

IIT, BOMBAY

M.Sc. PROJECT
AUTUMN 2016

Ab Initio Molecular Dynamics
using Fragment-Based
Electronic Structure
Calculations

Author:

Ankit Mahajan
Roll No. 120100027

Guide:

Prof. Srinivasan Iyengar
Department of Chemistry, Indiana University, Bloomington

Co-guide:

Prof. Aftab Alam
Department of Physics, IIT Bombay

Declaration

I, Ankit Mahajan, understand that plagiarism is defined as any one or the combination of the following:

1. Uncredited verbatim copying of individual sentences, paragraphs or illustrations (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.
2. Uncredited improper paraphrasing of pages or paragraphs (changing a few words or phrases, or rearranging the original sentence order)
3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did or wrote what. (Source: IEEE, The Institute, Dec. 2004)

I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks. I affirm that no portion of my work can be considered as plagiarism and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the seminar report may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Abstract

The goal of this project is to study and explore potential applications of a novel fragment-based electronic structure technique to *ab initio* molecular dynamics (AIMD). This method, referred to as PIE-ONIOM, uses an ONIOM-based fragmentation approach to do electronic structure calculations[16, 17] by making use of the set theoretic principle of inclusion-exclusion(PIE). Salient features of this method include a large reduction in computational cost without significant losses in accuracy and possibility of combining extended Lagrangian AIMD methods with post-Hartree-Fock electronic structure techniques. These features make this method attractive for large scale calculations. The focus in this project will be on efficiently accounting for strong correlation effects within this framework. DMRG has been established as an accurate method for electronic structure calculation, especially effective for systems with multiconfigurational character, which is exhibited by chemical systems at stretched geometries[2]. In MD, the system geometry can become stretched away from equilibrium. Use of DMRG, on top of an appropriate method to account for dynamic correlation, can thus lead to more accurate results. Recently, analytic DMRG gradients have been used for geometry optimizations[11, 19]. Availability of analytic gradients opens up the possibility of using DMRG for on-the-fly AIMD simulations, although this has not yet been explored.

In the first phase, I studied the theory of PIE-ONIOM method and its application to AIMD. I also learned about the details computation of analytic DMRG gradients. The quantum chemical computational software packages Orca and Psi4 have implemented DMRG calculations. I modified the original PIE-ONIOM code to call these packages for energy and gradient calculations, in addition to Gaussian. Beyond allowing the use of DMRG, this development is useful in itself. It makes possible the use of a combination methods implemented in multiple different platforms. I did benchmark calculations with the solvated Zundel cation ($\text{H}_{13}\text{O}_6^+$), using the modified code, to check for consistency across different packages. The results were found to be in good agreement.

The report is organized as follows: In the first chapter, ONIOM and fragment-based approaches to electronic structure calculation are introduced. Theory of static PIE-ONIOM calculations is reviewed in detail. Also included are the results of benchmark calculations, potential energy surface (PES) scans, with the solvated Zundel cation. In the second chapter, a brief discussion of *ab initio* molecular methods, with focus on Born-Oppenheimer Molecular Dynamics (BOMD) and Atom-centered Density Matrix Propagation (ADMP), is followed by applications of PIE-ONIOM to these methods.

Finally, a rough outline of the work to be done in the second phase of the project is sketched.

Contents

1	PIE-ONIOM Theory and calculations	5
1.1	Introduction	5
1.2	Hybrid methods	6
1.2.1	QM/MM	6
1.2.2	ONIOM	8
1.3	Fragment-based methods	11
1.3.1	Cardinality Guided Molecular Tailoring Approach (CG-MTA)	12
1.3.2	PIE-ONIOM	13
1.4	PIE-ONIOM implementation	15
1.4.1	Extension to multiple platforms	17
2	Applications of PIE-ONIOM to AIMD	21
2.1	<i>Ab initio</i> Molecular Dynamics[20]	21
2.1.1	Theory	22
2.2	AIMD with PIE-ONIOM	24
2.3	Future work	25

Chapter 1

PIE-ONIOM Theory and calculations

In this chapter, a brief overview of hybrid techniques is presented first. Then the concept of fragment based calculations using the principle of inclusion-exclusion is introduced and illustrated with concrete examples. The chapter concludes with results of benchmark calculations with the solvated Zundel cation.

1.1 Introduction

“The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.”^[5] Although the first part of this famous quote by Paul Dirac may be disputed by those grappling with the unification of quantum mechanics and general relativity, the second part is still true today. Since 1929 we have come a long way in the pursuit of search for fast approximate methods to solve the Schrödinger equation. Development of exceedingly speedy computers has played a crucial role in this process. But even with the fastest computer technology available today, there are still severe constraints on the size of the systems that can be treated accurately. Only systems with 18 electrons can be solved exactly (Full CI). This is due to the exponential growth of the Hilbert space with the number of electrons, making it very hard to get exact results as system size increases.

Correspondingly, the time complexity of electronic structure methods also scales unfavorably. The less accurate independent particles techniques like Hartree-Fock (HF) and Density Functional Theory (DFT) have computation times scaling as $O(N^4)$, while the more accurate correlated post Hartree-Fock (pHF) methods like DMRG, MP2 and CCSD(T) scale as $\Omega(N^4)$, $O(N^5)$ and $O(N^6)$ respectively.

A way to get around these difficulties is the composite quantum chemical approach, which involves making a series of smaller calculations to extrapolate to an accurate final result. This can be accomplished in many ways. In the following sections, we will look at two ways, *viz.* hybrid methods and fragment-based methods. The PIE-ONIOM method combines these two techniques.

1.2 Hybrid methods

In this composite approach, the large system is divided into layers which are treated at different levels of theory. These hybrid methods are based on the idea that only certain chemically active parts of the system need to be treated with accurate methods, while the rest of the system can be sufficiently described by less expensive methods. This approach has been widely used for simulating reactions in solutions. It has also played an important role in understanding reactions involving biological macromolecules, organic molecules and nanomaterials.

1.2.1 QM/MM

The first attempts at a hybrid approach involved a combination of molecular mechanics (MM) and quantum mechanics (QM). It was first used by Honig and Karplus[10] and developed further by many researchers[6, 28, 32, 33]. Karplus, Levitt and Warshel received the 2013 Nobel Prize in Chemistry for their contributions to this formalism. This set of methods is now collectively termed *QM/MM*. While MM methods are fast, they cannot account for reactive phenomena like formation or breaking of bonds. In the QM/MM approach, the system is divided into two parts: a QM region, which contains the chemically active part of the large system, and a MM region, which contains the remaining atoms. An effective Hamiltonian for the system can be written as[6]

$$\hat{H}_{eff} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \quad (1.1)$$

