



EINDHOVEN UNIVERSITY OF TECHNOLOGY

BACHELOR FINAL PROJECT

# Automated methods for multiscale modeling

*Anne-Wil van den Heuvel*  
1360183

supervised by  
Björn Baumeier

July 9, 2023

## **Abstract**

The understanding of the interaction between light and matter is essential in various applications, such as LEDs or Solar Cells. Simulating the electronic structure of materials to understand their interaction with light is important for designing materials with desirable properties. This report uses molecular simulations using VOTCA-XTP to execute a DFT calculation including embedding with Ewalds method. The goal is to create an automated workflow to execute these calculations to find the electron affinity and ionization potential. The influence of different input parameters was examined, such as the used basis set in the DFT calculation (svp, tzvp and qzvp) and the cutoff radius used for the embedding. Further steps are needed to clean the input data and to further examine the input parameter dependency.

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Organic semiconductors . . . . .	2
1.1.1	Organic Solar cells . . . . .	2
1.1.2	OLEDs . . . . .	2
1.2	Predicting properties . . . . .	3
<b>2</b>	<b>Theory</b>	<b>4</b>
2.1	Many-body Hamiltonian . . . . .	4
2.2	Hartree Fock (HF) . . . . .	5
2.3	Density functional theory (DFT) . . . . .	6
2.3.1	Kohn-Sham (KS) . . . . .	6
2.4	Numerical solutions of HF and KS . . . . .	6
2.5	Electron affinity (EA) and ionization potential (IP) . . . . .	7
2.5.1	Determining the EA and IP . . . . .	7
2.6	Embedding . . . . .	8
2.6.1	Embedding on crystals . . . . .	8
2.7	Population analysis . . . . .	9
2.8	Atomic polarizabilities . . . . .	9
<b>3</b>	<b>Method</b>	<b>10</b>
3.1	DFT calculation . . . . .	10
3.2	Embedding . . . . .	10
<b>4</b>	<b>Results and discussion</b>	<b>13</b>
4.1	Parameter dependency . . . . .	13
4.1.1	Basis set dependency population analysis . . . . .	13
4.1.2	Cutoff radius embedding . . . . .	15
4.2	Future steps . . . . .	17
<b>5</b>	<b>Conclusion</b>	<b>20</b>
5.1	Impact parameters . . . . .	20
5.2	Future steps . . . . .	20

# Chapter 1

## Introduction

The interaction between light and matter plays an essential role in various applications. Upon gaining a fundamental understanding of these interactions, a variety of materials can be designed which optimize properties for different applications. An example of the materials of interest is the group of organic semiconductors.

### 1.1 Organic semiconductors

Semiconductors are materials with an electric conductivity between metals and insulators. The band gaps of these materials are generally below 3 eV [1], which makes them attractive in the interaction with visible light, as these electromagnetic waves have an energy corresponding to the band gaps of these materials. Possible applications for these semiconductors are organic solar cells or organic light-emitting diodes (OLEDs).

#### 1.1.1 Organic Solar cells

The basic principle of solar cells is the use of energy from incoming light for charge separation within the solar cell. This process is schematically drawn in Figure 1.1. The photons excite electrons from the HOMO (Highest Occupied Molecular Orbital) level to the LUMO (Lowest Unoccupied Molecular Orbital) level, resulting in a vacant spot, a hole. This leads to an electron-hole pair, a so-called exciton. The exciton is held together through the Coulomb interaction, the charge needs to be separated for the electrons and holes to diffuse to the electrical contacts. In inorganic solar cells, this Coulomb interaction within the exciton is typically on the order of magnitude of the thermal energy, such that the charge separation can happen spontaneously. For inorganic solar cells, this Coulomb interaction is stronger, typically in the order of 1 eV, which makes the charge separation challenging.

Currently, mainly inorganic solar cells are used for large-scale applications. If more complex chemical compounds are used, the properties of the materials can be better adjusted to meet different requirements. Organic solar cells can have properties like printability, flexibility, or even semi-transparency [3]. This will allow for various applications. The disadvantage of these organic compounds is the increased complexity. Elaborate research is needed to understand the properties of the materials.

#### 1.1.2 OLEDs

From a general point of view, the working of a LED is the reverse of the working of a solar cell. Instead of using light to enhance a current, we are using a current to generate light. The energy gap between the HOMO level and LUMO level influences the wavelength of the light that is being created. OLEDs are used in displays and offer advantages when it comes to flexibility and being able to adjust the wavelength of the light. Again, the main disadvantage of OLEDs is the increased complexity of their

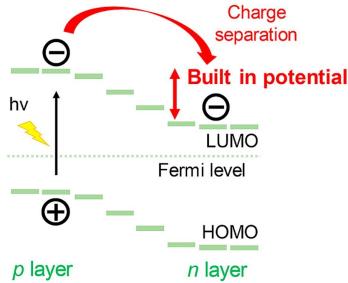


Figure 1.1: The charge separation in an organic solar cell [2]

materials. The change of one compound can influence the band gap and thus the light being produced [1]. Gaining an understanding of the electronic properties of these materials can help in predicting the band gap.

## 1.2 Predicting properties

Instead of determining the band gap of different materials through experiments and direct measurements, simulations to predict the properties of such complexes can be used. The properties of interest are determined at the subatomic scale and can be predicted through simulations of the interactions at this scale. This report strives to make a step forward toward an automated method of predicting those properties by running several calculations. The workflow can be used to select materials based on their desired properties quickly. The knowledge gained through the mathematical models will help in the understanding and improvement of the compounds for these applications. This saves a lot of lab work, as the first step in the selection of materials can be done by means of a calculation.

For this workflow, a basic theoretical understanding of subatomic interactions is required. The required methods are explained in Chapter 2. How this theory is implemented in the workflow is described in Chapter 3. A discussion on intermediate and end results can be found in Chapter 4 and the conclusion and further steps are presented in Chapter 5.

# Chapter 2

## Theory

To gain a fundamental understanding of the interaction between light and matter, the electronic structure of materials plays a crucial role. In general, this structure is described by many-electron wave functions, as they follow from solutions to the Schrödinger equation [4]:

$$\hat{H}\Psi = i\frac{\partial}{\partial t}\Psi. \quad (2.1)$$

For simplicity, Hartree atomic units are used throughout this report. In equation (2.1), the wave function  $\Psi$  is a function of time and the positions of electrons and nuclei in the material, and  $\hat{H}$  is the Hamiltonian operator. The Hamiltonian can be described as a combination of the kinetic energy and the potential energy of the particles. Unfortunately, the Schrödinger equation has only been solved exactly for systems of two particles, for example, the hydrogen atom [5]. To be able to describe systems of more than two interacting particles, an approximation of the many-body Schrödinger equation is needed. This chapter describes several theoretical methods that are used to be able to describe the electronic structure of materials. Firstly, it is described how an approximation of the many-electron wave function can be made using Slater determinants. Secondly, approximations based on a total electron density function are described. Finally, the implementation of previously described methods in combination with classical methods, so-called embedding, is covered.

