Benchmarking Basis Sets for Orbital Energies in Density Functional Theory: A Comparison of Plane Waves and Gaussians

Bachelor's thesis

by

Lukas Heimgaertner

May 8, 2025

Reviewer: Dr. Jan Wilhelm
Supervisors: Dr. Jan Wilhelm & Mia Schambeck

Contents

Chapter 1

Introduction

In the contemporary world, solar cells are of central importance to the supply of energy. As demand for energy continues to increase and the area suitable for photovoltaic installations is finite, improving solar cell efficiency is advantageous. Additionally, finding new cost-effective materials for solar cells is especially beneficial for less affluent regions, helping to promote the use of renewable energy in these areas.

A new class of low-cost solar cells based on halide perovskites disrupts this field. Metal-halide perovskites are a category of crystals which exhibit a perovskite structure. This structure can be best described with the chemical formula ABX_3. Where X represents halide anions (such as I^-, Br^-, Cl^-), B denotes a metal atom (e.g., Pb, Ti), and A refers to a molecular cation within the metallic framework. In Figure 1.1, the structure of a metal-halide perovskite is exemplified by $CH_3NH_3PbI_3$.
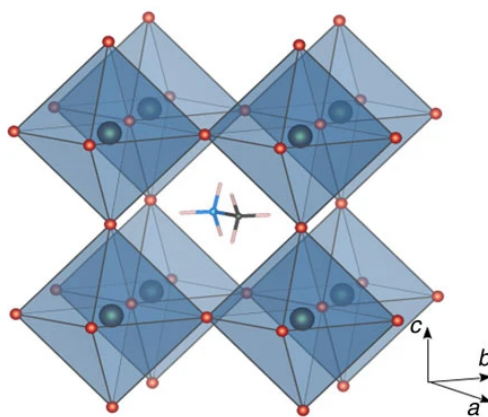
Figure 1.1: Ball-and-stick representation of MA–PbI3 in the high-temperature cubic phase: the dark spheres inside the octahedra represent Pb atoms, the red spheres at the octahedral corners are I atoms and the molecule inside the cuboctahedral cavity is CH3NH3. The plot and the description originate in [1].

Through the work of Shockley–Queisser [2], it is known that the maximum efficiency of solar cells is achieved for a band gap of 1.1 eV. Since perovskite solar cells (PSCs) are not only cost-

effective but also highly tunable, efforts are being made to adjust the band gap of PSCs to match this optimal 1.1 eV value. Filip et al. [1] used **D**ensity **F**unctional **T**eory (DFT) calculations to study the impact of the size of the molecular cation on the optical band gap. As seen in 1.2 **(a)** the bond angles of PbI$_3$ can be modulated, by inserting different sized molecules. After geometric optimizations, the optical band gap was calculated. Not only could the band gap be tuned by more than 1 eV, but a trend was also observed (c.f. Figure 1.1 **(c)**), which shows that larger ions result in smaller band gaps.
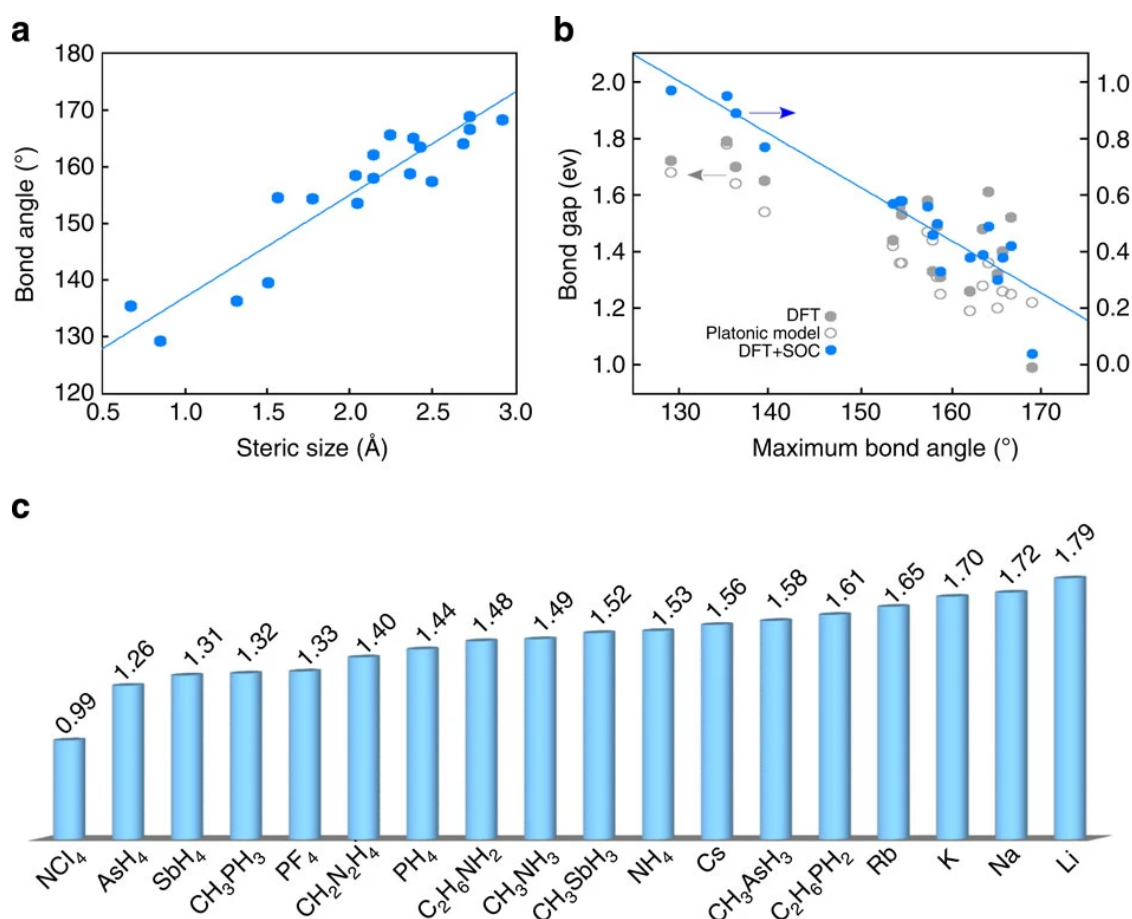
Figure 1.2: **(a)** Correlation between the bondangles of PbI3-based perovskites and the steric radius of the cation. As the size of the cation increases, the metal–halide–metal bonds tend to become collinear (180° angle). **(b)** Correlation between the DFT band gap and the largest metal–halide–metal bond angle in the structure. **(c)** Calculated band gaps (in eV) of all the PbI$_3$-based perovskites considered in [1]. The band gaps were obtained after full structural optimization within scalar relativistic DFT. The plots and parts of the description originate in [1].

Analogous to the band gap in semiconductors, molecules exhibit a HOMO-LUMO gap, which can also be calculated using DFT. Overall DFT is a widely utilized method for ab initio calculations of the structure of atoms, molecules, crystals, surfaces and their interactions [3]. It facilitates the determination of quantum mechanical properties, such as the ionization energy, based on the electronic density. A major advantage of DFT is that it does not necessitate the complete solution

of the Schrödinger equation for many-body systems, thereby significantly reducing the computing power required. The balance between efficiency and yet relatively high accuracy often makes DFT the method of choice, to accommodate experimental testing [4].

As DFT in practice employs a number of approximation and therefore the calculations are not exact, benchmarks are crucial for validating its accuracy and ensuring reliable results [5]. Motivated by the GW-100 benchmark, which compares different implementations of GW for a set of 100 molecules, we want to follow a similar approach for DFT calculations. The present study focuses in particular on the selection of basis set: Abinit uses a plane wave basis, whereas CP2K employs a Gaussians basis. This choice is motivated by the fact that plane waves are primarily employed for large periodic systems, whereas localized Gaussians are predominantly effective for atomic and molecular systems. For optimal converged calculations one would not expect that those two algorithms would differ in den results. Nevertheless, a considerable discrepancy in outcomes can be observed between different DFT codes. [5].

To further investigate this issue we repeat the central theorems behind DFT (c.f. 2), examine the different implementations of DFT in Abniti and CP2K (c.f. 2.6), and compare the two algorithms using the GW-100 molecule set. Specifically, we conduct several convergence tests to ensure that both algorithms converge within within themselves, as outlined in chapter 3. Once this is achieved we compare the results from Abinit and CP2K for the Highest Occupied Molecular Orbital (HOMO) (c.f. 4.1) and the Lowest Unoccupied Molecular Orbital (LUMO) (c.f. 4.2). Finally, in Chapter 5, we offer a conclusion to this thesis, including a summary of the results and suggestions for potential future investigations in this field.

Chapter 2

Theoretical foundations of DFT

The following chapter provides a brief introduction to the theoretical background of DFT. The discussion then continues with a comparison of the different DFT implementations in Abinit and CP2K, as well as an introduction to pseudopotentials and methods to reduce computational cost through the use of the Coulomb cutoff. The following chapter is based on [6], [7], [8], [9].

2.1 The many-electron problem

As a basis for our upcoming calculations, we first consider the simplest form of the Schrödinger equation.

$$\hat{H}\Psi_n = E_n\Psi_n \tag{2.1}$$

Where $\hat{H}$ is the quantum-mechanical Hamiltonian, containing the kinetic and potential energy operator

$$\hat{H} = \hat{T} + \hat{V}. \tag{2.2}$$

Since we do not want to limit ourselves to individual atoms in the following, but rather consider molecules, i.e. multi-particle systems, we must break down our eigenvalue equation 2.1 further. The wave function Ψ_n not only depends on a location variable, but is made up of the location coordinates of the electrons and those of the atomic nuclei.

$$\Psi_n \hat{=} \Psi(\mathbf{r}, \mathbf{R}), \tag{2.3}$$

where $\mathbf{r}/\mathbf{R}$ is the sum of all electron/nucleus position vectors. Given that electrons and nuclei interact with one another, it is necessary to introduce an interaction term, which ultimately leads to the many-body Hamiltonian

$$\hat{H} = \hat{T}_{nucl}(\mathbf{R}) + \hat{T}_e(\mathbf{R}) + \hat{V}_{Ne}(\mathbf{r}, \mathbf{R}) + \hat{V}_{ee}(\mathbf{r}) + \hat{V}_{NN}(\mathbf{R}). \tag{2.4}$$

The first two terms describe the kinetic energy operators of the nuclei and the electrons. The other three are the electrons-nuclei, electron-electron and nuclei-nuclei electrostatic Potentials.

$$\hat{T}_e(\mathbf{r}) = -\frac{\hbar^2}{2m} \sum_i \nabla^2_{\mathbf{r}_i} ,$$

$$\hat{T}_{nucl}(\mathbf{R}) = -\sum_A \frac{\hbar^2}{2M_A} \nabla^2_{\mathbf{R}_A} , \tag{2.5}$$

where m is the mass of an electron and M_A is the mass of the nucleus. $\mathbf{r}/\mathbf{R}$ describe the same vectors, as with the wave functions cf. 2.3. The particle-particle interactions are approximately described by Coulomb potentials [6][7]

$$\hat{V}_{Ne}(\mathbf{r}, \mathbf{R}) = -\sum_A \sum_i \frac{Z_A e^2}{|\mathbf{r}_i - \mathbf{R}_A|} ,$$

$$\hat{V}_{ee}(\mathbf{r}) = +\sum_i \sum_{j>i} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} , \tag{2.6}$$

$$\hat{V}_{NN}(\mathbf{R}) = +\sum_A \sum_{B>A} \frac{Z_A Z_B e^2}{|\mathbf{R}_A - \mathbf{R}_B|} ,$$

where A is the running index for the nuclei and i the running index for the electrons. Z_A/Z_B describes the atomic charge number of the considered atoms. The inner sum is restricted to $j > i / B > A$ to avoid double counting of the electron-electron/nucleus-nucleus interaction [6]. Combining the above we get the Hamiltonian of the system. In theory, it would be possible to solve the Schrödinger equation for an arbitrary many-body system and thereby obtain all of its physical properties. However, it is not feasible to separate the Hamiltonian into a sum of Hamiltonians of independent particles due to the presence of the electron-nucleus interactions. Consequently, one has to solve a coupled system of differential equations, which is too complex to be solved numerically. [7]

2.1.1　The Born-Oppenheimer approximation

In order to overcome the numerically unfeasible problem, some approximations are made, the so called Born-Oppenheimer approximation. The mass of protons and neutrons is 1836 times greater than that of electrons, resulting in a slower dynamic for the nuclei compared to electrons, which follow atomic movements instantaneously. That basically means the cores are frozen, compared to the electrons [7] .Therefore it is justifiable to separate the Hamiltonian into two independent parts, one for the electrons and the other for the nuclei. For electrons the Hamiltonian from Eq. 2.4 looses the kinetic energy operator of the nuclei $\hat{T}_{nucl}$ and $\hat{V}_{NN}(\mathbf{R})$ simplifies to $V_{NN,\mathbf{R}}$, [6][7]

$$\hat{H}_{e,\mathbf{R}} = \hat{T}_e(\mathbf{r}) + \hat{V}_{Ne,\mathbf{R}}(\mathbf{r}) + \hat{V}_{ee}(\mathbf{r}) + V_{NN,\mathbf{R}} . \tag{2.7}$$

Since the Coulomb Repulsion of the nuclei $V_{NN,\mathbf{R}}$ is a constant the Hamiltionian simplifies to

$$\hat{H}_{e^-} = \hat{T}_e(\mathbf{r}) + \hat{V}_{Ne,\mathbf{R}}(\mathbf{r}) + \hat{V}_{ee}(\mathbf{r}) . \tag{2.8}$$

We yield the Schroedinger equation for the many electron wavefunction

$$\left\{ -\frac{\hbar^2}{2m} \sum_i \nabla_{\mathbf{r}_i}^2 + -\sum_A \sum_i \frac{Z_A e^2}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_i \sum_{j>i} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \right\} \Psi_n = E_n \Psi_n . \tag{2.9}$$

An exact calculation for a many-body quantum system scales exponentially with system size. The calculations for this thesis are done using Noctua2 which has a peak computational performance of 5.4 PFLOPS. The biggest molecule in our data set has 218 electrons electrons which corresponds to

$$N_{\text{operations}} = \exp(218) \simeq 4.7 \cdot 10^{94} \tag{2.10}$$

Floating point operations. Even with the computational power of a supercomputer, the calculations would still take $2.0 \cdot 10^{61}$ times the age of the universe,

$$\begin{aligned} t &= 4.7 \cdot 10^{94} \, \text{FLOP} : 5.4 \cdot 10^{15} \, \text{FLOPS} = 8.8 \cdot 10^{78} \, \text{s} , \\ &8.8 \cdot 10^{78} \, \text{s} : \text{age}_{\text{Universe}} \simeq 2.0 \cdot 10^{61} \, \text{age}_{\text{Universe}} , \end{aligned} \tag{2.11}$$

which is an inconceivably vast number that is beyond the scope of a human lifespan. Therefore another approach to the many body quantum problem had to be designed.