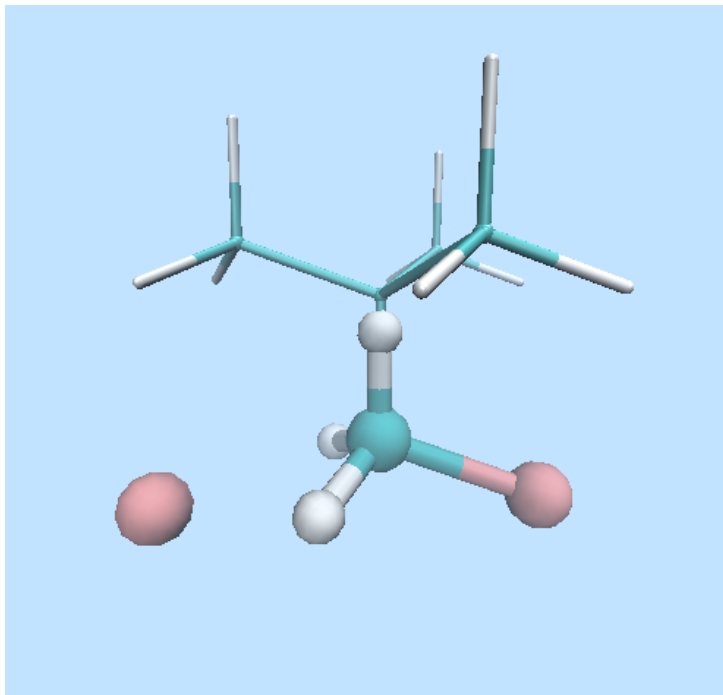


Figure 1.1: QM/MM regions of the S_N2 reaction complex: $\text{Cl}^- + \text{CH}_2(\text{CH}_3)_3\text{Cl}$. H along the C-C line is a link atom (ball-and-stick \rightarrow QM region and tube \rightarrow MM region)

If an *ab initio* QM method is used¹

$$\hat{H}_{QM} = -\frac{1}{2} \sum_i \nabla_i^2 + \sum_{ij} \frac{1}{r_{ij}} - \sum_{i\alpha} \frac{Z_\alpha}{r_{i\alpha}} + \sum_{\alpha\beta} \frac{Z_\alpha Z_\beta}{r_{\alpha\beta}} \quad (1.2)$$

where, i , j and α , β represent electronic and nuclear coordinates, respectively, inside the QM region.²

$$\hat{H}_{MM} = E_{MM}(R_M) \quad (1.3)$$

where, indices M refer to the atoms inside the MM region. This is a scalar operator. The interaction Hamiltonian can be modeled in different ways.

¹Born-Oppenheimer approximation is assumed throughout this chapter.

²In cases where the boundary between QM and MM regions cuts through covalent bonds, it is necessary to cap the dangling bonds with link atoms or localized orbitals, to avoid unphysical results. For the systems considered in this report, the system will always be divided so as not to cut through bonds. So I will gloss over this caveat here.

One possible form is given by³

$$\hat{H}_{QM-MM} = - \sum_{iM} \frac{q_M}{r_{iM}} - \sum_{\alpha M} \frac{Z_\alpha q_M}{r_{\alpha M}} + \sum_{\alpha M} \left(\frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^6} \right) \quad (1.4)$$

Here q_M is the charge on an atom in the MM region. The first term is the electrostatic interaction between electrons in the QM region and atoms in the MM region. The last two terms are electrostatic and Lennard-Jones interactions between the QM nuclei and MM atoms. If $|\psi\rangle$ is the state of the QM part, then

$$\begin{aligned} E_{QM/MM} &= \langle \psi | \hat{H}_{eff} | \psi \rangle = \langle \psi | \hat{H}_{QM} - \sum_{iM} \frac{q_M}{r_{iM}} + \sum_{\alpha M} \frac{Z_\alpha q_M}{r_{\alpha M}} | \psi \rangle + E_{MM} + E_{QM-MM} \\ &= E_{QM} + E_{MM} + E_{QM-MM} \end{aligned} \quad (1.5)$$

Note that \hat{H}_{MM} and the last term of \hat{H}_{QM-MM} (denoted by E_{QM-MM}) are constants, independent of electronic coordinates. This scheme of adding interactions with MM atomic charges to \hat{H}_{QM} is called electronic embedding (EE). Alternatively, this electrostatic interaction between QM electrons and MM atoms can be approximated as an MM interaction, using charge densities obtained from the QM calculation. This is called mechanical embedding (ME). EE is more accurate, but also more expensive, compared to ME. Many other variations have also been explored. Forces are obtained by differentiating this expression with respect to R_α and R_M .

$$\mathbf{F}_\alpha = - \frac{\partial E}{\partial \mathbf{R}_\alpha}, \quad \mathbf{F}_M = - \frac{\partial E}{\partial \mathbf{R}_M} \quad (1.6)$$

1.2.2 ONIOM

The QM/MM methods are *additive*, in that energies of QM and MM regions and their interactions are calculated separately and then added together. In 1995, Morokuma and co-workers proposed an extrapolative scheme to calculate the energy using an integrated QM/MM calculations[21]. In this method, called IMOMM- Integrated Molecular Orbital + Molecular Mechanics, the chemically active part of the system, called the model system, is again treated using more accurate QM methods. But in contrast to QM/MM methods, no separate calculation is performed on the complement of the model system.

³Again, if there are dangling bonds, then bonding terms need to be included in \hat{H}_{QM-MM}

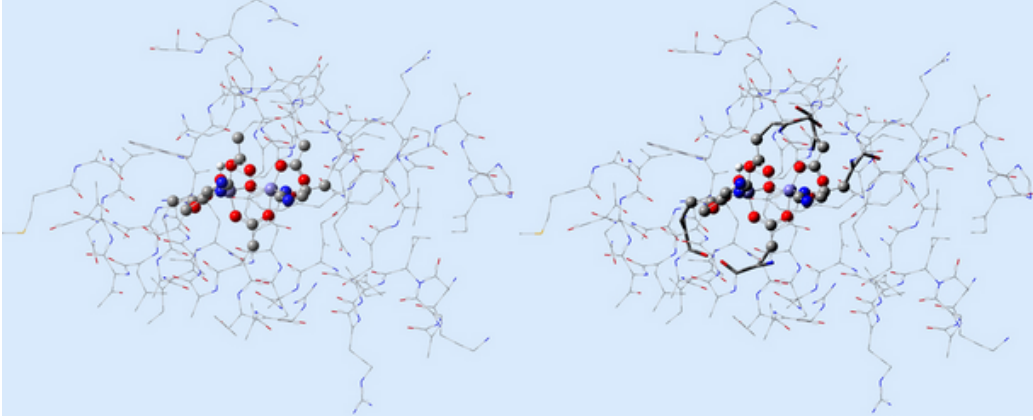


Figure 1.2: ONIOM layer definitions for Ribonucleotide Reductase (ball-and-stick \rightarrow model system, tube \rightarrow intermediate system, wireframe \rightarrow real system)

Left: 2-layer ONIOM, Right: 3-layer ONIOM (H atoms not shown)[13]

Instead both the model system and the whole system, referred to as the real system, are treated using MM. The total energy is then given by the expression

$$E_{IMOMM} = E_{QM,model} + E_{MM,real} - E_{MM,model} \quad (1.7)$$

The last term removes doubly-counted contribution of the model system. Note that E_{IMOMM} is the same as $E_{QM/MM}$ (1.5) if, $E_{MM,real} - E_{MM,model} = E_{MM} + E_{QM-MM}$. In this scheme, the interactions of the model system with other atoms doesn't need to be calculated separately, they are accounted for in the low level calculation of the real system. The method can be further modified by including the charges in the environment while calculating the model system energies (electronic embedding). One advantage of the extrapolative approach is that it only involves energy calculations of chemically realistic systems. So the low level calculations, can be done using methods other than MM. In particular, less accurate, fast QM methods can be employed for these calculations. This method is called *IMOMO* (Integrated Molecular Orbital + Molecular Orbital)[12].

$$E_{IMOMO} = E_{QM-high,model} + E_{QM-low,real} - E_{QM-low,model} \quad (1.8)$$

This formalism can be extended to allow more than 2 layers treated at different levels of theory. All such methods are collectively termed *ONIOM* (Our own N -layered Integrated molecular Orbital molecular Mechanics)[29]. For