### 2.1 Many-body Hamiltonian

We first look a bit closer at the Hamiltonian operator. For a many-body system, the Hamiltonian can be constructed out of the basic operators for kinetic ( $\hat{T}$ ) and potential ( $\hat{V}$ ) energy of charged particles with charge  $q$  and mass  $M$

$$\hat{T} = -\frac{1}{2} \sum_n \frac{1}{M_n} \Delta_{r_n} \quad (2.2)$$

$$\hat{V} = \frac{1}{2} \sum_{\alpha \neq \beta} \frac{q_\alpha q_\beta}{|r_\alpha - r_\beta|} \quad (2.3)$$

When we combine these two operators on a system consisting of  $M$  nuclei with charge  $Z_\alpha$  at position  $R_\alpha$ , ( $\alpha \in \{1, \dots, M\}$ ), and  $N$  electrons at positions  $r_i$ , ( $i \in \{1, \dots, N\}$ ), we end up with a non-relativistic Hamiltonian taking into account the nucleus-nucleus (nuc-nuc), nucleus-electron (nuc-el) and electron-electron (el-el) interactions:

$$\hat{H} = \hat{T}_{\text{nuc}} + \hat{T}_{\text{el}} + \hat{V}_{\text{nuc-nuc}} + \hat{V}_{\text{nuc-el}} + \hat{V}_{\text{el-el}} \quad (2.4)$$

$$\hat{H} = -\frac{1}{2} \sum_{\alpha=1}^M \frac{1}{M_\alpha} \Delta_{R_\alpha} - \frac{1}{2} \sum_{i=1}^N \Delta_{r_i} + \frac{1}{2} \sum_{\alpha \neq \beta} \frac{Z_\alpha Z_\beta}{|R_\alpha - R_\beta|} - \frac{1}{2} \sum_{\alpha=1}^M \sum_{i=1}^N \frac{Z_\alpha}{|R_\alpha - r_i|} + \frac{1}{2} \sum_{i \neq j} \frac{1}{|r_i - r_j|} \quad (2.5)$$

This Hamiltonian operator can be substituted into the Schrödinger equation (2.1), but as mentioned before, this has only been solved for systems with two particles. The standard method of solving such problems is by the use of separation of variables. We do this both with regard to the time dependency and to the motion of electrons and nuclei. The first makes the implicit assumption that the wave function  $\Psi$  can be written as a product of a spatial part  $\Phi$  and a time-dependent part  $U$ . If the Hamiltonian is not time-dependent, the time-dependent part will yield the expectation value of the Hamiltonian, say  $E$ . The spatial parts yield the time-independent Schrödinger equation,

$$\hat{H}\Phi = E\Phi. \quad (2.6)$$

The second separation rests on the Born-Oppenheimer approximation [6], which assumes that the electronic and nuclear motion can be decoupled. As the nuclei are much heavier than the electrons, we also assume that the nuclear motion is negligible. With this assumption, the term for the kinetic energy of the nuclei ( $T_{\text{nuc}}$ ) in the Hamiltonian (equation (2.5)) will be zero. The problem can be tackled in multiple ways, building from this simplified Hamiltonian with the time-independent Schrödinger equation.

## 2.2 Hartree Fock (HF)

As the electrons are fermions, the Pauli exclusion principle must be taken into account. This anti-symmetry is taken as a starting point for the Hartree-Fock (HF) principle, by using an estimated wave function ( $\Phi$ ) in the form of a determinant of single-electron wave functions ( $\phi_i$ ). By choosing the form of a determinant, the anti-symmetry requirement, namely the change of sign upon interchanging two electrons, is fulfilled. The determinant takes the following form:

$$\Phi(\vec{r}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(r_1) & \dots & \phi_1(r_N) \\ \vdots & \ddots & \vdots \\ \phi_N(r_1) & \dots & \phi_N(r_N) \end{vmatrix}. \quad (2.7)$$

This estimated wave function is then substituted into the time-independent Schrödinger equation (equation (2.6)) with the many-electron Hamiltonian. The variational theorem states that the energy will be minimal for the ground state wave function. Upon minimizing the energy under the constraint that the individual single-electron wave functions are normalized, a set of equations for the functions  $\phi_i$  is found, the so-called Hartree-Fock equations:

$$\left( -\frac{\Delta_r}{2} + \hat{V}_{\text{ext}}(r) + V_{\text{H}}(r) - \hat{V}_{\text{X}}(r) \right) \phi_i^{\text{HF}}(r) = \epsilon_i^{\text{HF}} \phi_i^{\text{HF}}(r), \quad (2.8)$$

where  $V_{\text{H}} = \int \frac{n(r)}{|r-r'|} dr'$  describes the Coulomb interaction between the electrons and  $\hat{V}_{\text{X}} = \int \frac{n(r,r')}{|r-r'|} \dots dr'$  the exchange potential operator. By this approach, the N-electron problem is reduced to a set of N single-electron problems. An estimation of the ground-state energy is given by

$$E_0^{\text{HF}} = \sum_{i=1}^N \epsilon_i^{\text{HF}} - \frac{1}{2}(E_{\text{H}} + E_{\text{x}}), \quad (2.9)$$

where  $E_{\text{H}}$  corresponds to the Coulomb interaction and  $E_{\text{x}}$  to the exchange energy. The difference between this estimation given by Hartree-Fock and the true (usually unknown) ground state energy is called correlation energy.

## 2.3 Density functional theory (DFT)

Another approach to tackle the many-electron Schrödinger equation is Density Functional Theory (DFT). The electron density is used instead of starting from an (estimated) many-electron wave function. So instead of substituting a trial wave function and minimizing the energy, the energy is minimized through a variational principle regarding the total electron density  $n(r)$ , with the constraint that  $\int n(r)d^3r = N$ . This method is allowed by the Hohenberg-Kohn theorem which states that the ground state of any interacting many-particle system with a given fixed inter-particle interaction is a unique functional of the electron density  $n(r)$  [7]. The energy functional

$$E[n] = \int V_{\text{ext}}(r)n(r)dr + \hat{T} + \hat{V}_{\text{el-el}} \quad (2.10)$$

is minimal for the ground state density  $n_0(r)$ . This is a theoretically exact relation.

### 2.3.1 Kohn-Sham (KS)

A problem occurs when one tries to minimize the energy, as the kinetic energy and the electron-electron interaction are not described in terms of the electron density. This can be solved by an approximation proposed by Kohn and Sham [8]. The kinetic energy is first described for a system of non-interacting electrons, as a function of the electron density. The electron-electron interaction is approximated by the Coulomb interaction, the Hartree energy. These equations are implemented in the energy functional that needs to be minimized. To correct for the fact that the kinetic energy is taken from a system of non-interacting electrons and that the electron-electron interaction is approximated by the Hartree energy, an additional term, the exchange-correlation functional is added. The exchange-correlation functional corrects for the errors that were made in these approximations. With this, the Kohn-Sham (KS) energy functional is

$$E^{\text{KS}} = T_s + \int V_{\text{ext}}(r)n(r)dr + E_H + E_{\text{xc}}, \quad (2.11)$$

with  $T_s$  the kinetic energy for a system of non-interacting electrons,  $E_H$  the Hartree energy, and  $E_{\text{xc}}$  the exchange-correlation functional. Minimization of equation (2.11) yields the KS-equations:

$$\left( -\frac{\Delta_r}{2} + \hat{V}_{\text{ext}}(r) + \hat{V}_H(r) + \frac{\delta E_{\text{xc}}}{\delta n}(r) \right) \phi_i^{\text{KS}}(r) = \epsilon_i^{\text{KS}} \phi_i^{\text{KS}}(r). \quad (2.12)$$

In principle, these equations are exact, though there is no practical implementation of finding the exchange-correlation functional [9].

