First-Principles Basin-Hopping for the Structure Determination of Atomic Clusters

Ralf Gehrke
Diplom-Physiker

**im Fachbereich Physik
der Freien Universität Berlin eingereichte
Inaugural-Dissertation
zur Erlangung des akademischen Grades
Doctor Rerum Naturalium**

**Berlin
November 2008**

Diese Arbeit wurde in der Zeit von Dezember 2004 bis November 2008 unter der Aufsicht von PD Dr. Karsten Reuter in der Theorieabteilung des Fritz-Haber-Institutes der Max-Planck-Gesellschaft durchgeführt.

Erstgutachter: PD Dr. Karsten Reuter

Zweitgutachter: Prof. Dr. Eberhard K.U. Gross

Disputationstermin: 26. Januar 2009

Abstract

Research on small particles containing up to a few tens of atoms is largely driven by their novel properties that are significantly affected by quantum effects, particularly in the interplay between structural and electronic degrees of freedom. Such clusters, thus, carry the potential of major technological advances for applications exploiting their already exemplified unique optical, magnetic, and chemical properties. Atomically resolved structural information is a key prerequisite towards employing these envisioned functionalities, considering that the latter will be tailored to the atomic scale. In this respect not only the ground state isomer will be of importance, but potentially all energetically low-lying metastable isomers.

A materials modeling targeting the identification of such relevant cluster isomers involves the global and local exploration of the corresponding vast configuration space, represented by the high-dimensional potential-energy surface (PES). The exponential growth of the number of local PES minima, i.e. metastable isomers, with increasing cluster size quickly limits approaches focusing only on structural motifs provided by chemical intuition.

In this thesis the problem is addressed with a first-principles Monte Carlo approach. The essential features herein are a quantitative quantum mechanical energetics from density-functional theory, together with a basin-hopping type sampling for the global exploration of the energy surface. This approach is applied to $Co_n^+Ar_m$ clusters, for which far-infrared vibrational spectroscopy data were provided by the group of Prof. Gerard Meijer at the Molecular Physics Department of the FHI. For the isomers obtained by the implemented first-principles sampling scheme, we calculated the infrared spectra, that are in turn compared to the experimental data. Apart from the thus gained structure information, we particularly elaborated on the role of the involved Ar probe atoms and its implication for the indirect structure determination using the vibrational fingerprints.

To suggest possible strategies for optimizing the numerically intense *ab initio* based sampling, the involved technical settings are analyzed in detail, using small silicon and copper clusters as illustrative benchmark systems. In order to assess and critically discuss the limitations and bottlenecks of the scheme, a central aspect of this part of the work was to establish a framework to quantitatively measure the sampling efficiency. The detailed insights obtained are then used to devise a robust self-adapting scheme for future applications in particular to systems involving intermediate size clusters.

“So einfach wie möglich. Aber nicht einfacher.”

Albert Einstein

Contents

List of Figures

List of Tables

1. Introduction

Clusters are aggregates of atoms and thus somewhere intermediate in size between individual atoms and bulk matter. One major characteristics that separates them from the bulk is the dramatic change in the electronic structure when the system size is reduced to few atoms, thus replacing the quasi-continuous density of states by a discrete energy level spectrum. Due to the intricate relationship between structural and electronic degrees of freedom, optical, magnetic and chemical properties do therefore not vary smoothly with size and shape of the cluster but are highly governed by quantum (size) effects. The study of these effects constitutes one field of cluster physics, since they open the door to novel applications, such as small electronic devices. A second important property of clusters is the large fraction of atoms being on the surface. Due to this favourable surface to volume ratio, metal clusters are for instance ideal candidates for applications in catalysis. Atomically resolved structural information is thereby a key prerequisite towards employing these envisioned functionalities, considering that the latter will be tailored to the atomic scale.

Experimentally, structural information can for instance be obtained by photoelectron spectroscopy, which exploits the fact that different isomers exhibit different electronic structures. The measured density of states can therefore serve as a characteristic fingerprint for the underlying atomic structure. In the case of reactivity experiments, the number of adsorbed atoms on a cluster is measured, which is obviously correlated with the number of adsorption sites and therefore with the cluster shape. Another probe of cluster structure is provided by mobility experiments in which clusters are accelerated by an electric field and then traverse a chamber containing an inert gas. The travelling time is hereby affected by the collisions of the clusters with the inert gas and thus correlates with the collision cross section, being sensitive to the atomic structure. Another possibility that is particularly sensitive to the internal cluster structure is the measurement of vibrational frequencies. A corresponding technique that has recently been successfully employed to determine the structure of cationic and neutral metal clusters is far-infrared (vibrational) resonance enhanced multiple photon dissociation (FIR-MPD) spectroscopy [1, 2, 3, 4, 5]. Common to all these experimental methods is that atomic information is not directly measured, but only quantities that are coupled to the geometric configuration. At this point, theory comes into play and provides the required tool to assign the experimental data to an underlying geometry. As starting point for any such theoretical study, structural motifs are therefore required, which in the easiest approach are merely guessed, guided by chemical intuition. While for small systems, containing only few atoms such an approach might seem sufficient at first glance, it is in practice even then likely that one misses isomers. This holds in particular for systems exhibiting strong Jahn-Teller distortions, with a concomitant number of different isomers all corre-

sponding to the same basic structural motif. What is thus required is a more systematic scheme which then involves the global and local exploration of the huge configurational space, represented by the potential-energy surface (PES) [6]. A central and challenging property of the PES is thereby the suspected exponential increase of the number of local minima with system size. This drastic growth has led to the famous Levinthal's paradox according to which a protein would never reach its native state within the lifetime of the universe if it would have to go through all local PES minima completely randomly [7, 8].

Correspondingly, approaches addressing the determination of cluster structures face a twofold challenge, the first being the calculation of the PES itself. To have predictive power and in view of the intricate quantum-size effects ruling clusters in the targeted size range, a quantum-mechanical treatment of the PES is obviously required. Accurate quantum-mechanical methods become unfortunately prohibitively expensive due to their unfavourable scaling behaviour. The second order perturbation theory by Møller and Plesset (MP2) [9], for instance, scales as $O(N^5)$ with the system size, thus quickly becoming prohibitively expensive in connection with the exponentially increasing number of PES minima. As an alternative, density-functional theory scales formally as $O(N^3)$ and can be implemented to reach $O(N)$ for larger systems. It thus represents a suitable compromise between accuracy and computational demand and is the technique employed in the present work. The second major theoretical problem is the systematic and efficient exploration of the thus defined PES. A classical optimization method for this purpose is simulated annealing, in which successive trial moves, corresponding to random distortions of the cluster geometry are performed. Accepting or rejecting such moves via a Boltzmann-factor, the system is then driven towards the ground state structure. An additional key ingredient of many unbiased sampling schemes is a local structural relaxation, ensuing the trial move, which thereby reduces the PES sampling to the local minima and thus facilitates the optimization tremendously. In combination with simulated annealing, this scheme has first been proposed as Monte-Carlo minimization or basin-hopping by D. Wales and L. Scheraga [10]. A primary target for which basin-hopping is employed is to identify the global PES minimum, i.e. the most stable geometric structure in the focus of atomic clusters. However, in experiment, thermodynamic or kinetic reasons might lead to a population of metastable isomers or superpositions of it, thus making energetically higher-lying isomers subject to fundamental interest as well. By suitably adjusting the temperature controlled acceptance scheme in the simulated annealing, basin-hopping can be tuned to achieve this goal, thus identifying isomers within a thermodynamically reasonable energy range above the ground state. In the present work, a corresponding basin-hopping scheme has been coupled to the DFT implementation FHI-aims [11] currently developed a the FHI Theory Department, with the goal to perform a conformational sampling of cluster structures based on an accurate *ab initio* PES.

Aiming at interpreting the FIR-MPD data of Co_n^+Ar complexes measured in the group of Prof. Gerard Meijer at the Molecular Physics Department of the FHI, we first applied the developed tool to obtain candidate structures for the ensuing electronic structure studies. In an FIR-MPD experiment, the vibrational spectrum is measured by

far-infrared absorption spectroscopy using argon as probe atoms. Irradiated by a free-electron laser, the Co_n^+Ar complex is heated up when the laser frequency is in resonance with a vibrational eigenmode. Multiple photons can then be absorbed, which in turn leads to the dissociation of the argon atom from the metal cluster. Recording the resulting abundance changes of the rare-gas complexes as a function of the IR frequency yields the desired spectra that can finally be compared to computed IR absorption spectra for the different isomer structures obtained in the preceding sampling. Tacitly assuming that the influence of the probe atom on the vibrational fingerprint can be neglected, this enables the structure determination of the bare cobalt clusters and provides valuable feedback on the accuracy of the energetic ordering provided by the employed approximate DFT exchange-correlation functional. Contrary to previous studies on V_n^+ $(n = 3 - 23)$ [1, 2], Nb_n^+ $(n = 5 - 9)$ [3, 4] and Ta_n^+ $(n = 6 - 20)$ [5], however, a strong influence of the measured spectra on the number of adsorbed argon atoms has this time been observed, which motivated us to explicitly elaborate on the role of the probe atom and its implications for the structure determination. This then also involved exhaustive sampling runs, with argon ligands explicitly taken into account.

Despite the fact that due to continuing methodological improvements and vast increase of computational power *ab initio* basin-hopping runs have nowadays become possible, one still has to recognize that due to the exponential growth of the number of PES minima with system size, this method quickly reaches its limit. Unfortunately, many studies in the literature merely provide the size of the largest system studied as some kind of performance indicator of the employed sampling scheme, or at best record as benchmark the mean first encounter of the global minimum, i.e. the number of local relaxations until the ground state has been identified for the first time. What is more or less completely missing is an exact measure of efficiency that is coupled to the sampling goal used in a practical application, namely not only to find the ground state but also to identify higher-lying isomers. Motivated by the application in the first part of this work, the aim of the second part is therefore to establish a framework to quantitatively analyze the efficiency of a sampling run. Apart from identifying inefficient settings and suggesting possible optimization strategies, one has to note in this respect that already exactly pinning down the efficiency of an *ab initio* basin-hopping run is a task of its own. On the one hand one needs to define a reasonable measurable quantity which correlates with the sampling success. On the other hand, hundreds of basin-hopping runs are required to get statistically meaningful averages. This can quickly become unfeasible with straightforward DFT, whereas use of simple model-potentials is uncertain to provide answers for the practical materials-science applications. To circumvent this problem, we therefore developed the concept of a hopping matrix, which records the transition probabilities between individual isomers and thus enables the extraction of statistically meaningful quantities at tremendously reduced computational demand. Once the connectivity of the low-lying isomers is therewith represented, the hopping matrix provides the additional possibility to analyze different kind of efficiency indicators without having to run further first-principles basin-hopping runs and thus allows to further disentangle and understand the sampling process.

After introducing the different methodologies and concepts to describe and explore

the configurational space of atomic clusters in Part I, the employed DFT implementation FHI-aims is presented in Part II. Special emphasis is hereby given on the analytical atomic forces, which are required for an efficient local structural relaxation, and the implementation of which constituted a major, preparatory part of this thesis. Part III then describes the comparison between the experimental and computed Co_n^+Ar IR-spectra based upon geometries obtained by first-principles basin-hopping. Using small silicon and copper clusters as illustrative benchmarks, Part IV finally addresses the performance of the applied basin-hopping algorithm by disentangling and critically discussing the different technical settings, and then presenting a self-adapting scheme based upon the obtained insights.

Part I.

Theoretical Background

2. Describing the Configurational Space of Atomic Clusters

2.1. The Born-Oppenheimer Surface

The starting point to get the energetics of an atomic cluster containing N_{el} electrons and N nuclei is the Schrödinger equation which in time-independent non-relativistic cases can be written as

$$\hat{H}\Psi(\{\mathbf{x}_i\},\{\mathbf{R}_\alpha\}) = E\Psi(\{\mathbf{x}_i\},\{\mathbf{R}_\alpha\}) \;, \tag{2.1}$$

where $\{\mathbf{x}_i\} \equiv \{\mathbf{x}_1,\ldots,\mathbf{x}_{N_{\mathrm{el}}}\} \equiv \{(\mathbf{r}_1,\sigma_1),\ldots,(\mathbf{r}_N,\sigma_{N_{\mathrm{el}}})\}$ is a short hand notation for all spatial and spin coordinates of the N_{el} electrons and $\{\mathbf{R}_\alpha\} \equiv \{\mathbf{R}_1,\ldots,\mathbf{R}_N\}$ denotes the spatial coordinates of all nuclei. The Hamiltonian contains the following terms

$$\hat{H} = T_{\mathrm{e}} + T_{\mathrm{N}} + V_{\mathrm{Ne}} + V_{\mathrm{ee}} + V_{\mathrm{NN}} \;, \tag{2.2}$$

with the individual contributions being the kinetic energy of the electrons T_{e} and of the nuclei T_{N}, the electrostatic interaction between the electrons V_{ee}, between the nuclei V_{NN} and between the electrons and nuclei V_{Ne}. Using atomic units, i.e. $\hbar = m_e = e = 4\pi\epsilon_0 = 1$, the terms have the following form:

$$T_{\mathrm{e}} = \sum_{i=1}^{N_{\mathrm{el}}} \frac{\mathbf{p}_i^2}{2} = -\frac{1}{2}\sum_{i=1}^{N_{\mathrm{el}}} \nabla_i^2 \;, \tag{2.3}$$

$$T_{\mathrm{N}} = \sum_{\alpha=1}^{N} \frac{\mathbf{p}_\alpha^2}{2M_\alpha} = -\frac{1}{2}\sum_{\alpha=1}^{N} \frac{1}{M_\alpha}\nabla_\alpha^2 \;, \tag{2.4}$$

$$V_{\mathrm{Ne}}(\{\mathbf{r}_i\},\{\mathbf{R}_\alpha\}) = -\sum_{i=1}^{N_{\mathrm{el}}}\sum_{\alpha=1}^{N} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} \;, \tag{2.5}$$

$$V_{\mathrm{ee}}(\{\mathbf{r}_i\}) = \sum_{i=1}^{N_{\mathrm{el}}}\sum_{j>i}^{N_{\mathrm{el}}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \;, \tag{2.6}$$

$$V_{\mathrm{NN}}(\{\mathbf{R}_\alpha\}) = \sum_{\alpha=1}^{N}\sum_{\beta>\alpha}^{N} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \;. \tag{2.7}$$

In principle, the solution of Eq. (2.1) accurately describes the energetics of a system. However, containing $3N + 3N_{\mathrm{el}}$ degrees of freedom, it is computationally unfeasible to

tackle Eq. (2.1) exactly and hence, approximations have to be applied. The first important one is the *Born-Oppenheimer Approximation* [12] which states that electronic and nuclear motions can be separated using the following ansatz for the total wavefunction:

$$\Psi(\{\mathbf{x}_i\},\{\mathbf{R}_\alpha\}) = \Psi_\mathrm{e}(\{\mathbf{x}_i\},\{\mathbf{R}_\alpha\})\Psi_\mathrm{N}(\{\mathbf{R}_\alpha\}) \;, \tag{2.8}$$

with the electronic part being a solution of the electronic Hamiltonian:

$$\hat{H}_\mathrm{e} = T_\mathrm{e} + V_\mathrm{Ne} + V_\mathrm{ee} \;, \tag{2.9}$$

$$\hat{H}_\mathrm{e}\Psi_\mathrm{e}(\{\mathbf{x}_i\},\{\mathbf{R}_\alpha\}) = E_\mathrm{e}(\{\mathbf{R}_\alpha\})\Psi_\mathrm{e}(\{\mathbf{x}_i\},\{\mathbf{R}_\alpha\}) \;. \tag{2.10}$$

The justification is that due to the huge ratio of the masses, electrons can instantaneously catch up with the movements of the nuclei. In the extreme case of a hydrogen atom, for instance, $M/m_e \approx 1800$. In other words, the ions are fixed from the electronic point of view or just move very slowly. Consequently, the nuclear positions appear only as parameters in the electronic wavefunction and the effect of the kinetic energy operator of the nuclei on the electronic part of the wavefunction can be neglected.

$$T_\mathrm{N}\,\Psi_e(\{\mathbf{x}_i\},\{\mathbf{R}_\alpha\})\Psi_\mathrm{N}(\{\mathbf{R}_\alpha\}) \approx \Psi_e(\{\mathbf{x}_i\},\{\mathbf{R}_\alpha\})\,T_\mathrm{N}\,\Psi_\mathrm{N}(\{\mathbf{R}_\alpha\}) \;. \tag{2.11}$$

Using this approximation, the ansatz (2.8) leads to the following equation for the nuclear part of the wave function:

$$(T_\mathrm{N} + V_\mathrm{NN}(\{\mathbf{R}_\alpha\}) + E_\mathrm{e}(\{\mathbf{R}_\alpha\}))\Psi_\mathrm{N} = E\Psi_\mathrm{N} \;. \tag{2.12}$$

So the ions move under the influence of the effective potential

$$V_\mathrm{BO}(\{\mathbf{R}_\alpha\}) \equiv V_\mathrm{NN}(\{\mathbf{R}_\alpha\}) + E_\mathrm{e}(\{\mathbf{R}_\alpha\})) \;, \tag{2.13}$$

which is called the *Born-Oppenheimer surface* or more generally the *potential energy surface* (PES). The total energy E_tot of the system is then the potential energy plus quantum mechanical corrections due to lattice vibrations.

The initial problem of solving equation (2.1) has thus been reduced to solving the electronic equation (2.10) in order to obtain $E_\mathrm{e}(\{\mathbf{R}_\alpha\})$ which in turn defines the PES according to Eq. (2.13). Though the number of degrees of freedom has therewith been reduced from $3N_\mathrm{el}+3N$ to $3N_\mathrm{el}$, a huge computational burden still remains and further approximations are required which will be discussed in chapter 4.

2.2. Fundamentals of Group Theory

Many common molecules, in particular small ones containing few atoms, possess symmetry, i.e. undergoing certain *symmetry operations*, like a rotation around a certain angle for instance, the molecule superimposes with itself. Using group theory, symmetry considerations are therefore on the one hand helpful to introduce a nomenclature for describing a molecular structure. Furthermore, features of the PES that will later on be

discussed, can be addressed more quantitatively exploiting the symmetry of the system taken into account. It is therefore of great importance to understand the fundamentals of group theory with which the symmetry properties of a molecule are described in a concise way (see e.g. [13]).

2.2.1. Symmetry Elements and Operations

The symmetry of a molecule can best be described by listing all *symmetry elements* which allow for symmetry operations. One example of a symmetry element is the *n-fold rotation axis* C_n. In such a case, the corresponding symmetry operation C_n^k is k times a rotation about an angle of $2\pi/n$ around the rotation axis which thus superimposes the molecule with itself.

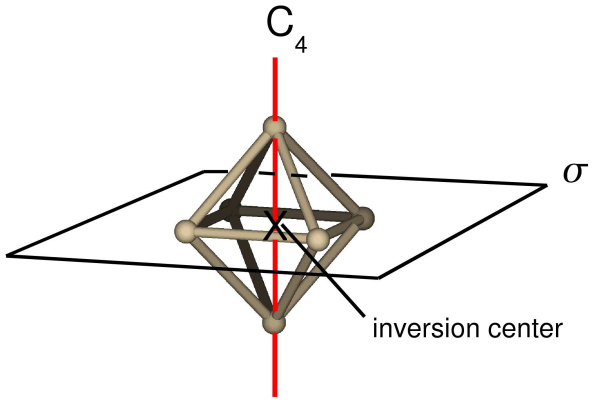


Figure 2.1.: Illustration of the symmetry elements exemplifying an octahedral structure.

