charges should rather be that the total charge of the system should be conserved. We do not know what effect this might have on the optimization of the MM system but it might be subtle. Due to the particular way in which the charges for the QM system are derived, one should not expect by default that the charges are a good representation of the true charge density of the QM system.

## 9.5 ONIOM-EE optimizations with rigid blocks

The micro-iterations for ONIOM-EE optimizations are currently done by using the GDIIS method and converting to the Conjugate Gradient (CG) method when the forces are very small [Vreven *et al.*, 2003]. In **Section 2.9** we described how a geometry optimization is performed for a QM/MM system. We will briefly recap the important details. Electronic embedding is currently done by adding the interaction of the point charges on the MM atoms to the QM Hamiltonian. This leads to an energy, which is then used to compute the exact electrostatic forces between the QM and MM systems. The sum of this and the force for the van der Waals interactions between all the nonbonded atoms in the entire system is then used when the force is computed during the micro-iterations. During the micro-iterations time is saved by instead of computing the exact force at every step, the change in the exact force is rather updated by an approximation based on the change in the interaction between the ESP derived charges and the stationary charges on the MM atoms as the atoms in the MM system move.

Having a very large negative electrostatic energy for the interaction between the QM and MM systems will also lead to a large electrostatic force, since the magnitude of the force in atomic units is just the energy in hartree divided by the distance in bohr between an atom in the QM system and an atom in the MM system. Approximating the change in the exact electrostatic force between the QM and MM systems by using point charges will result in a far smaller change in the exact force than should be the case. This will then lead to the MM system experiencing a large attractive force pulling it towards the QM system that will not change much as the MM atoms move. Using rigid blocks in the MM system will make the optimization more difficult, as the atoms of a monomer are not allowed to turn around the center of mass of the monomers or stretch their bond lengths. The only course for the MM system to follow is to plunge into the QM system. This would eventually lead to an enormous increase in the van der Waals interaction, which will make the QM and MM systems move as far away as possible from each other. The optimizer will break down since GDIIS is known to behave erratically when far from a minimum [Schlegel, 2003]. In **Appendix B.2** we illustrate this behavior for Cluster V of the HF clusters. When we removed the rigid block constraint, for some of the clusters, the QM and MM systems did not show this behavior, but the bond lengths in the MM system definitely stretched during the optimization.

A possible way to freeze internal coordinates in the MM system might be to use penalty functions [Leach, 2001], such as is done to freeze bond lengths, angles and torsion angles in normal restrained force field optimizations, however this has not been implemented in *Gaussian 03*. We however expect that by using penalty functions geometry optimizations,

where the user specifically wants to restrain the MM molecules' bond lengths, will be more efficient.

The reader might have noticed that so far we did not mention the CO/HF clusters. The problem with the CO/HF clusters was already encountered when point charges were derived earlier in **Chapter 8**. The charge schemes used gave erroneous results such as positive charges for both a carbon and a hydrogen atom that are known to both participate in a hydrogen bond. Using these charges in QM/MM calculations would have been a waste of time. This emphasizes the need for more accurate charge derivation schemes and the use of higher multipoles in charge density approximations.

## ***9.6 QM-QM interactions vs QM-MM interactions***

Previously in **Chapter 5** we used two different hybrid systems in an attempt to minimize the time needed for the optimization of the force field. By using two QM molecules and only one MM molecule in the HF trimer, we could determine the difference in the hydrogen bond distances in the hybrid system with respect to the MP2 results. We found that adding an MM molecule to a QM system weakened all of the hydrogen bonds in the system, resulting in a distortion of the geometry. MM point charges and van der Waals forces were therefore not adequate to give the correct hydrogen bond distances between the QM molecules in the QM system. Studying the difference in the hydrogen bond distances between two QM atoms and the hydrogen bond distance between a QM and MM atom in such a hybrid system, we found that the QM-QM hydrogen bond distance is shorter than the QM-MM hydrogen bond distance. We also found that in comparing the results for the optimization of the force field for the hybrid system containing only one MM molecule with the hybrid system containing one QM molecule, the  $R_0$ -values for H<sub>1</sub> and F<sub>1</sub> are 3.52 Å and 1.28 Å and 3.34 Å and 1.28 Å for the two systems respectively. The reason for this result is probably that the interaction between an MM fluorine atom and a QM hydrogen atom is more repulsive than between two MM atoms. This is related to an inaccurate description of the electrostatic interactions due to an inaccurate polarization of the QM wave function by the atomic point charges on the MM atoms, and also due to charge transfer effects from the QM to the MM system that of course cannot be described by molecular mechanics.

## 9.7 A possible solution for problems encountered with ONIOM-EE

We showed that the use of point charges could lead to errors and even optimizations failing. In this section, we suggest a method to improve the present ONIOM-EE method to obtain results for clusters that are more accurate. For this method to be successful, it should be able to:

1. Produce a more accurate perturbation on the QM Hamiltonian;
2. Incorporate multipole moments as far as possible;
3. Allow the MM system to be polarized by the QM system as well;
4. Keep the charge of the total system constant.

### 9.7.1 Presenting the anisotropy of a charge distribution

Before we discuss a possible solution to the problem, we need to review methods that are able to give more accurate representations of the charge-density. One of these methods is the central multipole expansion.

#### 9.7.1.1 The central multipole expansion

The central multipole expansion is based on the electric moments or multipoles of molecules. The multipole expansion contains the monopole or charge, the dipole, the quadrupole, the octupole etc. By using an expansion of all of these moments to a specific order, the anisotropy of a system's charge distribution can be accurately accounted for. A dipole can be represented by placing two charges an appropriate distance apart, a quadrupole by placing four charges appropriate distances apart, an octupole by using eight charges and so on.

Electrical multipole moments higher than the dipole are all tensor properties. A tensor is useful when the  $x$ -component of one vector influences not only the  $x$ -component of another vector, but also its  $y$  and  $z$  components and when the  $y$ -component of one vector influences not only the  $y$ -component of another vector but also its  $x$  and  $z$  components. The same applies to the  $z$ -component. Tensors are not matrices, but quantities that can be represented by appropriate matrices. The quadrupole tensor can be written as:

$$\Theta = \begin{pmatrix} \sum q_i x_i^2 & \sum q_i x_i y_i & \sum q_i x_i z_i \\ \sum q_i y_i x_i & \sum q_i y_i^2 & \sum q_i y_i z_i \\ \sum q_i z_i x_i & \sum q_i z_i y_i & \sum q_i z_i^2 \end{pmatrix} \quad (9.2)$$

where  $q_i$  is the charge and  $x_i, y_i$  and  $z_i$  are the coordinates [Leach, 2001].

#### 9.7.1.2 Distributed Multipole Analysis (DMA)

The well-known DMA method [Leach, 2001] calculates the multipoles from a quantum mechanical wave function in terms of Gaussian primitives. As we mentioned in **Chapter 2**, in general *ab initio* calculations use basis functions that are centered on atoms. DMA uses the product of basis functions that corresponds to a charge density at the atoms. This charge density is written as a multipole expansion at the site of the atoms. Then the multipole expansion for the same charge density at the atoms is expanded at various sites around the molecule leading to the creation of multipole moments at these sites. These multipole moments automatically incorporate the anisotropy of the charge distribution. It is also possible to obtain only the multipole moments of the charge density at the atomic sites by making use of AIM. AIM can be used to calculate the higher multipole moments on atoms based on an integration of the electron density over atomic basins.

#### 9.7.2 A new method for ONIOM-EE optimizations of vdW clusters

All the details regarding how ONIOM-EE calculations and geometry optimizations are performed in *Gaussian 03* can be found in **Section 2.9** and will not be repeated. We would like to suggest modifications to this scheme:

1. Instead of using an atomic point charge perturbation, it would be better to use many point charges on selected sites. These point charges can be incorporated as dummy atoms to represent dipoles, quadrupoles and higher multipoles for the molecules in the MM system. Instead of using only atomic point charges the perturbation on the QM wave function will now include multipole contributions, making it more accurate.
2. Instead of using the exact force in the micro-iterations, the force can be approximated by deriving point charges for the atoms or on other selected sites for the QM system. If the QM system is large, fast multipole methods (FMM) can be used to incorporate the electrostatic force in the micro-iterations.
3. In order to create a charge polarization of the QM system's charges on the MM system, we can use an approximation. Instead of only fitting charges to the QM system based on the electron density, it would also be possible to fit charges on the

atoms and sites in the MM system to give the same ESP at points between the MM and QM systems. The ESP is related to the electric field by a simple equation:

$$\frac{dESP(\mathbf{r})}{d\mathbf{r}} = -E(\mathbf{r}) \quad (9.3)$$

where  $E$  is the electric field at coordinate  $\mathbf{r}$ . The charges on the MM and the QM system should reproduce the same electric field at the selected points, which would lead to a resultant electric field of zero. The sites selected for point charges on each molecule can be on the atoms themselves, positioned an appropriate distance above or beneath the bonds, or they can be positioned on the bonds. The charges should be fitted to a quantum mechanically computed ESP for each molecule. For very large MM systems, such as in biological systems, fitting ESP charges to a large MM system used to be very difficult, but recently a method was developed whereby this could be done [Gascon *et al.*, 2006]. For sites not on atoms, dummy atoms need to be used. The addition of these dummy atoms can be done by means of a graphical user interface (GUI). In Table 9.3 we summarize the possible advantages and disadvantages of our proposed new method.

Table 9.3: Possible advantages and disadvantages of new method

Advantages	Disadvantages
Easy to use	Wrong determination of sites for charges can lead to the condition where some molecules are well described electrostatically and others not
Should be simple to program and incorporate in the Gaussian source code	Possible longer calculation times
The method can make use of FMM methods to incorporate the electrostatic perturbation of the QM on the MM as the MM moves with respect to the QM.	It might be that a new optimization algorithm will have to be designed for the method
Polarization of the QM on the MM and vice versa is consistent	The method will probably not be applicable when a very large molecule is used as the MM system

### 9.7.3 The problem of charge transfer

In HF clusters we found that charge transfer plays an important role. We already touched briefly on charge transfer in **Section 9.2**. To simulate charge transfer in a QM/MM system might be difficult and therefore we suggest rather using a QM/QM method when charge transfer is significant. There have been attempts to simulate charge transfer for QM/MM systems. One method is the *Generalized Hybrid Orbital* (GHO) method [Pu *et al.*, 2004] mentioned in **Chapter 2**, where an  $sp^3$ -hybrid orbital is used for each MM atom at the boundary. The QM system can then interact with this hybrid orbital and charge transfer can

take place. This method was designed as a substitute to link atoms for systems where covalent bonds cross the barrier between the MM and QM systems, but to our knowledge this and similar ideas have not been applied to nonbonded QM/MM systems.

## ***9.8 Summary, conclusions and future work***

In this chapter we emphasized various problems with the present way in which the ONIOM-EE method is incorporated in *Gaussian 03*, for the modeling of clusters. The problems and possible solutions can be summarized as follows:

**Problem:** Point charges on atoms are not a good description of the charge distribution.

**Solution:** When deriving point charges for a charge distribution we suggest that multipoles should also be derived. As we have seen, an addition of multipoles might give a better description of the charge density of a system.

**Problem:** The electronic perturbation is not very accurate.

**Solution:** We have seen for the HF dimer at least that using multipoles in terms of many point charges for the description of the MM system might solve this problem.

**Problem:** Because it is accepted by *Gaussian 03* that only the QM system should have a zero charge, one sometimes obtains weakly charged systems. This is of course unnatural and chemically incorrect. See Table 9.2.

**Solution:** The charge constraint should be on the whole system, not just the QM system. The QM system is not necessarily neutral.

**Problem:** The QM system is polarized by the MM system, but not vice versa, leading to incorrect interactions between the MM and QM systems.

**Solution:** The charges on the MM system should be modified in relation to the charges that are derived for the QM system. See **Section 9.6.2**.

**Problem:** Micro-iterations, when using electronic embedding, converge extremely slowly when the molecules in the MM system (with frozen bond lengths) move significantly with respect to the QM system. Under certain conditions, the optimizer can also behave erratically.

**Solution:** Instead of adding possible extra coordinates to the MM system, one could use penalty functions, such as are used to constrain bond lengths and other internal coordinates in general force field optimizations.

## CHAPTER 10

# Final conclusions and a summary of possible future work

### *10.1 Introduction*

In this chapter, we will summarize our main conclusions in **Section 10.2** and possible future work that might lead to interesting research, in **Section 10.3**.

### *10.2 Main conclusions and brief summary*

1. Reviewing the current literature made it clear that CO and CO/HF clusters are more of a challenge to model computationally than HF clusters. According to the literature reviewed, very little is known about CO and CO/HF clusters experimentally as well as computationally. The literature emphasizes that large basis sets and high levels of theory should be used when modeling vdW clusters, as it is only possible to account properly for electron correlation in this way. It was also clear from the literature that the carbon monoxide dimers are extremely challenging to model quantum mechanically.
2. We showed that using MP2, utilizing the 6-311+G(d,p) basis set, gives reasonably good results for the HF, CO and HF/CO clusters. However, if the aim is to obtain very accurate energies, we advise utilizing both a higher level of theory and a larger basis set.
3. In **Chapters 4, 6 and 8** we illustrated that because AIM analysis can only give a 3-dimensional analysis of the electron density topology, and NBO analysis is based on a single determinant wave function, it is better to use visualizations of the correlated many determinant wave function. This in turn leads to the definition of *intermolecular orbitals*. We showed that *intermolecular orbitals* could qualitatively explain hydrogen bonding in HF and CO/HF clusters and vdW interactions in CO clusters.
4. We gave a possible explanation for the weak ring current in the HF trimer in terms of the delocalized *intermolecular orbital* model and the arrangement of two shells



in a monomer in the trimer that are possibly in close proximity to each other in energy.

5. We showed that different charge schemes can give totally different charges both in magnitude and in sign for the carbon and oxygen atoms in CO clusters.
6. For the CO/HF clusters we showed that charge schemes can give erroneous charges. For example, for some charge schemes tested, positive charges were put on both a hydrogen and a carbon atom, whereas we know from the literature that the C...H bonded dimer is a global energy minimum. We suggested that this behavior was the result of trying to approximate the anisotropic charge density by using atomic centered point charges instead of incorporating higher multipoles.
7. Reviewing the literature on hybrid methods made it clear that a study such as the one described in this work, has to the best of our knowledge, never been performed before. Hybrid methods are usually used for large systems such as those of a biological nature and it seems that none of the clusters modeled in this work have been modeled with any hybrid method before.
8. Optimizing a force field by trial geometry optimizations while the bond lengths in the MM system are frozen by utilizing rigid blocks, results in a very slow convergence if the parameters are too far from being optimized. Depending on the parameters, such a convergence might take days even for relatively small systems. This is due to the way in which the MM bond lengths are frozen. Using rigid blocks also has the disadvantage that the atoms in a monomer cannot rotate around their center of mass. A better method might be to use penalty functions.
9. An alternative method for optimizing force field parameters was developed. The necessity for this method arose out of two observations: the energies for the perturbations of the wave functions of the clusters in the ONIOM-EE calculations were not comparable to the MP2 interaction energies; using atom-atom contacts was not a good way of assessing the force field quality for the CO clusters. The frozen geometry method we developed in **Chapters 5, 7 and 8** appears promising in this regard. A genetic algorithm is ideal for this case in order to find a global minimum.
10. Another method of force field optimization might be to relax the bond lengths of the monomers in the MM system during a geometry optimization. However, then the force constants for the bond stretches will probably also need to be reoptimized. Nevertheless, the geometry optimizations will be much faster in this case and convergence can be established in a much shorter time.
11. We could not find parameters that fit the MP2 results exactly, but for both the hydrogen fluoride and carbon monoxide clusters, we succeeded in obtaining a good

fit of the ONIOM steric energies to the MP2 BSSE corrected interaction energies for the HF and CO clusters. The force field could not be fully optimized for the CO/HF clusters.

12. We established that the main reason for the difficulty in optimizing the force field parameters for the CO/HF clusters by our frozen geometry method was the wrong signs of the atomic point charges that were used on the MM atoms. We showed that in some cases it is impossible to optimize the van der Waals parameters so that the positive electrostatic energy can be counterbalanced to give the MP2 interaction energy. None of the charge schemes tested derived the correct charges for the C $\cdots$ H bonded CO/HF cluster.
13. We showed for the HF dimer that approximating the charge distribution of the molecules in the MM system by point charges only can give an inaccurate perturbation on the QM Hamiltonian. We also showed that the MKS charges derived from the ESP based on the perturbed wave function were not always efficient enough to be used to approximate the change in the exact electrostatic force between the QM and MM systems during the micro-iterations.
14. We showed that in *Gaussian 03* weakly charged systems result from the fact that only the QM system was allowed to have a zero charge, instead of the whole system including the MM atoms.
15. We suggested a method to improve the present way in which ONIOM-EE is performed to make it more suitable for the modeling of vdW clusters. This method is described in detail in **Chapter 9**.
16. We believe that it might be easier and more accurate to use a QM/QM method with electronic or even mechanical embedding rather than a QM/MM method when accurate modeling of clusters, as described in this work, is essential. To accomplish this, one would have to determine a specific lower level QM system that could function together with a higher level QM system by using the S-value method mentioned in **Chapter 2**. The beauty of a QM/QM system is that no force field parameters need to be optimized for a specific system and it can be used in a more general way than QM/MM methods.
17. The intermolecular delocalized orbital model might be more suitable for the description of van der Waals and hydrogen bonds than current models and would lead to a unified theory of intermolecular interactions. This model for hydrogen bonds and van der Waals bonds is given in **Chapters 4, 6 and 8**.

### 10.3 Future work

Based on our studies, there is still a large scope for work in both the fields of vdW clusters and hybrid system development. In this section we will give a list of ideas that could be pursued.

1. Before accepting the intermolecular orbital description for intermolecular interactions, various other systems should be tested, such as rare-gas systems, systems with improper hydrogen bonds and even systems containing dihydrogen bonding. One can determine the relative electron populations for the delocalized orbitals at various separations of the monomers and see how the relative electron populations change. Instead of only viewing the change in electron density with respect to the separation between monomers, delocalized molecular orbitals can also be viewed at various separations of the monomers.
2. CO/HF clusters are good models for understanding intermolecular interactions in general, as both hydrogen and van der Waals bonds are present in these clusters. One could attempt to study these clusters in more depth.
3. The explanation for the ring current in the HF trimer should be tested on other systems to ascertain its validity. These systems could include the HBr trimer, for example.
4. The optimization of the force field for ONIOM-EE by the frozen geometry method, developed in this work, should be validated to see if it is adequate for applying to other systems as well.
5. The method mentioned in **Chapter 9** to improve ONIOM-EE optimizations for vdW clusters could be used as a starting point for a modification to the *Gaussian 03* or *Gaussian98* source codes to incorporate better charge density descriptions for the MM system in hybrid calculations.
6. Research could be done into incorporating a polarizable force field for the MM system in the hybrid system, together with the method mentioned above.
7. A vibrational analysis of the CO hexamer is still lacking due to the hardware limitations of our system. It might be possible to do this calculation by using nodes in parallel. We did not do this in this work, as MP2 is not optimized in *Gaussian 03* to make use of parallel computing. One could also use a system with more memory.
8. Force fields for other clusters modeled by QM/MM, such as H<sub>2</sub>O, NH<sub>3</sub>, BH<sub>3</sub> and larger molecules, could be optimized.

9. Different charge schemes could be tested for the derivation of point charges for systems where there is more electron localization (such as CO clusters) to more delocalization (such as HF clusters).
10. The *Gaussian 03* and *Gaussian98* programs could be modified to make periodic boundary conditions (PBC) possible for QM/MM calculations.
11. The hybrid optimizations in this work can be repeated by utilizing QM/QM systems. This will probably lead to a more accurate modeling of the clusters than can be obtained with a QM/MM system. For crystals and other periodic systems, PBC should be applied.

## References

- Abu-Awwad, F.M., (2002), *Chem. Phys. Lett.*, **360**, 340.
- Acevedo, A.J., Caballero, L.M., López, G.E., (1997), *J. Chem. Phys.*, **106**(17), 7257.
- Andersson, Y., Langreth, D.C., Lundqvist, B.I., (1996), *Phys. Rev. Lett.*, **76**(1), 102.
- Askarpour, V., Kieft, H., Clouter, M.J., (1989), *J. Chem. Phys.*, **90**(12), 7014.
- Bakowies, D., Thiel, W., (1996), *J. Phys. Chem.*, **100**, 10580.
- Bernard, S., Chiarotti, G.L., Scandolo, S., Tosatti, E., (1998), *Phys. Rev. Lett.*, **81**(10), 2092.
- Biegler-König, F.W., Bader, R.F.W., Tang, T-H., (1982), *J. Comput. Chem.*, **3**, 317.
- Bochenkova, A.V., Suhm, M.A., Granovsky, A.A., Nemukhin, A.V., (2004), *J. Chem. Phys.* **120**(8), 3732.
- Boys, S.F., Bernardi, F., (1970), *Mol. Phys.*, **19**, 553.
- Breneman, C. M., Wiberg, K. B., (1990), *J. Comput. Chem.*, **11**, 361.
- Bressanini, D., Zavaglia, M., Mella, M., Morosi, G., (2000), *J. Chem. Phys.*, **112**(2), 717.
- Carmichael, M., Chenoweth, K., Dykstra, C.E., (2004), *J. Phys. Chem. A*, **108**(15), 3143.
- Casadesús, R., Moreno, M., González-Lafont, Á., Lluch, J.M., Repasky, M.P., (2003), *J. Comput. Chem.*, **25**, 99.
- Chałasiński, G., Szcześniak, M.M., (1994), *Chem. Rev.*, **94**, 1723.
- Chałasiński, G., Szcześniak M.M., (2000), *Chem. Rev.*, **100**, 4227.
- Chandler, W.D., Johnson, K.E., Campbell, J.L.E., (1995), *Inorg. Chem.*, **34**, 4943.
- Collins, C.L., Morihashi, K., Yamaguchi, Y., Schaefer III, H.F., (1995), *J. Chem. Phys.*, **103**(14), 6051.
- Cornell, W.D., Ciplak, P., Bayly, C.I., Gould, I.R., Merz Jr., K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A., (1995), *J. Am. Chem. Soc.*, **117**, 5179.
- Cotton, F.A., (1990), *Chemical applications of group theory*, 3<sup>rd</sup> Edition, John Wiley and Sons:New York.
- Cui, Q., Karplus, M., (2000), *J. Chem. Phys.*, **112**(3), 1133.
- Cunha and Smith, (2005), Observatorio Nacional, Rio de Janeiro, Brazil Los Alamos National Laboratory, Preprint Archive, Astrophysics.
- Curtiss, L.A., Pochatko, D.J., Reed, A.E., Weinhold, F., (1985), *J. Chem. Phys.*, **82**(6), 2679.
- Dapprich, S., Komáromi, I., Byun, K.S., Morokuma, K., Frisch, M.J., (1999), *J. Mol. Struct.(Theochem)*, **461**, 1.
- Davidson, E.R., (1996), *Chem. Phys. Lett.*, **260**, 514.
- Dennington II, R., Keith, T., Millam, J., Eppinnett, K., (2003), *GaussView Version 3.09*, Semichem, Inc., Shawnee Mission, KS.

- Hovell, W.L., Gilliland, R., Desiraju, G.R., Parthasarathy, R., (1989), *J. Am. Chem. Soc.*, **111**, 8725.
- Doublerly, G.E., Miller, R.E., (2003), *J. Phys. Chem. B*, **107**, 4500.
- Dunitz, J.D., Gavezzotti, A., (2005), *Angew. Chem. Int. Ed.*, **44**, 1766.
- Dunning Jr, T.H., 1989, *J. Chem. Phys.*, **90**, 1007.
- Dunning Jr, T.H., Peterson, K.A., Woon, D.E., (1998), *Basis sets: Correlation Consistent Sets*, In *Encyclopedia of Computational Chemistry*, Schleyer, P.V.R., Allinger, N.L., Clark, T., Gasteiger, J., Kollman, P.A., Schaefer III, H.F., Schreiner, P.R., Eds., John Wiley & Sons: Chichester, UK, **1**, 88.
- Espinosa, E., Alkorta, I., Elguoro, J., Molins, E., (2002), *J. Chem. Phys.*, **117**(12), 5529.
- Evangelisti, S., (1997), *Chem. Phys.*, **218**, 21.
- Foulkes, W.M.C., Mitas, L., Needs, R. J., Rajagopal, G., (2001), *Rev. Mod. Phys.*, **73**(1), 33.
- Fraser, G.T., Pine, A.S., (1988), *J. Chem. Phys.*, **88**(7), 4147.
- Freindorf, M., Shao, Y., Furlani, T.R., Kong, J., (2005), *J. Comput. Chem.*, **26**, 1270.
- Friesner, R.A., Guallar, V., (2005), *Annu. Rev. Phys. Chem.*, **56**, 389.
- Frisch, M.J., Pople, J.A., Binkley, J.S., (1984), *J. Chem. Phys.*, **80**, 3265.
- Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, Jr., J.A., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J.B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M. C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe M., Gill, P.M.W., Johnson B., Chen, W., Wong, M.W., Gonzalez, C., Pople J. A., (2003), Gaussian 03, Revision B.05, Gaussian, Inc., Pittsburgh PA.

- Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Zakrewski, V.G., Montgomery Jr., J.A., Stratmann, R.E., Burant, J.C., Dapprich, S., Millam, J.M., Daniels, A.D., Kudin, K.N., Strain, M.C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G.A., Ayala, P.Y., Cui, Q., Morokuma, K., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Cioslowski, J., Otiz, J.V., Baboul, A.G., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P.M.W., Johnson, B.G., Chen, W., Wong, M.W., Andres, J.L., Head-Gordon, M., Replogle, E.S.W., Pople, J.A., (1998), *Gaussian98 Revision A.7.*, Gaussian, Inc., Pittsburgh PA.
- Froudakis, G.E., (2001), *Nano Letters*, **1**, 179.
- Gálvez, O., Gómez, P.C., Pacios, L.F., (2001), *J. Chem. Phys.* **115**(24), 11166.
- Gascon, J.A., Batista, V.S., (2004), *Biophys. J.* , **87**, 2931.
- Gascon, J.A., Leung, S.S.F, Batista, E.R., Batista, V.S., (2006), *J. Chem. Theory Comput.*, **2**(1), 175.
- Glendening, J. K., Reed, J. E., Carpenter, J.A., and F. Weinhold, *NBO 3.0.*, (2001), E.D. Theoretical Chemistry Institute, University of Wisconsin, Madison.
- Golebiowski, J., Lamare, V., Martins-Costa, M.T.C., (2001), *Chem. Phys.*, **272**, 47.
- Guedes, R.C., do Couto, P.C., Costa Cabral, B.J., (2003), *J. Chem. Phys.*, **118**(3), 1272.
- Hagler, A.R., Wilson, S., (1974), *Acta Crystallogr.*, **B30**, 1336.
- Han, H.S., Kim, K., (1997), *J. Mol. Struct.(Theochem)*, **418**, 1.
- Harcourt, R.D., (1999), *J. Phys. Chem. A*, **103**, 4293.
- Havenith, M., (2001), *Infrared spectroscopy of molecular clusters: an introduction to intermolecular forces*, Springer:New-York.
- Hehre, W.J., Ditchfield, R., Pople, J.A., (1969), *J. Chem. Phys.*, **51**, 2657.
- Hodges M.P., Stone, A.J., Lago, E.C., (1998), *J. Phys. Chem. A*, **102**, 2455.
- Hohenberg P., Kohn, W., (1964), *Phys. Rev.*, **136**, B864.
- Hopkins, B.W., Tschumper, G.S., (2004), *J. Phys. Chem. A*, **108**, 2941.
- Huber, K.P., Herzberg, G., (1979), *Molecular spectra and Molecular structure: Constants of Diatomic Molecules*, Van Nastrand: New York.
- Humbel, S., Sieber, S., Morokuma, K., (1996), *J. Chem. Phys.*, **105**(5), 1959.
- Izgorodina, E.I., Coote, M.L., (2006), *J. Phys. Chem. A.*, **110**, 2486.
- Jansen, G., Hesselmann, A., (2001), *J. Phys. Chem. A.*, **105**, 11156.
- Jensen, F., (2001), *Introduction to Computational Chemistry*, John Wiley & Sons, Chichester:UK.

- Jensen, L., van Duijnen, P.Th., Snijders, J.G., (2003), *J. Chem. Phys.*, **118**(2), 514.
- Jiao, H., Frapper, G., Halet, J.-F., Saillard, J.-Y., (2001), *J. Phys. Chem. A.*, **105**, 5945.
- Kaminski, G.A., Stern, H.A., Berne, B.J., Friesner, R.A., Cao, Y.X., Murphy, R.B., Zhou, R., Halgren, T.A., (2002), *J. Comput. Chem.*, **23**, 1515.
- Kamiya, M., Tsuneda, T., Hirao, K., (2002), *J. Chem. Phys.*, **117**(13), 6010.
- Karasawa, N., Dasgupta, S., Goddard, W.A., (1991), *J. Phys. Chem.*, **95**, 2260.
- Kanakarajan, K., Czarnik, A.W., (1986), *J. Org. Chem.*, **51**(26), 5241-3.
- Kendall, R.A., Dunning Jr., T.H., Harrison, R.J., (1992), *J. Chem. Phys.*, **100**, 7410.
- Kerdcharoen, T., Morokuma, K., (2002), *Chem. Phys. Lett.*, **355**, 257.
- Knowles, P., Schütz, M., Werner, H.-J., (2000), *Ab initio methods for electron correlation in molecules*, In *Modern Methods and Algorithms of Quantum Chemistry 2<sup>nd</sup> Edition*, Ed., J. Grotendorst, John von Neumann Institute for computing, **3**, 97.
- Kongsted, J., Osted, A., Mikkelsen, K.V., Christiansen, O., (2003), *J. Chem. Phys.*, **118**(4), 1620.
- Kongsted, J., Osted, A., Mikkelsen, K.V., Christiansen, O., (2003), *J. Phys. Chem. A.*, **107**, 2578.
- Krishnan, R., Binkley, J.S., Seeger, R., Pople, J.A., (1980), *J. Chem. Phys.*, **72**, 650.
- Kulkarni, A.D., Ganesh, V., Gadre, S.R., (2004), *J. Chem. Phys.*, **121**(11), 5043.
- Leach, A.R., (2001), *Molecular Modeling. Principles and Applications*, 2<sup>nd</sup> Edition, Prentice Hall, Harlow:England.
- Lehn, J.-M., (1978), *Pure Appl. Chem.*, **50**, 871.
- Leininger, M.L., Allen, W.D., Schaefer III, H.F., Sherrill, C.D., (2000), *J. Chem. Phys.*, **112**(21), 9213.
- Levine, I.N., (2000), *Quantum Chemistry*, 5<sup>th</sup> Edition, Prentice Hall, Englewood Cliffs, New Jersey:USA.
- Li, Q., Yin, P., Liu, Y., Tang, A.C., Zhang, H., Sun Y., (2003), *Chem. Phys. Lett.*, **375**, 470.
- Maerker, C., Schleyer, P.v.R., Liedl, K.R., Ha, T.-K., Quack, M., Suhm, M.A., (1997), *J. Comput. Chem.*, **18**(14), 1695.
- Magnasco, V., (2004), *Chem. Phys. Lett.*, **387**, 332.
- Martin, F., Zipse, H., (2005), *J. Comput. Chem.*, **26**, 97.
- Martín, M.E., Aguilar, M.A., Chalmet, S., Ruiz-López, M.F., (2002), *Chem. Phys.*, **284**, 607.
- Martins, J.B.L., Longo, E., Salmon, O.D.R., Espinoza, V.A.A., Taft, C.A., (2004), *Chem. Phys. Lett.*, **400**, 481.
- Masseras, F., Morokuma, K., (1995), *J. Comput. Chem.*, **16**(9), 1170.
- Mayo, S.L., Olafson, B.D., Goddard III, W.A., (1990), *J. Phys. Chem.*, **94**, 8897.
- McKellar, A.R.W., (2004), *J. Mol. Spec.*, **226**, 190.
- Meredith, A.W., Stone, A.J., (1998), *J. Phys. Chem. A*, **102**, 434.