## 2.4 Numerical solutions of HF and KS

The eigenvalue equations resulting from the Hartree-Fock or Kohn-Sham equations are in general solved numerically by substituting functions of the form

$$\phi_i = \sum_{\alpha} c_{\alpha} \chi_{\alpha}, \quad (2.13)$$

where the  $\chi_{\alpha}$  are functions from a basis set. The  $\phi_i$  are linear combinations of functions from the basis set. Substituting these functions in the HF or KS equations and integrating both sides yields

$$\int \chi_{\beta}(\underline{r}) H \sum_{\alpha} c_{\alpha}^i \chi_{\alpha}(\underline{r}) d^3r = \int \chi_{\beta}(\underline{r}) \sum_{\alpha} c_{\alpha}^i \chi_{\alpha}(\underline{r}) d^3r. \quad (2.14)$$

By taking the finite sum out of the integral, the equation simplifies to

$$\sum_{\alpha} c_{\alpha}^i \int \chi_{\beta}(\underline{r}) H \chi_{\alpha}(\underline{r}) d^3r = \epsilon_i \sum_{\alpha} c_{\alpha}^i \int \chi_{\beta}(\underline{r}) \chi_{\alpha}(\underline{r}) d^3r. \quad (2.15)$$

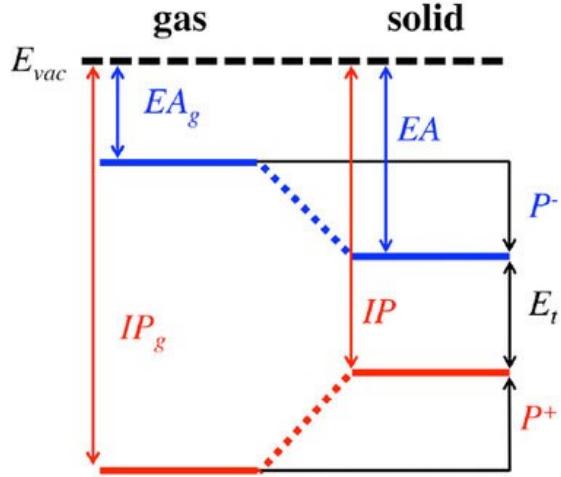


Figure 2.1: An energy-level diagram for a generic material, indicating the electron affinity (EA) and the ionization potential (IP) [10]

This leads to a Hamiltonian tensor ( $\underline{\underline{H}}$ ) and an overlap tensor ( $\underline{\underline{S}}$ )

$$\underline{\underline{H}}c^i = \epsilon_i \underline{\underline{S}}c^i, \quad (2.16)$$

and the eigenvalue problem can be solved for the  $c^i$ , giving the functions  $\phi_i$ .

## 2.5 Electron affinity (EA) and ionization potential (IP)

The electron affinity (EA) is a measure of the energy increase upon adding an electron to an isolated molecule, whereas the ionization potential (IP) is a measure of the energy it takes to remove an electron from such molecule. Here, the electron should completely be taken up into the system or completely removed. The electron affinity has a correlation to how easily the molecule reacts in a chemical reaction, as this generally entails the exchange of an electron. In Figure 2.1, an energy diagram indicating the EA and the IP is shown.

### 2.5.1 Determining the EA and IP

Through the calculations, the EA and the IP can be determined in different manners. As suggested in Figure 2.1, the EA can be found by taking minus the energy from the LUMO, whereas the IP can be found from minus the energy from the HOMO. When this is done on a system of  $C_2H_2$ , this gives an IP of 8.54 eV (as the  $\epsilon_{HOMO} = -8.54$  eV) and the EA gives 1.02 eV (as  $\epsilon_{LUMO} = -1.02$  eV). However, these results do not seem to match the results from experiments. From experiments, the IP for  $C_2H_2$  is found to be 11.4 eV [11]. Another way to determine the IP and EA from the DFT calculation is by taking the difference of the Self-Consistent Field energy,  $\Delta SCF$  for the states with charge 0, and  $\pm 1$ . In the case of the IP, the energy it takes to remove an electron from the system means taking the difference between the -1 and the neutral state ( $E_{-1} - E_0$ ). For the EA, this is calculated as the difference between neutral and +1 state ( $E_0 - E_{+1}$ ). This gives for  $C_2H_2$  the values EA = -3.25 eV and IP = 11.44 eV. As the IP correlates better with the results from experiments, this method is used in the remainder of the report. The IP and EA are determined through the DFT calculation for an isolated molecule. These values are used as a constant offset, where the contributions of the surroundings are calculated in a similar manner.

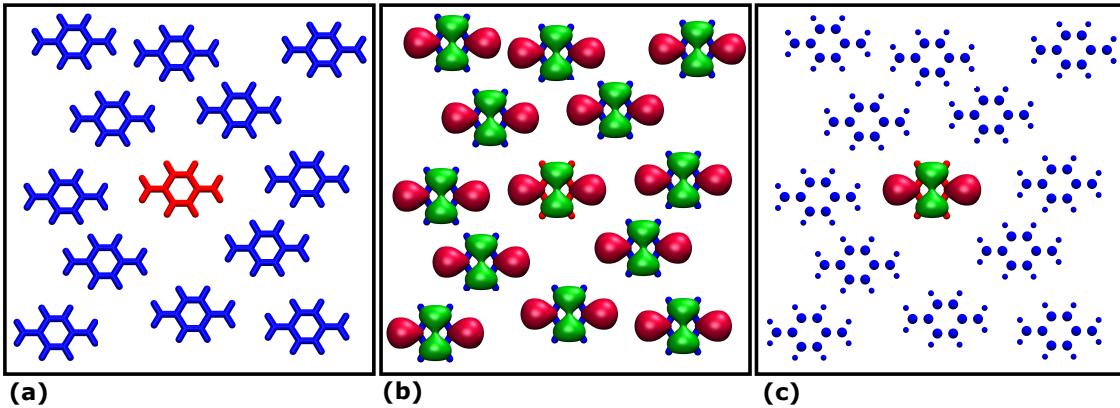


Figure 2.2: A schematic representation of the embedding process. (a) The region of interest is indicated in red, (b) considering the entire region with quantum techniques (c) only considering the region of interest with quantum techniques, whilst considering the surroundings as polarized molecules, inducing an external potential.