2.2 Density functional theory

The beauty of Density functional theory (DFT) is, that all properties of a wavefunction can be extracted from the ground state density of the electrons $n(\mathbf{r})$ [8]. It condenses the N variables needed for the exact solution of the wavefunction Ψ to a single variable $n(\mathbf{r})$, greatly streamlining the calculations and significantly improving their efficiency [6]. In the following we discuss the Hohnberg-Kohn theorems, which form the foundation of DFT. Furthermore we also look at the Kohn-Sham-ansatz with which the ground state functionals of many electron systems can be approximated [8].

2.2.1 Hohnberg-Kohn theorems

Starting from equation 2.8 we redefine the electron-nuclei interaction potential $\hat{V}_{Ne,R}(\mathbf{r})$

$$\hat{V}_{Ne,\mathbf{R}}(\mathbf{r}) := \hat{V}_{ext} , \tag{2.12}$$

where $\hat{V}_{ext}$ is

$$\hat{V}_{ext}(\mathbf{r}) = \int d^3\mathbf{r} \, v_{ext}(\mathbf{r}) \, \hat{n}(\mathbf{r}) = \sum_{i=1}^{N} v_{ext}(\mathbf{r}) , \tag{2.13}$$

with the density operator $\hat{n}(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r_i})$ and $\hat{v}_{ext}(\mathbf{r}) = -e^2 \sum_A /|\mathbf{r} - \mathbf{R}_A|$ is the electron-nuclei interaction of a single electron with all existing nuclei and $\hat{V}_{ext}(\mathbf{r})$ is the sum of $\hat{v}_{ext}(\mathbf{r})$ over all

electrons. In order to obtain the first Hohnberg-Kohn theorem we have to define the electron density of a system $n(\mathbf{r})$:

$$n(\mathbf{r}) = \langle \Psi | \hat{n}(\mathbf{r}) | \Psi \rangle = \int d\mathbf{r}_1 d\mathbf{r}_2 ... d\mathbf{r}_N \, \Psi^*(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) \, \hat{n}(\mathbf{r}) \, \Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) \tag{2.14}$$

which means, that $n(\mathbf{r})$ is the expection value of the electron density operator $\hat{n}(\mathbf{r})$ to a given wavefunction Ψ. The first Hohnberg-Kohn theorem (HKT) states:

First Hohenberg-Kohn theorem: For any system of interacting particles in an external potential $V_{ext}(\mathbf{r})$, the potential $V_{ext}(\mathbf{r})$ is determined uniquely, except for a constant, by the ground state particle density $n_0(\mathbf{r})$ [8][S.122, direct quote]

This suggests, that if we know the ground state density $n_0(\mathbf{r})$ of a system, the Hamiltonian and also the many-body-wavefunction of the whole system are known. Therefore all physical properties of a system are determined by the ground state density $n_0(\mathbf{r})$ [8]. The proof of the first HKT is quite simple however it doesn't provide any guidance on how to find the functional. Thus the second Hohnberg-Kohn theorem is needed:

Second Hohenberg-Kohn theorem: A universal functional for the energy E[n] in terms of the density n(\mathbf{r}) can be defined, valid for any external potential $V_{ext}(\mathbf{r})$. For any particular $V_{ext}(\mathbf{r})$, the exact ground state energy of the system is the global minimum value of this functional, and the density $n(\mathbf{r})$ that minimizes the functional is the exact ground state density $n_0(\mathbf{r})$ [8][S.122, direct quote].

$$E[\tilde{n}] = F[\tilde{n}] + \int d^3\mathbf{r} \, v_{ext}(\mathbf{r})\tilde{n}(\mathbf{r}) \tag{2.15}$$

This means by minimizing the functional $E[\tilde{n}]$ from equation 2.15 the ground state energy E is gained [6].

$$E = \min_{\tilde{n} \in L_N} E[\tilde{n}], \tag{2.16}$$

where L_N describes the function space containing all electron densities for N electrons

$$L_N = \left\{ \tilde{n} : \mathbb{R}^3 \to \mathbb{R} \text{ with } \int \tilde{n}(\mathbf{r}) \, d^3\mathbf{r} = N \text{ and } \tilde{n}(\mathbf{r}) \geq 0 \; \forall \mathbf{r} \in \mathbb{R}^3 \right\}. \tag{2.17}$$

The second Hohenberg-Kohn theorem enables us to calculate the properties of an arbitrary system, however the form of the functional $F[n]$ is unknown. Therefore approximations have to be found in order to enable DFT calculations [6].

2.2.2   The Functional F[n]

Since $F[n]$ contains all Energy terms except the external Energy $E_{ext}[n]$ it can be written as

$$F[n] = T[n] + E_H[n] + E_{xc}[n] \, . \tag{2.18}$$

The calculation of $T[n]$ is dependent upon the ground state wave function $\Psi[n]$

$$T[n] = \langle \Psi[n] | \, \hat{T} \, | \Psi[n] \rangle \, , \tag{2.19}$$

where $\Psi[n]$ can be calculated from the ground state density n cf. 2.2.1. The Hartree energy

$$E_H[n] = \frac{e^2}{2} \int \int d^3\mathbf{r} d^3\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{2.20}$$

covers the interaction Energy of the charge density with itself. The exchange-correlation functional $E_{xc}[n]$ contains all quantum-mechanical electron-electron interaction properties beyond the Hartree energy $E_H[n]$ [6].

$$E_{xc}[n] = \langle \Psi[n] | \, \hat{V}_{ee} \, | \Psi[n] \rangle \tag{2.21}$$

where $\Psi[n]$ is the ground state wave function and n is the ground state density.

2.2.3   Kohn-Sham DFT

The first approximation for $F[n]$ was historically proposed by Thomas and Fermi, though this approximation is imprecise due to the assumption that the electrons within an atom form a homogeneous electron gas, which ultimately results in an inaccurate representation of the systems that are to be described. A more precise approach was developed by Kohn and Sham, who reintroduced orbitals. It uses an auxiliary system to replace the interacting many-body problem. Kohn and Sham suppose that the ground state energy of the interacting many-body system doesn't differ from a chosen one without interaction. In order to do so, the many body terms, making the calculation so difficult, are condensed into a exchange-correlation functional of the density. This enables accurate calculation to the extent of the approximation of the exchange-correlation functional of the density [8].

The Schematic 2.1 represents the transition to the auxiliary system, which is valid in both directions to and from the auxiliary system. Where HK/HK_0 is the Hohenberg-Kohn-theorem applied to the initial-/auxiliary-system. The KS arrow connects both system which illustrates the equivalence of the original many-body problem and the independent Kohn-Sham approach.
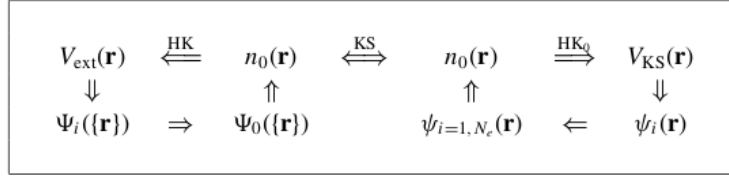
$$V_{ext}(\mathbf{r}) \overset{HK}{\Longleftarrow} n_0(\mathbf{r}) \overset{KS}{\Longleftrightarrow} n_0(\mathbf{r}) \overset{HK_0}{\Longrightarrow} V_{KS}(\mathbf{r})$$
$$\Downarrow \qquad \Uparrow \qquad \Uparrow \qquad \Downarrow$$
$$\Psi_i(\{\mathbf{r}\}) \Rightarrow \Psi_0(\{\mathbf{r}\}) \qquad \psi_{i=1,N_e}(\mathbf{r}) \Leftarrow \psi_i(\mathbf{r})$$

Figure 2.1: Schematic representation of the Kohn-Sham ansatz [8]

From the auxiliary approach an auxiliary Hamiltonian

$$\hat{H}_{aux}^{\sigma} = -\frac{1}{2}\nabla^2 + V^{\sigma}(\mathbf{r}) \tag{2.22}$$

follows [8]. Kohn and Sham also reintroduced the idea of orbitals, since in KS-DFT each electron is treated independently the electron density can be written as the sums of squares of the orbitals for each spin [8]

$$n(\mathbf{r}) = \sum_{\sigma} n(\mathbf{r}, \sigma) = \sum_{\sigma} \sum_{i=1}^{N^{\sigma}} \left| \psi_i^{\sigma}(\mathbf{r}) \right|^2 . \tag{2.23}$$

Since there are only two possibilities for the occupation of the spin σ, namely up and down, the density simplifies to

$$n(\mathbf{r}) = 2 \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2 , \tag{2.24}$$

where ψ_i are called Kohn-Sham orbitals. In case of molecules i is the i-th molecular orbital index. The KS-orbitals have the constraint to be orthonormal [6]

$$\langle \psi_i | \psi_j \rangle = \int d^3\mathbf{r} \, \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}) = \delta_{ij} . \tag{2.25}$$

Although the reintroduction of orbitals resulted in a slower algorithm compared to approaches such as the Thomas-Fermi approximation, the benefits of incorporating N orbitals outweighed the cost. Hence the kinetic energy of the interacting system's electrons T can be approximated with that of non-interacting electrons T_S [6][S.25-26]

$$T_S[\psi_q, \psi_2, ..., \psi_n] := 2 \sum_{i=1}^{N} -\frac{\hbar^2}{2m} \langle \psi_i | \nabla^2 | \psi_i \rangle ,$$
$$\langle \psi_i | \nabla^2 | \psi_i \rangle = \int d^3\mathbf{r} \, \psi_i^*(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) . \tag{2.26}$$

In reference to the total energy functional $E[n]$ derived from the second HKT, the KS approximation provides an expression for the functional $F[n]$. In order to accomplish that $F[n]$ has to be split into two parts. The kinetic energy functional $E[n]$ being the first cf. equation 2.26. The second part contains the Hartree energy $E_H[n]$ and the exchange-correlation energy $E_{xc}[n]$, which contains all electron-electron interactions beyond the Hartree Energy [6]. By plugging the above

into the second Hohnberg-Kohn theorem the following is obtained

$$E[n] = T[n] + E_H + E_{xc}[n] + \int d^3\mathbf{r}\, v_{ext}(\mathbf{r})n(\mathbf{r})\,, \tag{2.27}$$

with the definition of the external energy

$$E_{ext}[n] = \int d^3\mathbf{r}\, v_{ext}(\mathbf{r})n(\mathbf{r})\,. \tag{2.28}$$

The second HKT transforms to

$$E[n] = T[n] + E_H + E_{xc}[n] + E_{ext}[n]\,. \tag{2.29}$$

Since the kinetic energy functional of the interacting system can only be approximated by T_S cf. 2.26. Therefore the total energy functional $E[n]$ approximates to

$$E[n] \approx T_S[\psi_1, \psi_2, ..., \psi_N] + E_H + E_{xc}[n] + E_{ext}[n]\,. \tag{2.30}$$

To gather information about the Energy E the second part of the second HKT 2.16 is needed. Applying the minimization principle to equation 2.30 the below follows

$$E \approx \min_{\substack{orthonormal \\ \psi_1, \psi_2, ..., \psi_N}} [T_S[\psi_1, \psi_2, ..., \psi_N] + E_H[n] + E_{xc}[n] + E_{ext}[n]]\,. \tag{2.31}$$

It is possible to prove that the approximation can be written as an equal, however the proof wouldn't add any value to this discussion. The interested reader can find the proof in any generic DFT book [7].