Another element is the *plane of symmetry* σ, with the corresponding operation being a reflection of the molecule through this plane.

A third possible element is the *alternating axis of rotation* S_n. In that case, a rotation about an angle of $2\pi/n$ is carried out, being followed by a reflection of the molecule through a plane perpendicular to this axis. S_n^k corresponds to k times a rotation around the axis, each being followed by a reflection. A special case of an alternating axis of rotation is $S_n^{n/2}$, n being even, which is equivalent to an *inversion* of the molecule through its center. Furthermore, if k is even, then S_n^k is simply equivalent to an ordinary rotation C_n^k. For odd n, all S_n^k with $k \leq 2n-1$ are different operations, since then $S_n^{k>n}$ equals S_n^{k-n} followed by a reflection through the symmetry plane and thus is a new operation.

2.2.2. Symmetry Point Groups

The set of all possible symmetry operations of a molecule forms a *group* which is mathematically defined by the following properties:

- The *product* of two members $A \circ B = C$ also constitutes a member of the group. In particular, the product is *associative*, i.e. $A \circ (B \circ C) = (A \circ B) \circ C$.
- A special group element is the *identity* E, which leaves any member unchanged, i.e. $E \circ A = A \circ E = A$.
- Each member A can be assigned an inverse A^{-1} so that the product $A \circ A^{-1} = A^{-1} \circ A = E$ equals the identity.

In the case of symmetry operations, a product of two operations corresponds to the consecutive application of these two operations, like for instance the product $C_n^1 \circ C_n^1 = C_n^2$ corresponds to twice a rotation about an angle of $2\pi/n$. The net operation C_n^2 is hence a rotation about an angle of $2 \cdot 2\pi/n$ and therefore also part of the group. The identity E is a trivial operation that does not change the molecule at all. Obviously, for each operation one can find an operation that reverts the effect of the former one, thus being the corresponding inverse element. The inverse of the rotation C_n^1, for instance, is C_n^{n-1}.

The symmetry of the molecule is reflected by the point group of the constituent symmetry operations, with the number of different operations being an important quantity which is called the *order* of the point group.
The most common point groups describing the symmetry of molecules are:

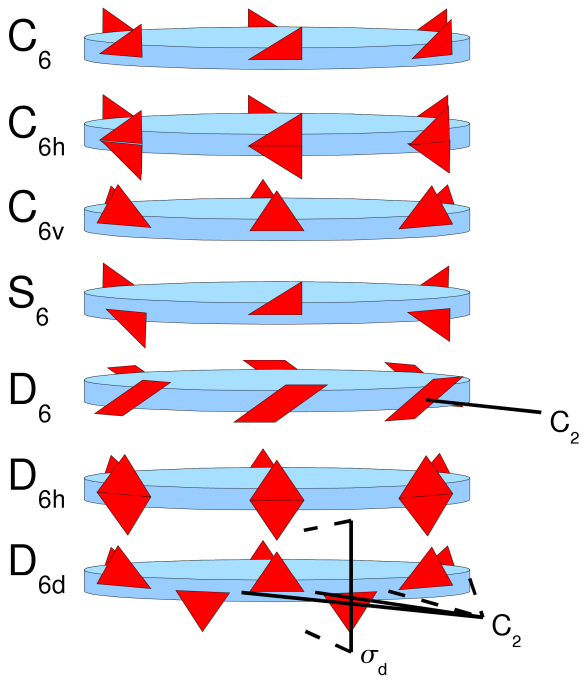


Figure 2.2.: Illustration of the most common symmetry point groups.

- C_n. This simple point group is made up of an n-fold-rotation axis which thus contains the elements $\{E, C_n^1, \cdots, C_n^{n-1}\} = \{E, \{C_n^k\}\}$. Correspondingly, it possesses the order n and is called a *cyclic* group.

- C_{nh}. Additional to an n-fold-rotation axis, the molecule described by this cyclic group exhibits a plane of symmetry σ_h perpendicular to the axis ("horizontal"). Consequently, the alternating axis appears as well as a symmetry operation, since $\sigma_h \circ C_n^k = S_n^k$, with k being odd. For even k, S_n^k is equivalent to C_n^k since the reflections at the horizontal plane even out ($\sigma_h \circ \sigma_h = E$). The special case of C_{1h} is commonly denoted as C_s. The group therewith contains the operations $\{E, \sigma_h, \{C_n^k\}, \{S_n^k\}\}$.

- C_{nv}. Additional to an n-fold-rotation axis, the molecule contains n vertical planes perpendicular to the axis. Summarized the operations are $\{E, \{\sigma_v^i\}, \{C_n^k\}\}$.

- S_n. A further cyclic group is obtained by having a pure alternating axis of rotation which thus leads to the operations $\{E, \{S_n^k\}\}$. This group is only new for even n, since for n being odd, S_n is equivalent to C_{nh}.

- D_n. Additional to an n-fold-rotation axis, the molecule described by the *dihedral* group features n two-fold rotation axes perpendicular to the principle symmetry

axis, which thus allows for the symmetry operations $\{\mathrm{E}, \{\mathrm{C}_n^k\}, \{(\mathrm{C}_2^1)^i\}\}$, with i being an index running over the different two-fold rotation axes.

- $\underline{\mathrm{D}_{n\mathrm{h}}}$. Additional to the symmetry operations of the D_n-group, the molecule possesses a plane of symmetry perpendicular to the principle axis which further introduces the alternating axis of symmetry and n vertical planes of symmetry. Hence, this dihedral group contains the operations $\{\mathrm{E}, \sigma_\mathrm{h}, \{\sigma_\mathrm{v}^i\}, \{\mathrm{C}_n^k\}, \{(\mathrm{C}_2^1)^i\}, \{\mathrm{S}_n^k\}\}$.

- $\underline{\mathrm{D}_{n\mathrm{d}}}$. Additional to the symmetry operations of the D_n-group, the system contains n vertical planes of symmetry bisecting the n two-fold rotation axes, resulting in the set of operations $\{\mathrm{E}, \sigma_\mathrm{h}, \{\sigma_\mathrm{d}^i\}, \{\mathrm{C}_n^k\}, \{(\mathrm{C}_2^1)^i\}, \{\mathrm{S}_n^k\}\}$.

The remaining point groups contain multiple rotation axes of order greater than two. The *tetrahedral* group T, for instance, possesses all rotation axes of a regular tetrahedron, while the *full tetrahedral* group T_d additionally contains all mirror planes. Similarly, the *full octahedral* group O_h comprises all symmetry elements possessed by a regular cube, and the *full icosahedral* group I_h those of an icosahedron.

2.3. Features of the Energy Surface

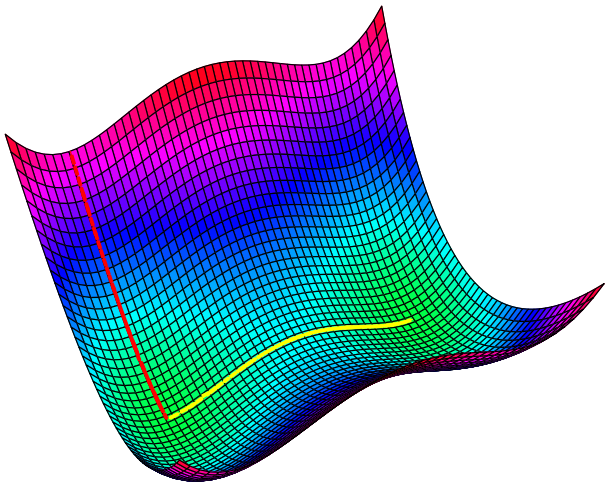


Figure 2.3: Model surface ($V = x^4 - x^2 + y^2$) exhibiting two local minima (green regions) that are connected by two steepest-descent paths crossing a transition state (yellow line). The red line exemplifies a further steepest-descent path leading to the left local minimum.

2.3.1. Stationary Points and Normal Modes

Within the Born-Oppenheimer approximation, the nuclear dynamics is typically treated semi-classically, i.e. the energy surface V_BO is obtained quantum-mechanically, whereas the atoms are treated as classical point masses in space. The forces are therefore obtained as total energy derivatives with respect to the atomic positions

$$\mathbf{F}_\alpha = -\frac{dV_\mathrm{BO}}{d\mathbf{R}_\alpha} \,, \tag{2.14}$$

and the atomic motion is given by the solution of the classical Newton equation

$$M_\alpha \ddot{\mathbf{R}}_\alpha = -\frac{d}{d\mathbf{R}_\alpha} V_{\mathrm{BO}}(\{\mathbf{R}_\alpha\}) \ . \tag{2.15}$$

Most interesting points on the PES are the *stationary points* which are configurations for which all forces vanish and the corresponding energy possesses an extremum. In particular, *local minima* exhibit a positive slope of the PES in all directions. Consequently, any small displacement of the atoms results in a restoring force that brings the atoms back to their equilibrium positions, the corresponding configuration is *stable* and called an *isomer* of the system in the context of molecules and clusters. In particular, the local minimum with the lowest total energy is the *global minimum*. In order to check the local stability of a stationary point quantitatively, a *normal mode* analysis can be performed. For that, molecular vibrations around the equilibrium position of the nuclei are considered. In the *harmonic approximation*, the PES can be expanded in a Taylor-series around $\{\mathbf{R}_{\alpha,0}\}$ and truncated after the second-order term. The first-order term is zero since the forces vanish at the equilibrium position $\{\mathbf{R}_{\alpha,0}\}$,

$$\begin{aligned} V_{\mathrm{BO}}(\{\mathbf{R}_\alpha\}) &= V_{\mathrm{BO}}(\{\mathbf{R}_{\alpha,0}\}) + \sum_\beta \underbrace{\left.\frac{dV_{\mathrm{BO}}}{d\mathbf{R}_\beta}\right|_{\{\mathbf{R}_{\alpha,\mathbf{0}}\}}}_{=0} \mathbf{q}_\beta + \frac{1}{2}\sum_{\beta\gamma} \underbrace{\left.\frac{d^2 V_{\mathrm{BO}}}{d\mathbf{R}_\beta d\mathbf{R}_\gamma}\right|_{\{\mathbf{R}_{\alpha,\mathbf{0}}\}}}_{\equiv H_{\beta\gamma}} \mathbf{q}_\beta \mathbf{q}_\gamma + \cdots \\ &= V_{\mathrm{BO}}(\{\mathbf{R}_{\alpha,0}\}) + \frac{1}{2}\sum_{\beta\gamma} H_{\beta\gamma}\mathbf{q}_\beta\mathbf{q}_\gamma + \cdots \ , \end{aligned} \tag{2.16}$$

where the *displacement vector* $\mathbf{q}_\alpha = \mathbf{R}_\alpha - \mathbf{R}_{\alpha,0}$ is defined as the change of the atomic positions with respect to their equilibrium configuration. Inserting this into Eq. (2.15) yields

$$M_\alpha \ddot{\mathbf{q}}_\alpha = -\frac{1}{2}\sum_{\beta\gamma} \frac{d}{d\mathbf{R}_{\alpha,0}} \left(H_{\beta\gamma}\mathbf{q}_\beta\mathbf{q}_\gamma\right) \ . \tag{2.17}$$

Since $\frac{d\mathbf{q}_\beta}{d\mathbf{R}_\alpha} = \delta_{\alpha\beta}$, the equation of motion further simplifies to

$$M_\alpha \ddot{\mathbf{q}}_\alpha = -\sum_\beta H_{\alpha\beta}\mathbf{q}_\beta \ . \tag{2.18}$$

A solution to the above equation can be obtained by the following ansatz

$$\mathbf{q}_\alpha = \mathbf{A}_\alpha \cos\left(\omega t + \epsilon\right) \ , \tag{2.19}$$

with the second time derivative

$$\ddot{\mathbf{q}}_\alpha = -\omega^2 \mathbf{A}_\alpha \cos\left(\omega t + \epsilon\right) \ . \tag{2.20}$$

Inserting this into Eq. (2.15) gives the following eigenvalue-equation for the $3M$ amplitudes $\{\mathbf{A}_\alpha\}$.

$$M_\alpha \omega^2 \mathbf{A}_\alpha = \sum_\beta H_{\alpha\beta} \mathbf{A}_\beta \tag{2.21}$$

Non-trivial solutions can be obtained by solving the secular equation

$$\det\left(H_{\alpha\beta} - \delta_{\alpha\beta} M_\alpha \omega^2\right) \stackrel{!}{=} 0\ . \tag{2.22}$$

For a system containing M atoms, the Hessian $H_{\alpha\beta}$ possesses $3N$ rows and columns and the secular equation yields $3N$ solutions ω_i. Within the harmonic approximation, the atoms are therefore oscillating around their equilibrium positions with an harmonic motion having the frequency $2\pi/\omega_i$ and a constant phase ϵ_i. The normalized eigenvectors $\mathbf{Q}_{\alpha,i} = \mathbf{A}_{\alpha,i}/\sqrt{\sum_\beta \mathbf{A}^2_{\beta,i}}$, the *eigenmodes*, describe the relative motion of the individual atoms, where the amplitude is uniquely defined by the boundary conditions, i.e. how far the individual atoms are initially displaced. Having the same frequency and phase, all atoms reach their maximum displacements simultaneously. A vibrational mode exhibiting the above described characteristics is called a *normal mode* of the system.

It can be shown that for a non-linear molecule, only $(3N-6)$ solutions have non-vanishing frequencies. The other six solutions correspond to the three translational and three rotational degrees of freedom of the system. In case of a linear molecule, there are $(3N-5)$ non-vanishing solutions, since a linear system only possesses two rotational degrees of freedom.

A local minimum possesses only positive eigenvalues ω_i^2. An imaginary frequency ω_i corresponds to a negative slope of the PES in direction of the eigenvector and thus indicates a local instability, since an infinitesimally small displacement of the atoms along the eigenvector brings the system out of the equilibrium position. The number of negative eigenvalues is often referred to as the *index* of the Hessian. An index of zero corresponds to a local minimum on the PES which hence represents a stable isomer of the system under consideration. Stationary points with indices of one or higher are called *saddle points*, in particular a saddle point with index one is a *transition state*. If two local minima are connected by any path on the PES, then according to the Murrel-Leidler theorem [14] there is a lower energy path connecting these two minima which involves a transition state. In other words, if the system undergoes a transition from one isomer to another, the path that involves the lowest energy barrier to surmount possesses a saddle point of index one at its energy maximum (see Fig. 2.3).

Steepest-Descent Path and Basin of Attraction

Loosely stated, a steepest-descent path starting from any configuration is defined by simply following the PES downhill along the forces until a stationary point is reached where the gradient vanishes. Numerically, one would therefore obtain the steepest-descent path by successively displacing the atoms by infinitesimal amounts along the force direction. One important feature is that the resulting path on the PES is uniquely defined and every point with a non-vanishing gradient in the configurational space lies exactly on one path [15]. In particular, all points whose steepest-descent paths converge to a certain

local minimum form its *basin of attraction* which thus comprises all configurations of a molecule that will relax into the corresponding isomer represented by the local minimum on the PES. Due to the uniqueness of the steepest-descent path, basins of attractions belonging to different local minima cannot interpenetrate and are a characteristic region around a local minimum.

At a transition state there are exactly two steepest-descent paths going in opposite directions along the eigenvector corresponding to the negative curvature and leading to two local minima. They constitute the minimum-energy path on the PES for a transition between these two local minima (see Fig. 2.3).

Scaling behaviour of the number of local minima with the system size

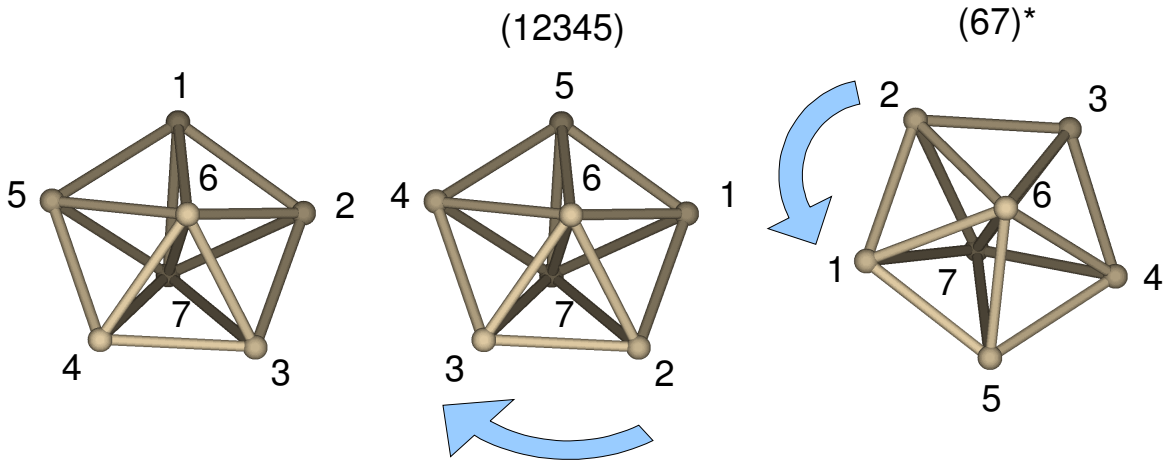


Figure 2.4.: Illustration of two permutation-inversions of a pentagonal bipyramid leading to the same local minimum. (12345) indicates a permutation that swaps atom 1 with atom 2, 2 with 3, 3 with 4, 4 with 5, 5 with 1, thus resulting in a pure rotation (indicated by the blue arrow), which in turn corresponds to the same local minimum. $(67)^*$ swaps atom 6 with 7 being followed by an inversion through the center which can also be superimposed with the unpermuted one by a pure rotation.

Regarding in particular the question how the configurational space of a system can be sampled efficiently, it is of fundamental interest to estimate how big the PES is in terms of the number of different isomers one expects to find. To address this question, one has to distinguish between two different aspects of this problem. Since the Hamiltonian of the system is invariant with respect to permutations of the nuclear coordinates of equivalent species and with respect to inversions of all the coordinates through the origin of a space-fixed coordinate system, there are many symmetry-equivalent isomers with the same geometry and energy, but corresponding to different points on the PES. Assuming the system contains N_A atoms of species A, N_B atoms of species B, N_C atoms of species C, and so forth, one would expect to have

$$2 \times N_\mathrm{A}! \times N_\mathrm{B}! \times N_\mathrm{C}! \times \cdots \tag{2.23}$$

symmetry-equivalent isomers, or *versions* of the structure. However, it can be shown [6, 16] that in case of a molecule with a point group of order o, there are o permutation-inversions that can be obtained from the non-permuted isomer by a pure rotation, thus not being a distinct local minimum of the isomer (see Fig. 2.4). Therefore, there are

$$N_\mathrm{versions} = 2 \times N_\mathrm{A}! \times N_\mathrm{B}! \times N_\mathrm{C}! \times \cdots / o \tag{2.24}$$

different versions of the same isomer that cannot be superimposed by a pure rotation. In other words, every isomer possesses N_versions different basins of attraction on the PES. Hence, the lower the symmetry of a structure, the more symmetry-equivalent local minima it possesses on the PES.

More difficult is the question how many isomers of different energy and geometric structures exists. According to empirical observations and heuristic estimates [17, 18, 19], the number of isomers grows exponentially with increasing cluster size. The basic idea of such an estimate is to divide the system into m equivalent subsystems each containing M atoms. Assuming that the stable configurations of each subsystem can be considered independent, the number of local minima n_min then behaves like

$$n_\mathrm{min}(mN) = n_\mathrm{min}(N)^m \ . \tag{2.25}$$

A solution of this equation is

$$n_\mathrm{min}(N) \propto \exp(\alpha N) \ , \tag{2.26}$$

where α is a system-dependent constant.