- MOLEKEL 4.0, Flükiger, P., Lüthi, H.P., Portmann, S., Weber, J., (2000), Swiss Center for Scientific Computing, Manno (Switzerland).
- Molecular Simulations Inc., (1998), *Cerius<sup>2</sup>, Software Environment for Chemical Computing*, San Diego.
- Molecular Simulations Inc., (2000), WebLab ViewerLite 4.0.
- Møller, C., Plesset, M.S., (1934), *Phys. Rev.*, **46**, 618.
- Montero, L., Roque, R., Rosado, A., Fernandez B.J., (1981), *J. Mol. Struct., THEOCHEM*, **2**(3- 4), 393.
- Morokuma, K., Musaev, D.G., Vreven, T., Basch, H., Torrent, M., Khoroshun, D.V., (2001), *IBM J. Res & Dev.*, **45**, 367.
- Muenter, J.S., (1970), *J. Mol. spectrosc.*, **55**, 490.
- Muigg, D., Denifl, G., Stamatovic, A., Echt, O., Märk, T.D., (1998), *Chem. Phys.*, **239**, 409.
- Muñoz-Losa, A.I., Fdez.-Galván, M.E., Martín, M.E., Aguilar M.A., (2003), *J. Chem. Phys. B*, **107**, 5043.
- Murphy, R.B., Philipp, D.M., Friesner, R.A., (2000), *J. Comput. Chem.*, **21**(16), 1442.
- NIST Computational Chemistry Comparison and Benchmark Database, (2005),  
NIST Standard Reference Database Number 101, Release 12, Aug 2005, Editor:  
Russell D. Johnson III, <http://srdata.nist.gov/cccbdb>.
- Olsen, J., Christiansen, O., Koch, H., Jørgenson, P., (1996), *J. Chem. Phys.*, **105**(12), 5082.
- Osted A., Kongsted, J., Mikkelsen, K.V., Christiansen, O., (2004), *J. Phys. Chem. A*, **108**, 8646.
- Parr, R.G., Ayers, P.W., Nalewajski, R.F., (2005), *J. Phys. Chem. A*, **109**, 3957.
- Peterson, K.A, Woon, D.E., Dunning Jr., T.H., (1994), *J. Chem. Phys.*, **100**, 7410.
- Pine, A.S., Howard, B.J., (1986), *J. Chem. Phys.*, **84**(2), 590.
- Popelier P., (2000), *Atoms in Molecules. An introduction*, Prentice Hall, Harlow:England.
- Pu, J., Gao, J., Truhlar, D.G., (2004), *J. Phys. Chem. A*, **108**, 632.
- Quack, M., Stohner, J., Suhm, M.A., (2001), *J. Mol. Struct.*, **599**, 381.
- Quack, M., Suhm, M.A., (1995), *Chem. Phys. Lett.*, **234**(1-3), 71.
- Rappé, A.K., Goddard III, W.A., (1991), *J. Phys. Chem.*, **95**, 3358.
- Rappé, A.K., Casewit, C.J., Collwell, K.S., Goddard III, W.A., Skiff, W.M., (1992), *J. Am. Chem. Soc.*, **114**, 10024.
- Rega, N., Iyengar, S.S., Voth, G.A., Schlegel, H.B., Vreven, T., Frisch, M.J., (2004), *J. Phys. Chem. B.*, **108**, 4210.
- Rehaman, A., Datta, A., Mallajosyula, S.S., Pati, K.S., (2006), *J. Chem. Theory Comput.*, **2**, 30.
- Ricca, A., Drocco, J.A., (2002), *Chem. Phys. Lett.*, **362**, 217.
- Riccardi, D., Li, G., Cui, Q., (2004), *J. Phys. Chem. B.*, **108**, 6467.

- Rincón, L., Almeida, R., García-Aldea, D., Diez y Riega, H., (2001), *J. Chem. Phys.*, **114**(13), 5552.
- Rode, M., Sadlej, J., Moszynski, R., Wormer, P.E.S., van der Avoird, A., (1999), *Chem. Phys. Lett.*, **314**, 326.
- Rode, M., Sadlej, J., Moszynski, R., Wormer, P.E.S., van der Avoird, A., (2001), *Chem. Phys. Lett.*, **334**, 424.
- Rybak, S., Jeziorski, B., Szalewicz, K., (1991), *J. Chem. Phys.*, **95**(9), 6576.
- Sabzyan, H., Noorbala, M.R., (2003), *J. Mol. Struct.(Theochem)*, **626**, 143.
- Saebo, S., Tong, W., Pulay, P., (1993), *J. Chem. Phys.*, **98**, 2170.
- Sangthong, W., Probst, M., Limtrakul, J., (2005), *J. Mol. Struct.*, **748**, 119.
- Sarsa, A., Bacic, Z., Moskowitz, J.W., Schmidt, K.E., (2002), *Phys. Rev. Lett.*, **88**(12), Art. No. 123401.
- Sato, T., Tsuneda, T., Hirao, K., (2005), *Mol. Phys.*, **103**(6-8), 1151.
- Schaftenaar, G., Noordik, J. H., (2000), *J. Comput. Aided Mol. Design*, **14**, 123.
- Schlegel, H.B., (1982), *J. Comput. Chem.*, **3**, 214.
- Schlegel, H.B., (2003), *J. Comput. Chem*, **24**, 1514.
- Schutz, M., Rauhut, G. Werner, H.-J., (1998), *J. Phys. Chem. A*, **102**, 5997.
- Seung-Hoon, K., (1998), *J. Kor. Chem. Soc.*, **42**(6), 629.
- Singh U.C., Kollman, P.A., (1984), *J. Comput. Chem.*, **5**, 129.
- Sklenak, S., Yao, Cukier, R.I., Yan, H., (2004), *J. Am. Chem. Soc.*, **126**, 14789.
- Suhm, M.A., (1995), *Ber. Bunsenges. Phys. Chem.*, **99**, 1159.
- Sun, H., Watts, R.O., Buck, U., (1992), *J. Chem. Phys.*, **96**(3), 1810.
- Surin, L.A., Fourzikov, D.N., Dumesh, B.S., Winnewisser, G., Tang, J., McKellar, A.R.W., (2004), *J. Mol. Spec.*, **223**, 132.
- Surin, L.A., Fourzikov, D.N., Lewen, F., Dumesh, B.S., Winnewisser, G. McKellar, A.R.W., (2003), *J. Mol. Spec.*, **222**, 93.
- Svensson, M., Humbel, S., Froese, R.D.J., Matsubara, T., Sieber S., Morokuma, K., (1996), *J. Phys. Chem.*, **100**, 19357.
- Tao, J., Perdew, J.P., (2005), *J. Chem. Phys.* **122**(11), 114102.
- Treusekol, P., Lewis, J.P., Limtrakul, J., Truong, T.N., (2001), *Chem. Phys. Lett.*, **350**, 128.
- Tschumper, G.S., Morokuma, K., (2002), *J. Mol. Struct.(Theochem)*, **592**, 137.
- Tuma, C., Boese, A.D., Handy, N.C., (1999), *Phys. Chem. Chem. Phys.*, **1**, 3939.
- Van der Pol, A., van der Avoird, A., Wormer, P.E.S., (1990), *J. Chem. Phys.*, **92**(12), 7498.
- van der Vaart, A., Kenneth Jr., M.M., (2002), *J. Chem. Phys.*, **116**(17), 7380.
- Vissers, G.M.W., Hesselmann, A., Jansen, G., Wormer P.E.S., van der Avoird, A., (2005), *J. Chem. Phys.*, **122**(05), 43061.
- Vreven, T., Mennuci, B., da Silva, C., Morokuma, K., Tomasi, J., (2001),

- J. Chem. Phys.*, **115**(1), 62.
- Vreven, T., Morokuma, K., Farkas, Ö., Schlegel, H.B., Frisch, M.J., (2003),  
*J. Comput. Chem.*, **24**, 760.
- Walker, K.A., Xia, C., McKellar, A.R.W., (2000), *J. Chem. Phys.*, **113**(16), 6618.
- Warshel, A., Levitt, M., (1976), *J. Mol. Bio.*, **106**(2), 421.
- Weinhold, F., (1997), *J. Mol. Struct.(Theochem)*, **398-399**, 181.
- Wierchowski, S.J., Kofke, D.A., Gao, J., (2003), *J. Chem. Phys.*, **119**(14), 7365.
- Wilson, A.K., van Mourik, T., Dunning Jr., T.H., (1997), *J. Mol. Struct.(Theochem)*, **388**,  
339.
- Wojtulewski, S., Gabrowski, S.J., (2002), *J. Mol. Struct.*, **605**, 235.
- Woon, D.E., Dunning Jr., T.H., (1993), *J. Chem. Phys.*, **98**, 1358.
- Wu, Q., Yang, W., (2002), *J. Chem. Phys.*, **116**(2), 515.
- Yamanaka, M., Mikami, K., (2002), *Organometallics*, **21**, 5847.
- Yu, Z., Chuang, C-C., Medley, P., Stone, T.A., Klemperer, W., (2004),  
*J. Chem. Phys.*, **120**(15), 6922.

## APPENDIX A

### Appendix A.1

Starting geometries for the pre-optimization of the HF clusters with 6-31G(d) are shown in Fig. A.1


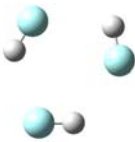



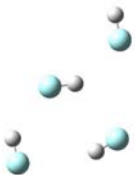


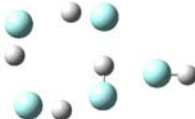



		
Cluster i	Cluster ii	Cluster iii
	 Use GDIIIS	
Cluster iv	Cluster v	Cluster vi
	 Use GDIIIS	 Use GDIIIS
Cluster vii	Cluster viii	Cluster ix
	 Use GDIIIS	 Use GDIIIS
Cluster x	Cluster xi	Cluster xii

Fig. A.1 Starting geometries for preoptimization with 6-31G(d)

## Appendix A.2

In Fig. A.2 we show the geometries of the HF clusters as optimized with 6-31G(d).

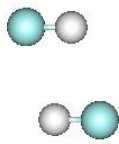
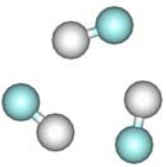
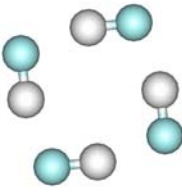
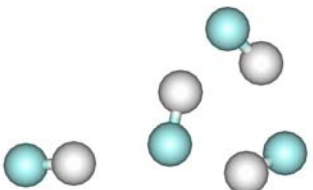
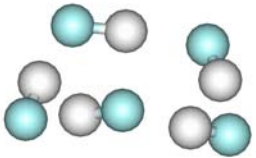
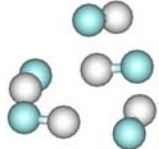
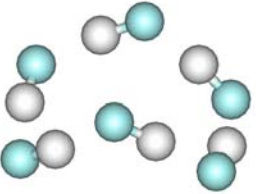
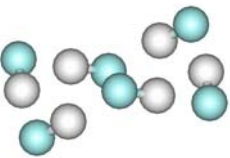
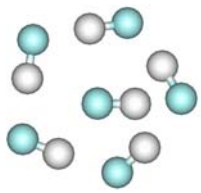


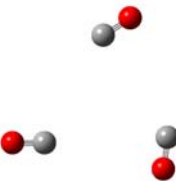
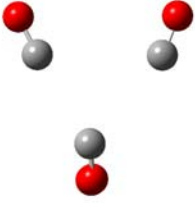
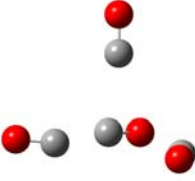
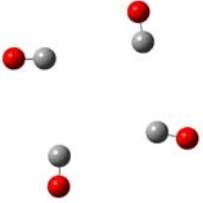
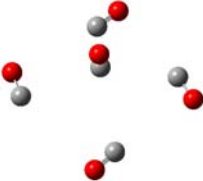
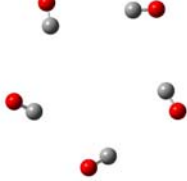
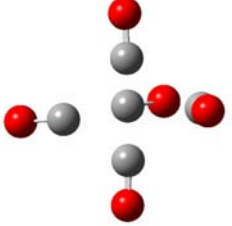
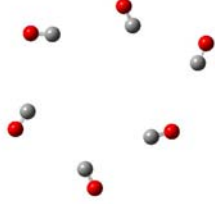
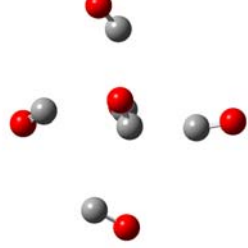
		
Dimer : -200.355526 hartree	Trimer: -300.591005 hartree	Tetramer 1: -400.801233 hartree
		
Tetramer 2: -400.782245 hartree	Pentamer 1: -500.004473 hartree	Pentamer 2: -500.995108 hartree
		
Hexamer 1: -601.206783 hartree	Hexamer 2: -601.206900 hartree	Hexamer 3: -601.201436 hartree.

Fig . A.2: HF clusters optimized with MP2 and 6-31G(d)

## Appendix A.3

For all CO clusters, very tight optimization conditions were used and force constants were calculated at each optimization step when possible. At least 100 optimization steps were used for each optimization. The clusters are shown in Table A.3. Only clusters i, ii, iii, iv, v, ix, xi were completely optimized. See **Chapter 6** for details.

Table A.3 Starting geometries of CO clusters for pre-optimization with 6-31G(d)

		
Cluster i	Cluster ii	Cluster iii
		
Cluster iv GDIIIS was used This cluster is tetrahedral	Cluster v Cluster iii was augmented with one monomer	Cluster vi GDIIIS was used
		
Cluster vii Cluster vi was augmented with one monomer	Cluster viii	Cluster ix Cluster iii was augmented with two monomers on each side of the plane
		
Cluster x: Organized specifically so that every pair of monomers is arranged in a T formation.	Cluster xi	

## Appendix A.4

In Fig. A.4 the geometries for the pre-optimized CO clusters, optimized with 6-31G(d) are shown.

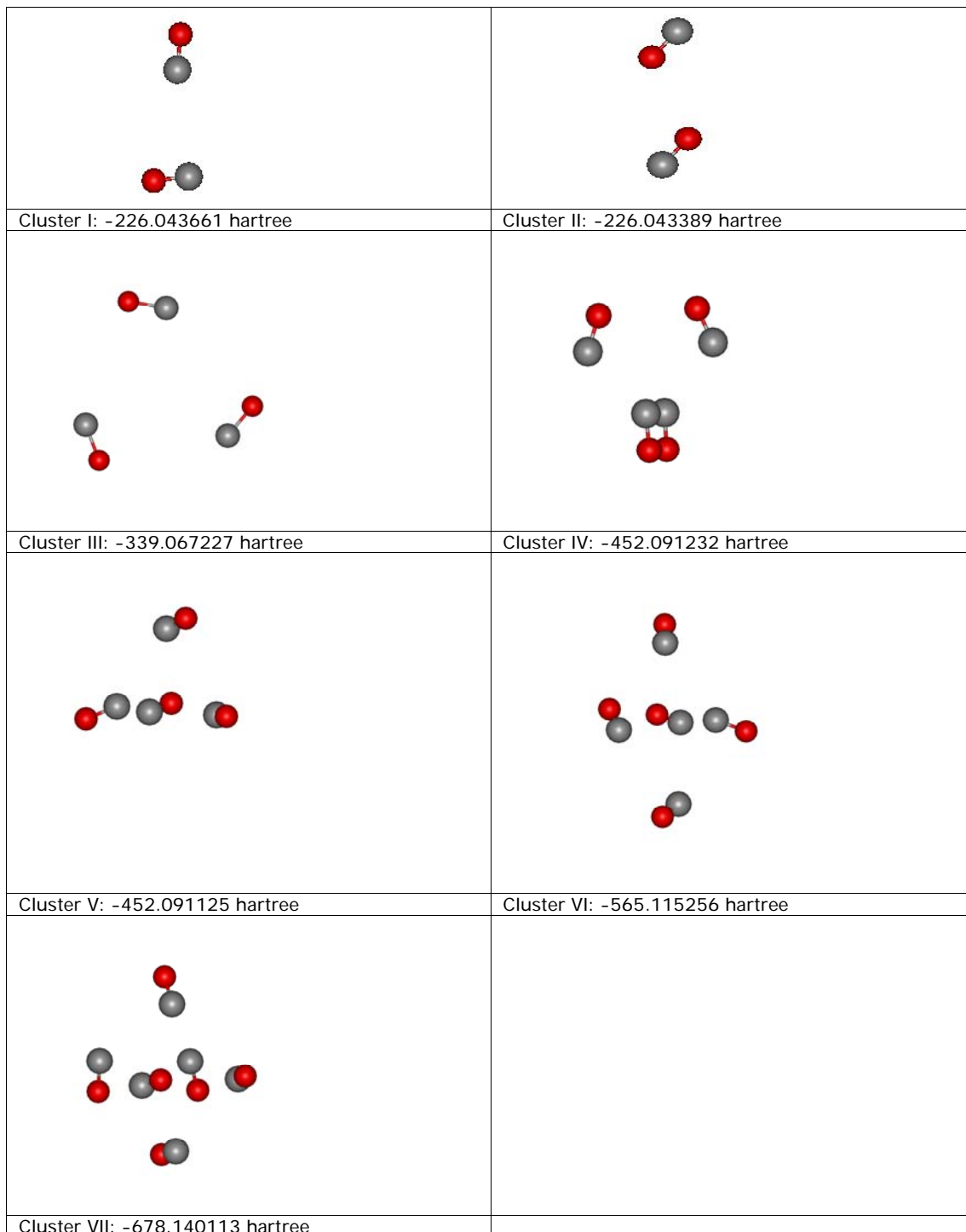


Fig. A.4 : CO clusters optimized with 6-31G(d)

All of these clusters have been shown to be minima on the PES by vibrational analysis. No imaginary frequencies were found.

### Appendix A.5

In Fig. A.5 we show the pre-optimized CO/HF clusters that were optimized with 6-31G(d)

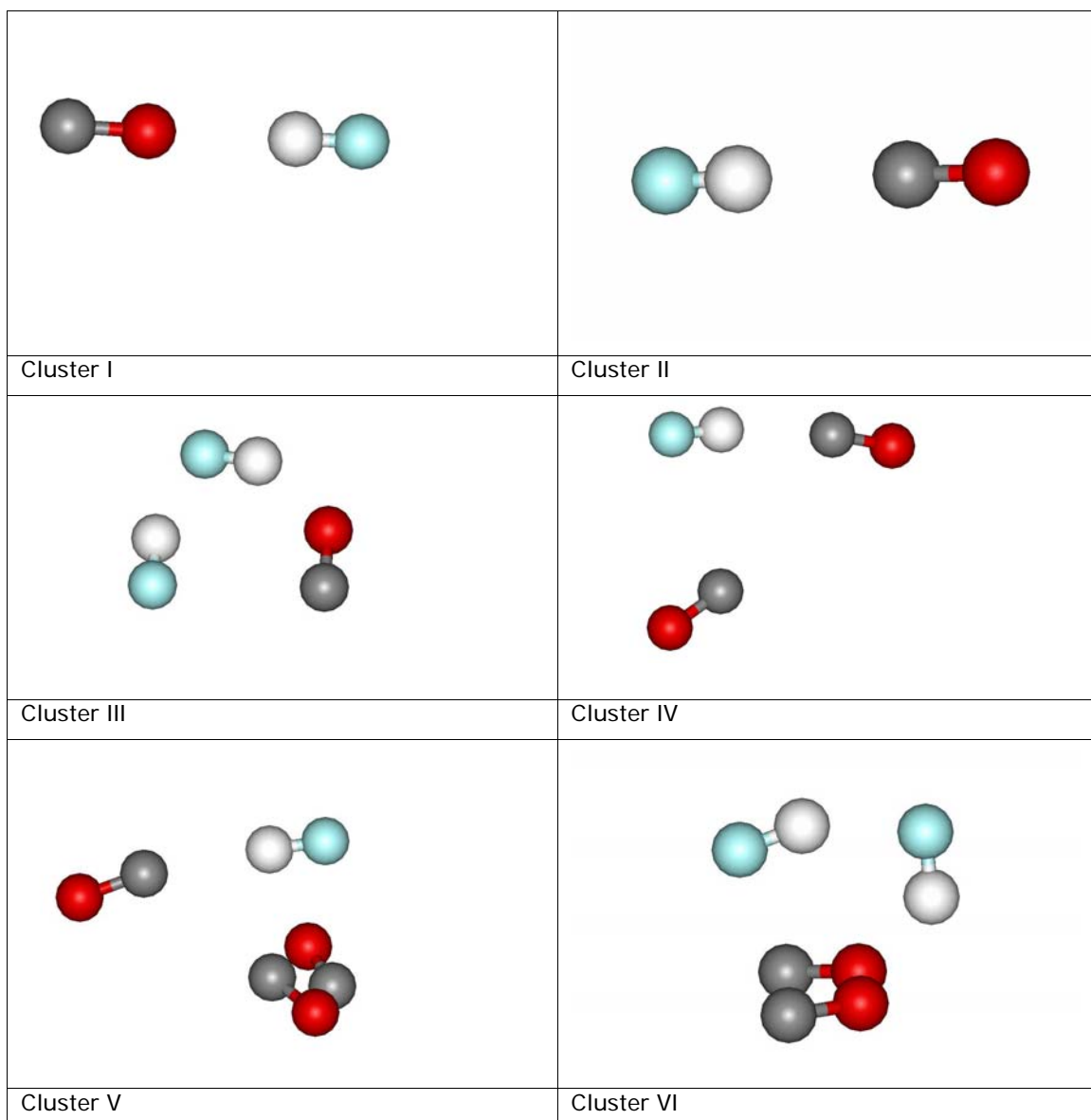


Fig. A.5 continued ...



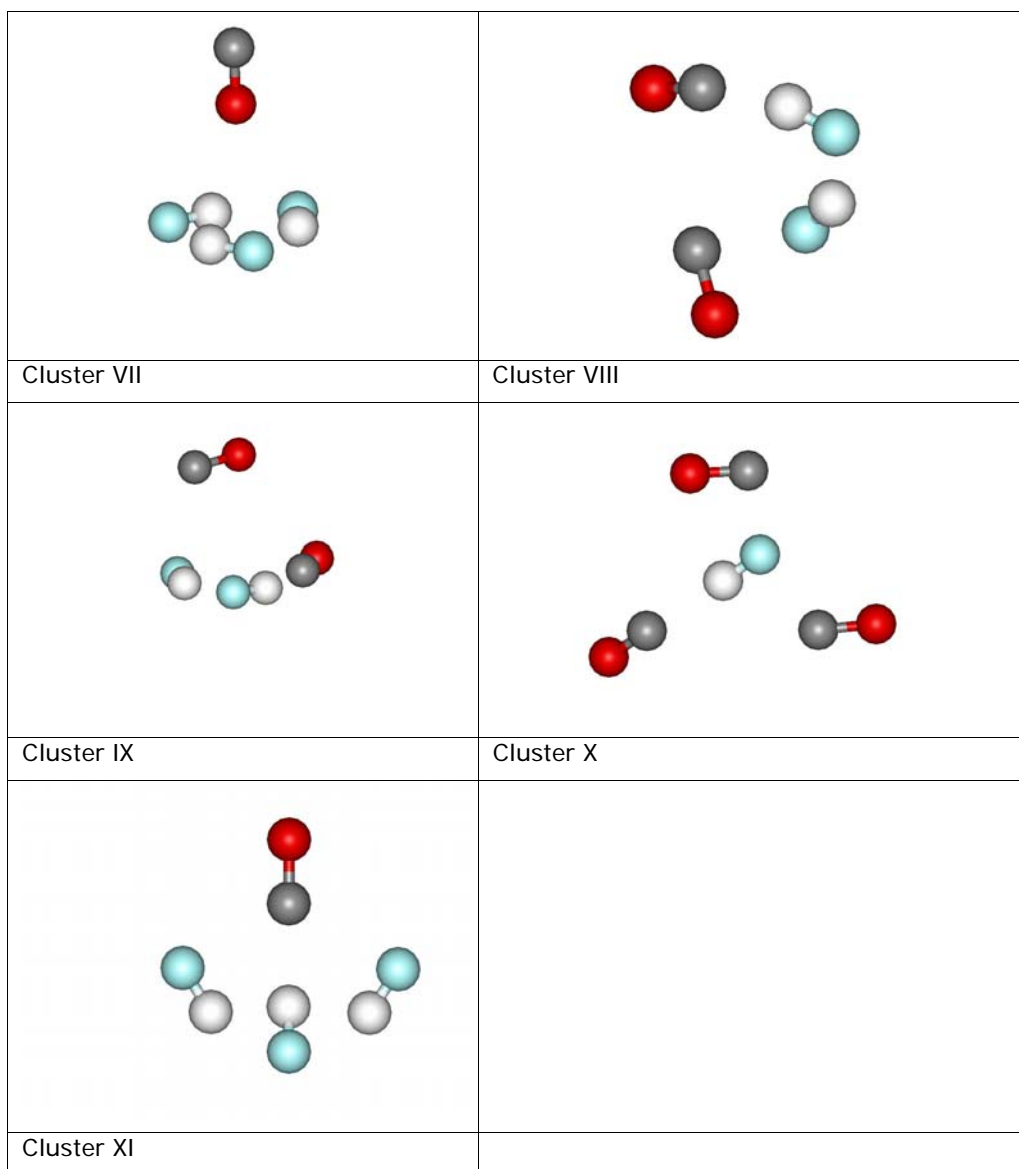


Fig. A.5: CO/HF clusters optimized with 6-31G(d). Cluster I: -213.20681848 hartree  
Cluster II: -213.21043766 hartree Cluster III: -313.40491410 hartree Cluster IV: -326.23646521 hartree  
Cluster V: -439.26144251 hartree Cluster VI: -426.43121828 hartree Cluster VII: -413.61576880 hartree  
Cluster VIII: -426.43735043 hartree Cluster IX: -426.43121828 hartree Cluster X: -439.26274252 hartree  
Cluster XI: -413.60411316 hartree

## APPENDIX B

### *Appendix B.1*

#### *Selected Gaussian 03 inputs*

In many cases, we found the *Gaussian 03* manual inadequate for our work. Most of the inputs were therefore discovered by trial and error. The Gaussian test jobs, as supplied with the software package, were more valuable than the manual. There are also mistakes in the manual that made it even harder to perform our calculations. All of the inputs given below have been tested by us, and should work in all cases. All ONIOM jobs were run on a single node. It is however possible to run them in parallel. It would be worthwhile to test the performance on parallel nodes. As our molecular systems were small compared to the usual systems treated with ONIOM, we did not think this was necessary.

*Input for an ONIOM optimization where the force field parameters are explicitly defined in the input. Micro-iterations are used.*

```
%mem=512MB
In order to speed up calculations and optimizations the maximum memory was set to 512
MB}
#ONIOM(MP2/6-311+G(d,p):UFF=softfirst)=(MK,Embedcharge)
opt=(Maxcycle=10, CalcFC, Vtight) nosymm
```

This line starts with a specification of ONIOM with the methods to be used in each layer separated by a semicolon. Instructions for each layer can be given before the semicolon such as modified basis sets, ECP's (Effective core potentials)<sup>4</sup> etc. for the QM method and for the UFF layer one can specify what force field parameters should have preference. In this case, we chose that the soft-parameters or those in the input file, should have preference above the hard parameters, which are the internal parameters of the program for the force field. The ONIOM setup specification is enclosed in parentheses. After this, the ONIOM options follow in a separate parenthesis. This states that MKS charges should be used and that the point charges in the MM should be incorporated in the QM Hamiltonian. If the Embedcharge keyword is omitted, mechanical embedding will be the result. To add more control over the charge fitting of the ESP, one can use the keyword `pop=(MK,dipole)` to constrain the charges to the dipole moment of the QM system. The `opt` keyword contains instructions for the QM and MM optimizers. The `maxcycle` option does not influence the microiteration optimizer, only the QM optimizer. The maximum optimization cycles for the QM to optimize

---

<sup>4</sup> ECP's in ONIOM should be specified by the keyword `genECP`.

will therefore be 10. CalcFC is again an option for the QM optimizer and can be omitted. It should speed up optimizations of the QM system. Vtight has an effect on the entire geometry optimization and also gives instructions to the micro-iteration optimizer to what convergence is allowed. The nosymm keyword can be omitted as Gaussian test the symmetry of the model and the real systems to see if they are the same. If not, Gaussian removes symmetry constraints automatically. Other options as in general *ab initio* calculations can also be given in this line. pop=full will give a full Mulliken population analysis of the QM wave function for example. It is not possible however to obtain a proper wave function as in general *ab initio* calculations with the keyword output=wfn.

```
-----  
  ONIOM optimization  
-----  
0,1,0,1,0,1
```

This is given in the sequence of:

Charge for low-level calculation on real system, Multiplicity for low-level calculation on real system, Charge for high-level calculation on model system, Multiplicity for high-level calculation on model system, Charge for low-level calculation on model system, Multiplicity for low-level calculation on model system.

Note that multiplicities need to be specified even if UFF is used in the low-level, otherwise the program will not accept the input.

Cartesian coordinates

F-F_0.0	0	1.29146	1.29146	0.00000	H
H-H_0.0	0	1.49594	0.37698	0.00000	H
F-F--0.435	-2	1.29146	-1.29146	0.00000	L
H-H_-0.433	-2	0.37698	-1.49594	0.00000	L
F-F--0.423	-3	-1.29146	-1.29146	0.00000	L
H-H_-0.421	-3	-1.49594	-0.37698	0.00000	L
H-H_-0.438	-4	-0.37698	1.49594	0.00000	L
F-F--0.434	-4	-1.29146	1.29146	0.00000	L

In column one the atom, force field type and point charge are given. The point charge is only for the MM parameter definition and if charges on the QM atoms are given, the software is supposed to ignore these. In column two are the instructions for the geometry optimizers are given. 0 means optimize. All the atoms that are part of rigid blocks have numbers smaller than -1. If the number is -1, the specific Cartesian coordinate will be frozen. Therefore, the second monomer in this cluster is a rigid block and also the third and fourth. Columns 3,4 and 5 contains the Cartesian coordinates of the starting geometry. Note that if a Z-matrix is used, the program automatically assumes by default that the first three atoms are in the high-layer. This cannot be changed. The last column specifies what atoms are part of the high (H) or low (L) layers.

```
NonBon 2 1 0.000 0.000 0.000 0.000 0.000 0.000
VdW F_ 3.364 0.05
VdW H_ 2.886 0.044
```

The NonBon function specifies both the type of vdW and electrostatics to be used in the force field. 2 means that geometric vdW parameters such as in UFF should be used. This is followed by a 1 which means that 1/R electrostatics should be used. Setting one of these values to zero instructs the program to “switch off” the vdW or the electrostatic calculation and in our case only the bond lengths are then considered. The rest of the terms are cutoff terms and is all set to zero for our purposes, as our molecules are simple.

VdW sets the vdW parameters of the force field type. The first number is the  $R_0$ -value and the second is the  $\epsilon$ -value. In the above example, both the F\_ and H\_ vdW parameters are set in the input.