## 2.6 Embedding

Whenever the size of the system grows, it becomes increasingly more computationally expensive to treat the system with the methods that were described in this chapter. To tackle this problem, a general method is to only consider a smaller region of the material using the methods as described, containing the quantum mechanical contributions. This is schematically drawn in Figure 2.2. The size of the region of interest (as indicated by (a)) is determined by setting a cutoff radius.

The surroundings can be treated as a classical system of charges or polarized molecules, more on this in section 2.8. The influence of the surrounding on the region of interest is being implemented as an external potential. The region of interest changes according to this potential and influences on its turn the surrounding. The mutual influences are calculated until an equilibrium is reached [9].

### 2.6.1 Embedding on crystals

The method changes slightly whenever the system is highly structured as a crystal. A crystal consists of repeating unit cells on a lattice. This is translational invariant, as long as one translates over an integer multiplication of the lattice vectors [12]. Because of this translational invariance, the cutoff radius that can practically be used for the embedding will be bounded by the length of the lattice vectors. If this cutoff radius is bigger than half of the length of the lattice vectors (in the case of a cubic lattice where all vectors have the same length and are orthogonal), then also the interactions with translated atoms will be taken into account, instead of just the closest neighbor. The total Coulomb energy of a crystal leads to the Ewald sum

$$E_{\text{coulomb}} = \frac{1}{2} \sum_{j=1}^N q_j \Phi(x_j), \quad (2.17)$$

which is only conditionally convergent. Ewald came up with a method to rewrite the Ewald sum into two convergent series. One series accounting for the short-range (described in real space) and one for the long-range (described in reciprocal space) contribution to the Coulomb energy. The reciprocal space is a Fourier-transformed space of the original (real space) lattice, as the long-range contribution has a high periodicity. This leads to

$$\Phi(x_j) = \Phi_{\text{real}}(x_j) + \Phi_{\text{reciprocal}}(x_j), \quad (2.18)$$

which is the sum of two convergent series.

## 2.7 Population analysis

Once the electron density is determined using a DFT calculation, population analysis can be done to determine the individual atomic charges. This can be done through different methods. This section will describe three common methods: CHELPG, Löwdin, and Mulliken. The first method is an example of potential derived charges, whereas Löwdin and Mulliken are both orbital-based methods. CHELPG (Charges from Electrostatic Potential using a Grid) uses a partitioning of the electrostatic potential as derived from the DFT calculation [13]. It uses a cubic grid to determine the charge at several points. This method has a smaller basis set dependency as compared to the other two orbital-based methods [14]. The Mulliken method uses a molecular orbital function as a linear combination of the separate atomic orbitals [15]. The charge  $q$  on an atom  $a$ ,  $q_a$  is determined by taking the atomic number  $Z_a$  and subtracting the number of electrons allocated to the atomic orbitals on the atom,

$$q_a = Z_a - \sum_{\mu\nu \in a} P_{\mu\nu} S_{\mu\nu}, \quad (2.19)$$

where  $S_{\mu\nu}$  is the overlap matrix of the separate atomic orbitals. The basic functions are based on an orbital partitioning scheme. The Mulliken method is simple, though the method is dependent on the set of basic functions used in the Mulliken calculation. The Löwdin method is an improved version of the Mulliken method. This method uses a different set of basic functions, symmetrically orthogonalized atomic orbitals [16].

## 2.8 Atomic polarizabilities

Once obtaining the charges of the individual atoms, a potential can be calculated from this. The potential at a point  $R$  from a system of particles at positions  $r_i$  is given by

$$V(R) = \sum_i \frac{1}{|R - r_i|}. \quad (2.20)$$

This potential can be expanded using a Taylor series about the center of the system. This expansion leads to a multipole expansion, which contains the electrostatic, dipole, quadrupole, and higher moments of interaction. The geometry is determined by these interactions, where the electrostatic moment is the most dominant term [17].

# Chapter 3

## Method

The previous chapter gave a brief overview of the theory that is being implemented by the programs that were used for this project. The project used the implementations from VOTCA-XTP [18] and Orca [19] [20]. In this chapter, it is described how everything comes together in a workflow. The workflow is split into two parts, which will be explained separately.

### 3.1 DFT calculation

The first step of the workflow is to execute the DFT calculation on the system. This is schematically drawn in Figure 3.1. The DFT calculation is done for three different states: neutral, +1, and -1. The DFT calculation is done by the implementation of Orca [19] [20]. This is done through the VOTCA framework using

```
xtp_tools -e dftgwbse -o OPTIONFILE.xml
```

where the options file specified to use the implementation of Orca. The input for this calculation is an XYZ file of the structure, which contains information on the atoms and their coordinates. The spin is set to 1 for the neutral calculation and to 2 for the  $\pm 1$  calculation. For all calculations, the used basis set is tzvp.

The next step is a conversion step, where the three generated log files are converted to MPS files.

```
xtp_tools -e log2mps -o OPTIONFILE.xml
```

These files contain information on the electronic structure of the system. This step saves the information from the log file in a different format, an MPS file, such that the calculator for the molecular polarizabilities can update them.

```
xtp_tools -e molpol -o OPTIONFILE.xml
```

This calculator uses the generated MPS files and the log file to calculate the molecular polarizabilities for every state. In the input MPS file, the polarizability is taken from a database, whilst in this step the polarizability is calculated specifically for the system. The output from this calculation is an MPS file with an updated polarizability tensor.

### 3.2 Embedding

The second part of the workflow consists of the calculation on the embedding and implementing Ewalds method. The first step is only done for the smaller compounds. The next step uses a crystal structure. It starts with preparing the input data. The data is an open-source database with crystal structures. The data is a crystal structure, of which only the coordinates and atoms within the primitive unit cell

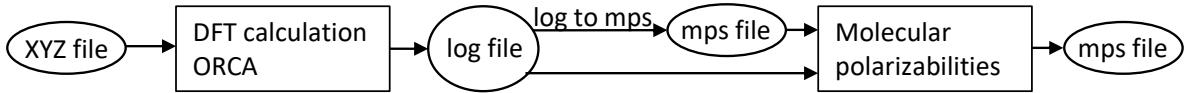


Figure 3.1: Schematic workflow of the DFT and related calculations. The calculations are indicated with rectangles and the files with ellipses.

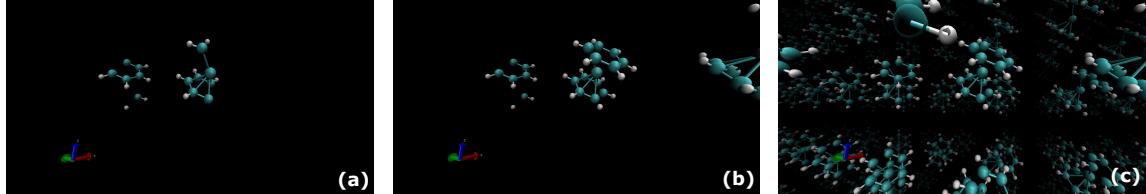


Figure 3.2: A visualization of a crystal structure, all made using VMD [21]. (a) One unit cell, where the atoms in the molecule are dislocated. (b) Two unit cells, where the molecular structure becomes visible. (c) Several copies of the unit cell giving an impression of the crystal structure

are saved. This may lead to molecules being separated over different unit cells, causing problems in the initial DFT calculation, as the equivalent images of atoms are not taken into account. In Figure 3.2, an example of such a structure is visible. The original unit cell contains four such molecules, of which one is visualized in Figure 3.2. To work around this problem, the data set is cleaned manually. The coordinates are translated such that the images of atoms corresponding to the same molecule have minimal distance.