2.2.4   Kohn-Sham equations

To obtain the ground state energy, in actual calculations the density is not optimized as shown in eq. 2.31, but the Kohn-Sham orbitals. Since the Kohn-Sham orbitals have to be orhonormal, we have to use Lagrange multiplier, in order to minimize the ground state energy, which leads to the Kohn-Sham equations [6]

$$\left\{ -\frac{\hbar^2}{2m}\nabla^2 + v_{ext}(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc}(\mathbf{r}) \right\} \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r})\,. \tag{2.32}$$

The KS equations are containing the external potential $v_{ext}(\mathbf{r})$ 2.13, the Hartree potential $v_H(\mathbf{r})$, being the potential to the corresponding Hartree energy cf. 2.20 and the exchange-correlation potential $v_{xc}(\mathbf{r})$ which can be computed by the functional derivative of the exchange-correlation functional $E_{xc}[n]$:

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta(\mathbf{r})} \tag{2.33}$$
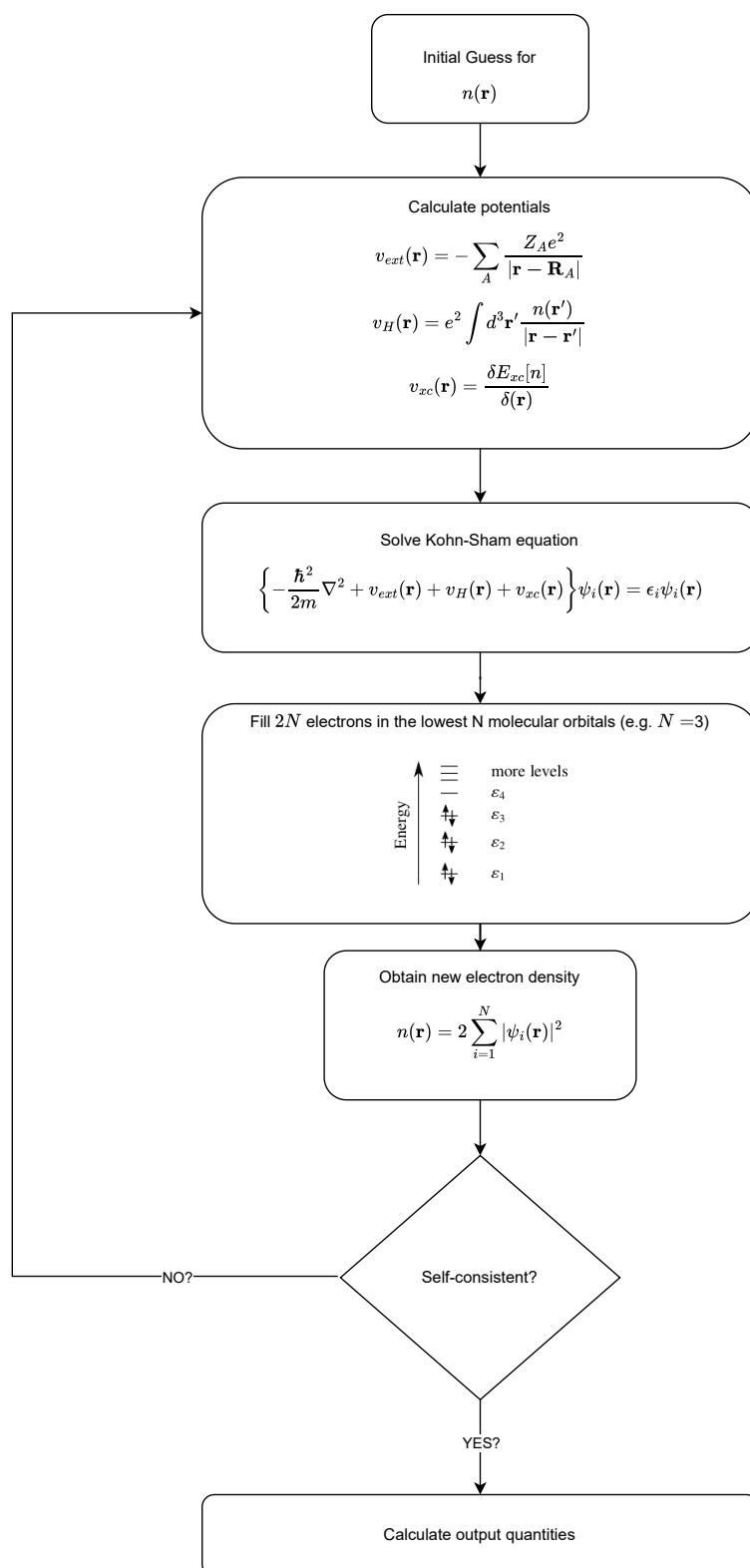
2.3 SCF-Cycle

Figure 2.2: SCF-Cycle drawn after [6], [8]

To obtain significant results from the Kohn-Sham equations, they have to be solved self-consistently. In order to perform an actual calculation, the effective Potential V_{eff} and the electron density $n(\mathbf{r})$ are successively varied, until the self consistency is reached [8]. Where the effective Potential is simply the sum of the three potentials of the KS-equation cf. 2.34

$$V_{\text{eff}}(\mathbf{r}) = v_{ext}(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc}(\mathbf{r}) \tag{2.34}$$

The flowchart 2.2 illustrates the mechanism of the SCF-cycle. In order to get the calculations started, it is mandatory to give an initial guess for the electron density $n(\mathbf{r})$, which determines the effective potential V_{eff}, from there on the computationally intensive KS-equation has to be solved, from which a new electron density can be derived, which starts a new cycle. These steps will be continued until the SCF-cycle is converged [8]. That also implies, that the KS-orbitals $\psi_1, \psi_2, ..., \psi_N$ minimize the energy functional $E[\psi_1, \psi_2, ..., \psi_N]$ and we can now calculate the ground state energy cf. Eq. 2.31. Since the computational cost of KS-DFT scales with $O(N^3)$ it is possible to do much more sophisticated calculations in terms of system size, compared to the brute-force approach discussed in chapter 2.1.1 [6].

Now the question arises, whether there is any connection between the ground state energy (HOMO) and the ionization potential. That's where Janak's theorem comes into play, it states that the KS self-energies ϵ_i correspond to the change in energy of the total energy ∂E due to the addition or removal of electrons ∂n_i

$$\frac{\partial E}{\partial n_i} = \epsilon_i \, . \tag{2.35}$$

This allows the electron affinity and ionization energy to be derived from the DFT calculations. The ionization energy, which corresponds to the removal of an electron, is therefore the energy of the highest occupied molecular orbital (HOMO)

$$E_{HOMO} = E_N - E_{N-1} \, . \tag{2.36}$$

The Lowest unoccupied molecular orbital (LUMO) energy however can't be directly calculated even if the funcitonal F[n] was exactly known. The LUMO energy follows from the gap Energy

$$E_{gap} = E_{HOMO} - E_{LUMO} + \Delta_{xc} \tag{2.37}$$

The Janak theorem is only valid for an infinitesimal change of the occupation number; therefore, DFT calculations are unable to resolve the ionization energy of, for example, an electron in close proximity to the core. As the exchange correlation functional can't be solved exactly, it is approximated. The KS-eigenvalues are only a approximation of the systems eigenvalues, which leads to the fact that the HOMO and LUMO energies aren't the exact ionization energy/electron affinity, but approximations too. [9]

2.4 Basis Sets

The Kohn-Sham equations are based on Kohn-Sham orbitals. The SCF cycle, when iterated, yields meaningful results only if the underlying basis functions, which represent the Kohn-Sham orbitals, are sufficiently accurate. The subsequent section provides an explanation of how basis sets are formed and the different basis sets that are available. A complete basis set is a linear combination of infinitely many functions, which in theory could exactly describe an orbital. However since in reality you can only use a finite amount of functions the accuracy of our calculations is dependent on the scale as well as the precision of the basis set. In general there are 2 different methods used to implement such basis sets, they are either composed of a linear combination of atomic orbitals (e.g. Slater Type Orbitals, Gaussian Type Orbitals) or plane waves. The better a single basis set function matches the unknown orbital function, the smaller the basis set can be, which is the reason for the different basis set approaches. In the following, we discuss a selection of different basis sets and their advantages and disadvantages [10].

2.4.1 Slater and Gaussian Type Orbitals

The most common basis set being composed of atomic orbitals are the Slater Type Orbitals (STO) and the Gausian Type Orbitals (GTO). STOs have the following form

$$\chi_{\xi,n,l,m}(r,\theta,\phi) = N Y_{l,m}(\theta,\phi) r^{n-1} e^{-\xi r}\,, \tag{2.38}$$

where N is the normalization constant, $Y_{l,m}$ spherical harmonic functions, r is the distance to the atomic nucleus and ξ is a shielding constant. Given that a single STO lacks radial nodes, the superposition of numerous STOs provides the basis set with radial nodes. The combination of numerous functions and the exponential dependency results in a rapid convergence of the basis set. However three- and four-center two-electron integrals cannot be solved analytically, hence the calculations with a STO basis set are quite expensive. Another numerically cheap basis set are the Gaussian Type Orbitals

$$\chi_{\xi,n,l,m}(r,\theta,\phi) = N Y_{l,m}(\theta,\phi) r^{2n-2-l} e^{-\xi r^2}\,. \tag{2.39}$$

In certain aspects, the GTOs are less effective than the STOs. Fundamentally, the GTOs are unable to accurately describe the desired behavior, given that, as the name Gaussian Type Orbitals suggests, they possess a zero slope in proximity to the atomic centre. The STOs however have a discontinuous derivative which is exactly what we were expecting at the nucleus. A further issue that arises from the r^2 dependency is that the GTOs decline too rapidly far away from the nucleus. These problems suggest, that the number of GTOs needed to properly describe an atom will largely exceed the number of needed STOs. Indeed to get comparable results three times more GTOs than STOs are required to get similar results. Since the computational cost is determined not only by the sheer number of basis functions, but also by the complexity of solving those functions, the

three times higher number of GTOs is compensated by the simplicity of their calculation. That's why Gaussian Basis sets are state of the art in electronic structure calculations. [10]

2.4.2 Plane Wave Basis

An alternative approach to the description of orbitals is through the use of plane waves. Rather than employing a linear combination of Gaussian functions, it is possible to utilize functions that are capable of describing the entire system. The reasoning behind this is that the valence electrons of metals exhibit a similar behavior to that of free electrons. This justifies the assumption of using the solution of free electrons as basis functions. The solution of the one dimensional Schrodinger equation for free particles gives us

$$\phi(x) = Ae^{ikx} + Be^{-ikx} \,, \tag{2.40}$$

$$\phi(x) = A\cos(kx) + B\sin(kx) \,. \tag{2.41}$$

The corresponding energy has a k^2 dependency

$$E = \frac{1}{2}k^2 \,, \tag{2.42}$$

for infinite systems the energy difference between adjacent levels vanishes, the orbitals merge into bands, which can be described by the plane waves. In three dimensions the plane wave formalism can be expressed as

$$\chi_k(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \,. \tag{2.43}$$

Since the $\mathbf{k}$ value corresponds to the frequency, a high $\mathbf{k}$ value indicates a rapid oscillation. The unit cell translation vector $\mathbf{t}$ determines the allowed $\mathbf{k}$ values

$$\mathbf{k} \cdot \mathbf{t} = 2\pi m \,, \tag{2.44}$$

with m being a positive integer. From equation 2.44 one can find that the typical spacing between two $\mathbf{k}$ vectors is about 0.01 eV. A typical cutoff energy is 200 eV, from which we can derive that an average plane wave basis set has up to 20000 functions. For small systems GTOs are favorable, however since the plane wave basis set only depends on the size of the unit cell, plane waves are favorable for large systems with many atoms. Normally plane waves are only used for lattice periodic calculations, however it is also possible to use them for molecular calculations. In order to do so, the molecule is placed in a unit cell so large that it doesn't interact with the molecule in the unit cell next to it. This so called supercell approach, as described above, requires a huge amount of plane waves, which is why in this case a GTO approach is preferred. This fact indeed is very interesting for this thesis, since it let's us assume, that there should be a significant difference in consumed RAM and computational cost between GTOs and plane waves [10].

2.5   Pseudo Potentials

Electrons in low orbitals, close to the atomic core are extremely localized. Therefore, especially in a plane wave basis, electrons are hard to describe. For a good approximation of a sharp peak many plane wave functions with a high frequency are needed. Since the frequency is determined by the cutoff energy, a high cutoff energy is necessary. At the end this would result in a huge basis set, which would make our calculations extremely costly. In order to overcome this problem pseudo potentials are introduced. The idea behind them is that electrons in proximity to the core have a negligibly small impact on the Coulomb interaction. Therefore the core electrons are combined into so called pseudo potentials. These are tabulated for any given molecule. Since near the core pseudo potentials are smoother than the Coulomb potentials, less plane wave functions are needed to converge the basis expansion [6]. Thus saving us a lot of computational time, while compromising our results only in a negligible manner.

2.6   Implementation of DFT in Abinit and CP2K

The forthcoming calculations employ two distinct implementations of Density Functional theory (DFT), namely Abinit and CP2K. Since the main focus of this thesis is comparing plane wave basis sets versus Gaussian Basis sets in Kohn-Sham DFT, pseudopotentials of the same class were chosen for both sets of calculations, the so called Goedecker-Teter-Hutter (GTH) pseudopotentials [11]. Which is very important, since the utilization of different pseudopotentials would highly impact the comparability. Since variation in results caused by using different pseudopotentials within the same program, such as Abinit, can be greater than the differences observed when comparing results from two different programs. The main challenge in both Abinit and CP2K is to converge the calculations. The difference between CP2k and Abinit is that Abinit using plane waves is converged in a different way compared to CP2K using Gaussian.