2.3.2. Global Topology of the Energy Surface

The overall shape of the energy surface is determined by the energetic ordering of the local minima and their connectivity through transition states. One feature to describe the global topology of a PES is for instance the *monotonic sequence* introduced by Kunz and Berry [20, 21]. It is defined as a sequence of local minima connected by transition states for which the energy of the minima monotonically decreases. All monotonic sequences leading to the same local minimum therewith define a *monotonic sequence basin* or *funnel*. Another common term is *superbasin* which emphasizes the analogy to a basin of attraction on the global scale of the PES. In contrast to the latter where every point of the configurational space with a non-vanishing gradient only belongs to one basin of attraction, a local minimum can belong to several monotonic sequence basins.

Systems with only one funnel converging to the global minimum are often referred to as *single-funnel systems*. They are obviously benign systems for a global optimization since the overall topology can guide the system towards the global minimum. Contrary to that, *multi-funnel systems* contain multiple sequence basins. Once in a wrong funnel, the system then has to surmount several energy barriers to climb out of it and reach the global minimum, which makes global optimization in general more difficult [7, 8].

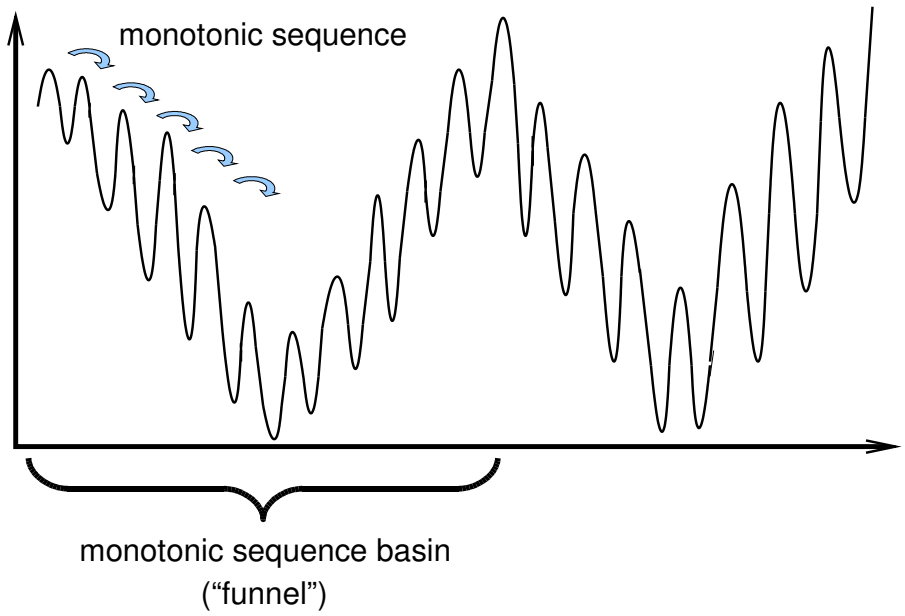


Figure 2.5.: Schematical picture of a multi-funnel potential energy surface with a depicted monotonic sequence basin.

2.4. Thermodynamics

At zero temperature, the most stable structure is the isomer with the lowest energy E_0 plus the *zero point energy* (*vide infra*) given by the vibrations. Increasing the temperature of the system, entropic effects start to play a role and the *free energy* F determines the stability of the system (see e.g.[22]).

$$F = E_0 - k_B T \ln Z \ , \tag{2.27}$$

where Z is the partition function of the system defined as

$$Z = \sum_i e^{-\frac{E_i}{k_B T}} \ , \tag{2.28}$$

with i being the index running over all possible states of the system and E_i the corresponding energy of the system. For an atomic cluster system, assuming that rotations and vibrations can be decoupled, the partition function separates into

$$Z = Z_{\text{trans}} Z_{\text{vib}} Z_{\text{rot}} \ , \tag{2.29}$$

where Z_{trans}, Z_{vib} and Z_{rot} are the partition functions due to the translational, vibrational and rotational degrees of freedom. The free energy then adopts the form

$$F = E + F_{\text{trans}} + F_{\text{vib}} + F_{\text{rot}}. \tag{2.30}$$

2.4.1. Translational Degrees of Freedom

The partition function Z_{trans} can be calculated in a straightforward manner and has the form [22]

$$Z_{\text{trans}} = V \left(\frac{M k_B T}{2\pi\hbar^2} \right)^{3/2} . \tag{2.31}$$

V is the volume of the box in which the cluster is placed. Since the only system-dependent parameter is the total mass M of the system, the translational degrees of freedom do not influence the energetic ordering of different isomers of the same system and can thus be neglected when addressing the energetic stability of free molecules.

2.4.2. Molecular Vibrations

The vibrational contribution to the free energy can be approximated using the vibrational frequencies obtained by the harmonic approximation. Using normal modes, the molecule can be considered as being composed of $(3N-6)$ independent one-dimensional harmonic oscillators with the energies $E_{n,i} = \left(n + \frac{1}{2}\right) \hbar\omega_i$, where n is the principal quantum number of the vibrational state. Hence the partition function further separates into [22]

$$Z_{\text{vib}} = \sum_{i=1}^{3N-6} Z_{\text{vib},i} . \tag{2.32}$$

Each partition function can easily be evaluated analytically,

$$Z_{\text{vib},i} = \sum_{n=0}^{\infty} e^{-\left(n+\frac{1}{2}\right)\frac{\hbar\omega_i}{k_B T}} = e^{-\frac{\hbar\omega_i}{2k_B T}} \sum_{n=0}^{\infty} e^{-n\frac{\hbar\omega_i}{k_B T}} = \frac{e^{-\frac{\hbar\omega_i}{2k_B T}}}{1 - e^{-\frac{\hbar\omega_i}{k_B T}}} , \tag{2.33}$$

$$Z_{\text{vib}} = \prod_{i=1}^{3N-6} \frac{e^{-\frac{\hbar\omega_i}{2k_B T}}}{1 - e^{-\frac{\hbar\omega_i}{k_B T}}} . \tag{2.34}$$

The resulting vibrational contribution to the free energy is

$$F_{\text{vib}} = -k_B T \ln Z_{\text{vib}} = \sum_{i=1}^{3N-6} \left(\frac{\hbar\omega_i}{2} + k_B T \ln\left(1 - e^{-\frac{\hbar\omega_i}{k_B T}}\right) \right) \tag{2.35}$$

$$= E_{\text{ZPE}} + k_B T \sum_{i=1}^{3N-6} \ln\left(1 - e^{-\frac{\hbar\omega_i}{k_B T}}\right) . \tag{2.36}$$

The first term corresponds to the *zero-point energy* since it contributes to the free energy already at zero temperature. It originates from the zero-point vibration of the molecules.

2.4.3. Molecular Rotations

As for the rotational contribution to the free energy, one can assume that the centrifugal forces acting upon the nuclei are negligible and do not alter the molecular structure significantly. A good approximation to the rotational partition function based upon classical mechanis is [22]

$$Z_{\mathrm{rot}} = \frac{\sqrt{\pi}}{\sigma}\left(\frac{2k_BT}{\hbar^2}\right)^{3/2}\sqrt{I_1I_2I_3}\,, \tag{2.37}$$

where I_1, I_2, I_3 are the principle moments of inertia, e.g the eigenvalues of the moment of inertia tensor. σ is the order of the rotational subgroup of the system, so the number of pure rotational symmetry operations.

3. Exploring the Configurational Space

Determining cluster structures requires to identify the local minima of the PES, which is a high-dimensional function of the atomic coordinates. Of fundamental interest is thereby without doubt the global minimum, which constitutes the most stable isomer at zero temperature. Energetically higher-lying and therewith metastable isomers, however, might be observed in experiment due to finite temperature or kinetic effects, thus additionally being in the focus of interest.

Hence, this work faces a twofold challenge. On the one hand, the PES needs to be evaluated, thus necessitating a method that yields the total energy for a given atomic configuration. Obtaining this energy within the framework of density-functional theory is the topic of chapter 4. On the other hand, the local information of a current configuration thereby gained needs to be exploited to efficiently sample the PES and finally identify the local minima. The goal of this chapter is to condense and classify most popular corresponding schemes to explore the huge configurational space.

3.1. Local Optimization Methods

Local optimization methods are deterministic schemes that guide the system to the next local minimum of the PES from an arbitrary starting point by following the surface downhill in some way. At each iteration step, local information like the energy, forces or the Hessian of the corresponding atomic configuration are taken into account to obtain the next structure until iteratively the forces vanish and the local minimum has been identified.

3.1.1. Steepest Descent

The simplest method to implement is the steepest descent method [22] that strictly proceeds downhill to reach the local minimum. In each iteration step, the atoms are displaced according to the forces acting on them

$$\mathbf{R}_{\alpha,i+1} = \mathbf{R}_{\alpha,i} + \gamma_i \mathbf{F}_\alpha(\{\mathbf{R}_{\alpha,i}\}) \,, \tag{3.1}$$

with γ_i being a technical step width parameter to adjust that does not necessarily have to be constant for the whole local optimization. If the step width is too small, many iteration steps are required and the convergence to the local minimum is rather slow. If on the other hand γ_i is too large, the system might start to oscillate around the local minimum.

3.1.2. Conjugate Gradient

The conjugate gradient scheme consists of succesive line minimizations along a search direction $\mathbf{G}_{\alpha,i}$ [22, 23, 24, 25, 26]

$$\gamma_i = \arg\min_{\gamma} E\left(\mathbf{R}_{\alpha,i} + \gamma \mathbf{G}_{\alpha,i}(\{\mathbf{R}_{\alpha,i}\})\right) \ , \tag{3.2}$$

where $\arg\min_\gamma E$ denotes the argument γ which minimizes the energy E. The atomic coordinates are then correspondingly updated to

$$\mathbf{R}_{\alpha,i+1} = \mathbf{R}_{\alpha,i} + \gamma_i \mathbf{G}_{\alpha,i} \ . \tag{3.3}$$

Similar to the steepest descent scheme, one starts along the atomic forces as first search direction, so $\mathbf{G}_{\alpha,0} = \mathbf{F}_\alpha(\{\mathbf{R}_{\alpha,0}\})$. Contrary to the former scheme, however, information from previous searches are accumulated so that new line searches do not destroy the progress of previous line searches as might be the case for ill-conditioned functions (see Fig. 3.1). In these cases, the steepest descent scheme follows a criss-cross pattern since every new line step only takes local information into account. The conjugate gradient scheme instead does not strictly follow the PES downhill but along a search direction that is somewhat perpendicular ("conjugate") to the previous search directions, which is achieved by adding a fraction of the previous search direction to the atomic forces

$$\mathbf{G}_{\alpha,i} = \mathbf{F}_{\alpha,i} + \beta_i \mathbf{G}_{\alpha,i-1} \ . \tag{3.4}$$

There exist different flavours of the conjugate gradient scheme, differing by the definition of β_i. Some of the names associated with these schemes are Fletcher-Reeves (FR) [24, 25], Polak-Ribière (PR) [24] and Hestenes-Stiefel (HS) [26]

$$\beta_i^{\mathrm{FR}} = \frac{\sum_\alpha \mathbf{F}_{\alpha,i} \cdot \mathbf{F}_{\alpha,i}}{\sum_\alpha \mathbf{F}_{\alpha,i-1} \cdot \mathbf{F}_{\alpha,i-1}} \ , \tag{3.5}$$

$$\beta_i^{\mathrm{PR}} = \frac{\sum_\alpha \mathbf{F}_{\alpha,i} \cdot (\mathbf{F}_{\alpha,i} - \mathbf{F}_{\alpha,i-1})}{\sum_\alpha \mathbf{F}_{\alpha,i-1} \cdot \mathbf{F}_{\alpha,i-1}} \ , \tag{3.6}$$

$$\beta_i^{\mathrm{HS}} = \frac{\sum_\alpha \mathbf{F}_{\alpha,i} \cdot (\mathbf{F}_{\alpha,i} - \mathbf{F}_{\alpha,i-1})}{\sum_\alpha \mathbf{G}_{\alpha,i-1} \cdot (\mathbf{F}_{\alpha,i} - \mathbf{F}_{\alpha,i-1})} \ . \tag{3.7}$$

The construction of the search directions $\mathbf{G}_i$ is based upon the assumption that the PES is harmonic. If this assumption holds exactly, the conjugate gradient scheme will converge to the local minimum in $3N$ iteration steps, thus being a direct minimization scheme. In particular, all aforementioned different schemes are then equivalent. Consequently, if the system is far away from the harmonic region, the search directions become unreasonable and it is recommended to start with a few steepest descent steps to bring the system closer to the local minimum.

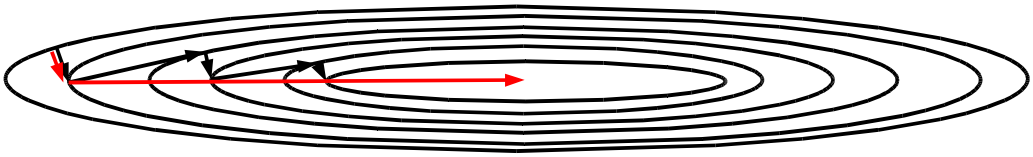


Figure 3.1.: Illustration of the steepest descent scheme (black arrows) for an ill-conditioned two-dimensional surface compared to conjugate gradient (red arrows).

3.1.3. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method

The BFGS method is a so-called quasi-Newton scheme that takes additional information in the form of the second derivative of the PES into account [27]. Knowing the Hessian matrix $H_{\alpha\beta}$, a new search direction $\mathbf{G}_{\beta,i}$ can be obtained by solving the Newton equation

$$\sum_{\beta} H_{\alpha\beta}\mathbf{G}_{\beta,i} = \mathbf{F}_{\alpha}(\{\mathbf{R}_{\alpha,i}\})\,. \tag{3.8}$$

The next atomic configuration can then be obtained by performing a line minimization as in the conjugate gradient scheme (Eq. (3.2)). If the PES were perfectly harmonic and the Hessian known exactly, the local minimum would be found within one line search. In practice, however, the calculation of the Hessian matrix in each iteration step can be prohibitively expensive, so that it is instead successively approximated in each iteration step, therewith being a *quasi*-Newton scheme. Since more information of the PES is taken into account, the BFGS method can be more efficient than steepest descent or even conjugate gradient, but it works only efficiently if the system is close to the local minimum, where the harmonic approximation is justified. It is therefore typically recommended to start with a few steepest descent steps to bring the system close to the local minimum as in the case of the conjugate gradient scheme.

3.2. Global Optimization Schemes

The above described local optimization methods are deterministic schemes that approach the next local minimum from a starting point constructively, since the local information obtained for a given configuration, like the energy and forces, uniquely guides the system to the next local minimum. In order to find all isomers, particularly the global minimum, a search algorithm is required that not only scans the local environment but samples the whole PES. Since local information is not enough to locate the global minimum of the system, stochastic methods are required that provide some recipe to systematically sample the huge configurational space. The global optimization of atomic clusters is assumed to be NP-hard (non-deterministic polynomial-time hard) [28], i.e. no algorithm is known that guarantees to locate the most stable isomer within polynomial time. This difficulty is typically ascribed to the exponential growth of the number of local minima with system size. Assuming the most primitive method that simply runs through all local

minima, the computational burden therefore quickly becomes unfeasible which connects to the famous Levinthals paradox according to which a protein would never reach its native state within the lifetime of universe in view of the tremendous number of different local minima [7, 8]. Correspondingly, one possible way out is that the global topology can serve as a guidance for the system to reach the global minimum. As already stated in the previous chapter, single-funnel systems are benign topologies in this sense.

One key ingredient of any sampling algorithm is the way new structures are generated. This so-called *trial move* corresponds to a jump of the system in the configurational space. A simple method for that is to randomly displace the atomic positions. After such a jump, a criterium is required which decides whether this new structure is preferred and the search should continue from there or if it should rather be discarded. Such an *acceptance criterium* is typically based upon the total energy, being a natural choice since it is the quantity to optimize. Of course, there is no need to confine to a single sequential run. Multiple search sequences can be performed and information can be exchanged to thus combine structures from different positions on the PES instead of a purely local search. The number of serial runs and the special scheme to cross them are examples of the large number of technical parameters that make up for the huge diversity of different optimization schemes of which the most archetypal ones will be described in the following.

3.2.1. Simulated Annealing

A classical optimization algorithm applied to many fields is the simulated annealing scheme that resorts to methods of statistical mechanics and is based upon the Metropolis-algorithm [29, 30]. Starting from an arbitrary configuration with a total energy E, a new structure is generated by randomly displacing the atoms, leading to a change of the total energy ΔE. If the energy has decreased, i.e. $\Delta E < 0$, the new structure is accepted and used as starting point for the next step. In the opposite case, however, the structure is not discarded unconditionally, but accepted with a probability of $P(\Delta E) = \exp(-\Delta E/k_B T)$. According to Metropolis, this acceptance criterium thus generates a canonical ensemble of atomic configurations at T. At zero temperature, only isomers that are lower in energy would be accepted which would be an intuitive choice at first glance since the system is thereby pushed towards the ground state. However, the system is then likely to get stuck in the wrong minimum. A finite temperature allows instead for controlled uphill steps thus enabling the system to get out of local minima. Starting from a high temperature, the system is then successively cooled down according to an *annealing schedule*. Coupled to the temperature are the random displacements of the individual atoms $\Delta\mathbf{R}_\alpha$ that obey a Gaussian distribution in the classical simulated annealing scheme [31]:

$$p(\Delta\mathbf{R}_\alpha) \propto \exp\left(-(\Delta\mathbf{R}_\alpha)^2/T\right) \tag{3.9}$$

With decreasing temperature, the step width is therewith reduced thus freezing the system in the ground state. The cooling rate needs to be inverse logarithmic in time to assure convergence

$$T \propto \frac{T_0}{\log(1+t)} \,. \tag{3.10}$$

Improvements can be made by adjusting the annealing scheme, as done in the *Fast Simulated Annealing* scheme, proposed by Szu *et al.* [32]. The Gaussian distribution of random displacements is replaced by a Cauchy distribution that contains longer tails additional to a Gaussian-like peak, thereby enabling the system to occasionally perform longer jumps in the configurational space. Instead of an inverse logarithmic cooling rate, the temperature can then be reduced linearly with time, converging to the ground state much faster. Both versions are special cases of the *Generalized Simulated Annealing* scheme, that introduces more technical parameters to tune the cooling rate [33]. Due to the reduction of the temperature, no canonical ensemble is generated anymore, so that unlike in the original Metropolis scheme, no thermodynamic quantities can be derived. The resulting simulated annealing scheme is a pure global optimization procedure.

3.2.2. Basin-Hopping

Loosely stated, the basin-hopping method is the simulated annealing method applied to local minima [10, 34, 35]. Additional to the trial move, a local structural relaxation is performed and the total energy of this local minimum is then assigned to the initial configuration. The PES is therewith effectively transformed into a set of interpenetrating staircases that are the basins of attractions introduced in chapter 2 (see Fig. 3.2)

$$\tilde{E}\{\mathbf{R}\} = \min E\{\mathbf{R}\} \,, \tag{3.11}$$

where min indicates a local structural relaxation.

This deformation is an example of the class of *hypersurface transformation* methods that aim at modifying the PES to facilitate the global optimization. In many cases, like e.g. the distance scaling method [36], the surface is smoothed out, thus decreasing the number of local minima. However, there is in general no guarantee that such a transformation does not change the global minimum, or even washes out relevant isomers. On the contrary, the basin-hopping method leaves the local minima unchanged but only removes the transition state regions, thus facilitating interbasin transitions. Hence, contrary to the original simulated annealing scheme, moves that lead the system to high points on the PES are much more likely to be accepted, since the ensuing local optimization allows the system to relax into the corresponding local minimum which is much lower in energy. This makes a complicated annealing schedule unnecessary, and in the classical BH scheme the temperature in the Boltzmann-criterium is simply kept constant [10], thereby generating a canonical ensemble of the transformed PES.