The next step is to create the topology for the embedding. A mapping is made to match the coordinates in the supercell to the XYZ file and the different MPS files using

```
xtp_map -c supercell.gro -t topol.tpr -f state.hdf5 -s map.xml
```

The topology needed for Ewald's method is created. This mapping is checked, with

```
xtp_run -e mapchecker -o mapchecker.xml -f state.hdf5
```

resulting in separate PDB files for every segment in the molecule. Once the topology is in place, the polarization of the periodic background is calculated.

```
xtp_run -e ewdbgpol
```

The next step is to run the calculation for the embedding, where the results from the DFT calculation are combined with the background polarization. To do this, several jobs need to be defined. This is done using

```
xtp_parallel -e ewald -o ewald.xml -f state.hdf5 -j write
```

In general, every job covers a different state a molecule is in. For the three states that were calculated for the DFT calculation, a separate job is created. If the unit cell consists of multiple molecules, they will all have three jobs. The jobs are executed, which consist of calculating the mutual interactions between the molecule in a particular state with the previously defined background.

```
xtp_parallel -e ewald -o ewald.xml -f state.hdf5 -j run
```

This leads to the total energy of the system.

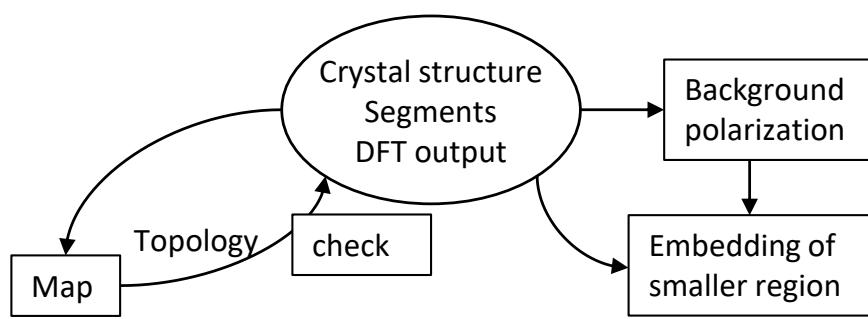


Figure 3.3: Schematic workflow of the embedding step in the workflow

# Chapter 4

## Results and discussion

The first part of the project consisted of getting acquainted with the code, by executing the different parts of the workflow separately and testing for parameter dependency. Some of these results are discussed in this chapter. Also, some results on the entire workflow and the future steps are discussed.

### 4.1 Parameter dependency

#### 4.1.1 Basis set dependency population analysis

The first DFT calculations were done on a series of isolated alkanes. The number of C atoms was increased up to 6, hexane ( $C_6H_{14}$ ). In this setting, the three different population analysis methods, as explained in section 2.7, were executed. This was done for different basis sets used in the initial DFT calculation. The three basis sets that are used are split valence polarization (svp), triple zeta valence potential (tzvp), and quadruple zeta valence potential (qzvp), in increasing order of functions. Firstly, a DFT calculation was done with the implementation from VOTCA-XTP, then the population analysis was done using the command

```
xtp_tools -e partialcharges -o OPTIONFILE.xml
```

where the option of the population analysis method was changed. This calculation outputs an MPS file, which was used to collect the partial charge on each atom in the molecule. The partial charges of the 'central' C atom are plotted against the different basis sets and the population analysis methods, see Figure 4.1. For the molecules with an even number of C atoms, the 'central' C was one of the two central C atoms. As the molecules are symmetric, the differences between both central atoms are not significant.

From Figure 4.1 it can be seen that the calculation using Löwdin with the qzvp basis set, gave interesting results. Where most partial charges for the central C atoms are negative (as to be expected, as the nuclei of C atoms are heavier than the nuclei of the H atoms), this combination consistently gives positive partial charges. To ensure that this behavior is not caused by the DFT implementation of VOTCA-XTP, the entire process was repeated using the implementation of Orca. This gave the exact same results. The calculations with the CHELPG method gave comparable results for the different basis sets used on the same system, as to be expected from the theory, see section 2.7. The method seems to be less dependent on the used basis set as compared to Mulliken and Löwdin. The CHELPG calculation on Butane, Figure 4.1(d), assigns a positive partial charge to the central C atoms. The same result was observed for the second and fifth C atoms for hexane, see Figure 4.2. Mulliken and Löwdin's calculation using the tzvp basis set gives the lowest partial charge.