Abinit:
If we recall the form of a plane wave and its energy dependency:

$$\chi_k(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}$$
$$E = \frac{1}{2}k^2 \tag{2.45}$$

It becomes apparent that by increasing the cutoff energy, a greater ensemble of plane wave functions is generated, thereby enabling a more precise description of the orbitals. However since k scales with $k \sim \sqrt{E}$ the calculations get increasingly expensive for large plane wave basis sets.

CP2K:
For CP2K there is not a simple variable, that we can scale in order to get better results. Instead

one has to choose different sized Gaussian basis sets, similarly, the greater the size of the basis set, the more precise the resulting calculations.

Another factor determining the accuracy of the calculations is the size of the unit cell. The unit cell is characterised by periodic repetition, a property that is particularly salient in the context of molecular calculations. It is crucial to ensure that the cell is sufficiently large to minimise or eliminate interactions with neighbouring molecules. It is therefore necessary to select a cell of sufficient size. The properties and criteria of the unit cell are the same for Gaussians and plane waves.

2.7   Coulomb Cutoff

Later we find that our calculations are limited by the size of the supercell, to address this problem we introduce the concept of the Coulomb Cutoff. Due to the periodic repetition of the supercell and the influence of objects in one unit cell on those in another, problems arise when simulating objects, such as molecules, that do not repeat infinitely or periodically. The initial thought would be, to enlarge the size of the unit cell in order to create more vacuum in between the neighboring objects, however this approach is very costly and also leads to undesired artifacts no matter how big the unit cell is. In order to maintain the advantages of the supercell approach and to minimise the unwanted artifacts, a strategy is now being developed to cut off the coulomb interaction outside the unit cell, the so-called coulomb cutoff approach [12]. To be exact the potential that is troubling us is the Hartree potential

$$V_H(\mathbf{r}) = \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \tag{2.46}$$

The Hartree potential is constructed through the introduction of a domain in which the potential acts normally, whilst outside this domain it is zero. In order to accomplish this, the potential must be divided in the following manner

$$V_H(\mathbf{r}) = \int d\mathbf{r}' n(\mathbf{r}') \cdot v_{tr}(\mathbf{r} - \mathbf{r}') \tag{2.47}$$

where $n(\mathbf{r})$ describes the density function and $v_{tr}(\mathbf{r})$ describes the domain dependent part of the Hartree potential

$$v_{tr} = \begin{cases} \frac{1}{r}, & \mathbf{r} \in \mathrm{D} \\ 0, & \mathbf{r} \notin \mathrm{D} \end{cases}$$

Rozzi et al. [12] demonstrate, that the occurrence of non-analytical behavior can be effectively eliminated by ensuring the adequate screening of both the ionic and Hartree potential. Therefore one can assume that by using the coulomb cutoff in our Abinit calculations the calculations should show better results even while using a smaller unit cell. Especially with molecules containing strongly electronegative atoms thus creating a dipole moment the coulomb cutoff should grant better results.

2.8 Mean Absolute Deviation

For the comparison of our Calculations we use the **M**ean **A**bsolute **D**eviation

$$\text{MAD} = \sum_{i=1}^{n} \frac{|y_i - x_i|}{n} \tag{2.48}$$

in the literature the MAD is also known as the MAE, the **M**ean **A**bsolute **E**rror.

Chapter 3

Convergence Tests

In order to get accurate calculations and results some convergence criteria have to be found. In the following we introduce and discuss the three most important convergence criteria used in this thesis on the basis of a C_6H_6 molecule. Starting with the **tol**erance on the **Dif**ference of total **e**nergy (toldfe) that determines when the SCF cycle comes to an end, the discussion continues with an examination of the scale of the system size and, consequently, the dimensions of the unit cell (acell) and the cutoff energy (ECUT). The findings of this chapter are used as input parameters for all the molecules contained in the GW-100 molecule set. The numerical accuracy of Abinit is 0.01 mHa which corresponds to ~ 0.3 meV. Consequently, all results are rounded up to the nearest 0.1meV.

3.1 Convergence of toldfe

The most apparent aspect is the convergence criteria of the SCF cycle, in Abinit this point is reached, if in 2 consecutive SCF cycles the difference of the total molecule energy is smaller than a predetermined value (toldfe), the SCF cycle stops and the calculation is considered converged. The following calculations are using a toldfe value of toldfe $= 1 \cdot 10^{-10}$ Ha. The reasoning for this value is a convergence test done with a C_6H_6 molecule c.f. figure 3.1.

Since the convergence of toldfe isn't directly relying on the system size (unitcell, cutoff energy), we started with general values that should be sufficient for molecular systems of this size [13]. Therefore a unit cell of $40\,a_b$ and a cutoff energy of 40 Ha were selected. In order to determine the converged tolfe value, 13 calculations were started. In each calculation the toldfe value was increased by a magnitude of 10^{-1} Ha. For each calculation the average of the energy levels including the HOMO was computed 3.1. Starting from a toldfe value of $1 \cdot 10^{-10}$ Ha, no change of the average orbital energy was observed c.f. 3.1. Therefore toldfe was assumed as converged.

The exact convergence of the before mentioned variables determining the system size is discussed in the following.
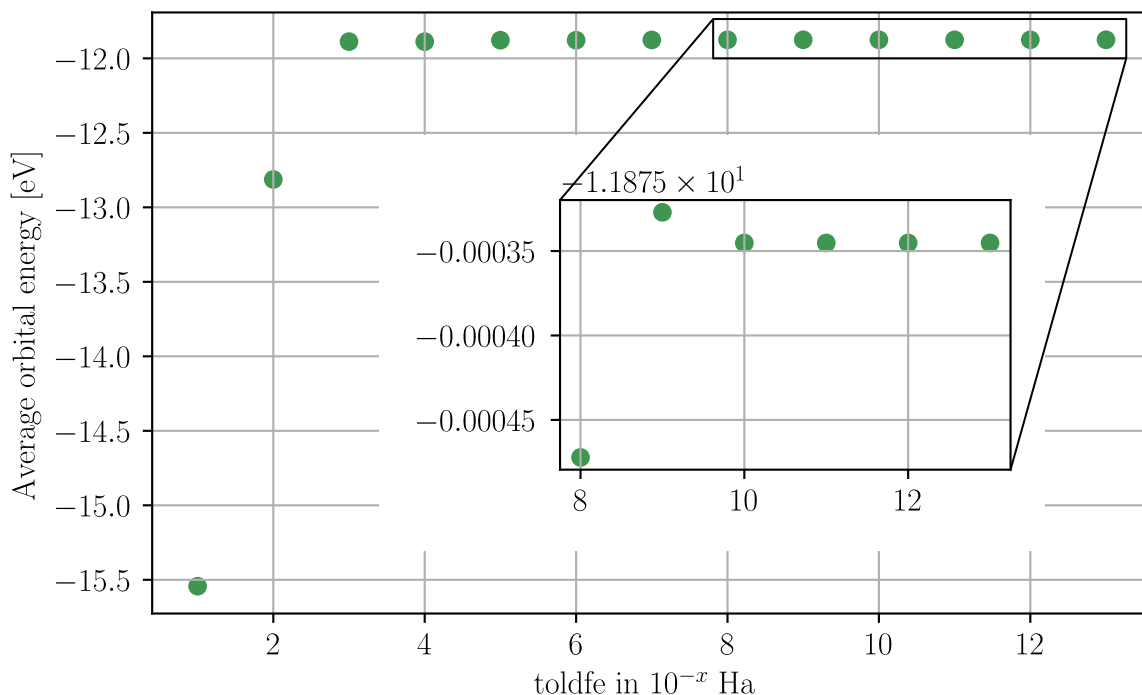
Figure 3.1: Plot of average orbital energy (up to ϵ_{HOMO}) of a C_6H_6 molecule against TOLerance on the DiFference of total Energy (toldfe). For tolfe values higher or equal to $1 \cdot 10^{10}$ Ha no improvements up to the the numerical limit of Abinit were found.

3.2 Convergence of the System size variables

Our objective is to achieve an overall precision in the comparison between Abinit and CP2K in the same region as the GW 100 Benchmark. The GW 100 Benchmark has a precision of a few meV [5]. To get our benchmark in the same region, the convergence tests have to be even more precise, as a matter of fact they should be 10 times more precise. Which leads to a convergence test accuracy of ~ 0.1 meV. However the numerical accuracy of Abinit is limited to 0.00001 Ha = 0.01 mHa $\simeq$ 0.3 meV. It undermines our desired accuracy by a factor of three, yet it still remains within an acceptable range. In the following we will see that the lack of numerical accuracy is negligible compared to the other challenges we are going to face.

Cutoff Energy:
The principal behind converging the cutoff energy is quite simple. At first we have to acknowledge the fact that the unit cell (acell) and the cutoff energy (ECUT) converge independently of each other [13]. Accordingly, to converge ECUT a small acell = $12\,a_b$ can be chosen, which, given the independence of the variables involved, allows for rapid and cost-effective yet accurate calculations. The next step is to increase ECUT step by step until the HOMO energy differs by a margin smaller than 0.01 mHa $\simeq$ 0.3 eV. In figure 3.2 the received data set is illustrated. From the subplot we can derive, that starting from a cutoff energy of 100 Ha, the HOMO energy is con-

stant at ∼ −3.7 eV. (except the one outlier at E_{CUT} = 195 Ha) Therefore the ionization energy (HOMO energy) varies by less than 0.3 meV and the cutoff energy is considered as converged at E_{CUT} = 100 Ha.
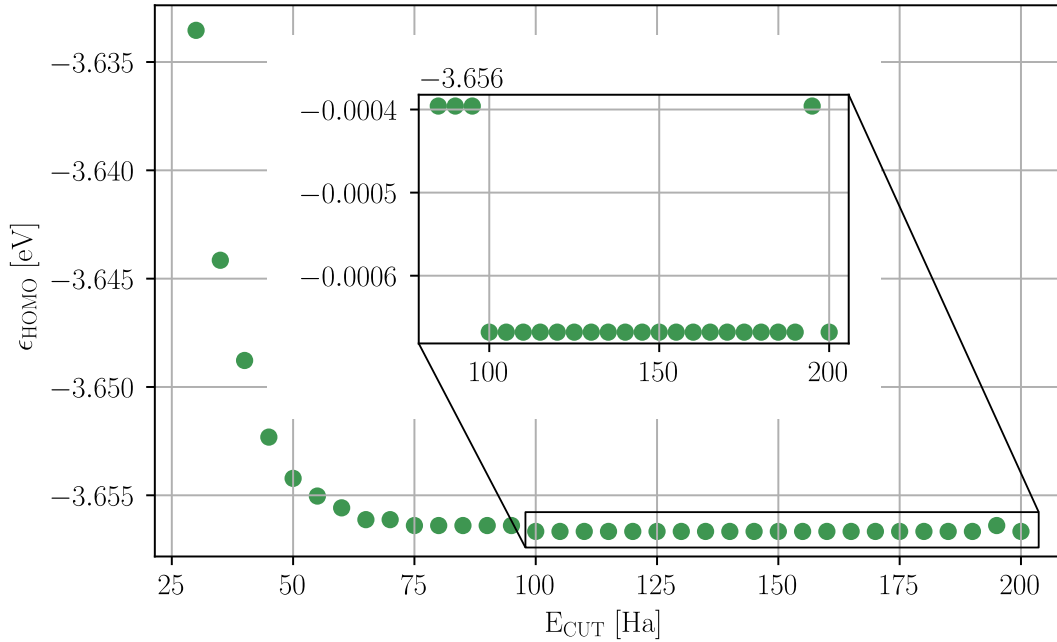
Figure 3.2: Plot of ϵ_{HOMO} against E_{CUT} of a C_6H_6 molecule. While slowly scaling E_{CUT} 5 Ha at a time, ϵ_{HOMO} decreases rapidly. At a cutoff energy of E_{CUT} = 100 Ha the HOMO energy is fully converged, displaying a value of ϵ_{HOMO} = 0.13438 Ha = 3.65668 eV.