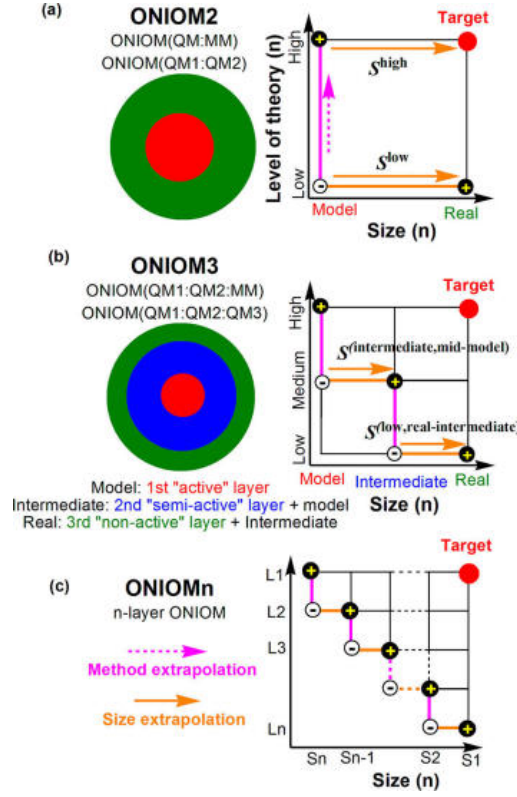


Figure 1.3: Schematic illustration of the extrapolation scheme for (a) two layer ONIOM, (b) three layer ONIOM, (c) n-layer ONIOM[3]

example, a three layer ONIOM energy is expressed as

$$E_{ONIOM3} = E_{high,model} + E_{medium,intermediate} - E_{medium,model} + E_{low,real} - E_{low,intermediate} \quad (1.9)$$

where, the system is divided into model and intermediate layers. The model system is treated with the most accurate theory, intermediate with an intermediate accurate theory and the whole system with the least accurate theory. The general expression for an n -layer treatment is given by

$$E_{ONIOMn} = \sum_{i=1}^n E\{level(i), model(n+1-i)\} - \sum_{i=2}^n E\{level(i), model(n+2-i)\} \quad (1.10)$$

where, the lower the i more accurate the theory and smaller the model system. Although implementation of an n -layer ONIOM method, for $n \geq 4$, is not difficult, only ONIOM2 and ONIOM3 are widely used in practice.

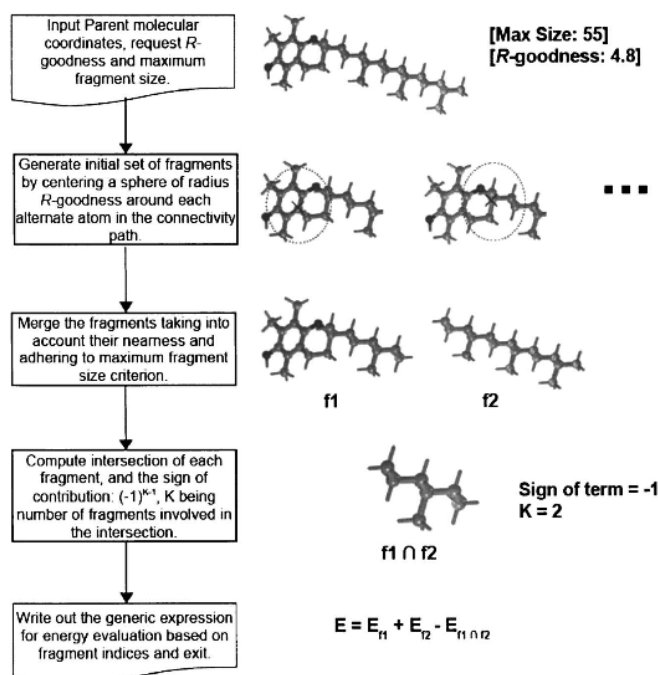
Error in an ONIOM calculation can be estimated using the S -value test. In an ONIOM2 calculation, it is given by the difference between the ONIOM energy and the high level energy of the real system.

$$\begin{aligned}\Delta &= E_{\text{ONIOM2}} - E_{\text{high,real}} \\ &= (E_{\text{low,real}} - E_{\text{low,model}}) - (E_{\text{high,real}} - E_{\text{high,model}}) \\ &= S_{\text{low}} - S_{\text{high}}\end{aligned}\tag{1.11}$$

The S -values (substituent values) are defined as the difference between real and model system energies at the given level of theory. They can be thought of as representing the effect of the environment on the model system. Thus if the low level theory mimics the S -value of the higher level theory, their combination would lead to a small error in the ONIOM2 calculation. This is known as the S -value test.

1.3 Fragment-based methods

Although ONIOM methods are useful for large scale calculations, the need to identify a chemically active region limits their applicability. In many cases it is necessary to treat multiple parts of the system with an accurate theory. In such cases another composite approach, involving breaking up a large system into fragments, is employed. These are based on the fundamental principle of locality or near-sightedness of electronic interactions⁴. Fragment-based methods scale approximately linearly with system size, because the size of fragments is relatively independent of system size. These methods also allow multi-level parallel implementations, leading to highly efficient calculations. A large number of fragment-based approaches have been developed, some dating as far back as 1960s^[15]. In this section, we will look at the Molecular Tailoring Approach (MTA), which serves as the fundamental building block for PIE-ONIOM.

Figure 1.4: Illustration of the automated fragmentation of α -tocopherol[8]

1.3.1 Cardinality Guided Molecular Tailoring Approach (CG-MTA)

This method was proposed by Gadre in 1994[7]. Since then many variants, modifying and improving the original technique, have been developed, CG-MTA being one of them. MTA implements a divide-and-conquer (DC) approach. A large system is broken into smaller overlapping fragments. Like almost all fragment-based methods, MTA prescribes an automated way of carrying out this fragmentation. It takes as input the maximum fragment size (number of atoms in a fragment) and a parameter called R -goodness (R_g). A sphere of radius R_g is centered on every atom and these spheres are merged with each other recursively based on their proximity to each other to create overlapping fragments until the desired size is reached. The creators of this approach proposed the R_g parameter to assess the quality of the fragmentation scheme. For an atom it is defined as the radius of the largest sphere that can be drawn around it while staying inside the fragment.

⁴There do exist systems where long-range delocalized interactions are important (e.g. highly conjugated systems). Fragmentation-based approaches cannot be applied in such cases.

If the atom is part of multiple fragments a maximum over all fragments is taken. The R_g value for the system is defined as the minimum of R_g values of all atoms. Intuitively, this parameter signifies the size of the environment about an atom, the interactions with which are treated accurately in the approach. Thus larger the R_g and smaller the number of atoms per fragment the more accurate will be the calculation. The fragments thus obtained are called *primary subsystems* (or fragments). Any dangling covalent bonds are capped with hydrogen link atoms along the bond, in a manner similar to the ONIOM methods. It is advisable not to break double bonds or aromatic rings. Overlaps between different fragments are also calculated and they are termed *derivative subsystems* (or fragments). This process is illustrated in fig (1.4).