#### *An ONIOM optimization with the default optimizer*

In this optimization if ONIOM-EE is used, electronic-embedding is only incorporated at the start and then the charges on the QM atoms are kept constant throughout the optimization. This is certainly wrong. However, we give the input for such an optimization as well.

```
%mem=512MB
#ONIOM(MP2/6-311+G(d,p):UFF=softfirst)=(MK,Embedcharge)
opt=(Maxcycle=120, CalcFC, Vtight, nomicro, modredun) nosymm
```

Most terms are the same as the first example except that now the opt-keyword is augmented with a nomicro and modredun options. The nomicro keyword instructs the program to use one optimizer (the Beryn-algorithm) for the QM and MM optimizations. The keyword modredun is used to freeze bond lengths or other internal coordinates and is only an instruction to the main optimizer, in this case the Beryn-algorithm.

```
-----  
ONIOM optimization without micro-iterations  
-----  
0,1,0,1,0,1
```

The same as in the first example.

F-F <sub>0</sub>	0	1.29146	1.29146	0.00000	H
H-H <sub>0</sub>	0	1.49594	0.37698	0.00000	H
F-F <sub>-</sub>	0	1.29146	-1.29146	0.00000	L
H-H <sub>-0.433</sub>	0	0.37698	-1.49594	0.00000	L
F-F <sub>--0.423</sub>	0	-1.29146	-1.29146	0.00000	L
H-H <sub>-0.421</sub>	0	-1.49594	-0.37698	0.00000	L
H-H <sub>-0.438</sub>	0	-0.37698	1.49594	0.00000	L
F-F <sub>--0.434</sub>	0	-1.29146	1.29146	0.00000	L

The same as in the first example. As no micro-iterations are used, zero is used in column two

3	4	0.9370	F
5	6	0.9370	F
7	8	0.9370	F

This states that the bond length between each atom pair in the MM should be frozen to a specific value. It is important to add the “F” to freeze these internal coordinates

```
NonBon 2 1 0.000 0.000 0.000 0.000 0.000 0.000  
VdW F- 3.364 0.05  
VdW H- 2.886 0.044
```

The same as in the first example.

## Appendix B.2

*Things to look out for when using ONIOM-EE in Gaussian 03:*

1. Using rigid blocks in the MM system will make convergence extremely slow, and there is not much one can do about it. Sometimes convergence can take a few days even for a small system. Micro-iterations can also fail in some cases and the program will not give an explanation for this and just aborts in the middle. See Fig. B.1. In Chapter 9 we gave a possible reason for the failure of the micro-iterations.

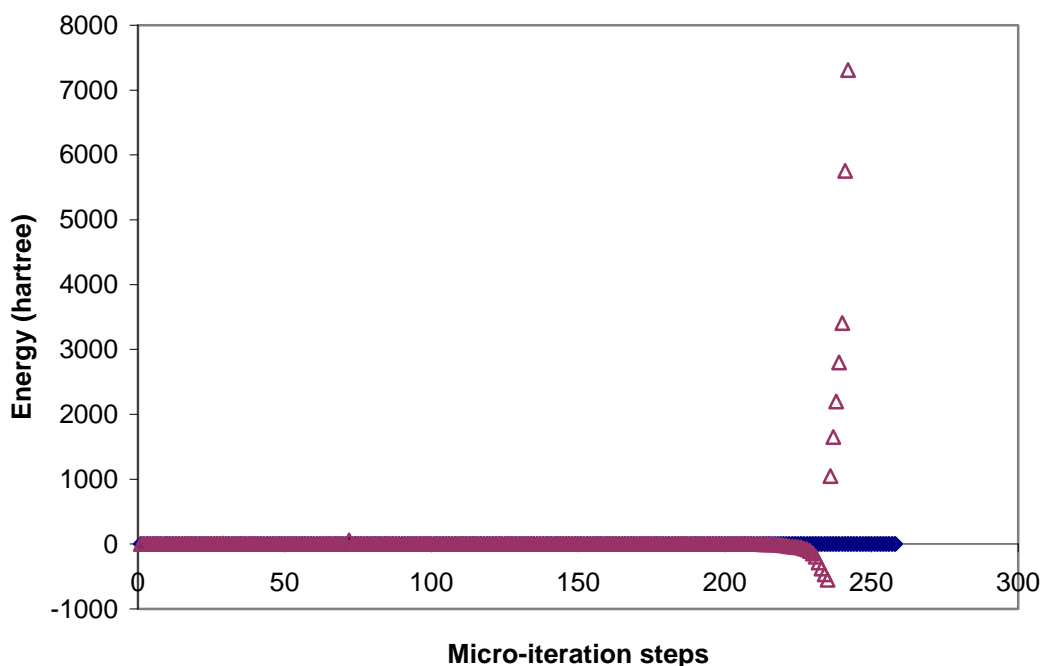


Fig. B.1: Micro-iterations in *Gaussian 03* break down for our HF pentamer for the default force field parameters (triangles), but not for our optimized parameters(diamonds). MKS charges based on the MP2 density were used.

2. Never use both `modredundant` and the flags (-2,-3 etc.) in the input for the same atoms. This will lead to erroneous results. Earlier in this work we used `modredundant` and the flags for the same atoms as we thought it would give more control over the bond length of the monomer in the rigid block. The program however will give a completely different output when `modredundant` is used together with the flags or when only flags are used. We think that this is because `modredundant` changes the MM systems' coordinates before Gaussian calculates the energy. The analytical derivatives of this energy are then used in the micro-iterations for the MM optimization, however the flags in the input instruct the micro-iteration optimizer to use the coordinates given in the input and not the modified coordinates that were determined by `modredundant`. Therefore, it is possible that the analytical gradient used in the MM optimization is the wrong gradient for the coordinates used for the MM system in the micro-iterations, leading to a totally inaccurate optimization.
3. Using Cartesian coordinates differing even at the fifth decimal might lead to a slightly different answer after an ONIOM-EE optimization. We therefore used the exact same coordinates for each specific cluster when optimizing the force field.

4. For short ONIOM optimizations, Molden can be used to view the entire optimization including the MM optimization. However, sometimes convergence is so slow that the Gaussian log file might be 1 GB in size, which can then only be viewed by Gaussview.
5. A problem with Gaussview is that if the QM system is automatically optimized at the start with respect to the optimized MM system, Gaussview will only show the geometry of the input structure and not the optimized structure. This might give the false impression, when optimizing a force field, that the parameters tested are indeed optimized, when the real geometry is completely different. Only if at least two steps are taken to optimize the QM system, will Gaussview show the correct final geometry. To make Gaussview show the correct geometry the following text can be added in the output file after the coordinates for the micro-iterations:

```
GradGradGrad  
Leave 3 lines open  
GradGradGrad
```

One can also take the final coordinates after the MM optimization and paste it in a .com file which can be viewed by Gaussview.

6. When comparing default force field calculations for CO clusters in *Gaussian 03* with the exact same coordinates in Cerius<sup>2</sup>, we find that *Gaussian 03* gives totally different results. In order to obtain the same results with *Gaussian 03* than with Cerius<sup>2</sup> one should add the keyword `geom=connectivity` in the root section and specify the connectivity explicitly before the force field parameter description. The force field types should also be given in the input as usual.
7. When looking at the output of a Gaussian ONIOM calculation/optimization the user will see that a basis set is mentioned with all MM calculations. This basis set is not actually used and it is just a trick to do the MM calculations with the program. This basis set is the STO-3G basis set.
8. Using explicitly defined force field parameters and using internal hard parameters might give different answers. We therefore used explicitly defined parameters in all cases.

These eight points are not discussed in the Gaussian-manual and so anyone wanting to use ONIOM-EE as we did, should familiarize himself/herself with this. We will compile a copy of these problems for Gaussian Inc.

### **Appendix B.3**

*A list of the most important IOps regarding ONIOM-EE*

Here is a list of most important IOps that can be used when doing an ONIOM-EE calculation with *Gaussian 03*.

IOP(1/52): Set type of ONIOM calculation

IOP(1/53): Instructions on reading from .rwf files and integration of energy, gradient and hessian. This integration simply means adding the individual terms as in the ONIOM methodology and not calculating integrals. See **Chapter 2**.

IOP(1/64): Molecular mechanics force field selection

IOP(1/65): Control over terms in force field such as torsions, stretches, out of plane etc.

IOP(1/66): Specify charge-equilibration. This only effects the MM parameter definition and was not used in our work to minimize the type of charges used in the optimizations.

IOP(4/110): This IOP scales rigid fragment steps in the micro-iterations. Scaling rigid fragments did not give a significant improvement over optimizations in our work.

IOP(1/67): Source of the MM parameters. This IOP need to be used when link-1 commands are used to do calculations with ONIOM. For example, the calculation of frequencies with modified MM parameters after an optimization should be done by using this IOP.

### **Appendix B.4**

*Difference between a geometry optimization with the default optimizer and the micro-iterations optimizer for ONIOM-EE.*

We will use the HF pentamer optimization as an example. The  $R_0$ -value was 3.34 Å for the fluorine atom-type and 1.28 Å for the hydrogen atom-type. These parameters were optimized, see **Chapter 4**. In Table B.2 we give information for this optimization for both optimization algorithms.



Table B.2: Comparison between two optimization algorithms used to optimize the HF pentamer

	<b>Micro-iterations</b>	<b>Default optimizer</b>
QM-bond length (Å)	0.931	0.923
CPU-time	16 min. 38.3 s	14 min. 5.8 s
High/model (hartree)	-100.760061	-100.742144
Low/model (hartree)	-0.474296	-0.457432
Low/real (hartree)	-0.044087	-0.032625

From this example, it is clear that the results for the two optimizers differ significantly when doing the same optimization. According to the literature [Vreven *et al.*, 2003], the optimization with micro-iterations should be the most accurate. Therefore, even if this optimization is slower than for the default algorithm we believe that to obtain accurate results the ONIOM-EE with micro-iterations should be used instead of ONIOM-EE with the default algorithm.

## APPENDIX C

### *Verifying the van der Waals and electrostatic interactions in Gaussian 03.*

*Gaussian 03* allows one to specify force field parameters in the input. This was not possible with *Gaussian98*. We therefore wanted to verify that the software calculates the correct van der Waals energy and electrostatic energy.

To verify the UFF Lennard-Jones parameters for the fluorine and hydrogen atoms we did electronic embedded singlepoint ONIOM calculations while the calculation of the van der Waals interactions and electrostatic interactions are explicitly switched off in the input. (The reader is referred to **Appendix B.1**). We use the HF dimer as model due to its simplicity. This allows us to calculate the electrostatic energy, obtained with the Coulomb potential, and the van der Waals energy by hand. See Table C.1.

Table C.1: Low/real energy values for different types of calculations to find the van der Waals energy

Calculation type	Energy
Coulomb interaction off (hartrees)	0.036160270216
Coulomb and van der Waals off (hartrees)	0.005800261488
Van der Waals energy (hartrees)	0.030360008728
Van der Waals energy (kcal/mol)	19.1

We use the 6-12 Lennard-Jones equation in Chapter 2 to calculate the van der Waals interaction between each atom in the HF dimer. The  $\epsilon$  (equilibrium well depth) and  $R_0$ -values are obtained from the UFF article [Rappé *et al.*, 1992]. The van der Waals interaction energy is calculated to be 19.1 kcal/mol, the same as predicted by *Gaussian 03* if possible rounding off errors is taken into consideration.

It is of note that we found a slight discrepancy in the geometries obtained when making use of the internal Gaussian defined UFF-parameters and when explicitly defining the parameters in the input as specified in Appendix B.2 point 8. Further investigation of this anomaly seems to point towards an error in the ONIOM method, as full UFF calculations with the internal parameters and the external parameters showed no difference.

Next, we did another two single point ONIOM calculations where we specify the charges on the QM atoms in the input and another calculation where the QM charges are set to zero. The difference between these energy values will give the energy for the Coulomb interaction. Table C.2 summarizes these results.

Table C.2: Table showing the results for the single point ONIOM calculations. Only the Low/Real values are shown.

System	Charges set to MKS-values	Charges set to zero	Difference
$E_{\text{real}}^{\text{low}}$ (hartree)	-0.131575	-0.126158	-0.005417

The difference between the low/real calculations are general UFF calculations and one should obtain the same value with a full UFF-calculation of the whole system. The difference of -0.00541662 hartree amounts to -3.4 kcal/mol. By using the equation for the Coulomb interaction as defined in the UFF-force field [Rappé *et al.*, 1992] and for a dielectric constant of 1, (See equation C.1) we calculated the Coulomb interaction as: -3.4 kcal/mol.

$$E_{el} = 332.0637 \left( \frac{q_i q_j}{R_{ij}} \right) \quad (\text{C.1})$$

where the  $q_i$  and  $q_j$  are the charges on atoms  $i$  and  $j$ .  $R_{ij}$  is the distance between atoms  $i$  and  $j$ .

The same type of calculation was performed to verify the vdW energy and electrostatic energy for the CO dimer.

## APPENDIX D

In this appendix we list all the nonbonded distances determined by ONIOM-EE and MP2 in Ångström for all the clusters. These nonbonded distances were used to optimize the force field based on geometrical parameters.

Nonbonded distances determined by ONIOM-EE and MP2 for all the clusters.

### *Hydrogen fluoride clusters*

	MP2 nonbonded distances (Å)	ONIOM-EE nonbonded distances obtained with default FF- parameters (Å)	Square of deviation	ONIOM-EE nonbonded distances obtained with optimized parameters (Å)	Square of the deviation
F1-F3	2.7878	2.8428	0.0030		
F1-H4	3.3587	2.5327	0.6823		
H2-F3	1.8745	2.6977	0.6776		
H2-H4	2.5103	2.6967	0.0348		
Dimer RMSD (Å)			<b>0.591</b>		

F1-F5	2.6678	3.1523	0.2349	2.4481	0.0482
F1-F6	2.6678	3.0994	0.1863	2.5032	0.0271
F5-F6	2.6678	3.0157	0.1211	2.4942	0.0301
F1-H4	1.8654	2.4858	0.3850	1.7502	0.0133
F1-H3	2.7440	3.5677	0.6784	2.7095	0.0012
H2-F5	1.8654	2.4389	0.3289	1.6345	0.0533
H2-F6	2.7440	3.2809	0.2882	2.5709	0.0300
H2-H3	2.1807	3.0267	0.7156	2.1270	0.0029
H2-H4	2.1807	2.8731	0.4794	2.0998	0.0066
H3-H4	2.1807	3.0783	0.8057	2.2431	0.0039
F5-H4	2.7440	3.3620	0.3819	2.6268	0.0137
H3-F6	1.8654	2.5053	0.4095	1.8166	0.0024
Trimer RMSD (Å)			<b>0.646</b>		<b>0.139</b>

F1-F3	2.5829			2.4653	0.0139
F1-H4	2.9336			3.0849	0.0229
F1-F5	3.6528			3.7930	0.0197
F1-H6	3.2486			3.4878	0.0572
F1-F8	2.5829			2.5275	0.0031
F1-H7	1.6809			71.6793	2.82E-06
H2-F3	1.6809			1.5828	0.0096
H2-H4	2.1817			2.3183	0.0186
H2-F5	3.2483			3.3532	0.0110
H2-H6	3.0854			3.2777	0.0370
H2-F8	2.9336			2.7906	0.0204
H2-H7	2.1817			2.1517	0.0009
F3-F5	2.5829			2.5116	0.0051
F3-H6	2.9336			2.8858	0.0023
F3-F8	3.6528			3.2747	0.1430
F3-H7	3.2486			3.0560	0.0371
F5-H4	1.6809			1.7025	0.0005
H4-H6	2.1817			2.2876	0.0112
H4-F8	3.2486			3.1451	0.0107
H4-H7	3.0854			3.2024	0.0137
F5-F8	2.5829			2.5180	0.0042
F5-H7	2.9336			3.0674	0.0179
H6-F8	1.6809			1.6935	0.0002
H6-H7	2.1817			2.4057	0.0502
Tetramer (III) RMSD (Å)					<b>0.146</b>

F1-F3	2.5475			2.4665	0.0066
F1-H4	3.0585			3.2568	0.0393
F1-F5	4.1220			4.4089	0.0823
F1-H6	4.0393			4.5481	0.2589
F1-F7	4.1220			4.3833	0.0683
F1-H8	3.4855			3.7957	0.0962
F1-F9	2.5475			2.5437	0.0000
F1-H10	1.6161			1.6588	0.0018
H2-F3	1.6161			1.5598	0.0032
H2-H4	2.2202			2.4080	0.0353
H2-F5	3.4855			3.7257	0.0577
H2-H6	3.5924			4.0193	0.1823
H2-F7	4.0393			4.1899	0.0227
H2-H8	3.5924			3.7851	0.0371
H2-F9	3.0585			2.8855	0.0299
H2-H10	2.2202			2.1731	0.0022
F3-F5	2.5475			2.5225	0.0006
F3-H6	3.0585			3.1025	0.0019
F3-F7	4.1220			3.8817	0.0578
F3-H8	4.0393			3.8386	0.0403
F3-F9	4.1220			3.6005	0.2720
F3-H10	3.4855			3.2196	0.0707
H4-F5	1.6161			1.6495	0.0011
H4-H6	2.2202			2.3658	0.0212
H4-F7	3.4855			3.4674	0.0003
H4-H8	3.5924			3.6497	0.0033
H4-F9	4.0393			3.8264	0.0453
H4-H10	3.5924			3.6667	0.0055
F5-F7	2.5475			2.5353	0.0001
F5-H8	3.0585			3.1159	0.0033
F5-F9	4.1220			3.9930	0.0167
F5-H10	4.0393			4.2189	0.0323
H6-F7	1.6161			1.6809	0.0042
H6-H8	2.2202			2.4059	0.0345
H6-F9	3.4855			3.5798	0.0089
H6-H10	3.5924			4.0039	0.1694
F7-F9	2.5475			2.5283	0.0004
F7-H10	3.0585			3.2659	0.0430
H8-F9	1.6161			1.6450	0.0008
H8-H10	2.2202			2.4721	0.0635
Pentamer (V) RMSD (Å)					<b>0.213</b>

F1-F3	2.6534			2.5178	0.0184
F1-H4	2.9639			2.9936	0.0009
F1-F5	3.6627			3.5913	0.0051
F1-H6	3.2569			3.3740	0.0137
F1-F8	2.6003			2.5291	0.0051
F1-H7	1.7040			1.6884	0.0002
F1-H9	4.1916			3.6717	0.2703
F1-F10	5.0204			4.4053	0.3783
H2-F3	1.7772			1.6354	0.0201
H2-H4	2.2580			2.2543	0.0000
H2-F5	3.2906			3.2114	0.0063
H2-H6	3.1257			3.2424	0.0136
H2-F8	2.9673			2.8812	0.0074
H2-H7	2.2199			2.2356	0.0002
H2-H9	3.2603			2.7615	0.2488
H2-F10	4.0963			3.5315	0.3190
F3-F5	2.5193			2.5231	0.0000
F3-H6	2.8939			3.0258	0.0174
F3-F8	3.6467			3.5267	0.0144
F3-H7	3.2814			3.2676	0.0002
F3-H9	1.8519			1.6984	0.0236
F3-F10	2.7722			2.6171	0.0240
H4-F5	1.5945			1.6694	0.0056
H4-H6	2.1058			2.3395	0.0546
H4-F8	3.1885			3.2499	0.0038
H4-H7	3.0655			3.2513	0.0345
H4-H9	2.5069			2.4076	0.0098
H4-F10	3.4023			3.3114	0.0083
F5-F8	2.5649			2.4967	0.0047
F5-H7	2.9316			2.9782	0.0022
F5-H9	4.0269			4.0767	0.0025
F5-F10	4.8801			4.9764	0.0093
H6-F8	1.6594			1.6868	0.0008
H6-H7	2.1740			2.3564	0.0333
H6-H9	4.6117			4.6908	0.0063
H6-F10	5.5075			5.6095	0.0104
F8-H9	5.4981			5.1893	0.0954
F8-F10	6.4187			6.0850	0.1113
H7-H9	5.1047			4.8183	0.0820
H7-F10	6.0123			5.6697	0.1174
Pentamer (VI) RMSD (Å)					<b>0.222</b>

F1-F2	2.5347			2.4749	0.0036
F1-F3	2.5347			2.5530	0.0003
F1-F4	4.3902			4.6647	0.0754
F1-F5	4.3902			4.6546	0.0699
F1-F6	5.0694			5.5403	0.2218
F1-H8	3.1434			3.3306	0.0350
F1-H9	4.5312			5.0898	0.3120
F1-H10	4.7104			5.3227	0.3749
F1-H11	3.6323			3.9403	0.0948
F1-H12	1.5947			1.6491	0.0030
F2-F3	4.3902			3.8131	0.3330
F2-F4	2.5347			2.5284	0.0000
F2-F5	5.0694			4.6766	0.1543
F2-F6	4.3902			4.3087	0.0066
F2-H7	1.5947			1.5589	0.0013
F2-H9	3.1433			3.2420	0.0097
F2-H10	4.5312			4.5246	0.0000
F2-H11	4.7104			4.4026	0.0947
F2-H12	3.6323			3.3307	0.0910
F3-F4	5.0694			4.7959	0.0748
F3-F5	2.5347			2.5282	0.0000
F3-F6	4.3902			4.3783	0.0001
F3-H7	3.1434			2.9550	0.0355
F3-H8	4.5312			4.2244	0.0941
F3-H9	4.7104			4.7326	0.0005
F3-H10	3.6323			3.7739	0.0201
F3-H11	1.5947			1.6244	0.0009
F4-F5	4.3902			4.2072	0.0335
F4-F6	2.5347			2.5160	0.0003
F4-H7	3.6323			3.8807	0.0617
F4-H8	1.5947			1.6280	0.0011
F4-H10	3.1433			3.1899	0.0022
F4-H11	4.5312			4.4680	0.0040
F4-H12	4.7104			4.7979	0.0077
F5-F6	2.5347			2.5185	0.0003
F5-H7	4.5312			4.6339	0.0106
F5-H8	4.7104			4.5581	0.0232
F5-H9	3.6323			3.6555	0.0005
F5-H10	1.5947			1.6292	0.0012
F5-H12	3.1434			3.3481	0.0419
F6-H7	4.7104			5.0813	0.1375
F6-H8	3.6323			3.7003	0.0046
F6-H9	1.5947			1.6276	0.0011
F6-H11	3.1434			3.2375	0.0089
F6-H12	4.5312			4.8640	0.1108
H8-H7	2.2611			2.4484	0.0351
H7-H9	3.9164			4.4086	0.2422
H7-H10	4.5223			5.0141	0.2419
H7-H11	3.9164			4.0674	0.0228
H7-H12	2.2611			2.1967	0.0042
H8-H9	2.2611			2.4296	0.0284
H8-H10	3.9164			4.0841	0.0281
H8-H11	4.5222			4.4763	0.0021
H8-H12	3.9164			3.9204	0.0000
H9-H10	2.2611			2.4073	0.0214
H9-H11	3.9164			4.0908	0.0304
H9-H12	4.5223			4.9179	0.1566
H10-H11	2.2611			2.4413	0.0325
H10-H12	3.9164			4.3950	0.2291
H11-H12	2.2611			2.5008	0.0574
Hexamer (VII) RMSD (Å)					<b>0.248</b>



## MP2 nonbonded distances

The convergence of the optimization was too slow for either the default UFF parameters or the optimized parameters. In this section the MP2 nonbonded distances are listed. See **Chapter 4**.

### Tetramer 2 (Cluster IV)

F1-F3	2.8059	F3-F5	2.5933
F1-H4	2.7562	F3-H6	2.7370
F1-F5	2.6674	F3-H7	1.8345
F1-H6	1.8606	F3-F8	2.7556
F1-H7	4.5407	H4-F5	1.7302
F1-F8	5.4304	H4-H6	2.0830
H2-F3	2.1345	H4-H7	2.5111
H2-H4	2.3641	H4-F8	3.4078
H2-F5	2.8685	F5-H7	4.2322
H2-H6	2.2809	F5-F8	5.1137
H2-H7	3.7395	H6-H7	4.5346
H2-F8	4.6020	H6-F8	5.4501

### Hexamer 2 (Cluster VIII)

F1-F3	2.5471	F4-F7	3.4641
F1-H4	3.0378	F7-H8	3.5413
F1-F5	4.1213	F4-F9	4.0121
F1-H6	4.0122	F4-H10	3.5413
F1-F7	4.1213	F4-F11	2.7700
F1-H8	3.4641	F4-H12	3.5046
F1-F9	2.5471	F5-H7	2.5471
F1-H10	1.6122	F5-H8	3.0380
F1-F11	3.0135	F5-F9	4.1213
F1-H12	3.7107	F5-H10	4.0122
F2-F3	1.6122	F5-F11	3.0135
F2-H4	2.1886	F5-H12	3.7107
F2-F5	3.4641	H6-F7	1.6122
F2-H6	3.5413	H6-H8	2.1886
F2-F7	4.0122	H6-F9	3.4641
F2-H8	3.5413	H6-H10	3.5413
F2-H9	3.0380	H6-F11	2.7700
F2-H10	2.1886	H6-H12	3.5046
F2-F11	2.7700	F7-F9	2.5471
F2-H12	3.5046	F7-H10	3.0380
F3-F5	2.5471	F7-H11	3.0135
F3-H6	3.0380	F7-H12	3.7107
F3-F7	4.1213	F8-F9	1.6122
F3-H8	4.0122	F8-H10	2.1886
F3-F9	4.1213	F8-F11	2.7700
F3-H10	3.4641	F8-H12	3.5045
F3-F11	3.0135	F9-F11	3.0135
F3-H12	3.7107	F9-H12	3.7107
F4-F5	1.6122	F10-F11	2.7700
F4-H6	2.1886	F10-F12	3.5046

## Carbon monoxide clusters

	MP2 nonbonded distances (Å)	ONIOM-EE Nonbonded distances obtained with default FF- parameters (Å)	Square of deviation	ONIOM-EE nonbonded distances obtained with optimized parameters (Å)	Square of the deviation
O1-C3	4.2616	4.0524	0.0438	3.6256	0.4046
O1-O4	5.1024	4.6320	0.2213	3.6834	2.0138
C2-C3	3.5899	3.7138	0.0153	3.8974	0.0946
C2-O4	4.2616	4.0283	0.0544	3.6074	0.4280
Cluster I RMSD (Å)			<b>0.289</b>		<b>0.857</b>

C1-O4	3.8530			3.4920	0.1303
C1-C3	4.7000			3.7341	0.9321
O2-O4	3.1944			4.3486	1.3321
O2-C3	3.8529			4.3426	0.2398
					0.6586
Cluster II RMSD (Å)					<b>0.812</b>

O1-C3	3.9608	4.0435	0.0068	3.8651	0.0092
O1-O4	5.0903	4.5408	0.3020	3.4363	2.7360
O1-C5	4.7585	4.3596	0.1591	3.9374	0.6743
O1-O6	5.0903	4.5356	0.3078	3.4656	2.6399
C2-C3	3.7040	3.7534	0.0024	3.7954	0.0084
C2-O4	4.7585	4.0509	0.5007	3.7053	1.1092
C2-C5	3.7040	3.6206	0.0069	3.7921	0.0078
C2-O6	3.9608	4.0457	0.0072	3.6568	0.0924
C5-C3	3.7040	3.7211	0.0003	3.7993	0.0091
C5-O4	3.9608	4.0583	0.0095	3.7708	0.0361
O6-C3	4.7585	4.0260	0.5366	3.7933	0.9315
O6-O4	5.0903	4.6207	0.2206	3.4236	2.7779
Cluster III RMSD (Å)			<b>0.414</b>		<b>0.959</b>

C1-O3	4.2041	4.6040	0.1599	3.6249	0.3354
C1-C4	3.5721	3.5429	0.0009	3.8368	0.0700
C1-O5	3.7989	4.0533	0.0647	4.2663	0.2184
C1-C6	3.9550	3.7209	0.0548	5.0852	1.2774
C1-C7	3.5721	3.6242	0.0027	3.8220	0.0624
C1-O8	4.2041	4.0717	0.0175	3.6141	0.3480
O2-O3	4.9988	5.6986	0.4898	3.5810	2.0100
O2-C4	4.2041	4.6547	0.2031	3.6418	0.3162
O2-O5	3.3202	4.4064	1.1799	3.3286	6.9756E-05
O2-C6	3.7989	4.1654	0.1343	4.2721	0.2238
O2-C7	4.2041	4.2820	0.0061	3.6667	0.2888
O2-O8	4.9988	4.4599	0.2904	3.5959	1.9681
O3-O5	4.9988	4.4782	0.2710	3.6011	1.9535
O3-C6	4.2041	4.0402	0.0269	3.6442	0.3134
O3-C7	3.7989	3.9693	0.0290	4.2733	0.2250
O3-O8	3.3202	5.0350	2.9405	3.3218	2.6634E-06
C4-O5	4.2041	4.0981	0.0112	3.6413	0.3168
C4-C6	3.5721	3.7054	0.0178	3.8141	0.0585
C4-C7	3.9550	3.6562	0.0893	5.1026	1.3170
C4-O8	3.7989	4.7113	0.8325	4.2709	0.2227
O5-C7	4.2041	4.7435	0.2910	3.6374	0.3211
O5-O8	4.9988	5.0400	0.0017	3.5946	1.9716
C7-C6	3.5721	3.6815	0.0121	3.8155	0.0592
C6-O8	4.2041	3.9480	0.0656	3.6413	0.3168
Cluster RMSD (Å)	IV		<b>0.547</b>		<b>0.769</b>

O1-C3	4.5840	4.5020	0.0067	4.5729	0.0001
O1-O4	5.0262	4.9837	0.0018	4.5851	0.1946
O1-C5	4.5840	4.5022	0.0067	4.5727	0.0001
O1-O6	5.0262	4.9847	0.0017	4.5848	0.1949
O1-C7	4.0002	3.8967	0.0107	3.6466	0.1250
O1-O8	5.0700	5.0261	0.0019	4.5738	0.2462
C2-C3	3.6450	3.5928	0.0027	3.7917	0.0216
C2-O4	3.9738	4.0122	0.0015	3.6241	0.1223
C2-C5	3.6447	3.5925	0.0027	3.7918	0.0217
C2-O6	3.9738	4.0126	0.0015	3.6239	0.1224
C2-C7	3.6260	3.6426	0.0003	3.6953	0.0048
C2-O8	4.5673	4.6947	0.0162	4.4017	0.0274
C3-C5	3.9519	3.6480	0.0923	3.8516	0.0101
C3-O6	3.8036	4.1853	0.1457	3.7922	0.0001
C3-C7	3.7532	3.6375	0.0134	3.8123	0.0035
C3-O8	3.8888	3.9793	0.0082	3.5793	0.0958
O4-C5	3.8036	4.1861	0.1463	3.7920	0.0001
O4-O6	3.3318	4.4461	1.2418	3.3958	0.0041
O4-C7	4.6001	4.7421	0.0202	4.5836	0.0003
O4-O8	4.6923	5.1121	0.1762	4.4459	0.0607
C5-C7	3.7532	3.6376	0.0134	3.8122	0.0035
C5-O8	3.8888	3.9793	0.0082	3.5793	0.0958
O6-C7	4.6001	4.7423	0.0202	4.5835	0.0003
O6-O8	4.6922	5.1119	0.1761	4.4459	0.0607
Cluster V RMSD (Å)			<b>0.297</b>		<b>0.243</b>