Unit Cell:

Following the same principal for the unit cell, we again chose a modest E_{CUT} = 10 Ha and increased acell step by step until we reached an acell value of 225 a_b despite the enormous unit cell, the HOMO energy didn't converge at all as seen in figure 3.3. At first glance the graph seems to have an approximately exponential decay, however if you take a closer look at the region of $120\,a_b − 220\,a_b$ it becomes evident, that the measured values aren't converging but oscillating in the interval

$$\epsilon_{HOMO} \in [−6.353\,\text{eV}, −6.350\,\text{eV}].   (3.1)$$

Therefore we can consider the unit cell as converged at acell = 150 a_b. The maximum deviation of the oscillating HOMO values is 0.00007 Ha ≃ 1.9 meV which exceeds our original aimed for precision by a factor of 10. Despite that enlarging the size of the unit cell even more would not make any sense since the computational cost would not be feasible. Which leaves us with an unit cell size of acell = 150 a_b ≃ 79.4 Å and a cutoff energy E_{CUT} = 100 Ha ≃ 2721 eV.
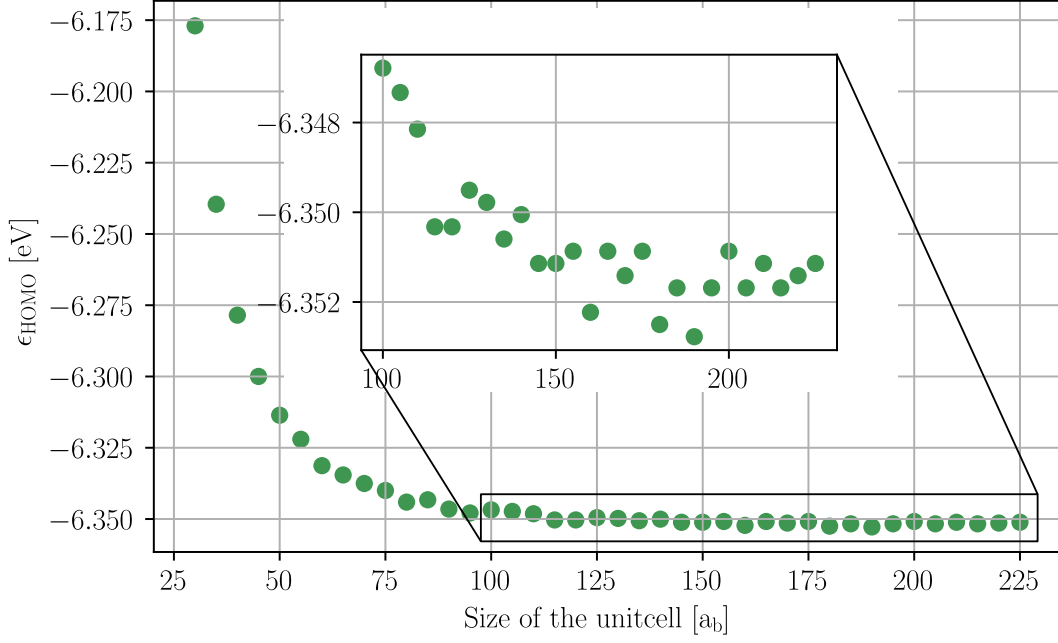
Figure 3.3: Plot of ϵ_{HOMO} against the size of the unit cell (acell) of a C_6H_6 molecule. While slowly scaling acell 5 a_b at a time, at first the ϵ_{HOMO} converges exponentially. This trend continues until acell = 120 a_b. From that point on, the HOMO energies start to oscillate in the interval $\epsilon_{\text{HOMO}} \in [-6.353\,\text{eV}, -6.350\,\text{eV}]$.

The initial calculations were computationally infeasible due to their size, which resulted in an Out-Of-Memory (OOM) error. Consequently, it is imperative to devise a method that reduces the memory requirements of our calculations while maintaining their accuracy.

ECUT

For the cutoff energy the reduction is quite simple, reducing ECUT from 100 Ha to 80 Ha leaves us with a HOMO energy difference of:

$$|\epsilon_{\text{HOMO}}^{E_{\text{CUT}}=100} - \epsilon_{\text{HOMO}}^{E_{\text{CUT}}=80}| = |-0.13438\,\text{Ha} - (-0.13437)\,\text{Ha}| = 0.01\,\text{mHa} \tag{3.2}$$

Taking into account the numerical accuracy of Abinit that leaves us with a accuracy of ~ 0.6 meV. In comparison to the before calculated accuracy of the unit cell of 1.9 meV, the error in ECUT is still considerably low, especially considering it only gets worse by scaling down acell.

ACELL

For acell the rescaling wasn't that straight forward. The simplest approach would be to reduce the size of the unit cell step by step until the calculations start again. That's exactly what we did. At an acell value of 120 a_b and $E_{\text{CUT}} = 80$ Ha our system was small enough for our calculations to

start again. Which leaves us with a unit cell precision of:

$$|\epsilon_{\text{HOMO, max}}^{\text{acell}=120} - \epsilon_{\text{HOMO, min}}^{\text{acell}=190}| = |-0.23337\,\text{Ha} - (-0.23346)\,\text{Ha}| = 0.09\,\text{mHa} \tag{3.3}$$

Since the error of the cut off energy $\Delta\epsilon_{E_{\text{CUT}}} \simeq 0.6$ meV and the error of unit cell $\Delta\epsilon_{\text{acell}} \simeq 2.5$ meV are in the order of a few meV, we estimate our total error for the Abinit vs. CP2K comparison to be at least the same magnitude.

3.3 Convergence of CP2K

In principle, a comparable convergence test would be required for CP2K as for Abinit; however, this would exceed the scope of this bachelor thesis. Therefore, the existing literature was consulted. The first part of our basis consists of TZVP-MOLOPT Gaussian basis, which itself is often regarded as a converged basis set as constructed in [14] and used in [5], [15]. The second part consists of the aug-cc-QZVP Basis, which is described and used in [16] and [17]. The paper [14] also uses the GTH pseudo potentials and suggests, that the usage of Gaussians and plane waves are equally valid. This proposition constitutes a solid foundation for the comparison that is the objective of this thesis.

Chapter 4

Comparing the GW-100 molecule set

To ensure comparability, the GW100 molecule set is used. The 100 molecules have a been carefully selected to give a useful representation of a broad spectrum of molecules. The ionization potentials range from ~ 4 eV for Rb_2 to ~ 25 eV for He. As organic compounds make up a large part of our world, a variety of Carbon based molecules eg. $C_2H_6, C_2H_4, C_2H_2, C_6H_6, C_8H_8, CO, CO_2, C_4$, are included, other organic molecules like alcohols ($CH_3OH, C_2H_5OH, HCOOH$), aldehydes and nitrogenous bases are also added. Another covalent bond-type of interest is metallic compounds such as $Ag_2, Li_2, K_2, Na_2, Rb_2$. Apart form covalent bonding also small metallic clusters such as Na_4, Na_6 are included, which form metallic bonds. In order to also cover ionic bonds alkaline metal halide compounds like LiF and KBr are added, which incorporate a small dipole moment. In contrast to them the alkaline earth metal compounds MgF_2 and MgO have a large dipole moment. The molecular structure, building the foundation of our calculations, are taken from experimental data [5].

In the following, we compare the aforementioned GW-100 molecule set, calculated using a plane wave basis in Abinit, against a Gaussian basis calculated in CP2K. The comparison focus' on the HOMO and LUMO energies.

4.1 HOMO

As the end of the convergence chapter suggests we expect a deviation between the CP2K and Abinit HOMO energy levels of about ~ 3 meV. The first Calculation was performed with the GTH-Pseudos for both Abinit and CP2K. In CP2K we used the TZVP-MOLOPT Basis combined with the aug-cc-QZVP Basis. The latter only contains Gauss exponents smaller $20\,a_b$. Since we are using pseudo potentials no strongly localized Gauss functions are needed. Until the last chapter the specifications of the CP2K calculations $C_0^{\text{CP2K,TZVP+aug-cc-QZVP}}$ don't change. In the following we primarily focus on the Abinit calculations. For the first Abinit calculation, $C_1^{\text{Abinit,80\,Ha,120\,a_b,no\,CC}}$ an unit cell size of $120\,a_b$ and a cutoff energy of $80\,Ha$ are selected, as suggested in the convergence

Chapter. The MAD of this calcualtion comes to

$$\text{MAD}^{\text{CP2K,TZVP+aug-cc-QZVP}}_{\text{Abinit,80 Ha,120 a}_{\text{b}}\text{,no CC}} = 7.5\,\text{meV}. \tag{4.1}$$

In order to ascertain that this represents the optimal convergence that can be achieved, further Abinit internal tests must be conducted. A subsequent question is how far the non-convergence of the calculations affected the mean deviation. In order to respond to this question, a calculation was performed on each molecule in the GW-100 set on two occasions. In theory, the optimal approach would be to augment the unit cell and the cut-off energy by a specific amount, and then to initiate the calculations once more. However as we discussed before the legacy calculations were already done with the biggest possible values. Therefore we had to scale them down in the following way. In the first calculation block $C_2^{\text{Abinit,80 Ha,100 a}_{\text{b}}\text{,no CC}}$ we only scaled the unit cell size, thus it was done with an acell value of $100\,\text{a}_{\text{b}}$ and an ECUT value of $80\,\text{Ha}$. The second calculations block $C_3^{\text{Abinit,60 Ha,120 a}_{\text{b}}\text{,no CC}}$ was done with an acell value of value of $120\,\text{a}_{\text{b}}$ and an ECUT value of $60\,\text{Ha}$, which means we only scaled the cutoff energy and left the unit cell constant. Those two calculations are compared to the previous calculation with an acell size of $120\,\text{a}_{\text{b}}$ and a energy cutoff of $80\,\text{Ha}$. The MAD

$$\text{MAD}^{\text{Abinit,80 Ha,100 a}_{\text{b}}\text{,no CC}}_{\text{Abinit,80 Ha,120 a}_{\text{b}}\text{,no CC}} = 1.5\,\text{meV} \tag{4.2}$$

for a smaller acell is relatively small compared to the MAD for a smaller ECUT

$$\text{MAD}^{\text{Abinit,60 Ha,120 a}_{\text{b}}\text{,no CC}}_{\text{Abinit,80 Ha,120 a}_{\text{b}}\text{,no CC}} = 8.5\,\text{meV}. \tag{4.3}$$

In contrast to our preliminary investigation, in which we had assumed that ECUT would be converged to a satisfactory extent, while acell would not be as converged, calculation 4.3 and 4.2 indicate that acell has a superior convergence quality compared to ECUT. Consequently, it is imperative that we formulate a strategy to enhance ECUT while preserving the relatively optimal convergence of the unit cell. The subsequent discourse encompass' a deliberation on the feasibility and potential methodologies associated with this endeavor.

4.1.1   Reduction of computational cost with coulomb cutoff

The incorporation of a coulomb cutoff in the calculation was shown to reduce computational cost and enable higher precision c.f. 2.7. Therefore it is added to our calculations. From legacy calculations in CP2K follows, that for an cutoff of $60\,\text{a}_{\text{b}}$ the unit cell size should be converged. This hypothesis is verified by a convergence test in which the size of the unit cell was scaled with and without the cutoff c.f. 4.1. The simulations utilizing the cutoff demonstrate not only accelerated convergence but also follow a linear pattern, with small oscillations for acell values greater than $40\,\text{a}_{\text{b}}$. This is in contrast to the non-cutoff calculations, which, as demonstrated in section 3.2, follow an exponential pattern that persists even for acell values well above $40\,\text{a}_{\text{b}}$. Consequently,

the unit cell is regarded as converged for the aforementioned value of $40\,a_b$. For our calculations, we use acell $= 60\,a_b$ instead of acell $= 40\,a_b$, even though the calculations already converge at $40\,a_b$, to provide a small buffer for possible deviations or errors in the convergence test.
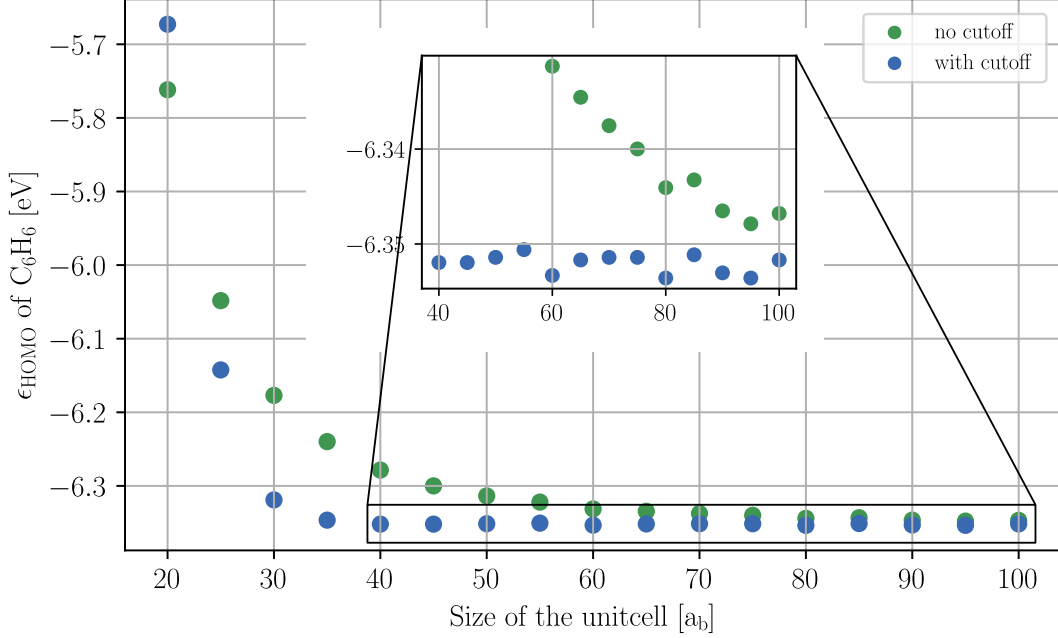
Figure 4.1: Plot of ϵ_{HOMO} against unit cell size (acell) for a C_6H_6 molecule. The blue graph corresponds to calculations with coulomb cutoff, the green graph to calculations without coulomb cutoff. The calculation with coulomb cutoff converge significantly faster. Form acell = 40 Ha the coulomb cuftoff calculation shows the same oscillating behavior as already seen in 3.2.