After creating the fragments, energies (or other desired properties like one electron densities) of the primary and derivative subsystems, including hydrogen link atoms, are calculated using an accurate electronic structure technique. These values are used to get the energy of the whole system. Since the fragments, in general, have non-zero overlap, care needs to be taken to avoid overcounting. This is accomplished using the set theoretic principle of inclusion exclusion. This principle relates the cardinalities of constituent sets and their intersections with that of their union (cardinality guided).

$$\begin{aligned} \left| \bigcup_{i=1}^n A_i \right| = & \sum_{i=1}^n |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cap A_j| + \sum_{1 \leq i < j < k \leq n} |A_i \cap A_j \cap A_k| - \dots \\ & + (-1)^{n-1} |A_1 \cap A_2 \cap \dots \cap A_n| \end{aligned} \quad (1.12)$$

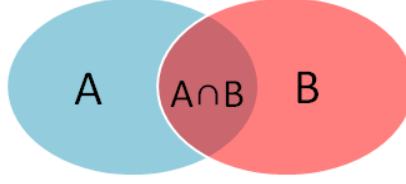
Accordingly the energy in the MTA approach is calculated as

$$\begin{aligned} E_{MTA} = & \sum_{i=1}^n E(A_i) - \sum_{1 \leq i < j \leq n} E(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} E(A_i \cap A_j \cap A_k) - \dots \\ & + (-1)^{n-1} E(A_1 \cap A_2 \cap \dots \cap A_n) \end{aligned} \quad (1.13)$$

This calculation of system energy using individual fragments energies is akin to a tailor stitching clothes together. Thus the name Molecular Tailoring Approach. Note that the hydrogen caps also get canceled in this equation.

1.3.2 PIE-ONIOM

MTA as originally proposed and as presented above does not take into account long range inter-fragment interactions, which limits its applicability.



Several modifications have been proposed to remedy this. Some of them make use of the multilayer formalism of ONIOM to include inter-fragment interactions. The PIE-ONIOM technique developed by Iyengar et.al[18] is one such technique. It uses the ONIOM formalism on top of the fragment approach. In ONIOM, a lower level calculation of the whole system is used to estimate these interactions. Consider a model system consisting of two overlapping fragments A and B . The MTA energy is given by

$$E_{MTA}^{high} = E^{high}(A) + E^{high}(B) - E^{high}(A \cap B) \quad (1.14)$$

It is feasible to do this calculation with an accurate QM theory, as only energies of smaller fragments are actually calculated. Inter-fragment interactions need to be added to this energy. This can be calculated as⁵

$$\begin{aligned} E_{frag-int}^{low} &= E^{low} - E_{MTA}^{low} \\ &= E^{low} - \{E^{low}(A) + E^{low}(B) - E^{low}(A \cap B)\} \end{aligned} \quad (1.15)$$

where E^{low} is the energy of the whole system calculated with a lower level of theory. Thus the total energy can be written as

$$\begin{aligned} E_{PIE-ONIOM} &= E_{MTA}^{high} + E_{frag-int}^{low} \\ &= E^{low} + \{E^{high}(A) - E^{low}(A)\} + \{E^{high}(B) - E^{low}(B)\} \\ &\quad - \{E^{high}(A \cap B) - E^{low}(A \cap B)\} \end{aligned} \quad (1.16)$$

This can be generalized to an arbitrary collection of fragments. Using a more general notation

$$\begin{aligned} E_{PIE-ONIOM} &= E^{level,0}(0) + \sum_{i=1}^n S(i) - \sum_{1 \leq i < j \leq n} S(i \cap j) + \sum_{1 \leq i < j < k \leq n} S(i \cap j \cap k) \\ &\quad - \dots + (-1)^{n-1} S(1 \cap 2 \cap \dots \cap n) \end{aligned} \quad (1.17)$$

⁵This is a semi-classical approximation. The quality of the approximation is assessed through benchmark calculations (*vide infra*).

with

$$S(X) = E^{level,1}(X) - E^{level,0}(X) \quad (1.18)$$

where X denotes both the primitive $(1, 2, \dots, n)$ and derivative $(i \cap j, i \cap j \cap k \dots)$ subsystems. Levels refer to the accuracy of the employed electronic structure calculation technique, $\{level, 1\}$ being more accurate than $\{level, 0\}$. This can be further modified by further subdividing the fragments and recursively calculating the corrections. The correction terms can be recursively extrapolated through η subfragments within a fragment i .

$$S(i) = \sum_{\alpha_i=0}^{\eta} \left(E_{recursive-fragment, \alpha_i}^{level, \alpha_i+1} - E_{recursive-fragment, \alpha_i}^{level, \alpha_i} \right) \quad (1.19)$$

where, α_i denotes the sub-fragment or layer. Another possible modification is to use electronic embedding when calculating fragment energies.

Availability of analytical gradients is absolutely essential for a method to be useful for AIMD. PIE-ONIOM gradients can be expressed as

$$\begin{aligned} \frac{\partial E_{PIE-ONIOM}}{\partial \mathbf{R}} &= \frac{\partial E^0}{\partial \mathbf{R}} + \sum_{i=1}^n \frac{\partial S(i)}{\partial \mathbf{R}} - \sum_{1 \leq i < j \leq n} \frac{\partial S(i \cap j)}{\partial \mathbf{R}} + \dots \\ &+ (-1)^{n-1} \frac{\partial S(1 \cap 2 \cap \dots \cap n)}{\partial \mathbf{R}} \end{aligned} \quad (1.20)$$

where, \mathbf{R} is the general system coordinate. If gradients for a fragment calculation are calculated using coordinates different from the global ones, they can be transformed back using the chain rule

$$\frac{\partial S(i)}{\partial \mathbf{R}} = \left(\frac{\partial E^{level,1}(i)}{\partial \mathbf{R}_i} - \frac{\partial E^{level,0}(i)}{\partial \mathbf{R}_i} \right) \frac{\partial \mathbf{R}_i}{\partial \mathbf{R}} \quad (1.21)$$

1.4 PIE-ONIOM implementation

The Iyengar group has developed a wrapper C++ code to do the PIE-ONIOM calculations. It takes as input the molecular geometry and information about the QM methods to be used at different levels. An automated fragmentation scheme has been implemented for water clusters. It is designed so that all hydrogen bond interactions are treated accurately. First, for all oxygen atoms, the minimum distance to another oxygen atom is calculated. All oxygen atoms within 10% of this distance are included in one fragment, while a maximum of two oxygen atoms are allowed in one fragment. All hydrogens within 1.4 Å distance from the oxygens in the fragment are also included. This process ensures that all hydrogen bonds will be treated with

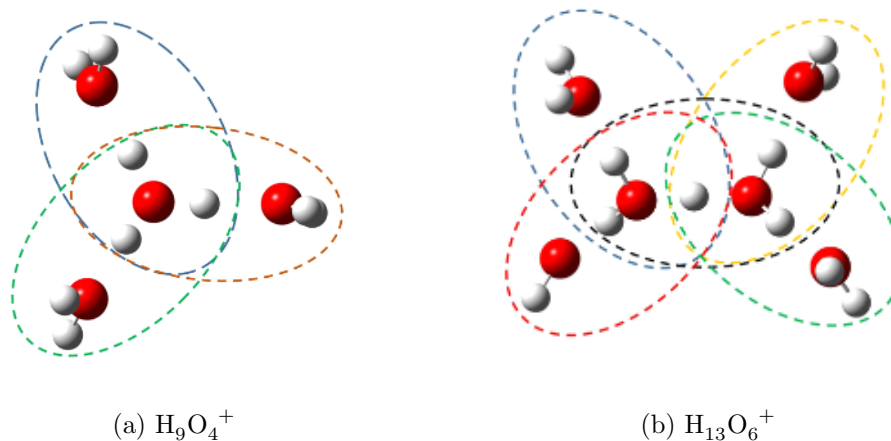


Figure 1.5: Fragmentation of water clusters

a high level calculation of some fragment. Fragments without any electrons are discarded. There is also an option of taking user input to allow customized fragmentation. Overlaps between primary fragments are calculated with an efficient binary bit manipulation algorithm. Once all the primary and derivative fragments are decided, the corresponding subsystems are created by capping any dangling bonds with hydrogen link atoms and their charges are calculated. QM electronic structure calculations of all the subsystems are then delegated to the desired software package, by creating appropriate input files. The energy of the whole system is calculated at the specified low level and energies of all the subsystems are calculated at both levels. The code is capable of using electronic embedding for the fragment calculations. This is accomplished by extracting the Mulliken charges from the low level calculation of the whole system. The results of electronic structure calculations, energies, forces or dipole moments, are extracted from the output files and combined according to the PIE-ONIOM equations. The code is also MPI-enabled, allowing it to exploit the inherent parallelizability of fragment-based calculations.