	MP2 nonbonded distances (Å)	ONIOM-EE Nonbonded distances obtained with default parameters (Å)	Square deviation of	ONIOM-EE nonbonded distances obtained with optimized parameters (Å)	Square of the deviation
O1-O10	3.4021	4.9930	2.5309	3.4119	0.0001
C2-O8	4.0033	4.8635	0.7399	3.6419	0.1306
O1-C9	3.8209	4.6337	0.6607	3.8608	0.0016
O1-C3	3.8959	4.5093	0.3762	3.4721	0.1796
O1-O4	4.9890	5.5726	0.3406	3.6421	1.8141
O1-C7	4.4445	3.8754	0.3238	3.5949	0.7218
O1-O6	4.8865	4.3574	0.2799	3.7241	1.3512
C3-O8	4.6295	4.1505	0.2295	4.5726	0.0032

O4-O10	4.9555	4.5228	0.1872	3.4856	2.1605
O8-O10	7.0079	7.4338	0.1814	6.0200	0.9761
O6-O10	4.7835	5.1820	0.1587	4.1843	0.3591
O1-C5	4.6653	4.2803	0.1482	4.4981	0.0280
O4-O8	4.8054	4.4249	0.1448	4.2660	0.2910
O4-C9	4.4727	4.0976	0.1407	3.5734	0.8086
O4-C7	3.8779	4.1764	0.0891	3.5896	0.0831
O6-O8	4.2574	4.5426	0.0814	3.3753	0.7781
C5-O8	3.7386	3.9895	0.0629	3.7190	0.0004
O6-C7	4.4550	4.2082	0.0609	3.7474	0.5008
C7-O10	6.4330	6.6735	0.0578	5.9129	0.2705
C2-O10	3.7368	3.9557	0.0479	3.6463	0.0082
C2-C9	3.8234	3.6061	0.0472	3.8897	0.0044
C2-C7	3.5821	3.7962	0.0458	4.0789	0.2468
O6-C9	3.9180	4.0904	0.0297	3.5271	0.1528
C5-C9	3.7274	3.5729	0.0239	3.8717	0.0208
C3-O6	4.6038	4.7496	0.0212	4.5598	0.0019
C2-C3	3.6684	3.5513	0.0137	4.4425	0.5992
O8-C9	6.5474	6.6589	0.0124	5.9426	0.3658
C3-O10	3.9667	4.0506	0.0070	3.6181	0.1215
C7-C9	6.0454	5.9701	0.0057	5.9218	0.0153
C3-C5	3.6226	3.6953	0.0053	4.4871	0.7474
C3-C7	3.6964	3.6247	0.0051	3.8106	0.0130
C3-C9	3.6063	3.6760	0.0049	4.0384	0.1867
C2-O6	3.9479	4.0136	0.0043	3.4566	0.2414
C2-O4	4.6762	4.6126	0.0040	4.4114	0.0701
C5-C7	3.7282	3.6674	0.0037	3.8467	0.0140
C2-C5	3.7446	3.7090	0.0013	4.4333	0.4743
C5-O10	4.6217	4.6546	0.0011	4.6018	0.0004
O4-C5	3.9822	3.9552	0.0007	3.4636	0.2690
O1-O8	5.0036	4.9789	0.0006	3.4738	2.3400
O4-O6	5.0732	5.0845	0.0001	3.7032	1.8769
Cluster RMSD (Å)	VI		<b>0.421</b>		<b>0.675</b>

O1-C3	3.8915	4.5553	0.4405	4.4522	0.3143
O1-O4	4.8569	5.6207	0.5834	3.4624	1.9444
O1-C5	4.0831	4.0996	0.0003	4.5989	0.2660
O1-O6	4.9818	4.2796	0.4930	5.1839	0.0409
O1-C7	4.6972	4.2043	0.2429	5.2298	0.2837
O1-O8	4.9377	5.3301	0.1540	4.2795	0.4332
O1-C9	3.8214	4.3960	0.3301	3.9643	0.0204
O1-O10	3.3448	4.5820	1.5306	3.4484	0.0107
O1-C11	5.9257	6.1161	0.0362	6.4080	0.2327
O1-O12	5.9592	6.9791	1.0402	6.3058	0.1201
C2-C3	3.6797	3.3911	0.0833	4.5263	0.7166
C2-O4	4.6697	4.5002	0.0287	3.5694	1.2107
C2-C5	3.5144	3.7541	0.0574	3.7562	0.0584
C2-O6	4.2423	4.2212	0.0004	4.4108	0.0284
C2-C7	3.6714	3.5572	0.0130	4.4384	0.5882
C2-O8	3.8635	4.6925	0.6873	3.4421	0.1775
C2-C9	3.9760	3.6105	0.1336	4.0467	0.0050
C2-O10	3.8204	4.0362	0.0466	3.4244	0.1569
C2-C11	5.3839	5.1759	0.0433	6.0591	0.4560
C2-O12	5.3062	6.0664	0.5779	5.8205	0.2644
C3-C5	5.3058	4.9071	0.1590	6.2088	0.8153
C3-O6	5.6609	5.8719	0.0445	5.9564	0.0873
C3-C7	3.5868	3.5526	0.0012	3.7815	0.0379
C3-O8	4.3028	4.2275	0.0057	3.6725	0.3972
C3-C9	3.6798	3.5265	0.0235	4.5113	0.6914
C3-O10	3.8913	4.5918	0.4906	3.4773	0.1714
C3-C11	3.5867	3.5980	0.0001	3.7673	0.0326
C3-O12	4.3006	4.4828	0.0332	4.4540	0.0235
O4-C5	6.4017	5.8712	0.2815	5.8184	0.3403
O4-O6	6.6813	6.8671	0.0345	5.7759	0.8197
O4-C7	4.1340	4.2506	0.0136	3.7096	0.1802
O4-O8	4.9927	4.7096	0.0802	3.3996	2.5380
O4-C9	4.6707	4.3383	0.1106	4.4228	0.0615
O4-O10	4.8575	5.4306	0.3284	3.3676	2.2199
O4-C11	4.1348	3.9414	0.0374	4.4692	0.1118
O4-O12	4.9913	4.7274	0.0697	4.9387	0.0028

C5-C7	4.3910	3.5999	0.6259	4.4671	0.0058
C5-O8	3.6694	4.1708	0.2514	3.4804	0.0357
C5-C9	3.5147	3.9294	0.1720	3.8404	0.1061
C5-O10	4.0826	3.9294	0.0235	3.6933	0.1515
C5-C11	4.3917	3.5183	0.7627	5.1951	0.6455
C5-O12	3.6690	4.0614	0.1539	4.3057	0.4054
O6-C7	4.1538	4.0572	0.0093	4.2089	0.0030
O6-O8	3.2238	4.5801	1.8396	3.3543	0.0170
O6-C9	4.2434	4.6207	0.1423	3.6133	0.3970
O6-O10	4.9819	4.7508	0.0534	3.4618	2.3106
O6-C11	4.1553	4.5589	0.1628	4.3076	0.0232
O6-O12	3.2243	5.0778	3.4356	3.3216	0.0095
C7-C9	5.3838	5.2016	0.0332	5.4599	0.0058
C7-O10	5.9248	6.0428	0.0139	4.4362	2.2159
C7-C11	3.8993	3.8277	0.0051	3.7758	0.0152
C7-O12	3.7778	4.9405	1.3517	3.6647	0.0128
O8-C9	5.3086	5.9943	0.4702	4.5423	0.5873
O8-O10	5.9603	6.8780	0.8421	3.5591	5.7658
O8-C11	3.7814	3.8904	0.0119	3.7444	0.0014
O8-O12	3.3299	4.9521	2.6316	3.4678	0.0190
C9-C11	3.6722	3.6853	0.0002	4.0975	0.1809
C9-O12	3.8642	3.9471	0.0069	3.8669	0.0000
O10-C11	4.6977	4.6212	0.0059	3.4798	1.4835
O10-O12	4.9378	4.7509	0.0349	3.3718	2.4522
Cluster VII RMSD (Å)			<b>0.595</b>		<b>0.738</b>

### *CO/HF clusters*

MP2 nonbonded atom-atom distances in Ångström

Cluster I	MP2 nonbonded distance (Å)
F1-O3	3.0970
F1-C4	4.2360
H2-O3	2.1814
H2-C4	3.3216

Cluster II	MP2 nonbonded distance (Å)
F1-C3	3.0484
F1-O4	4.1855
H2-C3	2.1262
H2-O4	3.2633

Cluster III	MP2 nonbonded distance (Å)
F6-F5	2.7623
F6-H3	3.3226
F6-O1	4.9598
F6-C2	5.9701
H4-F5	1.8468
H4-H3	2.4722
H4-O1	4.2664
H4-C2	5.3352
F5-O1	3.0001
F5-C2	4.1414
H3-O1	2.0791
H3-C2	3.2205

Cluster IV	MP2 nonbonded distance (Å)
F1-C3	3.0260
F1-O4	4.1602
F1-C5	3.2543
F1-O6	3.7267
H2-C3	2.1026
H2-O4	3.2376
H2-C5	3.0828
H2-O6	3.8127
C3-C5	3.6207
C3-O6	4.6939
O4-C5	4.2639
O4-O6	5.3928

Cluster V	MP2 nonbonded distance (Å)
C2-O4	4.0846
C2-C3	4.9438
C2-O6	4.0846
C2-C5	4.9438
C2-H7	3.5522
C2-F8	4.1649
O1-O4	3.1113
O1-C3	4.0846
O1-H7	2.4641
O1-F8	3.1644
O1-O6	3.1113
O1-C5	4.0846
C3-H7	3.5522
C3-F8	4.1649
C3-O6	4.0846
C3-C5	4.9438
O4-H7	2.4641
O4-F8	3.1644
O4-O6	3.1113
O4-C5	4.0846
H7-O6	2.4640
H7-C5	3.5522
F8-O6	3.1644
F8-C5	4.1649

Cluster VI	MP2 nonbonded distance (Å)	Cluster VII	MP2 nonbonded distance (Å)	Cluster VIII	MP2 nonbonded distance (Å)
F1-F6	2.7348	F1-F3	2.6713	F1-C7	2.9114
F1-H5	3.0337	F1-H4	2.7422	F1-O8	4.0469
F1-O8	3.4481	F1-F6	2.6713	F1-C4	3.7423
F1-C7	3.8840	F1-H5	1.8657	F1-O3	4.6484
F1-O3	3.2694	F1-O7	3.2273	F1-F6	2.7027
F1-C4	3.6703	F1-C8	4.2639	F1-H5	1.7845
H2-F6	1.8478	H2-C3	1.8657	H2-C7	1.9803
H2-H5	2.2874	H2-H4	2.1742	H2-O8	3.1157
H2-O8	3.2371	H2-F6	2.7422	H2-C4	3.4027
H2-C7	3.9166	H2-H5	2.1742	H2-O3	4.4382
H2-O3	3.1087	H2-O7	3.0878	H2-F6	3.2427
H2-C4	3.7401	H2-C8	4.1555	H2-H5	2.3970
F6-O8	2.9959	F3-F6	2.6713	C7-C4	3.5262
F6-C7	4.0381	F3-H5	2.7422	C7-O3	4.6620
F6-O3	3.1526	F3-O7	3.2273	C7-F6	4.8257
F6-C4	4.1480	F3-C8	4.2638	C7-H5	4.1319
H5-O8	2.2537	H4-F6	1.8657	O8-C4	4.0089
H5-C7	3.3643	H4-H5	2.1742	O8-O3	5.0911
H5-O3	2.7674	H4-O7	3.0878	O8-F6	5.8196
H5-C4	3.8619	H4-C8	4.1554	O8-H5	5.1890
O8-O3	3.1052	F6-O7	3.2273	C4-F6	3.1594
O8-C4	3.9964	F6-C8	4.2639	C4-H5	3.2062
C7-O3	3.7619	H5-O7	3.0878	O3-F6	3.4063
C7-C4	4.4388	H5-C8	4.1555	O3-H5	3.7399



Cluster IX	MP2 nonbonded distance (Å)	Cluster X	MP2 nonbonded distance (Å)	Cluster XI	MP2 nonbonded distance (Å)
F2-C7	2.9179	O2-O3	3.9855	F2-F4	2.7674
F2-O8	4.0520	O2-C4	3.9843	F2-H3	3.2135
F2-O4	3.6084	O2-F5	3.6210	F2-C7	4.5994
F2-C3	4.5022	O2-H6	3.7545	F2-O8	5.5731
F2-F5	2.7127	O2-C7	4.6565	F2-H6	4.2964
F2-H6	1.7938	O2-O8	5.3607	F2-F5	5.1966
H1-C7	1.9877	C1-O3	3.9842	H1-F4	1.8627
H1-O8	3.1224	C1-C4	3.6570	H1-H3	2.4088
H1-O4	3.2461	C1-C7	3.6318	H1-C7	3.9937
H1-C3	4.2770	C1-O8	4.2664	H1-O8	5.0446
H1-F5	3.2664	C1-F5	3.2356	H1-H6	3.3888
H1-H6	2.4147	C1-H6	3.0989	H1-F5	4.2964
C7-O4	3.2727	O3-C7	4.6565	F4-C7	2.8427
C7-C3	4.4086	O3-O8	5.3607	F4-O8	3.9778
C7-F5	4.8264	O3-F5	3.6210	F4-H6	1.8627
C7-H6	4.1296	O3-H6	3.7545	F4-F5	2.7674
O8-O4	3.7319	C4-C7	3.6318	H3-C7	1.9043
O8-C3	4.8012	C4-O8	4.2664	H3-O8	3.0394
O8-F5	5.8058	C4-F5	3.2356	H3-H6	2.4088
O8-H6	5.1753	C4-H6	3.0989	H3-F5	3.2135
O4-F5	3.2038	C7-H6	2.0766	C7-H6	3.9937
O4-H6	3.1716	C7-F5	3.0000	C7-F5	4.5994
C3-F5	3.3969	O8-F5	4.1310	O8-H6	5.0446
C3-H6	3.6619	O8-H6	3.2101	O8-F5	5.5731

The first observation we can make from Table 4.3 is that electron densities are unevenly distributed in the rings of the clusters with external monomers. It is as if some of the monomers have a larger electron density at their bond critical points than the others in the rings. Cluster VIII differs from the other clusters as we see that the external monomer's electron density at its bond critical point is almost as much as in the case of the isolated monomer. This shows that there is only a very weak interaction between the monomer and the ring. We investigated this further by plotting the second derivative of the electron density of this cluster in Fig. 4.9. The plot is in a plane spanned by the external monomer and a fluorine atom in the ring. The plot clearly shows that this interaction is a closed-shell interaction as the topology of the second derivative is almost spherical symmetrical around the fluorine atom in the external monomer and spherical symmetrical around the fluorine atom in the monomer in the ring.

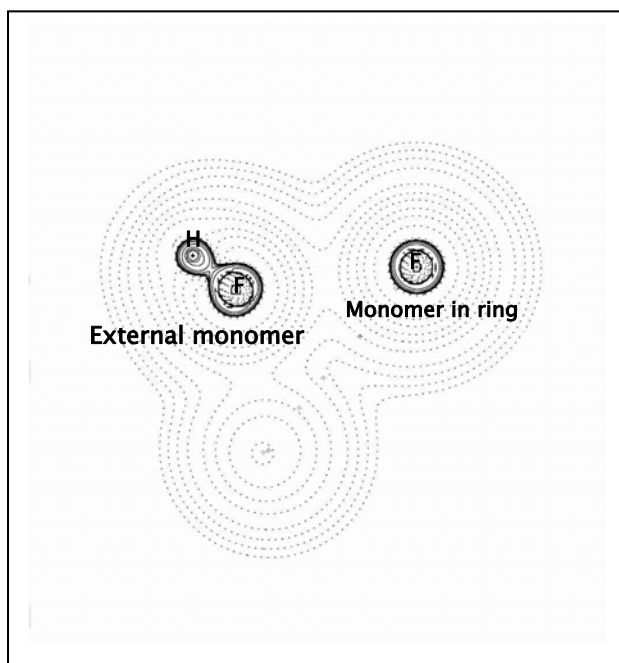


Fig. 4.9: Plot of the second derivative of the electron density of the external monomer and a monomer in the ring of cluster VIII.

Earlier we mentioned that the monomer in cluster VIII has a different interaction with the ring than in the other clusters with attached monomers.

Regarding Table 4.3, we also see an increase in the volume the fluorine occupies in a monomer in the cyclic clusters, with respect to a single isolated monomer, with an increase in cluster size. This can be deduced from the relative shift of the bond critical point (BCP). Although this shift is almost insignificant, it does allow a trend to be observed. This trend can be interpreted as a small shift of the electron density towards the hydrogen atom, possibly because of a decrease in effective electro-negativity of the fluorine atoms; however, this seems not to play a significant role. Our AIM analysis again proves our earlier explanation

that electron density moves out of valence bonds during cluster formation. The amplitude of maximum electron density in the valence bonds decreases, but its position with respect to the two atoms is not significantly affected.

With respect to the ring critical points, we see that the cyclic trimer has the largest electron density at its ring critical point. The angle at which each monomer is turned with respect to the other monomers in cyclic clusters probably determines the degree of overlap of the electron densities at the ring critical points and the larger the overlap of the electron densities of the monomers in the center of the ring, the larger the magnitude of the electron densities at the ring critical points. Following the literature, the trend in the density at the ring critical points, for the cyclic clusters with respect to their size, is the same as the trend in *Nuclear Independent Chemical Shift* (NICS) values, which is a measure of the strength of the ring current with respect to cluster size [Rehaman *et al.*, 2006, Maerker *et al.*, 1997]. Therefore, the electron density at ring critical points might be linked to ring currents. This should, however, be confirmed in future work. Note that the ring critical point in cluster VIII is between the fluorine atom in the external monomer and two fluorine atoms of the two adjacent monomers leading to a total of 5 ring critical points. All the other clusters with ring critical points have their ring critical points at the center of their rings. We also see that the clusters with external monomers have only a slight decrease in the electron density at the ring critical point. For example, adding an external monomer to the cyclic trimer distorts the electron density in such a way that the ring critical point still has a larger electron density than the cyclic tetramer. We might therefore expect that when adding an external monomer to the cyclic HF trimer as in cluster IV, the destabilization should not be that significant and if a ring current is present it would not be affected significantly.

#### ***4.6 Analysis of the electrostatics of the HF clusters***

To develop a hybrid QM/MM system for the HF clusters one needs to approximate the MM system's charge density by atomic point charges. It is important to realize that by putting a point charge on an atom, the charge distribution is approximated by a sphere, which can be a severe approximation due to the anisotropic nature of the charge density of atoms in molecules. The ONIOM-EE method in *Gaussian 03* can however only use atomic point charges. As charges are a controversial issue [Martin and Zipse, 2005], we determined not only the MKS charges, but as a reference for these charges, we also determined the *natural atomic orbital* (NAO) and the *Atoms in molecules* (AIM) charges for a few selected clusters.

As MKS charges are based on the Electrostatic Potential (ESP) of the electron density, we studied the ESP-maps to see if they are comparable to the charges. In Table 4.4, pictorial

representations of the ESP-maps, based specifically on the MP2 density of each cluster, are given together with the MKS charges based on the SCF and MP2 densities. It is important to note that charges derived from the SCF densities are different from charges derived from the MP2 densities. The SCF density are based on the ground state Hartree-Fock wave function, while the MP2 density is based on a first order corrected wave function, containing second excitations of the Hartree-Fock determinant.

Table 4.4: In each case the electrostatic potential mapped onto the total MP2 density is shown on the left side of the table and on the right are the Merz-Kollman-Singh (MKS) AIM and NAO charges for the atoms as numbered in Fig. 4.2. In this case, blue illustrates a positive potential and red a negative potential. All the charges are given in atomic units.

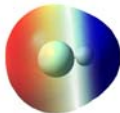
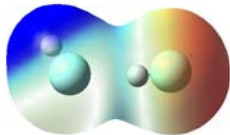
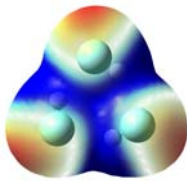
ESP mapped on total MP2 electron density	MKS charges calculated from the total SCF density	MKS charges calculated from the total MP2 density [AIM charges]	NAO charges
<b>Monomer</b> 	F1: -0.474 H2: 0.474	F1: -0.452 [-0.713] H2: 0.452 [0.711]	F1: -0.540 H2: 0.540
<b>I</b> 	F1 : -0.484 H2 : 0.458 F3 : -0.461 H4 : 0.487	F1: -0.464 [-0.739] H2: 0.438 [0.730] F3: -0.438 [-0.718] H4: 0.465 [0.728]	F1: -0.561 H2: 0.551 F3: -0.549 H4: 0.558
<b>II</b> 	F1 : -0.452 H2 : 0.452 H3 : 0.454 H4 : 0.451 F5 : -0.454 F6 : -0.452	F1: -0.427 H2: 0.428 H3: 0.430 H4: 0.426 F5: -0.429 F6: -0.427	F1: -0.568 H2: 0.568 H3: 0.568 H4: 0.568 F5: -0.568 F6: -0.568

Table 4.4 continued...

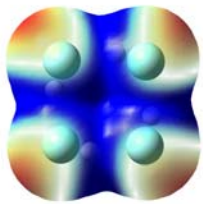
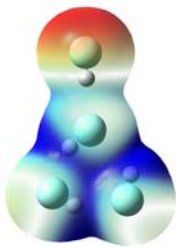
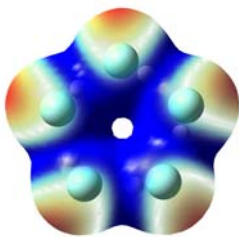
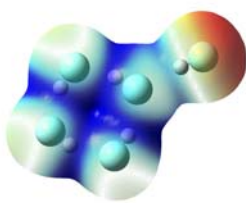
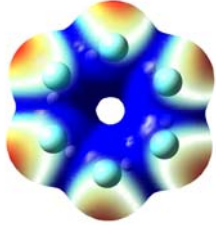
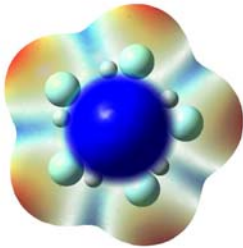
<p>III</p> 	<p>F1 : -0.436 H2 : 0.436 F3 : -0.435 H4 : 0.433 F5 : -0.423 H6 : 0.420 H7 : 0.438 F8 : -0.435</p>	<p>F1: -0.411 H2: 0.412 F3: -0.410 H4: 0.408 F5: -0.396 H6: 0.394 H7: 0.414 F8: -0.409</p>	<p>F1: -0.573 H2: 0.573 F3: -0.573 H4: 0.573 F5: -0.573 H6: 0.573 H7: 0.573 F8: -0.573</p>
<p>IV</p> 	<p>F1 : -0.455 H2 : 0.473 F3 : -0.476 H4 : 0.463 F5 : -0.445 H6 : 0.465 H7 : 0.455 F8 : -0.480</p>	<p>F1: -0.430 [-0.730] H2: 0.449 [0.740] F3: -0.451 [-0.763] H4: 0.437 [0.753] F5: -0.419 [-0.736] H6: 0.441 [0.749] H7: 0.430 [0.731] F8: -0.457 [-0.741]</p>	<p>F1 : -0.558 H2 : 0.568 F3 : -0.586 H4 : 0.574 F5 : -0.559 H6 : 0.572 H7 : 0.550 F8 : -0.562</p>
<p>V</p> 	<p>F1 : -0.440 H2 : 0.445 F3 : -0.443 H4 : 0.439 F5 : -0.421 H6 : 0.413 F7 : -0.421 H8 : 0.427 F9 : -0.430 H10 : 0.432</p>	<p>F1: -0.417 H2: 0.423 F3: -0.419 H4: 0.414 F5: -0.394 H6: 0.385 F7: -0.394 H8: 0.401 F9: -0.405 H10: 0.407</p>	<p>F1 : -0.575 H2 : 0.575 F3 : -0.575 H4 : 0.575 F5 : -0.575 H6 : 0.575 F7 : -0.575 H8 : 0.575 F9 : -0.575 H10 : 0.575</p>
<p>VI</p> 	<p>F1 : -0.431 H2 : 0.439 F3 : -0.432 H4 : 0.424 F5 : -0.421 H6 : 0.438 H7 : 0.438 F8 : -0.429 H9 : 0.453 F10 : -0.480</p>	<p>F1: -0.405 H2: 0.414 F3: -0.405 H4: 0.395 F5: -0.393 H6: 0.412 H7: 0.413 F8: -0.402 H9: 0.430 F10: -0.458</p>	<p>F1: -0.566 H2: 0.574 F3: -0.593 H4: 0.577 F5: -0.565 H6: 0.576 H7: 0.575 F8: -0.568 H9: 0.548 F10: -0.559</p>

Table 4.4 continued ...

<p>VII</p> 	<p>F1 : -0.445            F2 : -0.436            F3 : -0.436            F4 : -0.438            F5 : -0.434            F6 : -0.432            H7 : 0.445            H8 : 0.434            H9 : 0.438            H10 : 0.431            H11 : 0.433            H12 : 0.440</p>	<p>F1: -0.407            F2: -0.408            F3: -0.414            F4: -0.410            F5: -0.411            F6: -0.422            H7: 0.405            H8: 0.408            F9: 0.415            H10: 0.421            H11: 0.409            H12: 0.413</p>	<p>Not calculated</p>
<p>VIII</p> 	<p>F1: -0.448            H2: 0.449            F3: -0.447            H4: 0.451            F5: -0.445            H6: 0.448            F7: -0.448            H8: 0.452            F9: -0.450            H10: 0.455            F11: -0.544            H12: 0.527</p>	<p>F1: -0.421            H2: 0.423            F3: -0.420            H4: 0.425            F5: -0.418            H6: 0.422            F7: -0.422            H8: 0.426            F9: -0.424            H10: 0.429            F11: -0.523            H12: 0.504</p>	<p>Not calculated</p>

All the MKS charges conform to their ESP-maps, and no erroneous results were found. Obviously all the values for the charges for the different charge schemes are different. The main difference between the charge schemes is that MKS samples points of the electron density on and further than the overlap of the van der Waals radii of the atoms, whereas both NAO charges and AIM charges are based on the total charge of the atom, and therefore include the direct contribution of valence electrons to the atomic charge.

It is interesting to note that the difference in magnitude of the charges differs for the H2 and F3 atoms for the HF dimer for the various charge schemes. For the MKS charges based on the SCF density there is a slight difference between the absolute values of the charges on these two atoms. For the MKS charges based on the MP2 density, the absolute values of these charges are identical. The AIM charges show a larger difference in their absolute values. The absolute value of the hydrogen's atomic charge is larger than the absolute value of the fluorine's atomic charge. The NAO charges show a slight difference as well. The hydrogen is again more positive than the fluorine is negatively charged. This has to do with the

partitioning of the electron density. MKS does not partition the electron density as AIM and NAO charge-schemes do. Again, AIM and NAO charge schemes use localized approaches to derive the charges whereas MKS uses a more delocalized approach only dependent on the ESP of the wave function. For the HF clusters, MKS charges based on the MP2 density seem to be the best choice. It might also be that *electron correlation* leads to *electron delocalization* and this is why the SCF and MP2 charges differ in their absolute values. For the cyclic cluster NAO charges conform better to the symmetry point group of the molecule than MKS charges. Again, this is probably due to the localization of the correlated wave function onto the HF monomers.

In the clusters with external monomers attached to rings, there is a slight redistribution of charge from the ring to the external monomer. In the spirit of quantum mechanics, the likelihood of finding electrons is greater on the external monomer than on the ring or in other words there is a charge transfer from the ring to the external monomer. This is confirmed by a comparison between the total charge of the monomer and the total charge of the ring in each case.

#### ***4.7 Summary, conclusions and future work***

In this chapter we reported the optimization of eight HF clusters. The structural parameters compared very well to the experimental ones. The energies and vibrations also seemed reasonably accurate compared with other literature results and experimental results when the vibrations of the normal modes are scaled by a scaling factor. We then studied the bonding of the clusters. We gave an explanation regarding hydrogen bonding based on Mulliken population analysis results, visualizations of the MP2 wave function and the theory of *Atoms in Molecules*. We showed that for delocalized systems such as the cyclic HF clusters that NBO analysis is not feasible. NBO analysis could not give any explanation for why a ring current can be detected in the HF trimer and why no significant ring current can be calculated for the other cyclic clusters. By studying the *relative electron population* associated with the correlated wave function, we hypothesized that the reason for the ring current in the trimer can be due to the fact that fewer electrons need to be excited from a lower shell to a higher shell as this higher shell already contains sufficient electron density.

We reported charges derived from different charge derivation schemes and concluded that MKS charges based on the MP2 density do indeed give a better representation of the charge density of the systems than other charges. This was confirmed by studying the ESP-maps. See Table 4.4.

Future work might include the validation of the intermolecular orbital explanation and the reason given for the ring current in the HF trimer. In addition, it is interesting that cluster VIII explicitly “chose” not to form a hydrogen bond with one of the fluorine atoms in the ring, but rather to form weak interactions with all the fluorine atoms of the ring instead. We cannot explain this behavior from our work and future work might elucidate the reason for this. This can be accomplished by determining if a series of vdW bonds are more stable than a single hydrogen bond and if other factors are also present. It will also be important to validate the relationship between electron density at ring critical points and “aromaticity” in HF clusters or even other clusters. It is important to note that all the models discussed in this chapter are a consequence of scientific questioning. We can never assume that any model as such is “correct”. It can only be correct within the limits of the methods used.



## Chapter 5

# Optimization of the van der Waals parameters of the Universal Force Field (UFF) for the use in ONIOM-EE optimizations of hydrogen fluoride clusters

### *5.1 Introduction*

In this chapter, we report the optimization of van der Waals force field parameters to be used for UFF in QM/MM optimizations of hydrogen fluoride clusters by using the ONIOM-EE methodology.

In **Section 5.2** we will discuss the computational details, while in **Section 5.3** we will discuss our strategy to minimize trial optimizations. **Section 5.4** will contain details regarding the assessment of the quality of the force field parameters and in **Section 5.5** we will discuss the actual optimization of the force field parameters. In **Section 5.6** we will report the development of a method for force field optimizations in a hybrid system. The chapter will be concluded with **Section 5.7**.

### *5.2 Computational details*

All the ONIOM-EE optimizations were performed with the *Gaussian 03* set of programs. The computational details have been discussed in **Chapter 3 (Section 3.5)**.

The UFF [Rappé *et al.*, 1992] van der Waals parameters were modified manually for each ONIOM-EE optimization. Further details regarding the actual optimization of the van der Waals parameters will follow shortly. The nonbonded interactions were computed geometrically and a general  $1/R$  Coulomb description, was used for the electrostatics, both as implemented in UFF. Before attempting ONIOM calculations and optimizations, it was verified that *Gaussian 03* gives the correct energies for the van der Waals and electrostatic

interactions. The procedure for this can be found in **Appendix C**. We further had to find methods to minimize the trials needed to find optimized parameters for the force field, because geometry optimizations were laborious. One of these methods will be discussed in **Section 5.3**. In **Chapter 9** we will attempt to explain the slow convergence in the micro-iterations when using frozen bond lengths and electronic embedding.

We used very tight optimization conditions, as default optimization conditions were inadequate for our purpose<sup>1</sup>.

### ***5.3 Strategy to limit trials in optimization of Force field (FF) - parameters***

Optimizing van der Waals parameters for a force field by doing trial geometry optimizations is generally difficult, but can be done in a reasonable time. When the geometry optimization takes longer than an *ab initio* optimization, however, it is annoying and almost impossible to optimize the parameters. Since this is what we experienced, we tried another method. According to the literature one can freeze the MM system and make the QM system move as a rigid block during the micro-iterations. This has been tested and is known to lead to fast convergence [Vreven *et al.*, 2003]. It would therefore be better to use only one monomer with frozen Cartesian coordinates in the MM system and designate the remaining part of the system to the QM system. By doing this we increase the time necessary to compute the energy of the QM system, but we minimize the time taken for a geometry optimization. Our approach might seem strange at first, but the time to optimize rigid blocks in the MM system in the micro-iterations with electronic embedding, is at this stage more laborious than computing the energy of more molecules in the QM system. Our approach is a “trick” to make use of the properties of ONIOM-EE optimizations in *Gaussian 03* that has been developed and tested, instead of using a method that has never been used before. We therefore use a variation on our original hybrid system as discussed in **Chapter 3 (Section 3.5)**. In Fig. 5.1 we call the original hybrid system the *Final set* and the new hybrid system the *Intermediate set*.