In order to investigate the impact of the coulomb cutoff, as well as the reduction in unit cell size, the molecule set is re-computed on another occasion $C_4^{\text{Abinit,80\,Ha,60\,a_b,CC}}$. This is done with the legacy cutoff energy of 80 Ha and the previously discussed unit cell size of $60\,a_b$. The coulomb cutoff is also included. Comparing the results with our first calculation $C_1^{\text{Abinit,80\,Ha,120\,a_b,no\,CC}}$ we get a MAD of

$$\text{MAD}_{\text{Abinit,80\,Ha,120\,a_b,no\,CC}}^{\text{Abinit,80\,Ha,60\,a_b,CC}} = 1.7\,\text{meV}. \tag{4.4}$$

Despite the reduction of the unit cell's size by half, the difference between the MAD is only $\simeq 1.7\,\text{meV}$. As demonstrated in section 4.1 the unit cell size increased from $120\,a_b$ to $100\,a_b$ without the coulomb cutoff, resulting in a difference of $\simeq 1.5\,\text{meV}$. Which is only 0.2 meV less than the current difference, indicating that the calculations using the coulomb cutoff are successful. As previously hypothesized, the reduction in cell size is intended to free up computational capacity for the augmentation of ECUT. The present study aims to determine whether an increase in the cutoff energy yields more precise calculations.

After thourough testing, we could establish a maximum cutoff energy of 400 Ha, when using the coulomb cutoff. Therefore the molecule set is calculated once more with an unit cell size of $60\,a_b$, a cutoff energy of 400 Ha, aswell as incorporated coulomb cutoff, $C_5^{\text{Abinit,400\,Ha,60\,a_b,CC}}$. Comparing

the resultsof this calculation to the results our first calculations $C_1^{\text{Abinit,80 Ha,120 a}_b\text{,no CC}}$ we yield a MAD of

$$\text{MAD}_{\text{Abinit,80 Ha,120 a}_b\text{,no CC}}^{\text{Abinit,400 Ha,60 a}_b\text{,CC}} = 6.3 \, \text{meV}, \tag{4.5}$$

the increased MAD indicates, that the calculations with CC and a high cutoff energy could either yield better convergence with the CP2K calcualtions or worse. In order to obtain certainty, the latest calculations are compared to the CP2K calculations, which yields a MAD of

$$\text{MAD}_{\text{Abinit,400 Ha,60 a}_b\text{,CC}}^{\text{CP2K,TZVP+aug-cc-QZVP}} = 2.4 \, \text{meV}, \tag{4.6}$$

which increased the convergence compared to the non-cutoff calculations by more than 300%. In graph 4.2 the difference in IE for each molecule of the entire GW-100 set is illustrated.
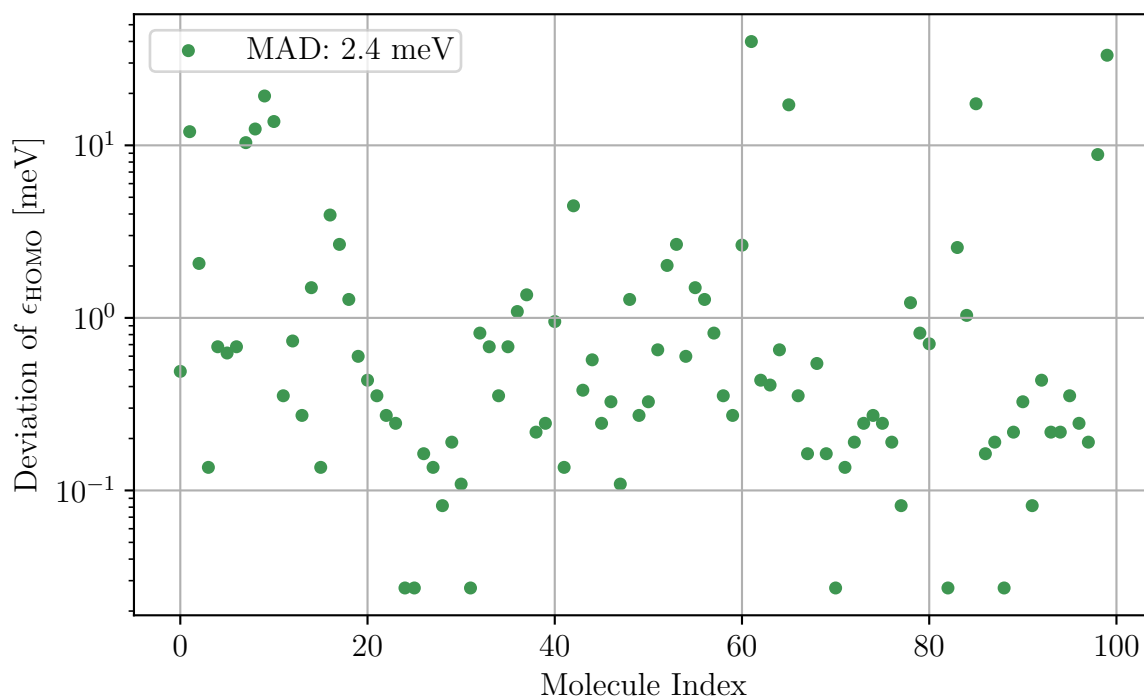
Figure 4.2: Plot of the deviation of ϵ_{HOMO} between Abinit and CP2K for the GW-100 molecule set. The order of the molecules on the x-axis is arbitrary, they are arranged in accordance with the sequence presented in table A.1. The MAD is 2.4 meV.

4.1.2 Non-Convergence of TZVP-MOLOPT Basis

As demonstrated in the following papers [14], [15] and [5], the TZVP-MOLOPT basis should be adequate for convergent calculations. However, upon utilizing the TZVP-MOLOPT basis, it was observed that convergence could not be achieved to a level that approximated the aforementioned convergence depicted. Comparing the CP2K TZVP-MOLOPT calculations $C_6^{\text{CP2K,TZVP}}$ with the

Abinit calculations $C_5^{\text{Abinit,400 Ha,60 a}_b}$ yields a MAD of

$$\text{MAD}_{\text{Abinit,400 Ha,60 a}_b\text{,CC}}^{\text{CP2K,TZVP}} = 48.1 \, \text{meV}, \tag{4.7}$$

It has been demonstrated that the convergence of the CP2K TZVP+aug-cc-QZVP calculations is exceeded by a factor of 20. This finding indicates that the TZVP-MOLOPT basis from paper [14] may not have fully converged, at least for the molecules contained in the GW-100 set.

4.2 LUMO

The following chapter presents a discussion of **L**owest **u**noccupied **m**olecular **o**rbital, which is analogous to that of the Highest occupied molecular orbital. Initially, the results are presented, followed by an analysis of their significance and meaningfulness. In 4.3 the LUMO energies of CP2K and Abinit are depicted next to each other and a rather interesting picture is showing. In contrast to the HOMO energies, which exhibited a high degree of similarity (max. ~ 40 meV deviation), the LUMO energies displayed a considerable degree of variation, with some even exhibiting different signs. Especially the LUMO energies with different signs have a huge deviation.
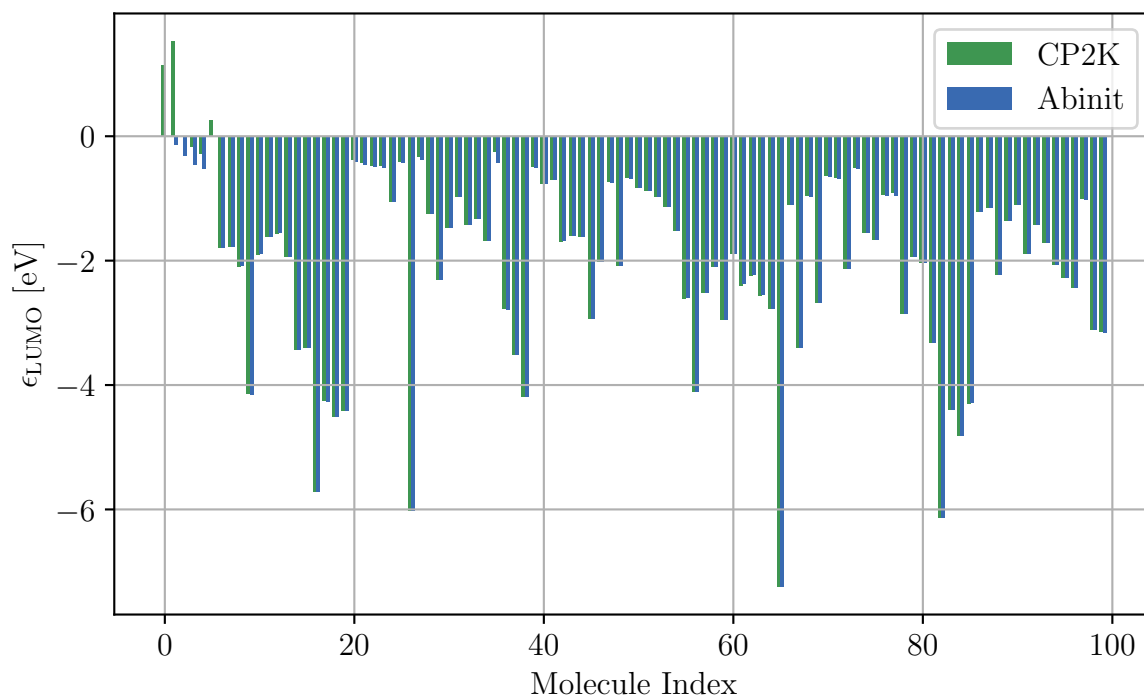
Figure 4.3: LUMO energy of Abinit and CP2K next to each other for GW-100 molecule set. The order of the molecules on the x-axis is arbitrary, they are arranged in accordance with the sequence presented in table A.1. Unbound states display a high deviation, strongly bound states a low deviation.

A more detailed examination of the LUMO energies of Abinit, which are in close proximity

to zero, reveals that the LUMO energies of CP2K predominantly exhibit an inverse sign and significant deviation. In contrast, strongly negative LUMO energies show very little deviation between the two algorithms. This finding indicates that strongly bound states exhibit only minor deviations, whereas unbound or weakly bound states are associated with significant deviations. In order to explain that behaviour one has to recall which basis sets Abinit and CP2K are using. Since Abinit uses a plane-wave basis sets, the physics of unbound states can be described more accurately as explained in chapter 2.4. Attempting to describe an unbound state with Gaussian functions is a difficult task, as it requires the stacking of an infinite number of Gaussians, which is why CP2K utilizing a Gaussian basis set fails to describe the unbound LUMOs. To validate this hypothesis we plotted the LUMO energies derived from Abinit against the LUMO energy difference of Abinit and CP2K, as seen in figure 4.4.

As expected, the Deviation of the LUMO energy follows an exponential trend. For the strongly bound states starting form roughly $-8\,\mathrm{eV}$ up to $-1\,\mathrm{eV}$ the MAD is

$$\mathrm{MAD}_{<-1\,\mathrm{eV}} = 3.1\,\mathrm{meV}, \tag{4.8}$$

which comes close to the MAD of our HOMO calculations. For weakly bound states, we can observe an increase in the deviation. If we include the weakly bound states into the MAD, we yield

$$\mathrm{MAD}_{\mathrm{Abinit},400\,\mathrm{Ha},60\,\mathrm{a_b},\mathrm{CC}}^{\mathrm{CP2K,TZVP+aug\text{-}cc\text{-}QZVP}} = 47.1\,\mathrm{meV}, \tag{4.9}$$

which is a substantial increase compared to the HOMO MAD. As discussed previously, this can be attributed to the limited ability of Gaussians to represent weakly bound states.
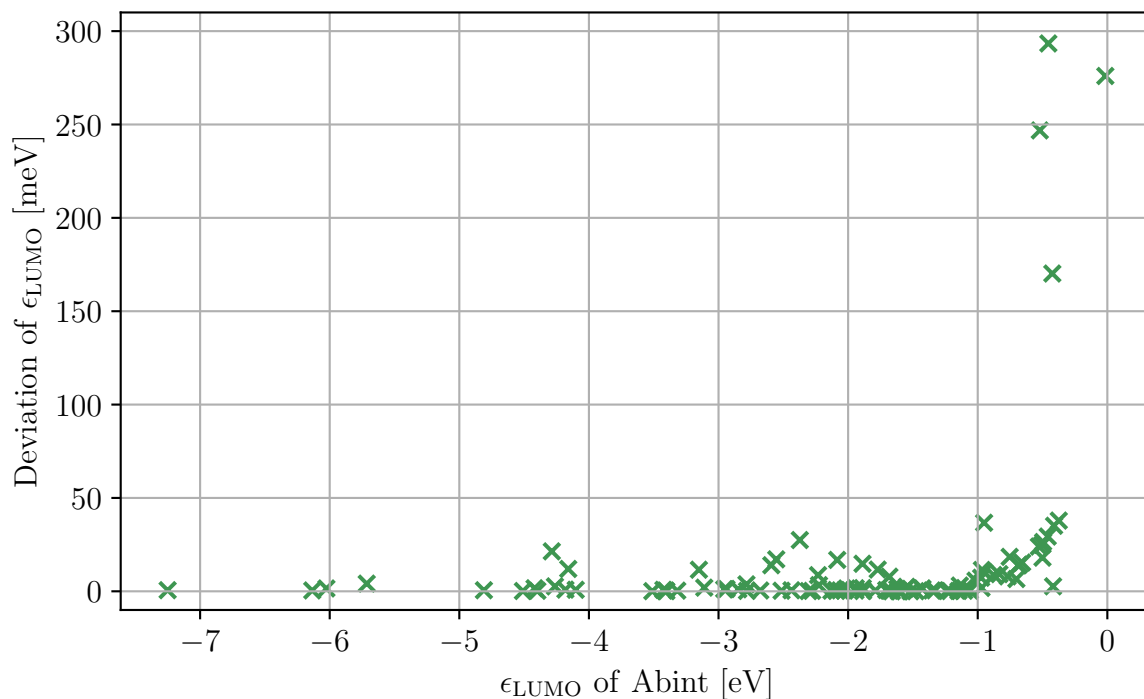
Figure 4.4: The deviation in ϵ_{LUMO} between Abinit and CP2K is illustrated as a function of LUMO energies of Abint. The values for Neon $\Delta\epsilon_{\text{HOMO}}^{\text{Ne}} = 1670.0\,\text{meV}$ and Helium $\Delta\epsilon_{\text{HOMO}}^{\text{He}} = 1155.4\,\text{meV}$ are not displayed in the plot, since their deviation is so big, including them would render the plot unreadable. In the case of unbound/weakly bound states, which correspond to high LUMO energies as calculated by Abinit, the deviation is observed to increase exponentially. For Abinit LUMO energies smaller than $-1\,\text{eV}$ the mean deviation is 3.1 meV, the deviation gets smaller as we get closer to the strongly bound states. The mean deviation of the whole molecule set is 47.1 meV.