Large scale benchmark calculations with H_9O_4^+ and $\text{H}_{13}\text{O}_6^+$ were reported in the first PIE-ONIOM publication[18]. The reduced dimensional potential energy surface (PES) scans of the Eigen and solvated Zundel systems, calculated using PIE-ONIOM, were found to be in good agreement (< 1 kcal/mol) with those calculated using high level theory on the whole system. The geometries scanned lie along the normal modes (calculated with B3LYP/6-31+G**) of the system about the optimized geometry (also

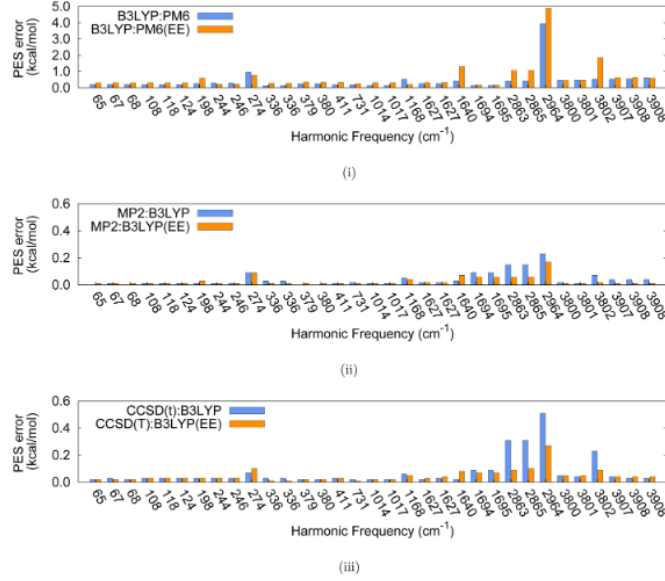


Figure 1.6: Potential energy surface errors for H_9O_4^+ . All B3LYP, MP2 and CCSD(T) calculations done with 6-31+G** basis set. A calculation with A as the high level of theory and B as the low level of theory is denoted as A:B. EE refers to electronic embedding.

calculated with B3LYP/6-31+G**). The graph below shows the errors for different PIE-ONIOM calculations for the Eigen system calculated as

$$\epsilon = \sum_i^{N_p} \left| \left(E_i^{frag} - E_{ref}^{frag} \right) - \left(E_i^{frag} - E_{ref}^{frag} \right) \right| \quad (1.22)$$

where, N_p is the number of calculations (21 geometries along each of the 33 modes) and ref denotes calculations on the optimized geometry. The electronic structure calculations were done using the Gaussian software.

1.4.1 Extension to multiple platforms

While the initial version of the PIE-ONIOM code only allowed calculations with the Gaussian package, I have modified the code to add the functionality of using Orca and Psi4. While the need to do this arose out of the absence of DMRG calculations in Gaussian, this development is quite useful in itself. One, it increases the number of methods that can be used in the PIE-ONIOM calculation. Second, it allows the use of multiple methods implemented in different packages in one calculation. Calculations at different levels of theory can be done with different packages, based on their efficiency.

This entailed an understanding of the original code. Different packages have different input and output file formats. There are even differences in their implementations of the same method. For example, implementations of the basis sets are often slightly different. They use different molecule orientations for calculating forces. So, in order to combine the results from different tools, they need to be converted to a form consistent across packages. This was done for some commonly used methods, *viz.* HF, DFT(B3LYP), MP2 and CCSD. Table (1.1) shows the details of one such calculation. These calculations were done on an Intel Core i7-4770 3.40 GHz processor.

Table 1.1: Details of a $\text{H}_{13}\text{O}_6^+$ MP2:B3LYP PIE-ONIOM calculation with PSI4. This calculation was done with the aug-cc-pVDZ basis

Entity	B3LYP	MP2
Whole system	-459.092029386	-
Primary Fragment 1	-153.2081112394	-152.8379018167
Primary Fragment 2	-153.1305816997	-152.7609535409
Primary Fragment 3	-153.1367243127	-152.7668477484
Primary Fragment 4	-153.1941978610	-152.8245727860
Primary Fragment 5	-153.1925218134	-152.8231040857
Derivative Fragment 1	-76.6532226680	-76.4673705372
Derivative Fragment 2	-76.7044716652	-76.5192670148
PIE-ONIOM energy	-457.985386	

The results below show sampling of the PES of the solvated Zundel cation with different combinations of methods and packages. All calculations were done using the aug-cc-pVDZ basis set. The geometries chosen are part of a AIMD simulation trajectory, which represents a curve on the PES. First, a B3LYP optimization was done to calculate the equilibrium geometry. All PES curves are calibrated by assigning a zero energy to this geometry. The geometry with the least energy out of the 21 (labeled 0 in the graphs) has RMS deviation of 0.18Å from the optimized geometry. These calculations were done without electronic embedding. Table (1.2) shows the RMS errors of the PIE-ONIOM surfaces calculated against the corresponding full system high level calculations.

Gradient calculations were also carried out using the same methods. The results were found to be in agreement across packages. For instance, the RMS difference between the full CCSD PSI4 gradients and CCSD(ORCA):MP2(PSI4) PIE-ONIOM gradients came out to be 0.044 Kcal mol⁻¹Bohr⁻¹ averaged over the same 21 geometries.

Table 1.2: Details of the PES calculations

Calculation	RMS Error (kcal/mol)	Time/Geometry (min)
CCSD(ORCA):B3LYP(PSI4)	0.263	3.67
CCSD(ORCA):MP2(PSI4)	0.056	2.89
CCSD(ORCA)	-	118.56
MP2(PSI4):HF(Gaussian)	0.376	-
MP2(PSI4):B3LYP(PSI4)	0.340	1.34
MP2(PSI4)	-	1.42

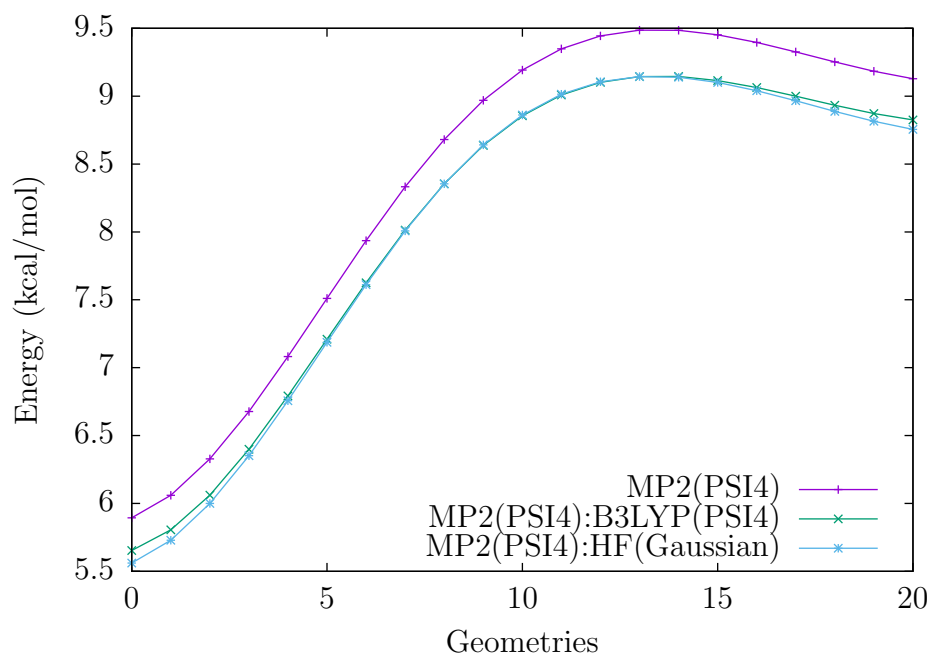


Figure 1.7: Sampling of the PES of the solvated Zundel cation. All calculations were done using aug-cc-pVDZ