A classical trial move scheme is to randomly vary the cartesian components of all atomic coordinates

$$\Delta\mathbf{R}_{\alpha,i} = 2 \cdot \mathbf{R}_0(\mathrm{rand} - 0.5) \,, \tag{3.12}$$

with $\mathbf{R}_0$ being the maximum step width and rand being a random number between zero and one.

With the basin-hopping scheme, Wales *et al.* identified in 1999 all hitherto known global minima of Lennard-Jones(LJ)-clusters containing up to 110 atoms and additionally new ones. The LJ_{38} is a well-known example of a double-funnel PES [37], that contains an icosahedral funnel with a large free energy, and an fcc funnel with a comparably smaller free energy but leading to the global minimum. The unmodified surface yields only a small overlap in the canonical occupation probabilities with respect to the temperature, so that the system is likely to get trapped in the wrong funnel. Transforming the PES like in the basin-hopping scheme, however, results in a broadening of the overlap region, and enables the system to climb out of the wrong funnel [37, 38] . Hence, it is not only the alleviated interbasin transition due to the removal of the transitions state regions, but also the easier transition between different funnels because of the modified thermodynamics that make up for the success of the basin-hopping scheme.

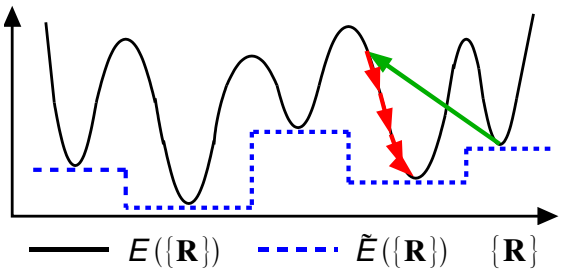


Figure 3.2.: The principle of the basin-hopping method. Depicted is a model energy surface together with its transformed landscape. The green arrow indicates a trial move performed on a local minimum, being followed by a local structural relaxation (red arrows).

3.2.3. Genetic Algorithms

The name of this class of algorithms comes from the mimicked Darwinistic principle of the survival of the fittest [39, 40]. Contrary to the basin-hopping scheme, the starting point is a *population* of randomly generated cluster structures instead of a single initial configuration. New cluster structures are hereby generated by choosing two candidate structures from the population, the *parents*, which are then *mated* to create a *child*. The mating has thereby to be appropriately designed to preserve structural properties of the parents during the mating procedure. A common choice for that is to cut the parent structures by a plane that is randomly oriented and cross the resulting halfs followed by a relaxation to generate a child (see Fig. 3.3). Local structural motifs from different points on the PES are therewith combined instead of performing a pure local

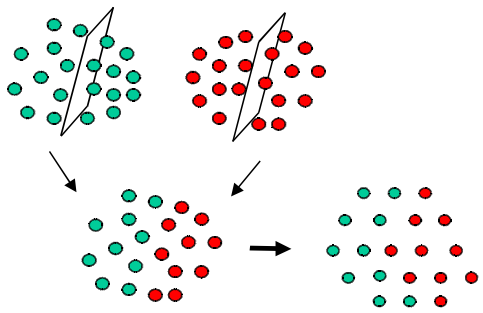


Figure 3.3.: Mating between two parent structures generating a child. After the mating, the new child is locally relaxed.

search like in the basin-hopping scheme.

A further technical aspect is the question which of the children are supposed to replace parents in the population. A simple rule of Deaven and Ho [39] is that a child has to be energetically lower than a parent structure to replace it. To keep the diversity of the population, a further constraint is that a child structure that is supposed to replace a parent structure needs additionally to be different enough from all the other members of the population. Deaven and Ho [39] therefore introduced a second energy criterium. A potentially new structure in the population needs to have an energy difference of ΔE to all other isomers in order to be considered different enough, where ΔE is another tunable parameter. With such an implementation of a genetic algorithm, Deaven and Ho independently discovered the fcc-structure for LJ_{38} in 1996 [40], and based upon a tight-binding scheme, the fullerene structure for C_{60} could be identified in 1995 [39]. It is worth pointing out that due to the involved local structural relaxation, a genetic algorithm is simply a different search algorithm applied to the same transformed PES like in the case of the basin-hopping scheme.

3.2.4. Further Algorithms

Many modifications of the above described basic types of global optimization schemes have been suggested in the literature, one of them is for instance the *minima hopping* algorithm by Stefan Goedecker [41, 42]. As indicated by the name, it can be considered as basin-hopping with a different type of move. Goedecker suggests using a molecular dynamics trajectory instead of random moves to generate new structures, thereby exploiting the *Bell-Evans-Polanyi*-principle [22, 43, 44]. This principle states that a local minimum behind a transition state is lower in energy the smaller the barrier to cross. By starting with a small kinetic energy, the MD-trajectory tends to surmount lower energy barriers and assuming the validity of the BEP-principle is therefore more likely to approach the ground state. Another key ingredient of the minima hopping method is a history feedback method which keeps track of all local minima identified during the sampling run. If the trajectory leads to an isomer that has already been found, the kinetic energy is slightly increased and higher energy barriers can be surmounted, thereby enabling the system to climb out of a wrong funnel if it has been exhaustively explored. Of course, there is no reason to constrain this kind of history feedback method to the application of MD-trajectories, it can also be combined with basin-hopping, which would simply be replacing MD-runs by random moves. Since detailed balance would then not be fulfilled, such a modified basin-hopping scheme would no longer generate a canonical ensemble, which would, however, not constitute a problem since thermodynamic quantities are not of interest in the case of global optimization. Instancing LJ_{38}, MD-steps were proven to be superior compared to random moves [41]. Averaged over a huge number of global optimization runs, roughly 34000 local minima using MD-trajectories were identified until the global minimum was found compared to 75000 in the case of random moves. However, Wales reports an average number of 2000 moves, including more efficient angular moves [10, 45] thus making basin-hopping an order of magnitude faster in this case and illustrating the intricate connection of the sampling scheme and

employed trial moves to the overall performance.

Another type of history feedback method is the concept of landscape paving [46]. The basic idea is to artificially increase the energy in regions that have already been explored. With that, moves that lead into such regions are likely to be rejected thus preventing the system from useless sampling already known regions of the PES. However, this modification, being a significant improvement at first glance, also bears some dangers. Considering the case of two-funnels that are connected by only one transition basin, the paving of this important basin prevents the system from interfunnel transitions, thus prohibiting the system to climb out of a potentially wrong funnel. Combined with basin-hopping, Zhan *et al.* suggested the *basin paving* method [47] that makes use of the histogram of already visited minima to construct an artificial increase of the PES. Using Lennard-Jones clusters, a statistically significant improvement for the global optimization could only be shown for very large systems containing around 150 atoms [47].

3.3. Summary

A concise overview of present sampling methods was given of which the basic ones are simulated annealing, basin-hopping and genetic algorithms. A few popular modifications have been presented which can be considered as derivations from basin-hopping with improvements that were proven to be advantageous for certain cases. However, the performance of the various schemes and their modifications can strongly depend upon both the actual system under consideration and upon the huge number of involved technical parameters, like e.g. the special type of move performed in the random move scheme. A critical analysis and discussion of the implemented sampling method of the present work, in particular regarding possible generic optimizations and showing clear limitations, is therefore part of this PhD project and will be presented in chapter 9.

4. Density-Functional Theory

This chapter deals with the question how to obtain the energetics of a system with a given atomic configuration as required to evaluate the PES. In this work, the method employed is density-functional theory. Before giving a short overview of other popular methods in order to properly classify DFT, some basic terminology will be introduced.

4.1. The Electronic Wave Function

The central quantity within electronic structure theory is the electronic wave function $\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_{N_\mathrm{el}})$ which yields the probability to find an electron at position $\mathbf{r}_1$ with spin σ_1, an electron at position $\mathbf{r}_2$ with spin σ_2 and so forth by

$$|\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_{N_\mathrm{el}})|^2 \, \mathrm{d}\mathbf{x}_1 \ldots \mathrm{d}\mathbf{x}_{N_\mathrm{el}} \, . \tag{4.1}$$

Being fermions, electrons obey the *Pauli principle* [48, 49] which states that two electrons of like spin cannot be at the same point in space. In terms of the wavefunction, this is reflected by the antisymmetry with respect to the space- and spin-coordinates

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_{N_\mathrm{el}}) = -\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_{N_\mathrm{el}}) \, , \tag{4.2}$$

which thus gives for two equal coordinates $\mathbf{x}_i = \mathbf{x}_j$

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_{N_\mathrm{el}}) = 0 \, , \tag{4.3}$$

and ensures that the probability to find two electrons of like spin at the same spatial position is zero as prescribed by the Pauli principle.

The *Variational Principle* [22, 50] states that the expectation value of any trial wave function $E\left[\Psi_\mathrm{trial}\right] = \langle \hat{H} \left| \Psi_\mathrm{trial} \right| \hat{H} \rangle$ is always an upper bound to the ground-state energy given by the expectation value of the ground-state wavefunction Ψ_0

$$E_0 \equiv E\left[\Psi_0\right] \leq E\left[\Psi_\mathrm{trial}\right] \, . \tag{4.4}$$

4.1.1. The Electron and Pair Density

The electron density at the position $\mathbf{r}$ is defined as the probability to find an electron of any spin at this point and is given by

$$\rho(\mathbf{r}) = N_\mathrm{el} \sum_{\sigma_1} \rho^{\sigma_1}(\mathbf{r}) = N_\mathrm{el} \sum_{\sigma_1} \int \ldots \int \mathrm{d}\mathbf{x}_2 \ldots \mathrm{d}\mathbf{x}_{N_\mathrm{el}} \left|\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_{N_\mathrm{el}})\right|^2 \, , \tag{4.5}$$

and is the sum of the individual *spin densities* ρ^σ which describe the probability to find an electron of a specific spin σ at the position $\mathbf{r}$. The prefactor N_{el} is the number of electrons in the system and ensures the right normalization

$$\int \mathrm{d}\mathbf{r}\rho(\mathbf{r})\mathrm{d}\mathbf{r} = N_{\mathrm{el}} \; . \tag{4.6}$$

In the same manner, the *pair density* at $\mathbf{x}_1$ and $\mathbf{x}_2$ is defined as the probability to find an electron with spin σ_1 at the point $\mathbf{r}_1$ and an electron with spin σ_2 at the point $\mathbf{r}_2$

$$\rho_2(\mathbf{x}_1, \mathbf{x}_2) = N_{\mathrm{el}}(N_{\mathrm{el}} - 1) \int \ldots \int \mathrm{d}\mathbf{x}_3 \ldots \mathrm{d}\mathbf{x}_{N_{\mathrm{el}}} \left|\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\mathrm{el}}})\right|^2 \; , \tag{4.7}$$

where the antisymmetry of the wavefunction due to the Pauli principle correspondingly leads to an antisymmetry of the pair-density

$$\rho_2(\mathbf{x}_1, \mathbf{x}_2) = -\rho_2(\mathbf{x}_2, \mathbf{x}_1) \; . \tag{4.8}$$

If electrons moved completely independently, the pair density would just be the product of the spin densities $\rho_2(\mathbf{x}_1, \mathbf{x}_2) = \rho(\mathbf{x}_1)\rho(\mathbf{x}_2)$. The probability to find an electron with spin σ_1 at the point $\mathbf{r}_1$ is then independent of the probability distribution of the second electron. Due to the Coulomb interaction, however, the movements of both electrons are *correlated*, so in general $\rho_2(\mathbf{x}_1, \mathbf{x}_2) \neq \rho(\mathbf{x}_1)\rho(\mathbf{x}_2)$.

4.2. Approximative Methods to solve the Many-Body Problem

One fundamental approach to solve the electronic Schrödinger equation (2.10) numerically is the *Hartree-Fock* method which approximates the electronic wavefunction by a single Slater-determinant of single particle wave functions, thus ensuring the antisymmetry of the wave function to fulfil the Pauli principle. Additional to the classical Coulomb-potential V_{ee}^c, this gives rise to a quantum mechanical contribution to the potential called *exchange potential* V_{X}. The aforementioned Coulomb correlation, however, is not captured, and the pair density of electrons of unequal spins which are not subject to the Pauli principle just separates into the individual spin densities. The missing energy is correspondingly defined as *correlation energy* E_{C} and is significant to obtain accurate results though being smaller compared to the exchange and classical Hartree energy. The corresponding potential is consequently termed *correlation potential* V_{C}.

Among the approaches to account for the correlation energy, the most popular ones based on the electronic wave function are second or fourth order perturbation theory by Møller and Plesset (MP2 or MP4) [9], configuration interaction (CI) [51], multi-configuration self-consistent field (MCSCF) [52], coupled cluster methods (CC) [53] or Quantum Monte Carlo (QMC) [54]. Though these methods can be quite accurate, they are computationally very demanding for larger systems. MP2 for instance, nominally scales as N^5, thus quickly becoming prohibitively expensive with increasing cluster size.

Density-functional theory (DFT) provides an alternative approach. The electronic Schrödinger equation and the corresponding wave function are replaced by a formalism based upon the much simpler electron density ρ.

4.3. The Thomas-Fermi Model

The original idea to use the electron density as the fundamental quantity goes back to 1927, when Thomas and Fermi approximated the distribution of electrons in an atom [55] and came up with their famous kinetic energy functional which is exact in the limit of a homogeneous electron gas

$$T_{\mathrm{TF}}(\rho) = C_F \int \mathrm{d}\mathbf{r} \rho^{5/3}(\mathbf{r})\, \mathrm{d}\mathbf{r}\,, \quad C_F = \frac{3}{10}(3\pi^2)^{2/3} = 2.871\,. \tag{4.9}$$

Additionally, the attractive electron-nucleus energy and the classical repulsive electron-electron Coulomb energy can be expressed as functions of $\rho(\mathbf{r})$.

$$V_{\mathrm{Ne}} = Z \int \mathrm{d}\mathbf{r}\, \frac{\rho(\mathbf{r})}{\mathbf{r}}\,, \tag{4.10}$$

$$V_{\mathrm{ee}}^{c} = \frac{1}{2} \int\int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2\, \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\,. \tag{4.11}$$

Neglecting the exchange-correlation energy, a simple formula for the total energy of an atom is derived exclusively in terms of the electron density

$$E_{\mathrm{TF}}[\rho] = C_F \int \mathrm{d}\mathbf{r}\, \rho^{5/3}(\mathbf{r}) + Z \int \mathrm{d}\mathbf{r}\, \frac{\rho(\mathbf{r})}{\mathbf{r}} + \frac{1}{2} \int\int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2\, \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\,. \tag{4.12}$$

Minimizing Eq. (4.12) with respect to the density under the normalization constraint Eq. (4.6) then yields the ground-state density ρ_0 and the corresponding ground-state energy $E[\rho_0]$. While simple, the Thomas-Fermi formula is of no practical use due to the underlying approximations of both a very crude kinetic energy functional and the complete neglection of the exchange-correlation energy.

4.4. The Hohenberg-Kohn Theorems

The Thomas-Fermi model was the first attempt to express the electronic energy in terms of the electron density only, thereby reducing the number of degrees of freedom from $3N_{\mathrm{el}}$ to 3. Nonetheless, this approximation did not become popular due to its low accuracy. The concept of the electron density was taken up again in 1964 by Hohenberg and Kohn [56] who founded the basics of an exact theory, the *density-functional theory*. It is based upon two theorems which prove that the electron density can in principle be used alone to obtain the ground-state energy of any system.

The first theorem states that the external potential V_{ext} of a system is, within an additive constant, uniquely defined by the electron density, where the external potential is

not confined to the Coulomb-potential of the nuclei. Since V_{ext} defines the Hamiltonian, which then determines the many-body wavefunction from which all observables can be obtained, it follows that the whole physics of a system, in particular the ground-state energy, can exclusively be obtained from the electron density. For non-degenerate systems, the proof is trivial and based upon the variational principle [57]. Consider two different external potentials V_{ext} and V'_{ext} that differ by more than just an additive constant but give rise to the same ground-state electron density ρ_0. The corresponding Hamiltonians $\hat{H}$ and $\hat{H}'$ then yield two different wavefunctions Ψ and Ψ'. Applying the variational principle to the unprimed Hamiltonian with Ψ' as trial wavefunction yields

$$E_0 < \langle \Psi' | \hat{H} | \Psi' \rangle = \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle \ , \tag{4.13}$$

which expands to

$$E_0 < E'_0 + \langle \Psi' | T + V_{\text{ee}} + V_{\text{ext}} - T - V_{\text{ee}} - V'_{\text{ext}} | \Psi' \rangle \ , \tag{4.14}$$

so that the expression can be simplified to

$$E_0 < E'_0 + \int \mathrm{d}\mathbf{r} \, \rho(\mathbf{r}) (V_{\text{ext}} - V'_{\text{ext}}) \ . \tag{4.15}$$

Now the roles are exchanged, so the variational principle is applied to the primed Hamiltonian using Ψ as trial wavefunction, thus yielding

$$E'_0 < E_0 - \int \mathrm{d}\mathbf{r} \, \rho(\mathbf{r}) (V_{\text{ext}} - V'_{\text{ext}}) \ . \tag{4.16}$$

Summing Eq. (4.14) and Eq. (4.15) finally gives the following contradiction

$$E_0 + E'_0 < E'_0 + E_0 \Longleftrightarrow 0 < 0 \ . \tag{4.17}$$

Hence, the basic assumption of two different external potentials giving rise to the same ground-state density was wrong. The total energy E_0 can therefore be expressed solely in terms of ρ

$$\begin{aligned} E_0[\rho] &= \int \mathrm{d}\mathbf{r} \, \rho(\mathbf{r}) V_{\text{Ne}} + T[\rho] + E_{\text{ee}}[\rho] & (4.18) \\ &= \int \mathrm{d}\mathbf{r} \, \rho(\mathbf{r}) V_{\text{Ne}} + F_{\text{HK}}[\rho] \ . & (4.19) \end{aligned}$$

$F_{\text{HK}}[\rho]$ is called the *universal functional* of ρ since it does not depend upon the external potential. The proof can be generalized to the case of degenerate systems [57].

The second Hohenberg-Kohn theorem is a reformulation of the variational principle, now being expressed in terms of the density. It states that the energy given by any trial density $\tilde{\rho}$ is an upper bound to the ground-state energy E_0

$$E_0 = E[\rho] \leq E[\tilde{\rho}] \ . \tag{4.20}$$

According to the first Hohenberg-Kohn theorem, the trial density $\tilde{\rho}$ is uniquely given by the external potential $\tilde{V}_{\text{ext}}$, which in turn defines the corresponding Hamiltonian $\hat{H}$

which yields the many-body wavefunction $\tilde{\Psi}$. Applying the variational principle Eq. (4.4) then gives the desired result

$$\left\langle \tilde{\psi} \right| \hat{H} \left| \tilde{\psi} \right\rangle = \int \mathrm{d}\mathbf{r}\, \tilde{\rho}(\mathbf{r}) V_{\mathrm{ext}}(\mathbf{r}) + F_{\mathrm{HK}}[\tilde{\rho}] = E[\tilde{\rho}] \geq E[\rho_0] \; . \tag{4.21}$$

Additionally, applying the variational principle to the Hohenberg-Kohn functional under the normalization constraint of the electron density and the assumption of the differentiability of E_{HK}

$$\delta \left\{ E[\rho] - \mu \left[\rho(\mathbf{r}) \mathrm{d}\mathbf{r} - N_{\mathrm{el}} \right] \right\} = 0 \; , \tag{4.22}$$

yields the *Euler-Lagrange equation*

$$\mu = \frac{\delta E[\rho]}{\delta \rho(\mathbf{r})} = V_{\mathrm{ext}}(\mathbf{r}) + \frac{\delta F_{\mathrm{HK}}[\rho]}{\delta \rho(\mathbf{r})} \; , \tag{4.23}$$

with μ being the chemical potential.