Chapter 5

Conclusion and Outlook

The initial premise of the study was to recreate the accuracy of the GW-100 Benchmarking. Following a process of fine-tuning the input parameters for our algorithms, this objective was successfully achieved. The final result yields a mean absolute deviation (MAD) of 2.4 meV, indicating that the utilization of Gaussians and plane waves does not significantly impact the accuracy of calculations pertaining to the HOMO energies of molecules.

However, a significant disparity is observed in the domain of computing costs. The mean core-hour utilization for a single molecule simulation in Abinit amounts to 396.8 core hours, whereas on average a simulation in CP2K necessitats 2.5 core hours, marking a 158-fold discrepancy. It is for this reason that plane waves are only of limited use for obtaining molecular HOMO energies. It must be acknowledged that the Abinit calculations could be executed with a greater degree of reduction in scale. During the convergence tests utilizing the Coulomb cutoff, calculations were performed for five molecules with a considerably reduced cutoff energy of 200 Ha. A comparison of the HOMO energies of these five molecules yields the following results

$$\mathrm{MAD}_{\mathrm{Abinit,400\,Ha,60\,a_b,CC}}^{\mathrm{Abinit,200\,Ha,60\,a_b,CC}} \leq 0.1\,\mathrm{meV}. \tag{5.1}$$

Therefore, there is a high probability that $C_7^{\mathrm{Abinit,200\,Ha,60\,a_b,CC}}$ could achive the same convergence with the CP2K calculations $C_0^{\mathrm{CP2K,TZVP+aug\text{-}cc\text{-}QZVP}}$ as $C_5^{\mathrm{Abinit,400\,Ha,60\,a_b,CC}}$. However, this hypothesis needs to be tested in detail, as it is possible that the convergence observed in the tests is merely a coincidence. Even with the reduced cutoff energy of 200 Ha, the plane wave Abinit calculations are still considerably more expensive than the CP2K calculations.

In the context of plane wave calculations, a notable advantage over Gaussian calculations is observed in the treatment of weakly bound states. As illustrated in 4.3, the Gaussians underperform significantly in this scenario. Consequently, when undertaking DFT calculations that involve weakly bound states, the utilization of a plane wave basis set may offer a distinct advantage. To summarize this thesis, the two different algorithms were optimized to a very high degree of convergence, which, with simple means, is unlikely to be significantly improved.

Appendix A

Additional data

The subsequent chapter comprises tabulated data illustrating the outcomes of the most significant calculations, specifically those pertaining to HOMO (A.1) and LUMO (A.2).

Abinit$_{C_1}$ refers to the Abinit calculations $C_1^{\text{Abinit,80 Ha,120 a}_b\text{,no CC}}$ featuring a unit cell size of $120\,a_b$, a cutoff energy of $80\,\text{Ha}$ and no coulomb cutoff.

Abinit$_{C_5}$ refers to the Abinit calculations $C_5^{\text{Abinit,400 Ha,60 a}_b\text{,CC}}$ featuring a unit cell size of $60\,a_b$, a cutoff energy of $400\,\text{Ha}$ and added coulomb cutoff.

CP2K$_{C_0}$ refers to the CP2K calculations $C_0^{\text{CP2K,TZVP+aug-cc-QZVP}}$ which utilizes a combination of the TZVP-MOLOPT Gaussian Basis and the aug-cc-QZVP Basis.

CP2K$_{C_6}$ refers to the CP2K calculations $C_6^{\text{CP2K,TZVP}}$ using the TZVP-MOLOPT Gaussian Basis.

Table A.1: HOMO Energies [eV] of the GW100 Set Calculated with Abinit using plane waves and CP2K using Gaussians.

	name	formula	Abinit$_{C_1}$	Abinit$_{C_5}$	CP2K$_{C_0}$	CP2K$_{C_6}$
1	helium	He	-15.742	-15.759	-15.758	-15.756
2	neon	Ne	-13.266	-13.332	-13.344	-13.331
3	argon	Ar	-10.268	-10.268	-10.266	-10.265
4	krypton	Kr	-9.245	-9.246	-9.245	-9.235
5	xenon	Xe	-8.234	-8.235	-8.236	-8.243
6	hydrogen	H$_2$	-10.374	-10.376	-10.376	-10.299
7	lithium dimer	Li$_2$	-3.224	-3.225	-3.225	-3.200
8	sodium dimer	Na$_2$	-3.139	-3.122	-3.132	-3.104
9	sodium tetramer	Na$_4$	-2.685	-2.675	-2.688	-2.663
10	sodium hexamer	Na$_6$	-2.994	-2.982	-2.995	-2.991
11	potassium dimer	K$_2$	-2.568	-2.571	-2.570	-2.524
12	rubidium dimer	Rb$_2$	-2.479	-2.481	-2.482	-2.400
13	nitrogen	N$_2$	-10.274	-10.276	-10.276	-10.283

continued on next page

Table A.1 continued from previous page

	name	formula	Abinit$_{C_1}$	Abinit$_{C_5}$	CP2K$_{C_0}$	CP2K$_{C_6}$
14	phosphorus dimer	P_2	-7.108	-7.110	-7.112	-7.124
15	arsenic dimer	As_2	-6.499	-6.501	-6.501	-6.446
16	fluorine	F_2	-9.418	-9.459	-9.463	-9.428
17	chlorine	Cl_2	-7.281	-7.283	-7.280	-7.235
18	bromine	Br_2	-6.795	-6.796	-6.795	-6.794
19	iodine	I_2	-6.251	-6.252	-6.253	-6.246
20	methane	CH_4	-9.448	-9.450	-9.449	-9.419
21	ethane	C_2H_6	-8.150	-8.153	-8.152	-8.111
22	propane	C_3H_8	-7.745	-7.748	-7.747	-7.701
23	butane	C_4H_{10}	-7.569	-7.572	-7.572	-7.524
24	ethylene	C_2H_4	-6.759	-6.760	-6.760	-6.709
25	ethyne	C_2H_2	-7.185	-7.186	-7.186	-7.160
26	tetracarbon	C_4	-7.255	-7.257	-7.257	-7.202
27	cyclopropane	C_3H_6	-7.046	-7.047	-7.047	-7.014
28	benzene	C_6H_6	-6.319	-6.322	-6.322	-6.268
29	cyclooctatetraene	C_8H_8	-5.287	-5.291	-5.291	-5.219
30	cyclopentadiene	C_5H_6	-5.390	-5.393	-5.393	-5.329
31	vinyl fluoride	C_2H_3F	-6.526	-6.536	-6.536	-6.493
32	vinyl chloride	C_2H_3Cl	-6.422	-6.424	-6.424	-6.371
33	vinyl bromide	C_2H_3Br	-5.822	-5.824	-5.823	-5.774
34	vinyl iodide	C_2H_3I	-6.037	-6.038	-6.039	-5.994
35	tetrafluoromethane	CF_4	-10.380	-10.417	-10.418	-10.378
36	tetrachloromethane	CCl_4	-7.650	-7.653	-7.652	-7.613
37	tetrabromomethane	CBr_4	-6.930	-6.934	-6.933	-6.907
38	tetraiodomethane	CI_4	-6.191	-6.195	-6.195	-6.165
39	silane	SiH_4	-8.524	-8.526	-8.526	-8.509
40	germane	GeH_4	-8.347	-8.349	-8.348	-8.334
41	disilane	Si_2H_6	-7.290	-7.293	-7.293	-7.276
42	pentasilane	Si_5H_{12}	-6.575	-6.575	-6.580	-6.564
43	lithium hydride	LiH	-4.361	-4.361	-4.361	-4.369
44	potassium hydride	KH	-3.477	-3.478	-3.477	-3.487
45	borane	BH_3	-8.494	-8.496	-8.495	-8.470
46	diborane	B_2H_6	-7.871	-7.874	-7.873	-7.855
47	ammonia	NH_3	-6.169	-6.169	-6.169	-6.064
48	hydrazoic acid	HN_3	-6.815	-6.817	-6.819	-6.791
49	phosphine	PH_3	-6.728	-6.729	-6.729	-6.698

Table A.1 continued from previous page

	name	formula	Abinit$_{C_1}$	Abinit$_{C_5}$	CP2K$_{C_0}$	CP2K$_{C_6}$
50	arsine	AsH_3	-6.787	-6.788	-6.788	-6.723
51	hydrogen sulfide	SH_2	-6.290	-6.291	-6.291	-6.187
52	hydrogen fluoride	HF	-9.642	-9.652	-9.654	-9.597
53	hydrogen chloride	HCl	-8.029	-8.030	-8.027	-7.963
54	lithium fluoride	LiF	-6.127	-6.122	-6.123	-6.032
55	magnesium fluoride	MgF_2	-8.291	-8.297	-8.299	-8.206
56	titanium tetrafluoride	TiF_4	-10.411	-10.429	-10.431	-10.409
57	aluminum fluoride	AlF_3	-9.677	-9.691	-9.692	-9.663
58	boron monofluoride	BF	-6.763	-6.772	-6.772	-6.709
59	sulfur tetrafluoride	SF_4	-8.210	-8.244	-8.244	-8.191
60	potassium bromide	BrK	-4.725	-4.726	-4.724	-4.662
61	gallium monochloride	GaCl	-6.613	-6.615	-6.575	-6.548
62	sodium chloride	NaCl	-5.289	-5.280	-5.280	-5.202
63	magnesium chloride	$MgCl_2$	-7.618	-7.621	-7.620	-7.587
64	aluminum iodide	AlI_3	-6.627	-6.630	-6.631	-6.623
65	boron nitride	BN	-7.473	-7.474	-7.492	-7.448
66	hydrogen cyanide	NCH	-9.025	-9.025	-9.026	-9.006
67	phosphorus mononitride	PN	-7.751	-7.752	-7.752	-7.728
68	hydrazine	N_2H_4	-5.281	-5.283	-5.282	-5.198
69	formaldehyde	H_2CO	-6.273	-6.278	-6.278	-6.206
70	methanol	CH_3OH	-6.350	-6.353	-6.353	-6.272
71	ethanol	C_2H_5OH	-6.161	-6.165	-6.165	-6.078
72	acetaldehyde	C_2H_4O	-5.979	-5.984	-5.984	-5.888
73	ethoxy ethane	$C_4H_{10}O_4$	-5.800	-5.805	-5.805	-5.707
74	formic acid	HCOOH	-6.947	-6.952	-6.952	-6.876
75	hydrogen peroxide	H_2O_2	-6.461	-6.468	-6.468	-6.416
76	water	H_2O	-7.263	-7.265	-7.265	-7.189
77	carbon dioxide	CO_2	-9.080	-9.088	-9.088	-9.057
78	carbon disulfide	CS_2	-6.788	-6.790	-6.791	-6.745
79	carbon oxide sulfide	OCS	-7.466	-7.471	-7.472	-7.435
80	carbon oxide selenide	OCSe	-6.933	-6.938	-6.938	-6.913
81	carbon monoxide	CO	-9.332	-9.338	-9.338	-9.256
82	ozone	O_3	-7.956	-7.966	-7.966	-7.930
83	sulfur dioxide	SO_2	-8.093	-8.100	-8.098	-8.062
84	beryllium monoxide	BeO	-6.138	-6.144	-6.145	-5.900
85	magnesium monoxide	MgO	-4.811	-4.815	-4.832	-4.761

continued on next page

Table A.1 continued from previous page

	name	formula	Abinit$_{C_1}$	Abinit$_{C_5}$	CP2K$_{C_0}$	CP2K$_{C_6}$
86	toluene	C_7H_8	-5.987	-5.991	-5.990	-5.936
87	ethylbenzene	C_8H_{10}	-5.991	-5.995	-5.995	-5.938
88	hexafluorobenzene	C_6F_6	-6.619	-6.653	-6.653	-6.642
89	phenol	C_6H_5OH	-5.625	-5.629	-5.629	-5.571
90	aniline	$C_6H_5NH_2$	-5.014	-5.018	-5.017	-4.945
91	pyridine	C_5H_5N	-5.931	-5.933	-5.933	-5.853
92	guanine	$C_5H_5N_5O$	-5.287	-5.293	-5.292	-5.224
93	adenine	$C_5H_5N_5O$	-5.523	-5.527	-5.526	-5.451
94	cytosine	$C_4H_5N_3O$	-5.717	-5.721	-5.721	-5.641
95	thymine	$C_5H_6N_2O_2$	-6.041	-6.048	-6.048	-5.984
96	uracil	$C_4H_4N_2O_2$	-6.278	-6.284	-6.284	-6.199
97	urea	CH_4N_2O	-5.935	-5.938	-5.937	-5.836
98	silver dimer	Ag_2	-5.198	-5.200	-5.209	-5.169
99	copper dimer	Cu_2	-4.650	-4.673	-4.707	-4.603
100	copper cyanide	$CuCN$	-6.774	-6.747	-6.726	-6.701