---

<sup>1</sup> In order to obtain good results with default optimization conditions it is important that one is as close as possible to the minimum otherwise we found that the optimizer will end up at various minima.

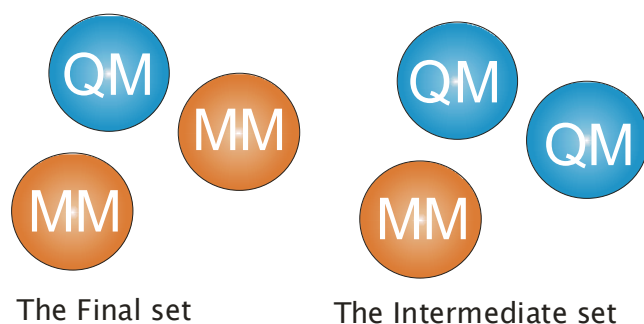


Fig. 5.1: Two different ways of representing the HF trimer in ONIOM-EE

We first tried to optimize the *Intermediate set* by systematically varying vdW parameters. In this set, the monomers in the QM system experience a quantum mechanical interaction between each other and a molecular mechanical interaction with the single MM monomer. So the only difference between the two sets is that in the *Final set* the hydrogen bonds are described by molecular QM/MM interactions whereas in the *Intermediate set* they are described by a QM-QM interaction and a QM-MM interaction.

#### ***5.4 Criteria for assessing the quality of the Force field (FF)***

Usually one gauges the quality of a force field by comparison of computational results to experimental data. For example, heats of formation, vibrations and geometries can be used. In our case, we had to rely on the MP2 optimized geometries. We assumed that the force field would be optimized when the root mean square deviation (RMSD) of the nonbonded distances with respect to the MP2 values, converged. We used the root mean square deviation (RMSD) as defined in **Chapter 2 (Section 2.10)**. Our aim was therefore to find parameters to come as close as possible to convergence in a reasonable time of *ca.* one month.

#### ***5.5 Optimization of FF-parameters***

##### *5.5.1 Testing for convergence*

The optimization of the eight HF clusters was simplified by using a small training set. The final optimized parameters were tested on the remaining clusters. The training set used contained the HF dimer, cyclic trimer, cyclic tetramer and cyclic pentamer. The change in the

geometries from their starting points after each ONIOM optimization was measured by a RMSD from the *ab initio* values of all the nonbonded distances in the cluster.<sup>2</sup>

Using the *Intermediate set* for the trimer, tetramer and pentamer, we optimized the Lennard-Jones parameters for these clusters by manually varying the parameters systematically and doing geometry optimizations. To simplify the task at hand, only the  $R_0$ -values or van der Waals equilibrium distances were optimized. Murphy and coworkers did the same when they optimized a force field for amino acids [Murphy *et al.*, 2000]. We eventually determined optimized parameters of 3.52 Å for the fluorine atom-type (F\_) and 1.28 Å for the hydrogen atom-type (H\_). These values were then used to optimize the van der Waals parameters for the *Final set* for all the clusters in the training set yielding 3.34 Å for F\_ and 1.28 Å for H\_. As the cyclic pentamer is the largest cluster in the training set, and therefore contributes the most to the overall RMSD, we used the change in the intermolecular distances of the pentamer as a measure of convergence.

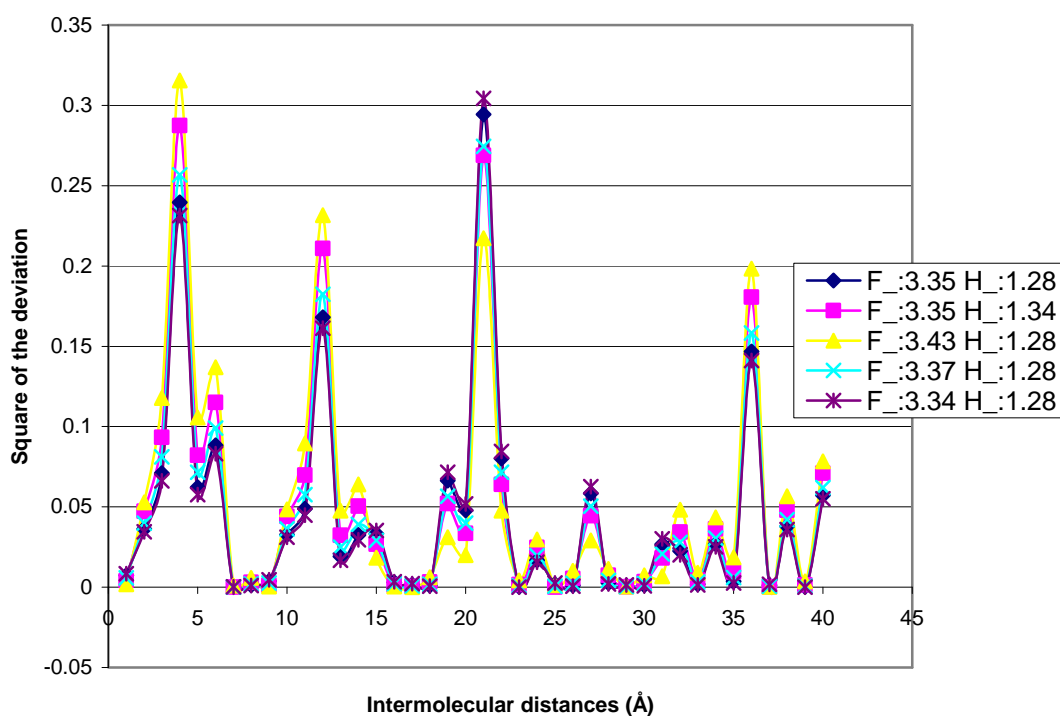


Fig. 5.2 : Graph showing the change in the different intermolecular distances of the cyclic HF-pentamer as they vary with the different  $R_0$ -Lennard-Jones parameters for each atom-type. The values for the parameters in the legend are in Ångström.

<sup>2</sup> A nonbonded distance is the distance between two nonbonded atoms. For example, for the cyclic tetramer we measured 24 nonbonded distances each time the parameters were changed. All these distances are listed in **Appendix D**.

Fig. 5.2 shows the RMSD for the separate nonbonded distances in the optimized pentamer for various selected  $R_0$ -values for the fluorine and the hydrogen atom-types. Here one can see that the changes in the intermolecular distances F1-H6 and F3-F9 are almost inversely proportional to each other with respect to the change in parameters. Therefore, if we lower the  $R_0$ -value of the fluorine atom-type and keep the  $R_0$ -value of the hydrogen atom-type constant, the F1-H6 intermolecular distance is “benefited” (more like in the MP2 optimized cluster) on this section of the optimization hyper-surface.<sup>3</sup> We also found that by keeping the  $R_0$ -value of the fluorine atom-type at 3.350 Å and increasing the  $R_0$ -value of the hydrogen atom-type, this worsened the force field. We took this behavior as a sign of convergence of the RMSD to a minimum.

One can argue that the  $R_0$ -value for the hydrogen atom is quite small. First, we need to consider that the vdW interaction is only dependent on the attractive dispersion interaction and the repulsive exchange interaction. As the fluorine atom of the HF monomer has a large electron-negativity the single electron on the hydrogen atom will not play a significant role in the attractive dispersion or repulsive exchange interactions. The point charge on the hydrogen atom is sufficient enough to describe its interaction with other nonbonded atoms.

### *5.5.2 Results for the newly determined $R_0$ -values vs the default $R_0$ -values*

The newly determined Lennard-Jones parameters were tested on the training set and compared to the results obtained when using the default parameters. Table 5.1 summarizes the results obtained for the training set.

---

<sup>3</sup> We also looked at a visual representation of the geometry changes to obtain an idea of how the parameters affect the pentamer’s geometry overall.

Table 5.1 : Results for the optimization of the training set using different Lennard-Jones parameters. The well-depths for the fluorine and hydrogen atoms were kept constant at the UFF values of 0.050 kcal/mol and 0.044 kcal/mol respectively. MP2 density based Merz-Kollman-Singh charges were used.

CLUSTER	RMSD FOR DEFAULT FORCE-FIELD PARAMETERS  (Å)	RMSD FOR IMPROVED TRAINING SET F <sub>-</sub> -R <sub>0</sub> =3.34Å H <sub>-</sub> -R <sub>0</sub> =1.28Å  (Å)	% improvement
DIMER	0.591	NA**	N/A
TRIMER	0.646	0.139	78.5%
TETRAMER	N/A*	0.146	N/A*
PENTAMER	N/A*	0.213	N/A*
TOTAL RMSD	0.633	0.184	70.9%

\* Micro-iterations did not converge

\*\* Problem with coordinate system as structure is linear

As can be seen from Table 5.1, the new parameters made a significant improvement for the HF trimer.

### 5.5.3 Results for the other clusters

The new van der Waals parameters were then tested on all of the remaining clusters. The final results, excluding the results for the training set, are summarized in Table 5.2.

Table 5.2 : RMSD for the remaining clusters, as calculated by using our Lennard Jones parameters and the UFF-default parameters. The total RMSD calculated for all the clusters is also shown.

	RMS-deviation for default parameters  (Å)	RMS-deviation for optimized parameters (Å)	% improvement
Cluster IV	Drastic change in overall geometry, see discussion	Drastic change in overall geometry, see discussion	N/A
Cluster VI	Drastic change in overall geometry, see discussion	0.222	N/A
Cluster VII	Micro-iterations did not converge	0.248	N/A
Cluster VIII	Drastic change in overall geometry	Drastic change in overall geometry	N/A
TOTAL RMSD	<b>0.633*</b>	<b>0.216</b>	<b>65.9%</b>

\* Only the HF dimer and HF trimer could be optimized

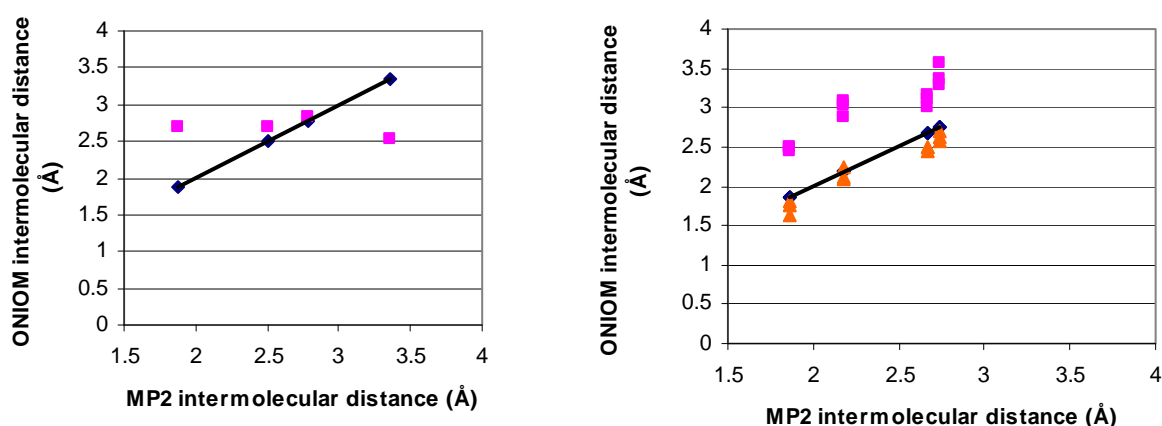
We can see from Table 5.2 that there are drastic changes in the geometries of some of the clusters. Cluster IV did not converge in time and its calculation was deliberately stopped.<sup>4</sup> This was taken as a sign that the micro-iterations had difficulty in optimizing the MM system due to the fact that the MM system moved significantly during the optimization. As we discussed in **Section 2.9**, the optimization algorithm is not able to give fast convergence for

<sup>4</sup> The Gaussian output file was 1 GB in size which is unusual and is a sign of an extremely slow convergence. It was so large that some text editors could not show the contents of the file.

systems where the molecules in the MM system are treated as rigid blocks and move significantly during the optimization. The significant movement of the MM system means the force field parameters for the cluster are poor. Cluster VI and VIII, however, did converge in a reasonable time. In cluster VI the external monomer was not incorporated into the ring, but in cluster VIII the external monomer was incorporated into the pentamer ring structure to form a planar cyclic hexamer. It is therefore possible that over time the monomer in cluster IV will also eventually insert itself into the ring and form the cyclic tetramer. This would depend on how feasible ring insertion is energetically according to the ONIOM-EE extrapolated energy. In Fig. 5.3 we illustrate the results graphically. A trend line is drawn through the datapoints of the MP2 values for the nonbonded (intermolecular) distances. Points close to the trend line are seen as good representations of the MP2 geometry. As we discussed earlier for cluster I, the HF dimer, a successful ONIOM-EE optimization could not be performed, because the optimized structure is linear. The optimized geometry using the default UFF van der Waals parameters, was parallel and shows obvious deviations from the MP2 optimized geometry, which is not parallel. We see however, that for the other clusters that could be optimized with our optimized force field parameters, this resulted in good fits to the MP2 optimized geometries as indicated by small deviations of the nonbonded distances from their MP2 values. As mentioned earlier, the default force field parameters could not be used to optimize Cluster II, V, VI and VII.

### 5.6 Optimizing the force field based on MP2 interaction energies<sup>5</sup>

As we mentioned previously, geometry optimizations were too laborious to be feasible for force field optimizations. We therefore looked for an alternative method. We developed a method based on the frozen geometries of the energy minima obtained with MP2.



<sup>5</sup> MP2 interaction energies are all corrected for BSSE.

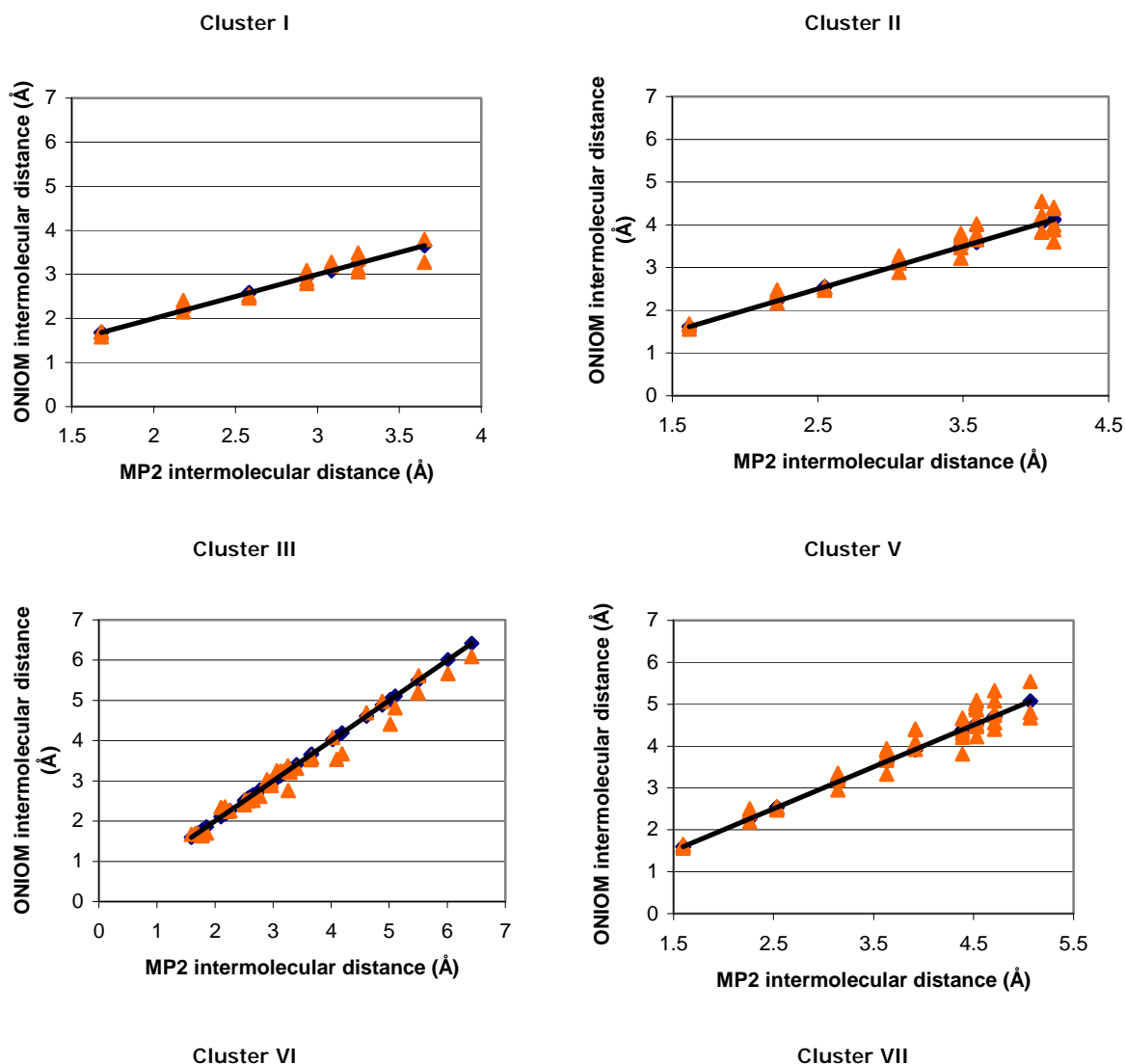


Fig. 5.3: Scattergrams of all the intermolecular distances for all the clusters optimized with ONIOM-EE together with the intermolecular distances optimized with MP2 in **Chapter 4**. A trend line is drawn through the data points of the MP2 results. The squares represent ONIOM calculations with the default UFF-parameters. Triangles represent ONIOM-EE calculations with the optimized parameters. Only the clusters which could be optimized to similar structures to the MP2 energy minima are shown.

This method is closely related to a method previously used to optimize force fields for crystals [Hagler and Wilson, 1974] and is also similar to the one used by Riccardi and coworkers, [Riccardi *et al.*, 2004] to optimize a force field for peptide dimers.

Our method can be seen as an attempt to optimize the force field parameters in order to give a quantum mechanical like interaction energy rather than by requiring the geometries of the clusters to be similar to the equilibrium geometries at the MP2 level of theory. The quality assessment of the force field will therefore change from a comparison of equilibrium geometries to comparing the ONIOM steric energies, *vide infra*, to the MP2 interaction energies and calculating an overall RMSD with respect to these energies. By using the overall RMSD, we automatically add the constraint that the trend in the ONIOM steric energies, with



respect to the cluster size, should be the same as for the MP2 BSSE corrected interaction energies.

To assist with the discussion we first define the ONIOM-EE interaction energy as:

$$E_{\text{interaction}}^{\text{ONIOM}} = E_{\text{real}}^{\text{vdW}} + E_{\text{MM}}^{\text{El}} + \Delta P \quad (5.2)$$

where  $E_{\text{real}}^{\text{vdW}}$  is the total van der Waals interaction between all the molecules in the entire system.  $E_{\text{MM}}^{\text{El}}$  is the Coulomb electrostatic energy for the electrostatic interactions of all the molecules in the MM system.  $\Delta P$  is the electrostatic perturbation, calculated as the difference in the value calculated by ONIOM-EE as  $E_{\text{model}}^{\text{high}}$ <sup>6</sup> (see **Section 2.8**) and the gas phase MP2 energy value of the isolated monomer in the model system. As we discussed in **Chapter 2**, when the MM system is optimized during the micro-iterations, the optimizer uses the ESP charges derived based on the perturbed QM wave function, as a perturbation to the exact gradient in the micro-iterations in order to optimize the MM system. If the energies for the interactions between these charges and the stationary charges on the MM atoms are not reasonably similar, one would also expect the exact force to change incorrectly during the microiterations leading to an inefficient optimization of the MM system. In **Chapter 9**, we will demonstrate that in *Gaussian 03* there is a large difference in the Coulomb electrostatic energy of the system when using point charges to describe the perturbed wave function, and the electrostatic energy calculated during the calculation for the model system. We will see that for the HF clusters this difference is so severe that it is impossible to obtain ONIOM interaction energy values comparable to the MP2 interaction energy. We therefore have to define another term which is unfortunately an approximation, but more suitable for our purposes. This term is the ONIOM steric energy and is defined as:

$$\text{ONIOM}(\text{steric}) = E_{\text{vdW}} + E_{\text{el}} \quad (5.3)$$

where  $E_{\text{el}}$  is the electrostatic energy of the whole system when using the ESP fitted charges on the QM molecule and the stationary charges on the MM molecules. The calculation of the ONIOM steric energy was performed by utilizing Cerius<sup>2</sup>. Before we commenced with the optimization of the force field, the ONIOM steric energies, when using the default UFF parameters, were compared to the ONIOM steric energies calculated by using the geometrically optimized  $R_0$ -values. See Table 5.3.

Table 5.3: ONIOM steric energies calculated by utilizing the default UFF  $R_0$ -values and the geometrically optimized  $R_0$ -values.

---

<sup>6</sup> It would be better to separate the energy for the model system from the energy for the electronic perturbation. However in *Gaussian 03*  $E_{\text{model}}^{\text{high}}$  is the sum of the energy for the model system and the energy for the electronic perturbation. The MKS charges for the QM atoms used during the optimization of the MM system in the micro-iterations are derived based on the ESP of this term's density.

Cluster	MP2 BSSE corrected interaction energy (kcal/mol)	ONIOM steric energy for default R <sub>0</sub> -values (kcal/mol)	ONIOM steric energy for geometrically optimized R <sub>0</sub> -parameters (kcal/mol)
I	-3.8	14.3	-4.6
II	-11.6	50.5	-12.5
III	-21.6	274.6	-21.1
IV	-15.0	86.2	-16.5
V	-30.1	566.0	-29.0
VI	-24.6	326.2	-22.9
RMSD (kcal/mol)	N/A	311.0	1.2

It seems as if the geometrically determined R<sub>0</sub>-values actually perform remarkably well when the quality of the force field is based on the fitting of the ONIOM steric energies to the MP2 interaction energies.

To optimize the R<sub>0</sub>-parameters we made use of the first six HF clusters that we optimized in **Chapter 4**. For the optimization of the parameters by the frozen geometry method, we used a genetic algorithm (GA) [Jensen, 2001]. GA's are discussed in **Chapter 2 (Section 2.8)**. For the GA utilized in this work, 16 chromosomes were used in the population, one-point cross-overs were used with a probability of 90% and the chromosome mutation and random selection probability were both set to 10%. Four preliminary runs with 10 generations per run were used. The maximum number of generations was set to 400, the convergence tolerance on the average deviation to  $1 \times 10^{-5}$  and the numeric precision to 12 decimal digits. The results obtained with the GA were reoptimized with a quasi-Newton optimization algorithm. The best R<sub>0</sub>-values that can be obtained for these six clusters are:

$$R_0(\text{H}_-) = 0.481 \text{ \AA}$$

$$R_0(\text{F}_-) = 3.386 \text{ \AA}$$

The R<sub>0</sub>-value for H<sub>-</sub> is now even smaller than when a series of geometry optimizations was used during the optimization of this value earlier. It is important not to interpret this value as a vdW radius, as this is only a value determined by fitting the ONIOM steric energy to the MP2 interaction energy and is also dependent on the point charge on the hydrogen atom, which is dependent on the accuracy of the charge-scheme used to derive this charge.

In Table 5.4 we summarize the ONIOM steric energies for these six clusters with the optimized parameters.

Table 5.4: ONIOM steric energies as calculated with the energetically optimized R<sub>0</sub>-values

Clusters	MP2 BSSE corrected interaction energy (kcal/mol)	ONIOM steric energy (kcal/mol)
I	-3.8	-4.6
II	-11.6	-12.1
III	-21.6	-21.4
IV	-15.0	-16.2
V	-30.1	-30.5
VI	-24.6	-23.4
<b>RMSD (kcal/mol)</b>	N/A	0.8

From Table 5.4 we see clearly that the fit is reasonable. The reason why it is not better is probably due to the use of point charges instead of a more elaborate description of the electrostatics in the clusters, especially in the HF dimer. For a more accurate description of the electrostatics, the van der Waals parameters would need to be reoptimized.

## 5.7 Summary and conclusions

Our aim in this chapter was to optimize the Lennard-Jones parameters for UFF specifically for use in a QM/MM hybrid system for hydrogen fluoride clusters of various sizes. We first optimized the  $R_0$ -values for the hydrogen and fluorine atom-types by using a series of geometry optimizations. For the method using a series of geometry optimizations, we attempted to accelerate the process of finding  $R_0$ -values by using an *Intermediate set* where only the van der Waals interactions between QM and MM systems were optimized and not the interaction between two or more MM molecules. Results pertaining to this system will be given and discussed in **Chapter 9**.

Finding it laborious to do geometry optimizations, we developed a method of force field optimization basing the quality of the force field only on the ability to reproduce the MP2 interaction energies. Comparing the  $R_0$ -values obtained with the two methods, we saw that when fitting the ONIOM steric energies for both sets of  $R_0$ -values that their ability to reproduce the MP2 interaction energy did not differ much. Future work might include validating all the force field optimization methods discussed in this chapter for other QM/MM systems. Probably the simplest way to optimize the force field will be to remove the constraints on the bond lengths of the molecules in the MM system and to use several geometry optimizations, varying the force field parameters for each. Judged on the similarity in the ONIOM steric energies obtained for the  $R_0$ -values of both force field optimization

methods, the use of the frozen geometry optimization method used in this chapter looks promising.

## Chapter 6

# *Ab initio* calculations and analysis of carbon monoxide clusters

### 6.1 Introduction

Carbon monoxide clusters are found in the earth's atmosphere and are a major component of interstellar clouds such as the Orion nebula. Carbon monoxide is also one of the most poisonous gases known to man.

At room temperature carbon monoxide is a gas, but at low temperatures it can occur as one of two polymorphic crystal structures. At approximately 61.55K the beta-phase of CO, which is an orientationally disordered hcp (hexagonally close packed) structure, converts to the orientationally ordered fcc (face centered cubic) alpha-phase [Askarpour *et al.*, 1989].

The only CO clusters investigated computationally [van der Pol *et al.*, 1990; Vissers *et al.*, 2005; Han and Kim, 1997; Meredith and Stone, 1998] and experimentally [Surin *et al.*, 2003; Surin *et al.*, 2004; McKellar, 2004] to date (*anno* 2006) are the CO dimers. Mass spectrometry data are also available for the CO trimer [Muigg *et al.*, 1998].

Presently it is accepted that there are two stable CO dimers; a global minimum, which is a C-bonded<sup>7</sup> slipped antiparallel structure, and a local minimum, which is an O-bonded slipped antiparallel structure. These are the only structures that can account for the experimentally observed stacks of rotational levels with different rotational constants in CO dimers [Vissers *et al.*, 2005; Surin *et al.*, 2003]. It has also been shown that CO polymerizes into linear chains under 5GPa pressure [Bernard *et al.*, 1998].

As mentioned in **Section 3.1**, it was only recently that researchers succeeded in developing a potential for the abovementioned two CO dimers that could account for the empirical rotational-vibrational<sup>8</sup> (rovibrational) results found experimentally. This was done by basing the potential on a DFT-SAPT PES [Vissers *et al.*, 2005]. (See **Section 2.6**, for an introduction to SAPT.) It was realized that higher order correlation effects were essential to describe the CO dimer [Rode *et al.*, 1999; Rode *et al.*, 2001]. CCSD(T) was shown to be inadequate in

---

<sup>7</sup> They are bonded in the sense of an intermolecular interaction. The distance between the carbon atoms is the smallest. In this work we will sometimes refer to intermolecular interactions as bonds.

<sup>8</sup> Rotational states are superimposed onto vibrational states. Each mode of vibration therefore has a number of rotational states.

giving a wave function with a good approximation (asymptotic behavior) to the interelectronic cusp in the exact wave function (see **Section 2.3**), even with extended basis sets. It was concluded that fifth-order excitations, such as are incorporated when performing an MP5 calculation, also play a significant role [Rode *et al.*, 1999; Rode *et al.*, 2001]. As far as we know at the time of writing (*anno* 2006), no CO clusters larger than dimers have been calculated with an *ab initio* method. However, one account was found where phase transitions in  $(\text{CO})_5$  and  $(\text{N}_2)_5$  were investigated [Acevedo *et al.*, 1997] by a Monte Carlo simulation based on an *ab initio* derived potential. This *ab initio* potential was derived assuming a T-structure for the global energy minimum C-bonded dimer and is therefore based on a model that does not conform to the most recent experimental results.

Unlike the hydrogen fluoride clusters, CO clusters are extremely weakly bound and can be utilized as a model for weak dispersion interactions, such as found in the clusters of rare gasses. In this chapter, **Section 6.2** will be devoted to discussing our *ab initio* method and its application to find starting geometries for modeling CO clusters. In **Section 6.3** we will discuss the structural properties of CO clusters and in **Section 6.4**, the interaction energies. In **Section 6.5** we will discuss the BSSE obtained when modeling these clusters. **Section 6.6** will be devoted to discussing the bonding in these clusters and in **Section 6.7** we will discuss the electrostatic properties of the clusters and the applicability of using atomic point charges to describe a charge distribution. In **Section 6.8** we discuss why a description of a covalently bonded CO dimer as a cluster, found in the literature, is misleading. The chapter will be concluded with **Section 6.9**.

## 6.2 Computational details

Our approach for the geometry optimizations is the same as discussed for the HF clusters in **Chapter 4** and only new details will be given in this section.

### 6.2.1 Pre-optimization of clusters with the 6-31G(d) basis set

The starting points for the pre-optimization of the CO dimers were based on information from the literature [Han and Kim, 1997]. However, it seems that no CO clusters<sup>9</sup> larger than the dimers have been optimized at a high level of theory before and therefore we based the starting points for the pre-optimizations on geometries obtained for  $\text{N}_2$ -clusters<sup>10</sup> [Li *et al.*,

---

<sup>9</sup> It should be mentioned that the CO trimer was found to be cyclic with a semi-empirical method [Montero *et al.*, 1981]. However, this is not a correlation consistent method.

<sup>10</sup>  $\text{N}_2$  is iso-electronic to CO.

2003]. Geometries of all the starting geometries, as well as information on their optimizations can be found in **Appendix A.3**.

In an effort to minimize the optimization times, we monitored the optimizations graphically. If geometries changed drastically from their starting points, such as when the starting geometry was planar and during the optimization one or more monomers moved out of the plane, we terminated the optimization prematurely. The aim was not to find all the energy minima on the PES, but to optimize a range of clusters of different sizes that could be used as a training set for force field optimizations in a QM/MM method. See **Chapter 7**.

In total, seven CO clusters were fully pre-optimized to local energy minima with the 6-31G(d) basis set. The internal geometries of the monomers were not constrained and symmetry constraints<sup>11</sup> were applied during the optimizations. In all cases, force constants were calculated at each optimization step to ensure that we obtained only local energy minima. Local energy minima were confirmed by the lack of imaginary frequencies in the vibrational analysis.

Eventually we optimized two CO dimers, one CO trimer, two CO tetramers, one CO pentamer and one CO hexamer at the MP2 level of theory with the 6-31G(d) basis set. These geometries can be found in **Appendix A.4**.