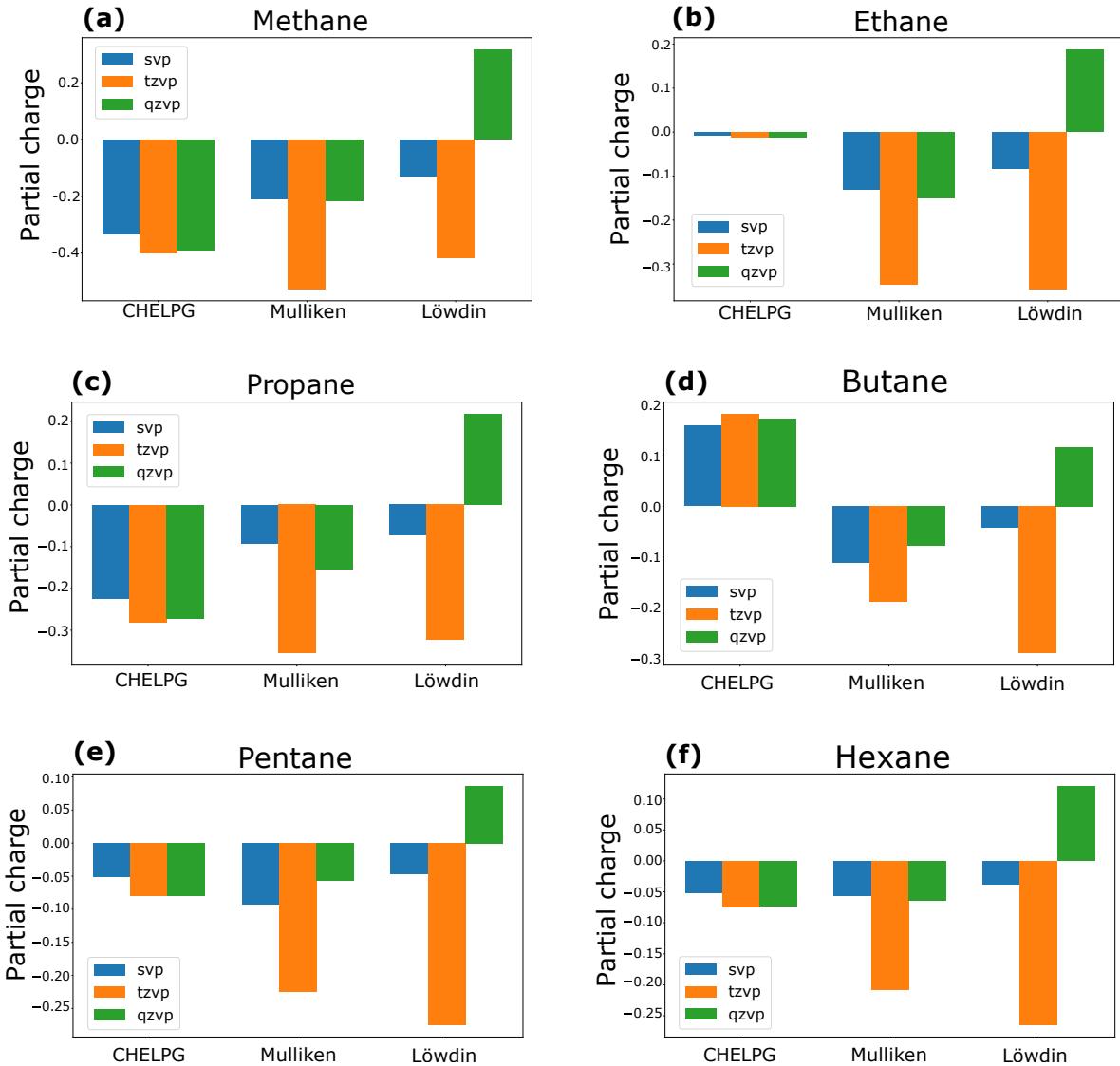


Figure 4.1: The partial charge of the 'central' C atom for different systems: (a) methane, (b) ethane, (c) propane, (d) butane, (e) pentane and (f) hexane. The calculations were repeated for three different basis sets (svp, tzvp and qzvp) and three different population analysis methods (CHELPG, Mulliken and Löwdin)

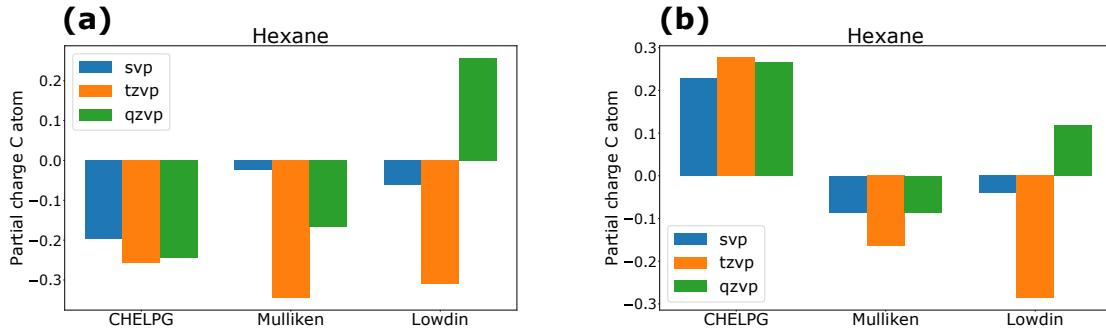


Figure 4.2: The partial charge of C atoms of hexane, again for three different basis sets (svp, tzvp, qzvp) and population analysis methods (CHELPG, Mulliken, Löwdin). (a) The first C atom, (b) the second atom. The third atom was considered as the 'central' C atom, so visible in figure 4.1 (f)). The other three C atoms were not plotted, as these are comparable to the first three, due to the symmetry of the molecule.

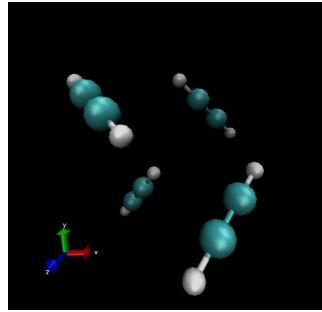


Figure 4.3: One unit cell of the used system of  $\text{C}_2\text{H}_2$  [21]

#### 4.1.2 Cutoff radius embedding

The total workflow was tested on a different system; a crystalline system of  $\text{C}_2\text{H}_2$ . The (linear) molecules are aligned as shown in Figure 4.3. Every unit cell consists of four such molecules with a different alignment. The input data was cleaned manually, such that the problem as described in Chapter 3 is taken care of. The dependency on the polar cutoff radius was examined, by varying this value from 1 nm to 3 nm. The EA and IP were plotted as a function of the polar cutoff radius, see Figures 4.4 and 4.5.

For every cutoff radius, five data points are available; one for every molecule and an additional molecule; a translation of the molecule located at the origin. This additional molecule is taken into account because the data for the molecule at the origin deviated from the other molecules, whilst the system seems to be highly symmetrical, and thus is expected to yield similar results. The different data points are best visible in the plots for the static moments. Here, the highest values for every cutoff radius correspond to the second, third, and fourth molecule, the lowest values correspond to the molecule located at the origin. The middle values correspond to a translated image of the molecule at the origin. In theory, the values of the molecule at the origin, and those for the translated image should be the same, as they represent the same molecule. However, this is not the case, which suggests there might be problems with molecules being centered at the origin. Further research on other systems is needed to test whether this is a consistent problem.

As the cutoff radius increases, the EA and IP seem to stabilize. The contribution to the EA and IP from a static charged and an induced charged surrounding can be considered separately, see Figures 4.4

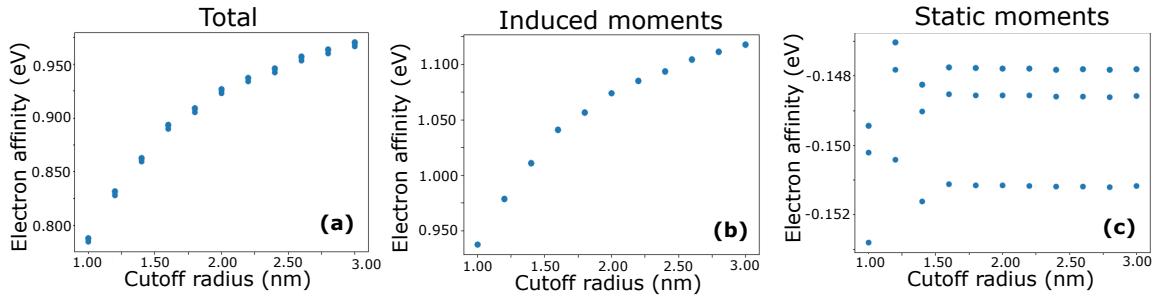


Figure 4.4: The EA as a function of the polar cutoff radius. (a) the total contribution from the embedding, (b) the contribution from the induced moments, (c) the static contributions