Table A.2: LUMO Energies [eV] of the GW100 Set Calculated with Abinit using plane waves and CP2K using Gaussians.

	name	formula	Abinit_{C_1}	Abinit_{C_5}	CP2K_{C_0}	CP2K_{C_6}
1	helium	He	-0.004	-0.007	1.149	14.123
2	neon	Ne	-0.149	-0.135	1.535	14.119
3	argon	Ar	-0.319	-0.320	0.000	11.007
4	krypton	Kr	-0.455	-0.455	-0.162	6.372
5	xenon	Xe	-0.520	-0.521	-0.274	2.303
6	hydrogen	H_2	-0.005	-0.016	0.260	1.721
7	lithium dimer	Li_2	-1.789	-1.790	-1.789	-1.721
8	sodium dimer	Na_2	-1.782	-1.772	-1.783	-1.698
9	sodium tetramer	Na_4	-2.100	-2.084	-2.101	-2.027
10	sodium hexamer	Na_6	-1.893	-1.889	-1.904	-1.844
11	potassium dimer	K_2	-1.608	-1.610	-1.610	-1.554
12	rubidium dimer	Rb_2	-1.558	-1.559	-1.561	-1.462
13	nitrogen	N_2	-1.940	-1.943	-1.944	-1.909
14	phosphorus dimer	P_2	-3.423	-3.425	-3.426	-3.412
15	arsenic dimer	As_2	-3.401	-3.403	-3.403	-3.335
16	fluorine	F_2	-5.654	-5.714	-5.719	-5.719
17	chlorine	Cl_2	-4.260	-4.262	-4.259	-4.190
18	bromine	Br_2	-4.508	-4.509	-4.509	-4.605
19	iodine	I_2	-4.418	-4.420	-4.421	-4.520
20	methane	CH_4	-0.411	-0.411	-0.376	2.499
21	ethane	C_2H_6	-0.459	-0.460	-0.431	1.518
22	propane	C_3H_8	-0.494	-0.496	-0.469	1.326
23	butane	C_4H_{10}	-0.502	-0.503	-0.479	1.114
24	ethylene	C_2H_4	-1.060	-1.060	-1.060	-0.917
25	ethyn	C_2H_2	-0.418	-0.419	-0.416	-0.225
26	tetracarbon	C_4	-6.024	-6.025	-6.027	-6.019
27	cyclopropane	C_3H_6	-0.374	-0.375	-0.337	1.412
28	benzene	C_6H_6	-1.236	-1.239	-1.238	-1.127
29	cyclooctatetraene	C_8H_8	-2.302	-2.306	-2.306	-2.212
30	cyclopentadiene	C_5H_6	-1.476	-1.478	-1.478	-1.370
31	vinyl fluoride	C_2H_3F	-0.964	-0.971	-0.969	-0.845
32	vinyl chloride	C_2H_3Cl	-1.425	-1.427	-1.425	-1.291
33	vinyl bromide	C_2H_3Br	-1.327	-1.329	-1.328	-1.174
34	vinyl iodide	C_2H_3I	-1.678	-1.679	-1.680	-1.558

continued on next page

Table A.2 continued from previous page

	name	formula	Abinit$_{C_1}$	Abinit$_{C_5}$	CP2K$_{C_0}$	CP2K$_{C_6}$
35	tetrafluoromethane	CF_4	-0.429	-0.424	-0.254	2.471
36	tetrachloromethane	CCl_4	-2.780	-2.785	-2.781	-2.607
37	tetrabromomethane	CBr_4	-3.507	-3.511	-3.511	-3.500
38	tetraiodomethane	CI_4	-4.177	-4.182	-4.182	-4.150
39	silane	SiH_4	-0.500	-0.500	-0.482	0.375
40	germane	GeH_4	-0.767	-0.768	-0.760	0.611
41	disilane	Si_2H_6	-0.700	-0.702	-0.695	-0.500
42	pentasilane	Si_5H_{12}	-1.687	-1.682	-1.690	-1.603
43	lithium hydride	LiH	-1.600	-1.601	-1.601	-1.479
44	potassium hydride	KH	-1.619	-1.621	-1.622	-1.591
45	borane	BH_3	-2.938	-2.939	-2.940	-2.887
46	diborane	B_2H_6	-2.010	-2.011	-2.012	-1.971
47	ammonia	NH_3	-0.753	-0.753	-0.735	0.860
48	hydrazoic acid	HN_3	-2.086	-2.089	-2.090	-2.027
49	phosphine	PH_3	-0.687	-0.687	-0.674	-0.058
50	arsine	AsH_3	-0.832	-0.833	-0.823	-0.037
51	hydrogen sulfide	SH_2	-0.877	-0.878	-0.869	-0.349
52	hydrogen fluoride	HF	-0.967	-0.974	-0.966	0.096
53	hydrogen chloride	HCl	-1.137	-1.139	-1.135	-0.450
54	lithium fluoride	LiF	-1.516	-1.519	-1.519	-1.447
55	magnesium fluoride	MgF_2	-2.625	-2.593	-2.607	-2.490
56	titanium tetrafluoride	TiF_4	-4.065	-4.102	-4.103	-4.091
57	aluminum fluoride	AlF_3	-2.506	-2.514	-2.513	-2.412
58	boron monofluoride	BF	-2.080	-2.091	-2.092	-2.019
59	sulfur tetrafluoride	SF_4	-2.923	-2.956	-2.955	-2.889
60	potassium bromide	BrK	-1.884	-1.887	-1.889	-1.893
61	gallium monochloride	$GaCl$	-2.371	-2.373	-2.400	-2.368
62	sodium chloride	$NaCl$	-2.233	-2.232	-2.241	-2.240
63	magnesium chloride	$MgCl_2$	-2.586	-2.555	-2.572	-2.510
64	aluminum iodide	AlI_3	-2.777	-2.780	-2.780	-2.782
65	boron nitride	BN	-7.249	-7.250	-7.250	-7.226
66	hydrogen cyanide	NCH	-1.100	-1.101	-1.100	-0.962
67	phosphorus mononitride	PN	-3.406	-3.408	-3.407	-3.354
68	hydrazine	N_2H_4	-0.967	-0.968	-0.956	0.283
69	formaldehyde	H_2CO	-2.672	-2.678	-2.677	-2.602
70	methanol	CH_3OH	-0.656	-0.657	-0.642	0.542

Table A.2 continued from previous page

	name	formula	Abinit$_{C_1}$	Abinit$_{C_5}$	CP2K$_{C_0}$	CP2K$_{C_6}$
71	ethanol	C_2H_5OH	-0.679	-0.680	-0.666	0.557
72	acetaldehyde	C_2H_4O	-2.124	-2.129	-2.129	-2.032
73	ethoxy ethane	$C_4H_{10}O_4$	-0.527	-0.528	-0.504	1.092
74	formic acid	$HCOOH$	-1.542	-1.548	-1.546	-1.473
75	hydrogen peroxide	H_2O_2	-1.653	-1.663	-1.663	-1.528
76	water	H_2O	-0.949	-0.950	-0.940	0.261
77	carbon dioxide	CO_2	-0.948	-0.951	-0.914	-0.301
78	carbon disulfide	CS_2	-2.845	-2.847	-2.846	-2.751
79	carbon oxide sulfide	OCS	-1.936	-1.941	-1.940	-1.863
80	carbon oxide selenide	$OCSe$	-2.022	-2.027	-2.027	-1.992
81	carbon monoxide	CO	-3.312	-3.318	-3.319	-3.271
82	ozone	O_3	-6.117	-6.135	-6.135	-6.117
83	sulfur dioxide	SO_2	-4.389	-4.398	-4.398	-4.357
84	beryllium monoxide	BeO	-4.800	-4.810	-4.810	-4.640
85	magnesium monoxide	MgO	-4.247	-4.286	-4.308	-4.231
86	toluene	C_7H_8	-1.209	-1.212	-1.212	-1.097
87	ethylbenzene	C_8H_{10}	-1.145	-1.149	-1.148	-1.037
88	hexafluorobenzene	C_6F_6	-2.196	-2.225	-2.222	-1.846
89	phenol	C_6H_5OH	-1.349	-1.352	-1.352	-1.243
90	aniline	$C_6H_5NH_2$	-1.102	-1.105	-1.104	-0.985
91	pyridine	C_5H_5N	-1.881	-1.884	-1.884	-1.783
92	guanine	$C_5H_5N_5O$	-1.423	-1.429	-1.428	-1.316
93	adenine	$C_5H_5N_5O$	-1.702	-1.706	-1.705	-1.609
94	cytosine	$C_4H_5N_3O$	-2.056	-2.061	-2.060	-1.967
95	thymine	$C_5H_6N_2O_2$	-2.271	-2.278	-2.277	-2.189
96	uracil	$C_4H_4N_2O_2$	-2.43	-2.437	-2.437	-2.355
97	urea	CH_4N_2O	-1.015	-1.017	-1.011	0.204
98	silver dimer	Ag_2	-3.108	-3.110	-3.112	-3.061
99	copper dimer	Cu_2	-3.261	-3.151	-3.140	-3.100
100	copper cyanide	$CuCN$	-4.247	-4.161	-4.148	-4.131

Bibliography

[1] M. R. Filip, G. E. Eperon, H. J. Snaith, and F. Giustino, *Steric engineering of metal-halide perovskites with tunable optical band gaps*, Nature communications **5**, 5757 (2014).

[2] W. Shockley and H. J. Queisser, *Detailed Balance Limit of Efficiency of p-n Junction Solar Cells*, Journal of Applied Physics **32**, 510–519 (1961).

[3] N. Argaman and G. Makov, *Density functional theory: An introduction*, American Journal of Physics **68**, 69–79 (2000).

[4] M. Orio, D. A. Pantazis, and F. Neese, *Density functional theory*, Photosynthesis research **102**, 443–453 (2009).

[5] M. J. Van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, *et al.*, *GW100: Benchmarking G0W0 for Molecular Systems*, J. Chem. Theory Comput. **11**, 5665–5687 (2015).

[6] J. Wilhelm, *computational electronic structure theory of solids and applications – Lecture Notes –*, , 2021.

[7] F. Bechstedt, *Many-Body Approach to Electronic Excitations*, Springer-Verlag Berlin Heidelberg, 2015.

[8] R. M. Martin, *Electronic Structure Basic Theory and Practical Methods*, Cambridge University Press, 2021.

[9] J. F. Janak, *Proof that $\frac{\partial E}{\partial n_i} = \epsilon$ in density-functional theory*, Phys. Rev. B **18**, 7165–7168 (1978).

[10] F. Jensen, *Introduction to computational chemistry*, John wiley & sons, 2007.

[11] M. Krack. *Goedecker-Teter-Hutter (GTH) pseudopotentials*. 2021. URL: https://htmlpreview.github.io/?https://github.com/cp2k/cp2k-data/blob/master/potentials/Goedecker/index.html.

[12] C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, *Exact Coulomb cutoff technique for supercell calculations*, Phys. Rev. B **73**, 205119 (2006).

[13] ABINIT. *Second (basic) tutorial*. Accessed: 2025-03-16. 2025. URL: https://docs.abinit.org/tutorial/base2/.

[14] J. VandeVondele and J. Hutter, *Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases*, The Journal of Chemical Physics **127**, 114105 (2007).

[15] J. Wilhelm, D. Golze, L. Talirz, J. Hutter, and C. A. Pignedoli, *Toward GW Calculations on Thousands of Atoms*, J. Phys. Chem. Lett. **9**, 306–312 (2018).

[16] J. Dunning Thom H., *Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen*, The Journal of Chemical Physics **90**, 1007–1023 (1989).

[17] R. A. Kendall, J. Dunning Thom H., and R. J. Harrison, *Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions*, The Journal of Chemical Physics **96**, 6796–6806 (1992).

Acknowledgments

This thesis would not have been possible without the support of the people around me. I would like to especially thank Mia Schambeck and Dr. Jan Wilhelm, who supported me throughout my research and writing process. Additionally, I would like to thank the entire team at Prof. Dr. Evers' chair, who warmly welcomed me and was always supportive in every regard.

Declaration of Authorship

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen Hilfsmittel als die angegebenen verwendet habe. Die Stellen, die anderen Werken (gilt ebenso für Werke aus elektronischen Datenbanken oder aus dem Internet) wörtlich oder sinngemäß entnommen sind, habe ich unter Angabe der Quelle und Einhaltung der Regeln wissenschaftlichen Zitierens kenntlich gemacht. Über wissenschaftlich korrektes Arbeiten wurde ich aufgeklärt. Diese Versicherung umfasst auch in der Arbeit verwendete bildliche Darstellungen, Tabellen, Kartenskizzen und gelieferte Zeichnungen. Ebenso versichere ich, dass die vorgelegten Druckexemplare und die abgegebene elektronische Version der Arbeit identisch sind, sowie, dass ich diese Arbeit nicht bereits an einer anderen Universität zur Erlangung eines akademischen Grades eingereicht habe. Mir ist bewusst, dass Täuschungen nach der für mich gültigen Studien- und Prüfungsordnung geahndet werden. Ich habe Kenntnis von den in § 26 Abs. 5 vorgesehenen Rechtsfolgen.

Unterschrift   Ort, Datum