4.5. The Kohn-Sham Equation

In principle, DFT is an exact theory, which means that if the universal functional F_{HK} were known, minimizing the energy functional would give the exact ground-state energy. Unfortunately, F_{HK} is not known, making a direct application of the Hohenberg-Kohn theorems impossible. In 1965, Kohn and Sham approximated F_{HK} [58], thus paving the way for tackling practical problems using DFT. Their strategy was to separate everything that can be expressed exactly and only approximate the missing terms. Introducing a reference system of non-interacting electrons described by single-particle states $\psi_i(\mathbf{r})$ the kinetic energy of that reference system can then be obtained exactly as

$$T_{\mathrm{S}} = -\frac{1}{2} \sum_{i}^{N_{\mathrm{el}}} \langle \phi_i | \nabla^2 | \phi_i \rangle \; , \tag{4.24}$$

with ϕ_i being the *Kohn-Sham orbitals*. The non-interacting kinetic energy T_{S} is not equal to the true kinetic energy T, with $T = T_{\mathrm{S}} + T_{\mathrm{C}}$. Furthermore, the classical Coulomb energy $J[\rho]$ can be expressed exactly with respect to the electron density. Therefore, Kohn and Sham rewrote the universal functional as

$$F_{\mathrm{HK}}[\rho] = T_{\mathrm{S}}[\rho] + J[\rho] + E_{\mathrm{XC}}[\rho] \; , \tag{4.25}$$

where the *exchange-correlation energy* $E_{\mathrm{XC}}[\rho]$ contains everything that is missing: The quantum-mechanical contributions due to exchange and correlation, and the remaining part of the kinetic energy not being captured by T_{S}

$$E_{\mathrm{XC}}[\rho] = (T[\rho] - T_{\mathrm{S}}[\rho]) + (E_{\mathrm{ee}}[\rho] - J[\rho]) \; , \tag{4.26}$$

which thus yields the *Kohn-Sham functional*

$$E^{\mathrm{KS}}[\rho] = T_{\mathrm{S}}[\rho] + J[\rho] + E_{\mathrm{XC}}[\rho] + E_{\mathrm{ext}}[\rho]\,. \tag{4.27}$$

Minimizing Eq. (4.27) with respect to the orbitals ϕ_i, using the expression of the electron density for a non-interacting system $\rho(\mathbf{r}) = \sum_i f_i \left|\phi_i(\mathbf{r})\right|^2$ and under the orthonormality constraint of the Kohn-Sham orbitals $\langle\phi_i|\phi_j\rangle = \delta_{ij}$, then results in the famous Kohn-Sham equation for the single-particle Kohn-Sham orbitals

$$\left[-\frac{1}{2}\nabla^2 + V_{\mathrm{eff}}(\mathbf{r})\right]\phi_i = \epsilon_i\phi_i\ , \tag{4.28}$$

with $\{\epsilon_i\}$ being the Kohn-Sham orbital energies which result as Lagrange-multipliers to ensure the orthonormality of the orbitals. V_{eff} is the effective potential, which contains the classical Coulomb potential, the exchange-correlation potential and the external potential $V_{\mathrm{ext}}(\mathbf{r})$,

$$V_{\mathrm{eff}}(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \frac{\rho(\mathbf{r})'}{|\mathbf{r}-\mathbf{r}'|} + V_{\mathrm{XC}}(\mathbf{r}) + V_{\mathrm{ext}}(\mathbf{r})\ . \tag{4.29}$$

The resulting electron density $\rho(\mathbf{r}) = \sum_i f_i \left|\phi_i(\mathbf{r})\right|^2$ is by contruction equal to the true electron density, where the occupation numbers $\{f_i\}$ are one for occupied states and zero otherwise. The exchange-correlation potential $V_{\mathrm{XC}}(\mathbf{r})$ is defined as functional derivative of the exchange-correlation energy with respect to the electron density

$$V_{\mathrm{XC}}(\mathbf{r}) \equiv \frac{\delta E_{\mathrm{XC}}}{\delta\rho(\mathbf{r})}\ . \tag{4.30}$$

Being an artificial reference system, the Kohn-Sham orbitals are no real single-particle wave functions but only approximations to them. Their introduction was in particular only necessary due to the lack of knowledge of the kinetic energy functional with respect to the electron density. Attempts to keep the density-functional theory purely density-based have been made by approximating the kinetic energy functional within the so-called *orbital-free density-functional theory* (see e.g. [59]). However, common functionals that go beyond the classical Thomas-Fermi model turned out to be little successful since the kinetic energy is in the same order of magnitude as the total energy, and errors in the approximations are therewith significant. Only in simple cases, like e.g. bulk aluminum, where the electron density resembles the homogeneous electron gas, quantitative results could be obtained [60]. For systems with localized electron densities, in particular transition metals that are of interest in the present work, the crude approximation of common existing kinetic energy functionals yields unsatisfactory results [61]. Therefore, to obtain quantitative results, classical density-functional theory based upon the Kohn-Sham equation is used, which is in principle an exact theory. Hence, if the exact forms of E_{XC} and V_{XC} were known, the Kohn-Sham scheme would lead to the exact total energy E_{tot} which in turn is the correct eigenvalue of the many-body Hamiltonian. Nevertheless, the unknown functional for the exchange-correlation energy E_{XC} and its corresponding potential V_{XC} necessitates approximations to them, therewith turning DFT into an approximative method in practice. The central goal of

modern density-functional theory is therefore to find better and better approximations to these two quantities.

4.6. Present-Day Exchange-Correlation Functionals

The quality of density-functional theory results depends upon the quality of the approximation of the exchange-correlation energy. Unfortunately, there is no systematic way to successively improve the results like in wavefunction based methods. Analogous to the Thomas-Fermi model, the first attempt to find an explicit expression for E_{XC} was based on the homogeneous electron gas. The basic assumption of the *Local-Density Approximation* (LDA) is that non-local effects of the real electron gas can be neglected and the exchange-correlation energy can be expressed as a sum over space where each point contributes to the energy with a value given by the homogeneous electron gas of the same density. In other words, the exchange-correlation energy density per particle per volume $\epsilon_{\mathrm{XC}}^{\mathrm{LDA}}(\rho(\mathbf{r}))$ in the local-density approximation equals to the true energy density $\epsilon_{\mathrm{XC}}(\rho(\mathbf{r}))$ of the corresponding homogeneous electron gas. The exchange-correlation energy is then given by

$$E_{\mathrm{XC}}^{\mathrm{LDA}}[\rho(\mathbf{r})] = \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r})\, \epsilon_{\mathrm{XC}}(\rho(\mathbf{r}))\ , \tag{4.31}$$

and the corresponding exchange-correlation potential by

$$V_{\mathrm{XC}}^{\mathrm{LDA}}(\mathbf{r}) = \frac{\delta E_{\mathrm{XC}}}{\delta \rho(\mathbf{r})} = \epsilon_{\mathrm{XC}}(\rho(\mathbf{r})) + \rho(\mathbf{r}) \frac{\partial \epsilon_{\mathrm{XC}}(\rho)}{\partial \rho}\ . \tag{4.32}$$

The exchange-correlation energy density $\epsilon_{\mathrm{XC}}(\rho(\mathbf{r}))$ can be separated into two parts, the exchange part ϵ_{X}, resulting from the Pauli principle, and the correlation part ϵ_{C}, resulting from the Coulomb correlation

$$\epsilon_{\mathrm{XC}}(\rho) = \epsilon_{\mathrm{X}}(\rho) + \epsilon_{\mathrm{C}}(\rho)\ , \tag{4.33}$$

where the exchange part $\epsilon_{\mathrm{X}}(\rho)$ can be expressed analytically for the homogeneous electron gas [55, 62]

$$\epsilon_{\mathrm{X}}(\rho(\mathbf{r})) = -\frac{3}{4} \left(\frac{3\rho(\mathbf{r})}{\pi} \right)^{\frac{1}{3}}\ . \tag{4.34}$$

For the correlation part $\epsilon_{\mathrm{C}}(\rho)$, no such analytical result is known. Expressions can be obtained for the high-density [63, 64] and the low-density [65, 66] limit. The intermediate range has only been assessed numerically by highly accurate quantum Monte Carlo calculations, e.g. by Ceperley and Alder [67]. In order to use this discrete set of values, a parameterization is required, of which the most popular ones are the one by Vosko, Wilk and Nusair [68] (VWN-LDA) and the one by Perdew and Wang [69] (PW-LDA). Both schemes give usually very similar results.

Though the local-density approximation is a crude approximation to the true exchange-correlation energy, is has been widely used until the 1990's with surprisingly good results.

This can partly be related to the fact that the magnitude of the exchange-correlation energy is rather small compared to the total energy so that errors in the approximation are less significant as in the case of the kinetic energy. Drawbacks are, however, the typical overestimation of binding energies and the underestimation of bond lengths. In particular, systems with electron density distributions far away from the homogeneous electron gas, like the atomic clusters of interest in this work, often yield disappointing results.

Improvements over LDA can be achieved by taking non-local information of the density-distribution explicitly into account. A straightforward correction to the local-density approximation is a formal expansion of E_{XC} in gradients of the density suggested by Hohenberg and Kohn [56], which yields in general a functional of the form

$$E_{\mathrm{XC}}^{\mathrm{DGE}}[\rho(\mathbf{r})] = \int \mathrm{d}\mathbf{r} \left[\epsilon_{\mathrm{XC}}^{(0)}(\rho) + \epsilon_{\mathrm{XC}}^{(1)}(\rho)\nabla\rho + \epsilon_{\mathrm{XC}}^{(2)}(\rho)\left|\nabla\rho\right|^2 + \ldots\right] . \tag{4.35}$$

However, truncating this expansion after the first order gives unsatisfying results, sometimes even being worse than LDA. Better approximations can be obtained by introducing a more generalized expansion. In this *Generalized-Gradient Approximation* (GGA), the exchange-correlation functional has the following form

$$E_{\mathrm{XC}}^{\mathrm{GGA}}[\rho(\mathbf{r})] = \int \mathrm{d}\mathbf{r} f\left(\rho(\mathbf{r}), \nabla\rho(\mathbf{r})\right) \mathrm{d}\mathbf{r} = \int \rho\epsilon_{\mathrm{XC}}^{\mathrm{GGA}}\left(\rho(\mathbf{r}), \nabla\rho(\mathbf{r})\right) , \tag{4.36}$$

which typically yields better binding energies. Similar to LDA, $E_{\mathrm{XC}}^{\mathrm{GGA}}$ can be separated into an exchange and a correlation part. There are many different flavours of GGA-functionals that are differing in the functional form of the exchange and correlation energy which is typically constructed by fulfilling known constraints of the exchange-correlation potential V_{XC}. One popular GGA-functional was proposed by Perdew, Burke and Ernzerhof in 1996 (PBE) [70] which is almost exclusively used in the present work. This functional is based upon the PW91 functional of Perdew and Wang [71] which is an analytical fit to a numerically determined first-principles GGA, but cast into a simpler functional form by neglecting energetically irrelevant terms.

4.7. Spin-Density-Functional Theory

Many systems of interest possess an odd number of electrons. While the results obtained by the formalism described so far are quite satisfactory for closed-shell systems, extensions to open-shell systems are required, in particular for transition metals which exhibit magnetic properties. In principle, density-functional theory is an exact theory and the total energy depends exclusively upon the total electron density which also holds for open-shell systems unless there is an explicit spin-dependence in the Hamiltonian, e.g. in the case of an external magnetic field. Experience has shown, however, that approximate functionals gain more flexibility when explicitly depending upon the individual spin densities ρ^σ. In the therewith resulting *Unrestricted Kohn-Sham* method [72, 73], two sets of Kohn-Sham orbitals $\phi_i^\uparrow$ and $\phi_i^\downarrow$ are introduced, and used to represent a spin-polarized system

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{eff}}^{\sigma}(\mathbf{r})\right]\phi_{i\sigma} = \epsilon_{i\sigma}\phi_{i\sigma} \ , \tag{4.37}$$

where the spin-dependency of the effective potential V_{eff}^{σ} enters through the exchange-correlation potential

$$V_{\text{XC}}^{\sigma}(\mathbf{r}) \equiv \frac{\delta E_{\text{XC}}}{\delta \rho^{\sigma}(\mathbf{r})} \ . \tag{4.38}$$

The spin-polarized version of LDA is called local-spin-density approximation (LSDA) [72] and the corresponding exchange-correlation functional is given by

$$E_{\text{XC}}^{\text{LSD}}[\rho^{\uparrow}(\mathbf{r}), \rho^{\downarrow}(\mathbf{r})] = \int \mathrm{d}\mathbf{r}\rho(\mathbf{r})\, \epsilon_{\text{XC}}(\rho^{\uparrow}(\mathbf{r}), \rho^{\downarrow}(\mathbf{r})) \ . \tag{4.39}$$

Similarly, the $E_{\text{XC}}^{\text{GGA}}$ in the spin-polarized case looks in general like

$$E_{\text{XC}}^{\text{GGA}}[\rho^{\uparrow}, \rho^{\downarrow}] = \int \mathrm{d}\mathbf{r}\rho\, \epsilon_{\text{XC}}(\rho^{\uparrow}, \rho^{\downarrow}, \nabla\rho^{\uparrow}\nabla\rho^{\downarrow}, \nabla\rho^{\downarrow}\nabla\rho^{\downarrow}, \nabla\rho^{\uparrow}\nabla\rho^{\downarrow}) \ . \tag{4.40}$$

with the special case of PBE

$$E_{\text{XC}}^{\text{PBE}}[\rho^{\uparrow}, \rho^{\downarrow}] = \int \mathrm{d}\mathbf{r}\rho\, \epsilon_{\text{XC}}(\rho^{\uparrow}, \rho^{\downarrow}, |\nabla\rho|^2) \ . \tag{4.41}$$

Part II.

The Density-Functional Theory Implementation FHI-aims

5. Solving the Kohn-Sham Equations with Numerical Atom-Centered Basis Sets

Due to the steadily increasing importance of density-functional theory, many different implementations are nowadays available. Though the underlying theory is always the same, differences arise from the actual strategy to put DFT into a practical tool. In a dominant fraction of existing implementations the main principle is always the same in that the Kohn-Sham orbitals ϕ_i are expanded into a set of basis functions φ_j

$$\phi_i(\mathbf{r}) = \sum_j c_{ij}\varphi_j(\mathbf{r}) \ . \tag{5.1}$$

A priori, the choice of a certain basis set is not prescribed and thus guided by the actual problem to tackle which leads to the manifold of electronic structure codes around. Historically, there are two main classes. *Plane waves* first arose in the context of solid state physics since they are naturally cast into periodic boundary conditions applied in solid state calculations. *Localized basis functions*, on the other hand, are suitable to treat finite systems like molecules, as the tail region of the eigenstates and the surrounding vacuum requires decaying basis sets for an efficient calculation. Common choices for the latter class of basis functions are analytically defined *Gaussians* or *Slater-like* orbitals. The implementation FHI-aims ("*ab initio* molecular simulations") developed at the Theory Department of the Fritz-Haber-Institute resorts to *numerically* defined *atom-centered* orbitals of the form

$$\varphi_i(\mathbf{r}) = \frac{u_i(r)}{r} Y_{\mathrm{lm}}(\vartheta, \phi) \ , \tag{5.2}$$

where $Y_{\mathrm{lm}}(\vartheta, \phi)$ are spherical harmonics. The radial part $u_i(r)$ is numerically tabulated, thus being very flexible since any kind of desired shape can be achieved. This paves the way for generating highly efficient species-dependent basis sets, which are furthermore constructed strictly hierarchical so that the accuracy can be continuously increased from tight-binding like to the meV level. The concept of atom-centered basis functions together with a well-defined control of convergence makes FHI-aims a suitable tool for the aspired atomic cluster calculations. The co-development of the electronic structure code, in particular the essential atomic forces for the sampling schemes (see chapter 6), constitutes an important, preparatory part of this PhD project.

5.1. The Eigenvalue Problem

The heart of FHI-aims is to solve the Kohn-Sham equation

$$\hat{h}^{\mathrm{KS}} |\phi_i\rangle = \epsilon_i |\phi_i\rangle \ , \tag{5.3}$$

with the single-particle Hamiltonian

$$\hat{h}^{\mathrm{KS}} = -\frac{1}{2}\nabla^2 + V_{\mathrm{eff}}([\rho], \mathbf{r}) \ , \tag{5.4}$$

where for simplicity the non-polarized case is considered here first. The Kohn-Sham orbitals lead to the electron density $\rho(\mathbf{r}) = \sum_i f_i |\phi_i(\mathbf{r})|^2$ which then determines the effective potential $V_{\mathrm{eff}}[\rho]$. Equations (5.3) and (5.4) have therefore to be solved self-consistently, i.e. the exact same eigenfunctions and electron density used to construct the correct Hamiltonian must re-emerge as its output. In each iteration step towards self-consistency, the orbitals are expanded in a set of basis functions φ_j.

$$\phi_i(\mathbf{r}) = \sum_j c_{ij}\varphi_j(\mathbf{r}) \ . \tag{5.5}$$

Inserting this ansatz into Eq. (5.3) and multiplying the equation with φ_i finally transforms the continuous differential equation into an algebraic generalized eigenvalue problem

$$\sum_j h_{ij}c_{jl} = \epsilon_l \sum_j s_{ij}c_{jl} \ . \tag{5.6}$$

The Hamiltonian and overlap matrix elements h_{ij} and s_{ij}, respectively, are hereby given by

$$h_{ij} = \int \mathrm{d}\mathbf{r} \left[\varphi_i(\mathbf{r})\hat{h}^{\mathrm{KS}}\varphi_j(\mathbf{r})\right] \ , \tag{5.7}$$

$$s_{ij} = \int \mathrm{d}\mathbf{r} \left[\varphi_i(\mathbf{r})\varphi_j(\mathbf{r})\right] \ . \tag{5.8}$$

The complex conjugate notation is not needed for finite systems and therewith omitted in this thesis.

As described in the previous chapter, the total energy of a system in DFT is given by the Kohn-Sham-Functional including the repulsive nuclei-nuclei interaction

$$E^{\mathrm{KS}} = T_{\mathrm{S}} + J_{\mathrm{H}} + E_{\mathrm{XC}} + E_{\mathrm{ext}} + E_{\mathrm{NN}} \ , \tag{5.9}$$

minimized over all possible densities according to the second Hohenberg-Kohn theorem. The discretization of the Hilbert space by the basis functions $\{\varphi_j\}$ leads to a minimization with respect to the expansion coefficients $\{c_{ij}\}$. The Kohn-Sham functional evaluated at the ground-state energy is therefore variational with respect to the $\{c_{ij}\}$, i.e.

$$\frac{\partial E_{\text{tot}}}{\partial c_{ij}} = 0 \tag{5.10}$$

The extension to the spin-polarized case is straightforward. Eq. (5.3) needs to be solved for both spin-channels, thus yielding two different sets of expansion coefficients $\{c_{i\sigma,j}\}$ for the corresponding orbitals $\{\phi_{i\sigma}\}$. Self-consistency then needs to be achieved for both spin-densities $\rho^{\uparrow}$ and $\rho^{\downarrow}$ due to the explicit dependency of the effective potential $V^{\sigma}_{\text{eff}}([\rho^{\uparrow}, \rho^{\downarrow}])$ on both quantities (see section 4.7).