### *6.2.3 Final optimization of clusters*

The same optimization conditions were used as with the 6-31G(d) basis set. The CO pentamer and hexamer however could only be optimized by using a single computer node each, due to the demanding nature of the calculations. Symmetry constraints were used for all clusters during the optimizations. All the pre-optimized clusters had distinct energy minima with the larger basis set. No imaginary frequencies were found for all the clusters up to the pentamer. With the hexamer, however, we were not able to do a vibrational analysis due to inadequacies of the hardware at our disposal.<sup>12</sup> We nevertheless kept the hexamer as a suitable cluster as it was optimized under very tight optimization conditions and therefore had a limited chance of being a transition state. On the above basis, six distinct minima were found. Where possible, force constants were calculated at each optimization step to steer the optimizer in the direction of a local energy minimum.<sup>13</sup>

---

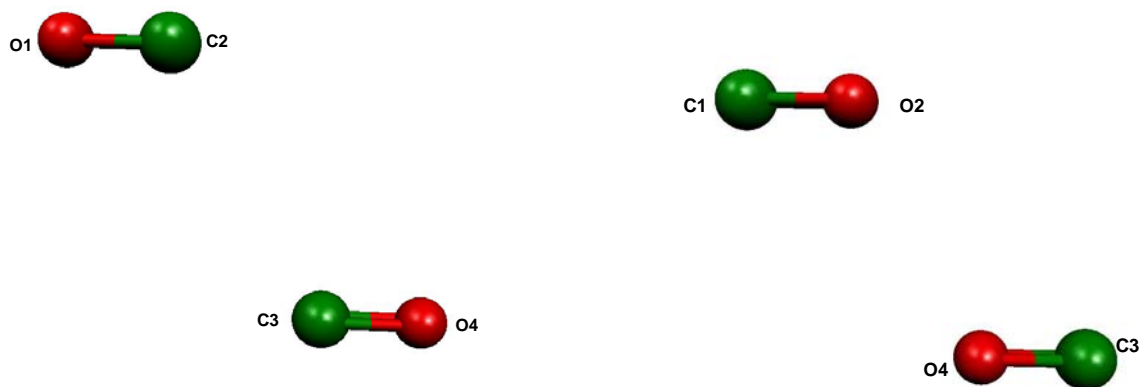
<sup>11</sup> By using symmetry constraints the optimization is speeded up as fewer degrees of freedom are allowed in the optimization. The symmetry is constrained to the symmetry of the starting geometry. If the cluster needs to move to a higher symmetry point group during the optimization, the Gaussian software will notify the user.

<sup>12</sup> Usually Gaussian does vibrational analysis directly in the memory to save time, but in this case the available memory (1GB of RAM) was inadequate.

<sup>13</sup> Force constants are related to the second derivative of the energy and give an indication of the curvature of the energy function on the PES.

### 6.3 Structural properties

The optimized structures for the 6-311+G(d,p) basis set are shown in Fig. 6.1. The hexamer is also shown, but as mentioned above, it is not necessarily a minimum on the PES. Important structural parameters<sup>14</sup> to aid in the description of the geometries are included in Table 6.1.



Cluster I: Antiparallel C-bonded dimer (I)  
Symmetry:  $C_{2h}$

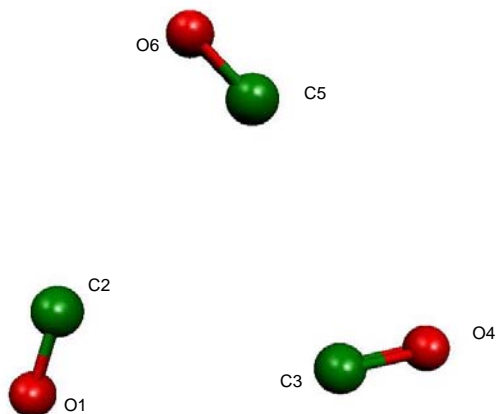
Cluster II: Antiparallel O-bonded dimer (II)  
Symmetry:  $C_{2h}$

Fig. 6.1 continued...

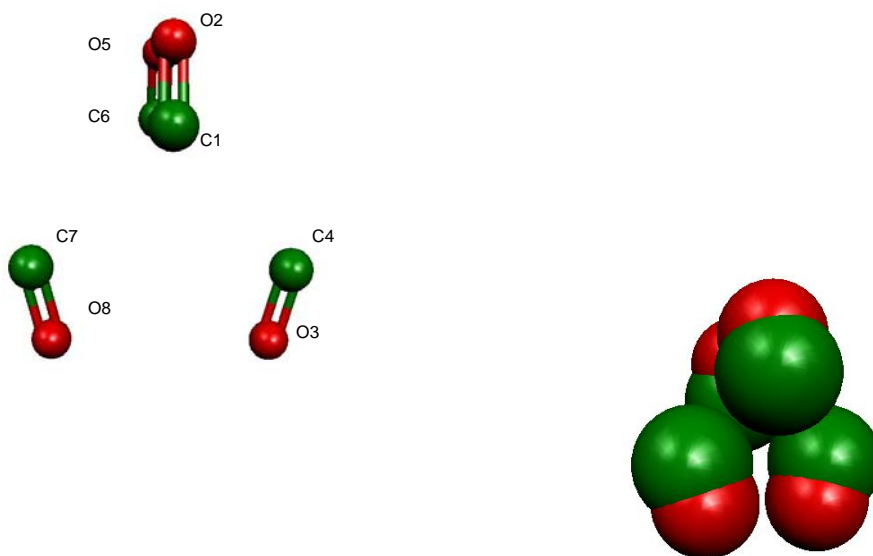
---

<sup>14</sup> All the geometric parameters are based on mass centers calculated for the monomers with a mass of 16.00 a.u. for  $^8\text{O}$  and 12.00 a.u. for  $^{12}\text{C}$ .



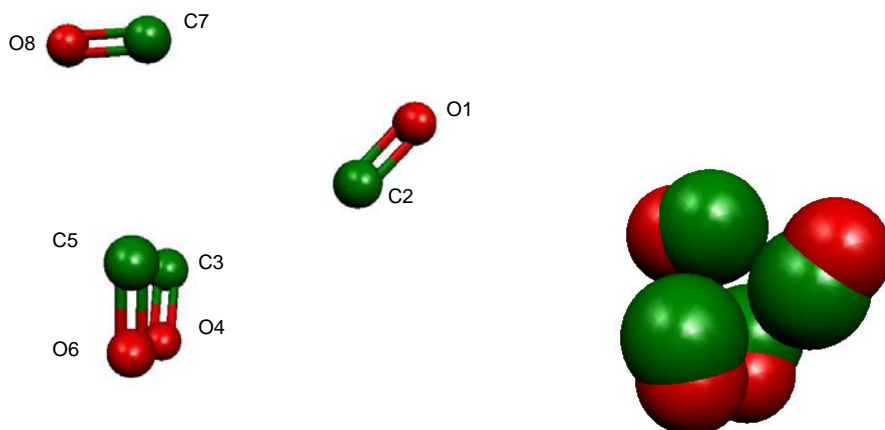


Cluster III: Cyclic trimer. Symmetry:  $C_{3v}$

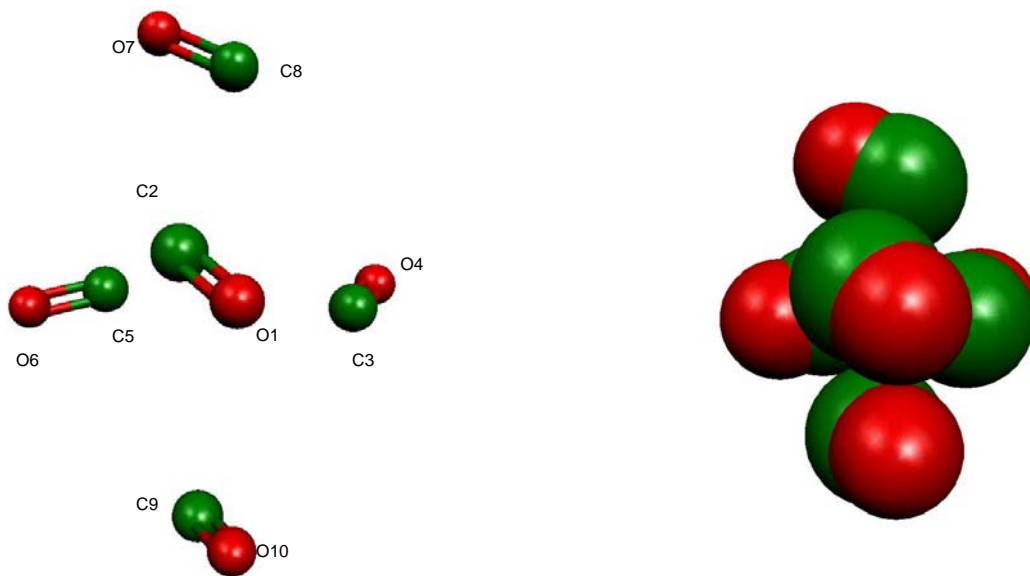


Cluster IV: Pseudo tetrahedral cluster. Symmetry:  $C_2$  (Almost  $D_{2d}$ )

Fig. 6.1 continued...

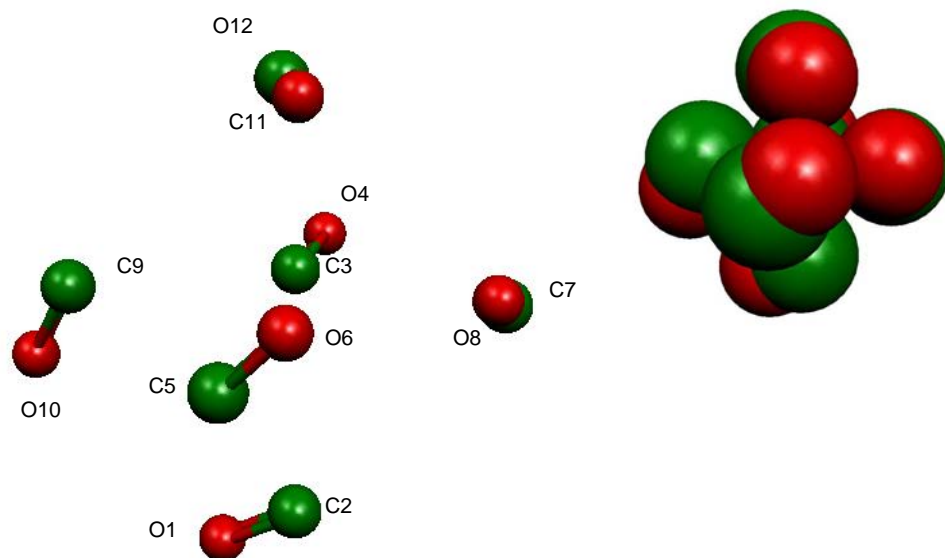


Cluster V. Symmetry:  $C_1$  (almost  $C_2$ )



Cluster VI: Pseudo trigonal bipyramidal. Symmetry:  $C_1$

Fig. 6.1 continued...



Cluster VII: Pseudo octahedral. Symmetry:  $C_1$

Fig 6.1: CO clusters optimized with MP2 and 6-311+G(d) in ball and stick models to demonstrate the packing of the monomers in the cluster and the numbering scheme. The space-filling models are in the same orientation as the ball and stick models. They are just used for a general representation to show the closeness of the packing. The symmetry is also given.

Table 6.1: Distances between the monomers. "Mon" denotes monomer.

Cluster	Distance between monomers (Å)
I	4.37
II	3.64
III	All: 4.44
IV	Mon(O5-C6)-Mon(O2-C1) : 3.59
	Mon(C7-O8)-Mon(C4-O3): 3.59
	Mon(O2-C1)-Mon(C4-O3): 4.43
V	Mon(O8-C7)-Mon(C2-O1): 4.38
	Mon(C5-O6)-Mon(C3-O4): 3.60
VI	Mon(C10-C9)-Mon(O8-C7): 6.54
	Mon(O1-C2)-Mon(C3-O4): 4.78
	Mon(O6-C5)-Mon(O1-C2): 4.35
	Mon(O6-C5)-Mon(C3-O4): 4.28
VII	Mon(O10-C9)-Mon(C5-O6): 4.27
	Mon(C5-O6)-Mon(C7-O8): 3.72
	Mon(C3-O4)-Mon(O8-C7): 4.31
	Mon(O1-C2)-Mon(C5-O6): 4.27
	Mon(O12-C11)-Mon(C3-O4): 4.31
	Mon(O12-C11)-Mon(O1-C2): 5.64
	Mon(O10-C9)-Mon(O8-C7): 5.64
	Mon(O4-C3)-Mon(O6-C5): 6.08

Both CO dimers and the CO trimer are in a single plane. The planarity and the symmetric nature of the CO trimer are striking as all the larger clusters show out of plane conformations. We see in Table 6.1 that the monomers are slightly further apart in the trimer than in either of the two dimers. There is probably a larger repulsive force on the monomers in the trimer than in the two dimers, but their arrangement in the trimer allows for a possible delocalization of electron density in a ring, which can lead to extra stability. The arrangement of the monomers is such that we find that in this arrangement the resultant dipole moment is zero. The quadrupole moments of each monomer are also close to minima as each pair of monomers are arranged in an almost T-structure.

Clusters IV and V have many similarities except that Cluster IV has a higher symmetry point group than Cluster V. Cluster V has a T-like arrangement for one pair of its monomers (C7-O8 and C2-O1), which, as we have already mentioned, is evidence of a minimum quadrupole-quadrupole interaction.

The pseudo trigonal bipyramidal cluster IV basically consists of two monomers on opposite sides of a trimer. Closer inspection shows that this cluster contains distorted CO trimers (T-formation) and dimers (antiparallel) as substructures.

We see that for cluster VII the monomers are arranged in a distorted octahedral structure with an elongated distance between Mon(O4-C3) and Mon(O6-C5). When the structure is closely inspected one finds that the structure can be divided into two trimers: Mon(O12-C11),

Mon(C5-O6) and Mon(C7-O8) are all part of one trimer, and the other trimer consists of the remaining monomers. The latter trimer's nonbonded distances are not all the same length as in the CO trimer (Cluster III) and the monomers are closer to each other than in Cluster III. The largest distance between two monomers in the cluster is smaller than in the pseudo trigonal bipyramidal structure.

It is evident that in the clusters larger than the CO dimer that many-body interactions will play a significant role in the stabilization of the clusters. Actually it is quite evident that as more monomers are added to the trimer to eventually form the tetramers, pentamer and the hexamer, we find a closer packing of the monomers (see Table 6.1). Although all the clusters have a unique arrangement of the monomers, we can always recover substructures in the clusters and all can be traced back to the arrangement of the CO dimer and the trimer. In the formation of the larger clusters there are probably first a formation of smaller units such as dimers and trimers and then the assembly of larger structures from these substructures.

The bond lengths for all the monomers in the clusters are almost identical to the gas phase bond length of carbon monoxide, of 1.140Å, calculated with the same method and basis set. The interactions between the monomers are therefore of a weak vdW nature.

Li and coworkers also found planar minima such as a square, pentagon and hexagon for the tetramers, pentamers and hexamers of N<sub>2</sub> clusters, whereas we found no planar minima, except for the trimer. The many-body effects in CO clusters are probably quite different to those in N<sub>2</sub> clusters leading to a different energy arrangement of the orbitals, making it energetically less feasible to form planar structures. CO has a small dipole moment (0.122D) [Muentner, J.S., 1970], whereas N<sub>2</sub> has no dipole moment. Minimizing the interactions between the dipole moments of the CO monomers, adds a further constraint to the minimum energy geometries that can be allowed for CO clusters. Furthermore, N<sub>2</sub> is symmetric with respect to a mirror plane perpendicular to the bond, whereas CO is not.

## 6.4 Electronic binding and interaction energies of clusters

In Table 6.2 we summarize the most important energetic terms for the CO clusters. All these terms were defined in both **Chapter 2** and **Chapter 4**. We did not include any binding energies in Table 6.2, as they do not differ significantly from the interaction energies. In other words, there is almost no one-body energies in the expansion of the binding energy of these clusters.

Table 6.2: Table showing the energetics of the clusters as calculated with MP2/6-311+G(d). Zero-point vibrational energy corrections were not applied.

Cluster	I	II	III	IV	V	VI	VII
Electronic energy (hartree)	-226.15773	-226.15721	-339.23658	-452.31725	-452.31687	-565.39725	-678.47850
BSSE (kcal/mol)	0.18	0.43	0.54	1.63	1.45	2.22	3.51
Interaction energy not corrected for BSSE (kcal/mol)	-0.46	-0.54	-1.30	-2.87	-2.60	-4.01	-5.96
Interaction energy corrected for BSSE (kcal/mol)	-0.28 [-0.38]*	-0.12 (-0.13) [-0.36]*	-0.77	-1.24	-1.19	-1.80	-2.44

[\*]: Interaction energy calculated with a CCSD(T) potential. [Vissers et al., 2005]

( ): Value obtained by Han and Kim [Han and Kim, 1997], converted from kJ/mol to kcal/mol by multiplying by 0.239001 kcal/kJ.

One can deduce from Table 6.2 that the optimization hyper-surface is extremely flat. This makes it more difficult to find individual local energy minima on the PES. For example, Cluster I is only 0.16 kcal/mol or 58.5 cm<sup>-1</sup> lower in interaction energy than Cluster II. Cluster IV is only 0.05 kcal/mol or 17.5 cm<sup>-1</sup> lower in interaction energy than Cluster V, according to our calculations.

In addition, we see that the interaction energy of Cluster I is reasonably close to the interaction energy determined by a CCSD(T) potential [Vissers *et al.*, 2005], however the interaction energy for Cluster II differs substantially from the interaction energy determined with the CCSD(T) potential [Vissers *et al.*, 2005]. We also see that, when correcting for

BSSE, the order of the energies of Cluster I and II changes. This is the same type of behavior found when comparing a T-shaped O-bonded dimer with a T-shaped C-bonded dimer [Han and Kim, 1997]. In Fig. 6.2 we show the difference in BSSE corrected interaction energies and the interaction energies without BSSE correction.

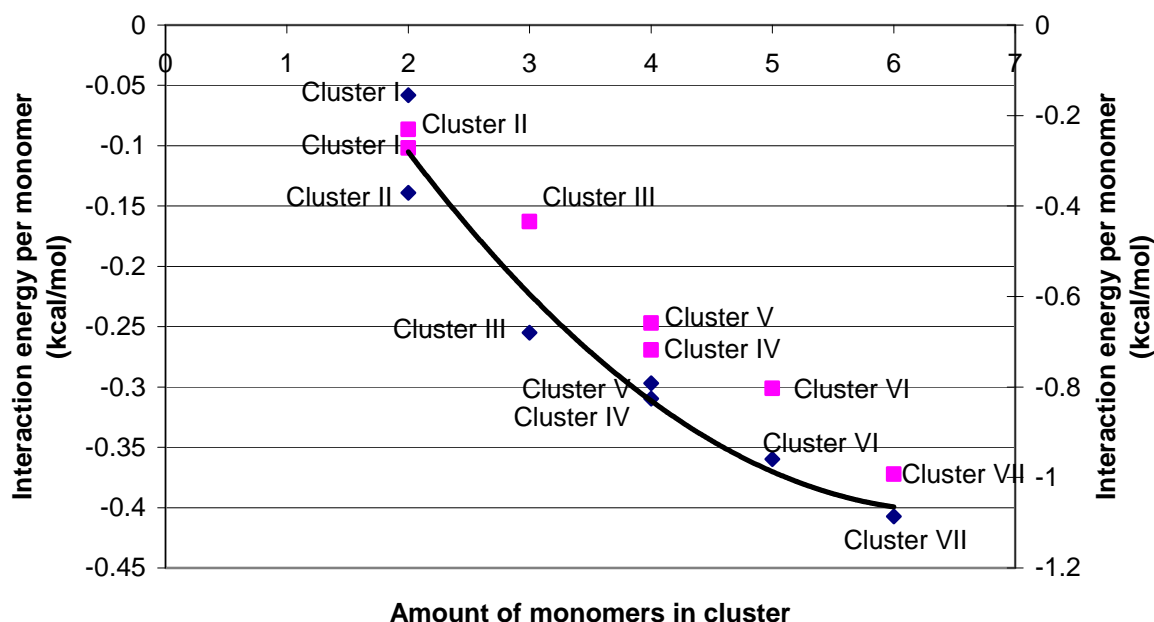


Fig. 6.2: Graph of the BSSE-corrected interaction energy per monomer and the interaction energy without a correction for BSSE per monomer. Diamonds: BSSE corrected interaction energy (shown on primary axis). Squares: Interaction energy without BSSE correction (shown on secondary axis). The trend line is drawn based on the data for the BSSE corrected interaction energy.

From Fig. 6.2 it is clear that including BSSE, while determining the interaction energies of the clusters, can be disastrous. As discussed in **Chapter 2**, we can only approximate the BSSE, as the counterpoise method only gives an upper limit for the BSSE. We observe that the interaction energy, and therefore the many-body interactions in the clusters, decline with an increase in cluster size.

It is worth mentioning that Han and coworker [Han and Kim, 1997] used the same basis set, 6-311+G(d), to calculate structures for the CO dimers. However, they did not identify the C-bonded dimer found in this study. It is interesting to note that the C-bonded dimer found in this study is 0.04 kcal/mol lower in energy than the lowest energy dimer they calculated. When Han and coworker published their paper, it was still uncertain if the C-bonded global energy dimer was T-shaped. This was only later proved later not to be the case by McKellar and coworkers [McKellar *et al.*, 2003]. It is thus extremely likely that the C-bonded dimer (Cluster I) that we found with our method and basis set is indeed the global energy

minimum.<sup>15</sup> Therefore MP2/6-311+G(d,p) seems to be adequate in obtaining structures corresponding to the experimentally determined geometries. Methods at a higher level of theory will of course give much more accurate energies.

We see in Fig 6.1 that the CO trimer is remarkably symmetric and from Fig 6.2 we see that the change in interaction energy from the dimer (Cluster I) to the trimer (Cluster III) per monomer is much larger than from the trimer to tetramer. This might be evidence of extra stabilization caused by adopting a planar conformation.

### 6.5 The importance of correcting for BSSE

In Table 6.2 we see that BSSE correction is important to obtain accurate interaction energies for the clusters. We therefore analyzed how the BSSE varies with cluster size. The results are shown in Fig. 6.3.

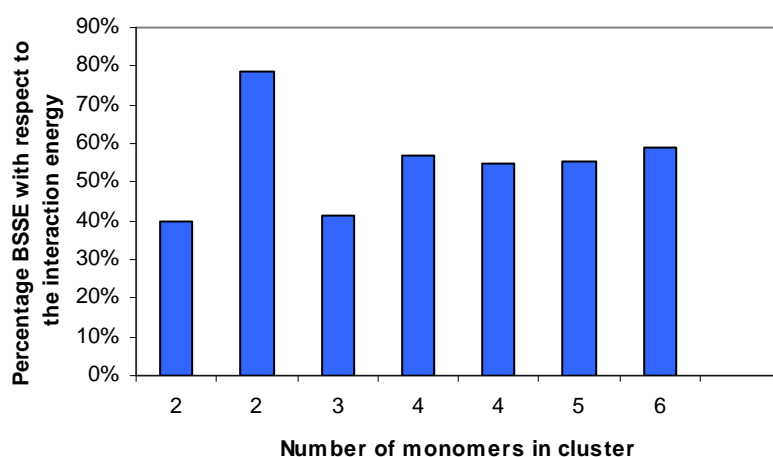


Fig. 6.3: Percentage BSSE with respect to the interaction energy for the CO clusters

It is obvious the percentage BSSE values, according to the counterpoise method, are alarmingly high for the O-bonded dimer, second from the left.<sup>16</sup> In Fig. 6.3 it seems as if the percentage BSSE, in relation to the interaction energy, is larger for the clusters larger than the trimer, but it seems as if the percentage error stays relatively constant for these clusters.

We did not expect to obtain a small BSSE, as our basis set was far too small to describe all of the higher correlation effects that are necessary to model CO clusters accurately (Rode *et al.*, 1999), but we were limited by our hardware and time.

<sup>15</sup> Han and Kim tested various other conformations of the dimers and also found other local minima for the CO dimer on the PES with the 6-311+G(d) basis set. We can therefore say, with reasonable certainty, that the C-bonded dimer that we found is the global energy minimum.

<sup>16</sup> We have already commented on this in **Chapter 3** and it will not be discussed again.



## 6.6 Analysis of bonding

### 6.6.1 Intermolecular orbitals

In **Chapter 4** we went to great lengths to show how the MP2 wave function can be used to explain intermolecular interactions. We will now show that our model regarding intermolecular orbitals is just as valid for CO clusters as for HF clusters. In Table 6.3 the delocalized orbitals for the C-bonded and O-bonded dimers are listed. The delocalized orbitals for an isolated monomer are also shown. In Table 6.3 the type of the intermolecular orbital is based on the signs of the wave function at the monomers. In **Chapter 4** we defined an *intermolecular antibonding orbital* as an intermolecular orbital where the wave function in the intermonomer space has a positive sign on one monomer and a negative sign on the other monomer.

It is evident from Table 6.3 that there is one antibonding intermolecular orbital for each bonding intermolecular orbital.

Table 6.3: Molecular orbital analysis for the C-bonded dimer, O-bonded dimer and an isolated monomer. The bonding and antibonding character is with respect to the two monomers for the dimers. For the monomer the characters of the orbitals denote the characters of the molecular orbitals.

C-bonded	O-bonded	Monomer
-0.552 (Antibonding)	-0.553 (Antibonding)	-0.557 (Bonding)
<b>-0.566 (Bonding)</b>	<b>-0.555 (Bonding)</b>	-0.636 (Bonding)
<b>-0.636 (Antibonding)</b>	<b>-0.628 (Antibonding)</b>	-0.636 (Bonding)
-0.637 (Antibonding)	-0.633 (Antibonding)	-0.805 (Bonding)
-0.638 (Bonding)	-0.635 (Bonding)	-1.519 (Bonding)
-0.638 (Bonding)	-0.638 (Bonding)	-11.371 (Nonbonding on C)
-0.806 (Antibonding)	-0.801 (Antibonding)	-20.670 (Nonbonding on O)
-0.807 (Bonding)	-0.805 (Bonding)	
-1.520 (Antibonding)	-1.516 (Antibonding)	
-1.520 (Bonding)	-1.517 (Bonding)	
-11.372 (Nonbonding on C)	-11.368 (Nonbonding on C)	
-11.372 (Nonbonding on C)	-11.368 (Nonbonding on C)	
-20.671 (Nonbonding on O)	-20.667 (Nonbonding on O)	
-20.671 (Nonbonding on O)	-20.667 (Nonbonding on O)	

When the energy values of the orbitals in the monomer are compared to those of the orbitals in the clusters, we see that the C-bonded dimer's energy values are similar to the isolated monomer's, whereas the O-bonded dimer's energy values are quite different. We see that the intermolecular antibonding orbitals of the C-bonded dimer are lower in energy than the O-bonded dimer's. The reason might be due to the Pauli-exchange repulsion in the space between the monomers. There are more electrons on the oxygen than on the carbon and therefore the electrons on the oxygen might show larger exchange repulsion than the electrons on the carbon in the intermonomer space. When the energy values for the C-bonded and O-bonded dimer are compared for each molecular orbital in Table 6.3 we see that the largest difference between the values for the C-bonded and O-bonded dimers is found for the orbitals that are shown in bold. In terms of the intermonomer space, a bonding intermolecular orbital is the result of dispersion and electrostatic interactions, whereas an antibonding intermolecular orbital can be interpreted as the result of the repulsion of the electrons with the same spin in the intermonomer space. Using this interpretation and considering the values in bold, we can say that the main reason for the larger stabilization of the C-bonded dimer compared to the stabilization in the O-bonded dimer is due to the larger attractive intermolecular forces in the C-bonded dimer and a smaller number of repulsive electron exchange in the intermonomer space. However, the difference in stabilization between the orbitals is very subtle. In Table 6.4 and 6.5 we show the *relative electron population* analysis for the two CO dimers. Again we made use of Mulliken population analysis.

Table 6.4: *Relative electron population* analysis for the oxygen in both CO dimers.

Oxygen	C-bonded dimer (a.u.)	O-bonded dimer (a.u.)
1sp	0	0
2sp	0	0
3sp	-0.0007	-0.0007
3d	0	0
4sp	0.0043	0.0041
5sp	-0.0092	0.1385

Table 6.5: *Relative electron population* analysis for the carbon in both CO dimers

Carbon	C-bonded dimer (a.u.)	O-bonded dimer (a.u.)
1sp	0	0
2sp	0.0004	0
3sp	-0.0002	-0.0028
3d	0.0002	0.0001
4sp	0.0000	0.0156
5sp	0.0056	0.0060

In tables 6.4 and 6.5 we first see a large number of electron density donated to the *5sp*-shell of the oxygen atom in the O-bonded dimer. The same can be said for the *4sp*-shell of the carbon atom in the O-bonded dimer. These results are due to orthogonal overcounting that is probably caused by the BSSE. However, we see with the C-bonded dimer that the situation is a little better and if we do the mathematics, we see that it is likely that electrons are both excited from the *5sp*-shell to the *4sp*-shell on oxygen and donated to the *5sp*-shell on carbon. Based on this assumption, the *4sp*-shell should be lower in energy than the *5sp*-shell in the cluster than in the monomer; otherwise the excitation of electrons cannot happen. Why then is there a flow of electrons from oxygen to carbon in the C-bonded dimer? Surely, this would make both carbon atoms more negative and then they should repel each other. How can this lead to extra stabilization? The only answer that we have at this stage is that intermolecular orbitals form between the two monomers and both carbon atoms donate an equal number of electron density into the intermolecular orbitals. The O-bonded dimer also has this behavior, except that it is exaggerated in the relative electron population analysis due to orthogonal overcounting. Obviously this will also influence the charges on the oxygen and carbon derived using the Mulliken population analysis, and hence any charge derivation scheme dependent on the basis set, *vide infra*. In Fig. 6.4 and 6.5 we show plots of the intermolecular bonding orbitals found for the CO dimers. VdW interactions are usually not associated with

orbital overlap<sup>17</sup>, but this is what we found. The molecular orbitals were visualized with a wave function cutoff of 0.01 a.u. as smaller cutoffs resulted in incomplete orbital visualizations. Usually, when one visualizes molecular orbitals the cutoff of the wave function is done at 0.001 a.u. as this includes more than 99% of the electron density found on the van der Waals surface around the molecules.