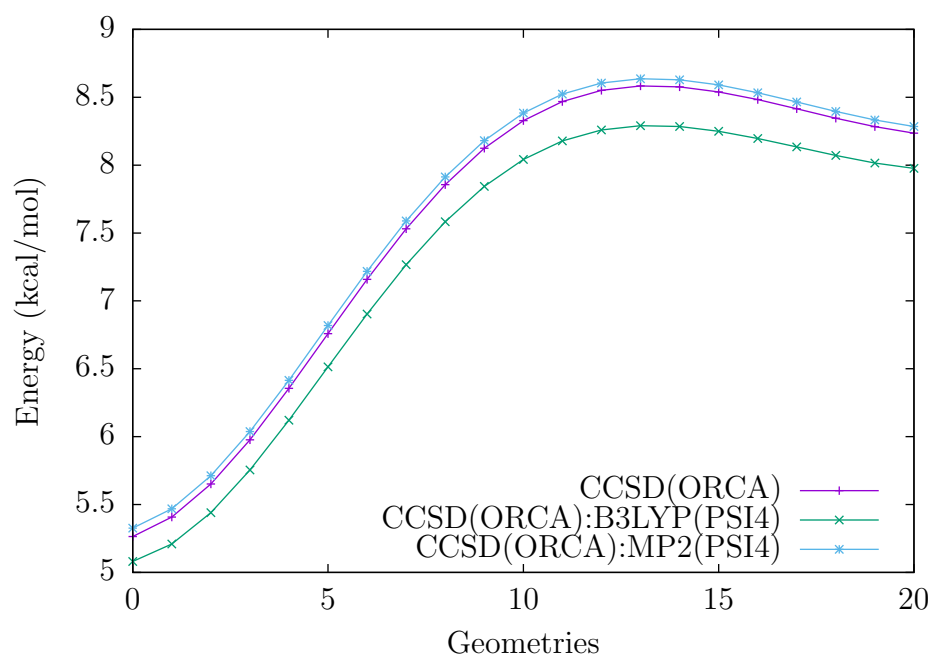


Figure 1.8: Sampling of the PES of the solvated Zundel cation. All calculations were done using aug-cc-pVDZ

Chapter 2

Applications of PIE-ONIOM to AIMD

In this chapter, theory of on-the-fly *ab initio* molecular dynamics¹ PIE-ONIOM is discussed first. In particular, Born-Oppenheimer Molecular Dynamics (BOMD) and Atom-centered Density Matrix Propagation (ADMP) methods are reviewed. Then the application of PIE-ONIOM to quantum molecular dynamics are also being studied. AIMD is discussed in the context of these two methods.

2.1 *Ab initio* Molecular Dynamics[20]

Traditionally, classical MD simulations have been done using force fields, which are potentials calculated either empirically or using multiple-body interactions modeled on accurate theoretical electronic structure calculations. Nuclear trajectories can then be calculated using Newton’s laws. Although this method has served as an important tool in the study of numerous chemical systems, it has several shortcomings. One of them being the need to parametrize the interactions for a vast number of atoms and settings. Another critical drawback is that it suffers from the *dimensionality bottleneck*. These methods calculate the global potential energy surface before calculating trajectories. In a classical trajectory calculation, this step requires most resources, especially for large systems. An N -body system has 3^{N-6} degrees of freedom. So calculation of the global PES scales roughly as 10^N , which is bound to be very expensive for large systems. *Ab initio* MD tackles these difficulties, while at the same time producing accurate results. In *on-the-*

¹Here, only the theory of *classical* MD (nuclei are treated classically) is presented. Applications of PIE-ONIOM to *quantum* MD are a subject of current research.

fly AIMD calculations the global potential energy surface is not calculated. Gradients at a point on the surface are calculated using QM methods and nuclei are propagated using these.

2.1.1 Theory

The time independent Schrödinger equation is

$$\hat{H}|\psi\rangle = i\hbar \frac{\partial}{\partial t}|\psi\rangle \quad (2.1)$$

In the position representation the Hamiltonian is given by

$$\hat{H} = - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 - \sum_i \frac{\hbar^2}{2m_i} \nabla_i^2 + \frac{1}{4\pi\epsilon_0} \sum_{i<j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (2.2)$$

$$- \frac{1}{4\pi\epsilon_0} \sum_{I,i} \frac{e^2 Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \frac{1}{4\pi\epsilon_0} \sum_{I<J} \frac{e^2 Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \quad (2.3)$$

$$= - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 + \hat{H}_e(\mathbf{r}_i, \mathbf{R}_I) \quad (2.4)$$

where, r_i and R_i denote electronic and nuclear degrees of freedom. This equation treats both electrons and nuclei quantum mechanically. The large number of degrees of freedom makes it impossible to solve this exactly in almost all cases. A number of ways of approximating this equation exist.

Ehrenfest MD is one of the oldest such approximations. The final equations are given as

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I \int \Psi^* \hat{H} \Psi d\mathbf{r} \quad (2.5)$$

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}_e \Psi \quad (2.6)$$

This coupled set of equations can be solved on-the-fly to calculate classical trajectories. The electronic wavefunction can be clamped to one state, to further simplify the calculation. One drawback of this method is that the length simulation time step is dictated by the fast electronic motion. This method fails to take advantage of the large difference in the time scales of nuclear and electronic motion.

A more widely used approach is the Born-Oppenheimer MD. Its equations are given by

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I \min \left\{ \int \Psi^* \hat{H} \Psi d\mathbf{r} \right\} \quad (2.7)$$

$$E_0\Psi_0 = \hat{H}\Psi_0 \quad (2.8)$$

Here, the electronic ground state is calculated self-consistently at each time step. This method can also be extended to calculate excited state dynamics. In this method the time-step is decided by considerations of the slow nuclear motion. But it comes at the expense of having to diagonalize the Hamiltonian at every time step.