Eq. (5.3) gives direct access to the Kohn-Sham eigenvalues ϵ_i. It is therefore convenient to rewrite the energy functional as

$$E^{\text{KS}} = \sum_i f_i \epsilon_i - \int \mathrm{d}\mathbf{r} \rho(\mathbf{r}) V_{\text{XC}}(\mathbf{r}) + E_{\text{XC}}[\rho] - \frac{1}{2} \int \mathrm{d}\mathbf{r} \rho(\mathbf{r}) V_{\text{H}}(\mathbf{r}) + E_{\text{NN}} \,, \tag{5.11}$$

which can trivially be extended to the spin-polarized case

$$E^{\text{KS}} = \sum_{i\sigma} f_{i\sigma} \epsilon_{i\sigma} - \sum_{\sigma} \int \mathrm{d}\mathbf{r} \rho^{\sigma}(\mathbf{r}) V^{\sigma}_{\text{XC}}(\mathbf{r}) + E_{\text{XC}}\left[\rho^{\uparrow}, \rho^{\downarrow}\right] - \frac{1}{2} \int \mathrm{d}\mathbf{r} \rho(\mathbf{r}) V_{\text{H}}(\mathbf{r}) + E_{\text{NN}} \,. \tag{5.12}$$

In principle, this reformulation is derived by summing up the Kohn-Sham equations, correcting the double-counting of the classical Hartree energy and replacing the exchange-correlation potential energy $\int \mathrm{d}\mathbf{r} \rho(\mathbf{r}) V_{\text{XC}}(\mathbf{r})$ by the exchange-correlation energy E_{XC} to obtain the correct total energy.

5.1.1. Achieving Self-Consistency

Density mixing

The main problem of an electronic structure calculation is to achieve self-consistency, i.e. the correct electronic Hamiltonian h_{ij} should yield as output the same wave functions, electronic charge density and potentials that were used to construct it, which means $\rho^{(n)}_{\text{out}} = \rho^{(n)}_{\text{in}}$. To achieve self-consistency, one therefore starts with an initial trial density $\rho^{(0)}$ which in FHI-aims is constructed by superimposing the densities of the constituent atoms. Solving the Kohn-Sham equation then yields orbitals $\{\phi_i\}^{(0)}$ that lead to an output electron density $\rho^{(0)}_{\text{out}}$ which is used to construct the next input density $\rho^{(1)}_{\text{in}}$. This whole procedure is repeated until finally the output density equals the input density and thus self-consistency is achieved. The construction of the next input density is a crucial technical detail that significantly determines the performance of the calculation. The simplest scheme is to linearly mix the output density with the previous input density

$$\rho^{(n)}_{\text{in}} = \alpha \rho^{(n)}_{\text{out}} + (1 - \alpha) \rho^{(n-1)}_{\text{out}} \,, \tag{5.13}$$

with α being an adjustable parameter. If chosen too small, convergence is poor since the electron density approaches the self-consistent density very slowly. Choosing a too

large value might destabilize the electron density, resulting into an oscillating behaviour or even no convergence at all.

More sophisticated is the *Pulay* mixing scheme [74, 75] that is exclusively used in the present work. In that scheme, the input charge densities are stored for a number of iteration steps. The key quantity in this mixing scheme is the *charge density residual* defined as

$$R[\rho_{\text{in}}] = \rho_{\text{out}}([\rho_{\text{in}}], \mathbf{r}) - \rho_{\text{in}}(\mathbf{r}) = \Delta\rho(\mathbf{r}) \ , \tag{5.14}$$

which describes the change of the electron charge density at every point in space after a self-consistency iteration. If self-consistency is achieved, the norm of the residual vanishes

$$\langle R[\rho_{\text{in}}] \, | R[\rho_{\text{in}}] \rangle = 0 \ , \tag{5.15}$$

where the norm corresponds to a simple integral

$$\langle R[\rho] \, | R[\rho] \rangle = \int \mathrm{d}\mathbf{r} (\Delta\rho)^2(\mathbf{r}) \ . \tag{5.16}$$

Hence, the density has converged to the self-consistent one at every point in space, since $\Delta\rho$ must equal zero everywhere. Achieving self-consistency therefore corresponds to minimizing the charge density residual.

A new optimal input charge density is obtained in each step as a linear combination of the input charge densities of all previous steps

$$\rho_{\text{in}}^{\text{opt}} = \sum_i \alpha_i \rho_{\text{in}}^{(i)} \ , \tag{5.17}$$

where the linear coefficients obviously have to fulfil the following condition to conserve the norm of the density

$$\sum_i \alpha_i = 1 \ . \tag{5.18}$$

The underlying assumption of the Pulay mixing scheme is the linearity of the residual so that the residual vector of the optimum charge density is given by

$$R[\rho_{\text{in}}^{\text{opt}}] = R\left[\sum_i \alpha_i \rho_{\text{in}}^{(i)}\right] = \sum_i \alpha_i R[\rho_{\text{in}}^{(i)}] \ . \tag{5.19}$$

Minimizing the norm of the residual

$$\left\langle R[\rho_{\text{in}}^{\text{opt}}] \, \middle| R[\rho_{\text{in}}^{\text{opt}}] \right\rangle \ , \tag{5.20}$$

then leads to a system of equations for the coefficients α_i which finally yields the optimum input density. To prevent the electron charge density from being trapped in a subspace spanned by the previous densities, a fraction of the residual vector is added to finally obtain the new input density which is then used to construct the next Hamiltonian

$$\rho_{\text{in}}^{(n+1)} = \rho_{\text{in}}^{\text{opt}} + \gamma R[\rho_{\text{in}}^{\text{opt}}] \,, \tag{5.21}$$

with γ being an adjustable parameter.

For the spin-polarized case, it is crucial to combine both spin-channels to one single residual vector

$$\left\langle R[\rho^{\uparrow}, \rho^{\downarrow}] \middle| R[\rho^{\uparrow}, \rho^{\downarrow}] \right\rangle = \sum_{\sigma} \int \mathrm{d}\mathbf{r} \left(\Delta\rho^{\sigma}\right)^2 (\mathbf{r}) \,, \tag{5.22}$$

which is then used to determine a single set of coefficients α_i for both spin-densities $\rho^{\uparrow}$ and $\rho^{\downarrow}$, thus ensuring stable convergence.

The number of stored densities and residuals is a key parameter and can significantly influence the performance. If chosen too small, too little information is taken into account and the convergence is slow. In case of too many previous densities taken into account, the individual densities might be too different, so that the assumption of linearity is not justified and the Pulay matrix becomes singular.

Occupation smearing

In systems with degenerate or quasi-degenerate eigenstates near the Fermi level, the occupation of the eigenstates might oscillate during the self-consistency iterations since the eigenstates permanently cross the Fermi level. Additional stability can then be achieved by occupying the states by a distribution $\{f_{i\sigma}\}$ which is slightly broadened about the Fermi level, thus weakening the effect of a level-crossing since the occupation then does not change discontinuosly. FHI-aims supports Fermi-smearing [76], Methfessel-Paxton [77] and Gaussian [78], where the latter is exclusively used in the present work when necessary. The occupation numbers are then given by

$$f_{i\sigma} = \frac{1}{2}\left(1 - \mathrm{erf}\left[\frac{\epsilon_{i\sigma} - \mu}{w}\right]\right) \,. \tag{5.23}$$

Since the PES is therewith effectively distorted, one aims at a value for the smearing width w as small as possible to not change the physics of the system significantly. Since clusters typically show Jahn-Teller distortions and therewith a HOMO-LUMO gap, w can typically be set to zero once the system is close to the local minimum. Small finite values for w in the range of tenths of eV are only chosen for the atomic configuration being far away from the local minimum.

5.2. Numerical Atom-Centered Basis Functions

5.2.1. The Definition of the Basis Functions

In FHI-aims the radial part of the basis function given by Eq. (5.2) is obtained by solving a radial Schrödinger-like equation on a logarithmic grid. A steeply increasing confinement potential is thereby used to spatially confine the basis functions to a local

region, thus allowing for an O(N)-like scaling of the integrations for large systems. The shape of the potential is chosen to ensure a continuous second derivative at the onset of the potential which then increases gradually to infinity at the cutoff-radius r_{cut}. One clear advantage over analytically defined basis functions like Gaussians or Slater-like functions is the greater flexibility since any kind of potential in the Schrödinger-like equation can be chosen. In FHI-aims the *minimal* basis of a species is constructed by choosing the effective Kohn-Sham potential of the corresponding non-polarized, spherically symmetric free atom. This basis thus facilitates the all-electron calculation since the oscillatory behaviour of the wavefunctions in the core-region are already well described at this level. Additionally the minimal basis avoids the atomic basis set superposition error (BSSE) which can typically be observed for analytical localized basis sets: When calculating the cohesive energy of an X_N-cluster

$$E_{\text{coh}} = \left[E_{\text{tot}}(X_N) - NE_{\text{tot}}(X)\right]/N \ , \tag{5.24}$$

the energy contribution of a constituent atom to the total energy E_{tot} might be variationally improved by basis functions sitting on adjacent atoms, thus leading to an overestimating of E_{coh}. Using atomic states, however, the total energy is already converged at the level of the minimal basis, and neighbouring basis functions have no effect. The BSSE might still occur and should be checked in case of *intermolecular* binding energies, though.

The minimal basis set is augmented by further classes of basis functions, one of which is formed by ion-like radial functions that are in particular suitable for describing a chemical bond as demonstrated by Delley [79]. These are states obtained from calculations of positive ions, which are supposed to describe the charge transfer of a system in a chemical bond. Hydrogen-like functions are used in addition, especially as *polarisation functions* for angular momenta beyond those present in the free atom itself. These functions are derived from one-electron atoms with an arbitrary nuclear charge, that does not need to be an integer. Since radial functions originating from different potentials are not necessarily orthonormal to one another even on the same atomic site, all on-site radial functions are orthonormalized explicitly using a simple Gram-Schmidt process [80].

Contrary to Gaussians or Slater-like functions, numerically tabulated basis functions do not allow for an analytical integration, thus requiring a numerical method that is described in section (5.3).

5.2.2. The Generation of Basis Functions

The major challenge when dealing with atom-centered basis functions is the systematic convergence towards a complete basis set limit (cbs) with an as small and thus as efficient basis set as possible. Contrary to plane waves, for which the cutoff energy can be successively increased until the desired accuracy in the total energy has been achieved, there is no such simple parameter for numerical basis functions. All one can be sure of is the variational principle according to which the total energy improves when a given basis set is augmented by further basis functions. In order to produce structure-independent,

transferable basis sets, that can later on be used for production, the strategy pursued by FHI-aims is therefore to choose the best basis functions out of a huge pool of candidate radial functions. This contains ion-like functions [79, 81] of different main and angular quantum numbers and hydrogen-like functions covering a broad range of effective charges [79]. Starting from the minimal free-atom basis of size $N_{\mathrm{min-basis}}$, the entire pool of candidate functions is sampled, adding each function to the given basis set in turn. The radial function with angular momentum l that gives the single largest improvement of a target total energy is added to the original basis set, increasing it to ($N_{\mathrm{min-basis}}$ + (2l+1)). This step is then repeated, adding again each candidate function with angular momentum $\tilde{l}$ to the basis set of size $(N_{\mathrm{min-basis}} + (2l+1))$ in turn and then increase it to size $(N_{\mathrm{min-basis}} + (2l+1) + (2\tilde{l}+1))$ with the next best function. The whole procedure is performed until no further significant total energy improvements result. Regarding the optimization target for a certain chemical element, the corresponding dimer is chosen which constitutes the simplest possible chemical bond and further a rather demanding test case for atom-centered orbitals [82], since no basis function overlap from further centers can accidentally improve the total energy. The improvement resulting from adding a certain trial basis function to the basis set is then defined as

$$\Delta_{\mathrm{basis}} = \frac{1}{N_d} \sum_{i}^{N_d} \left[\mathcal{E}_{\mathrm{basis}}(d_i) - \mathcal{E}_{\mathrm{cbs}}(d_i) \right] \ , \tag{5.25}$$

where $\mathcal{E}_{\mathrm{basis}}(d_i)$ denotes the non-self-consistent total energy for the dimer at the bond distance d_i. The non-self-consistent reference energy $\mathcal{E}_{\mathrm{cbs}}(d_i)$ for a converged basis set is obtained independently of the present procedure, by converging a very large and thus inefficient, but formally systematic basis set of confined atomic excited-state functions. The dimer distances $\{d_i\}$ are spread to sample the self-consistent LDA binding curve. Experience has shown [11] that non-selfconsistent energies are a sufficient optimization target and transferable to self-consistent calculations. Furthermore, potential instabilities of the self-consistency-cycle for pathological cases are therewith avoided.

The different basis functions resulting from the basis set generation typically arise in groups of different angular momenta, *spd* or *spdf*, and are thus organized in so-called *tiers* which contain a basis function of each angular momentum. The number of the tier thereby denotes the accuracy of the basis set. The different basis sets used in the present work are given in detail in Appendix A.

5.3. Numerical Integration

Numerical integrations are performed for several tasks. The dominant part in the computational effort constitutes the calculation of the overlap, Eq. (5.8), and Hamiltonian matrix, Eq. (5.7), which scale formally as $\mathrm{O}(N^3)$ with the system size. In FHI-aims , the numerical integration is based upon a partitioning technique [79, 83] which decomposes a function to pieces sitting on atoms that are then integrated by an atom-centered numerical grid. The distribution among the atoms is thereby done by an atom-centered partitioning function p_α

$$\int \mathrm{d}\mathbf{r} f(\mathbf{r}) = \sum_{\alpha} \int \mathrm{d}\mathbf{r}\; p_{\alpha}(\mathbf{r}) f(\mathbf{r}) \; . \tag{5.26}$$

The sum of all partitioning functions equals one which is achieved by the normalization

$$p_{\alpha}(\mathbf{r}) = \frac{g_{\alpha}(\mathbf{r})}{\sum_{\beta} g_{\beta}(\mathbf{r})} \; . \tag{5.27}$$

g_{α} is hereby strongly peaked close to the originating atom thus integrating the function preferably on the grid sitting on the next nearest atom. In FHI-aims this is achieved by an approach similar to the Hirshfeld partitioning scheme [84], which is based upon the electron density of non-spinpolarized, spherical free atoms $\rho_{\alpha}^{\mathrm{free}}$ as suggested by Delley [79]

$$g_{\alpha}(\mathbf{r}) = \frac{\rho_{\alpha}^{\mathrm{free}}}{r^2} \; . \tag{5.28}$$

Every single-atom centered integrand is then integrated over N_{r} radial integration shells each containing N_{ang} angular integration points,

$$\int \mathrm{d}\mathbf{r}\; p_{\alpha}(\mathbf{r}) f(\mathbf{r}) \approx \sum_{\mathbf{r}_{i\alpha}} \Delta\mathbf{r}_{i\alpha}\; p_{\alpha}(\mathbf{r}_{i\alpha}) f(\mathbf{r}_{i\alpha}) = \sum_{s}^{N_{\mathrm{r}}} \sum_{t}^{N_{\mathrm{ang}}} \Delta\mathbf{r}_{i\alpha}(s,t)\; p_{\alpha}(\mathbf{r}_{i\alpha}) f(\mathbf{r}_{i\alpha}(s,t)) \; , \tag{5.29}$$

where the grid weights $\mathbf{r}_{i\alpha}$ can simply be obtained by the analytically given integration points. Summarized, any kind of integral in FHI-aims is approximated by a discrete sum of the form

$$\int \mathrm{d}\mathbf{r} f(\mathbf{r}) \approx \sum_{\alpha} \sum_{\mathbf{r}_{i\alpha}} w_{\mathbf{r}_{i\alpha}} f(\mathbf{r}_{i\alpha}) \; , \tag{5.30}$$

with the integration weights $w_{\mathbf{r}_{i\alpha}}$ given by the product of the partition function and the weights of the individual atom-centered grid $w_{\mathbf{r}_{i\alpha}} = p_{\alpha}(\mathbf{r}) \Delta\mathbf{r}_{i\alpha}$.

The radial grid is defined as suggested by Baker [85]

$$r(s) = r_{\mathrm{outer}} \frac{\log\left(1 - \left[s/\left(N_{\mathrm{r}}+1\right)\right]^2\right)}{\log\left(1 - \left[N_{\mathrm{r}}/\left(N_{\mathrm{r}}+1\right)\right]^2\right)} \; , \tag{5.31}$$

which provides radial shells that are dense in the core region and then becomes successively coarser with increasing distance of the nucleus. Thus, the fast-varying localized wave function parts near the nuclei are captured, together with an efficient treatment of the smoother parts in the interstitial and far-field regions. r_{outer} gives the outermost radial shell and is chosen as 7 Å for all species and thus contains the whole basis function for cut-off radii $r_{\mathrm{cut}} < 7$Å that are used in this work. The number of radial grid points N_{r} further scales with the nuclear charge [85] so that heavier elements with steeper core states are integrated more accurately:

$$N_{\mathrm{r}} = 16.8 \left(Z+2\right)^{1/3} \; . \tag{5.32}$$

A uniform accuracy increase can be obtained by placing additional shells at integer fractions $N_{\mathrm{r,div}}$ of the original grid, e.g. at $s = \frac{1}{2}$, $s = \frac{3}{2}$, ... , $s = 2N_{\mathrm{r}} + \frac{1}{2}$ for $N_{\mathrm{r,div}} = 2$.

As angular grids, the *Lebedev* grids [86, 87, 88] are chosen in a special version provided by Delley [89]. The corresponding integration points have octahedral symmetry and are constructed in such a way as to integrate angular momentum functions up to a certain order exactly [90, 91].

The number of angular grid points is not chosen fixed for all distances. Close to the nuclei with a dense radial integration grid and small surfaces of the radial shells less angular points are required compared to large distances. In FHI-aims the required number N_{ang} for a given radial integration shell can therefore be determined adaptively by converging the initial overlap matrix elemets s_{ij} and the initial Hamiltonian matrix elements h_{ij} prior to the production run. An upper and lower bound is given by the parameter $N_{\mathrm{ang,max}}$ and $N_{\mathrm{ang,min}}$, respectively.

Formally, the numerical integration scales as $\mathrm{O}(N^3)$ with the system size, since all pairs of basis functions ($\propto N^2$) must be integrated across the entire system ($\propto N$). Due to the localization of the basis functions, this scaling is reduced to $\mathrm{O}(N)$ for large systems since the number of non-vanishing basis functions at a certain grid point becomes independent of the system size.