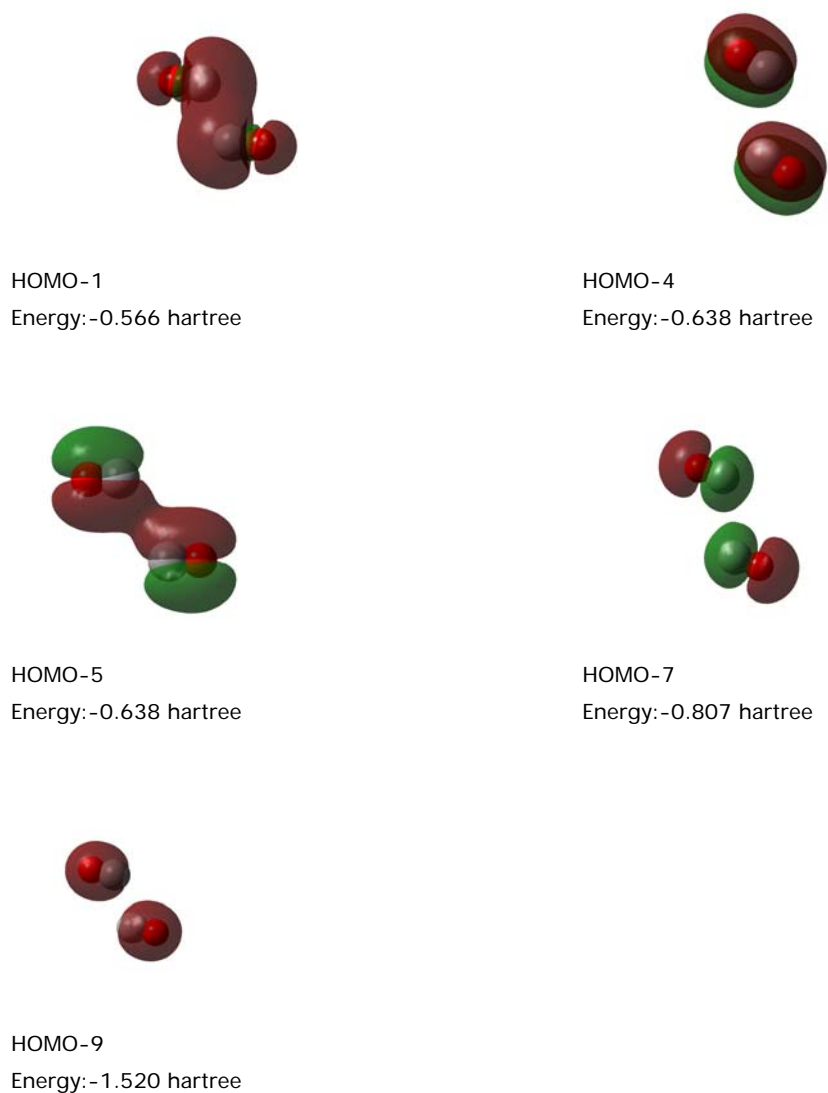
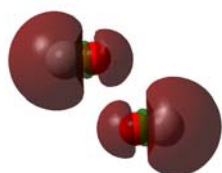


Fig. 6.4: Intermolecular bonding orbitals in the global energy minimum CO dimer (I). The energies of the orbitals are given in hartree. The wave function was cut off at 0.01a.u.

---

<sup>17</sup> Theories on vdW interactions might have been deduced from single determinant wave functions and this is why vdW interactions are usually seen as closed shell interactions.



HOMO-1  
Energy: -0.555 hartree



HOMO-4  
Energy: -0.635 hartree



HOMO-5  
Energy: -0.638 hartree



HOMO-7  
Energy: -0.805 hartree



HOMO-9  
Energy: -1.517 hartree

Fig. 6.5: Intermolecular bonding orbitals in the local energy minimum CO dimer (II). The energies are given in hartree. The wave function was cut off at 0.01a.u.

In the case of the CO clusters, it seems that the wave function occupies quite a large volume around the molecules. As found for the HF clusters, we can take this as a sign of higher energy spin orbitals mixing with lower lying spin orbitals. In this case, the orbitals are even bigger and more diffuse than in HF clusters, showing that the empty atomic orbitals in carbon monoxide monomers, usually considered high in energy by Hartree-Fock theory, are more able to mix with lower-lying orbitals than equivalent atomic orbitals in hydrogen fluoride monomers. This is probably the reason why only MP5 with a large basis set can give accurate results that are comparable to experiment [Rode *et al.*, 1999; Rode *et al.*, 2001]. Mixing these large diffuse orbitals with lower lying orbitals probably leads to a larger delocalization of the wave function around the nuclei. The mixing of the orbitals will depend mainly on electron correlation. The electrons should be arranged to be further apart in the CO monomer than predicted by Hartree-Fock theory. When two or more monomers approach each other, the large diffuse orbitals overlap to form an intermolecular orbital. Intermolecular orbitals can only be “created” when electrons have sufficient energy to move into them. This energy can

be supplied by electron correlation and/or polarization by an electric field. The temporary movement of electron density into intermolecular orbitals is probably the same as the dispersion interaction. The dispersion interaction has been described as a result of intermolecular correlation [Magnasco, 2004]. When sufficient electron density moves into the intermonomer space, the repulsive exchange interaction in the intermonomer space results in the division of the electron density into bonding and antibonding intermolecular orbitals. Electrostatic interactions in CO clusters are very weak and so only some electrons in the valence orbitals are excited to delocalized molecular orbitals and hence only a small number of electrons are expected to be found between the monomers. This is supported by AIM results, *vide infra*.

### 6.6.2 CO clusters other than the dimers

We visualized the molecular orbitals of the other clusters and found a relation between the number of intermolecular orbitals and the energy of a cluster. The larger the CO cluster, the more combinations of intermolecular orbitals that can form and hence the greater stabilization of the cluster as a whole. This is probably related to the increase in many-body interactions with cluster size. We plotted all orbitals in an effort to understand the bonding of the clusters; however, the complexity meant that an interpretation was beyond the scope of this work. A more detailed investigation as part of future work is warranted, along with a comparison to SAPT expansions of the interaction energy.

### 6.6.3 Atoms in Molecules (AIM)

An *Atoms in Molecules* (AIM) analysis was only done for the monomer and the two CO dimers, the trimer and the tetramer. Using the standard methodology of AIM, we define a *nonbonded* bond critical point as a critical point not centered on a monomer. The results are shown in Table 6.6.

Table 6.6: AIM results for the CO monomer and 4 selected CO clusters

BCP=bond critical point, RCP=ring critical point, CCP=cage critical point

	Monomer	I	II	III	IV
$\rho$ of BCP (e.bohr <sup>-3</sup> )	0.463	0.463	0.463	0.463	0.463
$\rho$ of nonbonded BCP (e.bohr <sup>-3</sup> )		0.003	0.004	0.003	All: 0.003
$\rho$ of RCP (e.bohr <sup>-3</sup> )				0.001	0.001
$\rho$ of CCP (e.bohr <sup>-3</sup> )					0.0008

A small amount of electron density is found between the monomers at a nonbonded BCP. Interestingly, there is a slightly larger electron density at the nonbonded BCP between the oxygen atoms in the O-bonded dimer, than between the carbon atoms in the C-bonded dimer. However, this should not be taken as a means of classifying the O-bonded dimer as lower in energy. The electron density at a nonbonded bond critical point is not directly related to the energy of the interaction between the monomers. We also need to consider the energies of the intermolecular orbitals in Table 6.2. We showed that the orbitals for the C-bonded dimer are all lower in energy than the O-bonded dimer's. The electron density as found by AIM analysis is a 3-dimensional function. It therefore omits the sign of the wave function, and thus is unable to separate the electron density at the nonbonded bond critical point into those electrons belonging to a bonding intermolecular orbital and those specifically belonging to antibonding intermolecular orbitals or, in other words, to the monomers. Therefore, the value for the electron density at the nonbonded critical point is made up of electron densities that stabilize both the intermolecular interaction and the valence bonds. The antibonding intermolecular orbitals do not contribute to the intermolecular interaction and therefore cannot be used for interpreting its strength.

It is interesting to note that the intermolecular interactions in Cluster IV are such that a cage critical point is found for this cluster.

Comparing our explanation for hydrogen bonding in HF clusters (**Section 4.5**) with our explanation of weak van der Waals interactions in CO clusters, we see that there are only two significant differences:

1. In HF clusters the wave function is not as diffuse as for CO clusters.
2. The electric field leading to polarization and excitation of valence electrons into intermolecular orbitals is much larger for HF clusters than for CO clusters.

Our study shows that returning to the basics of quantum mechanics can create a foundation in which nonbonded interactions can be explained based on quantum mechanical principles, rather than classical theories or theories based on uncorrelated Hartree-Fock wave functions.

There is only a small amount of literature on the nature of the vdW bond. One article [Sato *et al.*, 2005], stated that the vdW bond can be seen as the consequence of both electron correlation and long-range exchange interactions. It seems as if long-range exchange interactions are more important in rare-gas clusters than previously thought, as DFT functionals, which are unable to correctly model rare-gas dimers, have been shown to lack the ability to incorporate long-range exchange interactions.

## ***6.7 Investigating the electrostatics of the CO clusters***

In order to obtain data to be used for the MM description in a hybrid system, point charges need to be derived. In this section we will discuss the derivation of atomic point charges by the MKS scheme and as reference we will also report charges derived by other charge derivation schemes for selected clusters.



Table 6.7: ESP-maps of the ESP-potential derived from the MP2 density mapped onto the MP2 total density with cutoff values of 0.01au, and the charges calculated for the CO clusters using different electronic densities. Red shows a negative potential and blue a positive one. The sizes of the atoms in the molecules are scaled to their van der Waals radii. All charges are shown in a.u.

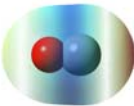
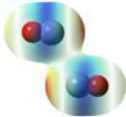
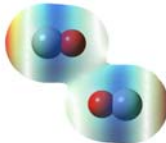
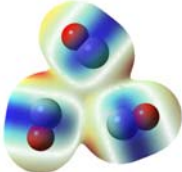
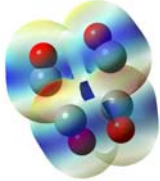
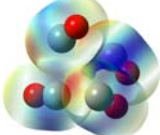
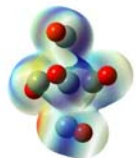
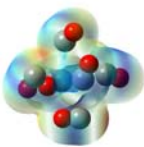
Cluster	ESP mapped onto the MP2 density	MKS charges based on the SCF density (a.u.)	MKS charges based on the MP2 density [AIM charges] (a.u.)	NAO charges (a.u.)
Monomer		C: 0.081 O: -0.081	C: -0.039 [1.105] O: 0.039 [-1.106]	C: 0.421 O: -0.421
I		O1: -0.119 C2: 0.119 C3: 0.117 O4: -0.117	O1: 0.002 [-1.104] C2: -0.003 [1.104] C3: -0.004 [-1.104] O4: 0.004 [1.104]	O1: -0.419 C2: 0.419 C3: 0.419 O4: -0.419
II		C1: 0.033 O2: -0.034 C3: 0.032 O4: -0.032	C1: -0.083 [1.106] O2: 0.083 [-1.106] C3: -0.084 [1.106] O4: 0.085 [-1.106]	C1: 0.423 O2: -0.423 C3: 0.423 O4: -0.423
III		O1: -0.134 C2: 0.135 C3: 0.128 O4: -0.129 C5: 0.132 O6: -0.132	O1: -0.011 C2: 0.011 C3: 0.013 O4: -0.014 C5: 0.016 O6: -0.015	O1: -0.418 C2: 0.418 C3: 0.418 O4: -0.418 C5: 0.418 O6: -0.418

Table 6.7 continued...

IV		C1: 0.130 O2:-0.130 O3:-0.129 C4: 0.129 O5:-0.130 C6: 0.130 C7: 0.131 O8:-0.131	C1: 0.014 O2:-0.014 O3:-0.012 C4: 0.012 O5:-0.014 C6: 0.014 C7: 0.014 O8:-0.015	Not calculated
V		O1:-0.154 C2: 0.169 C3: 0.142 O4:-0.133 C5: 0.142 O6:-0.133 C7: 0.052 O8:-0.086	O1:-0.037 C2: 0.052 C3: 0.025 O4:-0.017 C5: 0.025 O6:-0.017 C7:-0.057 O8: 0.026	Not calculated
VI		O1:-0.157 C2: 0.169 C3: 0.124 O4:-0.142 C5: 0.106 O6:-0.132 C7: 0.113 O8:-0.106 C9: 0.143 O10:-0.119	O1:-0.045 C2: 0.059 C3: 0.014 O4:-0.028 C5:-0.001 O6:-0.020 C7:-0.007 O8: 0.009 C9: 0.021 O10:-0.002	Not calculated
VII		O1:-0.130 C2: 0.119 C3: 0.207 O4: -0.177 C5: 0.158 O6: -0.168 C7: 0.078 O8: -0.078 C9: 0.119 O10: 0.130 C11: 0.078 O12: -0.077	O1:-0.018 C2: 0.005 C3: 0.100 O4:-0.063 C5: 0.054 O6:-0.065 C7:-0.039 O8: 0.039 C9: 0.005 O10:-0.018 C11:-0.040 O12: 0.040	Not calculated

As is immediately evident from Table 6.7, there are significant differences between the MKS charges based on the SCF and the MP2 electron densities, both in magnitude and in sign. It is well known that the dipole moment of CO changes sign [Jensen, 2001] when an MP2 correction is included. Interestingly, the charges for the two dimers (Cluster I and II) based on the SCF density have the same signs as the AIM and NAO charges. Immediately one is confronted with the question as to which charge scheme is the most correct. Analyzing the ESP-maps of the CO dimers again confirms that the MKS charges have the correct signs. Analyzing the RMSD of the charge fitting to the ESP (not shown), we found a worse result than for the HF clusters. This can be interpreted as the inability to approximate the ESP of CO clusters by atomic point charges. We also see that the AIM charges for the C-bonded and O-

bonded dimers do not differ by much, which is also the case for the NAO charges as well, whereas there is a large difference in the magnitude of the MKS charges for the CO dimers as based on the electron density. The O-bonded dimer's monomers are closer to each other than in the C-bonded dimer, which should result in stronger polarization of the electron densities, which in turn will influence the ESP. However, it nevertheless appears that the absolute values of the charges for the O-bonded dimer are too large. Again, the AIM and NAO charges fail to give a correct representation of the charge distributions due to their localized nature.

In this particular case, one should be careful to use atomic point charges for the description of a charge distribution. One can immediately see from the ESP-maps that the electron density is extremely anisotropic and to correct for this anisotropy in an accurate description of the charge-density, one should use point charges on the  $\pi$ -bond as well, or use multipole moments on the atoms and the bond. From standard chemical principles, carbon monoxide is expected to have a positive charge on the oxygen and a negative charge on the carbon satisfying the octet rule (see Fig. 6.6).

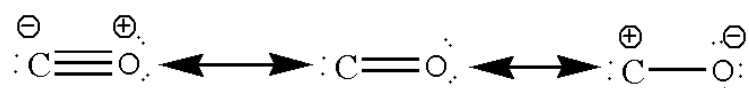


Fig. 6.6: Resonance structures for the carbon monoxide monomer

However, the AIM charges and NAO charges tell a different story, since it should also be clear now that one can never be certain about the hybridization by looking at atomic point charges derived by a charge derivation scheme. If the AIM charges are correct then the  $sp^3$ -hybridization on the right hand side in Fig. 6.6 should be favored. If the MKS charges based on the MP2 density are correct, then  $sp$ -hybridization, such as the left most structure in Fig. 6.6 should be favored.

We believe the MKS charges usually give a better representation of the charge-density than the other methods that localize electron density explicitly, at least for very simple molecules. MKS charges are usually also preferred for force fields [Jensen, 2001]. We should however emphasize again that point charges are a coarse approximation of the anisotropy of the electron density.

In the next chapter the MKS charges will be used as point charges to approximate the electron density during the micro-iterations. We would like to illustrate in more detail why atomic point charges can be expected to fail in a force field based on a simple Coulomb potential. By using point charges on atoms in a force field the lack of electron density in the center of the  $\pi$ -bonds in CO clusters is "ignored". If we consider the O-bonded dimer, for example, we see that the most stable structure according to Coulomb's law should be where the positive charge on the oxygen atom is directly opposite the negative charge of the carbon

atom, i.e., an antiparallel geometry. This however is not the case as we already confirmed. The MP2 calculations predict a *slipped* antiparallel geometry. Therefore, a simple Coulomb interaction will not be able to give a correct description of the electrostatics. In **Chapter 9** we will discuss methods to improve the description of the charge-density by point charges.

## 6.8 Oxocarbons vs CO clusters

In the literature there are references to compounds called oxocarbons, which are molecules containing only oxygen and carbon atoms. In a recent article [Sabzyan and Noorbala, 2003] calculations were done on compounds such as those shown in Fig. 6.7.

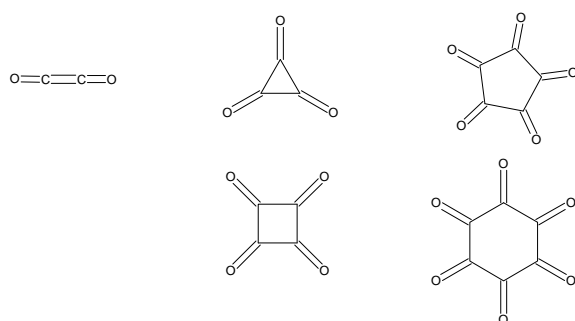


Fig. 6.7: Oxocarbons optimized by Sabzyan and coworker. with B3LYP/6-31G(d)

Sabzyan and coworker commented that these structures, which were optimized by using B3LYP/6-31G(d) calculations, cannot form spontaneously from CO monomers. They found interesting properties for the C-C bond of the dimer, which made them believe that the dimer can be regarded as a CO cluster. We argue that if the dimer is indeed a CO cluster, B3LYP would first of all not be a suitable method to do the calculations. B3LYP has been shown to give erroneous results for CO dimers [Han and Kim, 1997] such as positive binding energy values.

If these oxocarbons exist, and we know from experiment that hexaoxocyclohexane is stable [Kanakarajan and Czarnik, 1986], covalent orbital overlap will dominate and dispersion interactions are expected to not contribute significantly to the interaction energy. In such a case, DFT would indeed be an ideal method.

As far as we are concerned, CO clusters are not covalently bonded. We modeled all our CO clusters with the monomers at a van der Waals distance from each other. The molecules in the article by Sabzyan and coworker were modeled with the atoms at bonding distance from another.

Jiao and coworkers [Jiao *et al.*, 2001] report the optimization of tetraoxocyclobutane for the singlet and triplet states with the method and basis set used in our study. The bond lengths

they report show that this is a molecule rather than a cluster. It is therefore essential that one distinguish clearly between oxocarbons and CO clusters.

There are also accounts of atomic carbon-oxygen clusters [Evangelisti, 1997] in the literature. These are atomic clusters and should not be confused with molecular clusters such as those studied in our work. It is interesting to note that these carbon-oxygen clusters are investigated as rocket fuels, as they are extremely unstable and will release large amounts of energy.

## ***6.9 Conclusions and future work***

We have described the calculation of a range of CO clusters. The clusters larger than dimers, as far as we know, have never been studied before with a correlation consistent method such as MP2. We showed that intermolecular orbitals, based on the MP2 wave function, might play a role in describing the bonding of CO clusters. We showed that BSSE must be corrected for to obtain the C-bonded dimer as a global energy minimum. The use of atomic point charges cannot give a correct representation of the electrostatic potential of CO clusters and therefore will be inaccurate when used in a force field. However, for some of the clusters the point charges are very small and are expected not to make such a large difference to the interaction energy. Other ways of representing charge distributions should however be researched.

Future work should also include a more thorough investigation of the CO clusters and a comparison to SAPT results. A better description of the charge distribution in CO clusters would also be worthwhile.

In this work, we did not calculate the rotational energies of the dimers. A multitude of experimental information is available on the rotational energy levels [Surin *et al.*, 2003]. Future work should entail calculating these rotational energy levels for comparison. For accurate work a complete or near complete basis set should be used.

## Chapter 7

# Optimization of the van der Waals parameters of the Universal Force Field (UFF) for use in ONIOM-EE optimizations of carbon monoxide clusters

### *7.1 Introduction*

In this chapter, we will report the optimization of the van der Waals parameters for C<sub>1</sub> and O<sub>1</sub> in the UFF force field based on the CO clusters optimized in the previous chapter. From what we know at this stage, the force field optimization will not be affected much by the point charges, except where the point charges are unnaturally large. We expect that dispersion forces are the major attractive forces in these clusters, so the careful parameterization of van der Waals parameters will be important. We expect, however, that the use of atomic point charges, instead of various multipole moments, will make it almost impossible to obtain the MP2 geometries exactly.

In **Section 7.2** we will discuss the computational details and in **Section 7.3** some ways to assess the quality of the force field. In **Section 7.4** we will discuss the actual optimization of the force field parameters based on geometry optimizations, while we will look in more detail in **Section 7.5** at using atom-atom contacts in assessing the quality of the force field for CO clusters. In **Section 7.6** we will discuss an alternative way of optimizing the force field. **Section 7.7** will conclude this chapter.

### *7.2 Computational details*

All the ONIOM-EE optimizations were performed with the *Gaussian 03* set of programs. Computational details, including which molecules are in the MM and QM parts, have been discussed in **Chapter 3 (Section 3.5)** and will not be repeated. The nonbonded interactions were modeled geometrically, as in UFF, and the Coulomb interactions were made dependent

on the inverse of the distance between two charges. No cutoffs for van der Waals or Coulomb interactions were used. All the optimizations were done under tight optimization conditions, rather than very tight optimization conditions<sup>18</sup>. Force field single point calculations were done with Cerius<sup>2</sup> for comparison to *Gaussian 03*. For the force field calculations with Cerius<sup>2</sup> the point charges on all the atoms, as obtained from the MKS scheme, were used.

To standardize the force field results obtained by *Gaussian 03* with those from other software, it is important to set both the force field type and the connectivity explicitly in the input. The Gaussian software attempts to determine the connectivity automatically and then assigns a force field type, which may result in large differences in the results when using *Gaussian 03* compared to other force field software such as Cerius<sup>2</sup>, for the same calculation. This is discussed further in **Appendix B.2**.

### ***7.3 Quality assessment of the force field***

Assessing the quality of the force field for CO clusters is more involved than for HF clusters. The reason is that for most HF clusters, one can make use of the change in atom-atom contacts from their MP2 values as a measure of the accuracy of the geometry. However, the geometries of CO clusters are more diverse in terms of their geometries, and it would be difficult to use a RMSD for the atom-atom contacts as a measure of the quality of the force field such as we did in **Chapter 5**. This “problem” will become clear later. It is therefore not ideal to optimize the force field for CO clusters based on atom-atom contacts alone, but rather include mass-center mass-center distances in the assessment. However, we did not have time for this, so the force field was optimized based on the change in atom-atom distances, and a RMSD from the MP2 nonbonded distances was again used as a measure of the quality. For a definition of RMSD see **Section 5.3**. In **Section 7.5** we will report an alternative method of quality assessment that might be more trustworthy.

### ***7.4 Optimization of the force field parameters***

#### *7.4.1 Training set*

Due to the large number of CO clusters optimized with MP2 in the previous chapter, we decided to focus on just a few for finding optimized Lennard-Jones parameters. This training set consisted of the C-bonded dimer, the trimer, the pseudo tetrahedral cluster, and the pseudo octahedral structure. These clusters were randomly chosen.

---

<sup>18</sup> This was an attempt to make geometry optimizations faster.

### 7.4.3 Optimization of the force field

To simplify the optimization of the force field, we did not attempt to change the well depths ( $\epsilon$ -values) of the atom-types in the force field; only the van der Waals equilibrium distances ( $R_0$ -values). The default UFF van der Waals parameters can be found in Table 7.1.

Table 7.1: Default van der Waals parameters for C\_1 and O\_1 as in the original UFF article [Rappé *et al.*, 1992]. Some experimental values are also given.

Atom-type	$R_0$ (Å)	E (kcal/mol)	$R_0$ (Å) Experimental	$\epsilon$ (kcal/mol) Experimental
C_1	3.851	0.105	3.898 <sup>(a)</sup>	0.095 <sup>(a)</sup>
O_1	3.500	0.060	3.405 <sup>(b)</sup>	0.096 <sup>(b)</sup>

a) [Desiraju and Parthasarathy, 1989]

b) [Kassarasava *et al.*, 1991]

The optimization was started with the C-bonded dimer, as this dimer's structure is well known from experiment [McKellar *et al.*, 2003] and we again manually, but systematically, varied the  $R_0$ -values for the two atom-types until better parameters, as judged by the RMSD's with the MP2 nonbonded distances, were found. We realized that the only way to obtain a C-bonded slipped antiparallel geometry is to make the  $R_0$ -value for the C\_1 atom-type smaller than the  $R_0$ -value of the O\_1 atom-type. This is against experimental evidence, see Table 7.1, and conventional wisdom that states that the carbon atom should have the larger van der Waals radius. The reason for this anomaly is probably related to the fact that the van der Waals interaction is overcompensating for the lack of accuracy in the electrostatic interaction, as approximated by atomic point charges. We based our decisions on the choice of van der Waals parameters for the other CO clusters in the training set, on a larger O\_1 than C\_1 van der Waals radius. 14 sets of trial parameters were first tested for only the C-bonded dimer, the trimer and Cluster IV as, due to their size, they were quicker to calculate than the larger clusters. A choice of the best parameters, so obtained, was then used to find optimized parameters for the rest of the clusters in the training set.



The optimized Lennard-Jones parameters for the training set are:

C\_1: 3.7 Å

O\_1: 4.5 Å

Our results for the training set using the optimized parameters are summarized in Table 7.2.

Table 7.2: Results obtained with our optimized parameters for the training set

Cluster	RMSD for default force field parameters (Å)	RMSD for improved training set (Å)	% improvement
Cluster I	0.857	0.289	66.3
Cluster III	0.959	0.414	56.8
Cluster IV	0.769	0.547	28.9
Cluster VII	0.738	0.595	19.4%
Result for entire training set	0.780	0.556	28.7%

The force field parameters could not be optimized for Cluster V and II and worse results, based on our quality assessment, were obtained than when utilizing the UFF default parameters. Although the minimum energy structures obtained with these parameters showed better RMSD's, their overall geometries were different from the results obtained with MP2. With the UFF default parameters the RMSD's and the geometries obtained were not in accord to the MP2 results. This is probably due to the way in which the assessment of the force field was done. Applying the new Lennard-Jones parameters did manage to improve the force field for the training set, but the parameters were based on a larger  $R_0$ -value for C\_1 than for O\_1 and this is not chemically correct. As we mentioned earlier, this anomaly is caused by the inaccurate description of charge densities by point charges. The use of the parameters is therefore probably limited and possibly inaccurate for the other clusters in the training set. Nevertheless, for completeness we did geometry optimizations with these parameters for the remaining clusters. The results can be found in Table 7.3.

Table 7.3: Optimized parameters tested for the remaining clusters. The negative % improvement means a worsening of the values.

Cluster	RMSD for default force field parameters (Å)	RMSD for improved training set (Å)	% improvement
Cluster II	0.812	Monomers move to infinity	N/A
Cluster V	0.243	0.297	-22.2%
Cluster VI	0.675	0.423	37.3%
OVERALL †	0.701	0.495	29.4%

† For the optimizations with the new parameters, cluster II was not accounted for in the overall RMSD, since the monomers moved to infinity when the parameters of the improved training set were applied.

We observe from Table 7.2 and 7.3 that there are two cases where the UFF default parameters will actually be a better choice than our optimized parameters for an ONIOM-EE optimization. This is for Clusters V and II. In all the other cases our parameters fared better than the UFF default parameters, according to the RMSD's of our clusters from the MP2 nonbonded distances calculated for the clusters optimized in **Chapter 6**.

### ***7.5 The applicability of using atom-atom contacts for quality assessment***

In the assessment of the quality of the force field we made use of atom-atom contacts. Actually, other than for the HF clusters, this is not the ideal method to assess the quality. One can understand this by looking at the quality of the geometries as judged by a RMSD of the atom-atom contacts from the MP2 values with respect to a simple visual inspection. See Fig. 7.3.



RMSD = 0.547 Å

O\_1-R<sub>0</sub> 4.5 Å

C\_1-R<sub>0</sub> 3.7 Å

RMSD=0.570 Å

O\_1-R<sub>0</sub> 4.5 Å

C\_1-R<sub>0</sub> 4.0 Å

Fig 7.3: Diagrams showing the significant change in geometry when the Lennard-Jones parameters for the CO-tetramer (IV) are changed slightly.

The figure on the right in Fig. 7.3 seems closer to the MP2 structure determined in **Chapter 6**, although it has a worse RMSD from the intermolecular distances in the MP2 geometry. Therefore, a good RMSD does not necessarily mean a good geometry for the CO clusters. To assess the quality of the force field more accurately, the distances between mass centers of the monomers should also be included in the list of nonbonded distances when the RMSD's for the clusters are determined.

## 7.6 Force field optimization based on the ONIOM steric energies

As we had limited time to optimize the force field parameters, it was important to find a new way of optimizing the parameters. There are two problems with the present method of force field optimization for CO clusters:

1. Atom-atom contacts are not a good measure for quality assessment.
2. The geometry optimization of a MM system in the micro-iterations can be laborious due to reasons discussed in **Chapter 2**.

We thus saw the need for a better and faster method of optimizing van der Waals parameters and therefore we used a similar method to that discussed in **Section 5.6**. In this section, only the results for the method, applied to the CO clusters, will be given. Again, the same GA and quasi-Newton optimization algorithm methods were used as for the HF clusters in **Chapter 5**.

For the CO clusters, unlike the HF clusters, it was actually possible to optimize the force field by fitting the ONIOM interaction energy<sup>19</sup> directly to the MP2 BSSE corrected interaction energy, as the energy for the perturbation calculated on a QM level was small enough, due to the small charges on the MM atoms. For this, we found that the  $R_0$  of C\_1 is 3.457 Å and the  $R_0$  of O\_1 is 4.340 Å, which is still chemically incorrect. The RMSD with the MP2 interaction energy obtained was 0.565 kcal/mol. We therefore see that one cannot use the exact energy for the electronic perturbation, calculated on a QM level of theory, when optimizing the FF and one must rely on the approximation of this perturbation by point charges on an MM level of theory as previously found for HF clusters in **Chapter 5**. The reason for this is not clear, but in **Chapter 9** we will discuss a possible reason. We therefore had no choice but to base the electrostatic energy, as in **Chapter 5**, only on the interaction between the point charges derived for the QM atoms and the stationary point charges on the MM atoms. When the force field is optimized with a GA and then reoptimized with a quasi-Newton algorithm, we obtain the following  $R_0$ -values:

$R_0$  of C\_1: 3.687 Å

$R_0$  of O\_1 :2.948 Å

It is immediately evident that these values are not chemically incorrect, as the C\_1 atom-type's  $R_0$ -value is larger than the O\_1 atom-type's value, as is expected. Table 7.4 summarizes the results for the fitting of the  $R_0$ -values so that the ONIOM steric energy<sup>20</sup> matches the MP2 BSSE interaction energy for all the clusters as closely as possible. Both of these energies are given in Table 7.4.

---

<sup>19</sup> The ONIOM interaction was defined in Chapter 5 as the sum of the van der Waals energy, the electrostatic energy between the atoms in the MM system and the electrostatic energy associated with the perturbation of the MM charges on the QM atoms.

<sup>20</sup> The ONIOM steric energy was defined in **Chapter 5** and is the energy obtained by adding the van der Waals energy for a frozen geometry to the electrostatic energy as determined by the point charges on the QM and MM atoms at this geometry.

Table 7.4: Summary of the results for the optimization of the force field parameters. The terms are identical to the terms used in Section 5.6.

Cluster	MP2 interaction energy (kcal/mol)	ONIOM steric energy obtained with the optimized parameters (kcal/mol)	ONIOM steric energy obtained with default UFF parameters (kcal/mol)
I	-0.28	-0.17	-0.19
II	-0.12	-0.18	-0.22
III	-0.77	-0.51	-0.63
IV	-1.24	-1.17	-1.39
V	-1.19	-1.18	-1.44
VI	-1.80	-1.71	-2.09
VII	-2.44	-2.57	-3.06
Total RMSD (kcal/mol)	N/A	0.128	0.292

At first glance, the RMSD for the optimized parameters in Table 7.4 seems much better than for the HF clusters. However, one should keep in mind that the carbon monoxide clusters have a very shallow PES whereas the HF clusters do not and therefore a small change in energy can lead to a large change in geometry for the CO clusters. It is therefore not necessarily true that when the energy fitted parameters are used for a geometry optimization in ONIOM-EE, the geometries will be comparable to the MP2 ones in terms of the arrangement of the molecules in space.