Car and Parrinello[1] proposed a general formalism that combines the advantages of the previous approaches while also doing away with their shortcomings. They presented a way of calculating dynamics trajectories on the slow time scale of nuclear motion without diagonalizing the electronic Hamiltonian at every time step. This is achieved by converting the quantum mechanical problem of electronic wavefunction propagation into a classical one by writing the electronic degrees of freedom by the so called *fictitious* dynamic variables. They proposed a new class of extended Lagrangians (these methods are also referred to as extended Lagrangian methods)

$$\mathcal{L} = \sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \mu \langle \dot{\phi} | \dot{\phi} \rangle - \langle \Psi_0 | \hat{H}_e | \Psi_0 \rangle + \text{constraints} \quad (2.9)$$

where μ denotes the fictitious electronic mass and ϕ_i are the orbitals making up the electronic wavefunction (e.g. orbitals in a Slater determinant). Constraints are needed to conserve the orthonormality of orbitals. The corresponding Euler-Lagrange equations are

$$M_i \ddot{\mathbf{R}}_I(t) = - \frac{\partial}{\partial \mathbf{R}_I} \langle \Psi_0 | \hat{H}_e | \Psi_0 \rangle + \frac{\partial}{\partial \mathbf{R}_I} (\text{constraints}) \quad (2.10)$$

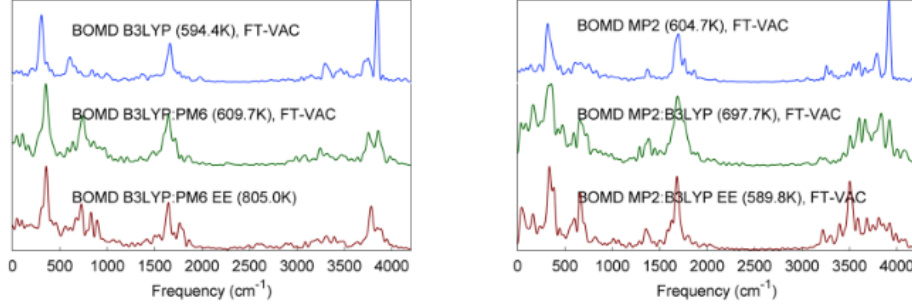
$$\mu \ddot{\phi}_i(t) = - \frac{\delta}{\delta \phi_i^*} \langle \Psi_0 | \hat{H}_e | \Psi_0 \rangle + \frac{\delta}{\delta \phi_i^*} (\text{constraints}) \quad (2.11)$$

It can be shown that the electronic wavefunction propagated using these equations stays close to the ground state and thus the BOMD surface.

Atom centered Density Matrix Propagation is one particular extended Lagrangian method which uses density matrices instead of molecular orbitals as the electronic dynamic variables[24–26]. It uses an atom-centered Gaussian orbital basis. The CP Lagrangian is written as

$$\mathcal{L} = \frac{1}{2} \text{Tr}(\mathbf{V}^T \mathbf{M} \mathbf{V}) + \frac{1}{2} \text{Tr} [\{\mu^{1/4} \mathbf{W} \mu^{1/4}\}^2] - E(\mathbf{R}, \mathbf{P}) - \text{Tr} [\mathbf{\Lambda}(\mathbf{P} \mathbf{P} - \mathbf{P})] \quad (2.12)$$

where, \mathbf{M} , \mathbf{R} and \mathbf{V} are the nuclear masses, positions and velocities, respectively. \mathbf{P} , \mathbf{W} and μ are the electronic density matrix, its velocity

Figure 2.1: Vibrational density of states of $\text{H}_{13}\text{O}_6^+$ [18]

and the fictitious mass diagonal matrix, respectively. $\mathbf{\Lambda}$ is the Lagrangian multiplier matrix to ensure the idempotency of \mathbf{P} and N-representability. The E-L equations are

$$\mathbf{M} \frac{d^2 \mathbf{R}}{dt^2} = - \frac{\partial E(\mathbf{R}, \mathbf{P})}{\partial \mathbf{R}} \Big|_{\mathbf{P}} \quad (2.13)$$

$$\mu^{1/2} \frac{d^2 \mathbf{P}}{dt^2} \mu^{1/2} = - \left[\frac{\partial E(\mathbf{R}, \mathbf{P})}{\partial \mathbf{P}} \Big|_{\mathbf{R}} + \mathbf{\Lambda} \mathbf{P} + \mathbf{P} \mathbf{\Lambda} - \mathbf{\Lambda} \right] \quad (2.14)$$

2.2 AIMD with PIE-ONIOM

PIE-ONIOM can be used in both BOMD and ADMP as an electronic structure method. In BOMD, the ground state energy and gradients can be calculated using PIE-ONIOM as outlined in the previous chapter. At present, long trajectories cannot be calculated as the fragment topology may change during dynamics, which would require dynamic fragmentation. For smaller trajectories (of the order of ps), energy conservation has been verified. Vibrational density of states can be calculated using the Fourier transform of the velocity autocorrelation function (FT-VAC) obtained during trajectory calculations. The graph below taken from ref [18] shows the agreement between vibrational density of states calculated using PIE-ONIOM and that calculated using the corresponding high level of theory.

$$I_V(\omega) = \lim_{T \rightarrow \infty} \int_{t=0}^{t=T} dt \exp(-i\omega t) \langle \mathbf{V}(0) \cdot \mathbf{V}(t) \rangle \quad (2.15)$$

Extended Lagrangian techniques can only be applied to one-particle formalisms. This restricts the use of pHF methods, like CCSD, MP2, in ELMD. PIE-ONIOM can be used to bridge this gap[16]. The lower level electronic

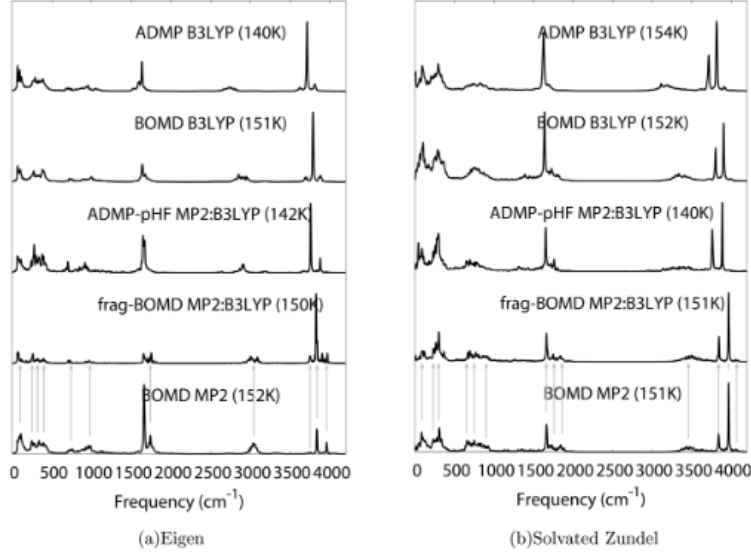


Figure 2.2: Vibrational density of states of (a) H_9O_4^+ , (b) $\text{H}_{13}\text{O}_6^+$ at 150K[16]

structure is propagated using ELMD while the fragments are treated using BOMD, which allows the use of pHF methods. The lower level full-system electronic structure is not converged at each step. In particular, ADMP has been used with PIE-ONIOM for MD trajectory calculations[16]. The EOM can be written as

$$\mathbf{M} \frac{d^2 \mathbf{R}}{dt^2} = - \frac{\partial E^{PIE-ONIOM}(\mathbf{R}, \mathbf{P}_{low0})}{\partial \mathbf{R}} \Big|_{\mathbf{P}_{low0}} \quad (2.16)$$

$$\mu_{low0}^{1/2} \frac{d^2 \mathbf{P}_{low0}}{dt^2} \mu_{low0}^{1/2} = - \left[\frac{\partial E(\mathbf{R}, \mathbf{P}_{low0})}{\partial \mathbf{P}_{low0}} \Big|_{\mathbf{R}} + \mathbf{\Lambda}_{low0} \mathbf{P}_{low0} + \mathbf{P}_{low0} \mathbf{\Lambda}_{low0} - \mathbf{\Lambda}_{low0} \right] \quad (2.17)$$

Again the vibrational density of states calculated using PIE-ONIOM was found to be in agreement with that calculated using the corresponding high level full-system calculation.