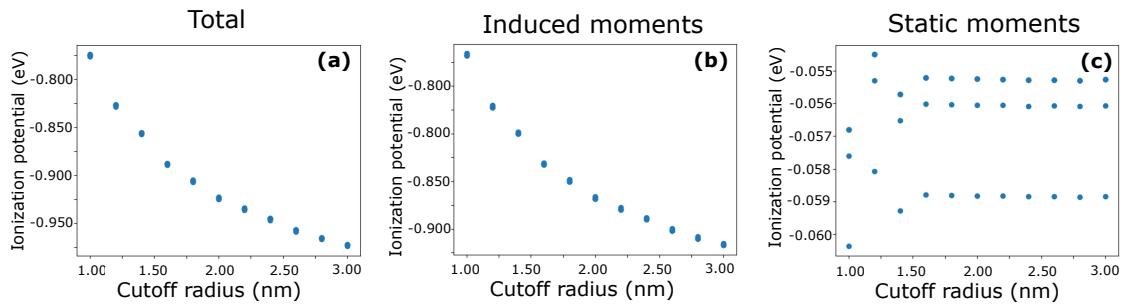


Figure 4.5: The IP as a function of the polar cutoff radius. (a) the total contribution from the embedding, (b) the contribution from the induced moments, (c) the static contributions

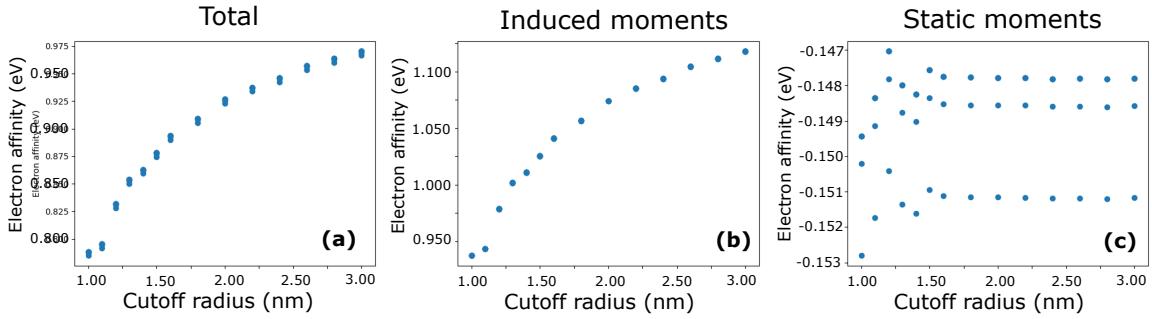


Figure 4.6: The EA as a function of the polar cutoff radius, with additional points at smaller cutoff radii to better examine the behavior for small cutoff radii. (a) the total contribution from the embedding, (b) the contribution from the induced moments, (c) the static contributions

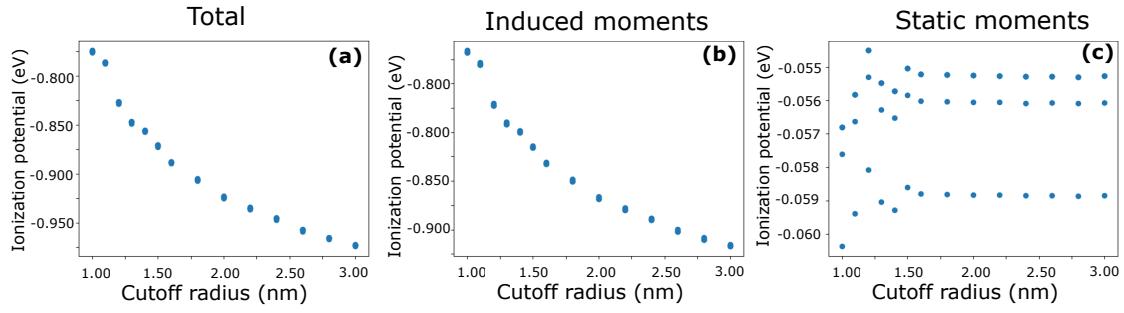


Figure 4.7: The IP as a function of the polar cutoff radius, with additional points at smaller cutoff radii to better examine the behavior for small cutoff radii. (a) the total contribution from the embedding, (b) the contribution from the induced moments, (c) the static contributions

(b), (c) and 4.5 (b), (c). The addition of these values yields the total contribution of the surroundings to the AE and IP, Figures 4.4 (a) and 4.5 (a). Especially on the static contribution, the behavior on small cutoff radii seems troublesome. To better evaluate this behavior, additional points were added to the calculation, see Figures 4.6 and 4.7.

This showed that the behavior at low cutoff radii is confusing. It was noted that in the calculations, the option to apply coarse-graining on the background was turned on. This means that the individual moments of the atoms on the background are first converted to averaged moments on the molecules. Especially for smaller cutoff radii, this may cause troublesome deviations. This option is relevant to decrease the computational time for systems with large molecules, but for the (relatively small) system of  $C_2H_2$ , this option is not necessary. In Figures 4.8 and 4.9 the same plots are made for when the option course-graining on the background is turned off. Here, the behavior at small cutoff radii is more in line with the overall behavior.

## 4.2 Future steps

These results were obtained by using Python code, which automated the commands to give to the VOTCA-XTP calculator. Several attempts were made to set up an algorithm to clean the input data such that the initial DFT calculations can use it. There should be an automated method to read the

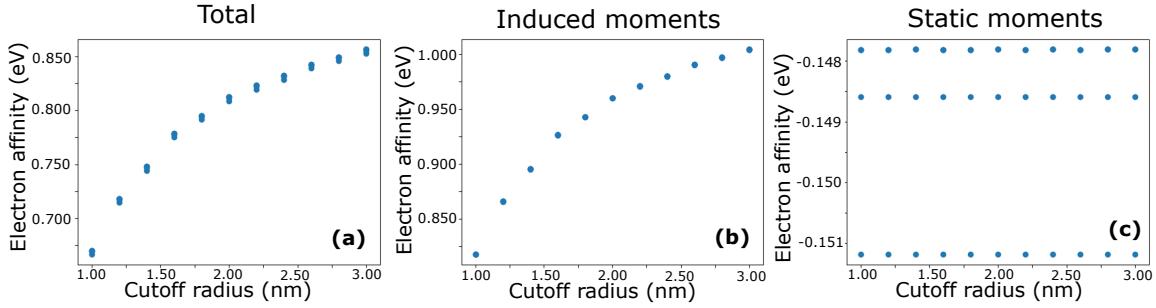


Figure 4.8: The EA as a function of the polar cutoff radius, without a coarse graining applied on the background (a) the total contribution from the embedding, (b) the contribution from the induced moments, (c) the static contributions