5.4. The Hartree-Potential

The calculation of the electrostatic potential V_{H} constitutes a further challenge. A direct integration of the Hartree potential in Eq. (4.29) would be prohibitively expensive as it requires a whole integration for each grid point. In FHI-aims the calculation is therefore much simplified by a multipole-decomposition method as described by Delley [79]. First, the superimposition of the non-polarized, spherical free atoms is subtracted from the electron density, leaving the much smoother difference electron density due to the chemical bond. The free electron density together with the corresponding potentials are calculated in the preparation of the production and are therewith known, thus allowing for a reconstruction of the full electrostatic potential .

$$\Delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \sum_{\alpha} \rho_{\alpha}^{\mathrm{free}}\left(|\mathbf{r} - \mathbf{R}_{\alpha}|\right) \ . \tag{5.33}$$

Using the partitioning scheme presented in the previous section, the electron density is allocated to the individual atoms, followed by a multipole expansion

$$\Delta\rho_{\alpha,lm}^{\mathrm{MP}}(r) = \int_{r=|\mathbf{r}-\mathbf{R}_{\alpha}|} \mathrm{d}^2\Omega_{\alpha}\, p_{\alpha}(\mathbf{r}) \cdot \Delta\rho(\mathbf{r}) \cdot Y_{\mathrm{lm}}(\Omega_{\alpha}) \ , \tag{5.34}$$

where Ω_{α} is a short-hand notation for the spherical coordinates (ϑ, φ) with respect to atom α. With the multipole components of the electron density $\Delta\rho_{\alpha,lm}^{\mathrm{MP}}$, the corresponding components of the Hartree potential $\Delta V_{\alpha,lm}$ can then efficiently be obtained by integrating the Poisson equation on a one-dimensional, logarithmic radial grid

$$\nabla^2(\Delta V_{\alpha,lm}) = -4\pi\Delta\rho_{\alpha,lm}^{\mathrm{MP}} \ . \tag{5.35}$$

The individual components are then reassembled to finally obtain the Hartree potential of the difference electron density

$$\Delta V_{\mathrm{H}}^{\mathrm{MP}}(\mathbf{r}) = \sum_{\alpha,lm}^{l_{\max}} \Delta V_{\alpha,lm}\left(|\mathbf{r}-\mathbf{R}_\alpha|\right) Y_{\mathrm{lm}}(\Omega_\alpha)\ . \tag{5.36}$$

The multipole expansion is truncated at a maximum angular momentum $l_{\max}$, thus introducing an expansion error. According to Dunlap *et al.* [92], the first-order expansion error can be eliminated by modifiying the electrostatic double-counting correction in the Kohn-Sham-functional, Eq. (5.11), like

$$-\frac{1}{2}\int \mathrm{d}\mathbf{r}\rho V_{\mathrm{H}}[\rho^{\mathrm{MP}}] \ \longrightarrow\ -\frac{1}{2}\int \mathrm{d}\mathbf{r}\rho^{\mathrm{MP}} V_{\mathrm{HH}}[\rho^{\mathrm{MP}}]\ , \tag{5.37}$$

with

$$\rho^{\mathrm{MP}}(\mathbf{r}) = \sum_\alpha \rho_\alpha^{\mathrm{free}}(|\mathbf{r}-\mathbf{R}_\alpha|) + \sum_{\alpha,lm} \Delta\rho_{\alpha,lm}^{\mathrm{MP}}(|\mathbf{r}-\mathbf{R}_\alpha|))Y_{\mathrm{lm}}(\Omega_\alpha)\ , \tag{5.38}$$

so that quadratic total energy convergence with respect to $l_{\max}$ can be achieved.

6. Energy Derivatives in FHI-aims

The local structural relaxation is a key ingredient to the success of many global optimization schemes discussed in chapter 3, since it effectively removes the transition state regions during the exploration of the PES. Efficient structural relaxation thereby requires the knowledge of the atomic forces which are the negative first derivative of the total energy. The simplest method to obtain energy derivatives is the finite difference scheme which approximates the gradient by displacing the atomic positions like

$$\frac{\mathrm{d}E_{\mathrm{tot}}}{\mathrm{d}\mathbf{R}_{\alpha,x}} \approx \frac{E_{\mathrm{tot}}(\mathbf{R}_{\alpha,x} + \Delta\mathbf{R}_{\alpha,x}) - E_{\mathrm{tot}}(\mathbf{R}_{\alpha,x} - \Delta\mathbf{R}_{\alpha,x})}{2\Delta\mathbf{R}_{\alpha,x}} \,. \tag{6.1}$$

Technical details on higher-order finite difference schemes are given in Appendix D. In principle, derivatives can thereby be evaluated up to any desired accuracy by choosing a significantly small displacement step width Δ. However, this requires the calculation of the total energy of all displaced atomic configurations for each single relaxation step, thus increasing the overall computational burden by a factor of $6N$ for the simplest finite difference scheme. Hence, analytical derivatives that can directly be obtained for a given atomic configurations are desirable, which significantly decrease the computational cost of a local structural relaxation. Accurate forces are furthermore a prerequisite for obtaining reliable vibrational spectra as used in the application described in chapter III. The implementation and critical examination of the accuracy of the analytical atomic forces in FHI-aims has therefore been a major part of the present work.

6.1. The Calculation of Analytical Atomic Forces

The forces are defined as the negative gradient of the total energy with respect to the nuclear coordinates

$$\mathbf{F}_{\alpha} = -\frac{\mathrm{d}E_{\mathrm{tot}}}{\mathrm{d}\mathbf{R}_{\alpha}} \,. \tag{6.2}$$

The total energy in FHI-aims is the minimum of the Kohn-Sham-functional E^{KS} under the orthonormalization constraint of the Kohn-Sham orbitals

$$E_{\mathrm{tot}} = \min_{\{c_{i\sigma,j}\}} \left(E^{\mathrm{KS}} - \sum_{i\sigma,j\sigma'} f_{i\sigma}\epsilon_{i\sigma}\delta_{i\sigma,j\sigma'} \left(\langle \phi_{i\sigma} \mid \phi_{j\sigma'} \rangle - 1 \right) \right) \tag{6.3}$$

$$= \min_{\{c_{i\sigma,j}\}} \left(E^{\mathrm{KS}} - \sum_{i\sigma} f_{i\sigma}\epsilon_{i\sigma} \left(\langle \phi_{i\sigma} \mid \phi_{i\sigma} \rangle - 1 \right) \right) = E_{\mathrm{tot}} \left[\{c_{i\sigma,j}\}, \{\mathbf{R}_{\alpha}\} \right] \,, \tag{6.4}$$

where the sum goes over all states $\phi_{i\sigma}$ with the eigenvalue $\epsilon_{i\sigma}$. From this, it follows that

$$\begin{aligned} \mathbf{F}_\alpha &= -\frac{\mathrm{d}E_{\mathrm{tot}}}{\mathrm{d}\mathbf{R}_\alpha} && (6.5) \\ &= -\frac{\partial E_{\mathrm{tot}}}{\partial \mathbf{R}_\alpha} - \sum_{i\sigma,j} \underbrace{\frac{\partial E_{\mathrm{tot}}}{\partial c_{i\sigma,j}}}_{=0} \frac{\partial c_{i\sigma,j}}{\partial \mathbf{R}_\alpha} \,. && (6.6) \end{aligned}$$

Since at self-consistency the total energy is variational with respect to $\{c_{i\sigma,j}\}$, the corresponding term for the forces vanishes and only the partial derivative remains

$$\mathbf{F}_\alpha = -\frac{\partial E_{\mathrm{tot}}}{\partial \mathbf{R}_\alpha} = -\frac{\partial E^{\mathrm{KS}}}{\partial \mathbf{R}_\alpha} - \frac{\partial}{\partial \mathbf{R}_\alpha} \sum_{i\sigma} \epsilon_{i\sigma} \left(\langle \phi_{i\sigma} | \, \phi_{i\sigma} \rangle - 1 \right), \tag{6.7}$$

with

$$E^{\mathrm{KS}} = T_{\mathrm{S}} + J_{\mathrm{H}} + E_{\mathrm{XC}} + E_{\mathrm{ext}} + E_{\mathrm{NN}} \,, \tag{6.8}$$

with the kinetic energy

$$T_{\mathrm{S}} = -\frac{1}{2} \sum_{i\sigma} f_{i\sigma} \, \langle \phi_{i\sigma} | \, \nabla^2 \, | \phi_{i\sigma} \rangle \,, \tag{6.9}$$

the Hartree energy

$$J_{\mathrm{H}} = \frac{1}{2} \int \int \mathrm{d}\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \,, \tag{6.10}$$

the exchange-correlation energy

$$E_{\mathrm{XC}} = \int \mathrm{d}\mathbf{r} \rho(\mathbf{r}) \epsilon_{\mathrm{XC}} \,, \tag{6.11}$$

the electron-nuclei energy

$$E_{\mathrm{ext}} = \int \mathrm{d}\mathbf{r} \rho(\mathbf{r}) V_{\mathrm{Ne}}(\mathbf{r}) \,, \tag{6.12}$$

with

$$V_{\mathrm{Ne}}(\mathbf{r}) = \sum_\beta \frac{Z_\beta}{|\mathbf{r} - \mathbf{R}_\beta|} \,, \tag{6.13}$$

the nuclei-nuclei energy

$$E_{\mathrm{NN}} = \sum_{\alpha < \beta} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \,. \tag{6.14}$$

It is worth to point out that it is the variational property of the energy functional that fortunately allows for a direct evaluation of the analytical derivative since the total derivative therewith reduces to the partial derivative, thus making the partial derivative of the variational parameters with respect to the nuclear coordinates $\partial c_{i\sigma,j} / \partial \mathbf{R}_\alpha$ dispensable which would not directly be accessible.

6.1.1. The Individual Derivative Terms

The Kinetic Energy

Just taking the partial derivative in a straightforward way yields

$$-\frac{\partial T_{\mathrm{S}}}{\partial \mathbf{R}_\alpha} = \frac{1}{2}\frac{\partial}{\partial \mathbf{R}_\alpha}\sum_{i\sigma} f_{i\sigma}\left\langle \phi_{i\sigma}\left|\nabla^2\right|\phi_{i\sigma}\right\rangle \tag{6.15}$$

$$= \frac{1}{2}\sum_{i\sigma} f_{i\sigma}\left(\left\langle \frac{\partial\phi_{i\sigma}}{\partial \mathbf{R}_\alpha}\right|\nabla^2\left|\phi_{i\sigma}\right\rangle + \left\langle \phi_{i\sigma}\left|\nabla^2\right|\frac{\partial\phi_{i\sigma}}{\partial \mathbf{R}_\alpha}\right\rangle\right) \tag{6.16}$$

$$= -2\sum_{i\sigma} f_{i\sigma}\left\langle \frac{\partial\phi_{i\sigma}}{\partial \mathbf{R}_\alpha}\right|-\frac{1}{2}\nabla^2\left|\phi_{i\sigma}\right\rangle \ . \tag{6.17}$$

The nuclear gradient of the orbitals can be obtained by

$$\frac{\partial\phi_{i\sigma}}{\partial \mathbf{R}_\alpha} = \sum_{j(\alpha)} c_{i\sigma,j}\,\frac{\partial}{\partial \mathbf{R}_\alpha}\varphi_j\left(|\mathbf{r}-\mathbf{R}_\alpha|\right) \ , \tag{6.18}$$

$$= -\sum_{j(\alpha)} c_{i\sigma,j}\nabla\varphi_j\left(|\mathbf{r}-\mathbf{R}_\alpha|\right) \ . \tag{6.19}$$

since

$$\frac{\partial}{\partial \mathbf{R}_\alpha}\varphi_j\left(|\mathbf{r}-\mathbf{R}_\alpha|\right) = -\nabla\varphi_j\left(|\mathbf{r}-\mathbf{R}_\alpha|\right), \tag{6.20}$$

and $j(\alpha)$ indicates a basis function sitting on atom α. The remaining basis functions do not contribute to the forces on atom α since they do not depend upon $\mathbf{R}_\alpha$. So the nuclear gradients can be expressed in terms of spatial gradients of the basis functions which can then easily be evaluated numerically.

The Hartree Energy

Similar to the kinetic energy, the Hartree energy can be derived in a straightforward way

$$-\frac{\partial J_{\mathrm{H}}}{\partial \mathbf{R}_\alpha} = -\frac{1}{2}\frac{\partial}{\partial \mathbf{R}_\alpha}\int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\,\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \ , \tag{6.21}$$

$$= -\int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\,\frac{\partial\rho(\mathbf{r})}{\partial \mathbf{R}_\alpha}\,\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \ , \tag{6.22}$$

$$= -\int \mathrm{d}\mathbf{r}\,\frac{\partial\rho(\mathbf{r})}{\partial \mathbf{R}_\alpha}\,V_{\mathrm{H}}(\mathbf{r}) \ . \tag{6.23}$$

Since $\rho(\mathbf{r}) = \sum_{i\sigma} f_{i\sigma}\phi_{i\sigma}(\mathbf{r})\cdot\phi_{i\sigma}(\mathbf{r})$, the derivative of the density is given by

$$\frac{\partial\rho(\mathbf{r})}{\partial \mathbf{R}_\alpha} = 2\sum_{i\sigma} f_{i\sigma}\frac{\partial\phi_{i\sigma}(\mathbf{r})}{\partial \mathbf{R}_\alpha}\cdot\phi_{i\sigma}(\mathbf{r}) \ , \tag{6.24}$$

which finally yields as derivative of the Hartree energy

$$-\frac{\partial J_{\mathrm{H}}}{\partial \mathbf{R}_\alpha} = -2\sum_{i\sigma} f_{i\sigma} \left\langle \frac{\partial \phi_{i\sigma}}{\partial \mathbf{R}_\alpha} \middle| V_{\mathrm{H}}(\mathbf{r}) \middle| \phi_{i\sigma} \right\rangle \ , \tag{6.25}$$

which boils down to calculating the nuclear gradients of the orbitals analogous to the kinetic energy term.

The Multipole-Correction Term

The above derivation of the Hartree energy was based upon the true electronic density. In FHI-aims, however, the Hartree potential is based upon a multipole expansion of the density ρ^{MP} instead of the true electron density ρ. Additionally, the first-order correction of the total energy by Dunlap [92] due to the multipole expansion error yields a further correction term. So the Hartree energy term as it is precisely implemented in FHI-aims is given by

$$J_{\mathrm{H}} = \int \mathrm{d}\mathbf{r} \rho V_{\mathrm{H}}[\rho^{\mathrm{MP}}] - \frac{1}{2}\int \mathrm{d}\mathbf{r} \rho^{\mathrm{MP}} V_{\mathrm{H}}[\rho^{\mathrm{MP}}] \ , \tag{6.26}$$

which then yields by a straightforward derivation the additional correction term

$$\mathbf{F}_{\mathrm{MP},\alpha} = -\int \mathrm{d}\mathbf{r} \left(\rho(\mathbf{r}) - \rho^{\mathrm{MP}}(\mathbf{r})\right) \frac{\partial V_{\mathrm{H}}[\rho^{\mathrm{MP}}]}{\partial \mathbf{R}_\alpha} \ , \tag{6.27}$$

which is in agreement with the expression given by Delley [93]. As one can immediately see, the multipole-correction term vanishes, if the multipole density would be equal to the true density. The nuclear gradients of the Hartree potential are then obtained in complete analogy to the nuclear gradients of the Kohn-Sham orbitals and can be boiled down to calculating the spatial gradients. Since the multipole expansion of the Hartree potential is composed of nuclear contributions

$$V_{\mathrm{H}}(\mathbf{r}) = \sum_\beta V_\beta \left(\mathbf{r} - \mathbf{R}_\beta\right) \ , \tag{6.28}$$

only the components sitting on atom α contribute to its forces

$$\frac{\partial V_{\mathrm{H}}}{\partial \mathbf{R}_\alpha} = \frac{\partial}{\partial \mathbf{R}_\alpha} V_\alpha \left(\mathbf{r} - \mathbf{R}_\alpha\right) = -\nabla V_\alpha \left(\mathbf{r} - \mathbf{R}_\alpha\right) \ . \tag{6.29}$$

The Exchange-Correlation Energy

Assuming here a local-density approximation of the exchange-correlation functional yields the corresponding force term which is given in the non-polarized case by

$$-\frac{\partial E_{\mathrm{XC}}}{\partial \mathbf{R}_\alpha} = -\int \mathrm{d}\mathbf{r} \frac{\partial \rho(\mathbf{r})}{\partial \mathbf{R}_\alpha} \left(\epsilon_{\mathrm{XC}}(\rho(\mathbf{r})) + \rho(\mathbf{r}) \frac{\partial \epsilon_{\mathrm{XC}}}{\partial \rho(\mathbf{r})}\right) \tag{6.30}$$

$$= -2\sum_i f_i \left\langle \frac{\partial \phi_i}{\partial \mathbf{R}_\alpha} \middle| V_{\mathrm{XC}}(\mathbf{r}) \middle| \phi_i \right\rangle \ . \tag{6.31}$$

In the spin-polarized case the expression is completely analogous

$$-\frac{\partial E_{\mathrm{XC}}}{\partial \mathbf{R}_\alpha} = -2\sum_{i\sigma} f_{i\sigma} \left\langle \frac{\partial \phi_{i\sigma}}{\partial \mathbf{R}_\alpha} \middle| V^{\sigma}_{\mathrm{XC}}(\mathbf{r}) \middle| \phi_{i\sigma} \right\rangle . \tag{6.32}$$

The Electron-Nuclei Energy

The electrostatic energy between the electrons and nuclei yields another force contribution

$$\begin{aligned} -\frac{\partial E_{\mathrm{ext}}}{\partial \mathbf{R}_\alpha} &= -\int \mathrm{d}\mathbf{r} \frac{\partial \rho(\mathbf{r})}{\partial \mathbf{R}_\alpha} V_{\mathrm{Ne}}(\mathbf{r}) - \int \mathrm{d}\mathbf{r} \rho(\mathbf{r}) \frac{\partial V_{\mathrm{Ne}}}{\partial \mathbf{R}_\alpha} & (6.33) \\ &= -2\sum_{i\sigma} f_{i\sigma} \left\langle \frac{\partial \phi_{i\sigma}}{\partial \mathbf{R}_\alpha} \middle| V_{\mathrm{Ne}}(\mathbf{r}) \middle| \phi_{i\sigma} \right\rangle - Z_\alpha \int \mathrm{d}\mathbf{r} \rho(\mathbf{r}) \frac{\mathbf{R}_\alpha - \mathbf{r}}{|\mathbf{R}_\alpha - \mathbf{r}|^3} . & (6.34) \end{aligned}$$

The Nuclei-Nuclei Energy

The electrostatic forces between the nuclei are trivially given by

$$-\frac{\partial E_{\mathrm{NN}}}{\partial \mathbf{R}_\alpha} = -Z_\alpha \int \mathrm{d}\mathbf{r} \rho(\mathbf{r}) \frac{\mathbf{R}_\alpha - \mathbf{r}}{|\mathbf{R}_\alpha - \mathbf{r}|^3} . \tag{6.35}$$

The Normalization Constraint

Finally the constraint term in the Lagrange functional needs to be taken into account

$$\frac{\partial}{\partial \mathbf{R}_\alpha} \sum_{i\sigma} f_{i\sigma} \epsilon_{i\sigma} \left(\langle \phi_{i\sigma} | \, \phi_{i\sigma} \rangle - 1 \right) = 2\sum_{i\sigma} f_{i\sigma} \left\langle \frac{\partial \phi_{i\sigma}}{\partial \mathbf{R}_\alpha} \middle| \epsilon_{i\sigma} \middle| \phi_{i\sigma} \right\rangle . \tag{6.36}$$

6.1.2. The Atomic Forces in the Case of LSD

Summing up the above derivative contributions Eqs. (6.15), (6.15), (6.25), (6.27), (6.32), (6.33), (6.35) and (6.36) finally results in a total force acting on atom α

$$\mathbf{F}_\alpha = \mathbf{F}_{\mathrm{HF},\alpha} + \mathbf{F}_{\mathrm{Pulay},\alpha} + \mathbf{F}_{\mathrm{MP},\alpha} . \tag{6.37}$$

These so-called *Hellman-Feynman* forces $\mathbf{F}_\alpha$ correspond to the classical forces by embedding each nucleus into the field of the electronic charge density and of all other nuclei [94, 95]

$$\mathbf{F}_{\mathrm{HF},\alpha} = Z_\alpha \sum_{\beta \neq \alpha} Z_\beta \frac{\mathbf{R}_\alpha - \mathbf{R}_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|^3} - Z_\alpha \int \mathrm{d}\mathbf{r} \rho(\mathbf{r}) \frac{\mathbf{R}_\alpha - \mathbf{r}}{|\mathbf{R}_\alpha - \mathbf{r}|^3} . \tag{6.38}$$

The *Pulay* forces $\mathbf{F}_{\mathrm{Pulay},\alpha}$ [96] result from the dependency of the basis functions on the nuclear coordinates and are given by

$$\mathbf{F}_{\mathrm{Pulay},\alpha} = -2\sum_{i\sigma}\left\langle \frac{\partial\phi_{i\sigma}}{\partial\mathbf{R}_\alpha}\middle|\,\hat{h}^{\mathrm{KS}} - \epsilon_{i\sigma}\middle|\,\phi_{i\sigma}\right\rangle \,, \tag{6.39}$$

This term vanishes in case of a complete Hilbert space spanned by the basis functions, because then the approximated eigenvalues $\epsilon_{i\sigma}$ would be the exact eigenvalues of $\hat{h}^{\mathrm{KS}}$ and $(\hat{h}^{\mathrm{KS}} - \epsilon_{i\sigma})\,|\phi_{i\sigma}\rangle$ would therewith vanish. In practice, however, this is basically never the case and for reasonable basis set sizes, this term is mandatory. Furthermore, in case of basis sets which do not depend upon the nuclear coordinates, like e.g. plane waves, this term would vanish as well because then the nuclear gradients of the Kohn-Sham orbitals are zero.