2.3 Future work

In the second phase of the project, I will focus on PIE-ONIOM calculations of strongly correlated systems. Accounting for non-covalent bonded interactions has been a subject of hot debate. Systems like stacked benzene rings, metallic complexes involving conjugated systems have been hard to describe

quantitatively with high accuracy quantum mechanical methods. I would like to try using DMRG with this fragment based approach to understand these systems better. I will also be doing some molecular dynamics simulations with PIE-ONIOM. Because of the availability of analytic gradients, DMRG can be used for dynamics as well. I would like to work on combining DMRG with other suitable methods to do efficient dynamics.

Bibliography

- [1] Richard Car and Mark Parrinello. “Unified approach for molecular dynamics and density-functional theory”. In: *Physical review letters* 55.22 (1985), p. 2471.
- [2] Garnet Kin-Lic Chan and Martin Head-Gordon. “Exact solution (within a triple-zeta, double polarization basis set) of the electronic Schrödinger equation for water”. In: *The Journal of chemical physics* 118.19 (2003), pp. 8551–8554.
- [3] Lung Wa Chung et al. “The ONIOM method and its applications”. In: *Chemical reviews* 115.12 (2015), pp. 5678–5796.
- [4] Stefan Dapprich et al. “A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives”. In: *Journal of Molecular Structure: THEOCHEM* 461 (1999), pp. 1–21.
- [5] Paul AM Dirac. “Quantum mechanics of many-electron systems”. In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Vol. 123. 792. The Royal Society. 1929, pp. 714–733.
- [6] Martin J Field, Paul A Bash, and Martin Karplus. “A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations”. In: *Journal of Computational Chemistry* 11.6 (1990), pp. 700–733.
- [7] Shridhar R Gadre, Rajendra N Shirsat, and Ajay C Limaye. “Molecular tailoring approach for simulation of electrostatic properties”. In: *The Journal of Physical Chemistry* 98.37 (1994), pp. 9165–9169.
- [8] V Ganesh et al. “Molecular tailoring approach for geometry optimization of large molecules: Energy evaluation and parallelization strategies”. In: *The Journal of chemical physics* 125.10 (2006), p. 104109.

- [9] Mark S Gordon et al. “Fragmentation methods: a route to accurate calculations on large systems”. In: *Chem. Rev* 112.1 (2012), pp. 632–672.
- [10] Barry Honig and Martin Karplus. “Implications of torsional potential of retinal isomers for visual excitation”. In: (1971).
- [11] Weifeng Hu and Garnet Kin-Lic Chan. “Excited-State Geometry Optimization with the Density Matrix Renormalization Group, as Applied to Polyenes”. In: *Journal of chemical theory and computation* 11.7 (2015), pp. 3000–3009.
- [12] Stéphane Humbel, Stefan Sieber, and Keiji Morokuma. “The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and SN2 reaction: RCl+ Cl⁻”. In: *The Journal of chemical physics* 105.5 (1996), pp. 1959–1967.
- [13] Gaussian Inc. *Investigating the Reactivity and Spectra of Large Molecules with ONIOM*. URL: http://www.gaussian.com/g_whitepap/oniom_technote.htm.
- [14] Srinivasan S Iyengar et al. “Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. II. Generalizations based on mass-weighting, idempotency, energy conservation and choice of initial conditions”. In: *The Journal of Chemical Physics* 115.22 (2001), pp. 10291–10302.
- [15] M Klessinger and R McWeeny. “Self-Consistent Group Calculations on Polyatomic Molecules. I. Basic Theory with an Application to Methane”. In: *The Journal of Chemical Physics* 42.10 (1965), pp. 3343–3354.
- [16] Junjie Li, Cody Haycraft, and Srinivasan Sesha Iyengar. “Hybrid extended Lagrangian, post-Hartree-Fock Born-Oppenheimer ab initio molecular dynamics using fragment-based electronic structure”. In: *Journal of chemical theory and computation* (2016).
- [17] Junjie Li and Srinivasan S Iyengar. “Ab Initio Molecular Dynamics Using Recursive, Spatially Separated, Overlapping Model Subsystems Mixed within an ONIOM-Based Fragmentation Energy Extrapolation Technique”. In: *Journal of chemical theory and computation* 11.9 (2015), pp. 3978–3991.

- [18] Junjie Li and Srinivasan S Iyengar. “Ab Initio Molecular Dynamics Using Recursive, Spatially Separated, Overlapping Model Subsystems Mixed within an ONIOM-Based Fragmentation Energy Extrapolation Technique”. In: *Journal of chemical theory and computation* 11.9 (2015), pp. 3978–3991.
- [19] Fengyi Liu et al. “Multireference ab initio density matrix renormalization group (DMRG)-CASSCF and DMRG-CASPT2 study on the photochromic ring opening of spiropyran”. In: *Journal of chemical theory and computation* 9.10 (2013), pp. 4462–4469.
- [20] Dominik Marx and Jürg Hutter. *Ab initio molecular dynamics: basic theory and advanced methods*. Cambridge University Press, 2009.
- [21] Feliu Maseras and Keiji Morokuma. “IMOMM: A new integrated ab initio+ molecular mechanics geometry optimization scheme of equilibrium structures and transition states”. In: *Journal of Computational Chemistry* 16.9 (1995), pp. 1170–1179.
- [22] Nicholas J Mayhall and Krishnan Raghavachari. “Molecules-in-molecules: An extrapolated fragment-based approach for accurate calculations on large molecules and materials”. In: *Journal of chemical theory and computation* 7.5 (2011), pp. 1336–1343.
- [23] Krishnan Raghavachari and Arjun Saha. “Accurate composite and fragment-based quantum chemical models for large molecules”. In: *Chemical reviews* 115.12 (2015), pp. 5643–5677.
- [24] H Bernhard Schlegel et al. “Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals”. In: *The Journal of Chemical Physics* 114.22 (2001), pp. 9758–9763.
- [25] H Bernhard Schlegel et al. “Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. III. Comparison with Born–Oppenheimer dynamics”. In: *The Journal of chemical physics* 117.19 (2002), pp. 8694–8704.
- [26] Ulrich Schollwöck. “The density-matrix renormalization group”. In: *Reviews of modern physics* 77.1 (2005), p. 259.
- [27] Ulrich Schollwöck. “The density-matrix renormalization group in the age of matrix product states”. In: *Annals of Physics* 326.1 (2011), pp. 96–192.

- [28] U Chandra Singh and Peter A Kollman. "A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the $\text{CH}_3\text{Cl} + \text{Cl}^-$ exchange reaction and gas phase protonation of polyethers". In: *Journal of Computational Chemistry* 7.6 (1986), pp. 718–730.
- [29] Mats Svensson et al. "ONIOM: A multilayered integrated MO+ MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and $\text{Pt}(\text{P}(\text{t-Bu})_3)_2 + \text{H}_2$ oxidative addition". In: *The Journal of Physical Chemistry* 100.50 (1996), pp. 19357–19363.
- [30] Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.
- [31] Thom Vreven and Keiji Morokuma. "Hybrid methods: Oniom (qm: mm) and qm/mm". In: *Annual reports in computational chemistry* 2 (2006), pp. 35–51.
- [32] A Warshel and M Karplus. "Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization". In: *Journal of the American Chemical Society* 94.16 (1972), pp. 5612–5625.
- [33] Arie Warshel and Michael Levitt. "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme". In: *Journal of molecular biology* 103.2 (1976), pp. 227–249.