Considering the values in Table 7.4, we see that for clusters I and III the UFF default parameters give better RMSD's than the optimized parameters. However, for the five remaining clusters, the values for the optimized parameters are better. We reported the same type of situation earlier in Section 7.4.3, when our optimized parameters were only better than the UFF default  $R_0$ -values for five of the seven clusters. The problem might be with the electrostatic energy, which is not accurate enough for a specific geometry, due to the fact that point charges are used on the atoms instead of a more elaborate description of the charge-density. Without a more accurate way of treating the electrostatics, it will be almost impossible to optimize this force field to give a better fit of the ONIOM steric energies to the MP2 BSSE corrected interaction energies. Some suggestions will be given in Chapter 9 of how to improve the fit.

The trends for the ONIOM steric energies for the default and optimized parameters as well as the MP2 BSSE corrected interaction energies are shown in Fig. 7.4. Note that the trends are reasonably close to the variation of the MP2 interaction energies in order of increasing cluster size. It is also evident that for smaller clusters the UFF default parameters give a closer fit to the MP2 interaction energies than the optimized parameters, but with the larger clusters the optimized parameters give an improved fit to the MP2 interaction energies.

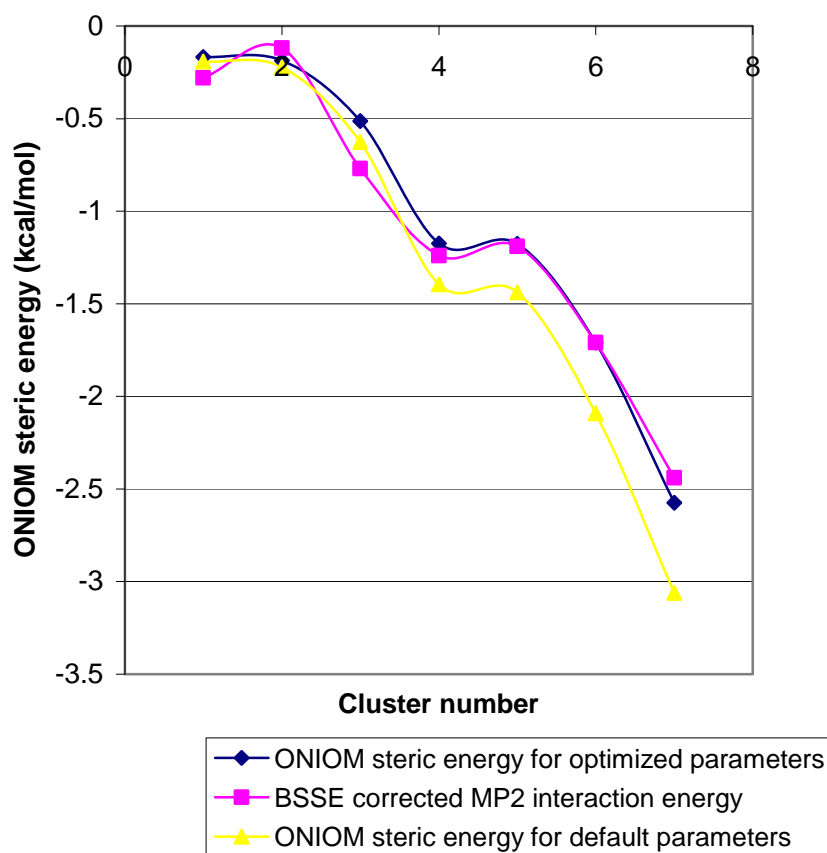


Fig. 7.4: Graph illustrating the similarities and differences between the fit of the ONIOM steric energies for the default UFF and optimized parameters. The MP2 BSSE corrected interaction energies are also shown. The clusters are numbered in the same sequence as previously given in **Chapter 6**. Cluster I is first and Cluster VII last.

## 7.7 Summary, conclusions and future work

In this chapter we aimed to optimize the  $R_0$ -values for the C\_1 and O\_1 force field atom-types in UFF in order to use these values in future QM/MM hybrid calculations and geometry optimizations. In our first attempt, we used the RMSD's of the atom-atom contact distances as obtained by geometry optimizations with the force field, with respect to the MP2 values, as a measure of the accuracy of the force field. However, we found that assessing the quality of the force field based on this is not ideal and distances between mass centers of monomers should possibly also be included in the list of nonbonded distances. Another problem was that the optimized parameters were based on the quality of the fit of the ONIOM-EE equilibrium geometry for the C-bonded dimer (Cluster I) to the MP2 equilibrium geometry.

In general, three main problems make the optimization of force field parameters with the geometric method difficult:

1. The ESP of the electrostatically perturbed QM system is not well described by point charges on atoms.
2. The PES of CO clusters is known to be extremely flat and therefore if a manual optimization is used, parameters should be varied in smaller steps than for the HF clusters. This will increase the time needed to find optimized parameters by geometry optimizations.
3. The quality assessment of the force field can be a problem.

In order to develop a more efficient force field optimization method, the fit of the ONIOM steric energy to the MP2 BSSE corrected interaction energy was used as a criterion for optimizing the force field. This approach again seems easier, but it is uncertain how well the geometries of the clusters will be improved when these parameters are used in an ONIOM-EE geometry optimization in *Gaussian 03*, as they are based on the interaction energy. However, the ONIOM steric energies gave a better fit to the MP2 interaction energies than the default UFF parameters, except for Clusters I and III. It is therefore not possible to obtain a good fit to the MP2 interaction energies by optimizing the vdW parameters, and hence the problem most likely lies with the calculation of the electrostatic energy. It is also evident that the exact energy for the perturbation, as given by *Gaussian 03*, is too large to be used directly for a frozen geometry optimization. In **Chapter 9** we will give a possible reason for this.

In future work, the  $R_0$ -values obtained for fitting the ONIOM steric energy to the MP2 interaction energy can be used to validate how well these values perform in ONIOM-EE geometry optimizations in *Gaussian 03*. Real improvement in the force field will, however,

only come with a better description of the charge-density distribution in the system. Future work may be directed towards this aim.



# Chapter 8

## CO/HF clusters: *ab initio* calculations and force field optimization for hybrid optimizations

### 8.1 Introduction

In this chapter we report the optimization and analysis of CO/HF clusters. The information that we obtain will be used to develop a force field for QM/MM hybrid optimizations of CO/HF clusters.

The only stable CO/HF-cluster known from experiment is the C $\cdots$ H bonded dimer [Fraser and Pine, 1988]. Computationally, clusters up to the trimers have been optimized [Seung-Hoon, 1998]. It is interesting to note that interstellar clouds can contain mixtures of HF and CO [Cunha and Smith, 2005]. It is therefore of importance to model CO/HF clusters not just from a theoretical point of view, in order to understand hydrogen bonding, but also from a physical and astronomical point of view. We are of the opinion that CO/HF clusters can be a useful model to describe the simultaneous effects of weak van der Waals interactions and hydrogen bonding in a cluster. In this chapter, **Section 8.2** will be devoted to discussing the optimization of these interesting clusters and **Section 8.3** to the energies of the clusters. **Section 8.4** will be used to discuss the bonding of the clusters in terms of MO theory and in **Section 8.5** we will compare the electrostatics of these clusters. In **Section 8.6** we will describe the development of a force field for a hybrid method for these clusters. We will conclude this chapter with a summary and a discussion of possible future work in **Section 8.7**.

### 8.2 Computational details

The general computational details are the same as in **Chapters 4** and **6** and will not be repeated.

### 8.2.1 Pre-optimization of clusters with the 6-31G(d) basis set

The starting points for the pre-optimization of the CO/HF clusters were partially obtained from the literature, as the minimum energy structures for the dimers have been calculated [Curtiss *et al.*, 1984] and are known to be linear. The geometries for the rest of the clusters were determined by attempting different combinations of the optimized geometries of the HF and CO clusters reported in **Chapter 4** and **Chapter 6** respectively. For example, for the tetramers we used tetrahedral and square conformations as starting points. Only the clusters that gave unambiguous energy minima were kept. The rest were discarded.

Again, just as with the HF and CO clusters, very tight optimization conditions were used. Optimizations were done by constraining the symmetry of each cluster to its lowest symmetry point group during the geometry optimization. No further constraints on the geometries of the clusters were applied. Energy minima were confirmed by a lack of imaginary frequencies in the vibrational analysis. Eventually we optimized 11 CO/HF clusters with the 6-31G(d) basis set. The pre-optimized geometries can be found in **Appendix A.5**.

### 8.2.3 Final optimization of clusters

The same optimization conditions were used as with the 6-31G(d) basis set. Symmetry constraints were used for all clusters during the geometry optimizations, except for one cluster shown as cluster V in Fig. 8.1. During the optimization, it converted from a  $C_s$  point group symmetry to a  $C_1$  symmetry. Again, energy minima of all the clusters were confirmed by a lack of imaginary frequencies in the vibrational analysis.

## 8.3 Structural properties

The optimized structures for the 6-311+G(d,p) basis set are shown in Fig. 8.1.

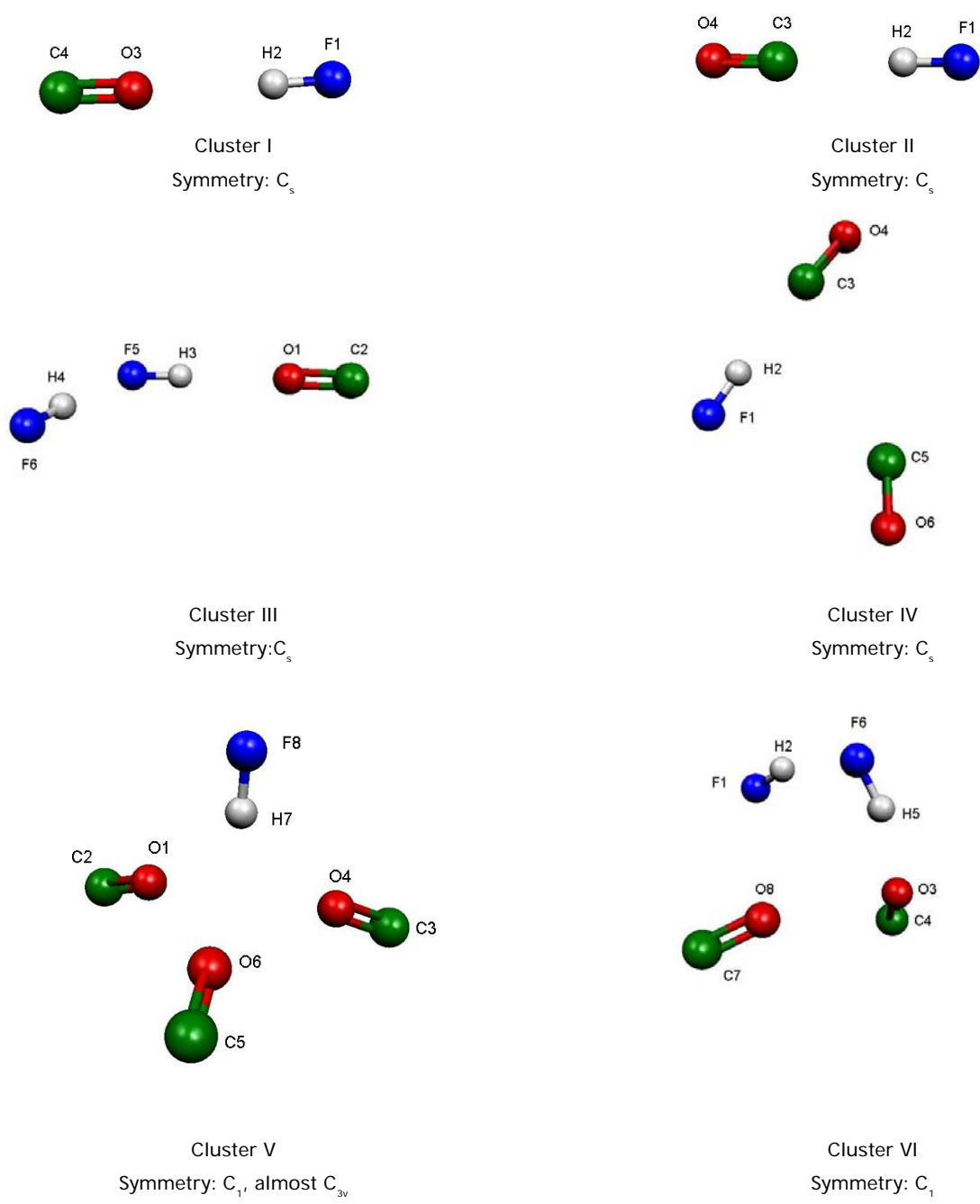
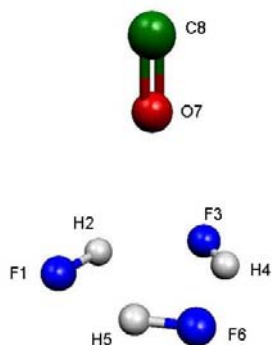
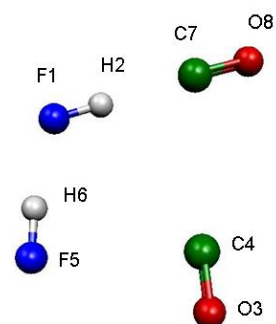


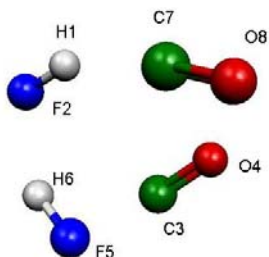
Fig. 8.1 continued ...



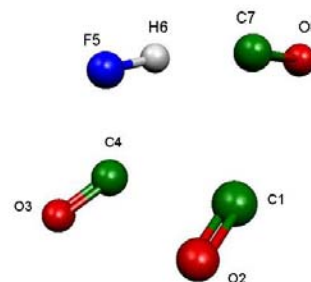
Cluster VII  
Symmetry:  $C_3$



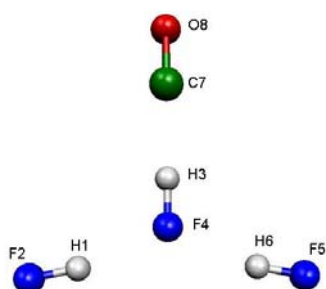
Cluster VIII:  
Symmetry:  $C_1$ , almost  $C_s$



Cluster IX  
Symmetry:  $C_1$ , almost  $C_s$



Cluster X:  
Symmetry:  $C_1$ , almost  $C_s$



Cluster XI  
Symmetry:  $C_{2v}$

Fig. 8.1: Minimum energy geometries obtained for the CO/HF clusters with 6-311+G(d,p). The symmetry point group for each cluster is also shown.

In Table 8.1 the monomers' bond lengths for each respective cluster in Fig. 8.1 are shown.

Table 8.1: Monomer bond lengths for the monomers in each cluster. Bond lengths are in Ångström.

Cluster	I	II	III	IV	V	
<u>Bond lengths</u>						
C-O	1.141	1.137	1.142	C3-O4: 1.137 O6-C5 :1.139	O1-C2:1.141 C3-O4:1.141 O6-C5:1.141	
H-F	0.918	0.922	F6-H4:0.922 F5-H3:0.921	F1-H2:0.924	F8-H7:0.918	
Cluster	VI	VII	VIII	IX	X	XI
<u>Bond lengths</u>						
C-O	O3-C4:1.141 O8-C7:1.141	O7-C8:1.140	O3-C4:1.139 C7-O8:1.136	C3-O4:1.141 C7-O8:1.136	C1-O2:1.140 O3-C4:1.140 C7-O8:1.137	O8-C7:1.135
H-F	F1-H2:0.923 F6-H5:0.922	F1-H2:0.929 F3-H4:0.929 F6-H5:0.929	F1-H2:0.931 F6-H5:0.925	H1-F2:0.930 F5-H6:0.924	F5-H6:0.925	H1-F2:0.921 H3-F4:0.938 F5-H6:0.921

We observe a large variation in the HF and CO monomer's bond lengths for the clusters. To save space, the nonbonded atom-atom distances are not included in Table 8.1, but can be found in **Appendix D**. As discussed in **Chapter 4**, the change in bond lengths for the monomers from the values for the bond lengths of the isolated HF and CO monomers is related to a one-body energy. We showed in **Chapter 4** that the valence bonds get weaker and therefore longer as electrons are delocalized into intermolecular orbitals.

### 8.2.2 Discussion of geometries

Our geometry for the C $\cdots$ H bonded cluster compares well with the experimental information available from Legon and coworkers [Legon *et al.*, 1981] who showed, by using microwave spectroscopy, that the stable C $\cdots$ H bonded dimer is a linear complex with a C-F distance of 3.0468 Å. Our value (see **Appendix D**) is 3.0484 Å.

Computationally the largest clusters that have been optimized are the trimers, of which we only optimized the O $\cdots$ H bonded one (Cluster III) [Seung-Hoon, 1998]<sup>21</sup>. Seung-Hoon optimized the trimers with SCF, MP2 and B3LYP with both the 6-31+G(d,p) and 6-311+G(d,p) basis sets [Seung-Hoon, 1998]. We confirmed that within three decimal places our results are the same as their results for MP2/6-311+G(d,p) and therefore if there are errors in our optimizations, they are insignificant.

<sup>21</sup> Our aim was not to identify all of the energy minima on the PES, but only to identify enough for the optimization and testing of our force field.

## 8.4 Analysis of electronic interaction and binding energies

In Table 8.2 we summarize the energies for the various CO/HF clusters optimized. Only the energy of the C $\cdots$ H bonded dimer has been determined experimentally. The most accurate value for the dissociation energy of  $732\pm 2$  cm $^{-1}$  [Oudejans and Miller, 2000] translates to  $2.093\pm 0.006$  kcal/mol as shown in terms of the binding energy in Table 8.2. Although this does not compare well to our value, we have not applied a zero-point vibration energy correction, which may change it slightly.

Another way of comparing our results to experimental results is by comparing the shifts in the frequencies of the stretching vibrational normal modes of the monomers in the optimized dimers from the frequencies of the stretching modes of isolated CO and HF monomers. For the C $\cdots$ H bonded dimer the HF stretch is *redshifted* by  $131.15$  cm $^{-1}$ , whereas the experimental *redshift* is  $117$  cm $^{-1}$  [Yu *et al.*, 2004] – a difference of  $0.040$  kcal/mol in energy. The CO stretch has a *blueshift* of  $24.4$  cm $^{-1}$ , which interestingly is the same as the experimentally determined *blueshift* [Yu *et al.*, 2004]. We should however warn that this is coincidental, as one would not expect MP2 to give such accurate values. *Gaussian 03* uses harmonic approximations of the vibrations while in actual fact they are not completely harmonic.

## 8.5 Analysis of bonding

As is the case with the CO and HF clusters discussed in **Chapters 4** and **6**, we found intermolecular delocalized orbitals when visualizing the MP2 wave function of the CO/HF clusters. However, we will not include the plots here, as nothing new could be determined from these in addition to what we have already found in **Chapters 4** and **6**.

Table 8.2: Energies of the CO/HF clusters as obtained at the MP2 level of theory. Zero-point vibrational energy corrections are not applied. The values are rounded off.

Cluster	Electronic energy (hartree)	Interaction energy <sup>(a)</sup> (kcal/mol)	Binding energy <sup>(a)</sup> (kcal/mol)	MP2 one-body energy <sup>(b)</sup> (kcal/mol)
I	-213.359316	-1.05	-1.04 (-1.67) <sup>(c)</sup>	0.002
II	-213.362593	-3.03	-3.00 (-2.10) <sup>(d)</sup>	0.03
III	-313.646425	-5.25	-5.21	0.04
IV	-326.442815	-3.91	-3.86	0.04
V	-439.520894	-2.11	-2.10	0.009
VI	-426.727082	-5.42	-5.36	0.05
VII	-413.938706	-11.70	-11.37	0.3
VIII	-426.732642	-9.52	-9.30	0.2
IX	-426.731661	-8.68	-8.49	0.2
X	-439.523505	-4.76	-4.71	0.06
XI	-413.936856	-11.57	-11.19	0.4

a) Both are corrected for BSSE.

b) MP2 one-body energy was defined in **Chapters 4** and **6**.

c) CCSD(T)/aug-cc-pVQZ result [Tuma *et al.*, 1999].

d) Experimental result converted from cm<sup>-1</sup> to kcal/mol [Oudejans and Miller, 2000]. See text.

In Table 8.3 the energy values of each *intermolecular orbital* in the two dimers are given together with their type, i.e., bonding or antibonding, based on the signs of the wave function in the intermonomer space.

Table 8.3: A list of the intermolecular orbitals of the CO/HF dimers and their energy eigenvalues in hartree. The type, i.e. antibonding or bonding, is given in terms of the intermolecular orbitals. The non-bonding orbitals are also nonbonding molecular orbitals, i.e., remains of the atomic core orbitals on the atoms.

Orbital	O · · · H bonded dimer	C · · · H bonded dimer
HOMO	<b>-0.579 (Antibonding)</b>	<b>-0.597 (Antibonding)</b>
HOMO-1	<b>-0.638 (Antibonding)</b>	<b>-0.635 (Antibonding)</b>
HOMO-2	<b>-0.638 (Antibonding)</b>	<b>-0.635 (Antibonding)</b>
HOMO-3	-0.664 (Bonding)	-0.665 (Bonding)
HOMO-4	-0.664 (Bonding)	-0.665 (Bonding)
HOMO-5	-0.755 (Bonding)	-0.760 (Bonding)
HOMO-6	-0.843 (Bonding)	-0.834 (Bonding)
HOMO-7	<b>-1.546 (Antibonding)</b>	<b>-1.550 (Antibonding)</b>
HOMO-8	-1.590 (Bonding)	-1.587 (Bonding)
HOMO-9	-11.399 (Nonbonding on C)	-11.403 (Nonbonding on C)
HOMO-10	-20.695 (Nonbonding on O)	-20.701 (Nonbonding on O)
HOMO-11	-26.285 (Nonbonding on F)	-26.282 (Nonbonding on F)

We see clearly that, unlike for the CO clusters, the number of antibonding and bonding intermolecular orbitals in the CO/HF dimer is not equal. We observe that there are 12 delocalized orbitals. Three of them are nonbonding remains of core atomic orbitals, leaving only four pairs of antibonding and bonding intermolecular orbitals, plus another intermolecular bonding orbital. Since an antibonding intermolecular orbital can be interpreted as a result of the Pauling electron-electron exchange in the intermonomer space, the lack of an antibonding intermolecular orbital can be interpreted as the lack of electron-electron exchange in the intermonomer space. As already mentioned, for the CO clusters (see **Chapter 6**) we found the same number of antibonding and bonding intermolecular orbitals. For the HF clusters, one of the antibonding intermolecular orbitals of the dimer was not completely antibonding, see HOMO-1 in Fig. 4.6. For the CO/HF dimers, we again observe that the number of bonding intermolecular orbitals is not equal to the number of antibonding intermolecular orbitals. Therefore, we conclude that this phenomenon is related to hydrogen bonds. One can rationalize this by realising that in a monomer the hydrogen atom has a very small electron density and most atomic orbitals, except for the valence orbitals, are empty, in which case the hydrogen atom can be used as a reservoir for electron density originating from the other monomer. By using these empty orbitals on the hydrogen atom, delocalization of



electron density of the “donor” monomer towards the “acceptor” monomer would not experience an electron-electron exchange interaction. The incorporation of the hydrogen atomic orbitals into intermolecular orbitals will depend on how energetically feasible it is and as in HOMO-1 of the HF dimer, antibonding character might still dominate in a specific intermolecular orbital. Adding the values for the energies of the occupied molecular orbitals for each cluster in Table 8.3, we obtain a larger stabilization in the C···H bonded dimer than in the O···H bonded dimer. It is important not to add the energies of the orbitals and compare this with the total electronic energy as the total MP2 energy is calculated in a specific way, see **Chapter 2 (Section 2.5.1)**.

It seems, based on the energies for the antibonding intermolecular orbitals in Table 8.3, that the reason why the C···H bonded dimer is more stable than the O···H bonded dimer is the exchange repulsion present between the electron densities of the monomers that constitute the HOMO of the cluster. The electrons responsible for this exchange originate from the fluorine, carbon and oxygen atoms. It has previously been shown [Curtiss *et al.*, 1985], that according to NBO theory, the reason for the C-bonded dimer being lower in energy than the O-bonded dimer, is due to the difference in their charge transfer energies. The stabilization obtained by transfer from the lone pair on the C-atom to the antibonding  $\sigma^*$ -orbital on the HF monomer is more than the stabilization obtained when a lone pair on the oxygen atom is transferred to the antibonding orbital on the HF monomer. This charge transfer can also be seen as the formation of an intermolecular orbital as we have seen for the HF clusters in **Chapter 4**. When the bonding intermolecular orbitals' energies in Table 8.3 are compared for the two dimers, we see, however, that the bonding intermolecular orbitals cannot be the main reason for this stability. In Table 8.3 the main difference in the orbital energies for the two dimers is found between the HOMO's of the dimers. Therefore, according to our results, the electron-electron exchange interaction between the molecular orbitals of the two monomers in the HOMO's of the clusters is the main difference between the stability of the dimers.

## 8.6 Atoms in molecules (AIM) analysis

An *Atoms in Molecules* (AIM) analysis was done on the two CO/HF dimers. In Table 8.4 we summarize properties of the dimers based on AIM theory. The other clusters show more complexity in their electron density topologies and therefore each cluster should be studied on its own. Time did not permit this, so in this section we limit ourselves to the two dimers.

The negative of the Laplacian shown in Table 8.4 is the negative of the second derivative of the electron density with respect to the coordinates, and acts as a magnifying glass to show a local increase and decrease in the electron density. If the negative of the Laplacian is

positive it means that there is an increase in local electron density at this point, but this does not mean that the increase is necessarily in all directions.

Table 8.4: AIM analysis for the CO/HF dimers. BCP=bond critical point. As we defined it in **Chapter 6**, a nonbonded BCP is a BCP that is not found on a monomer, but between two monomers.

Clusters	I	II
$\rho$ of the BCP (e.bohr <sup>-3</sup> )	F1-H2: 0.369 C3-O4: 0.460	F1-H2: 0.361 O3-C4: 0.467
Negative of Laplacian (e.bohr <sup>-5</sup> )	F1-H2: 2.837 C3-O4: -0.835	F1-H2: 2.776 O3-C4: -0.870
$\rho$ of the nonbonded BCP (e.bohr <sup>-3</sup> )	0.010	0.018
Negative of Laplacian (e.bohr <sup>-5</sup> )	-0.048	-0.061

We can see from the value of the negative of the Laplacian in Table 8.4 that the interaction in the CO monomers is a closed shell interaction suggesting an ionic bond, whereas the bond in the HF monomers is a shared, covalent interaction as expected.

In Fig. 8.2 we illustrate the gradient paths superimposed on the electron density, as determined by AIM analysis. Gradient paths are lines drawn orthogonal to the contours of constant electron density and their directions in 3-dimensions are the same as the directions of the gradients of the electron density at each point on the gradient path. With AIMPAC it is only possible to visualize the gradient paths in two dimensions. The gradient paths start from infinity and end at the nuclear attractors, where maximum electron density is found in all directions. Gradient paths can also end at bond critical points. These gradient paths are also called zero-flux lines as the flux, according to Gauss's law through the Gaussian surface, is zero [Popelier, 2000].

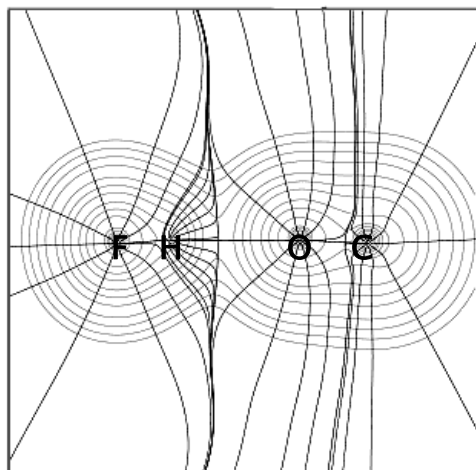


Fig. 8.2: Gradient paths superimposed on the electron density contours of the CO/HF O...H bonded linear dimer.

The zero-flux line defines the volume of the total electron density that belongs to the volume of an atom or in AIM terms, the atomic basin. In Fig. 8.3 we show the gradient paths as superimposed on the electron density of the C...H bonded dimer. Note the larger penetration of the electron density of the CO monomer into the electron density of the HF monomer leading to a more convex gradient path through the bond critical point than was found for the O...H bonded dimer. This might be the main cause for the larger stabilization in the C...H bonded dimer in comparison to the O...H bonded dimer.

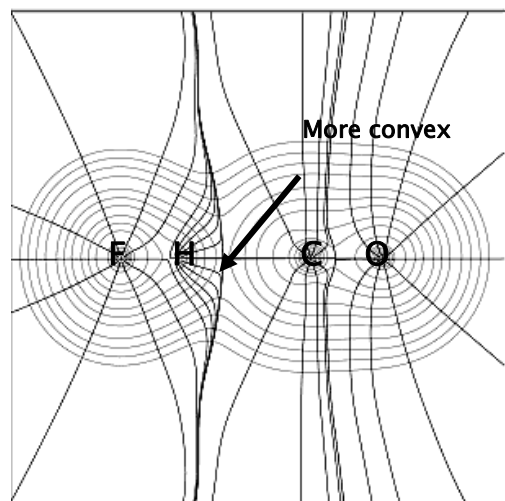


Fig. 8.3: Gradient paths superimposed on the electron density contours of the CO/HF C...H bonded linear dimer.

We therefore make the same conclusion as Galvez *et al.*, 2001 [Galvez *et al.*, 2001] for the HF dimer, that the stabilization in the CO/HF clusters is also due to the interpenetration of electron densities rather than charge transfer.

## ***8.7 Analysis of electrostatics.***

The ESP-maps based on the MP2 density for each cluster are given in Table 8.5, together with the MKS charges based on the SCF and MP2 densities. The AIM charges and NAO charges are given as a reference for some selected clusters. The MKS charges seem to agree with the ESP-maps. We see that the SCF and MP2 density based charges differ from each other. Both the NAO and AIM charge-schemes give a negative charge to the oxygen atom for Cluster I, while the MKS charge-scheme gives a positive charge to the oxygen. The NAO and AIM charge-schemes agree with the MKS charge-scheme in terms of that the carbon atom in Cluster II is positively charged. We see that no matter what charges are used, it would be difficult to cancel the strong repulsion between the positive charge of the carbon and hydrogen atoms in the C $\cdots$ H bonded dimer. This dimer is of course known to be the most stable one and a force field would have to have a very large vdW attractive force to cancel this repulsive electrostatic interaction. This is chemically incorrect. Therefore, designing a force field for the modeling of this cluster and other CO/HF clusters would fail unless the atomic point charge model is replaced by a more accurate multipole model. We will show in the next section that it is actually impossible, due to these positive charges, to obtain the C $\cdots$ H bonded dimer as the global minimum and the O $\cdots$ H bonded dimer as the local minimum with a basic force field such as UFF or a hybrid QM/MM method incorporating UFF.

## ***8.8 CO/HF force field optimization***

Optimizing a force field for ONIOM-EE in *Gaussian 03* is time consuming if one uses constrained bond lengths in the MM system. We could have attempted using several geometry optimizations to find the optimized parameters for the force field, but we already felt that the polarization of the wave function of the QM system by a positive charge on the carbon in the C $\cdots$ H bonded CO/HF cluster would be inaccurate. Therefore, for the optimization of the CO/HF clusters we only used the frozen geometry optimization described earlier in **Chapters 5** and **7**. This force field optimization method has the advantage that it is faster and the results are independent of the micro-iterations optimizer in *Gaussian 03*. The same GA and quasi-Newton algorithm methods were used for the optimization of the  $R_0$ -values.