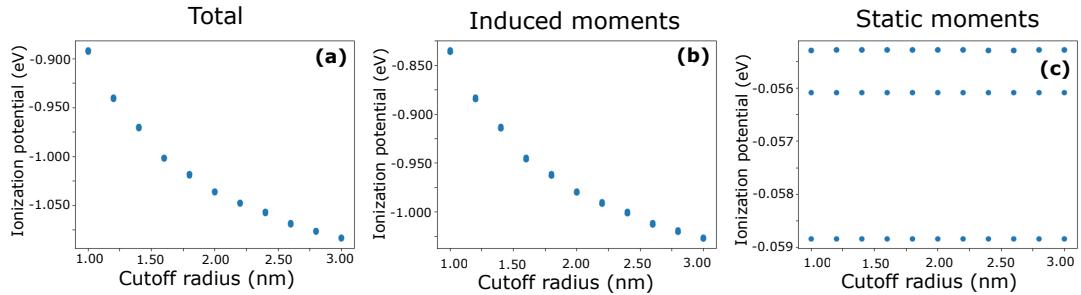


Figure 4.9: The IP as a function of the polar cutoff radius, without a coarse graining applied on the background (a) the total contribution from the embedding, (b) the contribution from the induced moments, (c) the static contributions

file containing the information on the crystal structure and the coordinates of the atoms within the unit cell, groups the individual atoms as molecules taking into account the periodic copies based on the crystal structure, and saves the coordinates in such way that the atoms within one molecule have a minimal mutual distance. This will prevent having to clean the input data manually, which is a tedious process when the molecule size increases. Such code is not in place yet but should be incorporated before the entire workflow can be applied to the database. Another improvement can be made upon realizing that the code overwrites a file if another file with the same name and directory already exists. To prevent files containing relevant data to be overwritten, several directories should be made. It is advisable to structure the documents from the beginning into different directories. Once a rigorous structure of directories is created, a more thorough examination of the parameter dependency on the code can be done.

# Chapter 5

# Conclusion

The goal of this project was to make a step towards an automated workflow for multiscale molecular modeling. In Chapter 3, the used workflow is explained. This workflow is automated using two Python codes, for the DFT and embedding calculation. These codes automate the commands that are given to the terminal of the VOTCA-XTP application. Through this workflow, the impact of the input parameters after the entire workflow could efficiently be determined.

## 5.1 Impact parameters

The impact of the basis set used and the population analysis on the individual atomic partial charges was determined. The DFT calculations on systems of isolated alkane molecules were done for different basis sets (svp, tzvp and qzvp). The results from these DFT calculations were examined using different population analysis methods (CHELPG, Mulliken, Löwdin). This yielded 9 different values for the partial charge for every atom. CHELPG showed the least basis set dependence, which is in line with the results from the theory. The combination of Löwdin with qzvp gave significantly different results as compared to the other values. More experiments are needed to explain this difference.

Another parameter dependency test was done on the embedding part of the code. The influence of the EA and IP on the polar cutoff radius was determined. This was done with coarse-graining of the background and without this option, thus taking the individual moments of the atoms on the background. Especially for the lower cutoff radii, the option without the coarse-graining of the background gave more reliable results. It is worth considering doing the calculations without this option, especially if one is using smaller cutoff radii. Of course, this only applies if this is reasonable for the size of the molecule. The influence of this option on larger systems is to be determined.

## 5.2 Future steps

As discussed in the previous chapter, several steps are necessary to improve the workflow and to adjust it such that it can be used more efficiently. More tests on the database, especially for systems of increasing molecule size, and the efficiency of the workflow should be tested. One of the first steps for this is by tackling the problem as discussed in Chapter 3, by cleaning the input from the database. Once an algorithm is in place to clean the input data, the workflow can be tested for systems with more atoms. The parameter dependency as done in this report will be interesting when the system size, thus the complexity, increases. After these necessary steps, the workflow can be used to do a more thorough determination of the parameter dependency. The results from this can be used to improve the workflow. Once the workflow has had more rounds of evaluation, the next step is to compare the outcomes of the code to the outcomes from experiments to determine the capability to predict

material properties. Once the correlation between the outcome of the code and the real-live properties is known, this workflow will be a tool for designing materials with desirable properties.

# Bibliography

- [1] José C.S. Costa, Ricardo J.S. Taveira, Carlos F.R.A.C. Lima, Adélio Mendes, and Luís M.N.B.F. Santos. Optical band gaps of organic semiconductor materials. *Optical Materials*, 58:51–60, 8 2016.
- [2] Seiichiro Izawa, Armand Perrot, Ji Hyun Lee, and Masahiro Hiramoto. Organic pn homojunction solar cell. *Organic Electronics*, 71:45–49, 8 2019.
- [3] Samuel E. Root, Suchol Savagatrup, Adam D. Printz, Daniel Rodriguez, and Darren J. Lipomi. Mechanical Properties of Organic Semiconductors for Stretchable, Highly Flexible, and Mechanically Robust Electronics, 5 2017.
- [4] E Schrödinger. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review*, 28(6):1049–1070, 12 1926.
- [5] David J Griffiths and Darrell F Schroeter. *Introduction to quantum mechanics*. Cambridge University Press,, Cambridge, United Kingdom :, 2018.
- [6] M. Born and R. Oppenheimer. Zur Quantentheorie der Moleküle. *Annalen der Physik*, 389(20):457–484, 1927.
- [7] P Hohenberg and W Kohn. Inhomogeneous Electron Gas. Technical report, Ecole Normale Supérieure, Paris, 6 1964.
- [8] W Kohn. Nobel Lecture: Electronic structure of matter-wave functions and density functionals\*. Technical report, University of California, Santa Barbara, 1998.
- [9] Gianluca Tirimbò and Björn Baumeier. Ab initio modeling of excitons: from perfect crystals to biomaterials. *Advances in Physics: X*, 6(1), 2021.
- [10] Gabriele D’Avino, Luca Muccioli, Frédéric Castet, Carl Poelking, Denis Andrienko, Zoltán G. Soos, Jérôme Cornil, and David Beljonne. Electrostatic phenomena in organic semiconductors: Fundamentals and implications for photovoltaics, 9 2016.
- [11] Pierre Plessis and Paul Marmet. Electrionization Study of Acetylene and Fragment Ions. Technical report, Institute of Astrophysics, National Research Council of Canada, Ottawa, 1986.
- [12] Charles Kittel. *Introduction to Solid State Physics*, volume 8. Wiley, 8 edition, 2005.
- [13] Thomas A. Manz. Seven confluence principles: a case study of standardized statistical analysis for 26 methods that assign net atomic charges in molecules. *RSC Advances*, 10(72):44121–44148, 11 2020.
- [14] F. Martin and H. Zipse. Charge distribution in the water molecule -A comparison of methods. *Journal of Computational Chemistry*, 26(1):97–105, 1 2005.

- [15] R. S. Mulliken. Electronic population analysis on LCAO-MO molecular wave functions. I. *The Journal of Chemical Physics*, 23(10):1833–1840, 1955.
- [16] Sasha C. North, Kameron R. Jorgensen, Jason Pricetolstoy, and Angela K. Wilson. Population analysis and the effects of Gaussian basis set quality and quantum mechanical approach: main group through heavy element species. *Frontiers in Chemistry*, 11, 2023.
- [17] Anthony Stone. *The Theory of Intermolecular Forces*. Oxford University Press, 2 edition, 2013.
- [18] Victor Rühle, Christoph Junghans, Alexander Lukyanov, Kurt Kremer, and Denis Andrienko. Versatile object-oriented toolkit for coarse-graining applications. *Journal of Chemical Theory and Computation*, 5(12):3211–3223, 12 2009.
- [19] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry*, 32(7):1456–1465, 5 2011.
- [20] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics*, 132(15), 4 2010.
- [21] W. Humphrey, A. Dalke, and K. Schulten. VMD-Visual Molecular Dynamics, 1996.