6.1.3. The GGA-Correction Term

In the GGA-functional suggested by Perdew, Burke and Ernzerhof, the exchange-correlation functional possesses the following functional form

$$E_{\mathrm{XC}} = \int \mathrm{d}\mathbf{r}\rho(\mathbf{r})\epsilon_{\mathrm{XC}}\left(\rho^{\uparrow},\rho^{\downarrow},|\nabla\rho|^2\right) \,. \tag{6.40}$$

Taking the derivative thus yields an additional term that arises due to the square of the charge density gradient

$$\frac{\partial E_{\mathrm{XC}}}{\partial\mathbf{R}_\alpha} = \frac{\partial E_{\mathrm{XC,LSD}}}{\partial\mathbf{R}_\alpha} + \int \mathrm{d}\mathbf{r}\rho(\mathbf{r})\frac{\partial\epsilon_{\mathrm{XC}}}{\partial\,|\nabla\rho|^2}\frac{\partial\,|\nabla\rho|^2}{\partial\mathbf{R}_\alpha} \,, \tag{6.41}$$

so another contribution to the overall forces is given by

$$\mathbf{F}_{\mathrm{GGA}} = -\int \mathrm{d}\mathbf{r}\rho(\mathbf{r})\frac{\partial\epsilon_{\mathrm{XC}}}{\partial\,|\nabla\rho|^2}\frac{\partial\,|\nabla\rho|^2}{\partial\mathbf{R}_\alpha} \,. \tag{6.42}$$

The calculation of the nuclear gradient of the square of the density gradient can be further broken down to

$$\frac{\partial\,|\nabla\rho|^2}{\partial\mathbf{R}_\alpha} = 2\,(\nabla\rho)^{\mathrm{T}}\,\frac{\partial\nabla\rho}{\partial\mathbf{R}_\alpha} \,, \tag{6.43}$$

Since $\nabla\rho(\mathbf{r}) = 2\sum_{i\sigma}\nabla\phi_{i\sigma}(\mathbf{r})\cdot\phi_{i\sigma}(\mathbf{r})$, the calculation of the matrix $\partial\nabla\rho(\mathbf{r})/\partial\mathbf{R}_\alpha$ yields

$$\frac{\partial\nabla\rho(\mathbf{r})}{\partial\mathbf{R}_\alpha} = 2\sum_{i\sigma}\left(\frac{\partial\nabla\phi_{i\sigma}(\mathbf{r})}{\partial\mathbf{R}_\alpha}\cdot\phi_{i\sigma}(\mathbf{r}) + \nabla\phi_{i\sigma}(\mathbf{r})\cdot\left(\frac{\partial\phi_{i\sigma}(\mathbf{r})}{\partial\mathbf{R}_\alpha}\right)^{\mathrm{T}}\right) \,, \tag{6.44}$$

where the second term corresponds to an outer product between two vectors thus giving a matrix. Analogous to the nuclear gradients of the orbitals that can be expressed in terms of the spatial gradients of the basis functions, the derivatives of the spatial gradient of the Kohn-Sham orbitals $\partial\nabla\phi_{i\sigma}(\mathbf{r})/\partial\mathbf{R}_\alpha$ can be boiled down to the spatial Hessians of the basis functions $\nabla^2\phi_j$.

6.1.4. The Atomic Forces in Connection with Smearing Methods

If an electronic smearing method is used to accelerate the convergence of the self-consistency cycle, the Lagrange functional is extended to another constraint to conserve the number of electrons

$$E_{\mathrm{tot}} = \min_{\{c_{i\sigma,j}\}} \left(E^{\mathrm{KS}} - \sum_{i\sigma} f_{i\sigma}\epsilon_{i\sigma} \left(\langle \phi_{i\sigma} \mid \phi_{i\sigma} \rangle - 1 \right) - \mu \left(\sum_{i\sigma} f_{i\sigma} - N_{\mathrm{el}} \right) \right) \tag{6.45}$$

$$= E_{\mathrm{tot}} \left[\{c_{i\sigma,j}\}, \{\mathbf{R}_\alpha\}, \{f_{i\sigma}\} \right] , \tag{6.46}$$

with μ being the chemical potential. It can be shown that the total energy is then not a variational quantity with respect to the partial occupancies [75]. Hence, the total derivative of the total energy does not reduce to the partial derivative as in the case without electronic smearing

$$\mathbf{F}_\alpha = -\frac{\mathrm{d}E_{\mathrm{tot}}}{\mathrm{d}\mathbf{R}_\alpha} \tag{6.47}$$

$$= -\frac{\partial E_{\mathrm{tot}}}{\partial \mathbf{R}_\alpha} - \sum_{i\sigma,j} \underbrace{\frac{\partial E_{\mathrm{tot}}}{\partial c_{i\sigma,j}}}_{=0} \frac{\partial c_{i\sigma,j}}{\partial \mathbf{R}_\alpha} - \sum_{i\sigma} \underbrace{\frac{\partial E_{\mathrm{tot}}}{\partial f_{i\sigma}}}_{\neq 0} \frac{\partial f_{i\sigma}}{\partial \mathbf{R}_\alpha} . \tag{6.48}$$

To obtain a variational quantity, one needs to consider the electronic *free energy* F

$$F = E - \sum_{i\sigma} wS(f_{i\sigma}) , \tag{6.49}$$

with a corresponding entropy term S. Using Gaussian smearing, the entropy is e.g. given by

$$S = \frac{1}{2\sqrt{\pi}} \exp\left(- \left(\frac{\epsilon_{i\sigma-\mu}}{w} \right)^2 \right) . \tag{6.50}$$

It can be shown that the free energy is therewith variational [75]. Since the entropy correction does not depend upon the variational parameters $\{c_{i\sigma,j}\}$, the corresponding partial derivatives of the free and total energy are identical. Hence, it follows that

$$\mathbf{F}_\alpha = -\frac{\partial F}{\partial \mathbf{R}_\alpha} . \tag{6.51}$$

As a result, the atomic forces are no longer consistent with the total energy but with the free energy which therefore needs to be used for a local structural relaxation scheme.

6.1.5. Grid Effects

The integrals to calculate the total energy and therewith the atomic forces are not calculated exactly, but approximated by a discrete sum according to Eq. (5.30)

$$I = \int \mathrm{d}\mathbf{r} f(\mathbf{r}, \{\mathbf{R}_\alpha\}) \tag{6.52}$$

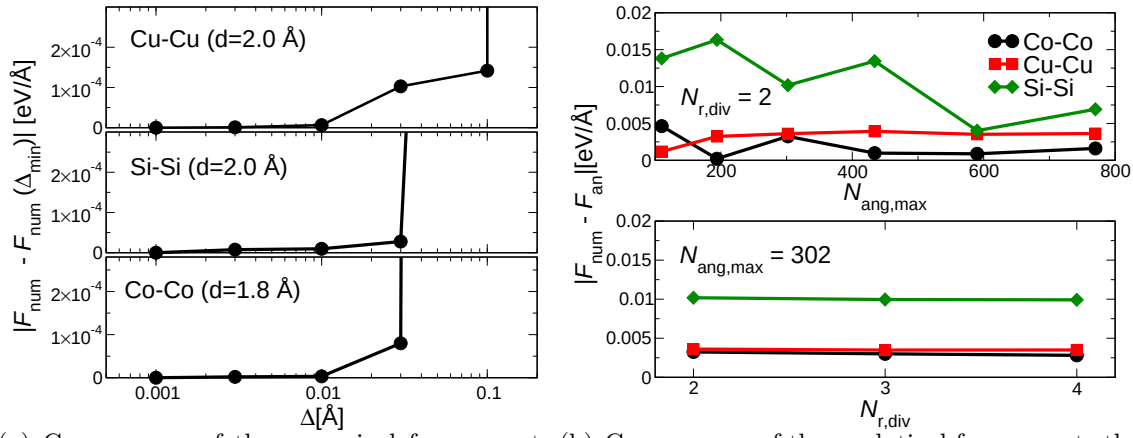
$$\approx \sum_\alpha \sum_{\mathbf{r}_{i\alpha}} w_{\mathbf{r}_{i\alpha}}(\mathbf{R}_\alpha) f(\mathbf{r}_{i\alpha}, \{\mathbf{R}_\alpha\}) \ . \tag{6.53}$$

Since the integration weights depend upon the atomic coordinates, they need to be taken into account explicitly for an exact derivative of the numerically evaluated total energy

$$\frac{\partial I}{\partial \mathbf{R}_\alpha} = \sum_\alpha \sum_{\mathbf{r}_{i,\alpha}} w_{\mathbf{r}_{i,\alpha}} \frac{\partial f}{\partial \mathbf{R}_\alpha} + \sum_\alpha \sum_{\mathbf{r}_{i,\alpha}} \frac{\partial w_{\mathbf{r}_{i,\alpha}}}{\partial \mathbf{R}_\alpha} f \ . \tag{6.54}$$

Hence, each integration term in the forces yields an additional correction term to capture the grid derivatives. This term can in principle be evaluated as the integration weights are given by a simple analytical formula. The actual calculation represents, however, a computational burden. Fortunately, grid effects are known to be negligible for grid densities used in practice [85] and turn out to play no role for the cases of the present work as well (see section 6.2).

6.2. Consistency of the Forces



(a) Convergence of the numerical forces w.r.t. the step width Δ instancing Cu_2, Si_2 and Co_2 for non-equilibrium bond-distances. Depicted are the differences in the numerical forces with respect to the smallest displacement applied $\Delta_{\min} = 10^{-3}$ Å .

(b) Convergence of the analytical forces w.r.t. the integration grid instancing Cu_2, Si_2 and Co_2 for non-equilibrium bond-distances. Depicted are the differences between the analytical and numerical forces obtained for $\Delta_{\min} = 10^{-2}$ Å. The upper panel presents the convergence w.r.t. $N_{\mathrm{ang,max}}$ with fixed $N_{\mathrm{r,div}}$. In the lower panel, $N_{\mathrm{ang,max}}$ is kept fixed to check the convergence w.r.t. the $N_{\mathrm{r,div}}$.

Figure 6.1.: Accuracy test of the analytical forces.

We illustrate the accuracy of the analytical atomic forces implemented in FHI-aims by a comparison with numerical forces obtained by the finite difference scheme for some

non-equilibrium dimers as test cases. Since the numerical forces can in principle be evaluated up to any desired accuracy by choosing a sufficiently small step width, it serves as a reliable reference value for the analytical forces. Caution is advised when converging the numerical force with respect to Δ, though. Since the integration is done on a discrete grid, a too small displacement might result in noise, thus yielding an unreasonable numerical force. Fig. 6.1(a) presents the convergence test for Δ for some non-equilibrium dimers of the species treated in the present work using a converged tier2 basis set. The numerical forces are hereby obtained by a finite difference scheme of sixth order (see Appendix D). The results clearly show that for all test cases, the numerical force is converged within $O(10^{-5})$ eV/Å at a step width of Δ= 0.01 Å, thus providing a sufficiently accurate reference value. With the optimized step width, the convergence of the analytical force with respect to the integration grid is then investigated. The results are presented in Fig. 6.1(b). For Cu_2 and Co_2, the accuracy is within $O(10^{-3})$ eV/Å and seems to be independent of the chosen integration grid. Hence in these cases, grid effects seem to play no role. In the case of Si_2, the agreement between analytical and numerical force is slightly worse and seems to be improved by augmenting the grid, indicating small grid effects. However, the difference is at most 10^{-2} eV/Å for even the smallest integration grid and therewith negligible. Since a typical force convergence criterium for a local relaxation scheme is 10^{-2} eV/Å, such an accuracy is sufficient. Figure 6.2 illustrates this at a local relaxation of the ground-state isomers of Si_7, Si_{10} and Cu_7. In all three cases, the change of the atomic configuration was only in the order of 10^{-3} Å at that force convergence level and therewith sufficiently converged.

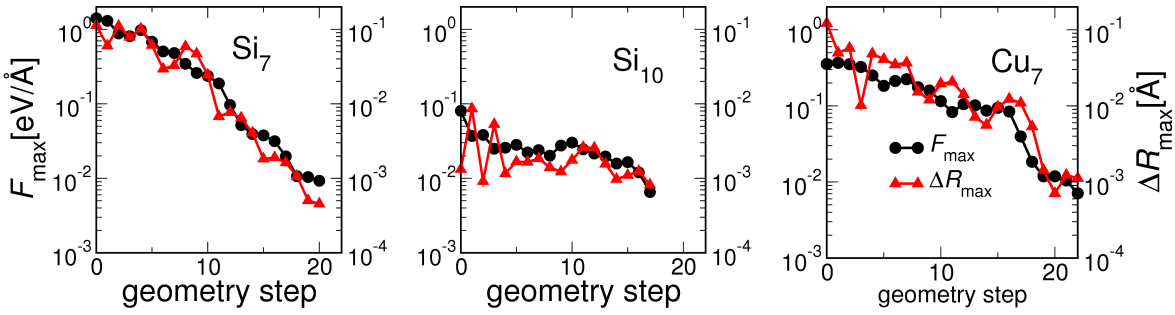


Figure 6.2.: Local relaxations for the ground-state isomers of Si_7, Si_{10} and Cu_7 with BFGS and PBE-DFT. Plottet are the maximum force component $F_{\max}$ and maximum displacement $\Delta R_{\max}$ of the atomic geometry vs. the geometry step.

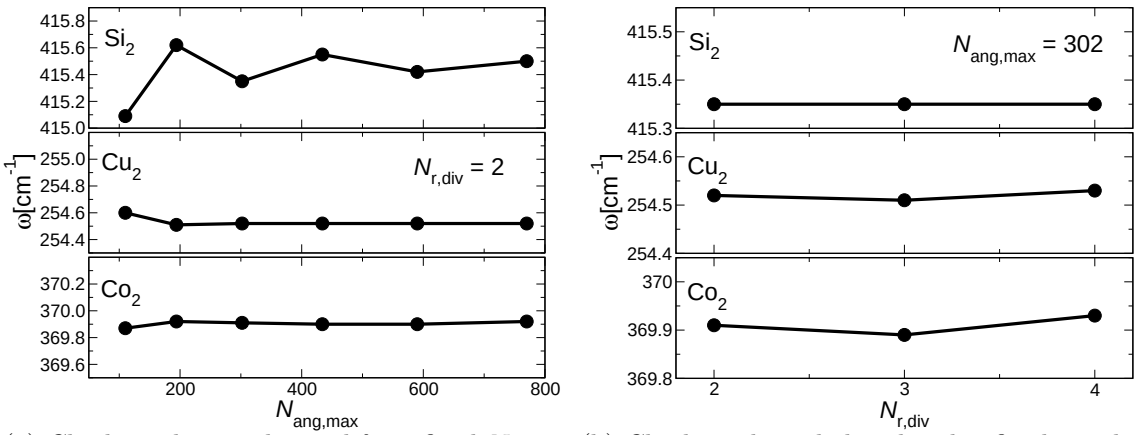
6.3. The Second Energy Derivative

The vibrational analysis, in particular the calculation of the IR-spectra in chapter III, necessitates the knowledge of the Hessian matrix, i.e. the second energy derivative with respect to the atomic positions. The Hessian matrix of the total energy can be obtained by taking the total derivative of the forces. Contrary to the total energy, the

atomic forces are not a variational quantity, though. Hence the total derivative does not reduce to the partial derivative, thus making a direct evaluation of the analytical Hessian impossible

$$\frac{\mathrm{d}E_{\mathrm{tot}}}{\mathrm{d}\mathbf{R}_\alpha \mathrm{d}\mathbf{R}_\beta} = -\frac{\mathrm{d}\mathbf{F}_\alpha}{\mathrm{d}\mathbf{R}_\beta} \tag{6.55}$$

$$= -\frac{\partial \mathbf{F}_\alpha}{\partial \mathbf{R}_\beta} - \sum_{i\sigma,j} \underbrace{\frac{\partial \mathbf{F}_\alpha}{\partial c_{i\sigma,j}}}_{\neq 0} \frac{\partial c_{i\sigma,j}}{\partial \mathbf{R}_\beta} \,. \tag{6.56}$$



(a) Checking the angular grid for a fixed $N_{\mathrm{r,div}}$.

(b) Checking the radial grid with a fixed angular grid.

Figure 6.3.: Convergence of the vibrational frequency w.r.t. to the integration grid instancing Cu_2, Si_2 and Co_2.

In the present work the Hessian is therefore obtained by taking numerical derivatives of the analytical forces. Since the vibrational analysis is only a post-processing step done at the local minimum, the additional computational burden due to the displacements remains small for the cluster sizes considered in the present work. Fig. 6.4 shows the vibrational frequencies for some dimers at their equilibrium bond distance with respect to different finite displacements Δ using a converged basis set tier2. Like in the case of numerical forces, the frequencies can clearly be converged

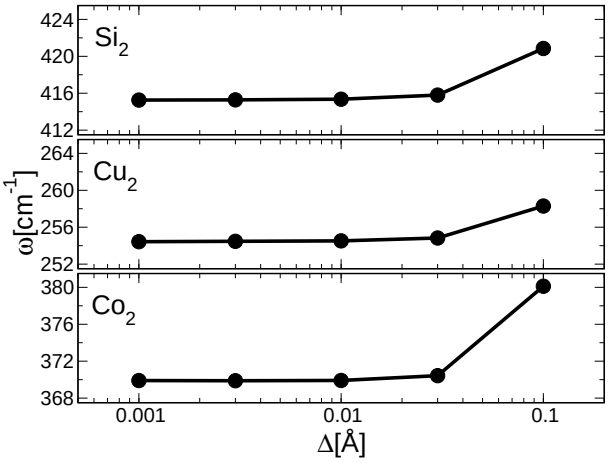


Figure 6.4.: Convergence of the vibrational frequency w.r.t. the step width Δ instancing Cu_2, Si_2 and Co_2.

with respect to Δ in all cases and grid effects play no role. With a displacement of Δ= 0.01 Å, the convergence with respect to the integration grid was explicitly checked and summarized in Fig. 6.3. In all cases, even for the silicon dimer with a considerably higher inaccuracy in the atomic forces, the vibrational frequencies are converged below 1 cm^{-1} and are therefore more than accurate enough for the requirements of the ensuing applications. It is worth to point out that the high accuracy of the atomic forces and the vibrational frequencies shown above of course only indicates a high consistency between the calculated derivatives and the underlying energy surface. If the latter describes the system poorly, the forces and vibrational frequencies are equally unphysical.